



**HAL**  
open science

# Modelling transcriptional regulation by DNA supercoiling in bacteria

Raphaël Forquet

► **To cite this version:**

Raphaël Forquet. Modelling transcriptional regulation by DNA supercoiling in bacteria. Molecular biology. Université de Lyon, 2021. English. NNT : 2021LYSEI102 . tel-03675197

**HAL Id: tel-03675197**

**<https://theses.hal.science/tel-03675197v1>**

Submitted on 23 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N d'ordre NNT : 2021LYSEI102

## THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON

opérée au sein de

**L'institut national des sciences appliquées de Lyon**

**École doctorale N°341:**

**Évolution, Écosystèmes, Microbiologie, Modélisation**

**Spécialité / discipline de doctorat:**

**Biomath-Bioinfo-Génomique évolutive**

Soutenue publiquement le 17/12/2021, par:

**Raphaël Forquet**

---

# Modelling transcriptional regulation by DNA supercoiling in bacteria

---

Devant le jury composé de:

Junier, Ivan	Chargé de recherche	TIMC	Rapporteur
Strick, Terence	Professeur	IBENS	Rapporteur
Muskhelishvili, Georgi	Professeur	AUG	Examineur
Reverchon, Sylvie	Professeure	INSA Lyon	Directrice de thèse
Meyer, Sam	Maître de Conférences	INSA Lyon	Co-directeur de thèse

## Département FEDORA – INSA Lyon - Ecoles Doctorales

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
<b>CHIMIE</b>	<b><u>CHIMIE DE LYON</u></b> <a href="https://www.edchimie-lyon.fr">https://www.edchimie-lyon.fr</a> Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage secretariat@edchimie-lyon.fr	<b>M. Stéphane DANIELE</b> C2P2-CPE LYON-UMR 5265 Bâtiment F308, BP 2077 43 Boulevard du 11 novembre 1918 69616 Villeurbanne <a href="mailto:directeur@edchimie-lyon.fr">directeur@edchimie-lyon.fr</a>
<b>E.E.A.</b>	<b><u>ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE</u></b> <a href="https://edeea.universite-lyon.fr">https://edeea.universite-lyon.fr</a> Sec. : Stéphanie CAUVIN Bâtiment Direction INSA Lyon Tél : 04.72.43.71.70 secretariat.edeea@insa-lyon.fr	<b>M. Philippe DELACHARTRE</b> INSA LYON Laboratoire CREATIS Bâtiment Blaise Pascal, 7 avenue Jean Capelle 69621 Villeurbanne CEDEX Tél : 04.72.43.88.63 <a href="mailto:philippe.delachartre@insa-lyon.fr">philippe.delachartre@insa-lyon.fr</a>
<b>E2M2</b>	<b><u>ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION</u></b> <a href="http://e2m2.universite-lyon.fr">http://e2m2.universite-lyon.fr</a> Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.e2m2@univ-lyon1.fr	<b>M. Philippe NORMAND</b> Université Claude Bernard Lyon 1 UMR 5557 Lab. d'Ecologie Microbienne Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69 622 Villeurbanne CEDEX <a href="mailto:philippe.normand@univ-lyon1.fr">philippe.normand@univ-lyon1.fr</a>
<b>EDISS</b>	<b><u>INTERDISCIPLINAIRE SCIENCES-SANTÉ</u></b> <a href="http://ediss.universite-lyon.fr">http://ediss.universite-lyon.fr</a> Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.ediss@univ-lyon1.fr	<b>Mme Sylvie RICARD-BLUM</b> Institut de Chimie et Biochimie Moléculaires et Supramoléculaires (ICBMS) - UMR 5246 CNRS - Université Lyon 1 Bâtiment Raulin - 2ème étage Nord 43 Boulevard du 11 novembre 1918 69622 Villeurbanne Cedex Tél : +33(0)4 72 44 82 32 <a href="mailto:sylvie.ricard-blum@univ-lyon1.fr">sylvie.ricard-blum@univ-lyon1.fr</a>
<b>INFOMATHS</b>	<b><u>INFORMATIQUE ET MATHÉMATIQUES</u></b> <a href="http://edinfomaths.universite-lyon.fr">http://edinfomaths.universite-lyon.fr</a> Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage Tél : 04.72.43.80.46 infomaths@univ-lyon1.fr	<b>M. Hamamache KHEDDOUCI</b> Université Claude Bernard Lyon 1 Bât. Nautibus 43, Boulevard du 11 novembre 1918 69 622 Villeurbanne Cedex France Tél : 04.72.44.83.69 <a href="mailto:hamamache.kheddouci@univ-lyon1.fr">hamamache.kheddouci@univ-lyon1.fr</a>
<b>Matériaux</b>	<b><u>MATÉRIAUX DE LYON</u></b> <a href="http://ed34.universite-lyon.fr">http://ed34.universite-lyon.fr</a> Sec. : Yann DE ORDENANA Tél : 04.72.18.62.44 yann.de-ordenana@ec-lyon.fr	<b>M. Stéphane BENAYOUN</b> Ecole Centrale de Lyon Laboratoire LTDS 36 avenue Guy de Collongue 69134 Ecully CEDEX Tél : 04.72.18.64.37 <a href="mailto:stephane.benayoun@ec-lyon.fr">stephane.benayoun@ec-lyon.fr</a>
<b>MEGA</b>	<b><u>MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE</u></b> <a href="http://edmega.universite-lyon.fr">http://edmega.universite-lyon.fr</a> Sec. : Stéphanie CAUVIN Tél : 04.72.43.71.70 Bâtiment Direction INSA Lyon mega@insa-lyon.fr	<b>M. Jocelyn BONJOUR</b> INSA Lyon Laboratoire CETHIL Bâtiment Sadi-Carnot 9, rue de la Physique 69621 Villeurbanne CEDEX <a href="mailto:jocelyn.bonjour@insa-lyon.fr">jocelyn.bonjour@insa-lyon.fr</a>
<b>ScSo</b>	<b><u>ScSo*</u></b> <a href="https://edsciencesociales.universite-lyon.fr">https://edsciencesociales.universite-lyon.fr</a> Sec. : Mélina FAVETON INSA : J.Y. TOUSSAINT Tél : 04.78.69.77.79 melina.faveton@univ-lyon2.fr	<b>M. Christian MONTES</b> Université Lumière Lyon 2 86 Rue Pasteur 69365 Lyon CEDEX 07 <a href="mailto:christian.montes@univ-lyon2.fr">christian.montes@univ-lyon2.fr</a>

\*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

*À mes parents...*

*To my parents...*

*“Le doute est père de la création.”*

Galilée

*“Doubt is the father of invention.”*

Galileo Galilei





## *Acknowledgments / Remerciements*

J'exprime tout d'abord ma gratitude à l'ensemble du jury pour avoir accepté d'évaluer cette thèse, et tout particulièrement Ivan Junier et Terence Strick pour avoir endossé le rôle chronophage et énergivore de rapporteur. Merci à Georgi Muskhelishvili, examinateur, pour la qualité de nos échanges, son intérêt et sa bienveillance durant ces années.

Je remercie les membres de mon comité de suivi individuel: mon tuteur de thèse Eric Tannier, Nicolas Parisot, Stéphan Lacroix et Olivier Espeli, pour avoir apporté un regard critique sur mes travaux chaque année, ainsi que pour leurs précieux conseils et encouragements.

J'adresse toute ma reconnaissance à mes directeurs de thèse, Sylvie Reverchon et Sam Meyer, pour m'avoir encadré avec bienveillance et sérieux durant toutes ces années, et sans qui rien de tout cela n'aurait été possible. Un grand merci pour votre disponibilité, votre patience, votre soutien dans les périodes difficiles, et pour toutes les compétences que vous m'avez transmises tout au long de cette aventure.

Je remercie l'ensemble de l'UMR MAP et plus particulièrement l'équipe CRP pour leur accueil chaleureux, ainsi que pour la qualité des échanges scientifiques. Je pense en particulier à William Nasser qui, malgré son poste de directeur d'unité, se rend toujours disponible pour aider à la résolution de problème et discuter science. Mais aussi à Florence Hommais que j'ai toujours eu plaisir à visiter (parasiter ?) durant mes journées en présentiel, pour sa bonne humeur et ses bons conseils, tout comme Zahar Haichar. Un grand merci à Jess, Camille (Tic & Tac), Amandine et Maiwenn, à la fois pour vos conseils à la paillasse (un bioinformaticien qui fait des manip, c'est pas toujours ça ?), votre dévouement à l'équipe, mais surtout pour vos qualités humaines, votre gaieté et vos plaisanteries pendant tous les repas, pauses café et autres partagés ensemble. Merci aux doctorants de l'équipe (Shiny, Diana, Théophile et Maiwenn), pour ces moments d'échanges, de soutien, de réflexion (de dépression ?) et de rire liés à la thèse, ses hauts et ses bas, et au fait que nous avons tous été dans le même bateau. Enfin, merci à Véro pour ta jovialité, Salima pour ta gentillesse et ta patience dans les démarches administratives, mais aussi à l'équipe AME (Maxime, Alvaro, Antonio, Elodie, Philippe...), nos voisins de palier, et aux différents stagiaires (Mathieu aka Atlantos, Bilel, Nicolas & Maxime que j'ai eu plaisir à encadrer, Gabin, Rémi,

Sarah...) qui ont contribué à créer une ambiance conviviale et chaleureuse au sein du laboratoire.

Enfin, car la vie doit bien continuer en dehors de la thèse, je remercie mes amis et ma famille, qui ont, sans nul doute, contribué indirectement à la réussite de ce travail de part le soutien et le bonheur qu'ils m'ont apporté. Dav (la quali), on en a fait du chemin ensemble, 7 ans de solide amitié et toujours fidèle au poste, dans l'intelligence comme dans la bêtise la plus profonde. Deh (toi et moi on a nul besoin de mots pour se comprendre), et Alix (le gang des pigeons), d'abord voisins d'évasion du confinement, puis très vite fusionnels et amis proches, cette période Covid infernale aura eu pour (seul ?) avantage d'avoir fait que nos chemins se croisent. Paulo et Antoine (les bagarreurs), merci pour tous ces moments conviviaux partagés au club de boxe et à l'extérieur. Mais aussi Marine & Ugo (les cocos, la bringue et la faille), Lola & Clément (nos virées sauvages, nos parties d'échecs et la moto), Milena, Alolo, et Mathou (mes bobos), Amandine et Maïwenn (les copines). Cha, qui malgré la situation actuelle, a partagé ma vie et mon quotidien tout au long de cette aventure, et a toujours su me soutenir et m'aimer malgré des périodes difficiles. Merci d'avoir été et d'être une personne exceptionnelle. Harold et Pascale, merci pour votre gentillesse et votre affection, et tous ces bons moments partagés ensemble. Enfin, je ne remercierai jamais assez ma famille pour qui ces quelques mots ne suffiront pas tant vous êtes des piliers dans ma vie, notamment mes frangins Vinc & Flo (les J les S), rien ne pourra jamais se mettre en travers de nos relations tant on on se connaît par coeur, on rit (les challengers de Colgan), on pleure, on crie (les SSB qui dégènèrent), on s'amuse (la station, la bagarre, la chacha ?) et on s'apaise (mention spéciale pour votre écoute ces derniers mois) depuis qu'on est gamins. Audrey, pour la station intergalactique, la geekerie et les débats sans fin, la bonne bouffe et les apéros qui partent en Y à Mouans (quoi on the beach ?). Mamish, pour avoir toujours été aux petits soins et cultivé mon âme d'enfant. Parrain, pour tes conseils et ton soutien en cette période de rédaction. Et le reste de toute la famille bien nombreuse pour tous ces étés et Noël (allo c'est quiii?) endiablés partagés ensemble. Enfin, Mam et Padre, au-delà de toutes les valeurs que vous m'avez inculquées, d'un soutien inégalable même quand vous n'approuviez pas, vous m'avez toujours donné les moyens de mener à bien mes projets personnels et professionnels. A vous mes parents je dédie donc cette thèse, même si aucun hommage ne pourrait être à la hauteur de l'amour que vous me portez.

# Résumé: modélisation de la régulation transcriptionnelle des gènes par le surenroulement de l'ADN chez les bactéries

Au cours de leur vie, les bactéries font face à des changements environnementaux, auxquels elles répondent en modifiant rapidement et globalement l'expression de leurs gènes. Pour ce faire, elles ont développé une variété de mécanismes de régulation intervenant à différents niveaux, en particulier durant l'étape d'initiation de la transcription. Les modèles de régulation classiques sont centrés sur les facteurs de transcription, qui reconnaissent des séquences spécifiques dans le promoteur des gènes. A plus grande échelle, cette régulation passe par des variations du surenroulement, qui est le produit du stress torsionnel subi par la double-hélice durant les transactions d'ADN ayant pour conséquence un changement d'hélicité et de vrillage. Le niveau de surenroulement est finement contrôlé par les topoisomérases, des enzymes essentielles et hautement conservées. Chez les bactéries, l'ADN est maintenu dans un état sous-enroulé (surenroulement négatif) par l'activité forte de l'ADN gyrase, tandis que la topoisomerase I (et IV dans une moindre mesure) relâchent l'ADN, c'est-à-dire enlèvent le surenroulement. Ce niveau varie néanmoins en réponse à une variété de facteurs, notamment aux stress environnementaux. En retour, ces variations sont associées à une réponse transcriptionnelle globale et complexe des gènes, comme le démontrent un nombre croissant d'études transcriptomiques utilisant des inhibiteurs de l'ADN gyrase pour induire une relaxation du chromosome. Ceci est observé non seulement chez la bactérie modèle *E. coli*, mais également dans d'autres bactéries très différentes en termes de phylogénie et mode de vie. Ainsi, des variations de surenroulement permettraient une reprogrammation rapide et globale de l'expression des gènes, par exemple en réponse à un stress environnemental. Et pourtant, la régulation transcriptionnelle des gènes par le surenroulement ne fait l'objet d'aucun modèle de régulation quantitatif à l'échelle d'un génome entier. Cela peut s'expliquer par le fait que le surenroulement affecte la transcription à plusieurs étapes du procédé via différents mécanismes, en particulier durant l'initiation, en modulant non seulement la fixation des facteurs de transcription sur l'ADN, mais également l'interaction ARN Polymérase-promoteur.

L'objectif de ma thèse est de caractériser quantitativement les mécanismes de réponse des gènes aux variations de surenroulement, et en particulier

d'identifier les éléments promoteur déterminants dans la sensibilité des gènes au surenroulement, en se focalisant sur l'interaction ARN Polymerase-ADN. Et ce, en combinant (i) l'obtention et l'analyse des données transcriptomiques disponibles fournissant la réponse des gènes à une relaxation du chromosome dans des bactéries phylogénétiquement distantes (ii) des études cinétiques de transcription sur promoteurs mutés, et (iii) une approche de modélisation thermodynamique de la transcription à l'échelle du génome entier.

Dans un premier temps, une base de données est construite dans l'objectif de rassembler les données nécessaires à la validation des modèles dans le plus grand nombre de bactéries possible, les mécanismes étudiés ayant une portée potentiellement large. Il s'agit notamment de la réponse transcriptomique à une relaxation globale du chromosome induite avec des inhibiteurs de gyrase, et la cartographie des débuts de transcription et des promoteurs associés. Des outils de programmation Python sont ensuite implémentés pour effectuer les analyses, permettant également la contribution à des projets secondaires.

Dans un second temps, une analyse statistique des données transcriptomiques disponibles est conduite dans la bactérie pathogène des plantes *D. dadantii*. Le but étant de caractériser ses unités de transcription ainsi que ses promoteurs, (i) pour la validation des modèles de régulation par le surenroulement, et (ii) pour contribuer à des progrès dans le domaine de la phytopathogénéicité, en cartographiant le transcriptome de ce modèle bactérien largement utilisé. En effet, *D. dadantii* est avantageuse car une quantité significative de données a été accumulée sur l'effet du surenroulement, en particulier durant le processus infectieux où il joue un rôle régulateur central en servant de senseur aux changements environnementaux rencontrés dans la plante.

Enfin, nous implémentons à l'échelle du génome entier deux modèles de régulation indépendants décrivant l'effet du surenroulement sur l'interaction ARN Polymerase-promoteur durant l'étape d'initiation de la transcription, indépendamment de la présence éventuelle d'autres régulateurs. En modifiant drastiquement l'énergie libre d'ouverture de la double-hélice, le premier modèle démontre en quoi des variations globales du niveau de surenroulement permettent d'activer sélectivement les promoteurs en fonction du contenu en G/C de leur discriminateur. En modifiant l'orientation des éléments -35 et -10 pour la fixation de l'ARN Polymerase, le deuxième modèle quantifie la contribution de la taille du spacer des promoteurs dans leur réponse au surenroulement. Ces modèles thermodynamiques sont basés sur

les propriétés physicochimiques connues de l'ADN et ne nécessitent pratiquement aucun paramètre ajustable. Les prédictions sont validées dans un premier temps en mesurant la réponse de promoteurs mutants à des variations de surenroulement. En étendant la validation des modèles à l'échelle du génome entier de bactéries différentes en termes de phylogénie et mode de vie sur la base des données transcriptomiques disponibles, nos résultats suggèrent que les mécanismes sous-jacents sont utilisés de manière universelle dans le royaume procaryote, en particulier en réponse à des changements environnementaux. Au-delà, ces travaux démontrent le caractère basal et ubiquitaire de la régulation par le surenroulement basée sur les propriétés fondamentales de l'ADN, et fournissent les premiers modèles permettant d'expliquer et de prédire quantitativement la réponse des gènes à des variations de surenroulement à l'échelle du génome entier, contribuant de manière significative à des progrès dans le domaine.

Mots-clés: Régulation transcriptionnelle, surenroulement de l'ADN, modélisation thermodynamique quantitative, discriminateur, spacer, phytopathogène, unité de transcription



## **Abstract: modelling transcriptional regulation by DNA supercoiling in bacteria**

Bacteria are exposed to environmental fluctuations, to which they respond by quick and global changes in gene expression. Usual models of transcriptional regulation are centred on transcription factors which recognise specific sequences in genes' promoters, but disregard the important role of DNA supercoiling (SC), an ubiquitous property of the double-helix resulting from torsional stress. SC acts as a global and ubiquitous regulator in response to environmental changes, as suggested by many recent transcriptomics studies. The objective of the present thesis is to develop quantitative models of this regulation mode, by identifying the promoter-sequence determinants of gene SC-sensitivity, based on RNA Polymerase-DNA interaction and independently from additional regulatory proteins. To this end, we combine (i) the analysis of available transcriptomic data under conditions of SC variations, (ii) transcription assays on mutant promoters, and (iii) a thermodynamic modelling of transcription at the genome-scale. We first characterise the transcriptome of *D. dadantii*, a phytopathogen in which extensive data have been accumulated regarding the role of SC during plant infection, defining its transcription units and promoters. We then present two models explaining how global SC variations can selectively activate/repress promoters, depending on (i) the G/C-content of their discriminator for SC-assisted promoter opening, and (ii) their spacer length for the SC-dependent orientation between -35 and -10 elements affecting RNA Polymerase binding. Transcription assays are conducted on mutant promoters, and quantitatively confirm the predictions of the models. The universality of these mechanisms is demonstrated by analysing transcriptomes of distant bacteria under conditions of SC variations. Altogether, these results show that SC, based on the fundamental properties of DNA, constitutes an ubiquitous regulation mode in the prokaryotic kingdom.

Keywords: Transcriptional regulation, DNA supercoiling, quantitative thermodynamic modeling, discriminator, spacer, phytopathogen, transcription unit





# Contents

<b>Acknowledgments / Remerciements</b>	<b>v</b>
<b>Résumé</b>	<b>vii</b>
<b>Abstract</b>	<b>xi</b>
<b>List of Figures</b>	<b>xv</b>
<b>List of Tables</b>	<b>xvii</b>
<b>List of Abbreviations</b>	<b>xix</b>
<b>1 Introduction and literature review</b>	<b>1</b>
1.1 Bacterial adaptation to environmental changes . . . . .	1
1.1.1 Environmental changes . . . . .	1
1.1.2 Bacterial genomes . . . . .	3
1.1.3 Bacterial transcription . . . . .	5
1.2 Regulation of transcription initiation . . . . .	8
1.2.1 Role of transcription factors . . . . .	8
1.2.2 Role of nucleoid-associated proteins . . . . .	12
1.2.3 Role of alternative $\sigma$ factors and small ligands . . . . .	17
1.2.4 Existing models of transcriptional regulation . . . . .	19
1.3 Transcriptional regulation by DNA supercoiling during environmental changes . . . . .	26
1.3.1 DNA supercoiling . . . . .	26
1.3.2 DNA topoisomerases . . . . .	30
1.3.3 DNA supercoiling measurement . . . . .	34
1.3.4 Review . . . . .	36
1.4 Thesis motivations and objectives . . . . .	46
<b>2 Development of an integrated database, implementation of programming tools and contribution to side projects</b>	<b>49</b>
2.1 Development of an integrated database . . . . .	49

2.2	Implementation of programming tools . . . . .	51
2.3	Side projects . . . . .	53
<b>3</b>	<b>Mapping the complex transcriptional landscape of the phytopathogenic bacterium <i>Dickeya dadantii</i></b>	<b>57</b>
3.1	Role of DNA supercoiling in <i>D. dadantii</i> . . . . .	57
3.2	Article . . . . .	59
<b>4</b>	<b>Role of the discriminator sequence in the supercoiling-sensitivity of bacterial promoters</b>	<b>89</b>
4.1	Inheritable increases of supercoiling during the long-term evolution experiment . . . . .	89
4.2	Transcription assays on mutant promoters . . . . .	92
4.3	Article . . . . .	94
<b>5</b>	<b>Quantitative contribution of the spacer length in the supercoiling-sensitivity of bacterial promoters</b>	<b>111</b>
5.1	Insights into seconeolitsine-induced Topoisomerase I inhibition in Gram-negative bacteria . . . . .	111
5.2	Article . . . . .	113
<b>6</b>	<b>Discussion and perspectives</b>	<b>141</b>
6.1	Future studies on <i>D. dadantii</i> transcriptome . . . . .	141
6.2	Limitations of our modelling and future directions . . . . .	143
6.3	Applications of our modelling in eukaryotes . . . . .	146
6.4	Applications of our modelling in archaea . . . . .	149
	<b>Bibliography</b>	<b>153</b>

# List of Figures

1.1	General features of bacterial genomes: example of the model bacterium <i>E. coli</i> K-12 MG1655. . . . .	4
1.2	Schematic representation of transcription in bacteria. . . . .	7
1.3	Regulation of promoters by transcription factors. . . . .	10
1.4	Chromosome organisation by nucleoid-associated proteins. . . . .	12
1.5	Structural and architectural properties of the main nucleoid-associated proteins. . . . .	15
1.6	Expression patterns of four nucleoid-associated proteins during the growth of <i>E. coli</i> . . . . .	16
1.7	Modelling of gene regulatory systems. . . . .	19
1.8	Statistical models of gene regulatory systems. . . . .	20
1.9	Differential equations-based models. . . . .	22
1.10	Boolean network-based models. . . . .	22
1.11	Thermodynamic modelling of a simple repression. . . . .	24
1.12	Twist and writhe in a supercoiled linear DNA molecule with constrained ends. . . . .	27
1.13	Different types of non-B DNA structures. . . . .	29
1.14	X-ray structures of the main bacterial topoisomerases. . . . .	31
1.15	Strand-passage mechanisms of class I and II topoisomerases. . . . .	32
1.16	Division of bacterial chromosomes into supercoiled topological domains. . . . .	33
1.17	Titration of SC density by sedimentation in ethidium bromide sucrose density gradients, and by one- or two-dimensional agarose gel electrophoresis. . . . .	35
2.1	Schematic representation of the database. . . . .	50
2.2	Schematic representation of the object oriented programming framework. . . . .	52
2.3	Impact of gene orientation on the distribution of SC generated by transcription. . . . .	54

3.1	Schematic representation of <i>Dickeya dadantii</i> infection process, and changes in chromosomal SC level in response to stress. . .	58
4.1	Principle of the long-term evolution experiment (LTEE). . . . .	91
4.2	Principle of the transcription assays performed to measure the response of mutant promoters to SC variations. . . . .	93
5.1	Induction of opposite SC variations using novobiocin and sec-neolitsine. . . . .	113
6.1	Theta-base online browser to investigate <i>B. thetaiotaomicron</i> transcriptomic features. . . . .	142
6.2	Co-transcriptional R-loop formation and biological roles. . . .	146
6.3	Common structural features between prokaryotic and RNAP II eukaryotic promoters, together with their SC-dependent opening profiles and expression levels. . . . .	148
6.4	Effects of heat and cold shock on plasmid topology in bacteria and in hyperthermophilic archaea. . . . .	151

# List of Tables

1.1	Bacterial genomes: number, geometry and size. . . . .	4
1.2	Main nucleoid-associated proteins (NAPs) of <i>E. coli</i> . . . . .	14
1.3	The seven $\sigma$ factors of <i>E. coli</i> . . . . .	17
1.4	Features of main bacterial topoisomerases. . . . .	30
2.1	Summary of the data used to investigate the link between promoter structure and transcriptomic response under SC variations. . . . .	56



# List of Abbreviations

<b>A</b>	<b>A</b> denine
<b>bp</b>	<b>B</b> ase pair(s)
<b>C</b>	<b>C</b> ytosine
<b>CODOs</b>	<b>C</b> oherent <b>D</b> omains of transcription
<b>DNA</b>	<b>D</b> eoxyribonucleic Acid
<b>dRNA-seq</b>	<b>D</b> ifferential RNA-seq
<b>FEDS</b>	<b>F</b> luorescent Evaluation of <b>D</b> N <b>S</b> upercoiling
<b>G</b>	<b>G</b> uanine
<b>GRN</b>	<b>G</b> ene <b>R</b> egulatory <b>N</b> etwork
<b>Inr</b>	<b>I</b> nitiator Element
<b>kb</b>	<b>k</b> ilobase(s)
<b>Lk</b>	<b>L</b> inking number
<b>LTTE</b>	<b>L</b> ong- <b>T</b> erm <b>E</b> volution <b>E</b> xperiment
<b>Mb</b>	<b>M</b> egabase(s)
<b>mRNA</b>	<b>M</b> essenger RNA
<b>NAP</b>	<b>N</b> ucleoid- <b>A</b> ssociated <b>P</b> rotein
<b>nm</b>	<b>N</b> anometer
<b>nt</b>	<b>N</b> ucleotide(s)
<b>OD</b>	<b>O</b> ptical <b>D</b> ensity
<b>ODE</b>	<b>O</b> rdinary <b>D</b> ifferential Equation
<b>PAI</b>	<b>P</b> athogenicity <b>I</b> sland
<b>PDE</b>	<b>P</b> artial <b>D</b> ifferential Equation
<b>PGA</b>	<b>P</b> olygalacturonic <b>A</b> cid
<b>RBS</b>	<b>R</b> ibosome- <b>B</b> inding <b>S</b> ite
<b>RNA</b>	<b>R</b> ibonucleic <b>A</b> cid
<b>RNAP</b>	<b>R</b> N <b>A</b> Polymerase
<b>RNA-seq</b>	<b>R</b> N <b>A</b> <b>S</b> equencing
<b>ROS</b>	<b>R</b> eactive <b>O</b> xygen <b>S</b> pecies
<b>RPKM</b>	<b>R</b> eads <b>P</b> er <b>K</b> ilobase <b>P</b> er <b>M</b> illion <b>M</b> apped <b>R</b> eads
<b>rRNA</b>	<b>R</b> ibosomal <b>R</b> N <b>A</b>
<b>SC</b>	<b>D</b> N <b>A</b> <b>S</b> upercoiling
<b>SD</b>	<b>S</b> hine- <b>D</b> algarno
<b>SMC</b>	<b>S</b> tructural <b>M</b> aintenance of <b>C</b> hromosome
<b>T</b>	<b>T</b> hymine
<b>TEX</b>	<b>T</b> erminator <b>E</b> xonuclease
<b>TF</b>	<b>T</b> ranscription <b>F</b> actor
<b>Topo</b>	<b>T</b> opoiso <b>m</b> erase
<b>tRNA</b>	<b>T</b> ransfer <b>R</b> N <b>A</b>
<b>TSC</b>	<b>T</b> ranscription- <b>S</b> upercoiling <b>C</b> oupling



<b>TSS</b>	<b>Transcription Start Site</b>
<b>TTS</b>	<b>Transcription Termination Site</b>
<b>TU</b>	<b>Transcription Unit</b>
<b>Tw</b>	<b>Twist</b>
<b>U</b>	<b>Uracil</b>
<b>Wr</b>	<b>Writhe</b>

# Chapter 1

## Introduction and literature review

### 1.1 Bacterial adaptation to environmental changes

#### 1.1.1 Environmental changes

Bacteria are unicellular, microscopic organisms, found among the earliest life forms on Earth [1], that played key roles in the history of its life. Billion years ago, during the the Great Oxygenation Event, a group of bacteria referred to as cyanobacteria, contributed to creating current Earth's atmosphere by performing photosynthesis, *i.e.* the production of nutrients and oxygen from carbon dioxide, water and sunlight [2]. Later on, at planetary time scales of billion years, the integration of bacteria into host cells led to the emergence of new intracellular components or organelles such as chloroplasts, which provided plants' photosynthetic capacity, or mitochondria, which brought the ability to generate the energy currency of animal and plant cells. According to this so-called endosymbiotic theory, all living plants and animals thus evolved from the inclusion of a bacterial cell inside another one [3]. Nowadays, bacterial activities are essential to the cycling of elements and consequently to life on Earth, *e.g.* by providing nutrients like nitrogen and phosphorus to plants, or by decomposing organic matter in soil and in oceans. They are also crucial for maintaining homeostasis in the human body (and in many other organisms), by living on and inside us by trillions, and contributing, *e.g.* to digestion and defence functions of the gastrointestinal tract, to repairing our skin, or tuning our immune system. Conversely, some bacteria are very destructive by causing severe infections, *e.g.* (in human) pneumonia, wound and bloodstream (sepsis) infections, sexually transmitted diseases, or (in plants) soft rot and ravages of crop fields. Despite their microscopic size imperceptible to human eyes, and apparent simplicity, bacteria

survive, thrive, fight with molecular spears and chemical weapons, swim with nanoscopic motors, but also sense, communicate, and remember.

Bacteria occupy a wide variety of habitats on Earth, including harsh environments, from deep-sea hydrothermal vents, to volcanic chimneys, including soil, Earth's crust, water, arctic ice and glaciers, interior or surface of plants, insects and animals. Their long evolutionary history exposed them to remarkably different environments, and developed their faculty to cope with many changing physicochemical conditions. The latter are of various types, including fluctuations in environmental factors such as temperature, oxygen, pressure, salt concentration, pH, light, and nutrient abundance or source. These are also of various timescales, including daily, weekly, monthly, seasonally variations and more. Some of them occur at a predictable frequency, whereas some others occur sporadically. Finally, environmental fluctuations can be gradual, or drastic. These characteristics are described in more detail elsewhere [4]. For the present thesis, environmental stresses deserve special attention. Although it has proven to be an elusive concept [5], here, inspired by Hoffmann and Parsons [6], we consider an environmental stress as an environmental factor (or stressor), causing sudden and potentially injurious changes in a biological system. In nature, many bacteria are exposed to constantly changing conditions, including stresses, and, for pathogens, switches from external environments to host-specific niches. For example, plant pathogenic bacteria of the *Dickeya* genus alternate between habitats such as soil or ground water, and surface/interior of plants [7]. In their host, these bacteria are exposed to many environmental stresses resulting partly from plant defences, such as exposition to antimicrobial agents, or abrupt changes in pH, osmotic pressure, oxygen availability, and nutrient sources [7]. In order to maintain cellular functions, but also for optimal energy management, *i.e.* to avoid wasteful consumption of resources and energy, bacteria must sense environmental changes, and adjust their physiology and metabolism to the new conditions. To this end, they have evolved a variety of responses, involving regulation of gene expression at different levels.

### 1.1.2 Bacterial genomes

A genome can be defined as the entire set of genetic information that produces and maintains an organism, and which manifests as DNA (Deoxyribonucleic Acid) composed of four different nucleotides - adenine (A), guanine (G), cytosine (C) and thymine (T) [8]. It consists of genes, plus non-coding sequences. Genes are segments of DNA directing the synthesis of RNA (Ribonucleic acid) composed of A, G, C and uracil (U) instead of T, that can be directly functional, or be the intermediate template for protein synthesis. The transcriptome is the set of all RNAs expressed by a genome, whereas the proteome refers to the full range of proteins produced by a genome. In contrast to eukaryotes wherein each cell's genome is stored within a nucleus, bacterial cells contain no inner membranes, and their genome is consequently located in a region of the cytoplasm called the nucleoid (see next section on nucleoid-associated proteins). The genome of bacteria usually consists of one circular DNA molecule (and, in some cases, more than one and/or linear DNA molecules, Tab. 1.1), even if they often harbour extragenomic DNA such as plasmids carrying accessory genes encoding, *e.g.* toxin-antitoxin systems, virulence factors, or peptides involved in antimicrobial resistance [9]. Bacterial genomes are small in length, typically under 10 megabases (Mb) (Tab. 1.1), and exhibit a high gene density, *i.e.* proportion of the genome that is composed of genes, with an average of 88% [10] (Fig. 1.1B). Base composition is highly variable across species, ranging from 13 to 75 G/C%, and genetic maps, *i.e.* gene order and content, remain fairly stable among related species [11]. These features contrast those of eukaryotic genomes, which are generally partitioned into multiple linear DNA molecules, and are of much larger size and complexity, from 125 Mb in the plant *Arabidopsis thaliana*, to 2.9 billion bases in human [12]. Unlike eukaryotes in which there is little association between genome size and gene density due to the abundance of non-coding DNA, in bacteria, gene number is tightly linked to genome size, and most genomic sequences correspond to protein-coding regions, with little non-coding DNA [13]. Finally, bacterial genomes are configured into domains related to replication and packaging (Fig. 1.1A).

Bacterium	Genome
<i>Escherichia coli</i>	One circular DNA molecule (4.6 Mb)
<i>Salmonella typhimurium</i>	One circular DNA molecule (4.7 Mb)
<i>Dickeya dadantii</i>	One circular DNA molecule (4.9 Mb)
<i>Mycoplasma pneumoniae</i>	One circular DNA molecule (0.8 Mb)
<i>Synechococcus</i> sp. PCC7002	One circular DNA molecule (2.7 Mb)
<i>Vibrio cholerae</i>	Two circular DNA molecules (2.9 Mb and 1,1 Mb)
<i>Agrobacterium tumefaciens</i>	One circular (3 Mb) and one linear DNA molecules (2.1 Mb)
<i>Borrelia burgdorferi</i>	One linear DNA molecule (0.9 Mb)

TABLE 1.1: Bacterial genomes: number, geometry and size (extragenomic DNA such as plasmids not shown). More genomes are described in [14].

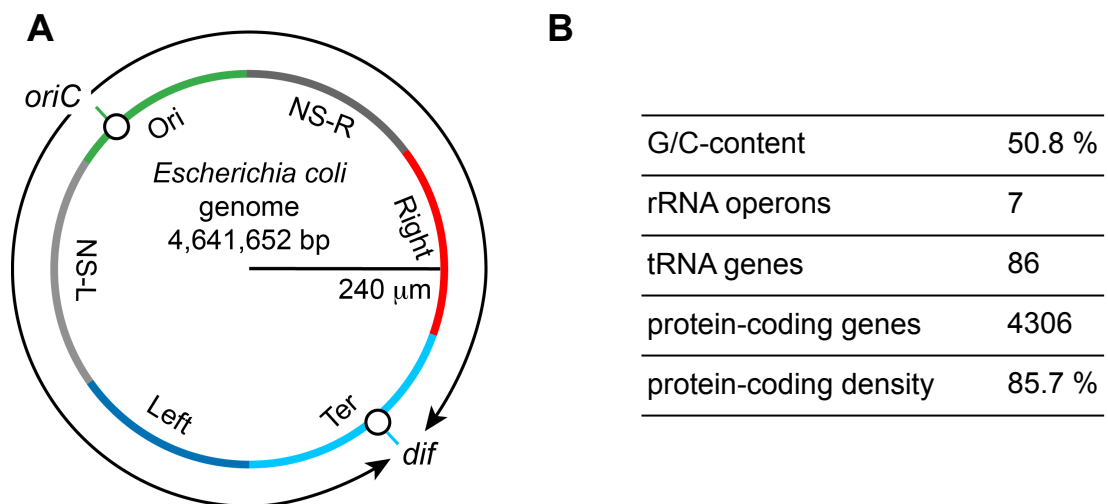


FIGURE 1.1: General features of bacterial genomes: example of the model bacterium *E. coli* K-12 MG1655. **(A)** Copied from [15]: the bi-directional DNA replication of bacterial genomes (arrows) initiates from a single origin of replication (*oriC*) until the converging replication forks meet in the replication termination region (*Ter*) at the site of chromosome decatenation of newly replicated DNA (*dif*). Colours represent regions of DNA that organise into spatially distinct macrodomains (see next section). NS: non-structured. **(B)** Adapted from [16]: rRNA operons refer to the operons encoding ribosomes, tRNA genes to the genes encoding transfer RNAs (see next subsection).

Gene expression occurs in two steps, transcription and translation, which are tightly coupled in space and time in bacteria [17], in contrast to eukaryotes, wherein the ribosomes, the key actors of translation, are separated from the genetic material (DNA and RNA) by the nucleus.

### 1.1.3 Bacterial transcription

During transcription, one strand of a DNA template is used to synthesise a complementary RNA molecule that serves either in translation (ribosomal and transfer RNAs), protein synthesis (messenger RNAs), or regulation (small RNAs). It is divided into three steps: initiation, subdivided itself into three stages, plus elongation and termination (Fig. 1.2).

In bacteria, transcription is carried out by a single RNA polymerase (RNAP), composed of a catalytic core with five sub-units ( $\alpha_2\beta\beta'\omega$ ) able to perform all steps of transcription except promoter-specific recognition, the region upstream of genes where initiation occurs [18]. The latter requires the association of RNAP to a sigma ( $\sigma$ ) factor, to form the holoenzyme ( $E\sigma$ ) (Fig. 1.2A). The sub-units of the core enzyme are highly conserved among bacteria [19], while  $\sigma$  factors are more variable [20]. For transcription to initiate its first stage, *i.e.* closed-complex formation, the  $\sigma$  factor recognises two well-conserved sequence elements, the -35 and -10 hexamers, separated by the spacer, which exhibits no consensus sequence but a consensus length of 17 nucleotides (nt) [21] (Fig. 1.2A-B). Some promoters may also contain an extended -10 element, generally in concert with a highly degenerated (or no) -35 box, and/or an UP element upstream of the -35 hexamer [22]. Once bound to the gene promoter, the holoenzyme undergoes a series of conformational changes, for DNA to be bent, wrapped and eventually get into the active site cleft of RNAP [18]. This ultimately leads to the open-complex formation, involving DNA melting over a distance of 13 base pairs (bp), from position -11 to +3, with respect to the transcription start site (TSS, +1) [18] (Fig. 1.2B). This includes the -10 hexamer, and an element located between the latter and the TSS, referred to as the discriminator, which exhibits no consensus sequence, and a length of typically six to eight nucleotides [21]. In a third and last stage, the transcription machinery has to escape promoter for productive RNA elongation, through DNA scrunching mechanism [23], generally involving a number of abortive initiation cycles whereby short RNA products are released (Fig. 1.2B). Importantly, depending on its sequence, every promoter has its own kinetic properties and activity, referred to as promoter strength, and the latter is usually (but not always) proportional to the proximity of promoter elements to consensus sequences [22].

During transcription elongation, the  $\sigma$  factor dissociates from RNAP, and one strand of the DNA template is used to synthesise a complementary RNA molecule as RNAP moves along DNA, until transcription termination occurs

(Fig. 1.2B). Two main mechanisms of transcription termination have been described in bacteria [24]. Rho-dependent termination requires a hexameric helicase, Rho (and its co factor NusG), which binds nascent RNA at specific termination (rut) sites, and translocates along the RNA in the 5' to 3' direction until it reaches the transcription complex, causing the release of the transcript and dissociation of RNA from DNA [24, 25]. Rho-independent or intrinsic termination involve short self-complementary G/C-rich sequences inducing stem-loop structures, followed by a poly(U)-tract in the RNA, causing a pausing of RNAP due to backtracking, and the formation of a hairpin-like RNA structure [24]. This eventually releases the nascent RNA from its complex with transcription machinery and gene template.

Finally, for messenger RNAs (mRNAs) encoding proteins, translation consists of three phases, namely initiation, elongation, and termination, similarly to transcription [26]. It is performed by ribosomes, that bind a short A/G-rich sequence referred to as Shine-Dalgarno (SD) sequence or ribosome-binding site (RBS), preceding the initiation codon (generally AUG). Amino acids are then added to the growing chain by aminoacyl-transfer RNAs (tRNAs), until ribosomes reach a stop codon. In bacteria, genes are generally grouped in operons that are coordinately expressed from a common promoter, resulting in a polycistronic transcript encoding multiple proteins [27].

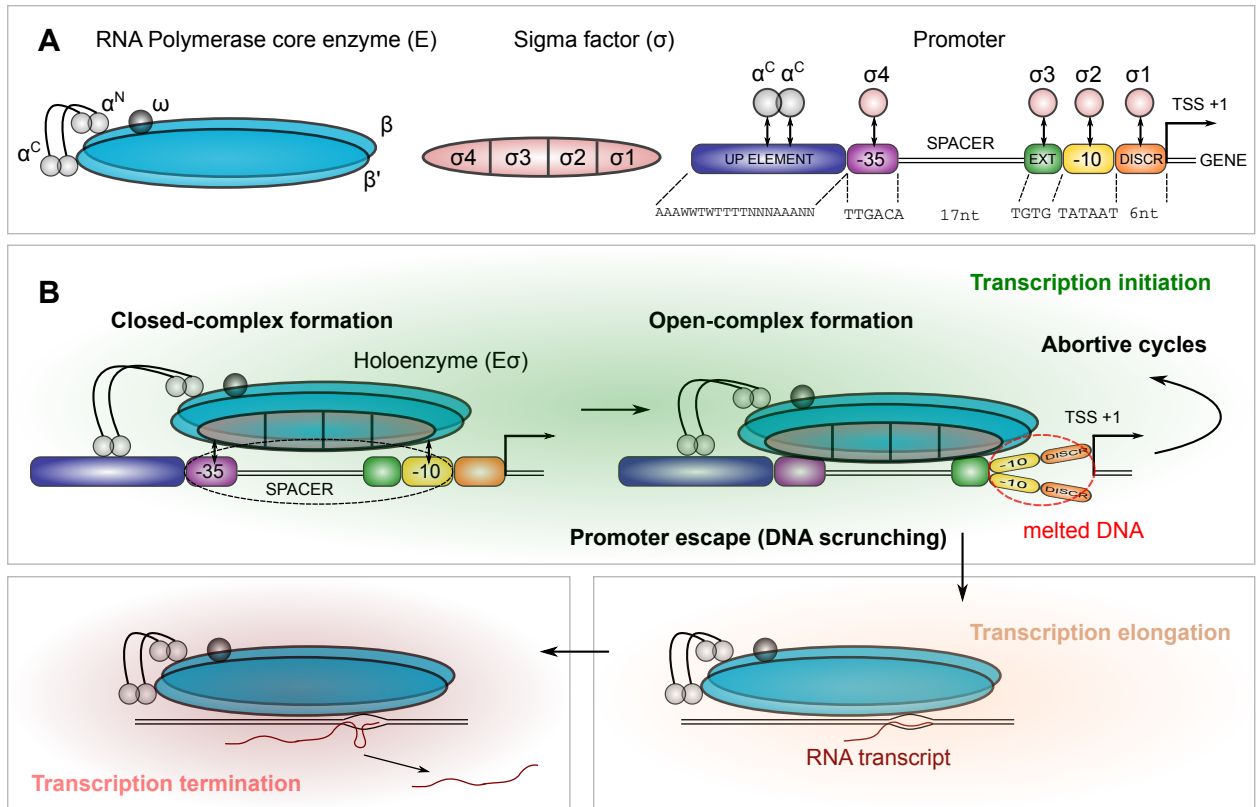


FIGURE 1.2: Adapted in part from [18, 21, 27]: schematic representation of transcription in bacteria. **(A)** Bacterial RNAP core enzyme (E) is composed of five sub-units ( $\alpha_2\beta\beta'\omega$ ), while  $\sigma$  factors ( $\sigma 70$ -related shown, see next section) comprise four domains ( $\sigma 1$  to 4). Bacterial promoters are constituted of the following elements: (optional) upstream promoter (UP), -35, spacer, (optional) extended -10 (EXT), -10, discriminator (DISCR), and transcription start site (TSS, denoted +1). Promoter consensus sequences and lengths (in nucleotides/nt) as well as interactions with transcription machinery are shown:  $\alpha^C$  refers to the carboxy-terminal domain of the  $\alpha$  sub-unit of RNAP,  $\alpha^N$  to its amino-terminal domain,  $\sigma 1$  to 4 to the four domains of  $\sigma$  factor. **(B)** For transcription to initiate, RNAP associates with a  $\sigma$  factor to form the holoenzyme  $E\sigma$ . In a first step, the latter binds gene promoter along -35 and -10 boxes separated by the spacer, to form a closed-complex. Then, it has to melt DNA over a distance of 13 base pairs, including -10 hexamer, discriminator element, and TSS, forming an open-complex. A scrunch complex is held at the promoter, resulting in possible cycles of abortive initiation whereby short RNA products are released, before the transcription machinery escapes from the promoter, dissociates from the  $\sigma$  factor and enters the elongation step. Finally, termination occurs when RNAP reaches a termination signal (Rho-dependent or intrinsic, see text), releasing gene DNA template and nascent RNA.



## 1.2 Regulation of transcription initiation

In response to environmental changes, bacterial gene expression can be regulated at the transcriptional level, but also at the post-transcriptional level, by regulating mRNA stability and translational efficiency, and at the post-translational level, by modulating protein degradation and activity. Extensive regulation occurs during transcription initiation [21, 22, 27], as it affects the very first step of the process and is thus presumably most interesting in terms of energy saving. However, accumulating studies suggest that a large part of regulation may also occur at the post-transcriptional and post-translational levels, which present the advantage of having an immediate impact on protein synthesis, quantity or activity, respectively [22]. Post-transcriptional regulation includes the modulation of RNA stability [28], which sets the number of both mRNA molecules that may be translated into functional proteins, and small RNAs that may exert a regulatory function. It also comprises variations in the accessibility of RBS to ribosomes by RNA-binding proteins [29], or by regulatory RNAs through double-stranded RNA structures [28]. Finally, post-translational regulation includes the modulation of protein degradation, sequestration [22], and post-translational modifications, *i.e.* the covalent modification of proteins by small chemical changes to the addition of entire functional groups or polypeptides, which impact protein function [30].

Transcription initiation is regulated by a variety of factors, that either modulate RNAP affinity for promoters, or RNAP activity itself. Some of them act at specific promoters, while some others have a global regulatory activity.

### 1.2.1 Role of transcription factors

As first discovered by Jacob and Monod [31], the most studied regulators are known as transcription factors (TFs), that either increase (activators) or decrease (repressors) transcription initiation frequency of certain promoters. TFs contain structural motifs (helix-turn-helix or, to a lower extent, ribbon-helix-helix) that bind to promoters containing specific sequences of more or less 4 bp referred to as operators. Since any given 4-bp sequence will arise once every  $4^4$  bp on average, to ensure binding of specific promoters, most TFs dimerise or multimerise, and bind direct or inverted repeats of operators, *e.g.* of 15- to 20-bp for dimers [22]. Generally, TFs bind near to, or overlapping RNAP binding sites. Repressors decrease promoter activity through

different mechanisms, including steric hindrance of RNAP, by binding operators overlapping -35 and -10 sites and thus blocking its access to promoter (Fig. 1.3A). This is the case of *E. coli* arginine repressor ArgR [32]. Some repressors operate through local DNA structure alterations, such as GalR, which induces DNA looping in the promoter region, preventing RNAP binding [33] (Fig. 1.3A). Inhibition of subsequent steps of transcription initiation by repressor binding to either or both RNAP and promoter is another possible regulatory mechanism. For instance, the protein p4 of phi29 bacteriophage binds upstream of RNAP, interacts with its C-terminal domain and inhibits promoter escape [34], while the Rv1222 protein of *Mycobacterium tuberculosis* binds to both RNAP and DNA, slowing down or preventing RNA elongation [35] (Fig. 1.3A). Finally, some repressors act by anti-activation mechanisms, such as occurs with CytR, the repressor of genes involved in uptake and catabolism of nucleosides, which simultaneously interacts with its operator and the adjacent activator CRP/CAP, to prevent RNAP binding on promoter favoured by the latter [36].

Activators increase promoter activity through three main mechanisms, namely class I or II activation, or activation by conformational changes. In class I activation, the activator binds to an operator located upstream of the promoter, and then recruits RNAP through interactions between the activating region of the activator and the  $\alpha^C$  carboxy-terminal domain of the  $\alpha$  sub-unit of RNAP, as occurs with *E. coli* CRP activator, e.g. at *lac* and *gal* operators [37] (Fig. 1.3B). In class II activation, the activator binds to an operator overlapping -35 promoter element to form direct interactions with the domain 4 of  $\sigma$  factor, and/or components of the RNAP, notably the  $\alpha^N$  amino-terminal domain of its  $\alpha$  sub-unit. This eventually facilitates RNAP recruitment for closed-complex formation, or promotes transition to open-complex formation (Fig. 1.3B). It is worth noting that class II activators may operate in combination with class I activators on the same promoter, and that some activators may act as both class I and class II activators, such as the global regulator CRP [38]. Finally, the third type of activation consists of promoter remodelling. For instance, the TFs of the MerR family which are described in more detail in Chapter 4, bind to a region between -35 and -10 elements of promoters with a sub-optimal spacer length, distort DNA, resulting in a better alignment of RNAP binding sites and thus facilitating its recruitment (Fig. 1.3B). Finally, similarly to indirect anti-activation mechanisms previously described, there are also some anti-repression mechanisms. For instance, at the *nir* promoter of *E. coli* encoding a nitrite reductase, class II

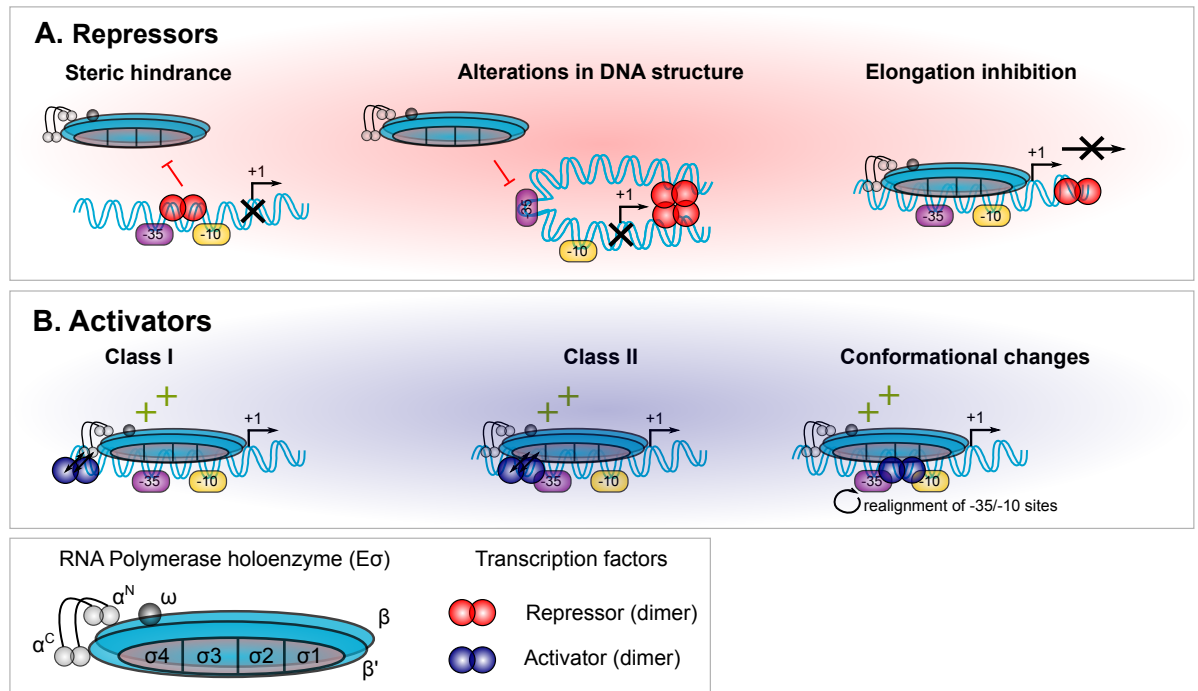


FIGURE 1.3: Adapted from [22]: regulation of promoters by transcription factors. **(A)** Mechanisms of repression (red), which include repression by steric hindrance, wherein a repressor binds operators overlapping -35 and -10 sites, preventing RNAP binding on promoter. But also alterations in DNA structure induced by a repressor (in the present case, looping) which may prevent RNAP binding. Finally, some mechanisms of repression consist of the inhibition of subsequent steps of transcription initiation, *e.g.* RNA elongation (present case), or promoter escape. **(B)** Mechanisms of activation (blue), which include class I activation, wherein an activator binds an operator located upstream of the promoter, facilitating RNAP binding through interactions with the carboxy-terminal domain of the  $\alpha$  sub-unit of RNAP. But also class II activation, wherein an activator binds an operator overlapping -35 element through interactions with domain 4 of the  $\sigma$  factor, and/or components of the RNAP, notably the  $\alpha^N$  amino-terminal domain of its  $\alpha$  sub-unit. Finally, some mechanisms induce conformational changes, in the present case, realigning -35 and -10 elements to promote RNAP binding on promoter.

activation, repression and indirect activation through anti-repression all contribute to regulate its activity [39]. The above regulation mechanisms facilitate RNAP recruitment at target promoter, rather than promoting transition from closed to open-complex, which usually does not require any activator. In contrast, promoters bound by RNAP associated to  $\sigma 54$  factor (see next subsection on the variability of  $\sigma$  factors) require a special class of activators, referred to as enhancer-binding proteins, which provide the energy required for promoter opening through ATP hydrolysis, since  $\sigma 54$  closed-complex is

unable to spontaneously isomerise to an open-complex [40].

Importantly, the affinity of TFs for operator sequences can be modulated by chemical modifications, *e.g.* phosphorylation-dephosphorylation [41], oxidation-reduction [42], or by reversible interaction with small molecules [22], in which case TFs act as allosteric proteins undergoing conformational changes once bound to inducers and/or co-repressors. Indeed, the activity of TFs is generally modulated by environmental changes, which, in turn, control the expression of specific genes required for the adaptation to the new physiological conditions. The most paradigmatic example is the *lac* operon, encoding the structural genes necessary to acquire and process lactose from the environment. In *E. coli*, in absence of the latter, *lac* operon is turned off by a repressor, LacI, which inhibits RNAP binding both by steric hindrance, and by forming a DNA loop through tetramerisation [43, 44]. In presence of lactose, LacI is bound by allolactose inducer, preventing its binding to the operator through conformational changes. To avoid wasteful consumption of resources and energy, since glucose is the preferred substrate, the *lac* operon then requires an additional activator, the cAMP-CRP complex. Activation by CRP requires the inducer cAMP, whose concentration is increased in absence of glucose [45].

Finally, it is important to note that (i) most TFs regulate several promoters, (ii) most promoters are regulated by more than one TF, (iii) most TFs are regulated by other TFs (auto-regulation in some cases), resulting in complex regulatory networks within the cell [46]. However, more than half of *E. coli* promoters are not targeted by any known TF [46], and entire organisms are almost devoid of them [47, 48] but exhibit nonetheless a complex regulation, indicating that other modes of regulation must exist. Beyond such regulators which usually bind to the promoters of a specific and limited set of functionally-related genes (except global TFs such as CRP), gene expression is regulated at a larger scale by nucleoid-associated proteins (NAPs). The latter act as global TFs by simultaneously controlling the transcription of many operons belonging to diverse functional categories, while shaping chromosome architecture.

## 1.2.2 Role of nucleoid-associated proteins

NAPs are architectural proteins of small size that ensure proper compaction of DNA inside bacterial cells, together with DNA supercoiling (see next section) [15]. Indeed, bacterial chromosomes would be longer than 1 mm in length if stretched out (Fig. 1.4A), while bacterial cells range from less than one to over ten micrometers in length [49]. As a consequence, the chromosome must be compacted more than 1000 fold to fit inside the cell (Fig. 1.4B).

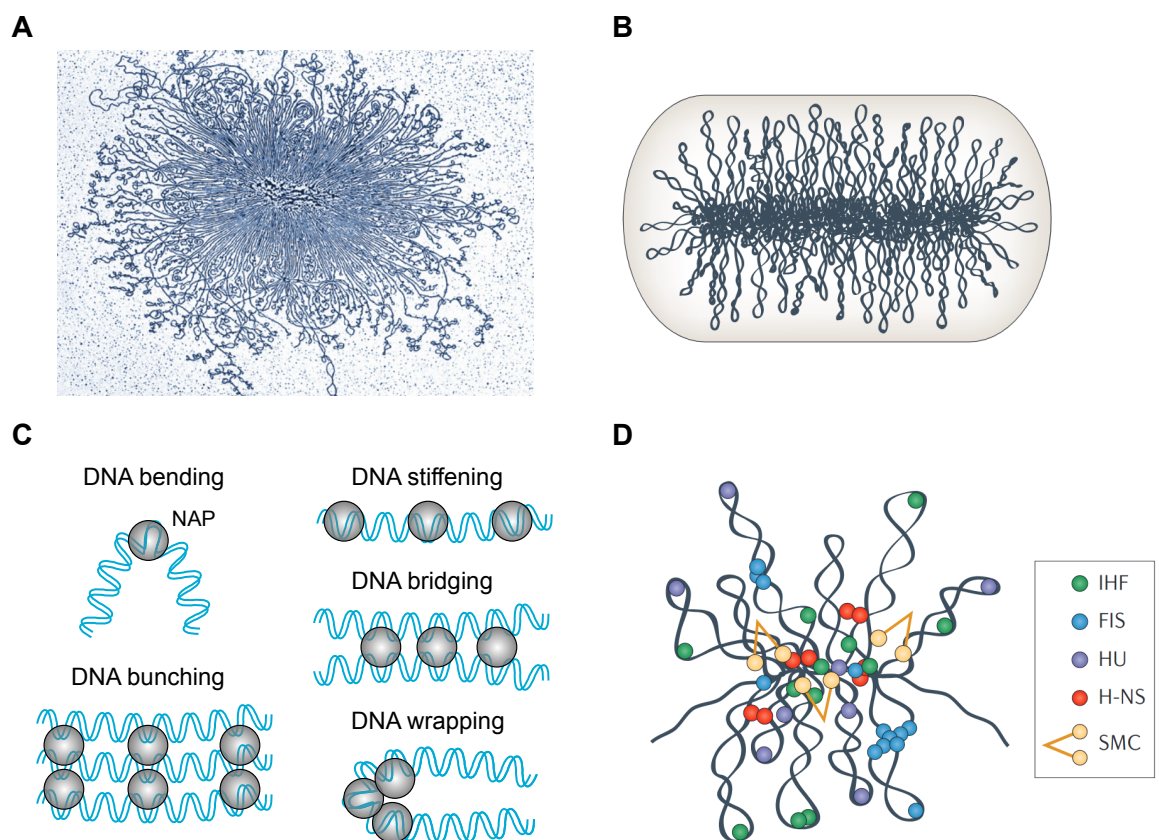


FIGURE 1.4: Chromosome organisation by nucleoid-associated proteins. **(A)** Copied from [49]: a gently lysed *E. coli* nucleoid visualised by transmission electron microscopy, of approximately 20 micrometers in length which is many times longer than the length of the cell (2-5 micrometers). **(B)** Copied from [49]: schematic representation of the compacted chromosome inside bacterial cell. **(C)** Adapted from [15]: once bound to more or less specific sites, NAPs shape DNA by bending it, stiffening it, bridging adjacent sites into parallel DNA segments, bunching several parallel DNA fragments together, or wrapping it by coherent bending. **(D)** Copied from [49]: schematic representation of the folded chromosome organised into looped domains by the binding of NAPs and SMC condensin complexes.



This compaction is hierarchically organised at multiple scales [15, 49]. At the kilobase (kb) scale, NAPs shape chromosome structure by bending, bridging, wrapping, looping and twisting DNA, frequently through dimerisation/-multimerisation (Fig. 1.4C-D and Fig. 1.5). At a larger scale (10 kb), plectone-mic loops arise from the presence of supercoiling, forming multiple topologically-constrained domains by the key action of NAPs (see next section and Fig. 1.4D). At the Mb scale, these loops organise into six spatially distinct macrodomains (Fig. 1.1A, Ori, Ter, Left, Right, NS-L, NS-R) wherein physical interactions among DNA sites are more frequent than between different macrodomains. Structural maintenance of chromosome (SMC) condensin complexes also play key roles in chromosome organisation and segregation, at multiple scales [50] (Fig. 1.1D). Eventually, bacterial chromosomes resemble rosettes with super-coiled loops of DNA (Fig. 1.1B and 1.1D).

All bacteria possess NAPs, some of which are unique for a given genus and/or species [51]. The main NAPs of *E. coli* include H-NS (histone-like nucleoid structuring protein), HU (heat-unstable protein), IHF (integration host factor), Fis (factor for inversion stimulation), Lrp (leucine-responsive regulatory protein), and Dps (DNA-binding protein from starved cells) which differ by their DNA-binding mode (Tab. 1.2 and Fig. 1.5). Most NAPs display low sequence specificity for binding, with a preference for A/T-rich sequences, and use indirect readout mechanisms, *i.e.* recognise and bind DNA with a specific structure and flexibility [52–54], although some of them have more sequence requirements (Tab. 1.2). There is a variety of NAPs, and their cellular abundance can vary in response to growth conditions, growth phase, and environmental changes in general [55] (Fig. 1.6).

NAP	DNA bending	DNA stiffening	DNA bridging	DNA bunching	DNA wrapping	Binding motif	Mass	Oligomer
H-NS	ND	yes	yes	ND	ND	AT-rich DNA/ TCGATAAAATT	15 kDa	Homodimer/ Heterodimer
HU	yes	yes	no	yes	ND	AT-rich/ curved DNA	9 kDa	Heterodimer
IHF	yes	ND	ND	ND	ND	(A/T)ATCAAN NNNTT (A/G)	11 kDa	Heterodimer
Fis	yes	ND	yes	ND	yes	A tracts/AT tracts	11 kDa	Homodimer
Lrp	ND	ND	yes	ND	yes	(T/C)AG(A/T/C) A(A/T)ATT(A/T) T(A/T/G)(A/G)	18 kDa	Homodimer/ Octamer
Dps	yes	ND	ND	ND	ND	[56]	19 kDa	Monomer / Dodecamer

TABLE 1.2: Adapted from [15, 51]: main nucleoid-associated proteins (NAPs) of *E. coli*. H-NS: histone-like nucleoid structuring protein, HU: heat-unstable protein, IHF: integration host factor, Fis: factor for inversion stimulation, Lrp: leucine-responsive regulatory protein, Dps: DNA-binding protein from starved cells, ND: not determined, N: A, T, C or G nucleotide.

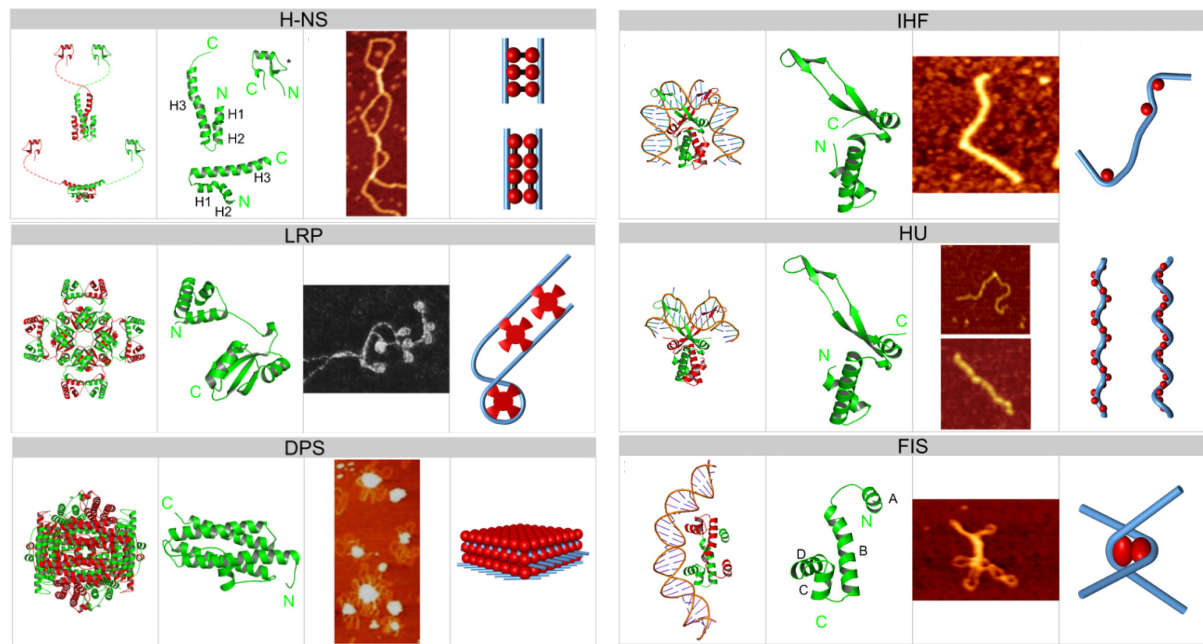


FIGURE 1.5: Copied from [57]: structural and architectural properties of the main nucleoid-associated proteins. Panels from left to right for each NAP: Protein Data Bank (PDB) structures (first two panels), microscopy images, schematic model for the mechanism of action. Dimeric H-NS from *E. coli* is depicted (followed by its monomeric components), which induces DNA loops as a consequence of DNA duplex bridging, as shown in the scanning force microscopy (SFM) image. Octameric LRP from *Pyrococcus furiosus* is depicted (followed by a monomeric Lrp sub-unit), which wraps and bridges duplex DNA, as shown in the electron microscopy (EM) image. Dodecameric Dps from *E. coli* is depicted (followed by a monomeric Dps sub-unit), which induces three-dimensional hexagonal Dps-DNA arrays, as shown in the SFM image. IHF-DNA complex is depicted (followed by a monomeric IHF sub-unit), which induces strong DNA bending (up to  $140^\circ$ ), as shown in the SFM image. HU-DNA complex is depicted (followed by a monomeric HU sub-unit), which induces weaker DNA bending than IHF, as shown in the SFM image, wherein HU bends DNA at low concentrations (top), and induces the formation of rigid DNA filaments (bottom) at high concentrations. Fis-DNA complex is depicted (followed by a monomeric Fis sub-unit), which induces weaker DNA bending than IHF yet stronger than HU (up to  $90^\circ$ ). Fis-Fis interactions lead to node formation, as shown in the SFM image.



Beyond their role in genome organisation and various other cellular processes such as DNA replication, recombination, and repair [55], NAPs share regulatory features with TFs [58], by inducing topological and/or structural changes at promoters that may modulate the binding of RNAP or TFs (Fig. 1.3) [51]. For instance, H-NS induces the formation of DNA-protein-DNA bridges (Fig. 1.4C and Fig. 1.5), trapping RNAP in a loop and preventing its binding on promoter (Fig. 1.3A), similarly to the LacI repressor [44]. Depending on their cellular abundance and number of target DNA sequences, NAPs can thus affect the transcription of many genes simultaneously, and ensure a quick and global response to various stimuli, as reviewed in the following: [51, 55]. Besides NAPs, the use of alternative  $\sigma$  factors provides another mechanism for bacterial cells to change their global gene expression program in response to environmental changes.

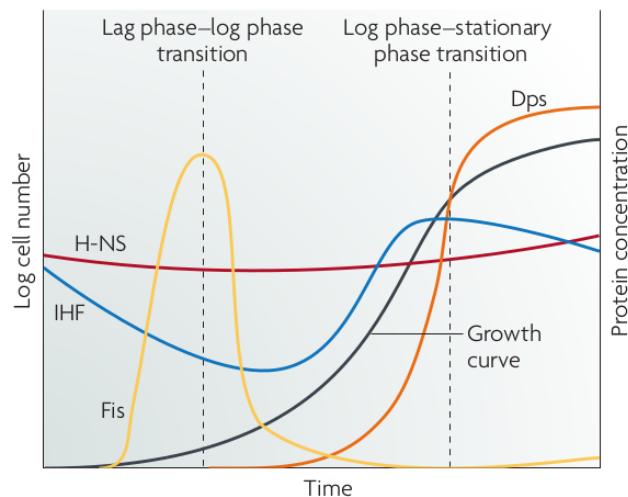


FIGURE 1.6: Copied from [51]: expression patterns of four nucleoid-associated proteins during the growth of *E. coli*. A typical bacterial growth curve is depicted, together with the different protein concentrations, and the growth phase transitions are indicated. H-NS concentration is found to be more or less constant throughout bacterial growth. IHF and Dps protein concentrations increase in stationary phase, whereas Fis is more abundant in exponential phase (such as HU which is not depicted).

### 1.2.3 Role of alternative $\sigma$ factors and small ligands

The role of  $\sigma$  factors, once associated to RNAP, is to ensure specific promoter binding of -35 and -10 hexamers, and orchestrate open-complex formation. Besides their housekeeping  $\sigma$  factor ( $\sigma^{70}$  in *E. coli*,  $\sigma^A$  in *B. subtilis* and cyanobacteria) recruiting RNAP at the majority of promoters, nearly all bacteria possess a variable number of alternative  $\sigma$  factors that recognise different classes of promoters. For instance, there are seven  $\sigma$  factors in *E. coli* (Tab. 1.3), and up to 109 in the myxobacterium *Sorangium cellulosum* [59]. The latter compete for binding to the limited pool of RNAP core enzymes, estimated about 2000 in *E. coli* cells, 700 of which are free [60]. In most conditions, the housekeeping  $\sigma$  factor is more abundant and out-competes alternative ones. However, when the concentration of an alternative  $\sigma$  factor increases through regulation of their transcription and translation, proteolysis, subcellular localisation, and sequestration by anti- $\sigma$  factors under specific conditions, it can then compete with and displace the housekeeping  $\sigma$  factor [22]. This provides a simple mechanism for reprogramming transcription to a different set of genes involved in specific cellular programs. For instance, in *E. coli*, while housekeeping  $\sigma^{70}$  factor is required for most transcription during growth,  $\sigma^{32}$  and  $\sigma^{38}$  activate the transcription of genes involved in heat shock stress response and upon entry to stationary phase, respectively (Tab. 1.3).

$\sigma$ factor	Function
$\sigma^{70} / \sigma^D$	Growth-related (housekeeping) genes
$\sigma^{54} / \sigma^N$	Nitrogen metabolism
$\sigma^{38} / \sigma^S$	Stationary phase and starvation
$\sigma^{32} / \sigma^H$	Heat shock response
$\sigma^{28} / \sigma^F$	Motility genes
$\sigma^{24} / \sigma^E$	Extracytoplasmic stress response
$\sigma^{19} / \sigma^{fecI}$	Ferric citrate transport

TABLE 1.3: Adapted from [22]: the seven  $\sigma$  factors of *E. coli*.

RNAP can also interact with small ligands modulating its global activity at numerous promoters. The best studied example is the alarmone ppGpp (guanosine 3' diphosphate, 5' diphosphate), which binds directly to the  $\beta$  and  $\beta'$  sub-units of RNAP, near its active site [61]. It exerts a global regulatory effect in cooperation with DksA, a small RNAP-binding protein reducing open-complex stability [62]. In *E. coli*, ppGpp production is directed by two different pathways, involving SpoT, or RelA, wherein pppGpp is first produced and then converted to ppGpp by GppA [63]. It is synthesised in

response to starvation for a variety of nutrients, and environmental changes that cause growth arrest in general. As a result, once bound to RNAP, ppGpp redirects transcription from growth-related genes, including mostly those involved in translation (*e.g.*, rRNA/tRNA) and DNA replication, to genes involved in starvation survival and stress resistance, including *rpoS*, the gene encoding the alternative  $\sigma^{38}$  factor binding promoters required in stationary phase and starvation (Tab. 1.3) [64]. Generally, its repressive effect occurs at promoters referred to as stringent promoters, characterised by a high-energy barrier for open-complex formation, due to a G/C-rich discriminator upstream the TSS, and a C nucleotide at position -1 [65]. In addition, ppGpp may act indirectly by redistributing available RNAP from stringent promoters to non-stringent ones [64], and by modulating  $\sigma$  factor competition, due to the induction of *rpoS*.

Promoter activity can also be regulated by reversible chemical modifications of bases of the DNA sequence, also referred to as epigenetic regulation [66], including mostly DNA methylation of adenines, modulating promoter affinity for RNAP or TFs. Finally, transcription initiation is affected at a global scale by a variety of environmental variables including nucleotide concentration, temperature, pressure and salt [67].

### 1.2.4 Existing models of transcriptional regulation

Gene regulation is essential to every biological process, determining everything from how cells sense and respond to environmental changes, to how they perform optimal energy management in general. Unravelling gene regulation requires the knowledge of which genes are expressed at which level, when and where, and how they interact with each other. Indeed, biological processes typically involve the interactions of a number of genes acting on each other in gene regulatory networks (GRNs). Due to their complex nature structured by positive and negative interactions between many different components, including DNA, RNA, proteins, small molecules and more, an intuitive understanding of their dynamics requires computational and modelling tools. The latter are also of crucial importance to provide mechanistic insights into how the components of a GRN are interrelated.

Generally, modelling of gene regulation involves the construction of an initial model based on available data and knowledge, with the purpose of simulating a given regulatory system for a variety of conditions experimentally reproducible. The resulting predictions are then confronted to expression data derived from experiments, providing an indication of model adequacy. The model is revised whenever predictions and experimental data do not match, and model adjustment is repeated until an optimal adequacy is obtained (Fig. 1.7). Eventually, this leads to new biological insights into the components of the regulatory system, and how they are interrelated, while allowing to simulate the influence of perturbations, *e.g.* knockdown or over-expression of specific genes [68].

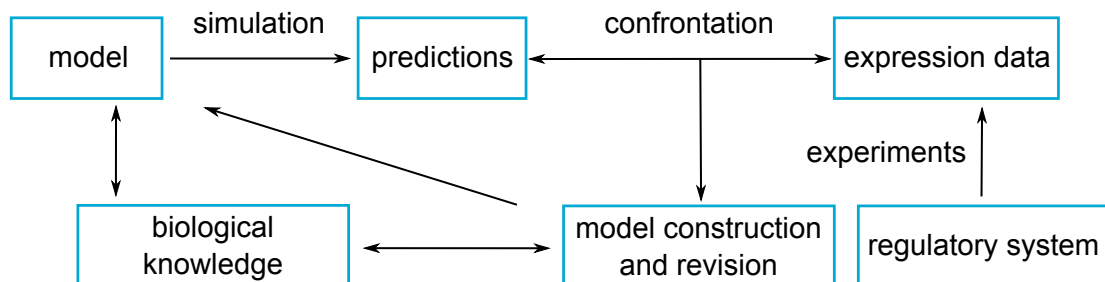


FIGURE 1.7: Adapted from [68]: modelling of gene regulatory systems. The latter are generally modelled based on biological knowledge, and simulations are then performed to obtain predictions, which are confronted to expression data derived from experiments. The model is repeatedly revised to increase its adequacy regarding experimental data. Biological knowledge is required and generated throughout the process.

Models of gene regulation are roughly divided into statistical and analytical models [69]. The first type emerged from the huge amounts of genome-scale data that have been made available with the advent of transcriptomic technologies, including the expression levels for all genes of an organism, or genome-wide occupancy of TFs and other DNA-binding proteins such as NAPs. Statistical models aim to identify GRNs underlying these genome-wide data, and the regulatory influences between its components, without providing biophysical explanations for the regulatory effects that come into play [70]. They are thus particularly used in systems biology, where the physiology of living systems is modelled as a whole rather than as a collection of single biological components. In such an approach, a GRN is generally composed of nodes representing genes and/or proteins, and edges representing regulatory influences between components, that can be either direct or indirect, and either positive or negative (Fig. 1.8). To infer a GRN, diverse and various GRN architectures exist which differ in their statistical approach, including system of equations, Boolean and Bayesian networks, information theory models and more, which are reviewed extensively elsewhere [68, 70]. The main drawbacks of these models are (i) the large data-sets on which they are based provide only an average picture of many cell states, at a given time point, and (ii) they usually fail to explain mechanistically the regulatory relations between the components of the GRN.

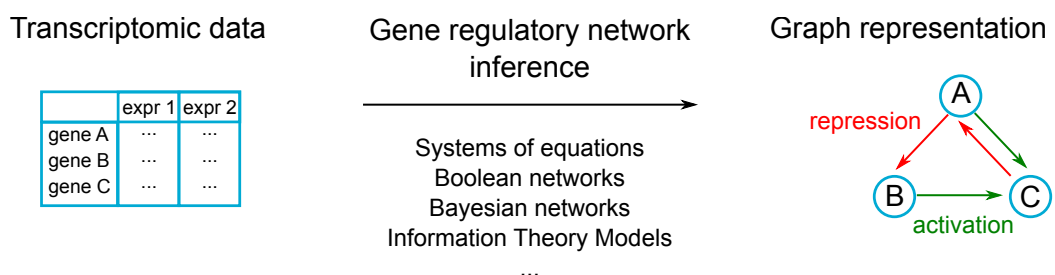


FIGURE 1.8: Adapted from [70]: statistical models of gene regulatory systems.

In contrast, analytical or mechanistic models generally focus on a small number of components, and aim to provide biophysical explanations to the regulatory effects linking them, requiring more knowledge and hypotheses about the system than statistical approaches [69]. They are based on a variety of mathematical methods, which include terms relating to the binding of regulators and RNAP to DNA, transcription rate, mRNA translation rate, mRNA and protein degradation, allowing to dissect GRNs and decipher regulation mechanisms. There are three main classes of mathematical approaches for analytical modelling: (i) differential equations models, (ii) Boolean models, and (iii) thermodynamic models [69]. They can be deterministic, *i.e.* produce the same exact outputs from a set of inputs, or stochastic, *i.e.* account for the inherent noise of biological systems, resulting in variable outputs from the same exact set of inputs. They can be discrete, such as Boolean models, *i.e.* represent time state or space as a discrete set of values, or continuous, such as differential equations-based models, *i.e.* use continuous values of time to represent dynamical changes occurring in the modelled system.

Differential equations are particularly suited to model temporally-evolving, dynamic systems. In such an approach, the concentration of RNAs, proteins, and other molecules is specified by differential equations as a function of the other components of the GRN, as the system evolves. These models use either Ordinary Differential Equations (ODE) which contain only one variable (time in most cases), or Partial Differential Equations (PDE) which involve multiple time and/or space-dependent variables [68, 69]. In the example below (Fig. 1.9), the transcription rate  $k$  of a gene is function of notably the proximity of promoter elements to consensus and the binding of regulatory elements. The resulting mRNAs are then translated into one or multiple copies of corresponding proteins with a rate  $t$ , that depends notably on the proximity of the RBS to consensus, and on the binding of regulatory molecules such as small RNAs. They then decay with a rate  $\lambda^m$ , which is also affected by a variety of factors. The proteins can further modulate the transcription of other genes, before decaying with a rate  $\lambda^p$ . The quantity of each mRNA and protein can then be modelled over time as a function of all those factors, with respect to the other components of the GRN. Differential equations models have a wide range of applications which are reviewed extensively elsewhere [68, 69], including the study of operons dynamics such as *lac / trp* operons [71]. The main drawbacks of these models are (i) the extensive knowledge of the system required, and (ii) the usually large number of parameters to be determined or arbitrarily set, although computational

methods have been developed to overcome this limitation [72, 73].

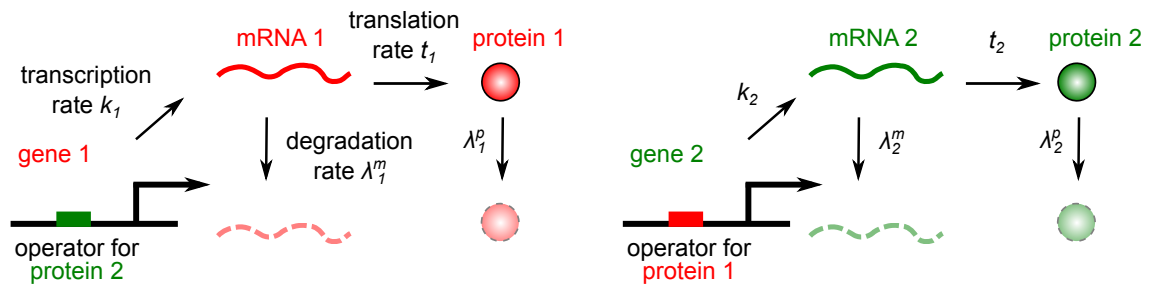


FIGURE 1.9: Adapted from [69]: differential equations-based models.

Concerning Boolean networks, they are used to study biological processes that exhibit on/off switch-like behaviours, such as bacterial competence, *i.e.* the ability to take up exogenous DNA, and gene transcription. In such models, at each discrete time step, each component of the GRN has one of only two states (on/off), the latter being determined by a Boolean function of the states of the other components of the GRN (Fig. 1.10). Despite their simplicity, Boolean networks share some features with biological systems, and may provide insights into their behaviour and dynamics, such as properties of self-organisation, stability, redundancy and periodicity, while not requiring extensive knowledge about the system and its components, *e.g.* nature of molecular interactions, values of mRNA/protein diffusion and decay. As a drawback of their simplicity, Boolean networks fail to provide a relevant description of biological systems that rely on fine details and kinetics, and are rather used for exploratory purposes [68, 69].

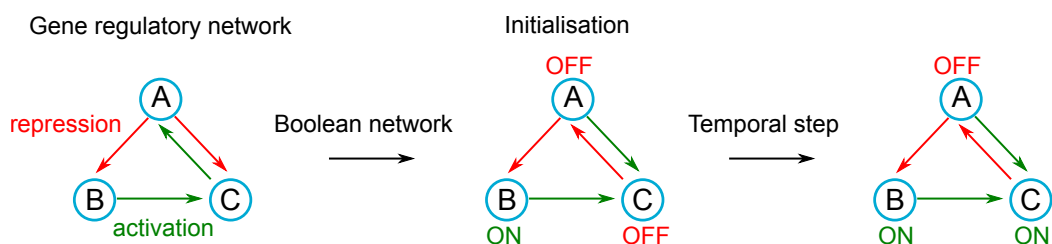


FIGURE 1.10: Boolean network-based models.



Finally, thermodynamic models rely on biophysical descriptions of DNA-protein interactions originating from statistical mechanics [69]. They have been selected for the present thesis for unravelling transcriptional regulation mechanisms. Thermodynamic models consider all possible states of a system, each of them being assigned a Boltzmann weight  $\omega = \Omega \times e^{-\beta\epsilon}$ , where  $\Omega$  is the multiplicity, *i.e.* the number of microstates corresponding to a particular macrostate (for simplicity, "state" will refer to a macrostate hereafter),  $\epsilon$  is the energy of that particular state, which depends on a variety of factors such as the concentration of regulatory elements and their binding site affinity,  $\beta = \frac{1}{k_B T}$ , where  $k_B$  is the Boltzmann's constant, and  $T$  is the system temperature [74]. According to the Boltzmann distribution, the probability of each state can then be calculated by dividing the Boltzmann weight of the state by the sum of the weights of all possible states. The probabilities of all transcriptionally active states in which promoter is bound by RNAP are summed, and the overall level of gene expression is assumed to be proportional to this probability [75]. As an illustration, in the case of a simple repression where RNAP and the repressor cannot bind promoter simultaneously, there are three possible states in the regulatory system: (i) the promoter is empty, (ii) the promoter is bound by RNAP, or (iii) the promoter is bound by the repressor (Fig. 1.11). The probability of promoter occupancy by RNAP  $P_{bound}$  is then given by the following equation [76]:

$$P_{bound} = \frac{\frac{P}{N_{NS}} e^{-\beta\Delta\epsilon_P}}{1 + \frac{P}{N_{NS}} e^{-\beta\Delta\epsilon_P} + \frac{R}{N_{NS}} e^{-\beta\Delta\epsilon_R}} \quad (1.1)$$

where  $P$  and  $R$  are the number of RNAP and repressors in the cell, respectively,  $N_{NS}$  is the number of non-specific binding sites for both RNAP and repressor, usually given by the size of the genome,  $\Delta\epsilon_P$  and  $\Delta\epsilon_R$  are the difference between the specific and non-specific binding energies ( $\Delta\epsilon_R = \epsilon_R^S - \epsilon_R^{NS}$ , and  $\Delta\epsilon_P = \epsilon_P^S - \epsilon_P^{NS}$ ) of  $P$  RNAP and  $R$  repressors. More regulatory systems together with their thermodynamic analysis are provided in the following [75], including more complex configurations such as dual and/or cooperative activation/repression. Thermodynamic models have a wide range of applications which are reviewed extensively elsewhere [77], including the study of the combinatorial control of the *lac* promoter by LacI and CRP [78], or the characterisation of the transcriptional regulation occurring at PR and PRM promoters of the lambda phage [79, 80].



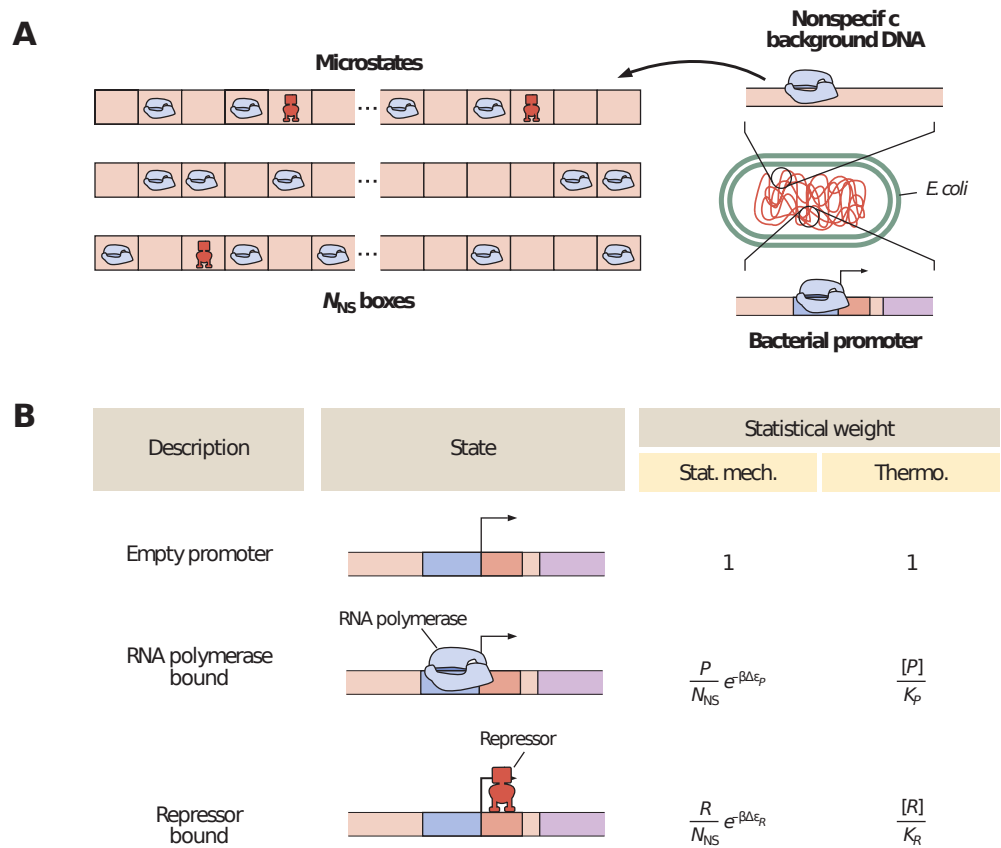


FIGURE 1.11: Copied from [76]: thermodynamic modelling of a simple repression. **(A)** The regulatory system is composed of  $P$  RNAP (in blue) and  $R$  repressors (in red) per cell, that either bind non-specifically to the genomic background on potential binding sites  $N_{NS}$  usually given by genome length, or compete for binding to the promoter of interest. **(B)** The different states of the system are depicted together with their Boltzmann weight (Stat. mech.) translated into thermodynamic formulations (Thermo.). The binding energies of the  $R$  repressors and  $P$  RNAP to their binding sites on the promoter are given by  $\epsilon_R$  and  $\epsilon_P$ , while  $\Delta\epsilon_R$  and  $\Delta\epsilon_P$  correspond to the difference between the specific and non-specific binding energies ( $\Delta\epsilon_R = \epsilon_R^S - \epsilon_R^{NS}$ , and  $\Delta\epsilon_P = \epsilon_P^S - \epsilon_P^{NS}$ ). More negative values of  $\Delta\epsilon$  indicate stronger binding. In thermodynamic formulation,  $[P]$  and  $[R]$  correspond to the cellular concentrations of RNAP and repressor, respectively, while  $K_P$  and  $K_R$  correspond to their dissociation constants, respectively, which correspond to the concentrations at which the promoter has a probability of being occupied of 0.5, reflecting their binding affinity.

In such models, gene expression is proportional to the equilibrium probability that promoter is bound by RNAP. The first key and questionable assumption is that of equilibrium itself, since transcription is multi-step, inherently out-of-equilibrium process [18] with irreversible stages, such as promoter escape and transcription elongation leading to mRNA production [22]. However, using a simplified model of the kinetics of transcription initiation, we can make the assumption that RNAP binding and unbinding occur at faster timescale than the rate of open-complex formation, promoter escape and transcription elongation. Due to the separation of time scales, the probability of RNAP binding on promoter can thus be obtained from its equilibrium value. This is the same for the concentration of a given regulator, which may change overtime, preventing the assumption of system equilibrium, although the rate at which it changes is slow compared to its rates of binding/unbinding, allowing the system to be considered nearly in equilibrium at any moment. The relevant separation of time scales thus renders the equilibrium assumption of statistical mechanics close to correct [74–77, 81].

The second key assumption concerns the proportionality between gene expression and promoter occupancy by RNAP ( $P_{bound}$ ) which is questionable, since there is a variety of mechanisms and factors intervening between RNAP binding and the synthesis of a functional gene product [22]. This includes open-complex formation, promoter escape, transcription elongation and termination (Fig. 1.2), but also mRNA/protein degradation and translation. If one of these steps may be rate-limiting rather than RNAP binding on promoter, these models would thus fail to explain experimental data. In spite of these caveats, thermodynamic models provide an instructive and predictive mechanistic framework to unravel regulation mechanisms, and are particularly useful for predicting gene expression variations (fold-changes) in transcriptional activity. Indeed, in the presence of unknown parameters, *e.g.* the rate of open complex formation, promoter escape or mRNA/protein degradation, as long as the latter do not depend upon the investigated regulation mechanisms, they are eliminated when fold-changes are computed. As an illustration, in the example above (Fig. 1.11), fold-changes will depend only on the ratio of  $P_{bound}$  in the presence of the repressor to  $P_{bound}$  in its absence [81] (see applications in Chapters 4 and 5).

Finally, there is one particular transcriptional regulator which was soon discovered to have a global and complex regulatory effect, and deserves special attention for the present thesis: DNA supercoiling.

## 1.3 Transcriptional regulation by DNA supercoiling during environmental changes

### 1.3.1 DNA supercoiling

DNA structure consists of two complementary strands held together by hydrogen bonding between complementary base pairs (two hydrogen bonds between A and T, three between C and G), and intertwined to form a right-handed helix. Soon after the discovery of this double-helix model of DNA [82], Watson and Crick pointed out themselves the topological problems associated with the intertwining of the two strands of DNA around each other [83]. Untangling these two strands is required for virtually every DNA transaction, notably for strands separation occurring during DNA replication, repair, recombination, and transcription. However, if this is possible for a linear DNA molecule in solution due to the free rotation of its ends, it is not the case for natural DNA. Indeed, bacterial chromosomes are circular DNA molecules which are covalently closed, preventing free-end rotation. Whereas eukaryotic chromosomes, although linear, consist of large DNA loops wrapped around histones, forming topological domains, *i.e.* constrained DNA regions equivalent to circular, topologically-closed DNA, due to the free-end rotation impossibility (Fig. 1.12) [84]. As a result of those topological constraints, virtually every DNA transaction generates an ubiquitous torsional stress referred to as DNA supercoiling (SC). Numerically, SC can be described with the linking number ( $Lk$ ), *i.e.* the number of crosses a single strand makes across the other in a closed DNA molecule. The linking number equals the sum of twist ( $Tw$ ), *i.e.* the total number of helical turns in the DNA molecule (Fig. 1.12), and writhe ( $Wr$ ), *i.e.* the total number of times the double-helix crosses itself, forming either plectonemes or solenoids (Fig. 1.12) [85]. For a topologically-closed DNA molecule, the linking number remains invariable, *i.e.* cannot be changed by any deformation of the DNA strands, even if complementary changes in  $Tw$  and  $Wr$  may still occur, as long as no break is introduced in one or both DNA strands (Fig. 1.12). For a given DNA molecule at rest, *i.e.* that is not supercoiled, its linking number  $Lk_0$  is defined by  $Lk_0 = \frac{N}{\gamma}$  where  $\gamma$  is the number of base pairs per helical turn (10.5 bp in average). The DNA molecule is supercoiled whenever its actual linking number differs from  $Lk_0$ . If  $Lk > Lk_0$  (twisting in a right-handed fashion, Fig. 1.12), DNA is over-wound ( $> 10.5$  bp per helical turn), *i.e.* positively supercoiled, whereas if  $Lk < Lk_0$  (twisting in a left-handed

fashion, Fig. 1.12), DNA is under-wound ( $< 10.5$  bp per helical turn), *i.e.* negatively supercoiled. The most common measurement for SC is the SC density  $\sigma$  which is normalised by DNA length, defined as  $\sigma = \frac{Lk - Lk_0}{Lk_0}$ . Importantly, *in vivo*, the twist/writhe partition in a supercoiled DNA is locally governed by many factors, including DNA sequence and protein-binding pattern, precluding the establishment of general predictive rules for its determination. To avoid confusion, it is noteworthy to mention that, by convention, microbiologists indicate an increase in absolute SC level (DNA relaxation) by a "-" sign, whereas a decrease (DNA overtwisting) is indicated by a "+" sign. This convention is used in Chapter 4, whereas in the upcoming subsection (review) and elsewhere in the thesis, we rather indicate an increase in absolute SC level (DNA relaxation) by a "+" sign, and a decrease in absolute SC level (DNA overtwisting) by a "-" sign.

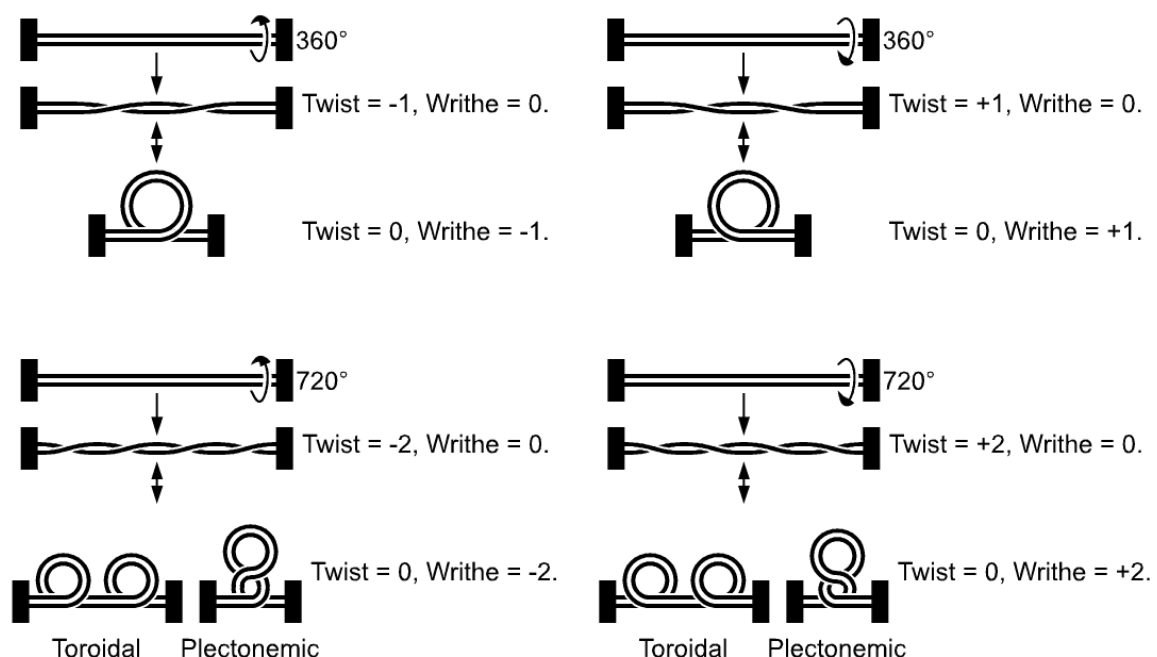


FIGURE 1.12: Copied from Wikipedia: twist and writhe in a supercoiled linear DNA molecule with constrained ends.

SC can induce structural transitions in DNA [86]. *In vivo*, DNA exists primarily in the standard right-handed B-form helix, although it can assume many other conformations characterised by different structural parameters [87]. Most of these alternative DNA structures are formed by specific DNA sequences (usually of repeated/symmetry nature), and are induced by various factors, including changes of temperature, ionic conditions, hydration, and SC [87]. Indeed, negative SC imposes DNA unwinding ( $Tw$  decrease), and as a consequence, to relieve the torsional stress, local changes in DNA structure with less twist than B-DNA may become energetically favourable. These alternate structures are reviewed elsewhere [85, 87, 88], and include the A-form, Z-form, H-form, cruciform, S-form, and G-quadruplex structure (Fig. 1.13). More details are provided in the next sections.

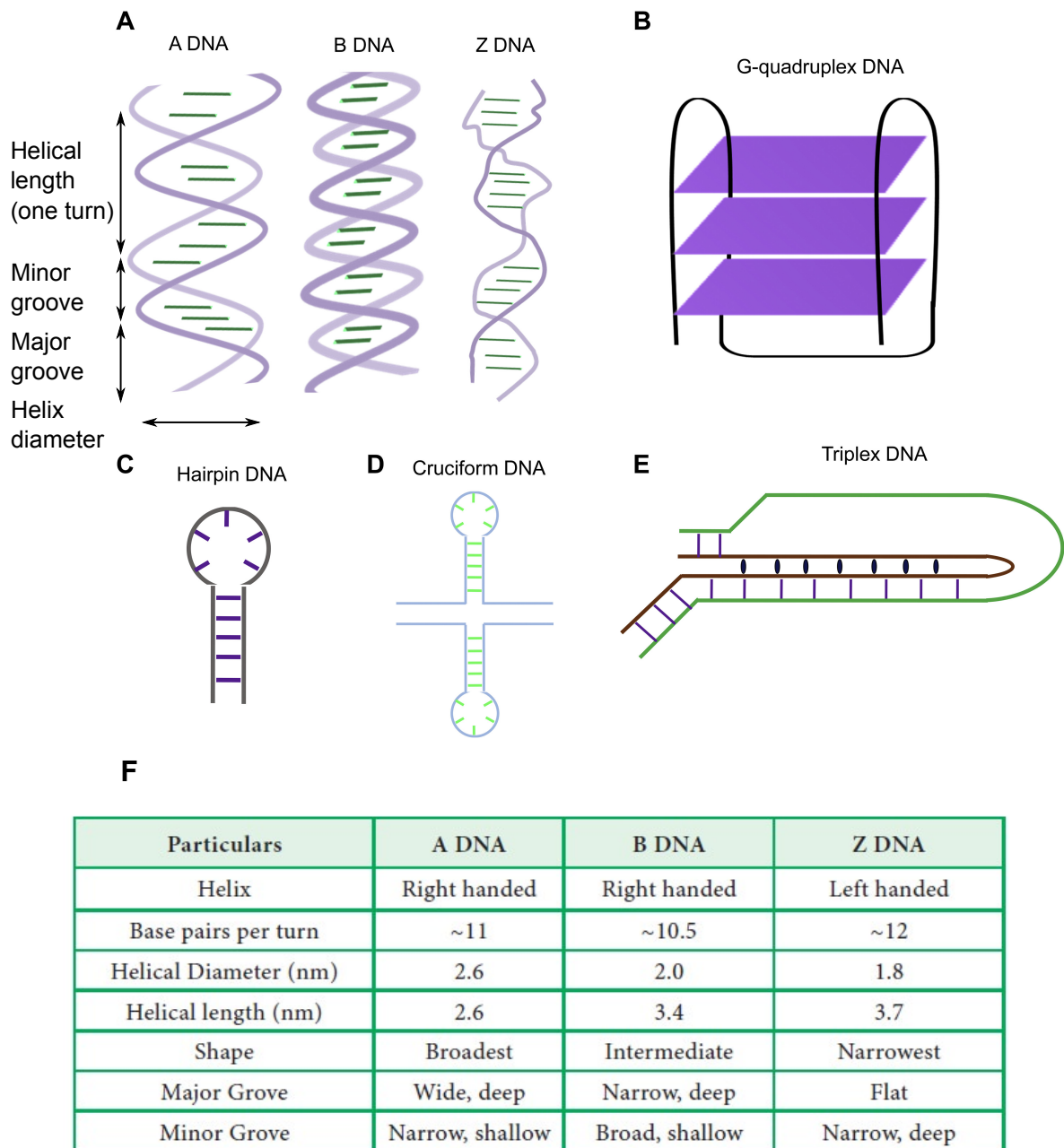


FIGURE 1.13: Adapted from [89]: different types of non-B DNA structures. **(A)** A-, B-, and Z-DNA. **(B)** G-quadruplex DNA formed by stretches of guanines, that assemble into a quartet structure (purple squares) connected by a loop DNA. **(C)** Hairpin structure formed by inverted repeats on a single-stranded DNA. **(D)** Cruciform structure formed by inverted repeats on a double-stranded DNA. **(E)** Triplex DNA formed by hydrogen bonding between a third strand with the duplex DNA. **(F)** Copied from Brainkart: properties of A-, B-, and Z-DNA.

### 1.3.2 DNA topoisomerases

SC is thus a fundamental property of DNA, and as a consequence, cells have developed a variety of enzymes, referred to as DNA topoisomerases, to solve the topological problems associated with DNA transactions. Topoisomerases (abbreviated "Topo" hereafter) are present in all domains of life and are highly conserved [90]. They modulate DNA linking number by catalysing the cleavage and the re-ligation of one (class I) or two (class II) DNA strand(s) [91]. Based on studies conducted in the model bacterium *E. coli*, the main bacterial topoisomerases include two class I topoisomerases (TopoI and III), and two class II topoisomerases (DNA gyrase and TopoIV) (Tab. 1.4) [92]. TopoI and TopoIII are closely related monomers (Fig. 1.14), that both relax negative supercoils, and decatenate single-stranded DNA (ssDNA) rings formed during recombination and replication. TopoIII is  $Mg^{2+}$ -dependent, while TopoI is activated by the latter, even though it does not require it [90]. Due to their slight structural differences (Fig. 1.14), TopoI is more efficient than TopoIII at relaxing negative supercoils, while TopoIII is more efficient at decatenating ssDNA rings [93]. As a consequence, TopoI is essentially involved in the regulation of SC homeostasis, while TopoIII ensures DNA decatenation. In contrast to TopoI, TopoIV, and DNA gyrase, TopoIII is present in a limited number in the cell and is not essential for bacterial viability [92].

Topoisomerase	Class	Gene	Sub-units	Function
Topoisomerase I	I	<i>topA</i>	$\alpha$	<b>Relaxation of (-) supercoils</b> DNA decatenation
Topoisomerase III	I	<i>topB</i>	$\alpha$	<b>DNA decatenation</b> Relaxation of (-) supercoils
DNA gyrase	II	<i>gyrA</i> ( $\alpha$ ) <i>gyrB</i> ( $\beta$ )	$\alpha_2\beta_2$	<b>Introduction of (-) supercoils</b> DNA decatenation Relaxation of (+/-) supercoils
Topoisomerase IV	II	<i>parC</i> ( $\alpha$ ) <i>parE</i> ( $\beta$ )	$\alpha_2\beta_2$	<b>DNA decatenation</b> Relaxation of (+/-) supercoils

TABLE 1.4: Features of main bacterial topoisomerases, which catalyse the cleavage and the re-ligation of one (class I) or two (class II) DNA strand(s). Their main function *in vivo* is highlighted in **bold**.



Among the class II Topos, DNA gyrase is a heterotetramer with two GyrA and two GyrB sub-units (Tab. 1.4 and Fig. 1.14), essential for bacterial viability, present in some archaea and in the organelles (chloroplasts and mitochondria) of some plants (more detail in Chapter 6), but absent from humans [91]. Its primary role is to introduce negative supercoils in DNA in a  $Mg^{2+}$  / ATP-dependent manner. It is also able, to a lower extent, and with less efficiency than TopoIV, to decatenate DNA, relax positive supercoils (also in a  $Mg^{2+}$  / ATP-dependent manner), and relax negative supercoils (in a  $Mg^{2+}$  / ATP-independent manner) [91]. TopoIV is also a heterotetramer with two ParC and two ParE sub-units (Fig. 1.14), which both relax positive/negative supercoils (with more efficiency on positive ones) and decatenate DNA, in a  $Mg^{2+}$  / ATP-dependent manner [91]. Its main role is to remove precatenane linkages after DNA replication, to allow DNA segregation [94]. The mechanisms of action of class I and II topoisomerases are presented in Fig. 1.15.

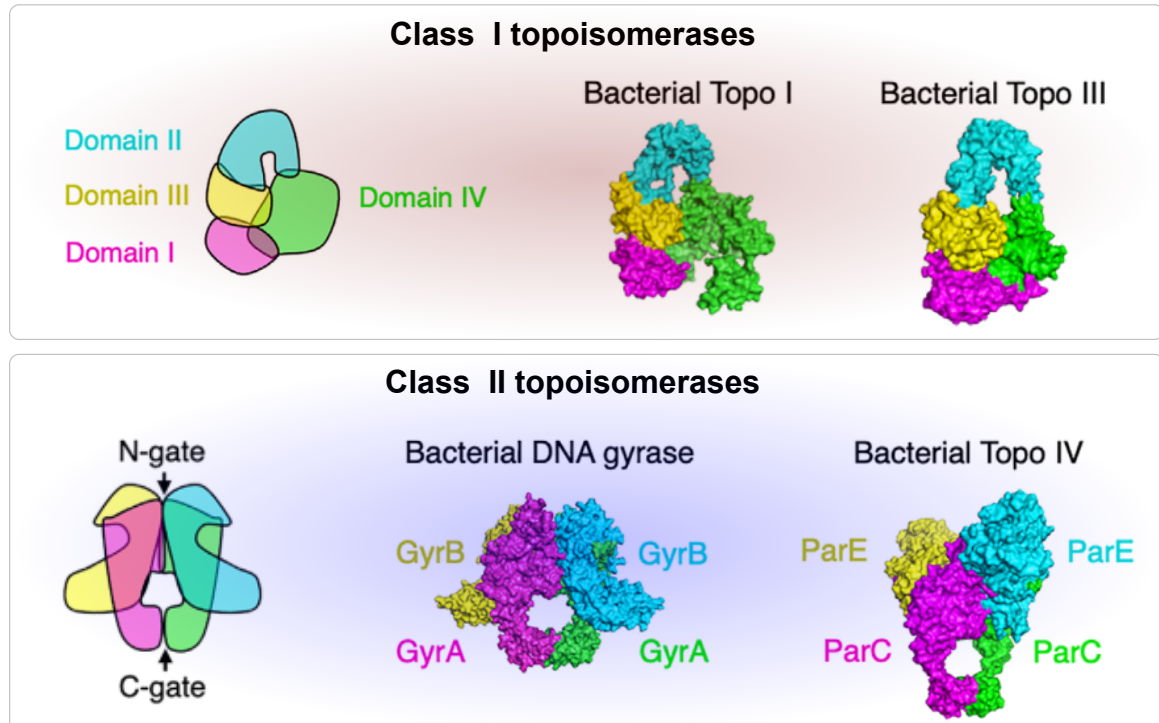


FIGURE 1.14: Adapted from [92]: X-ray structures of the main bacterial topoisomerases depending on their class (I or II), and for which a general schematic is provided. TopoI and III are monomers, while DNA gyrase and TopoIV are heterotetramers.



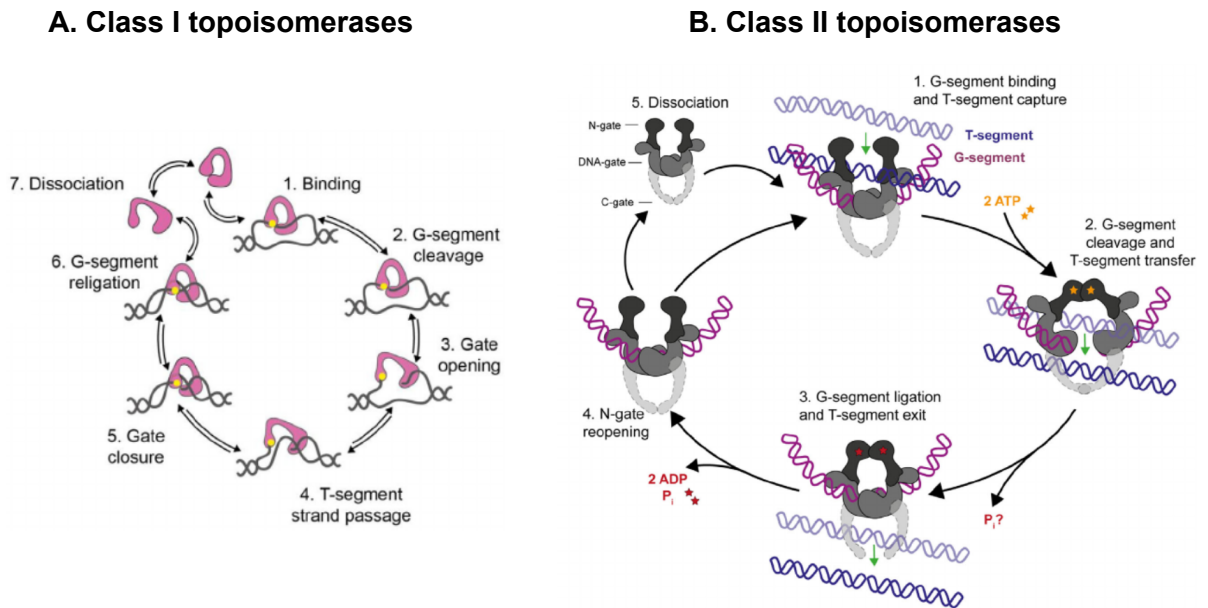


FIGURE 1.15: Adapted from [91]: strand-passage mechanisms of class I and II topoisomerases. **(A)** Class I topoisomerases: (1) bind a single DNA strand of the underwound duplex, known as the G (gate)-segment, (2) cleave it, (3) open their DNA-gate, (4) transfer the other strand, the T (transported)-segment, through the cleaved G-segment, (5) close their DNA gate, (6) re-ligate the G-segment. This induces a change in  $Lk$  by one. **(B)** Class II topoisomerases: (1) bind the G-segment at the DNA-gate and capture the T-segment, (2) cleave the G-segment while the T-segment is passed through the break, by dimerisation of the N-gate through ATP-binding, (3) re-ligate the G-segment while the T-segment exits through the C-gate, (4) open their N-gate through dissociation of ADP and  $P_i$ , and either capture a consecutive T-segment (1) by remaining bound to the G-segment, or (5) dissociate from the G-segment. In the case of DNA gyrase, the passage of the T-segment through the G-segment induces a decrease in  $Lk$  by two.

In bacterial cells, the level of SC is finely controlled mainly by the opposite activities of relaxing TopoI (and IV, to a lower extent), and negative supercoil-inducing DNA gyrase. Bacterial chromosomes are generally maintained in an underwound (negatively supercoiled) state by the extensive action of DNA gyrase, with an average of  $\sigma = -0.06$  for exponentially-growing *Escherichia coli* cells [88]. This both facilitates DNA melting required for transcription and replication, and contributes to genome compaction, with the combined action of NAPs [15]. Indeed, negative SC of bacterial chromosomes leads to the formation of plectonemes and solenoids through writhe

changes, that can be either in a free form, or constrained in topological domains with variable size (average of 10 kb) [95, 96] by interaction with proteins such as NAPs (Fig. 1.16) [97]. In *E. coli*, the effective (unconstrained) superhelicity was estimated about a half of the total as an average value [88, 98]. Beyond this direct regulation of SC, NAPs exert an indirect regulatory effect, by modulating the transcription of the genes encoding Topos, or by modulating their activity [15].

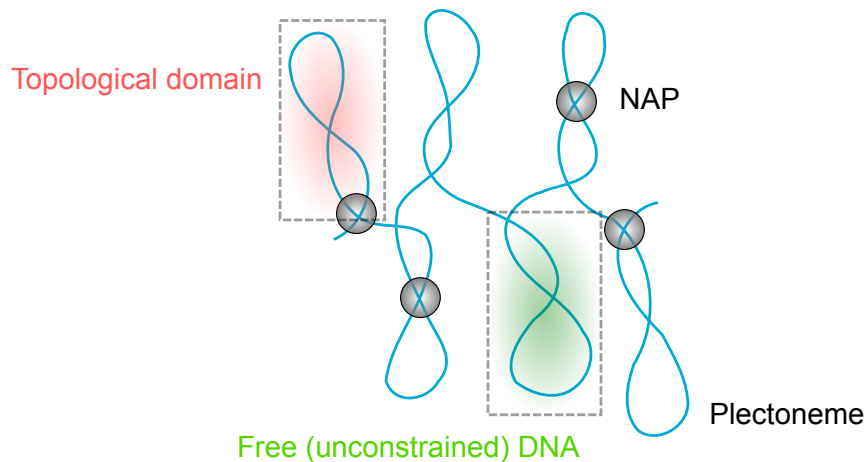


FIGURE 1.16: Adapted from [15]: division of bacterial chromosomes into supercoiled topological domains. Negative SC gives rise to plectonemes, half of which are present in a free, unconstrained form [98], whereas proteins such as NAPs (nucleoid-associated proteins) restrain the remaining half in independent topological domains. Only a fraction of the chromosome is shown for simplicity.

### 1.3.3 DNA supercoiling measurement

The level of SC can be measured using a variety of approaches. Historically, one of the first methods used was the titration of SC density by sedimentation in ethidium bromide sucrose density gradients, based on the faster sedimentation of supercoiled DNA molecules than relaxed ones due to their higher compaction (Fig. 1.17A). This allows to separate topoisomers, *i.e.* DNA molecules that are chemically identical but differ by their Lk [85]. However, such approaches are precise yet laborious, and were consequently replaced by electrophoretic methods, based on the faster migration of supercoiled DNA molecules than relaxed ones through an agarose gel. Due to the low resolution of standard agarose gels preventing the separation of highly supercoiled topoisomers, electrophoresis is performed in the presence of an intercalating agent, usually chloroquine, which decreases the migration rates of highly supercoiled DNA molecules (Fig. 1.17B-C). This allows to observe distinct DNA bands that correspond to topoisomers differing in their Lk [99]. An example is provided in Chapter 3 (Fig. 3.1). Nevertheless, one-dimensional electrophoresis does not allow the separation of complex mixtures including both positively and negatively supercoiled topoisomers, and the latter are generally analysed using two-dimensional agarose gel electrophoresis (Fig. 1.17D). In such an approach, a first separation is performed in a standard agarose gel, and co-migrating topoisomers of the same yet opposite Lk are further separated by a second migration in a direction perpendicular to the first one, in presence of chloroquine [85]. It is important to note that with such methods, the SC level of the chromosome is difficult to measure, and thus, its SC density is generally obtained from a plasmid population reflecting the average chromosomal SC level, which may introduce bias in some cases [88]. It is noteworthy to mention recently developed methods such as recombination assays [100], or fluorescent evaluation of DNA supercoiling (FEDS) [101], which relies on a reporter gene exclusively regulated by SC.

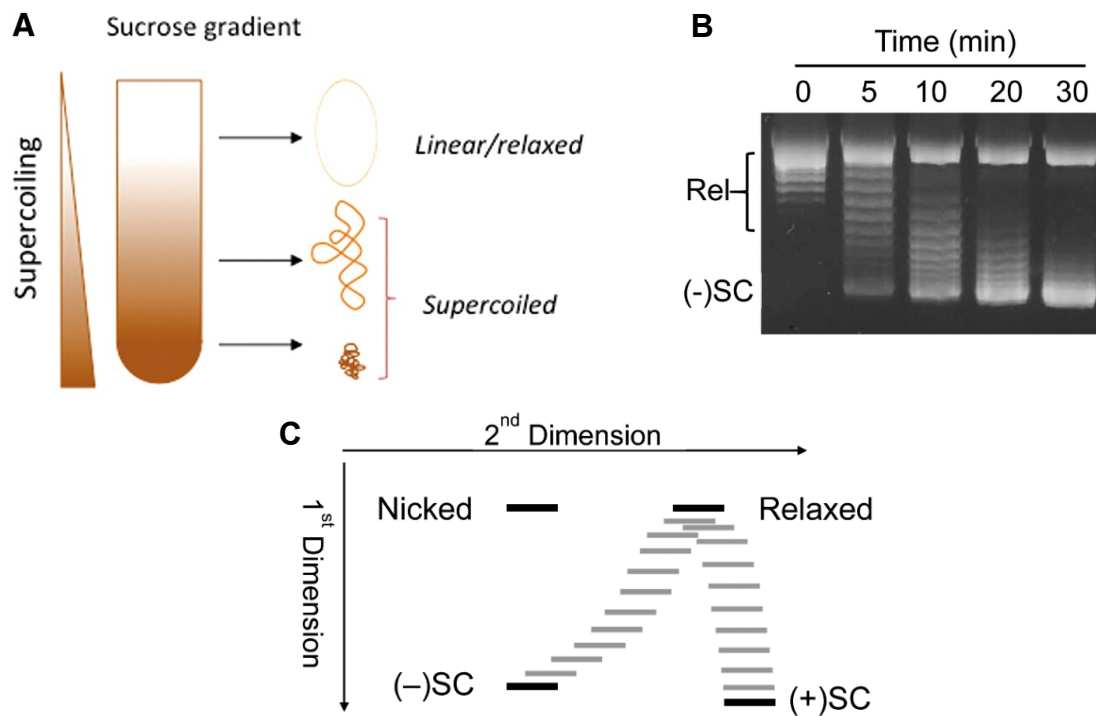


FIGURE 1.17: Titration of SC density by sedimentation in ethidium bromide sucrose density gradients, and by one- or two-dimensional agarose gel electrophoresis. **(A)** Copied from [102]: sedimentation in ethidium bromide sucrose density gradients, wherein supercoiled DNA molecules sediment faster than relaxed ones. **(B)** Copied from [103]: one-dimensional agarose gel electrophoresis. Relaxed plasmids are converted to negatively supercoiled molecules over time by *S. aureus* DNA gyrase. DNA bands were visualised by ultraviolet light after staining with ethidium bromide. **(C)** Copied from [103]: principle of the two-dimensional agarose gel electrophoresis in presence of chloroquine. Gray bands represent DNA topoisomers of intermediate SC level.

While the above methods are useful to determine the average SC level of circular DNA, they do not provide information (i) on the shape of the supercoiled molecules, and (ii) on the local SC levels. Electron microscopic techniques have thus been developed to address the first point, even if in such approaches, DNA conformation may change during sample preparation, and the sign of the crossing remains difficult to determine [85]. For the second point, *in vivo*, the local SC levels can be measured based on the rate of psoralen photobinding, which intercalates into DNA and preferentially binds underwound regions of the double-helix [85, 102]. This approach has been widely used in eukaryotes [102], and also in *E. coli*, in which the extent of psoralen crosslinking along its chromosome was measured by fragmenting DNA and hybridising DNA fragments to microarrays covering its entire

genome [104].

Beyond its role in genome compaction, SC level was soon discovered to vary in response to environmental changes, and to affect the expression of many promoters [105]. More recently, through the advent of transcriptomic technologies allowing genome-wide expression profiling, SC was shown to act as a global transcriptional regulator in many bacteria. Fast changes in SC levels may then contribute to adapting gene expression program of bacteria to environmental changes [106, 107]. In the following review published in the Computational and Structural Biotechnology Journal (CSBJ) [108], we compile (i) available evidence regarding the modulation of the SC level in response to environmental changes, (ii) available transcriptomic studies demonstrating the global regulatory effect of SC variations in distant bacteria, (iii) available mechanisms and computational models of transcriptional regulation by SC. We finally discuss the applications to pathogenic bacteria facing many environmental changes during their infection process, where SC may play a key role.

### 1.3.4 Review



journal homepage: [www.elsevier.com/locate/csbj](http://www.elsevier.com/locate/csbj)



Mini Review

## DNA Supercoiling: an Ancestral Regulator of Gene Expression in Pathogenic Bacteria?

Shiny Martis B.<sup>1</sup>, Raphaël Forquet<sup>1</sup>, Sylvie Reverchon, William Nasser, Sam Meyer\*

Université de Lyon, INSA Lyon, Université Claude Bernard Lyon 1, CNRS UMR5240, Laboratoire de Microbiologie, Adaptation et Pathogénie, 11 avenue Jean Capelle, 69621 Villeurbanne, France

ARTICLE INFO

Article history:

Received 7 June 2019  
 Received in revised form 17 July 2019  
 Accepted 24 July 2019  
 Available online 26 July 2019

Keywords:

DNA supercoiling  
 Transcription  
 Pathogenesis  
 Genetic regulation

ABSTRACT

DNA supercoiling acts as a global and ancestral regulator of bacterial gene expression. In this review, we advocate that it plays a pivotal role in host-pathogen interactions by transducing environmental signals to the bacterial chromosome and coordinating its transcriptional response. We present available evidence that DNA supercoiling is modulated by environmental stress conditions relevant to the infection process according to ancestral mechanisms, in zoopathogens as well as phytopathogens. We review the results of transcriptomics studies obtained in widely distant bacterial species, showing that such structural transitions of the chromosome are associated to a complex transcriptional response affecting a large fraction of the genome. Mechanisms and computational models of the transcriptional regulation by DNA supercoiling are then discussed, involving both basal interactions of RNA Polymerase with promoter DNA, and more specific interactions with regulatory proteins. A final part is specifically focused on the regulation of virulence genes within pathogenicity islands of several pathogenic bacterial species.

© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1. Introduction . . . . .	1047
2. DNA Supercoiling: A Global Regulator of Bacterial Gene Expression . . . . .	1048
2.1. DNA Supercoiling: A Relay of Environmental Signals to the Bacterial Chromosome . . . . .	1048
2.2. Global Transcriptional Response to Variations in DNA Supercoiling . . . . .	1049
3. Mechanisms and Models of Transcriptional Regulation by DNA Supercoiling . . . . .	1049
3.1. Basal Regulation of the RNA Polymerase-DNA Interaction . . . . .	1049
3.2. Specific Regulation Involving DNA Regulatory Proteins. . . . .	1050
3.3. Spatial Heterogeneities of DNA Supercoiling: The Transcription-Supercoiling Coupling . . . . .	1051
4. DNA Supercoiling and the Coordination of Virulence Programs . . . . .	1052
4.1. An Argument for DNA Supercoiling Being an Important Actor in Virulence Genetic Regulation . . . . .	1052
4.2. Widespread Evidence for a Regulatory Role of DNA Supercoiling in Virulence . . . . .	1052
5. Conclusion . . . . .	1053
Acknowledgements . . . . .	1053
References . . . . .	1053

### 1. Introduction

DNA supercoiling (SC) has received considerable attention in recent years as a global and ancestral actor in genetic regulation. This is especially conspicuous in bacteria [1–3], where the chromosome is maintained at an out-of-equilibrium level of negative SC by a finely controlled balance of topoisomerase activity. And yet, in contrast to

Abbreviations: SC, DNA supercoiling; TF, transcription factor; NAP, nucleoid-associated protein.

\* Corresponding author.

E-mail address: [sam.meyer@insa-lyon.fr](mailto:sam.meyer@insa-lyon.fr) (S. Meyer).

<sup>1</sup> Equal first authors.

<https://doi.org/10.1016/j.csbj.2019.07.013>

2001-0370/© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



classical regulation based on transcription factors, quantitative models of the regulatory mechanisms by SC are essentially lacking. A possible explanation for this shortcoming is that SC affects transcription at several stages of the process, and can also be involved in various and complex interactions with regulatory proteins. As a result, virtually every investigated promoter exhibits a distinct SC response, making it difficult to dissect and model the underlying mechanisms. In this review, we wish to summarise existing evidence and models suggesting a widespread role of SC in bacterial genetic regulation, and more specifically in bacterial virulence. This topic has already been addressed in previous extensive reviews focused either on its role in bacterial growth [4] or on specific promoters that were analysed in detail [5]. Here, we propose a complementary focus on proposed mechanistic and computational models of transcriptional regulation by SC as well as accumulating information obtained from transcriptomic data, which together underline the broad relevance of the investigated phenomenon in bacterial virulence and call for a combined experimental-theoretical research effort.

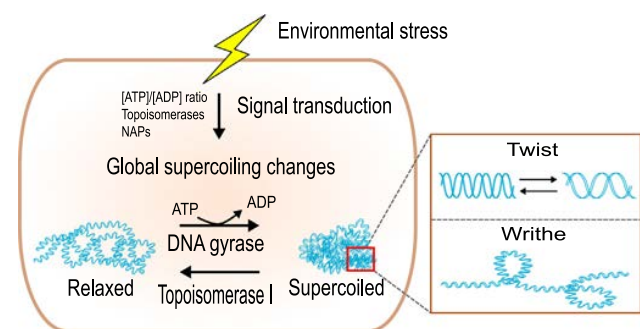
## 2. DNA Supercoiling: A Global Regulator of Bacterial Gene Expression

### 2.1. DNA Supercoiling: A Relay of Environmental Signals to the Bacterial Chromosome

As observed immediately following the discovery of the double-helical structure of DNA, virtually all DNA transactions face substantial topological constraints [6]. In mechanical terms, the latter give rise to a ubiquitous torsional stress, which in turn results in DNA supercoiling (SC), i.e., the deformation of the molecule either by rotation around its helical axis (over- or under-twisting) or by the winding of this helical axis itself (writhing), as illustrated in Fig. 1 [4,7].

Topoisomerases are the global regulators of SC and more generally, the solvers of topological problems associated with DNA transactions [8]. In bacteria, the two main topoisomerases are topoisomerase I (topo I) and DNA gyrase. The latter maintains the chromosomal DNA in an underwound state by introducing negative supercoils in an ATP-dependent manner, while conversely, topo I relaxes the DNA (i.e. removes negative supercoils) without any ATP requirement. The global negative SC level of the chromosome is thus primarily determined by the dynamic equilibrium between these two enzymes (Fig. 1). Additional actors play a more specific role: the ATP-consuming topoisomerase IV is primarily involved in solving topological problems associated with DNA replication and cell division [9], and abundant nucleoid-associated proteins (NAPs) contribute in distributing SC along the bacterial chromosome [4,10].

The negative SC level of the chromosome is finely controlled by the cell in response to environmental conditions, since almost all types of environmental challenges have been associated with SC variations, and in particular those most commonly encountered by pathogens



**Fig. 1.** DNA supercoiling acts as a sensor of environmental stress in the bacterial chromosome. Environmental cues are transduced by different mechanisms into global, stress-specific variations of the SC level. At a smaller scale, SC is distributed as twist and writhe deformations, which directly affect the transcriptional activity.

**Table 1**

Chromosomal supercoiling response to environmental stress conditions is conserved in distant bacterial species. Phyla: P: Proteobacteria, F: Firmicutes, A: Actinobacteria. SC variations: Rel (+): relaxation, Hyp (-): hyper-supercoiling.

Shock	Phylum	Species	SC change	Mechanism	Ref	
Heat	P	<i>Escherichia coli</i>	Rel (+)	Gyrase and topol activities Gyrase activity decrease	[12]	
		<i>Yersinia enterocolitica</i>			[13]	
		<i>Dickeya dadantii</i>			[14]	
	F	<i>Bacillus subtilis</i>	[15]			
	Cold	P	<i>Escherichia coli</i>	Hyp (-)	Gyrase activity increase / HU	[16]
F		<i>Bacillus subtilis</i>	[15]			
Acidic	P	<i>Escherichia coli</i>	Rel (+)	Gyrase activity decrease	[17,18]	
		<i>Salmonella typhimurium</i>			[18,19]	
		<i>Dickeya dadantii</i>			[20]	
Osmotic	P	<i>Escherichia coli</i>	Hyp (-)	[ATP]/[ADP] increase	[21]	
		<i>Salmonella typhimurium</i>			[22]	
		<i>Dickeya dadantii</i>			[23]	
		<i>Bacillus subtilis</i>			[15]	
		<i>Staphylococcus aureus</i>			[24]	
	A	<i>Streptomyces lividans</i>	[25]			
	Oxidative	P	<i>Escherichia coli</i>	Rel (+)	TopA activation by Fis	[26]
			<i>Dickeya dadantii</i>			[23]
	Anaerobic	P	<i>Escherichia coli</i>	Rel (+)	[ATP]/[ADP] decrease	[27]
			<i>Salmonella typhimurium</i>			[28]
<i>Bacillus subtilis</i>			[15]			
F		<i>Bacillus subtilis</i>	[15]			

during infection (Table 1). Depending on the applied stress, the chromosomal DNA experiences either a partial relaxation (+) or an increase in negative SC (-), which is usually rapid and transient. Importantly, in spite of strong differences in terms of phylogeny or lifestyle, the response to each specific stress is qualitatively similar in all investigated species, although these exhibit quantitatively different SC levels in standard growth conditions [3]. This observation suggests that SC is used in a wide range of bacteria to quickly transduce environmental signals toward the chromosome, with ancestral control mechanisms. Interestingly, environmental stresses have also been correlated with SC changes in archaeal species [11], suggesting that this notion could be extended to an even wider range of microorganisms.

What are the underlying mechanisms? The most clearly described pathway involves the modulation of gyrase activity by the energy charge of the cell through the [ATP]/[ADP] ratio [21,27,29]. When the latter is increased, gyrase introduces supercoils in the chromosomal DNA more actively and its negative SC level increases. Environmental stresses usually alter cellular metabolic fluxes and the energy charge in various ways [30], and this relatively simple, quick and general mechanism is indeed involved in the chromosomal response to a variety of conditions (Table 1). Other invoked mechanisms include (1) the action of NAPs (e.g. HU, FIS) either by direct interaction with DNA or as modulators of topoisomerase activity; (2) the regulation of topoisomerase expression (*topA*, *gyrA* and *gyrB* genes encoding topol and gyrase subunits respectively); (3) the modulation of topoisomerase activity through post-translational modifications [31].

It should be underlined that most bacteria from Table 1 are pathogenic or have pathogenic serotypes. Since the infection process can be assimilated to successive environmental changes to which the pathogen needs to adapt quickly, SC appears as a general candidate for the efficient transmission of stage-specific environmental signals toward the bacterial chromosome and thus also for its

transcriptional response, which in turn is critical for the subsequent steps of the infection.

### 2.2. Global Transcriptional Response to Variations in DNA Supercoiling

The transcriptional response to various stress conditions can be readily analysed from transcriptomic data, but it is then difficult to disentangle (1) the generic effect of the stress-induced SC variation on transcription and (2) the transcriptional effect of stress-specific pathways. The former contribution can however be analysed separately in transcriptomes obtained with gyrase inhibiting antibiotics. These are either aminocoumarins (novobiocin, coumermycin) which block the ATPase activity of class-II topoisomerases (gyrase and topoisomerase IV), or quinolones (ciprofloxacin, norfloxacin, nalidixic acid, oxolinic acid) which block their ligase activity, the latter resulting in many double-strand breaks and triggering a SOS response of the cell with pleiotropic effects [32]. When applied at a sublethal dosage, these drugs induce a sudden global relaxation of the chromosomal DNA, and the transcriptional response is then measured after a short time (usually 5–30 min), assuming that the latter then mostly reflects the direct effect of SC, rather than indirect effects influenced by the cell's response. Such data were obtained in many organisms (Table 2). Note that the reported number of affected genes is strongly variable; this variability might partly reflect actual differences between organisms, but is strongly affected by the experimental conditions and methods (relaxation level, transcriptomics technology, statistical analysis). Altogether, these data consistently demonstrate a very broad response to chromosomal relaxation, with a significant effect on more than one quarter of the genes. This response is complex, with some genes being upregulated and others downregulated. These affected genes are functionally diverse, including genes involved in essential functions (e.g. DNA replication, cell division), stress responses and metabolic pathways (e.g. stringent response, DNA repair pathway), as well as virulence. They are also usually scattered throughout the chromosome, highlighting that SC-mediated regulation acts in a global way, but follows a spatial organisation pattern involving large-scale responsive domains related to structural properties of DNA [20,33].

Since this global transcriptional response is observed in a wide range of species from different phyla (Table 2), SC might be considered as an ancestral and widespread mode of regulation in bacteria. This notion can be related to the fundamental and highly conserved character of topoisomerase enzymes themselves [8], and may even be extended to eukaryotes, albeit with different rules [7,34]. It does not mean however that the mechanism is identical in all bacteria. The longest-running evolution experiment [35] emphasized that mutations affecting SC

constitutes a “quick and efficient” way to modify the global expression pattern and gain substantial fitness, in this case by a mutation reducing topo I efficiency in less than 2000 generations [36]. It is therefore no surprise that fluctuations in topoisomerase structure and SC level were pointed in the close relatives *E. coli* and *S. typhimurium* [3] which have different lifestyles, and this is probably also frequent in different strains of the same species [19]. Changing the chromosomal SC level is thus a fundamental and generic way by which bacteria adapt to new environmental challenges, according to common ancestral rules. This extends in particular to genome-reduced bacteria almost devoid of TFs such as *Mycoplasma* or *Buchnera* [37–39], where transcriptional regulation remains poorly understood.

### 3. Mechanisms and Models of Transcriptional Regulation by DNA Supercoiling

The abovementioned transcriptional responses induced by SC variations differ qualitatively from those induced by classical transcriptional factors (TFs). The latter recognize, bind and regulate a specific subset of the genome defined by a well-defined (although often degenerate) target sequence motif, and their action can be modelled using classical thermodynamic models of activation or repression [55]. In contrast, as noted above, SC affects the transcriptome of all investigated species globally, without any identified promoter sequence determinant. And yet strikingly, regulatory models comparable to those involving TFs are essentially lacking. There are two reasons for this: first, experimentally, SC regulates gene activity in a continuous “more or less” manner as opposed to the stronger “on or off” mode of regulation by TFs [56]; second, SC can modulate transcription in a variety of ways, making them difficult to decipher. In the following, we discuss such mechanisms, most of which will be illustrated on the promoter of *pelle*, one of the major virulence genes of the phytopathogen *Dickeya dadantii* where SC-mediated regulation was studied extensively.

#### 3.1. Basal Regulation of the RNA Polymerase-DNA Interaction

The ancestral and global mode of regulation by SC results, in the first instance, from it affecting the interaction between DNA and RNA Polymerase (RNAP) itself, independently from any additional regulatory protein. But this basal regulation already involves distinct mechanisms occurring at successive steps of the complex process of bacterial transcription: closed-complex formation, open-complex formation, promoter clearance, and elongation [57].

The most clearly identified – and possibly strongest – effect of SC on transcription initiation results from the requirement for RNAP to open

**Table 2**

Transcriptomic response to variations of DNA supercoiling in bacteria. Phyla: P: Proteobacteria, F: Firmicutes, A: Actinobacteria, T: Tenericutes, C: Cyanobacteria. SC variations: Rel (+): relaxation, Hyp (–): hypersupercoiling. Transcriptomics technology: M: DNA Microarray, S: RNA Sequencing

Phylum	Species	SC change	Method	Genes significantly affected (% genome)	Technology	Ref	
P	<i>Escherichia coli</i>	Rel (+)	Norfloxacin	613 (15%)	M	[40]	
		Rel (+)	Novobiocin / pefloxacin	1957 (48%)	M	[41]	
		Rel (+)	Genetic engineering	740 (18%)	M	[42]	
	<i>Salmonella typhimurium</i>	Rel (+)	Genetic engineering / norfloxacin / novobiocin	306 (7%)	M	[43]	
		Rel (+)	Genetic engineering	499 (10%)	M	[44]	
		<i>Dickeya dadantii</i>	Rel (+)	Novobiocin	1461 (32%)	M	[20]
			Rel (+)		1212 (27%)	S	[45]
	<i>Haemophilus influenzae</i>	Rel (+)	Novobiocin / ciprofloxacin	640 (37%)	M	[46]	
		Rel (+)	Novobiocin	290 (14.2%)	M	[47]	
	F	<i>Streptococcus pneumoniae</i>	Hyp (–)	Seconeolitsin	545 (27%)	S	[48]
Rel (+)			Novobiocin	280 (11%)	M	[49]	
<i>Bacillus subtilis</i>		Rel (+)		1075 (24%)	M	[50]	
		<i>Streptomyces coelicolor</i>	Rel (+)	Novobiocin	121 (1.5%)	S	[51]
<i>Mycobacterium tuberculosis</i>	Rel (+)			Not provided	M	[52]	
T	<i>Mycoplasma pneumoniae</i>	Rel (+)	Novobiocin	469 (43%)	S	[37]	
C	<i>Synechocystis</i>	Rel (+)	Novobiocin	Several genes	M	[53]	
	<i>Synechococcus elongatus</i>	Rel (+)		Not provided	M	[54]	



the DNA strands and stabilize a “transcription bubble”, in order to gain access to the DNA bases in the template strand. In torsionally unconstrained DNA, this melting transition represents a substantial free energy cost of around 10  $k_B T$  (6 kcal/mol), which in eukaryotes is provided through ATP hydrolysis by the basal transcription factor TFIID [34]. Crucially, this cost reduces drastically when the double helix is destabilised by negative torsion at SC levels physiologically relevant in bacteria (Fig. 2A, upper panel), thus providing the physical basis for the bacterial transcription process that does not require any external energy consumption [57]. Interestingly, the same can become true of eukaryotic RNA Polymerase II when operating on a comparably negatively supercoiled template [34], whereas temperature may replace SC as the source of melting energy in thermophilic archaea relying on reverse gyrase [58,59]. This strong regulatory effect of SC can be observed in well-controlled in vitro transcription assays (Fig. 2A, lower panel), where plasmids carrying a model promoter are prepared at different supercoiling levels, resulting in a drastic variation of expression strength without any modification of the DNA sequence or regulator concentration. This activation curve is quantitatively reproduced [45] using thermodynamic models of DNA opening [60–62] relying on knowledge-based enthalpic and entropic parameters for base-pairing and stacking interactions of all base sequences.

How may this mechanism lead to transcriptional regulation, i.e., the selective activation of a subset of promoters by global SC variations? The opening curve is strongly dependent on the promoter base sequence (mostly though its GC content) in the  $-10$  region and that immediately downstream referred to as the “discriminator” [4,63–65]. Remarkably, although the latter region does not harbour any consensus binding signal for RNAP, mutation studies showed that it plays a predominant role in the SC-sensitivity of promoters: GC-rich discriminators are typically more activated by negative SC than AT-rich ones, the canonical example being those of stable RNAs strongly induced during exponential growth [4,64,65]. And yet, a systematic model of this regulation mechanism is still lacking. A possible obstacle is that the thermodynamic description used above might be insufficient, if the expression level of investigated promoters is limited not by the rate of promoter opening but of promoter escape. As an example, some promoters might attract RNAP and easily form an open-complex, but still exhibit low expression levels

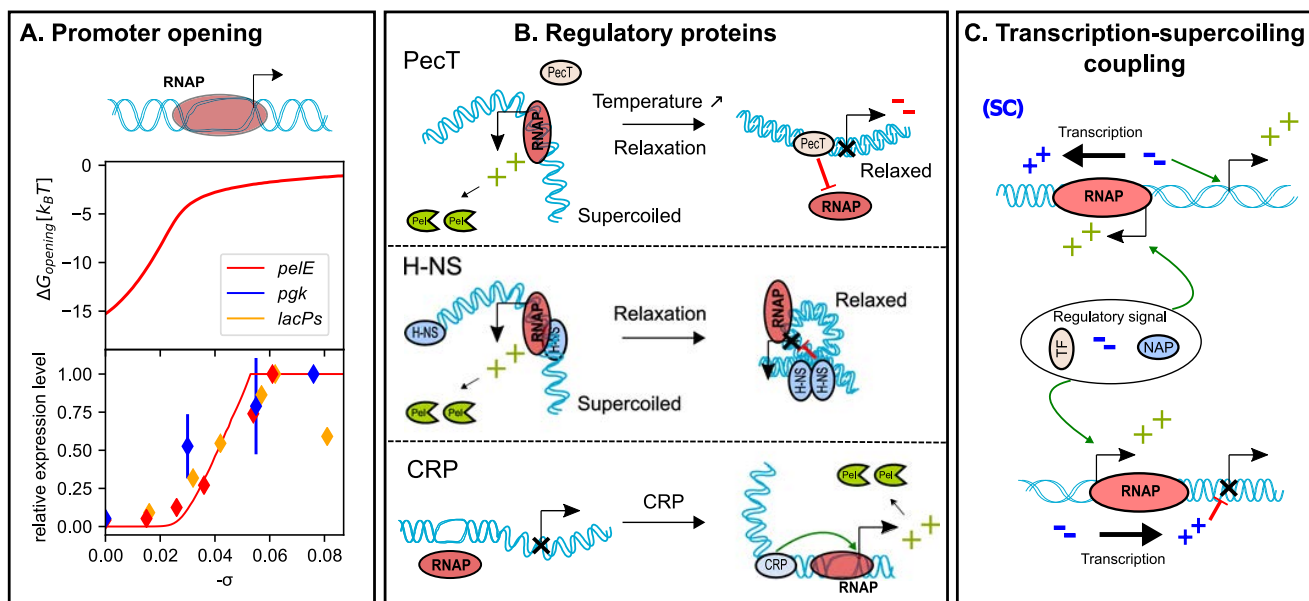
if the latter is too stable, resulting in abortive rather than processive transcription [4,66,67]. Since the influence of SC on these subtle kinetic steps was not dissected in detail, little can be predicted from such a scenario that may explain why promoters respond differently to SC variations [68], and some even in opposite ways [69–71]. Yet we still note that many in vitro investigated promoters do follow the behaviour expected from thermodynamic modelling (Fig. 2A, lower panel), which might thus account for at least a significant fraction of transcriptomic responses to SC variations.

A regulatory action of SC was proposed at least at two other transcription steps. RNAP binds promoters by recognising the  $-10$  and  $-35$  elements, which are separated by a spacer of variable size (15–20 nt, with an optimum at 17 nt). Because of the helical structure of DNA, these variations are associated to different relative angular positions of the recognition sites around the helical axis, possibly placing them out-of-phase for RNAP binding and closed-complex formation. A suitable (spacer length-dependent) level of torsional stress could then modulate this binding rate by untwisting the spacer DNA toward a favourable orientation [73,74]. This scenario was invoked to explain the response of several promoters to SC variations, including ribosomal RNAs which usually exhibit a suboptimal spacer length of 16 nt [73].

SC may also affect transcription at the elongation stage, where positive torsion hinders the progress of RNAP, as demonstrated from single-molecule experiments [75]. Such positive SC levels are not observed in vivo by usual techniques involving plasmids, but might be transiently generated locally downstream of transcribed genes (see paragraph 3.3 below). However, most *E. coli* genes were found to be transcribed at a comparable speed [76], suggesting that positive supercoils are readily eliminated by topoisomerases and do not play any regulatory role during elongation of these moderately expressed genes, whereas mechanical stalling of RNAP might occur at the most strongly expressed genes such as those of ribosomal RNAs [77].

### 3.2. Specific Regulation Involving DNA Regulatory Proteins

SC affects transcription via RNAP itself, but also via promoter-specific regulatory proteins (TFs). In classical thermodynamic regulation models [55], the latter act (as activators or repressors, depending



**Fig. 2.** Mechanisms of transcriptional regulation by DNA supercoiling. (A) Opening free energy of the *pelE* promoter of *D. dadantii* (upper panel) and expression level of three promoters from different species (lower panel, solid line for the model prediction) [23,45]; (B) Various mechanisms involving SC and NAPs/TFs at the *pelE* promoter [14,23,72]; (C) Local SC-mediated regulatory interaction: transcription from a given gene (left) can activate (top) or repress (bottom) its neighbour depending on their relative orientation [45].

on their action on RNAP activity) at distinct binding sites characterised by their affinity, the latter being inscribed in the genomic sequence and usually described by sequence motifs. In this simplistic view, the cell has limited regulatory freedom, since the only tunable parameter is the TF concentration. In the classical example of the global regulator CRP which binds (and usually activates) hundreds of promoters in *E. coli*, increasing its concentration (or rather that of its cofactor cAMP) may activate many of these target promoters, but in this model, the *relative* amount of CRP bound at these promoters, and thus also their relative activities, is out of the control of the cell. This view is increasingly challenged since additional layers of complexity were identified, including epigenetic modifications of the regulators or promoter DNA and post-transcriptional regulation. However, we note that many regulatory proteins recognize not only the base sequence, but also the DNA shape, a mechanism often referred to as indirect readout [78]. In that case, the activity of classical regulators is intrinsically modulated, not only by chemical (epigenetic) modifications requiring dedicated enzymes, but also by the ubiquitous mechanical deformations. A computational estimation of this recognition mode suggests that, in the case of CRP, the latter may in fact be the strongest determinant of its loose sequence selectivity [79]. Crucially, in contrast to the classical “static” sequence-motif model, this selectivity is now dependent on the mechanical state of chromosomal DNA, and thus subjected to cellular control through SC. Considering that the SC distribution itself is non-uniform and locally affected by NAPs, this additional mechanical dimension is probably a key contributor in the complexity of the binding selectivity of regulatory proteins.

This mechanism has not been investigated experimentally in many systems. One example is the thermoregulation of *pelE* by the repressor PecT, which is achieved not by a change in regulator concentration but rather by an increase of its binding affinity for the promoter resulting (at least partly) from temperature-induced chromosomal DNA relaxation (Fig. 2B, top) [14,80]. PecT belongs to the LysR-like family of TFs relying on indirect sequence readout, which is the largest TF family in enterobacteria [81], and similar mechanisms are used by other families [81–83]. Altogether, the role of DNA mechanics in TF binding selectivity is thus likely much more important than generally considered, although the precise mechanisms remain to be identified and modelled. Because most dimeric TFs are smaller than RNAP, they are probably less sensitive to the untwisting of bound DNA than the latter [79], but this sensitivity might be augmented substantially in the case of several binding sites arranged in helical phase. For example, H-NS was found to repress *pelE* on relaxed DNA (as most of its targets), but it is not the case on negatively supercoiled DNA [23]. A proposed explanation is that H-NS traps RNAP in a small loop at this promoter by bridging its two binding sites, which in the latter condition is prevented by their unfavourable helical phasing (Fig. 2B, middle). On the other hand, regulatory proteins also exert twist deformations themselves; a spectacular example is MerR, which in the presence of Hg<sub>2+</sub> binds and untwists the 19-bp spacer of the mercury resistance operon of Tn501 (a transposable element isolated from *Pseudomonas aeruginosa*), thus enabling RNAP binding [84].

In our opinion, the interplay between regulatory proteins and local SC may mostly rely on 3D modifications of the promoter conformation, i.e., writhe rather than twist. On the one hand, writhe could facilitate the formation of loops required for many regulatory interactions. These loops are favoured by the distinct mechanical properties of promoter DNA sequences [85], and were already included in regulatory models based on the transfer matrix formalism [86]. SC-dependent reduction of looping free energies can thus strongly modify the binding landscape of regulatory proteins, as already described [87,88]. Conversely, many regulatory proteins induce strong bends in DNA [83], e.g., CRP (~90°), LexA (~35°), as well as the NAPs FIS (~45°) and IHF (~180°). Since such deformations drastically reduce the energetic cost of DNA loops [89,90], they are expected to displace the twist-writhe equilibrium within the bound region in favour of the latter, and may thus induce local topological changes similar to those induced by SC [91]. In

summary, just like the NAPs are involved in a complex double-sided interaction with SC that shapes the global structure of the chromosome, a comparable effect probably occurs with many more regulatory proteins at the more local scale of gene promoters, with direct consequences for local transcription.

A final important ingredient is the widespread occurrence of structural transitions in genomic DNA. The latter can switch from double-stranded B-DNA to, among others, denaturated, Z-DNA, G-quadruplex or cruciform states. The rates of these different transitions can be computed by thermodynamic modelling [92], and depend not only on the DNA sequence, but also on torsional stress that destabilises B-DNA and can be accommodated more favourably in alternate states. It was also shown that these transitions occur predominantly at bacterial gene promoters [60] which they regulate according to various mechanisms, some involving TFs. Denaturated AT-rich regions located 50–200 bp upstream of the TSS can act as “sinks” for negative SC and impede the proper opening of the promoter by thermodynamic competition. This was shown to occur for *pelE*, which in vitro is not expressed in absence of CRP due to such upstream strand opening; when present, CRP not only favours the correct binding of RNAP, but also “closes” the upstream AT-rich tract, possibly by bending DNA (Fig. 2B, bottom) [23]. Since denaturation bubbles are extremely flexible, they may also strongly facilitate the formation of loops [89] required by TFs [86]. Additionally, some regulatory proteins may selectively bind non B-DNA regions, as occurs at the mammalian oncogene *cMYC* where negative SC triggering DNA melting is provided by adjacent transcription [60,93]. Finally, since rho-independent termination of transcription involves RNA hairpin structures, SC might also favour structural transitions in the DNA template itself at the transcription termination site, which could then modulate the termination rate, as already observed for the B-Z transition [94]. Since transcriptional read-through was highlighted as a widespread feature in bacterial genomes in recent years [95], an additional underestimated layer of regulation might thus also occur at this later stage of the transcription process.

### 3.3. Spatial Heterogeneities of DNA Supercoiling: The Transcription-Supercoiling Coupling

The intimate relationship between SC and transcription is not single-sided. In the elongation step, the helical structure of DNA imposes a fast rotation of the bulky RNAP relative to it (around two turns per second), but this movement is strongly hindered by the viscosity and crowding of the surrounding medium, resulting in an asymmetric accumulation of torsional stress from back to front, as recognised more than 30 years ago [96]. This phenomenon thus leads to an intrinsic dynamical coupling between SC and transcription highly dependent on gene orientations (Fig. 2C). This coupling regained significant interest in recent years since it was shown to underpin transcriptional bursting in bacteria [97], i.e., the nonlinear auto-induction of a promoter that can typically give rise to phenotypical heterogeneities among isogenic populations of cells [98]. Several theoretical models were proposed [45,77,99–101], most of them focusing on biophysical properties of transcription. A strong obstacle for their application in genomic regulation is the lack of experimental knowledge of the distribution of SC along a bacterial chromosome. A promising method involving the intercalating agent psoralen was developed in eukaryotes [7,102] but did not yet provide high-resolution data in bacteria [1]. Recently, indirect information was provided from binding distributions of topoisomerases obtained by ChIP-Seq at different resolutions [2,103]. These data, together with a systematic analysis of bacterial transcriptomes, confirmed that the distribution of SC along the chromosome is highly heterogeneous and strongly affected by gene orientations [104], leading to a fine-tuned and ancestral regulation of promoters depending on their genomic context [45]. In summary, the *local* level of SC experienced by a given promoter can strongly differ from the *global* (average) level of the chromosome, and depends on the orientation and activity of adjacent

genes, providing a strong mechanistic basis for the co-regulation of co-localised operons [105]. Direct evidence for these effects was obtained from experiments involving supercoiling-sensitive promoters inserted on the chromosome in different artificial configurations [70,104]. Comparable evidence is more difficult to establish for native promoters, but was highlighted in at least two examples of divergently organised operons: the *ilv* promoters of *E. coli* [106] and the *leu-500* promoter of *S. typhimurium* [107]. Interestingly, in these two examples, the local nature of SC (generated by the divergent genes) is combined with a complex regulatory mechanism involving DNA binding proteins. In the first example, the activation of *ilvP<sub>G</sub>* is prevented by the denaturation of an upstream AT-rich tract, except when the NAP IHF closes that region and favours the opening of the promoter (like CRP at the *pelE* promoter, Fig. 2B). The pattern is similar for *leu-500*, although the repression is here achieved by H-NS binding at an AT-rich tract, and relieved by the TF LeuO in presence of locally generated negative SC [5,108]. Since divergent genes involved in the same function and simultaneously expressed are commonly found in bacteria, including among those involved in pathogenicity (see below), these examples may only be the first of a large unexplored class.

The high density of bacterial genomes implies that the interaction between neighbouring genes could in fact give rise to a collective behaviour along larger distances, forming “topological domains”. Indeed, when promoters were displaced over the chromosome, their expression and supercoiling sensitivity were found to change depending on their location and neighbouring activity [109–112]. These domains, shaped by transcription and architectural proteins, remain a poorly defined notion in bacteria. Proposed lengthscales vary from 10 to 20 kb [109], to 50 kb [112] and up to hundreds of kilobases [20]; while the former may underpin an extension of the notion of operons [105,113], the latter probably reflect a higher order folding of bacterial chromatin involving different actors, and this hierarchical organisation remains to be characterised.

#### 4. DNA Supercoiling and the Coordination of Virulence Programs

##### 4.1. An Argument for DNA Supercoiling Being an Important Actor in Virulence Genetic Regulation

Most pathogenic bacteria exhibit close genomic proximity to non-pathogenic strains, with differences located at well-defined genomic regions (of a few up to hundreds of kilobases in size) called pathogenicity islands (PAIs), which contain the virulence genes involved in pathogenesis. These regions are harboured either on the chromosome or on plasmids, and are usually acquired by horizontal gene transfer (transformation, conjugation or phage-mediated transduction). As a result, different strains of a single species can present a remarkable diversity of pathogenic phenotypes (more than 10 for *Escherichia coli*), whereas a given virulence factor can be shared between different species [114]. This mechanism explains the rapid evolution of bacterial pathogens,

but also raises the question as to how the transferred genes are properly expressed after their integration into the distinct transcriptional regulatory network of the recipient cell. This problem is particularly acute for the bacterium, since any error in the expression time or strength of virulence factors immediately leads to the recognition and, ultimately, to the destruction of the invader by the host defence system [115,116].

At first glance, such drastic regulation of a few specific promoters seems to deviate from the global and non-specific regulation mode characteristic of SC. However, it appears equally incompatible with the sole action of strongly sequence-specific TFs, which would then be highly unstable during horizontal transfers between species, where these TFs are often evolutionarily distant [117]. As a matter of fact, many TFs involved in virulence indeed exhibit a weak sequence-specificity and are sensitive to the mechanical state of DNA [78], owing to an original regulatory mechanism affecting PAIs. Like other horizontally transferred regions, these usually exhibit a lower GC-content than the chromosomal average, and are therefore normally repressed by extensive binding of the NAP H-NS. Regulators can then activate the genes without any specific contact with RNAP, by competing with H-NS for promoter binding [114,118], which can be strongly dependent on the topological state of the region.

In most investigated species, the key signals triggering a quick activation or repression of virulence genes are precisely those environmental stress conditions that were shown to modulate the chromosome topology in various species (Table 3), e.g., a sharp acidity variation when *S. enterica* is transferred from the stomach to the intestine, or oxidative stress when *D. dadantii* leaves the plant apoplast. It is therefore no surprise that virulence genes from an increasing number of zoopathogenic or phytopathogenic species were shown to be directly regulated by SC, as summarized in Table 3. Does this mechanism play a role during the infection process, as these data suggest? In our opinion, based on the complex regulatory mechanisms illustrated above, SC is a good candidate to play the role of a basal and robust coordinator of virulence gene expression, by (1) modulating the simultaneous action of many (more specific) regulators at virulence promoters, such complexity being a characteristic feature of the latter, (2) co-regulating the adjacent genes of a PAI through the evolutionarily conserved transcription-supercoiling coupling. We present below some examples supporting this hypothesis, keeping in mind that existing results mostly concern individual genes, whereas the topological organisation of entire PAIs and its effect on their expression remain poorly understood [129].

##### 4.2. Widespread Evidence for a Regulatory Role of DNA Supercoiling in Virulence

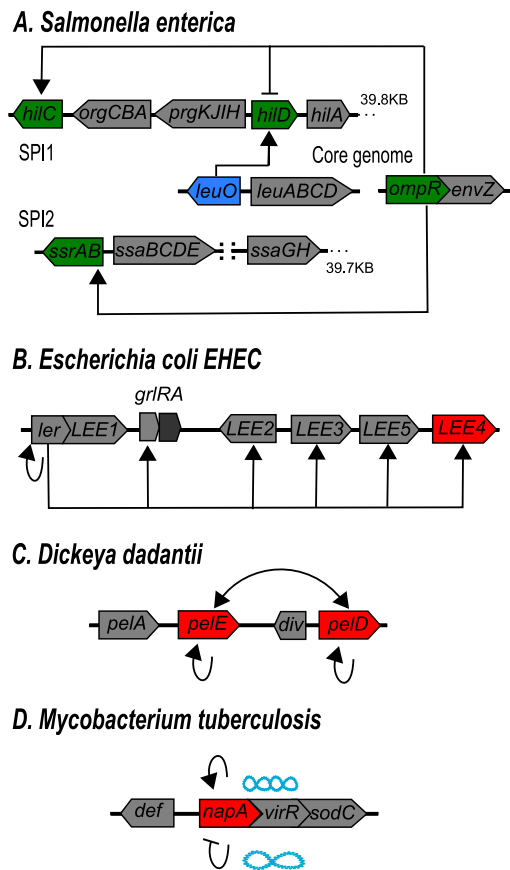
*Salmonella enterica* is one of the most studied pathogens, and this is also true of the regulation of its virulence system by SC (mostly in C. Dorman's laboratory). Several key virulence genes within its two largest PAIs (SPI-1 and SPI-2) are supercoiling-sensitive, as well as those of the

**Table 3**

Virulence genes are regulated by DNA supercoiling in various pathogenic species. Phyla: Proteobacteria (P), firmicutes (F), actinobacteria (A). Lifestyles: F = facultative, I = intracellular, E = extracellular, P = pathogen. Response to chromosomal DNA Relaxation by antibiotics: repressed (–) or activated (+), with corresponding reference(s).

Phylum	Family	Species	Tissue	Lifestyle	Stress encountered	Gene involved in virulence	Relax response	Ref	
P	Enterobacteriaceae	<i>Salmonella enterica</i>	gastrointestinal tracts	FIP	acid	<i>hilD</i> , <i>hilC</i> , <i>ssrAB</i>	+	[91]	
	Enterobacteriaceae	<i>Shigella flexneri</i>	intestinal epithelium	FIP	temperature	<i>virF</i>	+	[119]	
	Enterobacteriaceae	<i>E. coli</i> (EHEC)	intestinal epithelium	FIP	temperature	<i>espADB</i>	–	[120]	
	Pectobacteriaceae	<i>Dickeya dadantii</i>	plant apoplast	EP	acid,oxidative	<i>pelE</i>	–	[23]	
	Pseudomonadaceae	<i>Pseudomonas syringae</i>	plant apoplast	EP	oxidative	<i>avrPphB</i>	+	[121]	
	Vibrionaceae	<i>Vibrio cholerae</i>	small intestine	EP	acid	<i>acfA</i> , <i>acfD</i>	–	[122,123]	
	Alcaligenaceae	<i>Bordetella pertussis</i>	lung epithelial cells.	IP	temperature	<i>ptx</i>	–	[124]	
	Campylobacteriaceae	<i>Campylobacter jejuni</i>	digestive tract	IP	temperature, pH	<i>momp</i>	–	[125]	
	F	Staphylococcaceae	<i>Staphylococcus aureus</i>	respiratory tract, skin	EP	osmolarity	<i>spa</i> , <i>eta</i>	+	[126,49]
		Streptococcaceae	<i>Streptococcus pneumoniae</i>	respiratory tract, skin	EP	oxidative	<i>fatD</i>	–	[127]
A	Mycobacteriaceae	<i>Mycobacterium tuberculosis</i>	respiratory tract, skin	IP	oxidative	<i>virR</i> , <i>sodC</i>	–	[128]	





**Fig. 3.** Regulatory networks within several pathogenicity islands. In all species, key virulence genes are either relaxation-activated (green), relaxation-repressed (red), or regulated via the transcription-SC coupling (blue). Arrows indicate an activating (arrow) and bars a repressing effect. For *M. tuberculosis*, this effect depends on the SC level (repression on a relaxed template, activation on a supercoiled template). In *D. dadantii*, *pel* genes are self- and mutually-inductive via pectin degradation relieving the repression of both genes by the TF KdgR.

central virulence regulators OmpR and LeuO located in the core genome (Fig. 3A). In this representative example, key virulence functions were thus integrated into the pre-existing transcription regulatory network, whereby OmpR and LeuO, primarily devoted to other functions (the former is an abundant NAP-like protein), are specifically recruited at virulence promoters in a SC-dependent manner [5,91]. Interestingly, these regulators mostly target divergently oriented operons, and the same pattern is observed in *Vibrio cholerae* (*tcpPH*) [78] and *Shigella flexneri* (*icsA*, *icsB*) [117]. While gene orientations are indifferent to classical regulatory models, divergent promoters are the ones most sensitive to the transcription-supercoiling coupling [45], which is also involved in LeuO recruitment at its own (divergent) promoter [108]. Gene orientations within the PAIs are thus likely not accidental, but rather reflect the evolved infrastructure of a local coordinated and SC-coupled gene expression.

The virulence of *E. coli* EHEC (causing severe diarrhoea) depends on a secretion system encoded in the LEE operons (Fig. 3B), including many genes shared with other *E. coli* pathogenic strains, e.g. EPEC [114]. These operons are globally repressed by H-NS, whose binding is antagonised by the activator Ler encoded in LEE1 [118]. Ler itself belongs to the H-NS family and this competition may very well be affected by the topological state of the domain. Although such an effect was not investigated in detail, the expression of LEE4 was indeed found to be SC-sensitive [120], and the same could be true of the other operons.

In the phytopathogen *D. dadantii*, pectinolytic enzymes are the main virulence factors, responsible for the soft rot symptoms [115]. These are encoded by *pel* genes scattered in several PAIs along the chromosome [20]. *peIE* and *peID*, the two major members of this family, are paralogous genes that evolved from a unique ancestor but exhibit different expression patterns [130]. They are regulated by several NAPs (H-NS, FIS, IHF, Lrp) as well as many TFs (e.g., CRP, KdgR, PecT, PecS) but are also among the most SC-sensitive genes in the chromosome. We illustrated above how different regulatory proteins act in a complex combination, with SC modulating their relative affinities (Fig. 2B). Although no effect of locally generated SC was directly shown here, a divergent non-coding transcript (*div*) was recently identified upstream of the *peID* promoter, which “feeds” the latter with RNAP with a strong dependence on the 3D conformation of the promoter [130], possibly involving a local transmission of SC (Fig. 3C).

Finally, since topoisomerases are conserved among all bacteria, we may expect SC to play a role in the virulence of widely distant species. This was indeed recently demonstrated in the actinobacterium *Mycobacterium tuberculosis*, where the gene of a new NAP (NapA) was identified in the same operon as major virulence factors/regulators (*SodC*, *VirR*) [128]. Interestingly, NapA autoregulates its (divergently oriented) promoter in a SC-dependent manner (Fig. 3D): in this case, global SC variations induced by environmental conditions may act as a switch, turning the operon on or off with high specificity.

## 5. Conclusion

The discussed examples show that SC plays a direct role in the regulation of virulence in many species, albeit with a remarkable variety of mechanisms involving a combination of additional regulatory actors, and these mostly remain to be characterised. A final and striking example is the remarkably small and poorly characterised tenericute *Mycoplasma pneumoniae*, which is able to infect the human respiratory tract despite being almost devoid of TFs; the ancestral regulatory action of SC is thought to play an even more cardinal role in this case [37]. The increasing interest in SC in the genetic regulation community already results in new experimental techniques facilitating the mapping of supercoil distributions at higher resolution [1,2] as well as the development of computational models at various scales of detail [45,77,101,131], which together will help elucidating the pivotal regulatory action of SC in bacterial genetic regulation advocated here.

## Acknowledgements

This work has been funded by the Agence Nationale de la Recherche [ANR-18-CE45-0006-01 grant to S.M.], CNRS, INSA Lyon and Université de Lyon. We thank Georgi Muskhelishvili for his critical reading of the manuscript.

## References

- [1] Lal A, Dhar A, Trostel A, Kouzine F, Seshasayee ASN, Adhya S. Genome scale patterns of supercoiling in a bacterial chromosome. *Nat Commun* 2016;7:11055.
- [2] Sutormin D, Rubanova N, Logacheva M, Ghilarov D, Severinov K. Single-nucleotide-resolution mapping of DNA gyrase cleavage sites across the *Escherichia coli* genome. *Nucleic Acids Res* 2019;47:1373–88.
- [3] Rovinskiy NS, Agbleke AA, Chesnokova ON, Higgins NP. Supercoil levels in *E. coli* and *Salmonella* chromosomes are regulated by the C-terminal 35–38 amino acids of GyrA. *Microorganisms* 2019;7:81.
- [4] Travers A, Muskhelishvili G. DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nat Rev Microbiol* 2005;3:157–69.
- [5] Dorman CJ, Dorman MJ. DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. *Biophys Rev* 2016;8:209–20.
- [6] Watson JD, Crick FH. The structure of DNA. *Cold Spring Harb Symp Quant Biol* 1953;18:123–31.
- [7] Gilbert N, Allan J. Supercoiling in DNA and chromatin. *Curr Op Gen Dev* 2014;25:15–21.
- [8] Forterre P, Gabelle D. Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Res* 2009;37:679–92.

- [9] Zawadzki P, Stracy M, Ginda K, Zawadzka K, Lesterlin C, Kapanidis AN, et al. The localization and action of topoisomerase IV in *Escherichia coli* chromosome segregation is coordinated by the SMC complex, MukBEF. *Cell Rep* 2015;13:2587–96.
- [10] Dillon SC, Dorman CJ. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* 2010;8:185–95.
- [11] López-García P, Forterre P. DNA topology and the thermal stress response, a tale from mesophiles and hyperthermophiles. *BioEssays News Rev Mol Cell Dev Biol* 2000;22:738–46.
- [12] Ogata Y, Mizushima T, Kataoka K, Miki T, Sekimizu K. Identification of DNA topoisomerases involved in immediate and transient DNA relaxation induced by heat shock in *Escherichia coli*. *Mol Gen Genet MGG* 1994;244:451–5.
- [13] Rohde JR, Fox JM, Minnich SA. Thermoregulation in *Yersinia enterocolitica* is coincident with changes in DNA supercoiling. *Mol Microbiol* 1994;12:187–99.
- [14] Héraulot E, Reverchon S, Nasser W. Role of the LysR-type transcriptional regulator PecT and DNA supercoiling in the thermoregulation of *pel* genes, the major virulence factors in *Dickeya dadantii*. *Environ Microbiol* 2014;16:734–45.
- [15] Krispin O, Allmansberger R. Changes in DNA supercoiling as a response of *Bacillus subtilis* towards different kinds of stress. *FEMS Microbiol Lett* 1995;134:129–35.
- [16] 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* 1997;23:381–6.
- [17] Karem K, Foster JW. The influence of DNA topology on the environmental regulation of a pH-regulated locus in *Salmonella typhimurium*. *Mol Microbiol* 1993;10:75–86.
- [18] Quinn HJ, Cameron ADS, Dorman CJ. Bacterial regulon evolution: distinct responses and roles for the identical OmpR proteins of *Salmonella Typhimurium* and *Escherichia coli* in the acid stress response. *PLoS Genet* 2014;10.
- [19] Colgan AM, Quinn HJ, Kary SC, Mitchenall LA, Maxwell A, Cameron ADS, et al. Negative supercoiling of DNA by gyrase is inhibited in *Salmonella enterica* serovar *Typhimurium* during adaptation to acid stress. *Mol Microbiol* 2018;107:734–46.
- [20] Jiang X, Sobetzko P, Nasser W, Reverchon S, Muskhelishvili G. Chromosomal “stress-response” domains govern the spatiotemporal expression of the bacterial virulence program. *MBio* 2015;6.
- [21] Hsieh LS, Rouviere-Yaniv J, Drlica K. Bacterial DNA supercoiling and [ATP]/[ADP] ratio: changes associated with salt shock. *J Bacteriol* 1991;173:3914–7.
- [22] Higgins CF, Dorman CJ, Stirling DA, Waddell L, Booth IR, May G, et al. A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. *Cell* 1988;52:569–84.
- [23] Ouafa Z-A, Reverchon S, Lautier T, Muskhelishvili G, Nasser W. The nucleoid-associated proteins H-NS and FIS modulate the DNA supercoiling response of the *pel* genes, the major virulence factors in the plant pathogen bacterium *Dickeya dadantii*. *Nucl Ac Res* 2012;40:4306–19.
- [24] Sheehan BJ, Foster TJ, Dorman CJ, Park S, Stewart GS. Osmotic and growth-phase dependent regulation of the *etg* gene of *Staphylococcus aureus*: a role for DNA supercoiling. *Mol Gen Genet MGG* 1992;232:49–57.
- [25] Ali N, Herron PR, Evans MC, Dyson PJ. Osmotic regulation of the *Streptomyces lividans* thiothrepton-inducible promoter, *ptpA*. *Microbiology* 2002;148:381–90.
- [26] Weinstein-Fischer D, Elgrably-Weiss M, Altuvia S. *Escherichia coli* response to hydrogen peroxide: a role for DNA supercoiling, topoisomerase I and Fis. *Mol Microbiol* 2000;35:1413–20.
- [27] Hsieh LS, Burger RM, Drlica K. Bacterial DNA supercoiling and [ATP]/[ADP]. Changes associated with a transition to anaerobic growth. *J Mol Biol* 1991;219:443–50.
- [28] Cameron ADS, Stoebel DM, Dorman CJ. DNA supercoiling is differentially regulated by environmental factors and FIS in *Escherichia coli* and *Salmonella enterica*. *Mol Microbiol* 2011;80:85–101.
- [29] Balke VL, Gralla JD. Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *J Bacteriol* 1987;169:4499–506.
- [30] Hatfield GW, Benham CJ. DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu Rev Genet* 2002;36:175–203.
- [31] Zhou Q, Gomez Hernandez ME, Fernandez-Lima F, Tse-Dinh Y-C. Biochemical basis of *E. coli* topoisomerase I relaxation activity reduction by nonenzymatic lysine acetylation. *Int J Mol Sci* 2018;19.
- [32] Pommier Y, Leo E, Zhang H, Marchand C. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol* 2010;17:421–33.
- [33] de la Campa AG, Ferrándiz MJ, Martín-Galiano AJ, García MT, Tirado-Vélez JM. The transcriptome of *Streptococcus pneumoniae* induced by local and global changes in supercoiling. *Front Microbiol* 2017;8.
- [34] Leblanc BP, Benham CJ, Clark DJ. An initiation element in the yeast CUP1 promoter is recognized by RNA polymerase II in the absence of TATA box-binding protein if the DNA is negatively supercoiled. *Proc Natl Acad Sci U S A* 2000;97:10745–50.
- [35] Tenaillon O, Barrick JE, Ribick N, Deatherage DE, Blanchard JL, Dasgupta A, et al. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* 2016;536:165–70.
- [36] Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 2005;169:523–32.
- [37] Junier I, Unal EB, Yus E, Lloréns-Rico V, Serrano L. Insights into the mechanisms of basal coordination of transcription using a genome-reduced bacterium. *Cell Syst* 2016;2:391–401.
- [38] Dorman CJ. Regulation of transcription by DNA supercoiling in mycoplasma genitalium: global control in the smallest known self-replicating genome. *Mol Microbiol* 2011;81:302–4.
- [39] Brinza L, Calevro F, Charles H. Genomic analysis of the regulatory elements and links with intrinsic DNA structural properties in the shrunken genome of *Buchnera*. *BMC Genomics* 2013;14:73.
- [40] Jeong KS, Xie Y, Hiasa H, Khodursky AB. Analysis of pleiotropic transcriptional profiles: a case study of DNA Gyrase inhibition. *PLoS Genet* 2006;2:e152.
- [41] Cheung KJ, Badarinarayana V, Selinger DW, Janse D, Church GM. A microarray-based antibiotic screen identifies a regulatory role for supercoiling in the osmotic stress response of *Escherichia coli*. *Genome Res* 2003;13:206–15.
- [42] Blot N, Mavathur R, Geertz M, Travers A, Muskhelishvili G. Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep* 2006;7:710–5.
- [43] Peter BJ, Arsuaga J, Breier AM, Khodursky AB, Brown PO, Cozzarelli NR. Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biol* 2004;5:R87.
- [44] Webber MA, Ricci V, Whitehead R, Patel M, Fookes M, Ivens A, et al. Clinically relevant mutant DNA gyrase alters supercoiling, changes the transcriptome, and confers multidrug resistance. *MBio* 2013;4.
- [45] El Houdaigui B, Forquet R, Hindré T, Schneider D, Nasser W, Reverchon S, et al. Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Res* 2019;47:5648–57.
- [46] Gmuender H, Kuratli K, Di Padova K, Gray CP, Keck W, Evers S. Gene expression changes triggered by exposure of *Haemophilus influenzae* to Novobiocin or ciprofloxacin: combined transcription and translation analysis. *Genome Res* 2001;11:28–42.
- [47] Ferrándiz M-J, Martín-Galiano AJ, Schwartzman JB, de la Campa AG. The genome of *Streptococcus pneumoniae* is organized in topology-reacting gene clusters. *Nucleic Acids Res* 2010;38:3570–81.
- [48] Ferrándiz M-J, Martín-Galiano AJ, Aranz C, Camacho-Soguero I, Tirado-Vélez J-M, de la Campa AG. An increase in negative supercoiling in bacteria reveals topology-reacting gene clusters and a homeostatic response mediated by the DNA topoisomerase I gene. *Nucleic Acids Res* 2016;44:7292–303.
- [49] Schröder W, Bernhardt J, Marincola G, Klein-Hitpass L, Herbig A, Krupp G, et al. Altering gene expression by aminocoumarins: the role of DNA supercoiling in *Staphylococcus aureus*. *BMC Genomics* 2014;15:291.
- [50] Sioud M, Boudabous A, Cekaite L. Transcriptional responses of *Bacillus subtilis* and *thuringiensis* to antibiotics and anti-tumour drugs. *Int J Mol Med* 2009;23:33–9.
- [51] Szafraan MJ, Gongerowska M, Malecki T, Elliot MA, Jakimowicz D. Transcriptional response of *Streptomyces coelicolor* to rapid chromosome relaxation or long-term supercoiling imbalance. *BioRxiv* 2019;602359.
- [52] Boshoff HIM, Myers TG, Copp BR, McNeil MR, Wilson MA, Barry CE. The transcriptional responses of mycobacterium tuberculosis to inhibitors of metabolism: novel insights into drug mechanisms of action. *J Biol Chem* 2004;279:40174–84.
- [53] Prakash JSS, Sinetova M, Zorina A, Kupriyanova E, Suzuki I, Murata N, et al. DNA supercoiling regulates the stress-inducible expression of genes in the cyanobacterium *Synechocystis*. *Mol Biosyst* 2009;5:1904–12.
- [54] Vijayan V, Zuzow R, O’Shea EK. Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proc Natl Acad Sci* 2009;106:22564–8.
- [55] Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, et al. Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev* 2005;15:116–24.
- [56] Marr C, Geertz M, Hütt M-T, Muskhelishvili G. Dissecting the logical types of network control in gene expression profiles. *BMC Syst Biol* 2008;2:18.
- [57] Borukhov S, Nudler E. RNA polymerase: the vehicle of transcription. *Trends Microbiol* 2008;16:126–34.
- [58] Forterre P. A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet* 2002;18:236–7.
- [59] Meyer S, Jost D, Theodorakopoulos N, Peyrard M, Lavery R, Everaers R. Temperature dependence of the DNA double helix at the nanoscale: structure, elasticity, and fluctuations. *Biophys J* 2013;105:1904–14.
- [60] Du X, Wojtowicz D, Bowers AA, Levens D, Benham CJ, Przytycka TM. The genome-wide distribution of non-B DNA motifs is shaped by operon structure and suggests the transcriptional importance of non-B DNA structures in *Escherichia coli*. *Nucleic Acids Res* 2013;41:5965–77.
- [61] Jost D. Twist-DNA: computing base-pair and bubble opening probabilities in genomic superhelical DNA. *Bioinformatics* 2013;29:2479–81.
- [62] Alexandrov BS, Gelev V, Yoo SW, Alexandrov LB, Fukuyo Y, Bishop AR, et al. DNA dynamics play a role as a basal transcription factor in the positioning and regulation of gene transcription initiation. *Nucleic Acids Res* 2010;38:1790–5.
- [63] Travers AA. Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. *J Bacteriol* 1980;141:973–6.
- [64] Pemberton IK, Muskhelishvili G, Travers AA, Buckle M. The G+C-rich discriminator region of the *tyrT* promoter antagonises the formation of stable preinitiation complexes I. Edited by M. Yaniv. *J Mol Biol* 2000;299:859–64.
- [65] Auner H, Buckle M, Deufel A, Kutateladze T, Lazarus L, Mavathur R, et al. Mechanism of transcriptional activation by FIS: role of core promoter structure and (DNA) topology. *J Mol Biol* 2003;331:331–44.
- [66] Henderson KL, Felth LC, Molzahn CM, Shkel I, Wang S, Chhabra M, et al. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proc Natl Acad Sci* 2017;114:E3032–40.
- [67] Wood DC, Lebowitz J. Effect of supercoiling on the abortive initiation kinetics of the RNA-I promoter of ColE1 plasmid DNA. *J Biol Chem* 1984;259:11184–7.
- [68] Lim HM, Lewis DEA, Lee HJ, Liu M, Adhya S. Effect of varying the supercoiling of DNA on transcription and its regulation. *Biochemistry* 2003;42:10718–25.
- [69] Menzel R, Gellert M. Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. *Cell* 1983;34:105–13.
- [70] Dages S, Dages K, Zhi X, Leng F. Inhibition of the *gyrA* promoter by transcription-coupled DNA supercoiling in *Escherichia coli*. *Sci Rep* 2018;8:14759.
- [71] Jha RK, Tare P, Nagaraja V. Regulation of the *gyr* operon of mycobacterium tuberculosis by overlapping promoters, DNA topology, and reiterative transcription. *Biochem Biophys Res Commun* 2018;501:877–84.

[72] Lautier T, Blot N, Muskhelishvili G, Nasser W. Integration of two essential virulence modulating signals at the *Erwinia chrysanthemi* pel gene promoters: a role for Fis in the growth-phase regulation. *Mol Microbiol* 2007;66:1491–505.

[73] Wang JY, Syvanen M. DNA twist as a transcriptional sensor for environmental changes. *Mol Microbiol* 1992;6:1861–6.

[74] Unniraman S, Nagaraja V. Axial distortion as a sensor of supercoil changes: a molecular model for the homeostatic regulation of DNA gyrase. *J Genet* 2001;80:119–24.

[75] Ma J, Bai L, Wang MD. Transcription under torsion. *Science* 2013;340:1580–3.

[76] Chen H, Shiroguchi K, Ge H, Xie XS. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol Syst Biol* 2015;11:808.

[77] Sevier SA, Levine H. Properties of gene expression and chromatin structure with mechanically regulated elongation. *Nucleic Acids Res* 2018;46:5924–34.

[78] Dorman CJ, Dorman MJ. Control of virulence gene transcription by indirect readout in *Vibrio cholerae* and *Salmonella enterica* serovar Typhimurium. *Environ Microbiol* 2017;19:3834–45.

[79] Cevost J, Vaillant C, Meyer S. ThreDNA: predicting DNA mechanics' contribution to sequence selectivity of proteins along whole genomes. *Bioinformatics* 2018;34:609–16.

[80] Goldstein E, Drlica K. Regulation of bacterial DNA supercoiling: plasmid linking numbers vary with growth temperature. *Proc Natl Acad Sci U S A* 1984;81:4046–50.

[81] Pérez-Rueda E, Collado-Vides J. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res* 2000;28:1838–47.

[82] Hommais F, Oger-Desfeux C, Gijsegem FV, Castang S, Ligorì S, Expert D, et al. PecS is a global regulator of the symptomatic phase in the phytopathogenic bacterium *Erwinia chrysanthemi* 3937. *J Bacteriol* 2008;190:7508–22.

[83] Harteis S, Schneider S. Making the bend: DNA tertiary structure and protein-DNA interactions. *Int J Mol Sci* 2014;15:12335–63.

[84] Parkhill J, Brown NL. Site-specific insertion and deletion mutants in the mer promoter-operator region of Tn501; the nineteen base-pair spacer is essential for normal induction of the promoter by MerR. *Nucleic Acids Res* 1990;18:5157–62.

[85] Travers A, Muskhelishvili G. A common topology for bacterial and eukaryotic transcription initiation? *EMBO Rep* 2007;8:147–51.

[86] Teif VB. General transfer matrix formalism to calculate DNA-protein-drug binding in gene regulation: application to OR operator of phage  $\lambda$ . *Nucleic Acids Res* 2007;35:e80.

[87] Xiao B, Zhang H, Johnson RC, Marko JF. Force-driven unbinding of proteins HU and Fis from DNA quantified using a thermodynamic Maxwell relation. *Nucleic Acids Res* 2011;39:5568–77.

[88] Efremov AK, Yan J. Transfer-matrix calculations of the effects of tension and torque constraints on DNA-protein interactions. *Nucleic Acids Res* 2018;46:6504–27.

[89] Yan J, Marko JF. Localized single-stranded bubble mechanism for cyclization of short double Helix DNA. *Phys Rev Lett* 2004;93:108108.

[90] Du Q, Smith C, Shiffeldrim N, Vologodskii A, Vologodskii A. Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proc Natl Acad Sci* 2005;102:5397–402.

[91] Cameron ADS, Dorman CJ. A fundamental regulatory mechanism operating through OmpR and DNA topology controls expression of *Salmonella* pathogenicity Islands SPI-1 and SPI-2. *PLoS Genet* 2012;8:e1002615.

[92] Zhabinskaya D, Benham CJ. Theoretical analysis of competing conformational transitions in superhelical DNA. *PLoS Comput Biol* 2012;8:e1002484.

[93] Kouzine F, Sanford S, Elisha-Feil Z, Levens D. The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat Struct Mol Biol* 2008;15:146–54.

[94] Peck LJ, Wang JC. Transcriptional block caused by a negative supercoiling induced structural change in an alternating CG sequence. *Cell* 1985;40:129–37.

[95] Conway T, Creecy JP, Maddox SM, Grissom JE, Conkle TL, Shadid TM, et al. Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *MBio* 2014;5:e01442.

[96] Liu LF, Wang JC. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci* 1987;84:7024–7.

[97] Chong S, Chen C, Ge H, Xie XS. Mechanism of transcriptional bursting in bacteria. *Cell* 2014;158:314–26.

[98] Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science* 2002;297:1183–6.

[99] Meyer S, Beslon G. Torsion-mediated interaction between adjacent genes. *PLoS Comput Biol* 2014;10:e1003785.

[100] Brackley CA, Johnson J, Bentivoglio A, Corless S, Gilbert N, Gonnella G, et al. Stochastic model of supercoiling-dependent transcription. *Phys Rev Lett* 2016;117:018101.

[101] Ancona M, Bentivoglio A, Brackley CA, Gonnella G, Marenduzzo D, et al. Transcriptional Bursts in a Nonequilibrium model for gene regulation by supercoiling. *Biophys J* 2019;117(2):369–76.

[102] Kouzine F, Gupta A, Baranello L, Wojtowicz D, Ben-Aissa K, Liu J, et al. Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat Struct Mol Biol* 2013;20:396–403.

[103] Ahmed W, Sala C, Hegde SR, Jha RK, Cole ST, Nagaraja V. Transcription facilitated genome-wide recruitment of topoisomerase I and DNA gyrase. *PLoS Genet* 2017;13:e1006754.

[104] Sobetzko P. Transcription-coupled DNA supercoiling dictates the chromosomal arrangement of bacterial genes. *Nucleic Acids Res* 2016;44:1514–24.

[105] Junier I, Rivoire O. Conserved units of co-expression in bacterial genomes: an evolutionary insight into transcriptional regulation. *PLoS One* 2016;11:e0155740.

[106] Opel ML, Hatfield G. DNA supercoiling-dependent transcriptional coupling between the divergently transcribed promoters of the *ilvYC* operon of *Escherichia coli* is proportional to promoter strengths and transcript lengths. *Mol Microbiol* 2001;39:191–8.

[107] Wu H-Y, Tan J, Fang M. Long-range interaction between two promoters: activation of the *leu-500* promoter by a distant upstream promoter. *Cell* 1995;82:445–51.

[108] Fang M, Wu H-Y. A promoter relay mechanism for sequential gene activation. *J Bacteriol* 1998;180:626–33.

[109] Postow L, Hardy CD, Arsuaga J, Cozzarelli NR. Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev* 2004;18:1766–79.

[110] Gerganova V, Berger M, Zaldastanishvili E, Sobetzko P, Lafon C, Mourez M, et al. Chromosomal position shift of a regulatory gene alters the bacterial phenotype. *Nucleic Acids Res* 2015;43:8215–26.

[111] Brambilla E, Sclavi B, et al. G3 Genes *Genomes Genet* 2015;5:605–14.

[112] Ferrándiz M-J, Arnanz C, Martín-Galiano AJ, Rodríguez-Martín C, de la Campa AG. Role of global and local topology in the regulation of gene expression in *Streptococcus pneumoniae*. *PLoS One* 2014;9:e101574.

[113] Junier I, Frémont P, Rivoire O. Universal and idiosyncratic characteristic lengths in bacterial genomes. *Phys Biol* 2018;15:035001.

[114] Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. *Clin Microbiol Rev* 2004;17:14–56.

[115] Leonard S, Hommais F, Nasser W, Reverchon S. Plant-phytopathogen interactions: bacterial responses to environmental and plant stimuli. *Environ Microbiol* 2017;19:1689–716.

[116] Stecher B, Hardt W-D. Mechanisms controlling pathogen colonization of the gut. *Curr Opin Microbiol* 2011;14:82–91.

[117] Dorman CJ. Virulence gene regulation in *Shigella*. *EcoSal Plus* 2004;1.

[118] Connolly JPR, Finlay BB, Roe AJ. From ingestion to colonization: the influence of the host environment on regulation of the LEE encoded type III secretion system in enterohaemorrhagic *Escherichia coli*. *Front Microbiol* 2015;6.

[119] Dorman CJ, Bhriain NN, Higgins CF. DNA supercoiling and environmental regulation of virulence gene expression in *Shigella flexneri*. *Nature* 1990;344:789.

[120] Beltrametti F, Kresse AU, Guzmán CA. Transcriptional regulation of the *esp* genes of Enterohaemorrhagic *Escherichia coli*. *J Bacteriol* 1999;181:3409–18.

[121] Neale HC, Jackson RW, Preston GM, Arnold DL. Supercoiling of an excised genomic island represses effector gene expression to prevent activation of host resistance. *Mol Microbiol* 2018;110:444–54.

[122] Matson JS, Withey JH, DiRita VJ. Regulatory networks controlling *Vibrio cholerae* virulence gene expression. *Infect Immun* 2007;75:5542–9.

[123] Parsot C, Mekalanos JJ. Structural analysis of the *acfA* and *acfD* genes of *Vibrio cholerae*: effects of DNA topology and transcriptional activators on expression. *J Bacteriol* 1992;174:5211–8.

[124] Graeff-Wohlleben H, Deppisch H, Gross R. Global regulatory mechanisms affect virulence gene expression in *Bordetella pertussis*. *Mol Gen Genet MGG* 1995;247:86–94.

[125] Dedieu L, Pages J-M, Bolla J-M. Environmental regulation of campylobacter jejuni major outer membrane protein porin expression in *Escherichia coli* monitored by using green fluorescent protein. *Appl Environ Microbiol* 2002;68:4209–15.

[126] Fournier B. Protein A gene expression is regulated by DNA supercoiling which is modified by the ArlS-ArlR two-component system of *Staphylococcus aureus*. *Microbiology* 2004;150:3807–19.

[127] Ferrándiz M-J, de la Campa AG. The fluoroquinolone levofloxacin triggers the transcriptional activation of iron transport genes that contribute to cell death in *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* 2014;58:247–57.

[128] Datta C, Jha RK, Ganguly S, Nagaraja V. NapA (Rv0430), a novel nucleoid-associated protein that regulates a virulence operon in mycobacterium tuberculosis in a supercoiling-dependent manner. *J Mol Biol* 2019;431:1576–91.

[129] Martín-Galiano AJ, Ferrándiz MJ, de la Campa AG. Bridging chromosomal architecture and pathophysiology of *Streptococcus pneumoniae*. *Genome Biol Evol* 2017;9:350–61.

[130] Duprey A, Nasser W, Léonard S, Brochier-Armanet C, Reverchon S. Transcriptional start site turnover in the evolution of bacterial paralogous genes - the *peIE-peID* virulence genes in *Dickeya*. *FEBS J* 2016;283(22):4192–207.

[131] Lepage T, Junier I. A polymer model of bacterial supercoiled DNA including structural transitions of the double helix. *Phys A* 2019;121196.

## 1.4 Thesis motivations and objectives

The above review highlights the ubiquitous role of SC both as a relay of environmental signals to the chromosome, and as a global regulator of gene expression in bacteria, by affecting transcription at multiple steps through RNAP itself and regulatory proteins. However, systematic models of transcriptional regulation by SC comparable to those involving, *e.g.* TFs, remain essentially lacking. The main purpose of the present thesis is to decipher the underlying regulatory mechanisms, by developing quantitative models able to explain and predict gene response to SC variations. In particular, we focus our modelling on promoter-sequence determinants of gene SC-sensitivity, based on RNAP-DNA interaction, and independently from additional regulatory protein. Global-scale regulatory models are developed using a thermodynamic framework, and further validated at the genome level using a combination of (i) available transcriptomic responses under SC variations, and (ii) transcription assays on mutant promoters conducted specifically for the present thesis. Since the investigated mechanisms rely on the fundamental properties of DNA and highly conserved molecular actors (RNAP and topoisomerases), models are further validated in distant bacteria in terms of phylogeny and lifestyle.

To achieve genome-wide model validation, in Chapter 2, we first gather from different sources and for several bacteria (i) transcriptomic responses under SC variations, and (ii) promoter structures and locations, into an unified database. Programming tools are then implemented, in order to access the database and to perform various tasks, *e.g.* statistical analyses from the appropriate input datasets, and model construction and simulation. A subsection is then dedicated to side-projects to which I contributed through the present database and programming tools.

In Chapter 3, we characterise the transcriptome of *D. dadantii*, a phytopathogen in which SC plays a key role during plant infection. We therefore wished to extend model validation to this bacterium, which required the mapping of its transcription units, transcription start sites and promoters, based on available data. It was also a great opportunity to significantly contribute to further progress in the field of phytopathogenicity, by providing the first transcriptomic map of this model species of the *Dickeya* genus.

In Chapter 4, we present a quantitative mechanistic model of transcriptional



regulation by SC, relying on the step of promoter opening during open-complex formation. Based on transcription assays on mutant promoters, and genome-wide statistical analyses of transcriptomic responses under SC variations, it provides the predictive rules by which the sequence of the discriminator element, and notably its G/C-content, modulates the SC response of promoters.

Since SC affects transcription at multiple steps of the process beyond open-complex formation, in Chapter 5, we propose another model of transcriptional regulation by SC, relying on the orientation between -35 and -10 elements recognised by RNAP to form the closed-complex. It quantifies the contribution of the promoter spacer length in the SC response of promoters, based on a geometric effect and without any specific sequence requirement, in contrast to most existing models of transcriptional regulation. As a key advantage, the presented thermodynamic models of transcription depend essentially on DNA's measured physicochemical properties, and require almost no adjustable parameter.

Finally, in the Chapter 6, we present a summary of the results obtained, and propose some perspectives for future work.





## Chapter 2

# Development of an integrated database, implementation of programming tools and contribution to side projects

To achieve the thesis objectives, the first step was to collect and combine available data required for both the characterisation of *D. dadantii* transcriptome and the validation of the models of transcriptional regulation by SC, into an unified database. I then implemented Python programming tools, in order to access the database, and perform various tasks from the appropriate input datasets.

### 2.1 Development of an integrated database

A database can be defined as a logically coherent and organised collection of data, that can be easily accessed and managed. The results presented in this thesis involved many genome-scale analyses, which required to gather various data for several bacteria, including (Fig. 2.1):

- Genome sequence
- Annotation files (list of genes with coordinates, strand, possible protein product etc.)
- Gene expression data
- Promoter data (TSS locations, strand, genes controlled,  $\sigma$  factors and promoter elements)

These data were extracted from multiple sources (Fig. 2.1), mainly from literature and public databases, but also from team work and collaborations, and then uniformly formatted and integrated into a database.

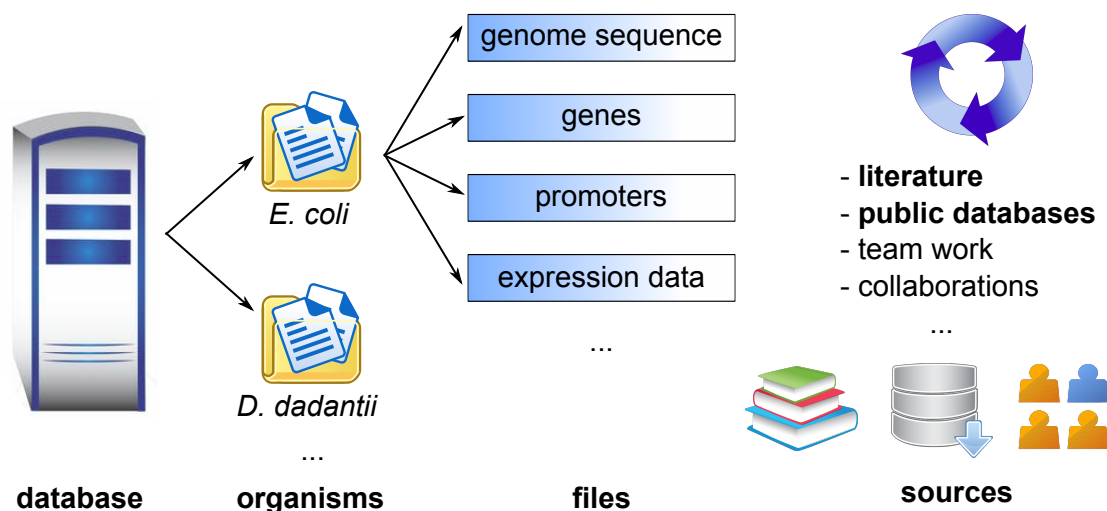


FIGURE 2.1: Schematic representation of the database. The latter is stored on a physical server, and contains one folder per organism, consisting of successive sub-folders containing informative files about the organism gathered from multiple sources, mainly from literature and public databases.

For instance, in order to analyse the statistical relation between promoter structure and gene response to topological changes for validating the models of transcriptional regulation by SC, I collected all available transcriptomic datasets from species where a TSS map was also available. Statistical procedures such as differential expression analysis were adjusted for each collected dataset due to strong differences in the experimental protocols (treatment with gyrase inhibitors vs mutation studies) and employed transcriptomic technologies (microarray vs RNA-seq). I also gathered gene responses to (i) environmental changes, where SC variations are induced naturally, and (ii) mutations in several NAPs, which modulate SC level. On the other hand, accurate and curated promoter maps are available only for *E. coli* [46], and remain incomplete or even lacking for other organisms, but were however required for the analysis. Consequently, genome-wide TSS maps were collected from the literature, and a prediction of promoter motifs was performed (see next section). The following table (Tab. 2.1) summarises most of the data collected for this specific purpose, whereas the additional datasets used for the characterisation of *D. dadantii* transcriptome are described in the corresponding chapter (Chapter 3).

## 2.2 Implementation of programming tools

In order to access the database and perform automatised genome-scale analyses in various organisms, the next step was to implement programming tools. While all species have their own ways of communicating, as human beings, we use different languages to communicate with each other. Similarly, as developers, we use programming languages to interact with computers, to which we transmit a code or program, *i.e.* a set of instructions to perform a specific task that may be time-consuming and inaccurate by hand. Like human languages, there are many different programming languages (estimated about 9000 by the Free-online Dictionary of Computing), that differ from their syntax, *i.e.* structure of the code, and semantics, *i.e.* meaning of the words. Among them, I chose a high-level programming language, *i.e.* more similar to human languages, Python, by which programs are easier to write, read, and maintain, although requiring a compiler or interpreter to translate the code into machine language. In contrast, low-level programming languages are faster and more memory efficient since they do not require any compiler to execute programs, although writing, reading, maintaining and debugging codes becomes more difficult. Beyond its advantages as a high level-programming language, Python is widely used in bioinformatics since it is open-source, and comes with a lot of modular and reusable sets of tools such as Biopython. Among the various existing approaches, I used an Object-Oriented Programming paradigm, which is a way of designing and structuring programs. It relies on the concept of classes and objects: an object is an instance of a class, representing a real-world entity, and having its own attributes, *i.e.* a set of properties, and methods, *i.e.* a set of functions which perform some useful action, which are both defined inside the class. For example (Fig. 2.2), the object corresponding to the *E. coli* genome is an instance of the genome class, inside which a method allows to load its genome sequence from the appropriate input file. In this framework, programs are divided into small parts (classes, objects and methods), easy to use and reuse, simplifying code development and maintenance. Altogether, using Python object-oriented programming, programs were developed to access the database, define classes and methods, allowing to create objects corresponding to the investigated biological entities, *e.g.* organisms, genes, promoters, integrate data from different sources, and eventually perform various genome-scale analyses required to achieve the thesis objectives (Fig. 2.2).

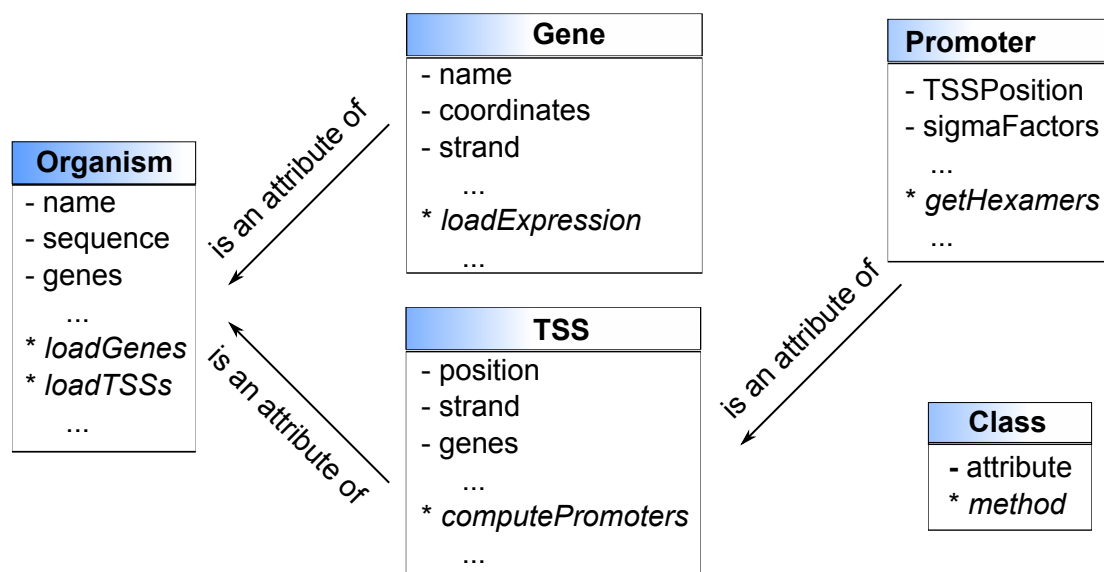


FIGURE 2.2: Schematic representation of the object oriented programming framework. Each class (in blue) contains methods (preceded by \*) to create objects and load their attributes (preceded by -) from the appropriate files stored in the database.

As an example, to predict promoter elements from TSS maps (Tab. 2.1), a tool was implemented by Antoine Heurtel, a student working in the team, based on bTSSfinder software [125]. The latter relies on position weight matrices, and predicts  $\sigma$  factor binding and promoter elements for a given sequence, for both enterobacteria and cyanobacteria. The developed tool requires a genome object, and an input TSS map. It then extracts sequences upstream of experimentally determined TSS positions, and predicts the presence of promoter elements, in which case it becomes a TSS attribute. Then, to investigate the link between promoter spacer length and gene response to a condition, I implemented a tool which takes as inputs (i) a genome object, (ii) a promoter map, and (iii) an expression dataset. Genes are first classified as significantly activated/repressed based on an user-defined threshold on the adjusted  $P$ -value, and on the log-fold-change (expression variation) value. Then, for each gene controlled by a TSS with promoter elements, the spacer length is computed from -35 and -10 coordinates, allowing to test the statistical relation between gene response to a given condition, and spacer length. Other examples are provided in the next subsection.

Once the database was developed and the programming tools implemented, this allowed us to (i) characterise the *D. dadantii* transcriptome, and (ii) validate the models of transcriptional regulation by SC, as developed in the upcoming chapters. Beyond these main thesis objectives, due to the practical aspect of this database and tools to integrate diverse datasets and perform various statistical analyses, I also contributed to side projects, which are summarised in the following subsection.

## 2.3 Side projects

As discovered early in the twin supercoiling domain model [126], transcription is an important source of SC. As RNAP cannot rotate freely around DNA template due to large viscosity, and because the same constraint applies for DNA rotation, it generates torsional stress during RNA elongation, negative (underwound) upstream and positive (overwound) downstream [127]. Those supercoils are able to diffuse along double-helix at distances of several kb, and reach nearby promoters, due to the high gene density of bacterial genomes [128, 129] (Fig. 2.3). Since SC in turn affects transcription [108], it was thus hypothesised that the relative distances and orientations between adjacent genes (convergent, divergent or tandem arrangements, Fig. 2.3) was a strong determinant of their response to SC variations, by modulating the distribution of SC at promoters. Due to the complex nature of this so-called transcription-supercoiling coupling (TSC), a stochastic regulatory model was developed by Bilal El Houdaigui and Sam Meyer, with the aim of predicting gene responses to SC variations, depending on their relative orientation. To validate model predictions at the genome-scale, the developed database and tools were particularly useful. Indeed, gene orientations can be deduced from genome annotation files present in the database, whereas promoter responses to SC variations can be inferred from the collected transcriptomic datasets (Tab. 2.1). By creating an automated tool, this allowed to validate the statistical relation between gene orientation and promoter response to SC variations in several organisms and experimental conditions, validating the TSC-mediated regulation, and leading to the contribution to an article published in Nucleic Acids Research [115].

Another side project concerns the investigation of the phenotypic and transcriptional effects of IHF mutation on *D. dadantii* virulence, a plant pathogenic bacterium which is presented in more detail in the next chapter (Chapter 3).

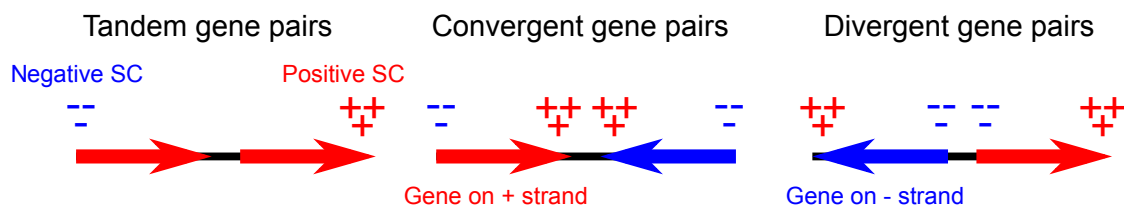


FIGURE 2.3: Impact of gene orientation on the distribution of SC generated by transcription. Elongating RNAP generate negative SC upstream, and positive downstream [126].

IHF is one of the major bacterial NAPs found in Proteobacteria [55]. It is a heterodimer formed by IHF $\alpha$  and IHF $\beta$  sub-units encoded by *ihfA* and *ihfB* genes, respectively, which induces sharp bends in DNA in a sequence-dependent manner (Tab. 1.2). In *E. coli* and *S. typhimurium*, IHF contributes to genome organisation, DNA replication and rearrangements, but also exerts a global transcriptional effect, by facilitating contacts between regulatory proteins and RNAP [51]. Its concentration within the cell varies during bacterial growth, increasing during the transition to stationary phase [51]. Since IHF has an important role in the infection process of *S. typhimurium* [130], its contribution to *D. dadantii* virulence and global gene expression was thus investigated based on a combination of phenotypic and RNA-seq approaches, and using bacteria carrying a mutation in *ihfA*, precluding the IHF heterodimer formation. The tools implemented allowed (i) to process RNA-seq data and define a list of genes significantly activated/repressed by IHF, (ii) to perform functional enrichment analyses to identify the biological functions regulated by IHF, and (iii) to integrate IHF predicted binding sites and the transcriptional response of genes, to identify putative direct vs indirect regulatory effects. Besides, in *D. dadantii*, changes in NAP binding, SC, physicochemical sequence properties and other factors were shown to organise coherent domains of gene expression (CODOs) ranging from several to hundred kb, and emerging transiently in particular conditions [119]. These CODOs have been the subject of a review published in *Microorganisms* to which I contributed [131]. The role of IHF in the organisation of CODOs was also investigated, involving many statistical analyses described in the corresponding article, and performed with the present database and tools. Altogether, this led to the contribution to an article published in *Nucleic Acids Research* [122].

Finally, I contributed to an article which is currently under revision for publication in *Nucleic Acids Research* [121], which characterises the transcriptional effect of the non-marketed antibiotic seconeolitsine in *D. dadantii*, providing the first insights into topoisomerase I inhibition in Gram-negative bacteria, and reassessing the genes' intrinsic property of "supercoiling-sensitivity" (more details in Chapter 5).



Species	NCBI reference genome	Expression condition and reference	Transcription start sites map reference
<i>Escherichia coli</i>	NC000913.2	norfloxacin (SC +) [109]	[46] [110]
		cold shock (SC -) [111] [112]	
		oxidative shock (SC +) [112]	
		heat shock (SC +) [113] [112]	
		osmotic shock (SC -) [113]	
		acidic shock (SC +) [114]	
	CP000819.1	experimental evolution (SC -) [115]	
<i>Salmonella typhimurium</i>	NC016856.1	novobiocin (SC +) [116]	[117]
	FQ312003.1	mutation study (SC +) [118]	
<i>Dickeya dadantii</i>	NC014500.1	novobiocin (SC +) [115] [119]	[120]
		seconeolitsin (SC -) [121]	
		NAP mutants (Fis, H-NS and IHF) [119] [122]	
		oxidative shock (SC +) [119]	
		acidic shock (SC +) [119]	
		osmotic shock (SC -) [119]	
<i>Synechococcus elongatus</i>	SYNPCC7942	correlation SC wave- form [123]	[124]
<i>Mycoplasma pneumoniae</i>	NC000912.1	novobiocin (SC +) [48]	[48]

TABLE 2.1: Summary of the data used to investigate the link between promoter structure and transcriptomic response under SC variations. A blank cell indicates a value identical to previous row. SC + refers to an increase in absolute SC level (DNA relaxation), SC - to a decrease (DNA overtwisting), as stated in Chapter 1. Several references are indicated when multiple datasets were available.

## Chapter 3

# Mapping the complex transcriptional landscape of the phytopathogenic bacterium *Dickeya dadantii*

In order to validate the models of transcriptional regulation by SC, we wished to extend our analyses to the phytopathogen *D. dadantii* in which SC plays a key role during plant infection, and for which extensive transcriptomic data have been accumulated in the team, including the genome-wide response of genes to SC variations. Yet the map of its transcriptome, including transcription units, transcription start sites and promoters, is lacking, but however required to validate the models. A part of my thesis was consequently dedicated to the characterisation of its transcriptome, which was also a great opportunity to significantly contribute to further progress in the field of phytopathogenicity, by providing the first transcriptomic map of this widely used model species of the *Dickeya* genus.

### 3.1 Role of DNA supercoiling in *D. dadantii*

Bacteria of the *Dickeya* genus are Gram-negative, necrotrophic phytopathogens, *i.e.* that kill their host and then feed on the dead tissues, causing soft rot in a wide range of plant hosts worldwide, including agriculturally important crops [132]. This severe disease leads to tissue maceration and eventually plant death [133]. Besides, the ability of *Dickeya* species to colonise various ecological niches such as soil, ground water, surfaces and interior of plant hosts and insects [7, 134], makes them remarkably versatile models for the

scientific community. *Dickeya dadantii* strain 3937 is the main model used for studying the regulation of gene expression during the plant pathogenic process, since its genome was the first to be sequenced and curated in the *Dickeya* genus [135].

As introduced in the review from Chapter 1 [108], pathogenic bacteria must cope with numerous environmental variations during infection, including changing sources or scarcity of nutrients and oligoelements, and various

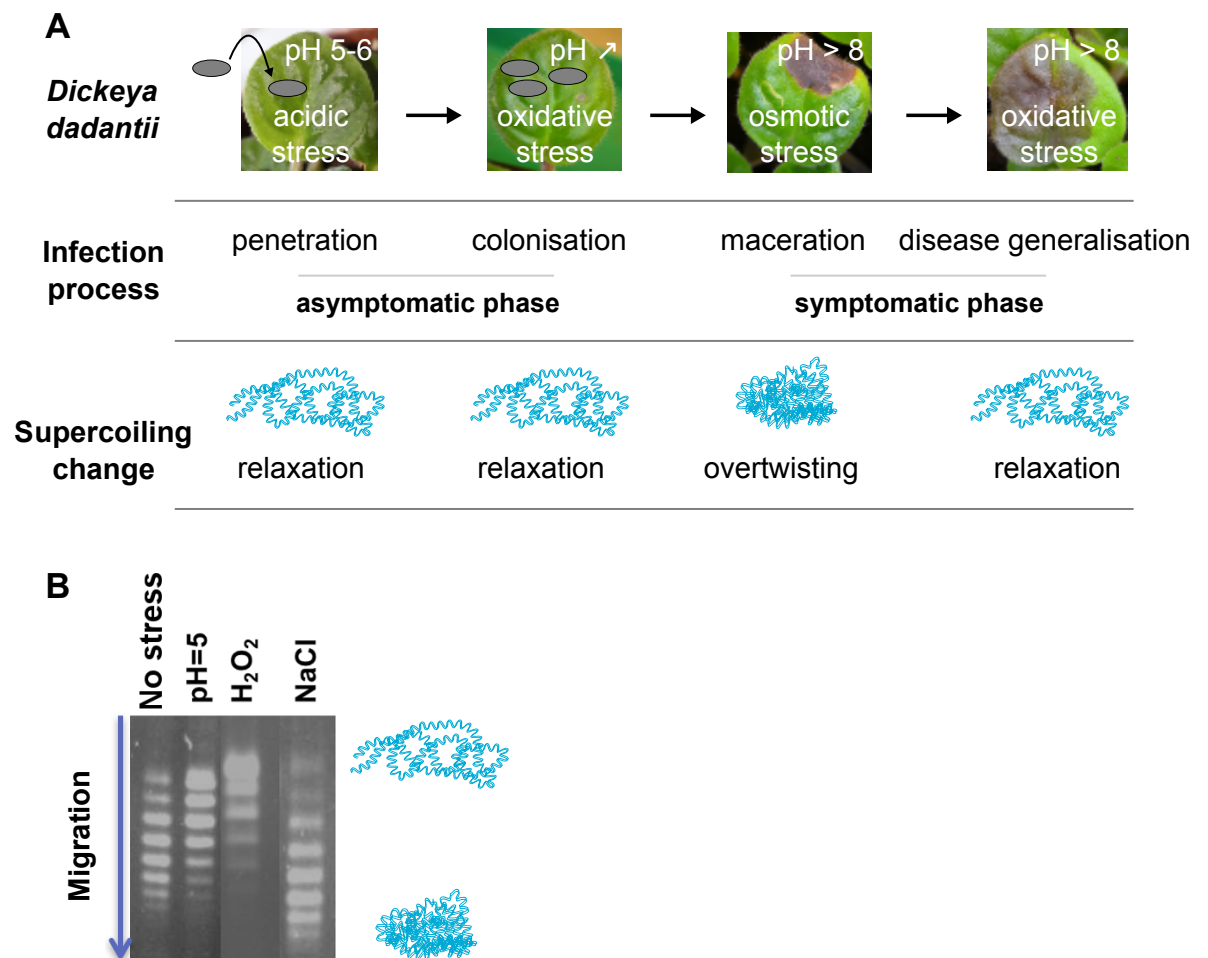


FIGURE 3.1: Schematic representation of *Dickeya dadantii* infection process, and changes in chromosomal SC level in response to stress. **(A)** During the asymptomatic phase comprising plant penetration and colonisation, bacteria experience acidic and oxidative stresses, respectively. During the symptomatic phase comprising plant tissue maceration and disease generalisation, bacteria experience osmotic and oxidative stresses, respectively. **(B)** Adapted from [136]: these various stresses distinctly affect SC level, as shown by the analysis of plasmid pUC18 topoisomers separated on agarose gels containing 2.5 µg/ml chloroquine.

types of stress and host defence reactions. The infection process of *D. dadantii* is summarised in the upcoming article, and is described in detail elsewhere [133]. Briefly, its pathogenesis comprises an asymptomatic phase, where bacteria penetrate and colonise plant tissues, and a symptomatic phase where enzymatic maceration of plant tissues is induced until soft rot generalisation occurs (Fig. 3.1). At each stage of the process, bacteria are exposed to changing conditions, including (i) an acidic stress resulting from the low pH of plant apoplasts, *i.e.* the invaded free diffusional space outside the plasma membrane [137], (ii) an oxidative stress resulting from plant defences reducing oxygen availability [138], (iii) an osmotic stress resulting from the release of plant cell components induced by degrading enzymes [133] (Fig. 3.1). For successful infection, bacteria must sense and respond to each of these stresses, requiring a subtle coordination of gene expression in time and space, including genes involved in the core metabolism, in the general response to stresses but also specifically in the encountered stress, and in the virulence program and acquisition of adaptive traits [7]. SC thus appears as an excellent candidate for both relaying the infection stage-specific environmental signals to the bacterial chromosome, and adapting its global transcriptional program [108]. Indeed, all those environmental stresses have been associated with rapid and transient SC variations, and in particular in *D. dadantii* [136] (Fig. 3.1). In turn, novobiocin-induced DNA relaxation triggers a global and complex response of genes in this phytopathogen [115, 119], including those encoding the main virulence factors, namely pectinolytic enzymes (*pel* genes) [108]. Thus, it is likely that SC contributes to global gene expression reprogramming in response to environmental changes during *D. dadantii* infection process, highlighting the relevance of including this species in the validation of the models of transcriptional regulation by SC. To that end, the first step was to obtain a map of its transcriptome, which is the focus of the following article currently under a second round of revision for publication in the mBio journal [120].

## 3.2 Article

*N.B.* Supplementary figures and tables can be found online on bioRxiv [120].

# 1 Mapping the complex transcriptional 2 landscape of the phytopathogenic 3 bacterium *Dickeya dadantii*

4 Raphaël Forquet,<sup>a</sup> Xuejiao Jiang,<sup>a</sup> William Nasser,<sup>a</sup> Florence Hommais,<sup>a</sup> Sylvie  
5 Reverchon,<sup>a</sup> Sam Meyer<sup>a</sup>

6 Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, CNRS, UMR5240, MAP, F-69622, France<sup>a</sup>

7 **ABSTRACT** *Dickeya dadantii* is a phytopathogenic bacterium that causes soft rot in  
8 a wide range of plant hosts worldwide, and a model organism for studying virulence  
9 gene regulation. The present study provides a comprehensive and annotated tran-  
10 scriptomic map of *D. dadantii* obtained by a computational method combining five  
11 independent transcriptomic datasets: (i) paired-end RNA-seq data for a precise re-  
12 construction of the RNA landscape; (ii) DNA microarray data providing transcriptional  
13 responses to a broad variety of environmental conditions; (iii) long-read Nanopore  
14 native RNA-seq data for isoform-level transcriptome validation and determination of  
15 transcription termination sites; (iv) dRNA-seq data for the precise mapping of transcrip-  
16 tion start sites; (v) *in planta* DNA microarray data for a comparison of gene expression  
17 profiles between *in vitro* experiments and the early stages of plant infection. Our re-  
18 sults show that transcription units sometimes coincide with predicted operons but are  
19 generally longer, most of them comprising internal promoters and terminators that  
20 generate alternative transcripts of variable gene composition. We characterise the  
21 occurrence of transcriptional read-through at terminators, which might play a basal  
22 regulation role and explain the extent of transcription beyond the scale of operons.  
23 We finally highlight the presence of noncontiguous operons and excludons in the *D.*  
24 *dadantii* genome, novel genomic arrangements that might contribute to the basal co-  
25 ordination of transcription. The highlighted transcriptional organisation may allow *D.*  
26 *dadantii* to finely adjust its gene expression programme for a rapid adaptation to fast  
27 changing environments.

28 **IMPORTANCE** This is the first transcriptomic map of a *Dickeya* species. It may there-  
29 fore significantly contribute to further progress in the field of phytopathogenicity. It  
30 is also one of the first reported applications of long-read Nanopore native RNA-seq in  
31 prokaryotes. Our findings yield insights into basal rules of coordination of transcrip-  
32 tion that might be valid for other bacteria, and may raise interest in the field of micro-  
33 biology in general. In particular, we demonstrate that gene expression is coordinated  
34 at the scale of transcription units rather than operons, which are larger functional ge-  
35 nomic units capable of generating transcripts with variable gene composition for a  
36 fine-tuning of gene expression in response to environmental changes. In line with re-  
37 cent studies, our findings indicate that the canonical operon model is insufficient to  
38 explain the complexity of bacterial transcriptomes.

39 **KEYWORDS:** phytopathogen, transcriptional regulation, transcription unit,  
40 transcriptional read-through, transcription start and termination sites

**Compiled** October 13, 2021

This is a draft manuscript, pre-submission

Address correspondence to Raphaël Forquet,  
raphael.forquet@insa-lyon.fr, Sam Meyer,  
sam.meyer@insa-lyon.fr.

Forquet et al.

#### 41 INTRODUCTION

42 Classically, bacterial transcription is described with the model of Jacob and Monod  
43 based on operons, defined as sets of contiguous and functionally-related genes co-  
44 transcribed from a single promoter up to a single terminator (1). In recent years how-  
45 ever, accumulating studies demonstrated that most operons actually comprise inter-  
46 nal promoters and terminators, generating transcripts of variable gene composition,  
47 generally in a condition-dependent manner (2, 3, 4, 5). This phenomenon, also known  
48 as suboperonic regulation (6), might be compared to alternative splicing in eukary-  
49 otes (7) and demonstrates a higher complexity of bacterial transcriptional landscapes  
50 than previously thought. Besides, transcription has been shown to extend beyond  
51 operons (3, 8), the latter being actually part of larger functional genomic units, referred  
52 to as transcription units (TUs) throughout the manuscript.

53 While transcriptomic maps have been established for various bacteria including  
54 *Escherichia coli* (9), *Salmonella enterica* (10), *Bacillus subtilis* (2), *Streptococcus pneumo-*  
55 *niae* (4), *Campylobacter jejuni* (11), *Clostridium beijerinckii* (12), *Mycobacterium tuberculo-*  
56 *sis* (13), *Mycoplasma pneumoniae* (14), and the phytopathogen *Xanthomonas campestris* (15,  
57 16), they are still lacking for *Dickeya species*. This study aims to provide the first com-  
58 prehensive and annotated transcriptomic map of *Dickeya dadantii*, a Gram-negative  
59 phytopathogenic bacterium representative of the *Dickeya* genus that causes soft rot,  
60 a severe disease leading to tissue maceration and eventually plant death (17) in a wide  
61 range of plant hosts worldwide, including agriculturally important crops (18, 19, 20, 21,  
62 22).

63 The infection process involves an asymptomatic phase, where bacteria remain  
64 latent, penetrate and colonise plant tissues, consuming simple sugars and small solu-  
65 ble oligosaccharides available in the plant apoplast to grow exponentially (23). In this  
66 compartment, bacteria are exposed to acidic conditions (24) and oxidative stress (25)  
67 resulting from plant defences. When all nutrients are consumed in the apoplast, the  
68 symptomatic phase initiates. Bacteria produce plant cell wall degrading enzymes (mainly  
69 pectinases) leading to the soft rot symptoms, and start cleaving pectin, which is used  
70 as a secondary carbon source for a new round of growth (26). By causing a total de-  
71 struction of plant cells, the maceration of plant tissues releases both vacuolar and  
72 cytoplasmic components in the apoplast, exposing the bacteria to osmotic stress (23).

73 In order to characterise the *D. dadantii* transcriptional landscape, we used a com-  
74 bination of transcriptomic data generated *in vitro* in a broad range of growth and  
75 stress conditions reflecting some of the key environmental signals encountered dur-  
76 ing the plant infection process, and ensuring optimal reproducibility and quality of  
77 analysed RNAs (27, 28). Different techniques were used, providing complementary  
78 knowledge: high-resolution Illumina paired-end RNA-seq; DNA microarray; Nanopore  
79 native RNA-seq; dRNA-seq. These data were combined using an integrative computa-  
80 tional method developed for this study, allowing the inference of the RNA landscape  
81 and a validation of co-expression occurring among genes of the same TU. This anal-  
82 ysis provides a detailed and annotated map of the TUs defining the *D. dadantii* tran-  
83 scriptome, *i.e.*, the sets of contiguous co-expressed genes. We then quantitatively map  
84 transcription start and termination sites in the investigated conditions, and analyse the  
85 associated predicted promoter and terminator motifs. We show that TUs sometimes  
86 coincide with predicted operons but are generally longer, most of them exhibiting in-  
87 ternal promoters and terminators. We characterise the occurrence of transcriptional  
88 read-through at terminators, a mechanism proposed as a basal coordinator and reg-  
89 ulator of gene expression yet never explored in phytopathogens and still poorly un-  
90 derstood across genomes in general. We finally detect putative noncontiguous oper-

Mapping of *Dickeya dadantii* transcriptional landscape

91 ons and excludons in the *D. dadantii* genome. In order to validate the obtained tran-  
92 scriptional map, we analyse available *in planta* expression data, and show that TUs  
93 inferred from *in vitro* cultures are also co-expressed during the early stages of plant  
94 infection (29), suggesting that many of the analysed features are used by *D. dadantii*  
95 in the pathogenic context. This transcriptomic map might serve as a community re-  
96 source to help elucidating the regulation of *D. dadantii* gene expression, including its  
97 virulence programme. It also provides insights into basal rules of coordination of tran-  
98 scription that might be valid for other bacteria, specifically for other *Dickeya* species  
99 for which a core genome of 1300 genes was identified by comparative genomics (30).

## 100 RESULTS AND DISCUSSION

101 **Characterisation of *Dickeya dadantii* transcription units.** In order to generate  
102 a biologically relevant transcriptional map of *D. dadantii*, we combined and integrated  
103 four sets of transcriptomic data obtained from *in vitro* cultures subjected to differ-  
104 ent sugar sources, environmental stress factors (acidic, oxidative, osmotic stress), and  
105 variations of DNA supercoiling, reflecting a variety of conditions also encountered by  
106 bacteria in the course of plant infection. A fifth set obtained from bacteria grown *in*  
107 *planta* was used for validation. These data were collected by different experimental  
108 methods providing complementary information, as follows (a more detailed descrip-  
109 tion of the datasets is provided in Materials and Methods).

110 Dataset 1 was generated from high-resolution Illumina paired-end, strand-specific  
111 RNA-seq covering 6 growth conditions: M63 minimal medium supplemented with  
112 sucrose, addition of polygalacturonate (PGA), a pectic polymer present in plant cell  
113 wall (31), and treatment by novobiocin, which induces a global and transient chro-  
114 mosomal DNA relaxation (32) in exponential or in early stationary phase. By providing  
115 short but precise sequencing reads at single base-pair resolution and high sequencing  
116 depth, this dataset yields precise and quantitative information on the RNA landscape.

117 Dataset 2 was generated from DNA microarray data covering 32 growth condi-  
118 tions, involving the presence of PGA and leaf extracts, and in each medium, a separate  
119 exposure to acidic, oxidative or osmotic stresses (28). This dataset provides a quanti-  
120 tative catalogue of genes' responses to a more comprehensive and detailed range of  
121 conditions than dataset 1, albeit of weaker spatial resolution.

122 Dataset 3 was generated from long-read Nanopore native RNA-seq in M63 min-  
123 imal medium supplemented with glucose and PGA, pooled from samples obtained  
124 in both exponential and early stationary phases. This method allows native RNAs to  
125 be sequenced directly as near full-length transcripts from the 3' to 5' direction, with a  
126 weaker depth than the previous datasets. Only a few transcriptomes were analysed by  
127 this technique, mostly from viral and eukaryotic organisms (33, 34, 35, 36), and, to our  
128 knowledge, a single prokaryotic one (37). This dataset provides a direct isoform-level  
129 validation of the TUs, and an accurate definition of transcription termination sites.

130 Dataset 4 was generated from differential RNA sequencing (dRNA-seq) experiments  
131 carried out on four samples obtained by pooling RNAs from the large variety of envi-  
132 ronmental conditions of dataset 2 followed by treatment with Terminator exonuclease  
133 (TEX) prior to sequencing. TEX enzyme degrades processed 5'-monophosphate RNAs  
134 and consequently enriches the samples in primary 5'-triphosphate end transcripts (38),  
135 thus locating transcription start sites at single-nucleotide resolution.

136 Finally, dataset 5 was generated from *in planta* DNA microarray data, 6 and 24  
137 hours post-inoculation of the model plant *Arabidopsis thaliana* (29), during the early  
138 stages of infection. Bacterial RNAs are difficult to isolate from plant tissues, especially  
139 during the symptomatic phase where phenolic compounds accumulate in decaying



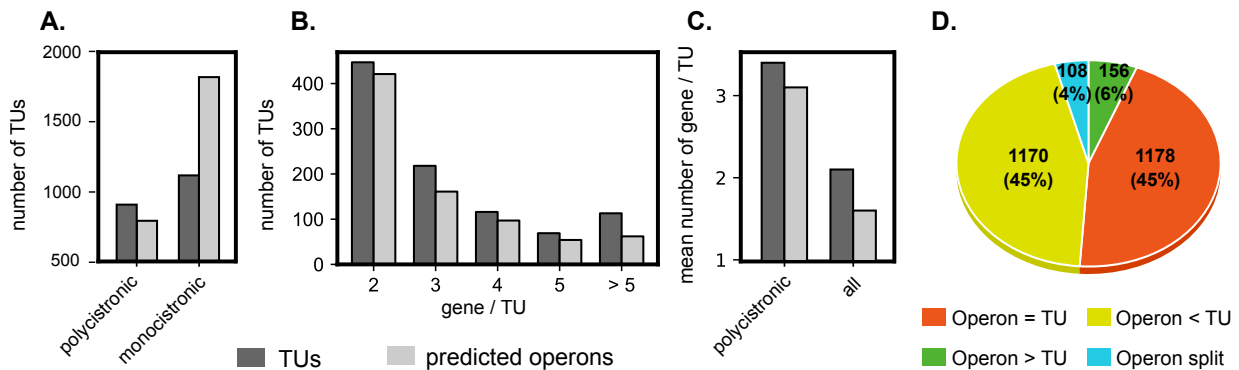
Forquet et al.

140 tissues, explaining the lack of transcriptomic data during the late stages of infection.  
141 In spite of a limited variety of conditions, this dataset allows a comparison of gene  
142 expression profiles between *in vitro* and *in planta* experiments, and was used to val-  
143 idate the level of co-expression of genes within TUs during the early stages of plant  
144 infection.

145 This collection of diverse and complementary transcriptomic datasets provided a  
146 solid ground for precisely characterising the *D. dadantii* transcription units, rather than  
147 basing our analysis on genomic data alone as in most operon predictors (intergenic  
148 distances between genes, functional links among products). The employed algorithm  
149 is described in details in Materials and Methods. Shortly, in a first step, we analysed  
150 the RNA landscape from Illumina paired-end strand-specific RNA-Seq (dataset 1), en-  
151 suring good resolution and sufficient sequencing depth to obtain a quantitative signal  
152 for all genes. These data also allowed us to uncover 50 putative coding genes previ-  
153 ously unannotated, most of which exhibiting sequence homology with proteins from  
154 the *Dickeya* genus (Supplementary Tab. S1D). Putative TUs were defined by fusing ad-  
155 jacent genes as long as RNA fragments were found in their intergenic region, a signa-  
156 ture of co-transcription. Secondly, if genes within the same putative TU are indeed  
157 co-transcribed, they should exhibit strong correlation of expression in a wider range  
158 of conditions than those of dataset 1. This analysis was carried using the diversity of  
159 samples in our DNA microarray data (dataset 2), based on a customised hierarchical  
160 clustering framework (39). This second criterion (correlation of expression) provided  
161 an orthogonal cross-validation compared to the first one (intergenic RNA signal), and  
162 yielded a total of 2028 putative TUs along the *D. dadantii* genome. In a third step,  
163 these TUs were validated based on Nanopore native RNA-seq (dataset 3). We tested  
164 the presence of long native RNA reads overlapping adjacent genes belonging to the  
165 same TU, thus yielding a direct evidence of co-transcription. For 16% of adjacent gene  
166 pairs, no conclusion could be drawn because of insufficient coverage. For the others,  
167 co-transcription was confirmed in 92% of the cases; for the remaining 8%, the absence  
168 of a common RNA might be indicative of false positives, but for some of them, may  
169 also be due to the weak number of culture conditions included in dataset 3. Since  
170 the large majority of TUs defined from datasets 1 and 2 match the observations of  
171 Nanopore native RNA-seq, we favoured the latter hypothesis and retained all of them,  
172 with a confidence level reflecting the presence or absence of overlapping RNA reads  
173 (Supplementary Tab. S1A).

174 With this approach, we mapped the first layer of transcription organisation in *D.*  
175 *dadantii*. According to our findings, the 4211 protein-coding genes are organised into  
176 2028 transcription units (provided in Supplementary Tab. S1A), among which 1118 are  
177 monocistronic and 910 are polycistronic, ranging from 2 to 28 genes (Fig. 1A, 1B and  
178 Supplementary Tab. S5). At the genomic scale, we compared our results with those  
179 of Rockhopper, a popular operon predictor that uses expression data as well as ge-  
180 nomic information as input (40). 45% of predicted operons exactly coincide with a TU  
181 in our analysis (Fig. 1D), including known examples such as *smtAmukFEB* involved in  
182 chromosome partitioning (Fig. 2A) (41). Besides, many identified TUs are likely oper-  
183 ons of unknown functions and features (Fig. 2B), which represent interesting starting  
184 points to discover new transcriptional functional units. Remarkably, TUs are generally  
185 longer than predicted operons: the average TU (including monocistronic ones) con-  
186 tains 2.1 genes and the average polycistronic TU contains 3.4 genes, against 1.6 and  
187 3.1 respectively for predicted operons (Fig. 1C). Almost three quarters (73.5%) of all  
188 genes are co-transcribed in TUs, against 56.9% for predicted operons (Fig. 1A, Supple-  
189 mentary Tab. S5). Our results indicate that TUs are indeed larger functional genomic



Mapping of *Dickeya dadantii* transcriptional landscape

**FIG 1** (A) Repartition of monocistronic and polycistronic TUs identified by our analysis and comparison to predicted operons. (B) Size distributions. (C) Average number of genes per TU. (D) Fate of predicted operons that are mostly found as or within TUs in our algorithm.

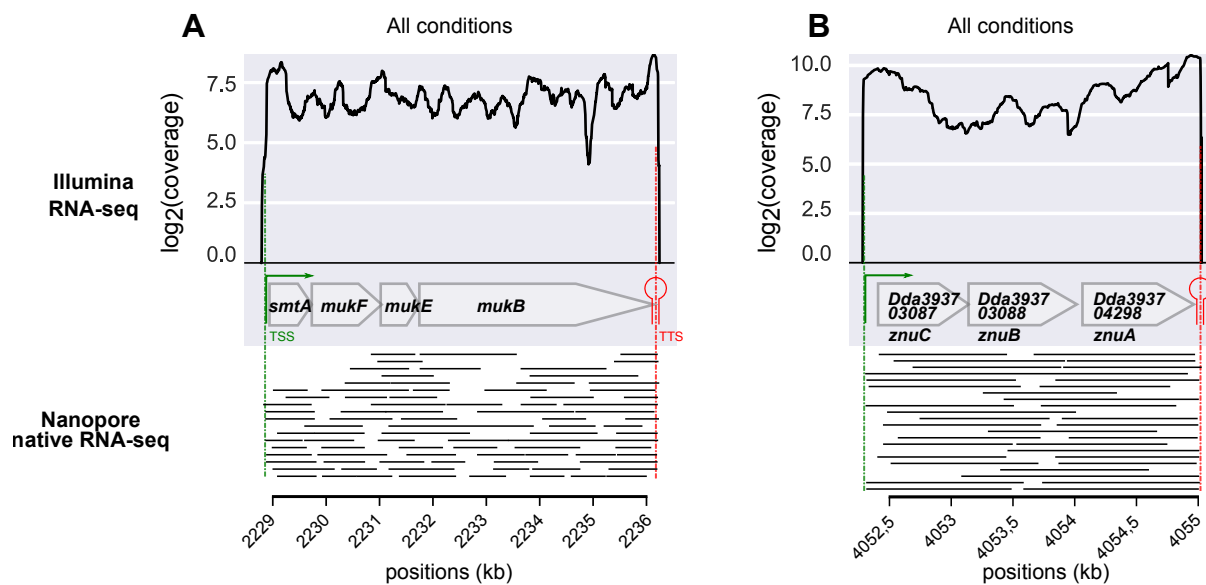
units, since 45% of predicted operons are extended by at least one gene (Fig. 1D), in agreement with recent findings in *E. coli* based on long-read sequencing (3).

As an example, the *sapABCD* operon encoding a transporter involved in antimicrobial peptide resistance and virulence in numerous bacteria including *D. dadantii* (42) is extended to include the enoyl-acyl carrier protein reductase *fabI* that catalyses an essential step in the biosynthesis of fatty acids of the membrane (43) (Fig. 3B). It might be noted that *fabI* has a different genomic location in *E. coli* and is consequently not co-transcribed with *sapABCD* in that species (44) although this synteny is conserved in other *Dickeya* genomes, showing that TUs can merge and/or vary over time at the evolutionary scale. Since these genes are functionally unrelated (except for a general relation with the membrane), the biological relevance and putative role of this event requires further investigation.

The *glg* genes involved in glycogen metabolism constitute another instructive example. They were initially classified in two separate operons in *E. coli* (45), and later identified as a single TU involving alternative transcripts of variable gene composition depending on growth conditions (46). The latter is also true in *D. dadantii* according to our findings (Fig. 3C), illustrating how transcription extends beyond the scale of the operon.

**Genome-wide identification of *D. dadantii* transcription start and termination sites.** Once *D. dadantii* transcription units were defined, the next step was to elaborate a map of transcription start sites (TSSs) and transcription termination sites (TTSs) for each TU along the genome. First, as mentioned above, dRNA-seq experiments were carried out to build a large library of 9313 putative TSSs at high-resolution (38) covering a wide range of *in vitro* cultures under growth and stress conditions also encountered during plant infection (dataset 4, Supplementary Tab. S2A). These were obtained by treating the RNA samples with TEX prior to sequencing, and the TSSer workflow was applied for a precise determination of TSS positions (48), followed by visual curation (Materials and Methods). For TTSs, two sets of putative positions were generated based on (i) Nanopore native RNA-seq (dataset 3), where transcripts are sequenced from the 3' ends, allowing the detection of 1165 TTS positions based on the enrichment of these ends downstream of gene stops (Supplementary Tab. S2D); (ii) genome-wide predictions of termination sites, based on the two main mechanisms of

Forquet et al.



**FIG 2** Transcription units identified by our approach, and coinciding with operons. (A) Example of a known operon (*smtAmukFEB*). The bottom panel shows the coordinates of the long native RNA reads sequenced by Nanopore. (B) Identification of a new TU exhibiting uniform read coverage and strong internal cross-correlations (Supplementary Fig. S1A), clearly indicative of an operon. Its function was unknown but a homology analysis revealed that it corresponds to the cluster of genes *znuCBA*, a  $Zn^{2+}$  uptake system (47). Long reads are observed for all adjacent gene pairs in Nanopore native RNA-seq data, and even a fragment carrying the three genes for *znuCBA*.

transcription termination in bacteria. 3564 Rho-independent (intrinsic) TTSs and 5851 Rho-dependent (regulated) TTSs (49) were predicted using ARNold (50) and RhoTermPredict (51) programmes respectively (Supplementary Tabs. S2B and S2C).

A quantitative mapping of the transcription landscape was then performed in order to estimate the contribution of each TSS/TTS to its TU. While most comparable maps define TSSs/TTSs by their position only, we exploited the complementarity of the input data to also systematically analyse their magnitude (or strength) in the investigated conditions. The +TEX libraries, Nanopore reads and TTS predictions are not suitable for the latter purpose, which required building a second list of TSSs and TTSs of poorer resolution but quantitative magnitude from the non-treated paired-end RNA-seq data (dataset 1). Briefly, TSSs and TTSs were defined based on the enrichment in RNA fragment starts and stops upstream of gene starts and downstream of gene stops respectively, and the number of fragments associated to these sites across all samples was considered as the global strength. The lists obtained with the three methods (from datasets 1, 3 and 4) were then merged into a unified list of TSSs/TTSSs of optimal spatial resolution, quantitative magnitude, and with an estimated level of confidence depending on the level of agreement between these datasets (see Materials and Methods). These TSSs and TTSs were then assigned to the TUs. In order to eliminate many very weak internal TTSs/TTSSs (most of which likely have poor biological relevance), the latter were retained only if they yielded at least 15% of the total start/stop magnitude of the TU and were thus used at least in some of the investigated conditions. As a result, we defined a total of 2595 TSSs and 1699 TTSs (including internal ones) over all TUs (Supplementary Tab. S1A to S1C). Inevitably, some alternate TSSs/TTSSs may be absent from these lists if they are specifically used in conditions not

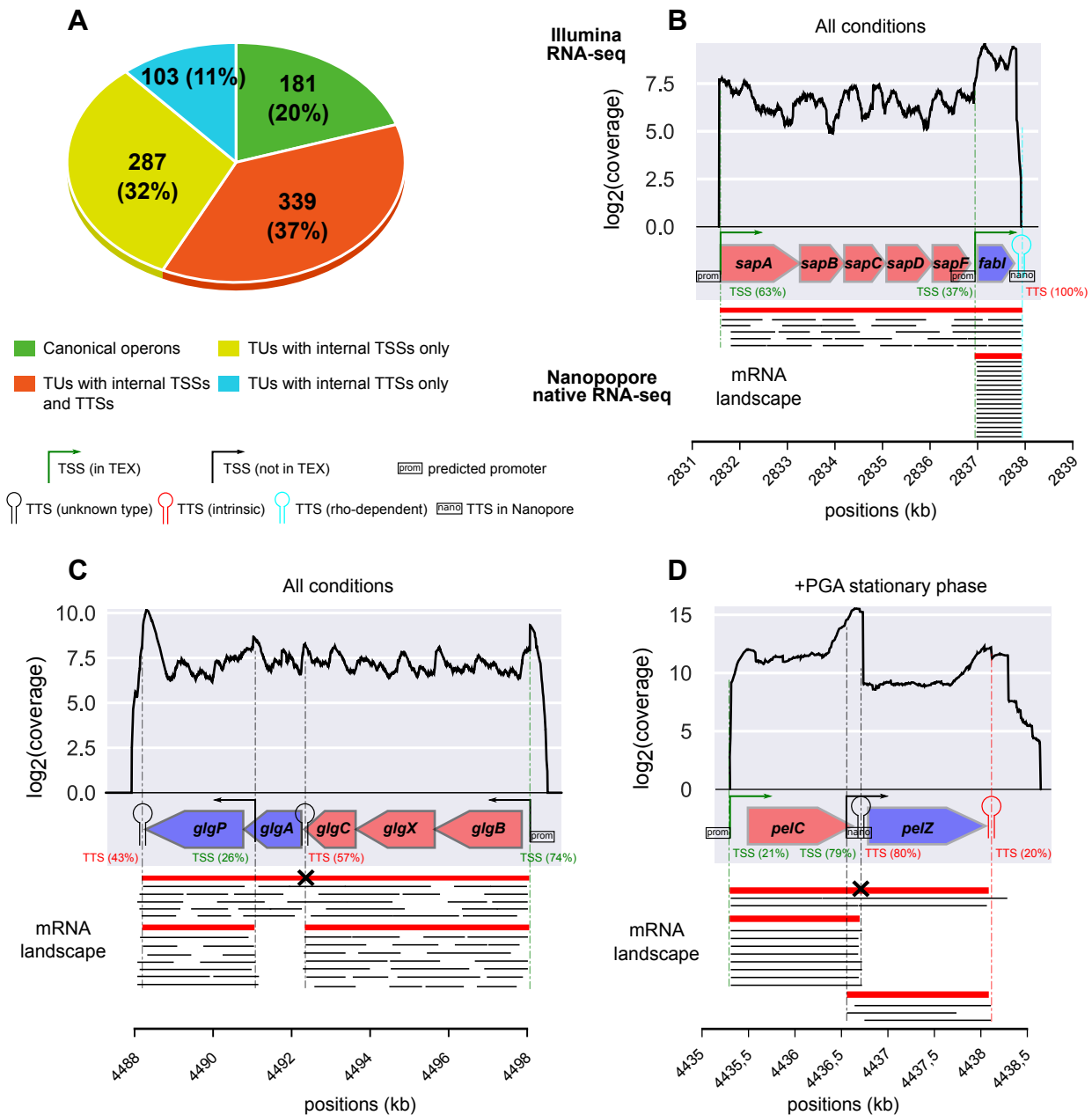
Mapping of *Dickeya dadantii* transcriptional landscape

246 included in our datasets. Finally, a scan for promoter motifs, conducted with bTSS-  
247 finder (52), identified promoters upstream of 1848 (71%) TSSs in total (Supplementary  
248 Tab. S1B and Fig. 3). The absence of detected promoters for the remaining 29% TSSs  
249 was expected due to the limitations of such predictors (53). To evaluate the quality  
250 of our TSS definition, we compiled all experimentally determined TSSs in *D. dadantii*  
251 (by primer extension), and compared their positions to our findings (Supplementary  
252 Tab. S3). 45% displayed exactly the same position, 38% were distant by less than 5 nu-  
253 cleotides, and only 17% were distant by more than 6 nucleotides. Manually-annotated  
254 promoter elements from these studies also match our findings well (Supplementary  
255 Tab. S3).

256 **Characterisation of a complex transcriptional landscape.** The quantitative map-  
257 ping of TSSs and TTSs allowed us to refine the comparison of TUs and operons pre-  
258 sented above. According to our findings, only 20% of polycistronic TUs (181) exhibit  
259 a single promoter and terminator (Fig. 2 and 3A) and thus fit into the classical defi-  
260 nition of operons, and only 47% of these (85) are predicted as such by Rockhopper.  
261 The 80% remaining TUs (729) are complex (Fig. 3A). 32% (287) have at least one internal  
262 TSS without any internal TTS, such as *sapABCDffabl* (Fig. 3B). 37% (339) have both  
263 internal TSS(s) and TTS(s), such as *glgBXCAP* (Fig. 3C) and *pelCZ* (Fig. 3D). Finally, 11%  
264 (103) have at least one internal TTS without any internal TSS such as *rhlB-gppA-pehV*  
265 (Fig. 4A), *pelD-paeY-pemA* (Fig. 4B) and *gcvTHP* (Fig. 6). Most *D. dadantii* TUs can conse-  
266 quently generate alternative transcripts of variable gene composition, resulting in a  
267 dense and complex transcriptional landscape.

268 A notable feature of complex TUs is the heterogeneity of transcription levels along  
269 the genes due to internal TSSs / TTSs, usually in a condition-dependent manner, re-  
270 sulting in a moderate correlation in the expression of genes within the TU (9). As  
271 an example, in the *sapABCDffabl* TU (Fig. 3B), *fabI* is expressed both as part of the  
272 entire transcript and as an independent transcript generated from a strong internal  
273 TSS, explaining the lower correlation between *fabI* and the remaining genes (Supple-  
274 mentary Fig. S1B). In *glgBXCAP* (Fig. 3C), alternative transcripts of variable gene com-  
275 position can be generated depending on TSS and TTS usage. Another example rele-  
276 vant to plant infection is the *pelCZ* cluster (Fig. 3D) encoding two endopeptidase lyases  
277 secreted by *D. dadantii* which degrade pectin contained in plant cell walls (56). The  
278 substrates of Pel enzymes are pectic oligomers, e.g. PGA, that act as inducers of *pel*  
279 expression (31). The *pelCZ* genes were previously shown by Northern blotting to be co-  
280 transcribed into a single polycistronic transcript under inducing conditions by PGA, in  
281 addition to the two monocistronic mRNAs encoded by *pelC* or *pelZ* under non-inducing  
282 conditions (55). Our present findings are in full agreement with these observations,  
283 as *pelCZ* is detected as a single TU harbouring one internal TSS and TTS, each giving  
284 rise to monocistronic transcripts. In our data, *pelCZ* expression profiles are similar in  
285 presence or absence of PGA in spite of a drastically different global expression level  
286 (Supplementary Fig. S1D), suggesting that in absence of inducer, this very low level  
287 previously prevented a reliable detection of the entire transcript. Altogether, our find-  
288 ings clearly indicate that the canonical operon model is insufficient to explain the com-  
289 plexity of the *D. dadantii* transcriptional landscape, in line with results in many other  
290 organisms (2, 3, 4, 5). The existence of alternative entry and exit points for RNA Poly-  
291 merase inside TUs allows the cells to adjust the relative expression level of adjacent  
292 genes within a global coordination of expression of the entire TU (Fig. 3) that may al-  
293 low, in the case of *D. dadantii* during plant infection, a rapid adaptation to changing  
294 environment.

Forquet et al.



**FIG 3** (A) Characteristics of TUs. (B) The *sapABCDF* and *fabI* genes, predicted by Rockhopper (40) as two separate (red and blue) operons, were identified as a single TU, with a strong internal TSS expressing *fabI* alone. The bottom panel indicates the different isoforms (red) and the long reads sequenced by Nanopore native RNA-seq (black). The latter overlap all adjacent gene pairs, providing a direct evidence for co-transcription. (C) The *glg* genes were identified as a single TU (involving several isoforms) containing two separate predicted operons (blue and red genes), as suggested by the uniform read coverage, long reads from Nanopore native RNA-seq (bottom), and in line with results in *E. coli* (3, 54, 46). (D) Identification of the *pelCZ* TU with different isoforms depending on the condition, as previously determined (55). The two genes are split into different operons by Rockhopper. A strong internal TSS, followed by a strong TTS, contributes to the complexity of its expression (see text). Long reads corresponding to the different mRNA isoforms (*pelC*, *pelZ*, or *pelCZ*) are observed.

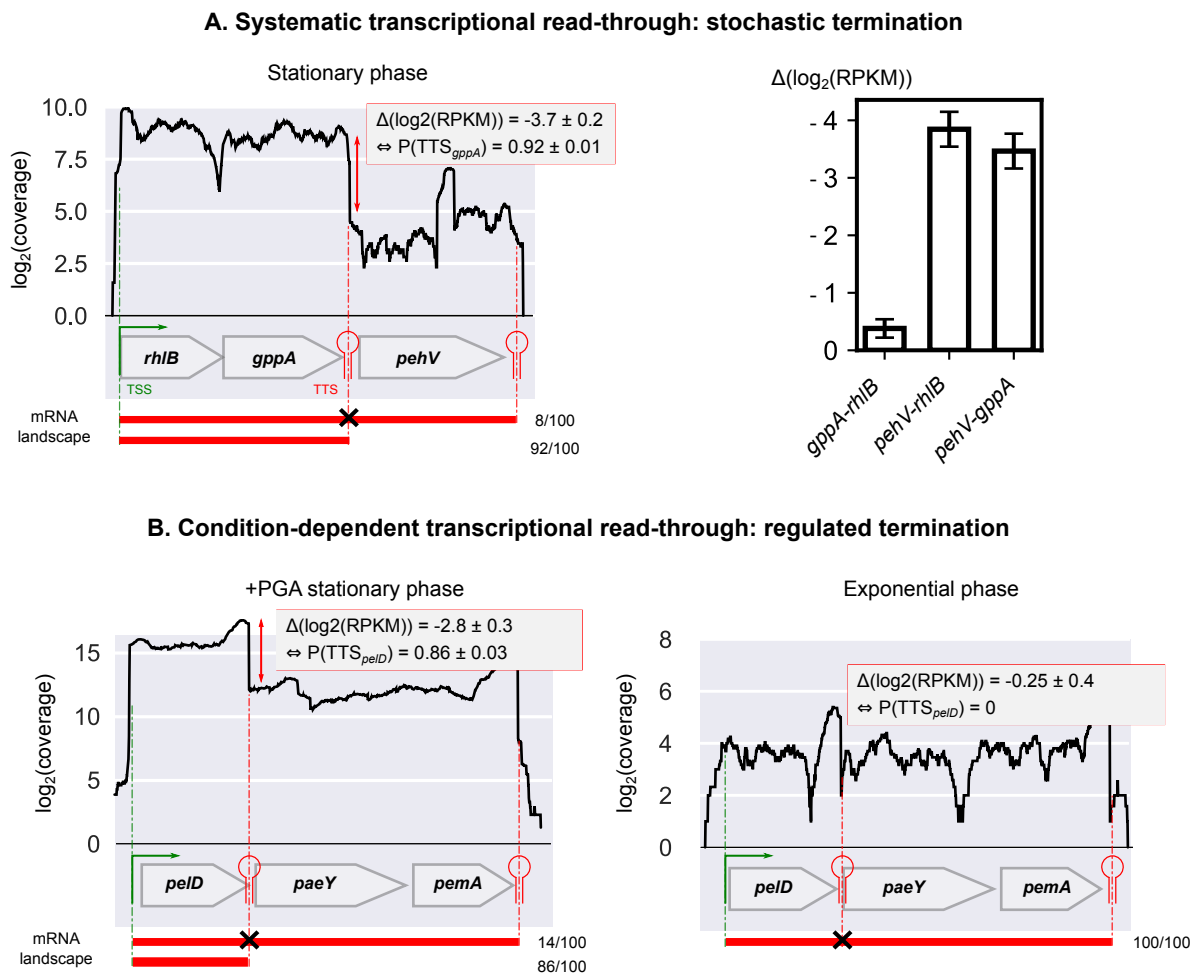
Mapping of *Dickeya dadantii* transcriptional landscape

**Transcriptional read-through, the root of transcription extension ?** We showed that transcription units comprise predicted operons, yet are generally longer. This extension of transcription might, in part, result from the ability of RNA Polymerase to stochastically override an imperfect terminator by a mechanism referred to as transcriptional read-through (3, 8). The latter has long been identified in specific operons (57, 58, 59) and was shown more recently to be widespread in bacterial genomes (2, 3, 8), where it may in fact play a basal coordination and regulation role (5). A condition-independent rate of stochastic termination might result in the co-expression of the genes located before and after the TTS (as in a classical operon), but with a reduced transcriptional level of the latter, a mechanism possibly relevant to functionally related genes that must be expressed at different strengths while keeping a constant ratio (59). The termination efficiency can also be subject to regulation, depending on environmental conditions and metabolic needs, resulting in a variable degree of read-through and thus of relative expression levels (57, 58). Such conditional read-through can involve Rho and other proteins assisting termination (60, 61, 62, 63) as well as other conditional premature termination mechanisms such as attenuation (64, 65), T-box conditional termination (66, 67) and riboswitches (68, 69).

An example of condition-independent read-through occurs at the *rhIB-gppA-pehV* TU (Fig. 4A and Supplementary Fig. S1C). The *rhIB* gene encodes a component of the RNA degradosome (70, 71) whereas *gppA* encodes guanosine 5'-triphosphate 3'-diphosphate (pppGpp) pyrophosphatase involved in bacterial stringent response (72) and the *pehV* gene encodes a polygalacturonase involved in pectin degradation (73). These genes are functionally unrelated (except for a distant link to nutritional stress) yet appear co-transcribed, which is in fact quite frequent among operons (41, 74). This TU exhibits a variable expression level (by up to 50%) across the sampled conditions, but the internal (relative) expression pattern is condition-independent: *rhIB* and *gppA* are expressed at a similar level, whereas *pehV* is systematically less transcribed (Fig. 4A and Supplementary Fig. S1C). This observation is correlated with the presence of an intrinsic internal TTS downstream of *gppA*. By computing the expression ratio of *pehV* compared to *rhIB/gppA*, we inferred the associated termination probability (or terminator strength) and found a constant value  $P(TTS_{gppA}) = 92 \pm 1\%$  (95% confidence interval) characteristic of a non-conditional transcriptional read-through. Thus, the three genes are co-transcribed from a single promoter of condition-dependent activity, with a reduced transcriptional level of *pehV* exhibiting a constant ratio (8%) compared to the other genes. The biological relevance of this mechanism remains to be clarified. In *E. coli*, *rhIB* and *gppA* were also recently shown to be co-transcribed (3, 54). Another example of condition-independent read-through occurs at the *gcvTHP* TU involved in glycine cleavage (75) (Fig. 6). We detected an internal TTS downstream of *gcvH* in accordance with studies in *E. coli* (3, 54) and inferred its termination probability  $P(TTS_{gcvH}) = 71 \pm 22\%$  (95% confidence interval), based on the expression ratio of *gcvP* compared to *gcvT* and *gcvH* across RNA-seq conditions. It is unclear whether this variability is due to RNA-seq signal variations or a weak regulation of the termination rate. The GcvT, H, and P proteins are part of the glycine cleavage system with GcvL (76), and GcvP activity might be required at lower concentration in the investigated conditions.

By definition, all identified internal TTSs (549) experience transcriptional read-through. As a rough estimate, condition-independent read-through was detected for 77 (14%) of internal TTSs, based on the constant expression ratio of the genes located downstream vs upstream across RNA-seq conditions (Fig. 4, Materials and Methods). The remaining internal TTSs rather experience condition-dependent read-through; how-

Forquet et al.



**FIG 4** Quantification of transcriptional read-through. (A) Non-conditional read-through: example of the *rhIB-gppA-pehV* TU (left panel). The first two genes are homogeneously transcribed among conditions, resulting in an expression variation  $\Delta(\log_2(\text{RPKM}))$  close to 0 (right panel, 95% confidence intervals are shown), while the intrinsic TTS downstream of *gppA* is stochastically overstepped in  $8 \pm 1\%$  of transcripts ( $P(\text{TTS}_{gppA}) = 0.92 \pm 0.01$ ), resulting in two different isoforms (red). (B) Condition-dependent read-through: example of the *pelD-paeY-pemA* TU. A TTS is identified downstream of *pelD* in agreement with previous studies (75). Its termination probability is regulated and depends on growth phase and presence of PGA ( $0.86 \pm 0.03$  vs 0), besides a global up- or down-regulation of the whole TU. All mRNA isoforms are observed in Nanopore native RNA-seq data (Supplementary Fig. S2A and S2B).



Mapping of *Dickeya dadantii* transcriptional landscape

345 ever, the systematic estimation of stochastic termination rates at internal TTSs is del-  
346 icate based on our data only, due to the limited number of RNA-seq conditions and  
347 the presence of nearby TSSs that contribute to the heterogeneous expression levels  
348 along the TU, as illustrated by *pelCZ* (Fig. 3D).

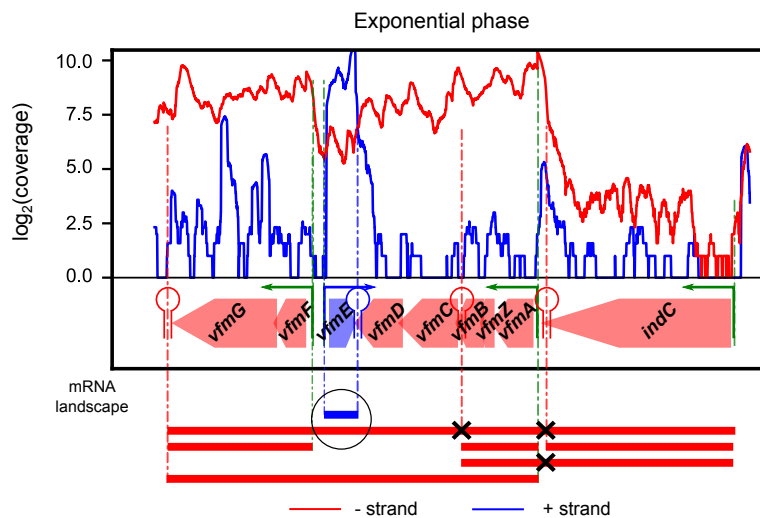
349 An example of condition-dependent read-through occurs at the *pelD-paeY-pemA*  
350 TU (Fig. 4B and Supplementary Fig. S2B), which is identified by our approach but was  
351 also characterised by Northern blotting (77). It encodes three genes involved in pectin  
352 degradation. In the initial step of pectinolysis occurring in plants, *paeY* (acetyltransferase)  
353 and *pemA* (methyltransferase) remove acetyl and methyl groups from pectin, which can  
354 then be efficiently degraded by the pectate lyase *pelD* (17). The *pelD* gene is essen-  
355 tially transcribed as a monocistronic RNA, although its terminator (predicted as intrinsic)  
356 can be overstepped to generate a polycistronic transcript comprising the three  
357 genes (77). In exponential phase, the three genes are homogeneously (but weakly)  
358 transcribed as a unique polycistronic RNA, suggesting that the internal TTS is not ef-  
359 ficient ( $P(TTS_{pelD}) = 0\%$ ). In stationary phase in presence of PGA, the whole TU is  
360 up-regulated, and the internal TTS becomes more efficient ( $P(TTS_{pelD}) = 86 \pm 3\%$ , 95%  
361 confidence interval), resulting in the extensive synthesis of the *pelD* monocistronic RNA  
362 and a lower expression level of the two downstream genes. The regulation events oc-  
363 ccurring at this TTS remain to be characterised, but may adjust the relative expression  
364 levels of the genes following metabolic needs, since *PelD* has a predominant role in  
365 pectin degradation and virulence (78, 79) and must likely be required at much higher  
366 concentrations than the two other enzymes. In addition, the fact that *pemA* is differ-  
367 entially expressed depending on the degree of pectin methylation (80) highlights the  
368 relevance of adjusting the relative expression levels of the three genes depending on  
369 plant cell-wall composition.

370 Another example occurs at the *cytABCD* TU (Supplementary Fig. S2C and S3A). In  
371 addition to plants, *D. dadantii* is able to infect insects (81), during which this TU ex-  
372 presses four insecticidal toxins and was previously shown to produce a polycistronic  
373 mRNA comprising the four genes, besides the possible existence of alternative iso-  
374 forms (82). The sequencing coverage together with the putative internal intrinsic TTS  
375 detected after *cytA* are clearly indicative of a condition-dependent read-through, with  
376 termination occurring less efficiently at *cytA* in stationary phase in presence of PGA  
377 compared to exponential phase. This variation in termination efficiency at *cytA* asso-  
378 ciated to an environmental change may again allow tuning the relative amounts of  
379 the corresponding toxins, especially if a precise and condition-dependent balance be-  
380 tween them is required for optimal activity during the insect infection process (82).  
381 Interestingly, this cluster of four genes was acquired by horizontal transfer. Since  
382 transcriptional read-through partly relies on basal RNA Polymerase / TTS interactions,  
383 it might be conserved during horizontal transfer among bacterial species without re-  
384 quiring an independent acquisition of regulatory signals and their integration in the  
385 transcriptional regulatory network of the recipient cell.

386 **Detection of putative excludons and noncontiguous transcriptions units.** All  
387 previous examples involved genes located on the same DNA strand; yet recent studies  
388 also describe interactions between overlapping antisense coding transcripts, involved  
389 in a mutual regulation. In particular, noncontiguous operons refer to operons that con-  
390 tain a gene or group of genes that is transcribed in the opposite direction (83). 83 TUs  
391 with such features were found in the *D. dadantii* genome (provided in Supplementary  
392 Tab. S4A). Among them, an example is the *indCvfmAZBCDFG* TU encoding a compo-  
393 nent of the *vfm* quorum sensing system required for the production of plant cell wall-  
394 degrading enzymes (Fig. 5 and Supplementary Fig. S2D) (84). The *vfmE* gene, located



Forquet et al.



**FIG 5** Existence of a potential noncontiguous transcription unit in the *vfm* locus. The *vfmE* gene is transcribed in the opposite direction of the *indCvfmAZBCDFG* TU, generating two overlapping mRNAs (as shown in red/blue for the  $-/+$  strand) that might be involved in a mutual regulation (see text). All mRNA isoforms are observed in Nanopore native RNA-seq data (Supplementary Fig. S2D), including a long native RNA read on the negative strand between *vfmD* and *vfmF*.

395 on the opposite strand and within this TU, is also part of this system and known to  
 396 encode a transcriptional activator of the *vfm* locus (of the AraC family). Since all genes  
 397 of the TU are co-transcribed within a single mRNA, it is likely that these two overlap-  
 398 ping antisense transcripts could negatively regulate each other, e.g. by transcriptional  
 399 interference (RNA Polymerases collision) or RnasIII-mediated double-stranded RNA  
 400 processing (85). An expression increase of the *vfm* locus would then reduce the expres-  
 401 sion of *vfmE*, and in turn its own expression, forming a genome-embedded negative  
 402 feedback loop controlling the production of quorum sensing signal and plant cell-wall  
 403 degrading enzymes (86).

404 Finally, “excludons” refer to genomic regions in which convergent or divergent  
 405 genes display overlapping transcription (87). From the map of transcription start and  
 406 termination sites, we found 160 putative convergent excludons (overlapping 3' UTRs)  
 407 and 63 putative divergent excludons (overlapping 5' UTRs) (provided in Supplemen-  
 408 tary Tab. S4B). An example is the divergent excludon between *greB* and *ompRenvZ*  
 409 transcription units, encoding a transcript cleavage factor required for effective tran-  
 410 scription elongation (88) and a two-component signal transduction system involved  
 411 in osmotic stress response (89), respectively (Supplementary Fig. S3B). Both TUs com-  
 412 prise long 5'UTRs, forming a region of overlapping transcription that was previously  
 413 identified in *E. coli* (90) and might underpin a mutual post-transcriptional regulation.

414 ***In planta* co-expression validation of the transcription units.** While our tran-  
 415 scriptional map was inferred from *in vitro* cultures, where RNAs could be extracted  
 416 with optimal quality and reproducibility, we wished to test if the identified TUs could  
 417 play a role in conditions of plant infection. We analysed a set of expression data ob-  
 418 tained *in planta* by DNA microarrays, during the early stages of *Arabidopsis thaliana*  
 419 infection (dataset 5) (29), 6 hours post-inoculation, during the epiphytic colonisation  
 420 of leaf surface, and 24 hours post-inoculation during leaf invasion, just before the  
 421 onset of visible symptoms. Overall, among the 50% gene pairs most correlated *in*  
 422 *planta*, 80% belong to the same TUs, suggesting that co-transcription of these genes

Mapping of *Dickeya dadantii* transcriptional landscape

423 may indeed likely occur in these conditions (Supplementary Fig S4A). As an example,  
424 in *cytABCD*, the four genes are also highly correlated *in planta*, while this correlation  
425 immediately drops in surrounding isodirectional TUs (Supplementary Fig. S4B), as we  
426 expected. However, comparable correlations might also arise between other genes  
427 that are not transcribed together, but share the same transcriptional regulators, in  
428 particular those involved in virulence such as KdgR, PecT, PecS (91), thus accounting  
429 for the 20% strongly correlated gene pairs not located in the same TU. For example,  
430 in the *pelCZ* complex TU involved in pectinolysis, both genes are strongly correlated *in*  
431 *planta* (Supplementary Fig. S4C), as expected from the previous *in vitro* observations  
432 (especially with PGA, Supplementary Fig. S4C), but the adjacent *pelB* gene is also cor-  
433 related, whereas *crp* and *mrcA* are not. This is not surprising, since most *pel* are par-  
434 alogous genes with similar regulators and are strongly induced by pectin. The same  
435 pattern is observed for the *pelD-paeY-pemA* TU (Supplementary Fig. S4D), with respect  
436 to the *pelE* and *pelA* genes located upstream on the same strand. Because of the lim-  
437 ited spatial resolution of microarrays and the weak number of investigated conditions,  
438 it is not possible to systematically distinguish the effects of these two mechanisms at  
439 the genomic scale from these data, but a survey of representative TUs confirmed that  
440 they usually coincide with correlated blocks of genes (as observed with *cytABCD*), even  
441 when the latter do not belong to the same functional pathways.

442 As an example, the complex TU *sufABCDE-ltdC* is composed of two functionally  
443 unrelated operons (Supplementary Fig. S4E). *sufABCDE* encodes components of the  
444 iron-sulfur cluster assembly machinery (92), which is required to synthesise and repair  
445 damaged iron-sulfur clusters under conditions of oxidative stress or iron limitation,  
446 and is therefore critical for *D. dadantii* virulence (91). In contrast, *ltdC* (previously *ycfS*),  
447 encodes a L,D-transpeptidase crucial for bacterial envelope assembly, by catalysing  
448 the attachment of the major outer-membrane protein Lpp to peptidoglycan (93). Ac-  
449 cording to our findings above, *sufABCDE* and *ltdC* can be transcribed together, with  
450 an internal TTS and TSS located between them. *In planta*, the seven genes are indeed  
451 strongly co-expressed, with a slight decrease for *ltdC*, in full agreement with the iden-  
452 tified transcriptional map (Supplementary Fig S4E). It is conceivable that these genes  
453 are required under a common set of conditions encountered during plant infection,  
454 which was favoured by their inclusion in the same transcript, while the presence of al-  
455 ternative TSS and TTS might still allow separate expression when required. Indeed, the  
456 *sufABCDE* operon is controlled by three transcriptional regulators, Fur, OxyR and IscR,  
457 which respectively sense iron limitation, oxidative stress and intracellular iron-sulfur  
458 cluster status (94). Each of them contributes to the activation of the *suf* promoter by  
459 oxidative stress occurring during plant penetration and colonisation (25): the repres-  
460 sor Fur is inactivated by reactive oxygen species (ROS); the activator OxyR becomes  
461 active through the oxidation of two cysteine residues and the formation of a disulfide  
462 bond; IscR becomes an activator of *suf* promoter after destruction of its iron-sulfur  
463 cluster by ROS (94). On the other hand, the activity of L,D-transpeptidases involves a  
464 catalytic cysteine residue that must be reduced (95), which is challenging under oxida-  
465 tive stress. The expression of *ltdC* from the *suf* promoter, which is strongly activated  
466 in the latter condition, is therefore biologically meaningful. Interestingly, in *E. coli*, the  
467 *suf* operon is also located upstream of a gene encoding a L,D-transpeptidase (*ltdA*),  
468 the two operons being also transcribed both together and separately (54).

469 **Concluding statement** In this study, we combined five transcriptomic datasets  
470 yielding complementary information and designed to provide a catalogue of genes' re-  
471 sponses and RNA landscapes to various growth and stress conditions, including one  
472 of the first applications of Nanopore native RNA-seq to prokaryotic transcriptomes.

Forquet et al.

473 Their integration through a computational method developed for this study allowed  
474 us to precisely determine and annotate the transcriptomic map of *D. dadantii*, the first  
475 of its kind in the *Dickeya* genus. The analysis of *in planta* DNA microarray data suggests  
476 that the identified TUs are also co-expressed during the early stages of plant infection,  
477 although a more refined *in planta* analysis would require higher-resolution transcrip-  
478 tomic data. Beyond its practical aspect as a community resource to help the scientific  
479 community to unravel gene regulation, including the virulence programme of this and  
480 related species, the obtained transcriptional map clearly indicates, after others, that  
481 the canonical operon model is insufficient to account for the complexity of bacterial  
482 transcription. The ability of the cell to differentially express genes of the same operon  
483 depending on metabolic needs and environmental conditions was first described with  
484 suboperonic regulation years ago. Later, with the emergence of next-generation se-  
485 quencing, transcriptomic analyses confirmed at the genomic scale that most operons  
486 were able to generate alternative transcripts of variable gene composition. Transcrip-  
487 tional read-through at terminators is another mechanism that might play a basal coordi-  
488 nation and regulation role, and explain the extent of transcription beyond the scale  
489 of operons. Recent findings include noncontiguous operons and excludons, where  
490 the expression of an operon transcript can be mutually regulated with that of a gene  
491 located on the opposite strand at the same locus. For such features, the putative cat-  
492 alogue provided here may be used as a starting point for further investigation, and in  
493 particular, might be combined with the *D. dadantii* non-coding RNA landscape (96) for  
494 a comprehensive analysis of transcriptional regulation in this bacterium. Altogether,  
495 our findings provide insights into the mechanisms of basal coordination of transcrip-  
496 tion and might contribute to the revision of the canonical view of operon structure  
497 and transcription organisation.

#### 498 MATERIALS AND METHODS

499 **Bacterial strain, genome annotation and genome-wide predictions of oper-**  
500 **ons.** The genome sequence and annotation files from *Dickeya dadantii* strain 3937  
501 were obtained from NCBI under accession NC\_014500.1 (97). This work focused on  
502 coding genes only (CDS, representing 4211 genes over 4411 in total). *D. dadantii* oper-  
503 ons were predicted using Rockhopper, a recent computational tool for operon predic-  
504 tion based on RNA-seq expression data as well as genomic and functional informa-  
505 tion (40), by providing dataset 1 as input.

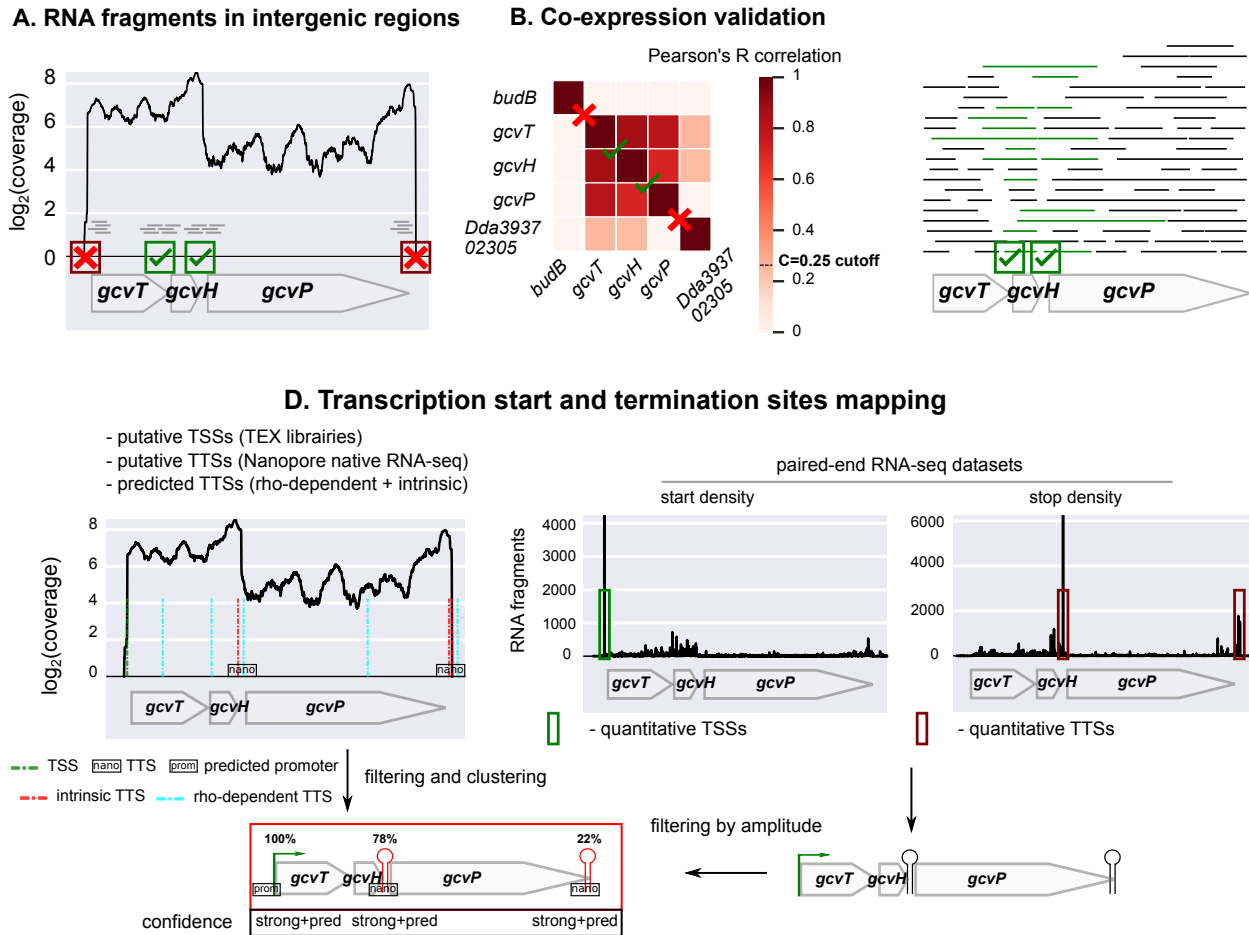
506 **RNA-sequencing data (dataset 1), definition of putative transcription units**  
507 **based on intergenic signals, and identification of unannotated genes.** Strand-specific,  
508 paired-end RNA-seq processed data used in this study are described in (98). Transcrip-  
509 tomes were obtained in 6 conditions (with two biological replicates each) including var-  
510 ious growth (M63 medium supplemented with sucrose, in exponential or stationary  
511 phase, in presence or absence of PGA) and DNA supercoiling conditions (novobiocin  
512 shock). For each sample, RNA fragments were inferred from paired-end reads infor-  
513 mation, and genome-wide coverage was computed from resulting RNA fragments co-  
514 ordinates using a Python home-made script.

515 To define putative transcription units, separately for each strand, adjacent genes  
516 were fused in the same putative TU as long as the coverage was greater than 0 at each  
517 position of their intergenic region (independently of its size) for at least half of the  
518 samples (Fig. 6A).

519 Unannotated genes were defined as DNA regions outside of known coding se-  
520 quences, longer than the first centile (1%) of *D. dadantii* gene lengths (192 bp), with an  
521 average coverage significantly different from 0 (with 99% confidence, *i.e.*, > 9 at each

Mapping of *Dickeya dadantii* transcriptional landscape

522 position) in all samples, and with a coding sequence predicted by Prodigal (99), result-  
 523 ing in 50 unannotated genes. A search for homolog proteins was performed using  
 524 PSI-BLAST based on the non-redundant protein database (Supplementary Tab. S1D).



**FIG 6** Algorithm for the characterisation of *D. dadantii* transcription units. (A) Definition of putative TUs based on RNA-seq coverage (dataset 1) in intergenic regions, with isodirectional genes being split when the coverage drops to zero (here before *gcvT* and after *gcvP*). (B) Validation based on correlation of expression across 32 conditions (dataset 2). The genes of identified putative TUs are correlated, in contrast to surrounding isodirectional genes (*budB* and *Dda3937\_02305*). (C) Validation based on Nanopore native RNA-seq, based on the presence of overlapping RNA reads between adjacent gene pairs, yielding a direct evidence for co-transcription. (D) TSS and TTS mapping based on dRNA-seq (TEX libraries, dataset 4), Nanopore native RNA-seq (dataset 3), TTS predictions, promoter predictions, and paired-end RNA-seq data (dataset 1). First, putative TSSs and TTSs of high resolution but qualitative strength were defined from an analysis of TEX libraries and Nanopore native RNA-seq, respectively, and rho-dependent/intrinsic terminations were predicted. Second, a list of TSSs and TTSs of quantitative strength but poorer resolution was defined from the enrichment of RNA-seq paired-end fragment starts (start density) and stops (stop density) upstream of gene starts and downstream of gene stops, respectively. Third, only TSSs and TTSs with sufficient strength were retained, and compared to the closest TEX TSS / Nanopore TTS / predicted hairpin loop, in order to define their exact position and level of confidence. Finally, promoters were predicted for the retained TSSs. As a result of the analysis, this TU included the *gcvTHP* genes, the first two genes being expressed both as part of the entire transcript and as an independent transcript generated from a strong internal TTS (76% of total magnitude), explaining the lower correlation between *gcvP* and the remaining genes.

Forquet et al.

**In vitro DNA microarray data (dataset 2) and co-expression validation of the putative transcription units using hierarchical clustering.** Microarray processed data used in this study are described elsewhere (27). They comprise 32 *in vitro* conditions (with two biological replicates each) including various growth and stress conditions encountered by *D. dadantii* during plant infection: cells were harvested in M63 (minimal) medium supplemented with sucrose, in exponential or stationary phase, in presence or in absence of PGA or leaf extract, and exposed or not to environmental perturbations (acidic, osmotic and oxidative stress). Pearson's correlation coefficients were computed among all gene pairs over all conditions on the logarithm of the normalised expression level (derived from probes intensity). For each putative TU, adjacent genes were grouped into clusters based on this correlation, using a hierarchical clustering framework constrained to group adjacent genes only, with a custom Python script. At each iteration of the algorithm, the median of cross-correlations among all clusters (or genes) was computed, and the adjacent clusters with maximal median were fused. The hierarchical clustering ends when a cutoff value  $C$  for the correlation is reached (Fig. 6B). If the agglomeration of all genes of the TU is achieved without reaching  $C$ , the TU is validated. Otherwise, the final clusters are considered as separate TUs. A high  $C$  value results in short highly correlated TUs, whereas a low  $C$  value yields longer moderately correlated TUs (Supplementary Tab. S5). We defined the value  $C = 0.25$  such that 20% of operon predictions were discarded (Supplementary Fig. S5), since it is the number of false predictions (*i.e.* specificity) evaluated for Rockhopper in *E. coli*, a *D. dadantii* enterobacterium relative. Varying the precise value of  $C$  did not qualitatively change the main results (Supplementary Tab. S5). The identified TUs exhibit a similar length distribution as those reported in *E. coli* (9, 3).

**Nanopore native RNA sequencing (dataset 3), validation of the mRNA landscape, and genome-wide identification of putative transcription termination sites.**

*D. dadantii* cultures were grown in M63 medium supplemented with 0.2% glucose and 0.2% PGA, until the early exponential phase ( $A_{600nm} = 0.2$ , condition 1), or the early stationary phase ( $A_{600nm} = 1.8$ , condition 2). RNAs were extracted using a frozen acid-phenol method, as previously described (100), and treated successively with Roche and Biolabs DNases. Two samples were prepared: 50  $\mu\text{g}$  of RNAs from each condition were pulled into one sample (sample 1), whereas the other one contained 100  $\mu\text{g}$  of RNAs from condition 2 (sample 2). Both samples were then supplied to Vertis Biotechnologie AG for Nanopore native RNA-seq: total RNA preparations were first examined by capillary electrophoresis, and ribosomal RNA molecules were depleted for sample 1 only using an in-house developed protocol (recovery rate = 84%). RNA 3'ends were then poly(A)-tailed using poly(A) polymerase, and the Direct RNA sequencing kit (SQK-RNA002) was used to prepare the library for 1D sequencing on the Oxford Nanopore sequencing device. The direct RNA libraries were sequenced on a MinION device (MIN-101B) using standard settings. Basecalling of the fast5 files was performed using Guppy (version 3.6.1) with the following settings: `-flowcell FLO-MIN106 -kit SQK-RNA002 -cpu_threads_per_caller 12 -compress_fastq -reverse_sequence true -trim_strategy rna`. Reads smaller than 50 nt were removed. 466 393 and 556 850 reads were generated from sample 1 and 2, respectively. Raw read sequencing data are available in the EBI Gene Expression (ArrayExpress) database under accession E-MTAB-10482. Quality control was performed on both datasets using Nanopack (101). Long-reads from the fastq files were mapped to *Dickeya dadantii* strain 3937 genome (NCBI accession number: NC\_014500.1) (97) using minimap2 (release minimap2-2.17 (r941)) (102). Output alignments in PAF and SAM format were generated with the recommended options for noisy Nanopore native RNA-seq, adapted to bacteria (no splic-



Mapping of *Dickeya dadantii* transcriptional landscape

ing) (-ax map-ont -k14). Secondary alignments were not reported for sample 2 due to multiple secondary alignments in ribosomal RNAs regions (-secondary=no). In total, 382 290 and 392 743 alignments were generated (77 and 67% mappability) from sample 1 and 2, respectively. Alignment files were further sorted, indexed and analysed with SAMtools. Alignments from both samples were merged into one PAF file, and the latter was used for further analyses.

For each TU previously defined with datasets 1 and 2 (Fig. 6A and 6B), the presence of long overlapping native RNA reads was investigated using a Python home-made script for adjacent gene pairs belonging to the same TU (Fig. 6C). If at least one RNA read overlapped the two adjacent genes, their co-transcription was validated (quoted "validated" in Supplementary Tab. S1A). If the signal was too weak for the investigated genes (read counts <9, not significantly different from 0 with 99% confidence), no conclusion could be drawn (quoted "weak signal" in Supplementary Tab. S1A). Otherwise, if no overlapping RNA was found, it was not validated (quoted "invalidated" in Supplementary Tab. S1A), which might also be due to the low number of conditions tested.

For the determination of TTSSs, for each position of the genome, we computed the total number of RNA fragments ending at this particular position, using a Python home-made script. From this stop density, we defined putative TTSSs as positions downstream of gene stop codons (up to 100 bp, based on 3'UTR lengths in *E. coli*) enriched for RNA fragments stops, respectively. In each of these regions, we started from site  $i$  with highest stop signal  $k_i$  on 5-bp centred windows (due to the low sequencing depth). For the position  $i$  to be considered as a putative TTS, we imposed  $k_i$  to be significantly different from 0 (with 95% confidence, > 6). TTSSs obtained with this approach are provided in Supplementary Tab. S2D.

**Differential RNA-sequencing experiments and genome-wide identification of putative transcription start sites (dataset 4).** RNAs from dataset 2 (27) (*in vitro* DNA microarray data) were pooled into four samples S1 to S4, resulting in a combination of stress ( $pH$ ,  $NaCl$ ,  $H_2O_2$ ) and growth conditions: exponential phase with (S1) or without (S2) stress, transition to stationary phase with (S3) or without (S4) stress. Those samples were then supplied to Vertis Biotechnologie AG for TEX treatment and Illumina sequencing. Briefly, ribosomal RNA molecules were depleted from the total RNA samples using the Ribo-Zero rRNA Removal Kit for bacteria (Epicentre), and small RNAs (< 200 nt) were discarded using the RNeasy MinElute Cleanup Kit (Qiagen). For the generation of TSS cDNA libraries, the samples were first fragmented using RNase III, poly(A)-tailed using poly(A) polymerase, split into two halves, with one half being treated with Terminator exonuclease (+TEX, Epicentre), while the other one was left untreated (-TEX). 5'PPP structures were then converted into 5'P ends using RNA 5' Polyphosphatase (5'PP, Epicentre), to which RNA adapters were ligated. First-strand cDNAs were synthesised using an oligo(dT)-adapter primer and the M-MLV reverse transcriptase, PCR-amplified using a high fidelity DNA polymerase, purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics), and sequenced on an Illumina NextSeq 500 system (60 bp read length, single-end, strand-specific protocol). Sequencing reads were trimmed to remove poly(A) tails and adapters. The fastq sequencing files are available in the EBI Gene Expression (ArrayExpress) database under accession E-MTAB-9075. Putative TSS positions were then determined based on the enrichment of sequencing reads in TEX-treated samples (+TEX) compared to non-treated ones (-TEX) using TSSer, an automated annotation programme from dRNA-seq data with default parameters: TSS positions within 5 bases on the same strand were clustered together and the position with the highest amount of read increase in the +TEX library was retained. TSSs obtained with such approach are provided in Supplemen-

Forquet et al.

625 tary Tab. S2A.

626 **In planta DNA microarray data (dataset 5) and co-expression validation of the**  
627 **transcription units inferred from in vitro conditions.** Microarray processed data  
628 used in this study are described in (29). They comprise two conditions: bacteria were  
629 collected 6 hours post-inoculation of the model plant *Arabidopsis thaliana* by wild-type  
630 *D. dadantii* during the epiphytic colonisation of the leaf surfaces (5 replicates), and 24  
631 hours post-inoculation during the leaf invasion (4 replicates). Pearson's correlation co-  
632 efficient was computed among all gene pairs over the two conditions on the logarithm  
633 of the normalised expression level (derived from probes intensity) (Supplementary  
634 Fig. S4).

635 **Genome-wide detection of transcription start and termination sites from**  
636 **RNA-seq data, mapping to the transcription units.** We computed the densities of  
637 RNA fragments starting and ending at each position of the genome, across all RNA-seq  
638 samples (dataset 1) (Fig. 6D). In order to retain only TSSs and TTSs relevant to protein-  
639 coding genes, we focused on regions located upstream of gene start codons (up to 250  
640 bp, based on 5'UTR lengths in *E. coli*), and downstream of gene stop codons (up to 100  
641 bp, based on 3'UTR lengths in *E. coli*), respectively. In each of these regions, putative  
642 TSSs/TTSs were defined as sites  $i$  with highest start/stop signal  $k_i$ . To differentiate a  
643 TSS/TTS at position  $i$  from the noise, we imposed two successive conditions: (i)  $k_i$  is  
644 significantly different from 0 (with 99% confidence,  $k_i > 9$ ); (ii)  $k_i$  is greater or equal  
645 than a density cutoff value  $D$ . The latter was set as ten times the median of the den-  
646 sity values of the region investigated for TSSs, and five times for TTSs, showing that  
647 the recorded transcripts indeed start/stop at that precise position, rather than along a  
648 poorly defined starting/stopping region. In that case, the position  $i$  was considered as  
649 a putative TSS/TTS, of strength  $k_i$ . Setting a low density cutoff  $D$  would tend to include  
650 false positives resulting from RNA-seq signal variations (noise), whereas a high cutoff  
651 would exclude weakly expressed TSSs/TTSs. We selected the value of  $D$  (i) such that  
652 TSSs and TTSs were detected for known operons and experimentally characterised  
653 TUs (described in the manuscript) and (ii) by visually curating the density graphs and  
654 excluding many positions obviously associated to RNA-Seq signal variations.

655 TSSs/TTSs positions were then compared among datasets to evaluate their con-  
656 fidence level. For each TSS identified with this approach, if a putative TSS obtained  
657 from dataset 4 (TEX libraries) was close enough ( $\pm 20bp$ ), its position was retained (as-  
658 suming a higher precision and resolution). In addition, a scan for promoter motifs  
659 was conducted with bTSSfinder (52). For TTSs, the same method was applied using  
660 the position of the closest predicted hairpin loop ( $\pm 50$  bp), or TTS positions obtained  
661 from Nanopore native RNA-seq data (dataset 3). TSSs and TTSs were then assigned to  
662 the TUs, and only internal TSSs and TTSs with 15% relative amplitude (*i.e.*  $\frac{k_i}{\sum(k_i)}$ ) were  
663 retained, resulting in a total of 2595 TSSs and 1699 TTSs over all TUs. Setting a low rel-  
664 ative amplitude cutoff would tend to retain all TSSs / TTSs, including many very weak  
665 ones mostly due to noise. We selected the relative amplitude cutoff value (i) based on  
666 a collection of known operons and TUs (shown in the manuscript), and (ii) such that  
667 the total number of TSSs and TTSs identified was consistent with those reported re-  
668 cently in *E. coli* (3, 54). If no TSS/TTS was found from dataset 1, we indicated the closest  
669 putative one from dataset 3/4 with a lower confidence level. The lists are provided in  
670 Supplementary Tab. S1A to S1C.

671 **Detection of transcriptional readthrough at internal TTSs** For each internal  
672 TTS, the expression ratio  $\Delta(\log_2(RPKM))$  of the gene located downstream compared  
673 to the gene located upstream was computed across RNA-seq conditions. We imposed  
674 two successive conditions to consider the transcriptional read-through at this TTS as



Mapping of *Dickeya dadantii* transcriptional landscape

675 condition-independent: (1)  $\Delta(\log_2(RPKM)) \leq -0.5$  for at least 8 samples over 12 cor-  
676 responding at least to a termination probability  $P(TTS) = 71\%$ , (2) standard error of  
677 the mean  $\sigma P(TTS) \leq 12.5\%$  corresponding to a relatively constant mean expression  
678 ratio and subsequent termination probability.

#### 679 **Availability of data and materials.**

- 680 • *Dickeya dadantii* strain 3937 genome sequence and annotation files: NCBI ac-  
681 cession number NC\_014500.1 (97).
- 682 • RNA-seq data (dataset 1): EBI Gene Expression (ArrayExpress) accession num-  
683 ber E-MTAB-7650 (98).
- 684 • *In vitro* microarray data (dataset 2): EBI Gene Expression (ArrayExpress) acces-  
685 sion number E-MTAB-541 (27).
- 686 • Nanopore native RNA sequencing (dataset 3): EBI Gene Expression (ArrayEx-  
687 press) accession number E-MTAB-10482. Note: not publicly available until the  
688 manuscript is accepted. Please use login details Username =  
689 Reviewer\_E-MTAB-10482, Password = gdrof3hg.
- 690 • Differential RNA-seq data (dataset 4): EBI Gene Expression (ArrayExpress) acces-  
691 sion number E-MTAB-9075. Note: not publicly available until the manuscript is  
692 accepted. Please use login details Username = Reviewer\_E-MTAB-9075, Pass-  
693 word = jenuxmon.
- 694 • *In planta* microarray data (dataset 5): NCBI Gene Expression Omnibus (GEO)  
695 accession number GSE94713 (29).

#### 696 **SUPPLEMENTAL MATERIAL**

697 **SUPPLEMENTARY FIGURE S1.** (A) Co-expression validation of *znuCBA* TU with *in vitro*  
698 DNA microarray data (dataset 2): the three genes exhibit strong internal cross-correlations  
699 clearly indicative of an operon. (B) Same for *sapABCDFfabI* TU: the six genes are co-  
700 expressed, with a reduced correlation of *fabI* due to the presence of a strong inter-  
701 nal TSS (Fig. 3B). (C) Same for *rhlB-gppA-pehV* TU: the three genes are co-expressed,  
702 with a reduced transcriptional level of *pehV* (Fig. 4A) and a reduced correlation due to  
703 condition-independent read-through at the *gppA* intrinsic terminator. (D) Effect of PGA  
704 on *pelCZ* TU: *pelC* and *pelZ* expression profiles are similar in absence (left) or presence  
705 (right) of PGA in stationary phase, in spite of a drastically different global expression  
706 level.

707 **SUPPLEMENTARY FIGURE S2.** Co-transcription and mRNA landscape validation  
708 with Nanopore native RNA-seq for (A) *rhlB-gppA-pehV* TU with condition-independent  
709 read-through at the intrinsic TTS downstream of *gppA*; (B) *pelD-paeY-pemA* TU with  
710 condition-dependent read-through downstream of *pelD* internal TTS; (C) *cytABCD* TU  
711 with condition-dependent read-through downstream of *cytA* internal TTS. Those inter-  
712 nal TTSs are occasionally overstepped, resulting in different transcripts isoforms  
713 (as shown in red) which are all detected as long native RNA reads (black). (D) *indCvf-*  
714 *mAZBCDFG* noncontiguous TU (black Nanopore reads on the negative strand), with  
715 *vfmE* being transcribed on the opposite strand (blue Nanopore reads on the positive  
716 strand), resulting in overlapping antisense transcripts.

717 **SUPPLEMENTARY FIGURE S3.** (A) Quantification of condition-dependent tran-  
718 scriptional read-through: example of the *cytABCD* TU. A putative Rho-independent TTS  
719 is identified downstream of *cytA* although not validated. Its probability of termination  
720 (inferred from the expression variation  $\Delta(\log_2(RPKM))$  of *cytA* compared to the other  
721 genes) is regulated and depends both on the growth phase and the presence of PGA  
722 ( $P(TTS_{cytA}) = 0.78 \pm 0.03$  vs  $0.51 \pm 0.03$ ) besides a global up-regulation of the whole TU.  
723 (B) The *greB* and *ompRenvZ* transcription units form a potential divergent excludon:

Forquet et al.

724 long 5'UTRs overlapping transcripts are generated by *ompR* and *greB* divergent genes  
725 and might form a dsRNA that could prevent each other transcription. In *E. coli*, *ompR*  
726 and *envZ* are part of the same operon (red), and *greB* is transcribed alone (blue). Such  
727 genomic region forming a dsRNA was also identified in *E. coli* (90).

728 **SUPPLEMENTARY FIGURE S4.** *In planta* DNA microarray data, 6 hours post-inoculation  
729 (hpi) of the plant *Arabidopsis thaliana* (epiphytic colonisation of the leaf surfaces, 5  
730 replicates), and 24 hpi (leaf invasion, 4 replicates). (A) Distribution of co-expression  
731 correlation coefficients among (blue) all genes and (red) genes belonging to the same  
732 TU. Among the 50% most correlated genes *in planta*, 80% belong to the same TUs, with  
733 the example TUs from the manuscript (*smtA-mukFEB*, *znuCBA*, *sapABCDF-fabl*, *glgBXCAP*,  
734 *pelCZ*, *rhlB-gppA-pehV*, *pelD-paeY-pemA*, *cytABCD*) displaying a median correlation of 0.9  
735 *in planta*. (B) Pearson's co-expression correlation coefficients of *cytABCD* TU with sur-  
736 rounding isodirectional (on the same strand) TUs. (C) Same for *pelCZ* TU. (D) Same  
737 for *pelD-paeY-pemA* TU. (E) Identification of *sufABCDSE-ltdC* complex TU, composed of  
738 operons of apparently unrelated functions, exhibiting a strong internal TSS (51% total  
739 magnitude) upstream of *ltdC* (previously *ycfS*) and a strong internal TTS (52% total  
740 magnitude) downstream *sufE*, allowing separate transcriptions. The seven genes are  
741 highly correlated *in planta*.

742 **SUPPLEMENTARY FIGURE S5.** Co-expression validation of transcription units for  
743 different correlation thresholds *C*. TUs obtained with high *C* values are more highly  
744 correlated but shorter. Putative TUs are obtained from step 1 of the analysis (inter-  
745 genic signal), without any requirement on the correlation of expression. With the cho-  
746 sen value ( $C = 0.25$ ), TUs group around three times more gene pairs than predicted  
747 operons by Rockhopper. The value of *C* was chosen such that 20% of operon predic-  
748 tions were discarded, since it is the number of false predictions of Rockhopper in *E.*  
749 *coli*, a *D. dadantii* enterobacterium relative.

750 **SUPPLEMENTARY TABLE S1.** (A) *Dickeya dadantii* transcription units defined by  
751 our approach. (B) TSSs across TUs. (C) TTSs across TUs. (D) Unannotated protein-  
752 coding genes.

753 **SUPPLEMENTARY TABLE S2.** (A) Putative TSSs identified by differential RNA-seq  
754 (TEX treatment) under a wide range of environmental conditions. (B) Genomic po-  
755 sition and secondary structure of putative TTSs: intrinsic terminators predicted by  
756 ARNold (Erpin and RNAmotif algorithms). (C) Genomic position of putative TTSs: Rho-  
757 dependent terminators predicted by RhoTermPredict. (D) Putative TTSs identified by  
758 Nanopore native RNA-seq.

759 **SUPPLEMENTARY TABLE S3.** TSS validation, based on all published TSSs to date  
760 and to our knowledge in *D. dadantii*.

761 **SUPPLEMENTARY TABLE S4.** (A) Catalogue of putative noncontiguous transcrip-  
762 tion units. (B) Catalogue of putative excludons.

763 **SUPPLEMENTARY TABLE S5.** Catalogue of transcription unit architecture. Puta-  
764 tive TUs are obtained from the first step of the approach (analysis of intergenic signal).  
765 Varying the precise value of the correlation threshold *C* for co-expression validation  
766 (step 2) does not change the results qualitatively. A larger *C* value results in shorter but  
767 more highly correlated TUs. Final TUs obtained with  $C = 0.25$  are longer than predicted  
768 operons and exhibit a similar length distribution as those reported in *E. coli* (9, 3).

## 769 ACKNOWLEDGEMENTS

770 We thank the whole CRP team for useful discussions as well as Ivan Junier.

Mapping of *Dickeya dadantii* transcriptional landscape

## 771 FUNDING

772 R.F was funded by a research allocation from the French Research Ministry. This work  
773 also benefited from INSA Lyon grants [BQR 2016 to S.M.]; IXXI; Agence Nationale de  
774 la Recherche [ANR-18-CE45-0006-01 to S.M.], Breakthrough Phytobiome IDEX LYON  
775 project, Université de Lyon Programme d'investissements d'Avenir [ANR16-IDEX-0005  
776 to S.R.]; Centre National de la Recherche Scientifique [to S.R., F.H. W.N. and S.M.]; Uni-  
777 versité Claude Bernard Lyon 1 [to S.R., F.H. W.N. and S.M.].

## 778 CONFLICT OF INTEREST STATEMENT

779 None declared.

## 780 REFERENCES

- Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3:318–356. doi:10.1016/s0022-2836(61)80072-7.
- Nicolas P, Mäder U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, Bidnenko E, Marchadier E, Hoebeke M, Aymerich S, Becher D, Bisicchia P, Botella E, Delumeau O, Doherty G, Denham EL, Fogg MJ, Fromion V, Goelzer A, Hansen A, Härtig E, Harwood CR, Homuth G, Jarmer H, Jules M, Klipp E, Le Chat L, Lecoine F, Lewis P, Liebermeister W, March A, Mars RAT, Nannapaneni P, Noone D, Pohl S, Rinn B, Rügheimer F, Sappa PK, Samson F, Schaffer M, Schwikowski B, Steil L, Stülke J, Wiegert T, Devine KM, Wilkinson AJ, van Dijk JM, Hecker M, Völker U, Bessières P, Noirot P. Mar 2012. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Sci (New York, N.Y.)* 335 (6072):1103–1106. doi:10.1126/science.1206848.
- Yan B, Boitano M, Clark TA, Ettwiller L. Sep 2018. SMRT-Cappable-seq reveals complex operon variants in bacteria. *Nat Commun* 9 (1):1–11. doi:10.1038/s41467-018-05997-6.
- Warrier I, Ram-Mohan N, Zhu Z, Hazery A, Echlin H, Rosch J, Meyer MM, Opijnen Tv. Dec 2018. The Transcriptional landscape of *Streptococcus pneumoniae* TIGR4 reveals a complex operon architecture and abundant riboregulation critical for growth and virulence. *PLOS Pathog* 14 (12):e1007461. doi:10.1371/journal.ppat.1007461.
- Mejía-Almonte C, Busby SJW, Wade JT, van Helden J, Arkin AP, Stormo GD, Eilbeck K, Palsson BO, Galagan JE, Collado-Vides J. Jul 2020. Redefining fundamental concepts of transcription initiation in bacteria. *Nat Rev Genet* p 1–16. doi:10.1038/s41576-020-0254-8.
- Adhya S. Jun 2003. Suboperonic regulatory signals. *Sci STKE: signal transduction knowledge environment* 2003 (185):pe22. doi:10.1126/stke.2003.185.pe22.
- Kornbliht AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. Mar 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* 14 (3):153–165. doi:10.1038/nrm3525.
- Junier I, Rivoire O. 2016. Conserved Units of Co-Expression in Bacterial Genomes: An Evolutionary Insight into Transcriptional Regulation. *PLOS One* 11 (5):e0155740. doi:10.1371/journal.pone.0155740.
- Conway T, Creecy JP, Maddox SM, Grissom JE, Conkle TL, Shadid TM, Teramoto J, San Miguel P, Shimada T, Ishihama A, Mori H, Wanner BL. 2014. Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *mBio* 5 (4):01442–01414. doi:10.1128/mBio.01442-14.
- Kröger C, Dillon SC, Cameron ADS, Papenfort K, Sivasankaran SK, Hokamp K, Chao Y, Sittka A, Hébrard M, Händler K, Colgan A, Leekitcharoenphon P, Langridge GC, Lohan AJ, Loftus B, Lucchini S, Ussery DW, Dorman CJ, Thomson NR, Vogel J, Hinton JCD. May 2012. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci United States Am* 109 (20):E1277–1286. doi:10.1073/pnas.1201061109.
- Dugar G, Herbig A, Förstner KU, Heidrich N, Reinhardt R, Nieselt K, Sharma CM. May 2013. High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. *PLOS genetics* 9 (5):e1003495. doi:10.1371/journal.pgen.1003495.
- Wang Y, Li X, Mao Y, Blaschek HP. Sep 2011. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. *BMC Genom* 12:479. doi:10.1186/1471-2164-12-479.
- Uplekar S, Rougemont J, Cole ST, Sala C. Jan 2013. High-resolution transcriptome and genome-wide dynamics of RNA polymerase and NusA in *Mycobacterium tuberculosis*. *Nucleic Acids Res* 41 (2):961–977. doi:10.1093/nar/gks1260.
- Güell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kühner S, Rode M, Suyama M, Schmidt S, Gavin AC, Bork P, Serrano L. Nov 2009. Transcriptome complexity in a genome-reduced bacterium. *Sci (New York, N.Y.)* 326 (5957):1268–1271. doi:10.1126/science.1176951.
- Schmidtke C, Findeiss S, Sharma CM, Kuhfuss J, Hoffmann S, Vogel J, Stadler PF, Bonas U. Mar 2012. Genome-wide transcriptome analysis of the plant pathogen *Xanthomonas* identifies sRNAs with putative virulence functions. *Nucleic Acids Res* 40 (5):2020–2031. doi:10.1093/nar/gkr904.
- Alkhateeb RS, Vorhölter FJ, Rückert C, Mentz A, Wibberg D, Hublik G, Niehaus K, Pühler A. May 2016. Genome wide transcription start sites analysis of *Xanthomonas campestris* pv. *campestris* B100 with insights into the gum gene cluster directing the biosynthesis of the exopolysaccharide xanthan. *J Biotechnol* 225:18–28. doi:10.1016/j.jbiotec.2016.03.020.
- Hugouvieux-Cotte-Pattat N, Condemine G, Gueguen E, Shevchik VE. 2020. *Dickeya* Plant Pathogens, p 1–10. *In* eLS. American Cancer Society. doi:10.1002/9780470015902.a0028932.
- Fujikawa T, Ota N, Sasaki M, Nakamura T, Iwanami T. Jul 2019. Emergence of apple bacterial quick decline caused by *Dickeya dadantii* in Japan. *J Gen Plant Pathol* 85 (4):314–319. doi:10.1007/s10327-019-00852-y.
- Toth IK, Wolf JMvd, Saddler G, Lojkowska E, Hélias V, Pirhonen M, Tsror (Lahkim) L, Elphinstone JG. 2011. *Dickeya* species: an emerging problem for potato production in Europe. *Plant Pathol* 60 (3):385–399. doi:10.1111/j.1365-3059.2011.02427.x.
- Jiang HH, Hao JJ, Johnson SB, Brueggeman RS, Secor G. Jul 2016. First Report of *Dickeya dianthicola* Causing Blackleg and Bacterial Soft Rot on Potato in Maine. *Plant Dis* 100 (11):2320–2320. doi:

Forquet et al.

- 10.1094/PDIS-12-15-1513-PDN.
21. **Pu XM, Zhou JN, Lin BR, Shen HF.** Dec 2012. First Report of Bacterial Foot Rot of Rice Caused by a *Dickeya zeae* in China. *Plant Dis* 96 (12):1818. doi:10.1094/PDIS-03-12-0315-PDN.
  22. **Ma B, Hibbing ME, Kim HS, Reedy RM, Yedidia I, Breuer J, Breuer J, Glasner JD, Perna NT, Kelman A, Charkowski AO.** Sep 2007. Host range and molecular phylogenies of the soft rot enterobacterial genera *pectobacterium* and *dickeya*. *Phytopathology* 97 (9):1150–1163. doi:10.1094/PHYTO-97-9-1150.
  23. **Reverchon S, Muskhelishvili G, Nasser W.** 2016. Virulence Program of a Bacterial Plant Pathogen: The *Dickeya* Model. *Prog Mol Biol Transl Sci* 142:51–92. doi:10.1016/bs.pmbts.2016.05.005.
  24. **Grignon C, Sentenac H.** 1991. pH and Ionic Conditions in the Apoplast. *Annu Rev Plant Physiol Plant Mol Biol* 42 (1):103–128. doi:10.1146/annurev.pp.42.060191.000535.
  25. **Lamb C, Dixon RA.** 1997. The Oxidative Burst in Plant Disease Resistance. *Annu Rev Plant Physiol Plant Mol Biol* 48 (1):251–275. doi:10.1146/annurev.arplant.48.1.251.
  26. **Lebeau A, Reverchon S, Gaubert S, Kraepiel Y, Simond-Côte E, Nasser W, Van Gijsegem F.** Mar 2008. The GacA global regulator is required for the appropriate expression of *Erwinia chrysanthemi* 3937 pathogenicity genes during plant infection. *Environ Microbiol* 10 (3):545–559. doi:10.1111/j.1462-2920.2007.01473.x.
  27. **Jiang X, Sobetzko P, Nasser W, Reverchon S, Muskhelishvili G.** Apr 2015. Chromosomal “Stress-Response” Domains Govern the Spatiotemporal Expression of the Bacterial Virulence Program. *mBio* 6 (3). doi:10.1128/mBio.00353-15.
  28. **Jiang X, Zghidi-Abouzid O, Oger-Desfeux C, Hommais F, Greliche N, Muskhelishvili G, Nasser W, Reverchon S.** 2016. Global transcriptional response of *Dickeya dadantii* to environmental stimuli relevant to the plant infection. *Environ Microbiol* 18 (11):3651–3672. doi:10.1111/1462-2920.13267.
  29. **Pédron J, Chapelle E, Alunni B, Van Gijsegem F.** 2018. Transcriptome analysis of the *Dickeya dadantii* PecS regulon during the early stages of interaction with *Arabidopsis thaliana*. *Mol Plant Pathol* 19 (3):647–663. doi:10.1111/mpp.12549.
  30. **Duprey A, Taib N, Leonard S, Garin T, Flandrois JP, Nasser W, Brochier-Armanet C, Reverchon S.** 2019. The phytopathogenic nature of *Dickeya aquatica* 174/2 and the dynamic early evolution of *Dickeya* pathogenicity. *Environ Microbiol* 21 (8):2809–2835. doi:10.1111/1462-2920.14627.
  31. **Nasser W, Condemine G, Plantier R, Anker D, Robert-Baudouy J.** Jun 1991. Inducing properties of analogs of 2-keto-3-deoxygluconate on the expression of pectinase genes of *Erwinia chrysanthemi*. *FEMS Microbiol Lett* 81 (1):73–78. doi:10.1111/j.1574-6968.1991.tb04715.x.
  32. **Ouafa ZA, Reverchon S, Lautier T, Muskhelishvili G, Nasser W.** 2012. The nucleoid-associated proteins H-NS and FIS modulate the DNA supercoiling response of the *pel* genes, the major virulence factors in the plant pathogen bacterium *Dickeya dadantii*. *Nucleic Acids Res* 40 (10):4306–4319. doi:10.1093/nar/gks014.
  33. **Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG, Mohr I, Wilson AC.** Feb 2019. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat Commun* 10 (1):754. doi:10.1038/s41467-019-08734-9.
  34. **Jenjaroenpun P, Wongsurawat T, Pereira R, Patumcharoenpol P, Ussery DW, Nielsen J, Nookaew I.** Apr 2018. Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN.PK113-7D. *Nucleic Acids Res* 46 (7):e38. doi:10.1093/nar/gky014.
  35. **Parker MT, Knop K, Sherwood AV, Schurch NJ, Mackinnon K, Gould PD, Hall AJ, Barton GJ, Simpson GG.** Jan 2020. Nanopore direct RNA sequencing maps the complexity of *Arabidopsis* mRNA processing and m6A modification. *eLife* 9:e49658. doi:10.7554/eLife.49658.
  36. **Workman RE, Tang AD, Tang PS, Jain M, Tyson JR, Razaghi R, Zuzarte PC, Gilpatrick T, Payne A, Quick J, Sadowski N, Holmes N, de Jesus JG, Jones KL, Soulette CM, Snutch TP, Loman N, Paten B, Loose M, Simpson JT, Olsen HE, Brooks AN, Akeson M, Timp W.** Dec 2019. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods* 16 (12):1297–1305. doi:10.1038/s41592-019-0617-2.
  37. **Pitt ME, Nguyen SH, Duarte TPS, Teng H, Blaskovich MAT, Cooper MA, Coin LJM.** Feb 2020. Evaluating the genome and resistome of extensively drug-resistant *Klebsiella pneumoniae* using native DNA and RNA Nanopore sequencing. *GigaScience* 9 (giaa002). doi:10.1093/gigascience/giaa002.
  38. **Sharma CM, Vogel J.** Jun 2014. Differential RNA-seq: the approach behind and the biological insight gained. *Curr Opin Microbiol* 19:97–105. doi:10.1016/j.mib.2014.06.010.
  39. **Oyelade J, Isewon I, Oladipupo F, Aromolaran O, Uwoghien E, Ameh F, Achas M, Adebisi E.** Nov 2016. Clustering Algorithms: Their Application to Gene Expression Data. *Bioinform Biol Insights* 10:237–253. doi:10.4137/BBI.S38316.
  40. **Tjaden B.** Apr 2019. A computational system for identifying operons based on RNA-seq data. *Methods* doi:10.1016/j.ymeth.2019.03.026.
  41. **Yamanaka K, Ogura T, Niki H, Hiraga S.** Nov 1995. Characterization of the *smtA* gene encoding an S-adenosylmethionine-dependent methyltransferase of *Escherichia coli*. *FEMS microbiology letters* 133 (1-2):59–63. doi:10.1111/j.1574-6968.1995.tb07861.x.
  42. **López-Solanilla E, García-Olmedo F, Rodríguez-Palenzuela P.** Jun 1998. Inactivation of the *sapA* to *sapF* locus of *Erwinia chrysanthemi* reveals common features in plant and animal bacterial pathogenesis. *The Plant Cell* 10 (6):917–924. doi:10.1105/tpc.10.6.917.
  43. **Cronan JE, Thomas J.** 2009. Bacterial Fatty Acid Synthesis and its Relationships with Polyketide Synthetic Pathways. *Methods enzymology* 459:395–433. doi:10.1016/S0076-6879(09)04617-5.
  44. **Bergler H, Fuchsbichler S, Högenauer G, Turnowsky F.** Dec 1996. The enoyl-[acyl-carrier-protein] reductase (FabI) of *Escherichia coli*, which catalyzes a key regulatory step in fatty acid biosynthesis, accepts NADH and NADPH as cofactors and is inhibited by palmitoyl-CoA. *Eur J Biochem* 242 (3):689–694. doi:10.1111/j.1432-1033.1996.0689x.
  45. **Preiss J.** Aug 2009. Glycogen: Biosynthesis and Regulation. *EcoSal Plus* 3 (2). doi:10.1128/ecosalplus.4.7.4.
  46. **Montero M, Almagro G, Eydallin G, Viale AM, Muñoz FJ, Bahaji A, Li J, Rahimpour M, Baroja-Fernández E, Pozueta-Romero J.** Jan 2011. *Escherichia coli* glycogen genes are organized in a single *glgBCAP* transcriptional unit possessing an alternative suboperonic promoter within *glgC* that directs *glgAP* expression. *The Biochem J* 433 (1):107–117. doi:10.1042/BJ20101186.
  47. **Patzter SI, Hantke K.** Aug 2000. The zinc-responsive regulator Zur and its control of the *znu* gene cluster encoding the ZnuABC zinc uptake system in *Escherichia coli*. *The J Biol Chem* 275 (32):24321–24332. doi:10.1074/jbc.M001775200.
  48. **Jorjani H, Zavolan M.** Apr 2014. TSSer: an automated method to identify transcription start sites in prokaryotic genomes from differential RNA sequencing data. *Bioinform (Oxford, England)* 30 (7):971–974. doi:10.1093/bioinformatics/btt752.
  49. **Ray-Soni A, Bellecourt MJ, Landick R.** 2016. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu Rev Biochem* 85 (1):319–347. doi:10.1146/annurev-biochem-060815-



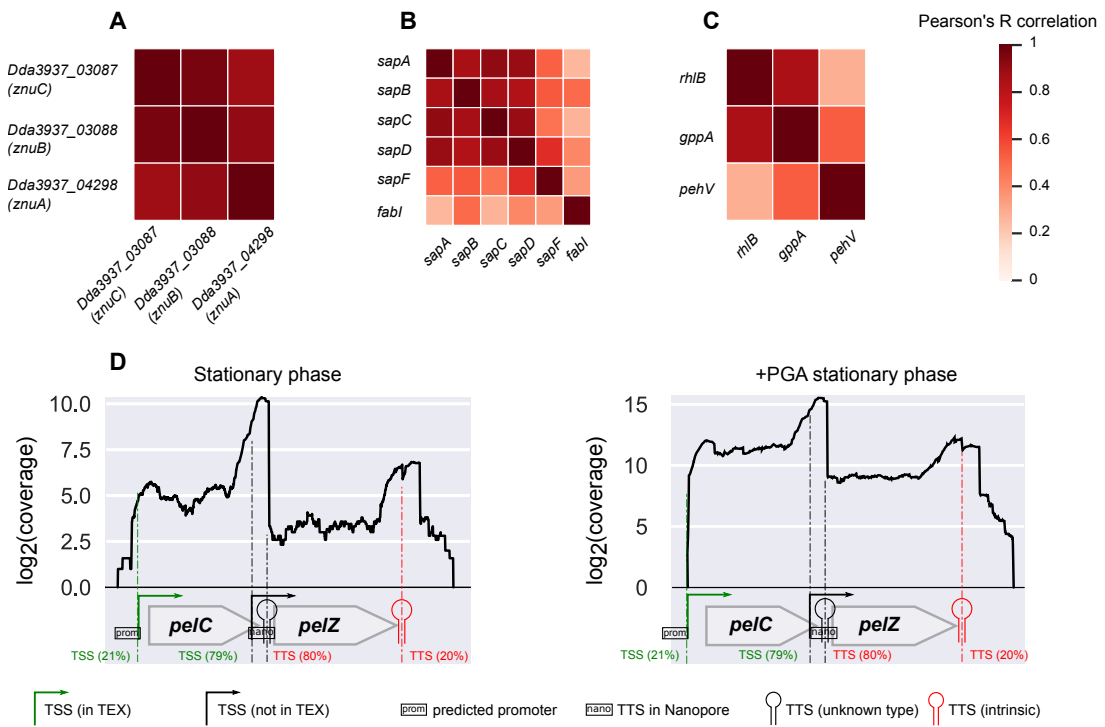
Mapping of *Dickeya dadantii* transcriptional landscape

- 014844.
50. **Naville M, Ghuillot-Gaudeffroy A, Marchais A, Gautheret D.** Feb 2011. ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA biology* 8 (1):11–13. doi: [10.4161/rna.8.1.13346](https://doi.org/10.4161/rna.8.1.13346).
  51. **Di Salvo M, Puccio S, Peano C, Lacour S, Alifano P.** Mar 2019. RhoTermPredict: an algorithm for predicting Rho-dependent transcription terminators based on *Escherichia coli*, *Bacillus subtilis* and *Salmonella enterica* databases. *BMC Bioinform* 20 (1):117. doi: [10.1186/s12859-019-2704-x](https://doi.org/10.1186/s12859-019-2704-x).
  52. **Shahmuradov IA, Mohamad Razali R, Bougouffa S, Radovanovic A, Bajic VB.** Feb 2017. bTSSfinder: a novel tool for the prediction of promoters in cyanobacteria and *Escherichia coli*. *Bioinformatics* 33 (3):334–340. doi: [10.1093/bioinformatics/btw629](https://doi.org/10.1093/bioinformatics/btw629).
  53. **Silva SdAe, Echeverrigaray S.** Nov 2012. Bacterial Promoter Features Description and Their Application on *E. coli* in silico Prediction and Recognition Approaches. *Bioinformatics* doi: [10.5772/48149](https://doi.org/10.5772/48149).
  54. **Ju X, Li D, Liu S.** Nov 2019. Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. *Nat Microbiol* 4 (11):1907–1918. doi: [10.1038/s41564-019-0500-z](https://doi.org/10.1038/s41564-019-0500-z).
  55. **Pissavin C, Robert-Baudouy J, Hugouvieux-Cotte-Pattat N.** Dec 1996. Regulation of *pelZ*, a gene of the *pelB-pelC* cluster encoding a new pectate lyase of *Erwinia chrysanthemi* 3937. *J Bacteriol* 178 (24):7187–7196. doi: [10.1128/jb.178.24.7187-7196.1996](https://doi.org/10.1128/jb.178.24.7187-7196.1996).
  56. **Garibaldi A, Bateman DF.** Jan 1971. Pectic enzymes produced by *Erwinia chrysanthemi* and their effects on plant tissue. *Physiol Plant Pathol* 1 (1):25–40. doi: [10.1016/0048-4059\(71\)90037-3](https://doi.org/10.1016/0048-4059(71)90037-3).
  57. **Gusarov I, Nudler E.** Nov 2001. Control of Intrinsic Transcription Termination by N and NusA: The Basic Mechanisms. *Cell* 107 (4):437–449. doi: [10.1016/S0092-8674\(01\)00582-7](https://doi.org/10.1016/S0092-8674(01)00582-7).
  58. **Morita T, Ueda M, Kubo K, Aiba H.** Aug 2015. Insights into transcription termination of Hfq-binding sRNAs of *Escherichia coli* and characterization of readthrough products. *RNA* 21 (8):1490–1501. doi: [10.1261/rna.051870.115](https://doi.org/10.1261/rna.051870.115).
  59. **Stringer AM, Currenti S, Bonocora RP, Baranowski C, Petrone BL, Palumbo MJ, Reilly AA, Zhang Z, Erill I, Wade JT.** Feb 2014. Genome-Scale Analyses of *Escherichia coli* and *Salmonella enterica* AraC Reveal Noncanonical Targets and an Expanded Core Regulon. *J Bacteriol* 196 (3):660–671. doi: [10.1128/JB.01007-13](https://doi.org/10.1128/JB.01007-13).
  60. **Boudvillain M, Figueroa-Bossi N, Bossi L.** Apr 2013. Terminator still moving forward: expanding roles for Rho factor. *Curr Opin Microbiol* 16 (2):118–124. doi: [10.1016/j.mib.2012.12.003](https://doi.org/10.1016/j.mib.2012.12.003).
  61. **Burns CM, Richardson LV, Richardson JP.** May 1998. Combinatorial effects of NusA and NusG on transcription elongation and rho-dependent termination in *Escherichia coli*. *J Mol Biol* 278 (2):307–316. doi: [10.1006/jmbi.1998.1691](https://doi.org/10.1006/jmbi.1998.1691).
  62. **Mondal S, Yakhnin AV, Babitzke P.** 2017. Modular Organization of the NusA- and NusG-Stimulated RNA Polymerase Pause Signal That Participates in the *Bacillus subtilis* *trp* Operon Attenuation Mechanism. *J Bacteriol* 199 (14). doi: [10.1128/JB.00223-17](https://doi.org/10.1128/JB.00223-17).
  63. **Lawson MR, Berger JM.** Jun 2019. Tuning the sequence specificity of a transcription terminator. *Curr Genet* 65 (3):729–733. doi: [10.1007/s00294-019-00939-1](https://doi.org/10.1007/s00294-019-00939-1).
  64. **Merino E, Yanofsky C.** May 2005. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends genetics: TIG* 21 (5):260–264. doi: [10.1016/j.tig.2005.03.002](https://doi.org/10.1016/j.tig.2005.03.002).
  65. **Turnbough CL.** 2019. Regulation of Bacterial Gene Expression by Transcription Attenuation. *Microbiol molecular biology reviews: MMBR* 83 (3). doi: [10.1128/MMBR.00019-19](https://doi.org/10.1128/MMBR.00019-19).
  66. **Green NJ, Grundy FJ, Henkin TM.** Jan 2010. The T box mechanism: tRNA as a regulatory molecule. *FEBS letters* 584 (2):318–324. doi: [10.1016/j.febslet.2009.11.056](https://doi.org/10.1016/j.febslet.2009.11.056).
  67. **Zhang J, Chetnani B, Cormack ED, Alonso D, Liu W, Mondragón A, Fei J.** Sep 2018. Specific structural elements of the T-box riboswitch drive the two-step binding of the tRNA ligand. *eLife* 7:e39518. doi: [10.7554/eLife.39518](https://doi.org/10.7554/eLife.39518).
  68. **Millman A, Dar D, Shamir M, Sorek R.** Jan 2017. Computational prediction of regulatory, premature transcription termination in bacteria. *Nucleic Acids Res* 45 (2):886–893. doi: [10.1093/nar/gkw749](https://doi.org/10.1093/nar/gkw749).
  69. **Proshkin S, Mironov A, Nudler E.** Oct 2014. Riboswitches in regulation of Rho-dependent transcription termination. *Biochimica Et Biophys Acta* 1839 (10):974–977. doi: [10.1016/j.bbagrm.2014.04.002](https://doi.org/10.1016/j.bbagrm.2014.04.002).
  70. **Coburn GA, Miao X, Briant DJ, Mackie GA.** Oct 1999. Reconstitution of a minimal RNA degradosome demonstrates functional coordination between a 3' exonuclease and a DEAD-box RNA helicase. *Genes & Dev* 13 (19):2594–2603. doi: [10.1101/gad.13.19.2594](https://doi.org/10.1101/gad.13.19.2594).
  71. **Py B, Higgins CF, Krisch HM, Carpousis AJ.** May 1996. A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature* 381 (6578):169–172. doi: [10.1038/381169a0](https://doi.org/10.1038/381169a0).
  72. **Hauryliuk V, Atkinson GC, Murakami KS, Tenson T, Gerdes K.** May 2015. Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Microbiol* 13 (5):298–309. doi: [10.1038/nrmicro3448](https://doi.org/10.1038/nrmicro3448).
  73. **Nasser W, Shevchik VE, Hugouvieux-Cotte-Pattat N.** Nov 1999. Analysis of three clustered polygalacturonase genes in *Erwinia chrysanthemi* 3937 revealed an anti-repressor function for the PecS regulator. *Mol Microbiol* 34 (4):641–650. doi: [10.1046/j.1365-2958.1999.01609.x](https://doi.org/10.1046/j.1365-2958.1999.01609.x).
  74. **de Lorenzo V, Danchin A.** Sep 2008. Synthetic biology: discovering new worlds and new words. The new and not so new aspects of this emerging research field. *EMBO Reports* 9 (9):822–827. doi: [10.1038/embor.2008.159](https://doi.org/10.1038/embor.2008.159).
  75. **Stauffer LT, Fogarty SJ, Stauffer GV.** May 1994. Characterization of the *Escherichia coli* *gcv* operon. *Gene* 142 (1):17–22. doi: [10.1016/0378-1119\(94\)90349-2](https://doi.org/10.1016/0378-1119(94)90349-2).
  76. **Kikuchi G, Motokawa Y, Yoshida T, Hiraga K.** Jul 2008. Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. *Proc Jpn Acad Ser B, Phys Biol Sci* 84 (7):246–263. doi: [10.2183/pjab.84.246](https://doi.org/10.2183/pjab.84.246).
  77. **Shevchik VE, Hugouvieux-Cotte-Pattat N.** Jun 1997. Identification of a bacterial pectin acetyl esterase in *Erwinia chrysanthemi* 3937. *Mol Microbiol* 24 (6):1285–1301. doi: [10.1046/j.1365-2958.1997.4331800.x](https://doi.org/10.1046/j.1365-2958.1997.4331800.x).
  78. **Boccaro M, Diolez A, Rouve M, Kotoujansky A.** Jul 1988. The role of individual pectate lyases of *Erwinia chrysanthemi* strain 3937 in pathogenicity on saintpaulia plants. *Physiol Mol Plant Pathol* 33 (1):95–104. doi: [10.1016/0885-5765\(88\)90046-X](https://doi.org/10.1016/0885-5765(88)90046-X).
  79. **Beaulieu C.** 1993. Pathogenic Behavior of Pectinase-Defective *Erwinia chrysanthemi* Mutants on Different Plants. *Mol Plant-Microbe Interactions* 6 (2):197. doi: [10.1094/MPMI-6-197](https://doi.org/10.1094/MPMI-6-197).
  80. **Dorel C, Hugouvieux-Cotte-Pattat N, Robert-Baudouy J, Ljkwaska E.** Jul 1996. Production of *Erwinia chrysanthemi* pectinases in potato tubers showing high or low level of resistance to soft-rot. *Eur J Plant Pathol* 102 (6):511–517. doi: [10.1007/BF01877017](https://doi.org/10.1007/BF01877017).
  81. **Grenier AM, Dupont G, Pagès S, Condemine G, Rahbé Y.** Mar 2006. The Phytopathogen *Dickeya dadantii* (*Erwinia chrysanthemi* 3937) Is a Pathogen of the Pea Aphid. *Appl Environ Microbiol* 72 (3):1956–1965. doi: [10.1128/AEM.72.3.1956-1965.2006](https://doi.org/10.1128/AEM.72.3.1956-1965.2006).
  82. **Costechareyre D, Dridi B, Rahbé Y, Condemine G.** Dec 2010. Cyt toxin expression reveals an inverse regulation of insect and plant virulence factors of *Dickeya dadantii*. *Environ Microbiol* 12 (12):3290–3301. doi: [10.1111/j.1462-2920.2010.02305.x](https://doi.org/10.1111/j.1462-2920.2010.02305.x).

Forquet et al.

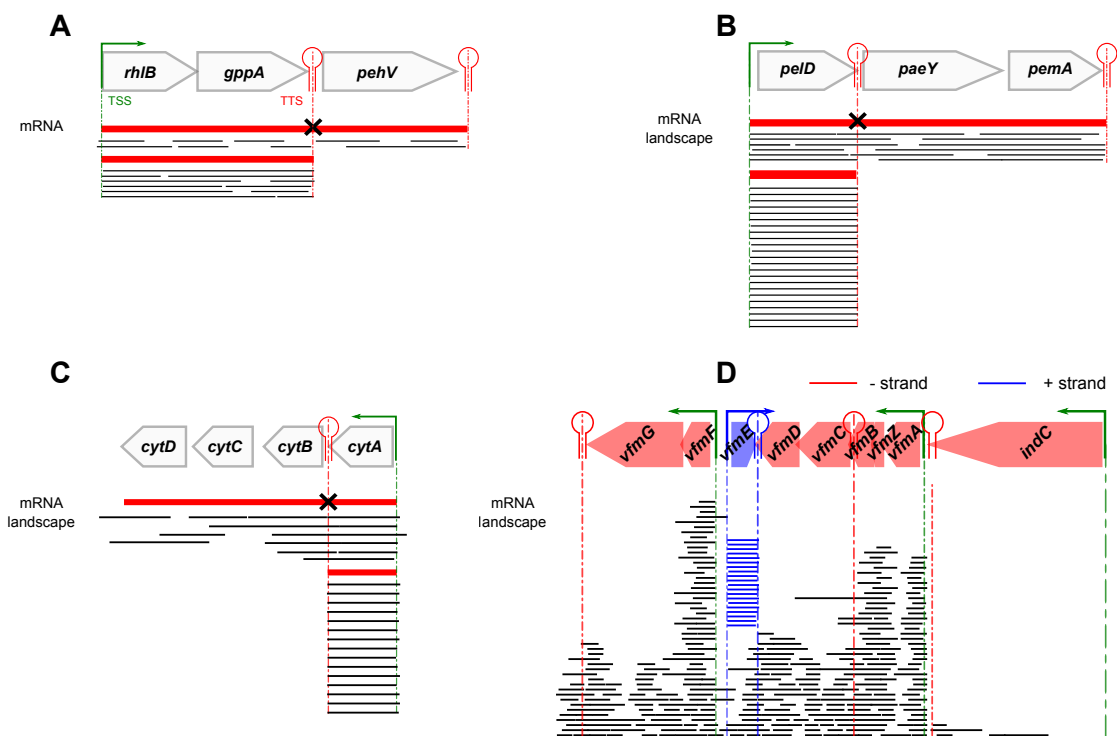
83. **Sáenz-Lahoya S, Bitarte N, García B, Burgui S, Vergara-Irigaray M, Valle J, Solano C, Toledo-Arana A, Lasa I.** 2019. Noncontiguous operon is a genetic organization for coordinating bacterial gene expression. *Proc Natl Acad Sci United States Am* 116 (5):1733–1738. doi: [10.1073/pnas.1812746116](https://doi.org/10.1073/pnas.1812746116).
84. **Nasser W, Dorel C, Wawrzyniak J, Van Gijsegem F, Groleau MC, Déziel E, Reverchon S.** Mar 2013. Vfm a new quorum sensing system controls the virulence of *Dickeya dadantii*. *Environ Microbiol* 15 (3):865–880. doi: [10.1111/1462-2920.12049](https://doi.org/10.1111/1462-2920.12049).
85. **Toledo-Arana A, Lasa I.** 2020. Advances in bacterial transcriptome understanding: From overlapping transcription to the excludon concept. *Mol Microbiol* 113 (3):593–602. doi: [10.1111/mmi.14456](https://doi.org/10.1111/mmi.14456).
86. **Baltenneck J, Reverchon S, Hommais F.** Jan 2021. Quorum Sensing Regulation in Phytopathogenic Bacteria. *Microorganisms* 9 (2). doi: [10.3390/microorganisms9020239](https://doi.org/10.3390/microorganisms9020239).
87. **Sesto N, Wurtzel O, Archambaud C, Sorek R, Cossart P.** Feb 2013. The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. *Nat Rev Microbiol* 11 (2):75–82. doi: [10.1038/nrmicro2934](https://doi.org/10.1038/nrmicro2934).
88. **Hsu LM, Vo NV, Chamberlin MJ.** Dec 1995. *Escherichia coli* transcript cleavage factors GreA and GreB stimulate promoter escape and gene expression in vivo and in vitro. *Proc Natl Acad Sci United States Am* 92 (25):11588–11592. doi: [10.1073/pnas.92.25.11588](https://doi.org/10.1073/pnas.92.25.11588).
89. **Cai SJ, Inouye M.** Jul 2002. EnvZ-OmpR interaction and osmoregulation in *Escherichia coli*. *The J Biol Chem* 277 (27):24155–24161. doi: [10.1074/jbc.M110715200](https://doi.org/10.1074/jbc.M110715200).
90. **Lybecker M, Zimmermann B, Bilusic I, Tukhtubaeva N, Schroeder R.** Feb 2014. The double-stranded transcriptome of *Escherichia coli*. *Proc Natl Acad Sci United States Am* 111 (8):3134–3139. doi: [10.1073/pnas.1315974111](https://doi.org/10.1073/pnas.1315974111).
91. **Reverchon S, Nasser W.** Oct 2013. *Dickeya* ecology, environment sensing and regulation of virulence programme. *Environ Microbiol Reports* 5 (5):622–636. doi: [10.1111/1758-2229.12073](https://doi.org/10.1111/1758-2229.12073).
92. **Nachin L, Loiseau L, Expert D, Barras F.** Feb 2003. SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress. *The EMBO J* 22 (3):427–437. doi: [10.1093/emboj/cdg061](https://doi.org/10.1093/emboj/cdg061).
93. **Magnet S, Bellais S, Dubost L, Fourgeaud M, Mainardi JL, Petit-Frère S, Marie A, Mengin-Lecreux D, Arthur M, Gutmann L.** May 2007. Identification of the I,d-Transpeptidases Responsible for Attachment of the Braun Lipoprotein to *Escherichia coli* Peptidoglycan. *J Bacteriol* 189 (10):3927–3931. doi: [10.1128/JB.00084-07](https://doi.org/10.1128/JB.00084-07).
94. **Roche B, Aussel L, Ezraty B, Mandin P, Py B, Barras F.** Mar 2013. Iron/sulfur proteins biogenesis in prokaryotes: Formation, regulation and diversity. *Biochimica et Biophys Acta (BBA) - Bioenerg* 1827 (3):455–469. doi: [10.1016/j.bbabi.2012.12.010](https://doi.org/10.1016/j.bbabi.2012.12.010).
95. **Collet JF, Cho SH, Iorga BI, Goemans CV.** Aug 2020. How the assembly and protection of the bacterial cell envelope depend on cysteine residues. *The J Biol Chem* 295 (34):11984–11994. doi: [10.1074/jbc.REV120.011201](https://doi.org/10.1074/jbc.REV120.011201).
96. **Leonard S, Meyer S, Lacour S, Nasser W, Hommais F, Reverchon S.** Sep 2019. APERO: a genome-wide approach for identifying bacterial small RNAs from RNA-Seq data. *Nucleic Acids Res* 47 (15):e88. doi: [10.1093/nar/gkz485](https://doi.org/10.1093/nar/gkz485).
97. **Glasner JD, Yang CH, Reverchon S, Hugouvieux-Cotte-Pattat N, Condemine G, Bohin JP, Van Gijsegem F, Yang S, Franza T, Expert D, Plunkett G, San Francisco MJ, Charkowski AO, Py B, Bell K, Rauscher L, Rodriguez-Palenzuela P, Toussaint A, Holeva MC, He SY, Douet V, Boccara M, Blanco C, Toth I, Anderson BD, Biehl BS, Mau B, Flynn SM, Barras F, Lindeberg M, Birch PRJ, Tsuyumu S, Shi X, Hibbing M, Yap MN, Carpentier M, Dassa E, Umehara M, Kim JF, Rusch M, Soni P, Mayhew GF, Fouts DE, Gill SR, Blattner FR, Keen NT, Perna NT.** Apr 2011. Genome Sequence of the Plant-Pathogenic Bacterium *Dickeya dadantii* 3937. *J Bacteriol* 193 (8):2076–2077. doi: [10.1128/JB.01513-10](https://doi.org/10.1128/JB.01513-10).
98. **El Houdaigui B, Forquet R, Hindré T, Schneider D, Nasser W, Reverchon S, Meyer S.** 2019. Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Res* 47 (11):5648–5657. doi: [10.1093/nar/gkz300](https://doi.org/10.1093/nar/gkz300).
99. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** Mar 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* 11:119. doi: [10.1186/1471-2105-11-119](https://doi.org/10.1186/1471-2105-11-119).
100. **Hommais F, Oger-Desfeux C, Gijsegem FV, Castang S, Ligor S, Expert D, Nasser W, Reverchon S.** Nov 2008. PecS Is a Global Regulator of the Symptomatic Phase in the Phytopathogenic Bacterium *Erwinia chrysanthemi* 3937. *J Bacteriol* 190 (22):7508–7522. doi: [10.1128/JB.00553-08](https://doi.org/10.1128/JB.00553-08).
101. **De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C.** Aug 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34 (15):2666–2669. doi: [10.1093/bioinformatics/bty149](https://doi.org/10.1093/bioinformatics/bty149).
102. **Li H.** Sep 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34 (18):3094–3100. doi: [10.1093/bioinformatics/bty191](https://doi.org/10.1093/bioinformatics/bty191).

## SUPPLEMENTARY FIGURE S1

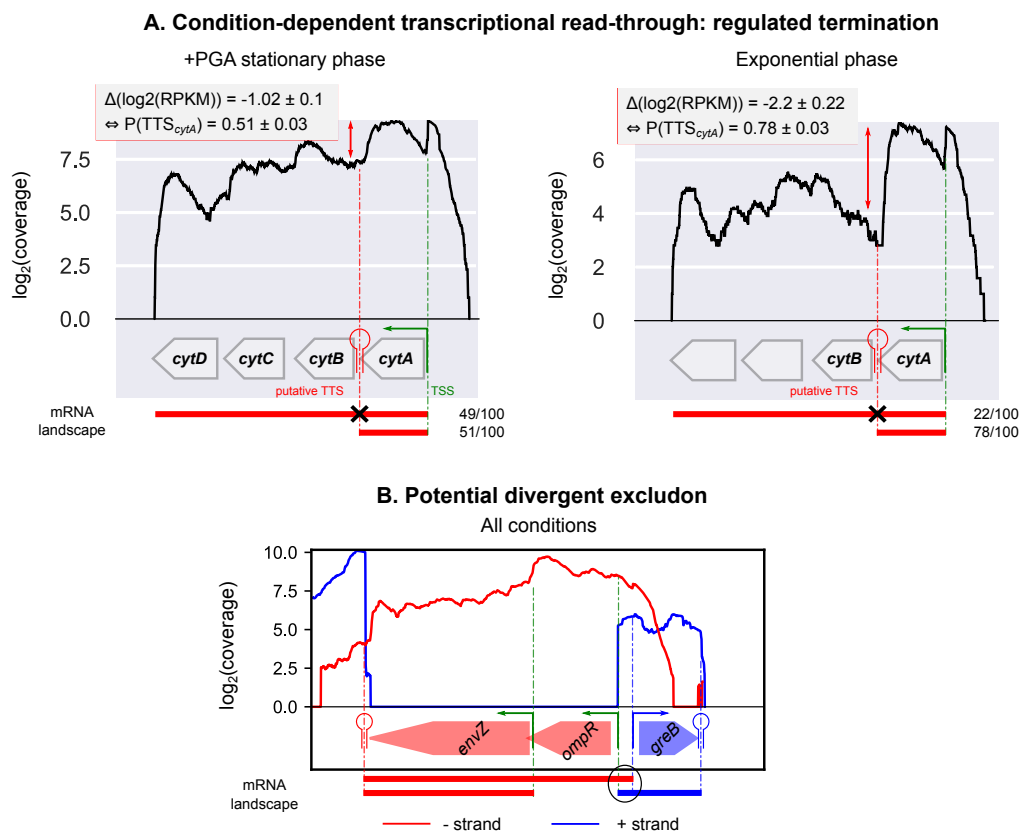




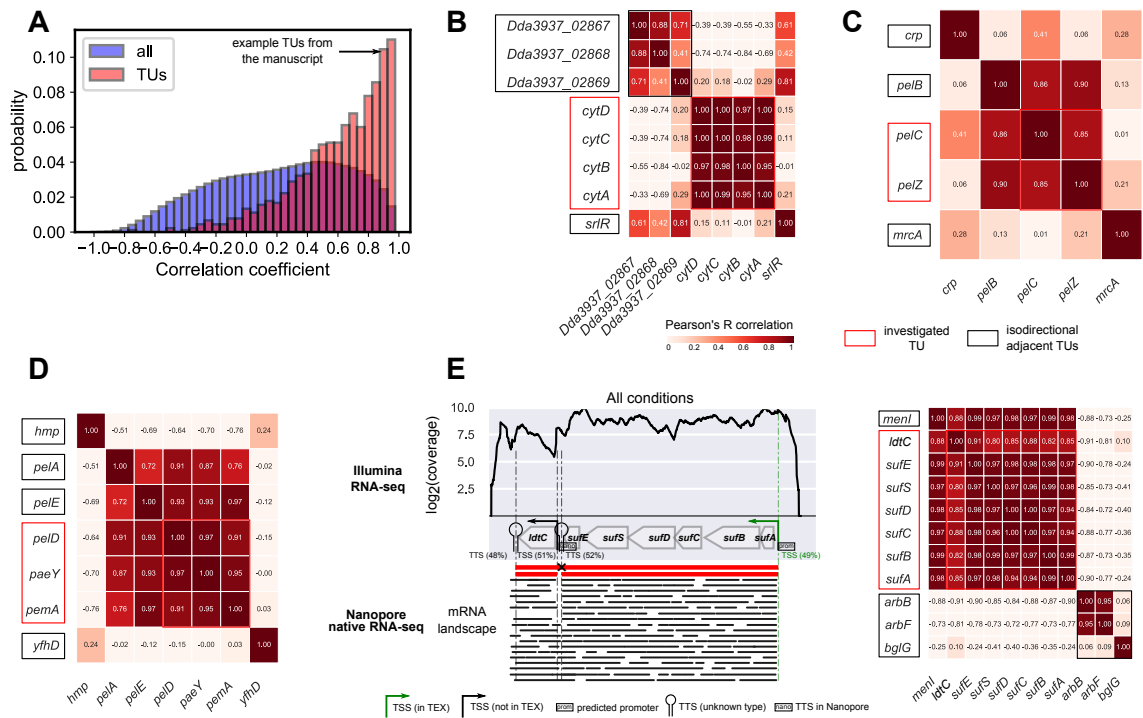
**SUPPLEMENTARY FIGURE S2**  
**Nanopore native RNA-seq**



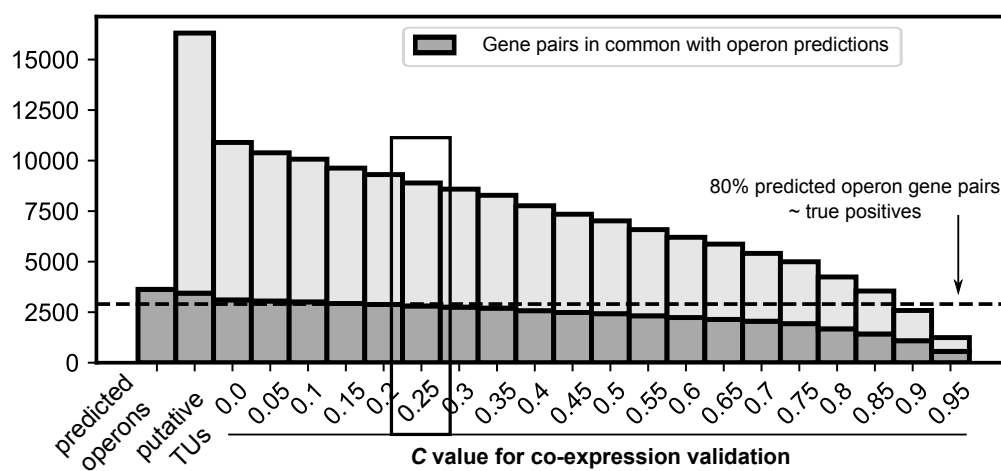
## SUPPLEMENTARY FIGURE S3



SUPPLEMENTARY FIGURE S4



## SUPPLEMENTARY FIGURE S5

Number of gene pairs predicted to be  
part of the same TU

## Chapter 4

# Role of the discriminator sequence in the supercoiling-sensitivity of bacterial promoters

Once *D. dadantii* transcription units were defined (Chapter 3), and the transcriptomic response of genes to SC variations together with TSS maps and promoters elements collected for several bacteria (Chapter 2), this allowed us to develop and validate global-scale, thermodynamic models of transcriptional regulation by SC. Since the latter affects transcription at many steps of the process, the upcoming chapters present two independent models based on distinct mechanisms which explain how SC modulates RNAP-promoter interactions, based on DNA physicochemical properties.

### 4.1 Inheritable increases of supercoiling during the long-term evolution experiment

The first model relies on the step of promoter opening during open-complex formation, and is described extensively in the following article. Here, we focus rather on the experimental data used in common for the validation of both models presented in the present Chapter and Chapter 5.

While most available transcriptomic data were obtained by sublethal antibiotic shocks with gyrase inhibitors inducing rapid and transient DNA relaxation [105, 108] (Tab. 2.1), the response of genes to the opposite SC variation,

*i.e.* DNA overtwisting, was derived from the long-term evolution experiment (LTEE) in *E. coli*. In the latter, inheritable increases of negative SC induce global modifications of the expression programme, and substantial fitness gains [139]. Indeed, at the scale of populations, environmental changes may result in the rapid emergence and spread of new traits [4]. In the LTEE which started in 1988 under the direction of Richard Lenski, an ancestral strain is evolved in a nutrient-poor medium where it alternates between a lag phase, exponential growth, and stationary phase due to depletion of the limiting glucose, until it is transferred into a fresh medium the next day [140] (Fig. 4.1A). Due to an average generation time of 20 minutes for *E. coli*, and the 100-fold dilution applied for each daily transfer, 6.7 generations are engendered each day, and thus, the populations studied reached 50K generations in 2010 [141]. Mutations are accumulated over generations, some of which are advantageous and increase fitness, *i.e.* the ability of bacteria to adjust their metabolism to suit new environmental conditions. In the LTEE, the gain in population fitness in the new environment is assessed by mixing the evolved strain with the ancestral strain, and by measuring the relative abundance of evolved cells compared to ancestral cells after one day of competition [142] (Fig. 4.1B). Gains in population fitness were quickly observed in the evolved strains, before 2K and 20K generations, due to the acquisition of 6 and 45 mutations, respectively [140]. Among them, a mutation in *topA* before 2K generations, and a second mutation in *fis* before 20K generations both led to an increased SC level (DNA overtwisting) in the evolved cells [140], demonstrating that SC is a key target of selection for adapting the global expression programme to new conditions, since it affects the expression of many genes [108]. The LTEE thus provides the transcriptomic response of genes to DNA overtwisting by comparing gene expression levels in the evolved strains compared to the ancestor [115]. Although it is important to note that in contrast to the rapid and transient DNA relaxation induced by a shock with gyrase inhibitors, here, the response of genes is rather adaptive, and the other accumulated mutations may also contribute to the observed response (see upcoming article).

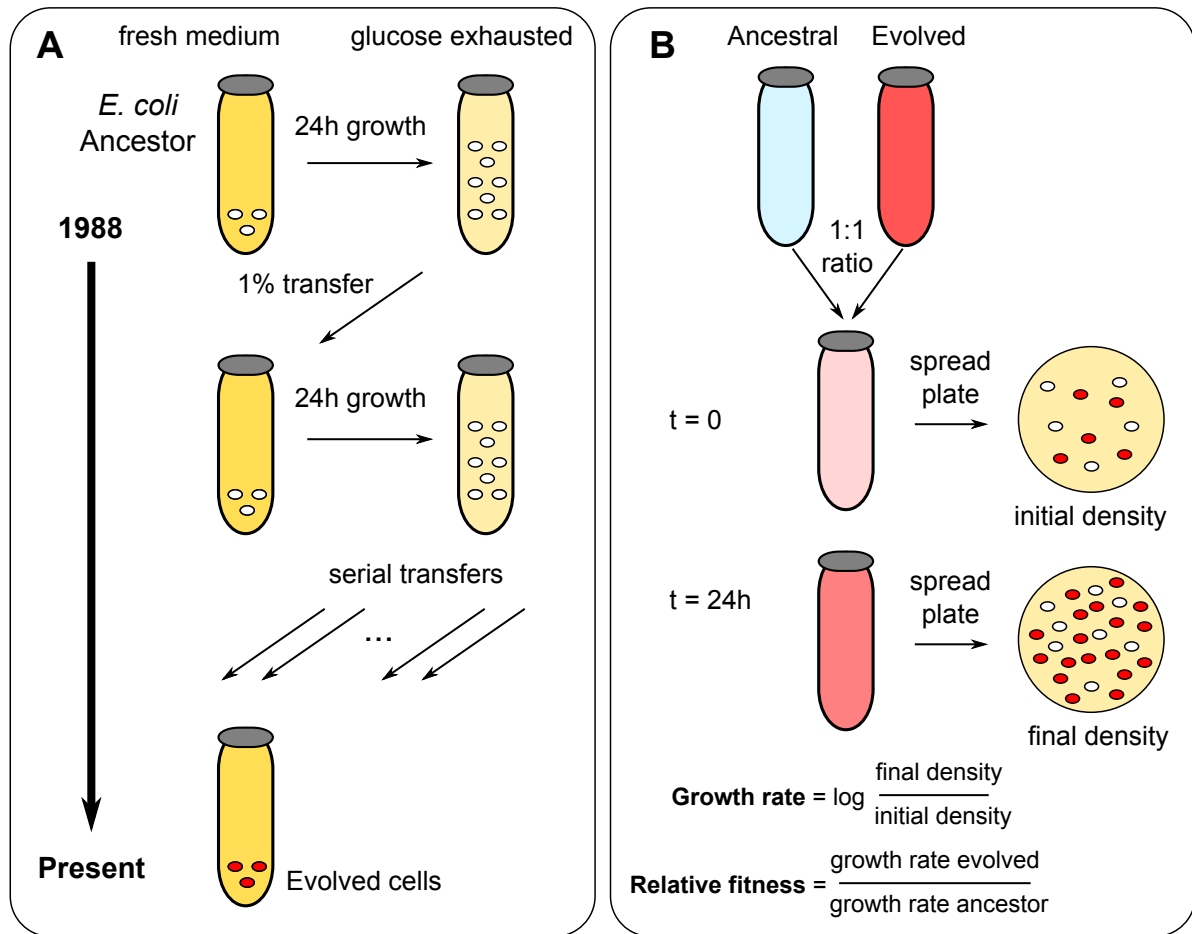


FIGURE 4.1: Principle of the long-term evolution experiment (LTEE). **(A)** The latter has been running since 1988 under the direction of Richard Lenski, where serial transfers of *E. coli* cultures grown in nutrient-poor medium supplemented with glucose are performed each day. An average of 6.7 generations are engendered each day due to the 100-fold dilution (see text). Mutations are accumulated over generations, some of which are beneficial and increase fitness. **(B)** Adapted from [142]: measurement of relative fitness of evolved strains compared to ancestors, by competition, in the same environment used for the LTEE. The competitors are first grown separately to ensure their acclimation to the conditions, then diluted 100-fold and mixed with a 1:1 ratio, and grown together for 24 hours. Initial and final densities are estimated by diluting and spreading the cells on agar plates that distinguish the evolved and ancestral cells by colony colour, which differs owing to an engineered marker. In the present case, white and red colonies correspond to Ara + (ancestor) and Ara - (evolved strain) phenotypes, respectively. This allows to measure growth rates, and relative fitness, as indicated.



## 4.2 Transcription assays on mutant promoters

While only statistical features emerge from the investigation of transcriptomic data, rather than a predictive signal dictating the response of each promoter, a complementary approach required transcription assays on individual promoters. To that end, I tested the validity of our models by measuring the regulatory effect of SC variations on mutant promoters of protein-encoding genes previously shown to be SC-sensitive. A first family was constructed based on the *pheP* promoter of *E. coli*, which is indeed SC-sensitive [109], and not regulated by any TF [46]. It was consequently an interesting target for the investigated regulatory mechanisms based on RNAP-DNA interaction. A second family was constructed based on the *pelE* and *pelD* promoters of *D. dadantii*, which control the expression of paralogous virulence genes encoding similar pectinolytic enzymes. Both are also SC-sensitive [119], but in contrast to *pheP* promoter, they exhibit a high regulation complexity with more than 10 identified TFs [7]. The SC response of such promoters may then include the contribution of regulatory proteins whose binding is sensitive to DNA topology [108]. They were consequently good candidates to test the independence between the investigated mechanisms based on RNAP-DNA interaction, and those relying on regulatory proteins. Further details on promoter sequences are provided in the upcoming section and chapter.

For the quantitative measurement of promoter expression levels, they were fused on pUcTerLuc plasmids (a pUC18 derivative) in front of a luciferase reporter gene, encoding firefly luciferase generating luminescence from luciferin substrate, and whose signal is proportional to promoter transcriptional rate. This luminescent reporter was retained to monitor the dynamics of transcription, in particular in response to rapid and transient SC variations. Indeed, it has a short lifetime, in contrast to fluorescent reporters which are generally more stable and less suited to report quick changes in transcriptional rates. I also checked that the expression patterns and SC responses of two promoters were consistent when inserted either in plasmid-borne or in chromosomal luciferase fusions (more details provided in the upcoming article).

The plasmids were then transformed into *E. coli* or *D. dadantii* cells, and grown in a Tecan Spark microplate reader (Fig. 4.2). This system was retained since it allows to test a wide range of growth conditions and promoters simultaneously (96 wells), while increasing the number of biological replicates.

Moreover, the conditions are generally uniform and stable across the samples, and temperature, shaking speed and amplitude are finely controlled throughout the experiment. The use of a humidity cassette also provides a protection against evaporation. Above all, it enables a fully automated monitoring of  $OD_{600nm}$  and luminescence signal intensities, reflecting bacterial growth and promoter transcriptional rates, respectively.

During the live-cell kinetic assay, rapid and transient DNA relaxation was then induced by a shock of the well-established gyrase inhibitor novobiocin at sub-lethal concentrations [105, 108]. It also affects TopoIV with a very low affinity [143], in contrast to quinolones, although this effect is probably insignificant at the employed dosage. The shock is usually applied in early exponential phase, where gyrase is more active. The quantitative response of promoters to novobiocin-induced DNA relaxation was then derived from the luminescence values reflecting transcriptional rates. The procedure is described in more detail in the upcoming articles, wherein the most relevant experiments are presented. Importantly, the precise and quantitative comparison of promoter expression levels required high data reproducibility. Reaching the latter was challenging, and thus a wide variety of conditions have been tested, and many protocol adjustments have been performed until I obtained the consistent results presented in the corresponding articles.

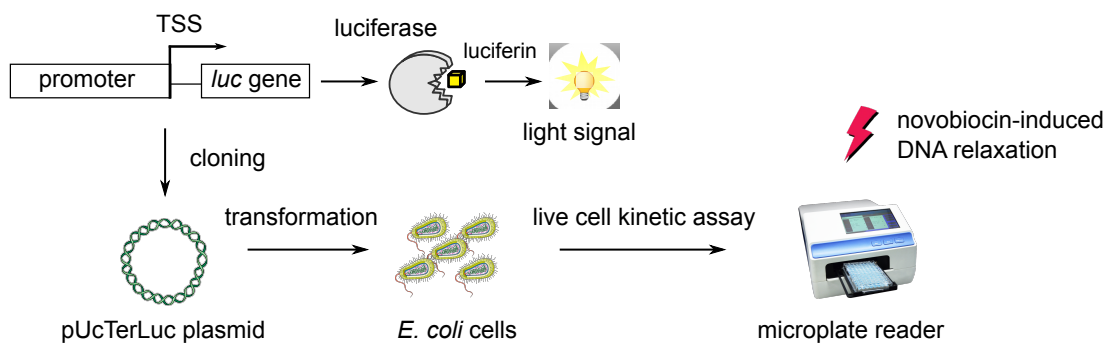


FIGURE 4.2: Principle of the transcription assays performed to measure the response of mutant promoters to SC variations. Promoters control the expression of a *luc* gene encoding firefly luciferase, generating luminescence from luciferin substrate. The constructs are cloned into pUcTerLuc plasmids, transformed into *E. coli* cells, and grown in a microplate reader to follow bacterial growth and promoter expression, with the measurement of the optical density at 600nm and the luminescence emission, respectively. A novobiocin shock is applied during growth to measure expression variations in response to DNA relaxation.

Based on a combination of such data and more, in the following article published in the mSystems journal [144], we present and valid a regulatory model of transcriptional regulation by SC, relying on the step of promoter opening during open-complex formation.

### **4.3 Article**



# Role of the Discriminator Sequence in the Supercoiling Sensitivity of Bacterial Promoters

 Raphaël Forquet,<sup>a</sup>  Maïwenn Pineau,<sup>a</sup>  William Nasser,<sup>a</sup>  Sylvie Reverchon,<sup>a</sup>  Sam Meyer<sup>a</sup>

<sup>a</sup>Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, CNRS, UMR5240, MAP, Lyon, France

**ABSTRACT** DNA supercoiling acts as a global transcriptional regulator that contributes to the rapid transcriptional response of bacteria to many environmental changes. Although a large fraction of promoters from phylogenetically distant species respond to superhelical variations, the sequence or structural determinants of this behavior remain elusive. Here, we focus on the sequence of the “discriminator” element that was shown to modulate this response in several promoters. We develop a quantitative thermodynamic model of this regulatory effect, focusing on open complex formation during transcription initiation independently from promoter-specific regulatory proteins. We analyze previous and new expression data and show that the model predictions quantitatively match the *in vitro* and *in vivo* supercoiling response of selected promoters with mutated discriminator sequences. We then test the universality of this mechanism by a statistical analysis of promoter sequences from transcriptomes of phylogenetically distant bacteria under conditions of supercoiling variations (i) by gyrase inhibitors, (ii) by environmental stresses, or (iii) inherited in the longest-running evolution experiment. In all cases, we identify a robust and significant sequence signature in the discriminator region, suggesting that supercoiling-modulated promoter opening underpins a ubiquitous regulatory mechanism in the prokaryotic kingdom based on the fundamental mechanical properties of DNA and its basal interaction with RNA polymerase.

**IMPORTANCE** In this study, we highlight the role of the discriminator as a global sensor of supercoiling variations and propose the first quantitative regulatory model of this principle, based on the specific step of promoter opening during transcription initiation. It defines the predictive rule by which DNA supercoiling quantitatively modulates the expression rate of bacterial promoters, depending on the G/C content of their discriminator and independently from promoter-specific regulatory proteins. This basal mechanism affects a wide range of species, which is tested by an extensive analysis of global high-throughput expression data. Altogether, our results confirm and provide a quantitative framework for the long-proposed notion that the discriminator sequence is a significant determinant of promoter supercoiling sensitivity, underpinning the ubiquitous regulatory action of DNA supercoiling on the core transcriptional machinery, in particular in response to quick environmental changes.

**KEYWORDS** DNA supercoiling, transcriptional regulation, quantitative modeling, discriminator, stress response, evolution, biophysics, computational biology


Bacteria encounter rapid changes of environmental conditions (availability of nutrients, physical or chemical stresses) to which they respond by quick and global modifications of their transcriptional program. Inspired by early studies, current mechanistic models of this regulatory action are mostly based on transcription factors (TFs) that bind at specific promoters and interact with RNA polymerase (RNAP). However, more than half of *Escherichia coli* promoters are not targeted by any known TF (1), and

**Citation** Forquet R, Pineau M, Nasser W, Reverchon S, Meyer S. 2021. Role of the discriminator sequence in the supercoiling sensitivity of bacterial promoters. *mSystems* 6: e00978-21. <https://doi.org/10.1128/mSystems.00978-21>.

**Editor** Pedro H. Oliveira, Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Université Évry, Université Paris-Saclay

**Copyright** © 2021 Forquet et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Sam Meyer, [sam.meyer@insa-lyon.fr](mailto:sam.meyer@insa-lyon.fr).

 The first predictive and quantitative relationship between the promoter sequence and its supercoiling sensitivity, valid in a wide range of bacterial species

**Received** 28 July 2021

**Accepted** 3 August 2021

**Published** 24 August 2021

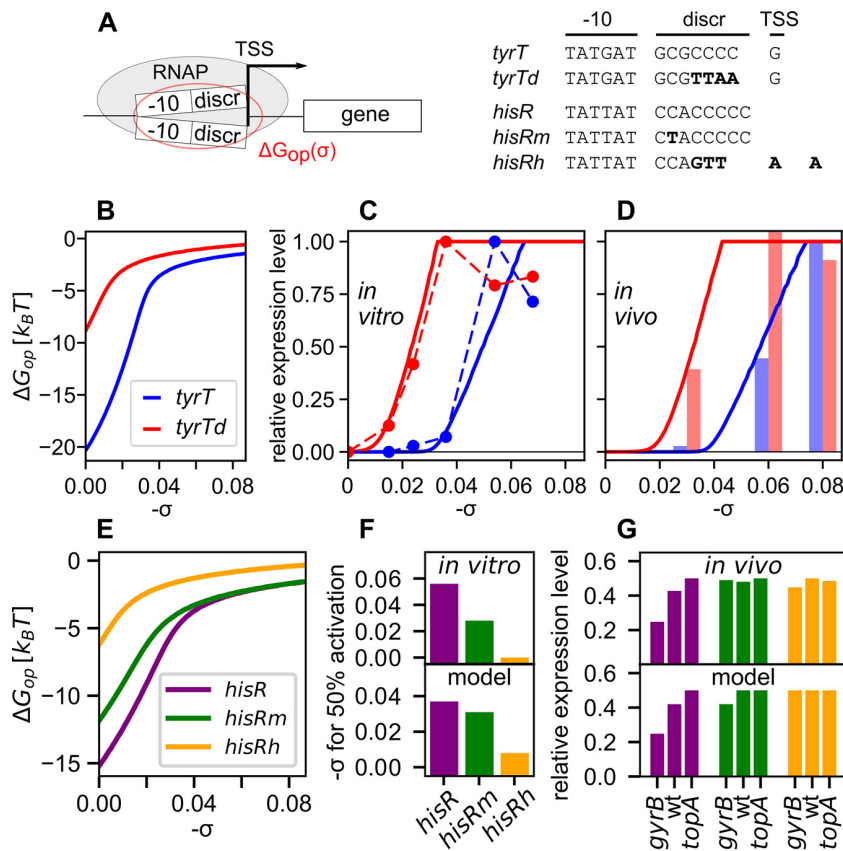
entire organisms are even almost devoid of them (2, 3) but nonetheless exhibit a complex regulation. Global transcriptional control has been further explained by variations in RNAP composition (sigma factors [4]) or abundance (5) depending on growth conditions as well as RNAP-binding regulatory molecules such as ppGpp (6).

Besides this variability of the transcription machinery, the physical state of the DNA template itself is subject to cellular control through DNA supercoiling (SC), i.e., the over- or underwinding of the double helix by the action of topoisomerase enzymes and architectural proteins (7–9). In bacteria, the chromosome is maintained at a negative SC level by the action of the DNA gyrase, which changes in response to environmental cues (9). This level was soon discovered to affect the expression of many promoters both *in vitro* and *in vivo* (10–14). Mechanistic studies showed that, besides modulating the binding of regulatory proteins (15), it could influence the activity of RNAP itself and, thus, could act as a global transcriptional factor (7–9). Accordingly, whole-genome analyses of the transcriptional response to DNA relaxation induced by gyrase inhibitors exhibited a broad response, providing lists of “supercoiling-sensitive genes” (3, 16–19).

In spite of its importance, no sequence or structural signature was ever clearly identified in support of the latter property. A possible reason is that SC affects transcription at many successive steps of the process, e.g., open complex formation (20, 21), promoter escape (10), elongation, and termination (22), and their combined action eluded the identification of simple determinants of supercoiling sensitivity. Additionally, transcription in turn affects the local level of SC (23), and, consequently, the response of a given promoter depends quite strongly on its genomic and physiological context (24, 25). Altogether, the complexity of the interaction between SC and transcription explains why there still are no models able to predict, even qualitatively, the response of a given promoter to variations of SC (9). The development of such predictive models is highly desirable considering the universality of superhelical variations in the prokaryotic kingdom.

One particular mechanism identified early as a putative strong factor in this response occurs at the step of open complex formation during transcription initiation (26). The unwinding of DNA strongly facilitates its denaturation and, thus, the formation of the “transcription bubble” by RNAP (11). Since this constraint affects all promoters, it may have a widespread effect on gene expression, yet the question then arises of how it may lead to transcriptional regulation, i.e., the selective activation/repression of a subset of promoters by global SC variations. An important observation was made when analyzing several stable RNA promoters as well as the *fis* promoter, which are both strongly SC sensitive and subject to stringent control (20, 27–30). Both properties are correlated with the presence of a G/C-rich discriminator sequence located between the  $-10$  element and the transcription start site (TSS) (31), which is denatured in the open complex. The discriminator has a variable length of 5 to 8 nucleotides (nt) and does not harbor any consensus sequence but is bound by the  $\sigma$ 1.2 domain of RNAP (32). Thus, it was postulated that the unusually high G/C content of these promoters affects the formation and stability of the open complex, which may then be modulated by SC, in contrast to mutant promoters containing an A/T-rich discriminator (20, 21, 30). However, it is not yet clear if this regulation mechanism is a specificity of some unusually G/C-rich promoters or a general regulatory principle by which SC quantitatively modulates the expression rate of bacterial promoters in a global and predictable manner.

In this paper, we consider the latter hypothesis and propose the first quantitative model of this mechanism, based on the free energy required to open the transcription bubble and related to the G/C content of the discriminator sequence. We show that it quantitatively recapitulates the *in vitro* and *in vivo* SC response of several promoters with mutant discriminator sequences, where the specific effect of this mechanism can be distinguished from other regulatory contributions of SC. Given its potentially broad regulatory effect, we then develop a statistical analysis of genome-wide expression data obtained after DNA relaxation by gyrase inhibitors and show that the discriminator indeed emerges as a primary location of global promoter selectivity under these conditions. We show that this sequence determinant is robustly detected in a series of phylogenetically distant bacterial species, and finally, we analyze this contribution under physiologically relevant



**FIG 1** (A) Sequences from wild-type *tyrT* and *hisR* promoters, the mutant *tyrTd* promoter with A/T-rich discriminator (21), and the mutants *hisRm* and *hisRh*, with 1 and 5 substitutions, C/G→A/T, in the discriminator (20). For *hisRh*, a shift in the transcription start site (TSS) (3 nt upstream) was observed. (B) Transcription bubble opening free energies of *tyrT* (B) and *tyrTd* promoters, computed from a thermodynamic model of DNA (see the text). (C and D) Transcription model predictions (solid lines) compared to the *in vitro* (dots) (C) and *in vivo* (bars) (D) expression data from reference 21. Data and computed values of the *tyrT* promoter are shown in blue, and those of *tyrTd* are in red. (E) Transcription bubble opening free energies of *hisR*, *hisRm*, and *hisRh* promoters. (F and G) Transcription model predictions compared to the *in vitro* (F) and *in vivo* (G) expression data from reference 20. Data and computed values of the *hisR*, *hisRm*, and *hisRh* promoters are shown in purple, green, and orange, respectively.

conditions involving SC changes, induced either transiently in response to environmental stress or inheritably in the longest-running evolution experiment. Altogether, this study highlights the role of the discriminator, previously observed in a few promoters, as a global sensor of SC variations that acts independently from promoter-specific regulatory proteins and according to a predictive rule inscribed in its physical properties.

## RESULTS

**Regulatory effect of the discriminator sequence in stable RNA promoters.** We first developed a quantitative model of SC-dependent transcriptional regulation based on the discriminator sequence. Negative SC destabilizes the double helix and facilitates the melting of the transcription bubble during open complex formation, which encompasses this sequence as shown in Fig. 1A. The melting energy is computed in Fig. 1B for the *tyrT* promoter (Fig. 1A) of the tyrosine tRNA operon, using a physical model of DNA denaturation (see Materials and Methods). Based on that curve, variations of the SC level should then directly affect the opening facility of promoters and, thus, their expression, and such a dependence was indeed observed for the *tyrT* promoter (blue) in both *in vitro* (Fig. 1C) or *in vivo* (Fig. 1D) transcription assays (21) (the *in vivo* SC levels are taken from reference 17). Further, the DNA denaturation energy is known to be



strongly dependent on the proportion of G/C bases, and while the A/T-rich sequence of the  $-10$  hexamer is relatively constrained due to its role in promoter recognition by the sigma factor, replacing four C/G by A/T nucleotides in the discriminator (*tyrTd* mutant) indeed strongly shifts the opening curve to the left (Fig. 1B, red curve), i.e., favors DNA opening already at weaker SC levels. Strikingly, the resulting transcriptional activation curves (Fig. 1C and D) closely follow the thermodynamic predictions.

We propose a thermodynamic model of this regulation step, based primarily on the promoter DNA opening curves (Fig. 1B), which is described in detail in Materials and Methods. It involves a single unknown parameter, representing opening assistance by RNAP, which was fitted on the data of Fig. 1 and kept constant henceforth for all promoters (thus neglecting the sequence dependence of the interaction of the discriminator with RNAP). The model reproduces most features of *in vitro* and *in vivo* activation curves of the analyzed promoters based on *tyrT* (solid lines in Fig. 1C and D). We tested it further using a similar data set collected independently based on the promoter of *hisR*, the histidine tRNA of *S. enterica* (20). *In vitro* (Fig. 1F), the expression increases with negative SC, both in the WT and in mutant promoters of variable G/C richness in the discriminator, closely following the DNA opening curves of the associated sequences (Fig. 1E), and, thus, are approximately reproduced by the model without any parameter adjustment. *In vivo*, only the native promoter was affected (Fig. 1G) in topoisomerase mutant strains exhibiting a global SC shift either in the direction of DNA relaxation (*gyrB* mutant) or SC increase (*topA*). This feature was reproduced using the experimentally measured SC levels of these strains (33), suggesting that the two A/T-rich mutant promoters have reached a plateau where the denaturation energy and, hence, the expression level is almost independent of SC.

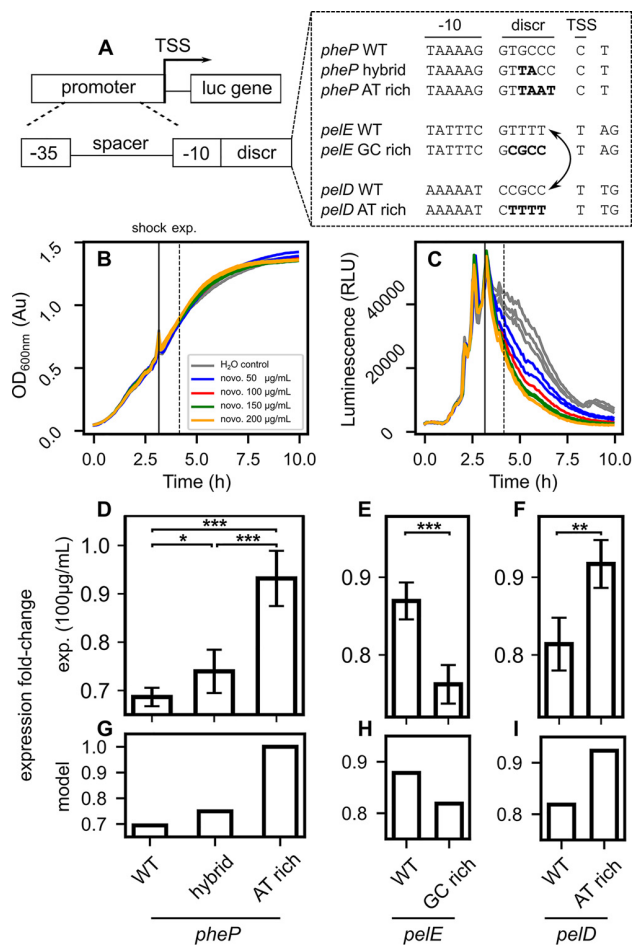
The model was kept voluntarily as simple as possible, since this mechanism is only one of the multiple steps by which SC affects transcription (as further developed in Discussion) and a reduced number of adjustable parameters was a key advantage. The approximations used in the modeling of this specific step as well as those other contributing factors may explain the slight discrepancies with the data (see details in Materials and Methods), but the clear overall agreement supports the notion that the proposed mechanism is the primary contributor in the SC sensitivity of promoters controlled by the discriminator sequence.

**Validation of model predictions on mutant mRNA promoters.** We then further tested the validity of the model by measuring the regulatory effect (expression fold change) of superhelical variations on mutant promoters of protein-coding genes with different features. Two families of synthetic promoters were constructed (Fig. 2A; see also Table S1 in the supplemental material). The first family is based on the *pheP* promoter of *E. coli*, which is SC sensitive (16, 17) and not regulated by any identified TF (1) and is an interesting candidate for our regulation mechanism based on the basal interaction with RNAP; these promoters were analyzed in LB medium, where gyrase activity is high (7). The second family is made of the paralogous virulence genes *pelD-pelE* of the enterobacterial phytopathogen *Dickeya dadantii*, encoding similar pectinolytic enzymes; in contrast to *pheP*, these genes exhibit a high regulation complexity, with more than 10 identified TFs, and both are supercoiling sensitive (34) but harbor different discriminators. These promoters were analyzed in minimal medium, which is closer to their physiologically relevant conditions (plant apoplast).

Promoters were fused on plasmids in front of a luciferase reporter gene (Fig. 2A), and their expression was analyzed in *E. coli* cells in a microplate reader after treatment by novobiocin, which relaxes the chromosomal DNA by inhibiting gyrase and, to a lesser extent, topoisomerase IV (35). The employed plasmids are well established as reflecting the average SC level of the chromosome (36), in particular during DNA relaxation by novobiocin (34, 37, 38).

We first checked that the presence of the plasmids did not affect bacterial growth and that the expression patterns of two promoters as well as their response to novobiocin were consistent when inserted either in plasmid-borne or in chromosomal luciferase





**FIG 2** DNA relaxation response of promoters with mutated discriminators. (A) Promoter sequences were derived from *pheP* (*E. coli*) and *pelD-pelE* (*D. dadantii*), with mutated discriminators of various G/C contents. (B) Bacterial growth monitored in a microplate reader (*E. coli* bacteria carrying plasmids with *pheP* hybrid promoter in rich medium). A novobiocin shock was applied in mid-exponential phase (different sublethal concentrations are shown). The slight increase at shock time is an optical artifact due to the opening of the recorder. (C) Expression of the *pheP* hybrid promoter monitored by luminescence (see all raw data points in Fig. S2). (D) Expression fold changes in response to relaxation computed 60 min after novobiocin shock (100  $\mu$ g/ml) in *pheP*-derived promoters. As expected, the repression factor reduces with increasing A/T%. (E and F) The DNA relaxation response of *pelE* (E) and *pelD* (F) are reversed when a tetranucleotide is swapped between their discriminators, with low and high G/C content, respectively. (G) Expression fold changes in response to relaxation predicted by the model reproduce the experimental observations on *pheP*-derived promoters as well as *pelE* (H)- and *pelD* (I)-derived promoters, assuming a weak relaxation compatible with the observed repression levels (see the text). Error bars represent 95% confidence intervals, and stars indicate the level of statistical significance (see Materials and Methods).

fusions (Fig. S1). These observations match previous similar comparisons involving other promoters and plasmids (30, 39) and confirm that the reduction in luminescence observed following the shock (raw data in Fig. S2) is due to SC-dependent transcriptional regulation rather than plasmid-specific effects. We then compared the relative effect of the novobiocin shock on the different plasmid-borne promoters. For the *pheP*-derived promoters (Fig. 2D), we found that the expression fold change (treated versus nontreated wells) was strongest for the native G/C-rich promoter and significantly reduced for the hybrid promoters (with two mutated nucleotides in the discriminator), whereas the A/T-rich discriminator (with four mutated nucleotides) was weakly sensitive to DNA relaxation. Thus, as already suggested *in vitro* with the *hisR* promoter (Fig. 1F), the SC sensitivity is progressively modulated by the discriminator G/C% *in vivo*. Similarly, swapping four nucleotides

between the discriminators of *pelE* and *pelD* (Fig. 2E and F) strikingly reversed their response to DNA relaxation. The relatively modest (but highly significant) repression levels are partly due to a buffering effect of the reporter system. Since the exact SC levels are not known under these conditions, we fitted the data using three adjustable parameters (an initial SC level for each growth medium and a common relaxation magnitude), which allowed us to reproduce the results with good accuracy (using, as expected, a stronger SC level in rich medium; Fig. 2G to I and Materials and Methods). Note that the direction of the promoters' predicted response is inscribed in their sequences and therefore is qualitatively robust when the exact value of these parameters is varied.

These results show that the effect of the discriminator on the SC sensitivity is not specific to G/C-rich ones (such as those of stable RNAs or *fis*) but is a quantitative effect that is progressively modulated by the G/C% and equally affects promoters with a naturally low G/C%, such as *pelE*, as expected from our modeling. It affects promoters of diverse biological functions and regulation complexities and is detectable under different physiological conditions (rich versus minimal medium). Based on these observations on a few selected promoters, and since the proposed mechanism of open complex formation is involved in RNAP-promoter interaction independently from additional regulatory proteins, we now enlarge the scale of the analysis to entire genomes.

#### **The discriminator is a primary location of promoter selectivity by DNA relaxation.**

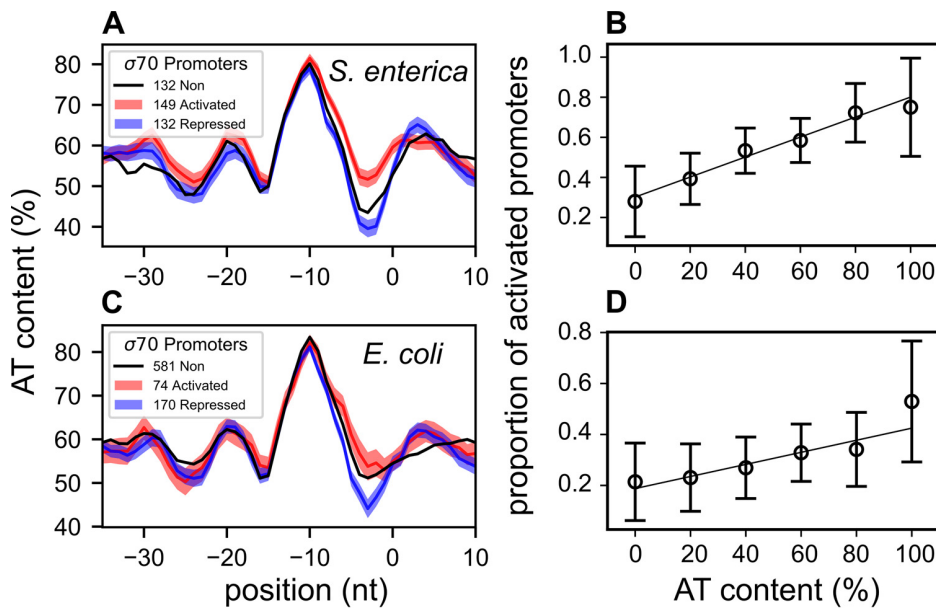
We first looked at the variability of discriminator G/C contents among mRNA promoters in various species based on available TSS maps (Fig. S3). These distributions are wide, and like *pheP* and *pelD*, a large class of promoters have G/C-rich discriminators. Based on the previous analysis, we hypothesized that such promoters would be more repressed by a DNA relaxation induced by gyrase inhibitors than those harboring an A/T-rich discriminator. However, in contrast to the mutation data described above, here the compared promoters differ by many additional factors beyond their discriminator sequence (upstream and downstream sequences, genomic context, binding of regulatory proteins, etc.), which may contribute to their supercoiling response; therefore, we looked for a statistical relation rather than a prediction valid for all analyzed promoters.

We aligned all  $\sigma^{70}$  promoters of *Salmonella enterica* and looked at their average A/T% profile (Fig. 3A) depending on their response 20 min after a novobiocin shock (19). Strikingly, although this content exhibits a characteristic nonuniform pattern along the promoter (with an expected peak at the  $-10$  element), the signals of the two groups of promoters are indistinguishable everywhere except in the region between  $-10$  and  $+1$ , precisely where we expected the observed difference ( $P < 10^{-5}$  around position  $-2$ ; Table S2). This observation, obtained independently from the mutation studies described above, confirms that the discriminator region is a primary location of selectivity for the relaxation response. As a comparison, no significant difference is detected at the  $-10$  element, suggesting that this selectivity is not related to a difference in sigma factor usage. Further, classifying the promoters based on their discriminator sequence composition (Fig. 3B) exhibits a clear and highly significant (approximately linear) effect on the proportion of activated promoters (correlation  $P < 10^{-4}$ ).

#### **A robust relation observed across phylogenetically distant bacterial phyla.**

Since the investigated mechanism relies on highly conserved molecular actors, RNAP and topoisomerases, it might affect a broad range of bacterial species. We therefore tested the validity of our observations in all organisms where a transcriptome obtained after DNA relaxation was available together with an accurate TSS map (from independent studies). The list of references of the employed data is summarized in Table S2, and the table of detailed promoter sequences is in Table S3.

Transcriptomic data were obtained in *E. coli* with DNA microarrays after norfloxacin shock in two alternate topoisomerase mutant strains (40), resulting in a strong magnitude of DNA relaxation (17). In spite of strong differences in the experimental protocol compared to the *S. enterica* data set, the obtained pattern is remarkably similar (Fig. 3C and D). Importantly, whereas in the first experiment (treated versus nontreated cells) this pattern might include contributions from SC-independent drug response pathways, here the two compared samples received exactly the same treatment, and any

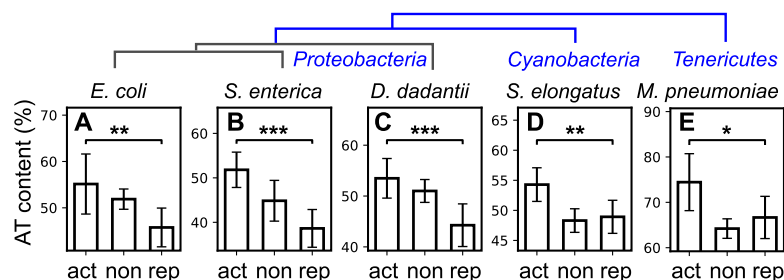


**FIG 3** Genome-wide relation between discriminator sequence and promoter selectivity during DNA relaxation. (A) Average A/T% profiles of *S. enterica*  $\sigma^{70}$  promoters along 5-nt-centered windows, depending on their response to novobiocin-induced chromosomal relaxation (19) (activated, significantly upregulated promoters; repressed, significantly downregulated promoters; non, not significantly affected). Colored-shaded areas represent  $\pm$  one standard error (67% confidence intervals). The profiles are very similar except in the discriminator region (between  $-10$  and  $+1$  positions). (B) Proportion of activated promoters among those responsive to DNA relaxation, depending on their A/T% in a 5-nt window centered around position  $-2$ . Error bars represent 95% confidence intervals. The resulting linear regression is highly significant ( $P < 10^{-4}$ ). (C) Average A/T% profiles of *E. coli*  $\sigma^{70}$  promoters depending on their response to norfloxacin-induced DNA relaxation (LZ54 versus LZ41 strains [17]). The resulting pattern is very similar to that observed in *S. enterica* in spite of strong differences in protocols. (D) Same as panel B but for the *E. coli* data ( $P = 0.011$ ).

such unwanted contribution should not be apparent. The slightly weaker observed effect might also be due to the lower sensitivity of the employed transcriptomic technology.

In *D. dadantii*, the response to relaxation by novobiocin was monitored in minimal medium (25) based on identified gene promoters (41). It exhibits the same pattern (Fig. 4C, more details are given in Fig. S4) as in *E. coli* (Fig. 4A) and *S. enterica* (Fig. 4B), suggesting that the investigated mechanism is valid for a broad range of enterobacteria of diverse lifestyles. Note that in Fig. 4 and later figures, genes not significantly affected by DNA relaxation were shown for qualitative comparison purpose but are heterogeneous among data sets and should not be used for rigorous statistical comparisons (heterogeneous and unknown false-negative rates).

Data were also available for two species of drastically larger evolutionary distance, the cyanobacterium *Synechococcus elongatus* and the small tenericute *Mycoplasma pneumoniae*. In these species, because the sigma factors differ from those of enterobacteria, the alignment of promoter elements was obtained with a poorer definition (promoters aligned at the TSS; see Materials and Methods). We nevertheless looked for sequence signatures comparable to those observed previously. In *Synechococcus elongatus*, where SC was shown to be a major determinant of circadian oscillatory genomic expression (42), the transcriptomic response to DNA relaxation was not monitored directly, but the phasing of gene expression in this oscillation can be used as an indirect proxy of this response (42), although many other metabolic signals may be equally correlated and could contribute to this signal. As a result of the analysis, a similar difference of discriminator sequence was detected as in enterobacteria (Fig. 4D) of slightly lower magnitude and at a position slightly shifted after the TSS (Fig. S4), possibly due to the poorer resolution of the analysis and the additional regulatory mechanisms



**FIG 4** Robust statistical relation between discriminator A/T% and promoter's response to DNA relaxation (act, activated; non, no significant variation; rep, repressed) is observed in phylogenetically distant bacterial species: (A) *E. coli* ( $P = 0.010$ , relaxation by norfloxacin in LZ54 versus LZ41 mutant strains) (17); (B) *S. Typhimurium* ( $P < 10^{-5}$ ); (C) *D. dadantii* ( $P < 10^{-3}$ ); (D) *S. elongatus* ( $P = 0.004$ ); and (E) *M. pneumoniae* ( $P = 0.029$ ). In enterobacteria, only  $\sigma^{70}$  promoters were considered and were aligned at the  $-10$  element. In the two other species, all promoters were aligned at their annotated TSS, resulting in a poorer definition of the signal and positional shifts. A/T% are computed in a 5-nt window centered around position  $-2$  in the discriminator region, except for *S. elongatus* (position  $+4$  after the TSS). Error bars represent 95% confidence intervals, and stars indicate the level of statistical significance (see Materials and Methods). A schematic phylogeny is depicted above.

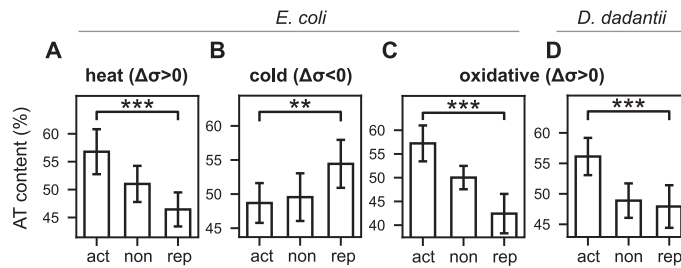
involved. In the small tenericute *Mycoplasm pneumoniae*, in which transcriptional regulation is poorly understood due to the quasi-absence of TFs (43), the response to novobiocin was also monitored (3). Although the signal is also weakened by the spatial resolution and by the lower number of promoters, it is still significant at the same location in the discriminator as in enterobacteria (Fig. 4E).

Altogether, the same signature is robustly and consistently observed in available data sets obtained after DNA relaxation in enterobacteria, and, with limitations due to the available definition of promoters and heterogeneity of the analyzed data, in two phylogenetically distant species that differ widely from the others in terms of lifestyle and average G/C content (in particular, *M. pneumoniae* has very few promoters with strongly G/C-rich discriminators; Fig. S3). These results suggest that the ancestral infrastructural constraint of DNA opening, coupled with the conserved activity of topoisomerases, indeed underpins a global regulatory mechanism throughout the prokaryotic kingdom.

#### Global response to stress conditions and inheritable supercoiling variations.

While sublethal antibiotic shocks are the classical method of choice to specifically induce rapid DNA relaxation (9, 44), under natural conditions the latter is rather triggered by sudden changes of environmental conditions, especially by physicochemical stress factors like temperature, acidity, oxidative agents, etc. The resulting rapid SC variations were found to be conserved even in phylogenetically distant species, e.g., increase of negative SC by cold shock, DNA relaxation by heat shock, or oxidative stress (9). We therefore tested if the sequence signature expected from the analysis described above could be detected in published transcriptomic data, although other stress-specific pathways contribute to the response and might hide this signature. Such data were obtained under various conditions (9); in the following, we focus our analysis on temperature and oxidative stress, where (i) the associated SC variations are well documented; (ii) there is no indication of ppGpp induction (see Discussion); and (iii) under each condition, two independent data sets were available and gave similar results.

Heat and cold shocks both put the bacteria under stress while affecting the SC level in opposite directions (relaxation and overtwisting, respectively; Table S2). The analysis of the corresponding transcriptomic data sets (45, 46) clearly confirms the expectations, with G/C-rich discriminators being repressed and activated with a linear dependence in the sequence content (Fig. 5A to C; see also the spatial patterns in Fig. S4). Similar signals were observed based on independent data sets obtained under the same conditions (47 and data not shown). In the case of oxidative stress (induced by  $H_2O_2$ ) associated with DNA relaxation, the response was analyzed in the enterobacteria *E. coli* and *D. dadantii* (18, 47), where the pattern is indeed very similar and matches



**FIG 5** Relation between discriminator sequence and response to SC variations induced by environmental stress conditions (act, activated; non, no significant variation; rep, repressed). A/T% are computed in a 5-nt window centered around position  $-2$  in the discriminator region. (A) During heat shock in *E. coli* (45), triggering a transient DNA relaxation ( $\Delta\sigma > 0$ ), activated promoters have discriminators with higher A/T% than repressed ones ( $P < 10^{-5}$ ), as expected from the presented model. (B) In a cold shock in *E. coli* (46) inducing an opposite SC variation (increase in negative SC,  $\Delta\sigma < 0$ ), the relation is reversed, with a preference of G/C-rich discriminators among activated promoters ( $P = 0.007$ ), as we expected. (C and D) Same as panels A and B but during an oxidative shock in *E. coli* (47) inducing DNA relaxation ( $\sigma > 0$ ,  $P < 10^{-7}$ ) (C), and in *D. dadantii* ( $P < 10^{-4}$ ) (D) (18), where the shock was shown to induce the same SC response (34), showing the conservation of the mechanism. Error bars represent 95% confidence intervals, and stars indicate the level of statistical significance (see Materials and Methods).

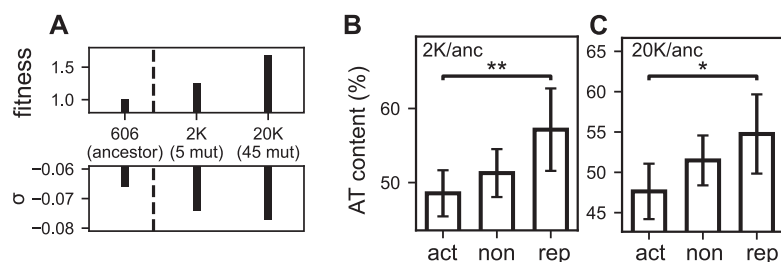
the expectations. Altogether, this analysis suggests that, beyond stress-specific regulation pathways mediated by dedicated regulatory proteins, the SC variations induced under these conditions play a direct role in the resulting global reprogramming of gene expression by modulating the RNAP-promoter interaction through the discriminator sequence. Under other stress conditions (osmotic or acidic stress) that we analyzed, the signal was species or data set dependent (data not shown), suggesting that other regulation mechanisms play a stronger role.

Finally, we address the question of whether the investigated mechanism is involved not only in transient responses but also in inheritable modifications of the expression program. In the longest-running evolution experiment with *E. coli* (48), point mutations inducing variations of the SC level were indeed quickly and naturally selected (49), as they provided substantial fitness gains that were attributed to the resulting global change of the transcriptional landscape (25). In the investigated conditions of growth in nutrient-poor medium, a first mutation (in *topA*, among 6 in total) before 2,000 generations and a second mutation (in *fis*, among 45 in total) before 20,000 generations both lead to an inheritable increase of negative SC (Fig. 6A). Based on the modeling, these mutations should predominantly enhance the expression of promoters with G/C-rich discriminators in the evolved strains. Such a tendency is indeed observed in both available transcriptomes that we analyzed, obtained either after 2,000 generations, where the signal is strongest (Fig. 6B) ( $P = 0.005$ ), or after 20,000 generations ( $P = 0.011$ ) (Fig. 6C, and Table S2), where 43 accumulated mutations besides these two affecting SC probably contribute to rewiring the regulatory network and blurring the signal. The detected signature suggests that the proposed biophysical regulatory mechanism not only is involved in rapid changes of gene expression but also may be used as a driving force in the evolution of genomes.

## DISCUSSION

In this work, we propose a simple thermodynamic model of open complex formation that quantitatively accounts for transcriptional regulation by SC based on the discriminator sequence. Our analysis confirmed and gave a quantitative content to the long-proposed notion that the discriminator sequence is a significant determinant of promoter supercoiling sensitivity. The statistical analysis of promoter sequences, carried out in various species and experimental conditions, highlights the widespread relevance of this mechanism in the genome-wide response to transient or inheritable variations of SC levels.





**FIG 6** (A) In the longest-running evolution experiment (48), two point mutations naturally acquired by *E. coli* (49) induced successive increases of negative SC, one in *topA* before 2,000 generations (among the five observed) and one in *fis* before 20,000 generations (among the 45 observed), and are associated with fitness gains through modifications of global gene expression. Adapted from reference 25. (B) Proportion of A/T content in the discriminator (5-nt window centered around position  $-2$ ) of promoters activated, repressed, or not significantly affected in the evolved strain 2K compared to the ancestor. As expected from our modeling for an increase in negative SC, activated promoters with G/C-rich discriminators are more activated ( $P = 0.005$ ). (C) In the 20K evolved strain, the same difference is observed ( $P = 0.011$ ), although less significant, possibly due to many other mutations affecting the regulatory network. Error bars represent 95% confidence intervals, and stars indicate the level of statistical significance (see Materials and Methods).

Interestingly, a global analysis of  $\sigma^{70}$ -dependent promoter sequences in *E. coli* yields a significant negative statistical relation between the A/T% at the discriminator and the RNAP binding score at the  $-10$  element (as computed from its sequence motif, Pearson's  $R = -0.18$ ,  $P < 10^{-16}$ ), suggesting that intrinsically attractive promoters have higher G/C-rich discriminators and, thus, are more difficult to open. This observation suggests that open complex formation is used as a general regulation mechanism for highly expressed operons, as occurs in rRNA promoters (7) (although a high affinity at the  $-10$  element does not imply a high expression level). However, we did not observe any A/T% difference at the  $-10$  element between promoters activated and repressed by SC (Fig. 3A and C), suggesting that high RNAP affinity and SC-mediated regulation are independent. While this study is focused on the specific role of SC, the general relation between RNAP affinity and the discriminator sequence might also involve other regulation mechanisms (including ppGpp; see below).

**Quantification and limitations of the regulatory mechanism.** A major difficulty when analyzing SC-induced regulation is that it affects the transcription process at multiple steps from the binding of regulators to the activity of RNAP itself during transcription initiation (10), elongation, and termination (22). While we focused our analysis on the discriminator sequence, the reader should keep in mind that many other mechanisms contribute to enhancing the complexity of this regulation: (i) the influence of DNA conformation on its interaction with regulatory proteins (9); (ii) competing structural transitions (denaturation, cruciform exclusion, G-quadruplex, and Z-DNA) occurring in nearby regions depending on the SC level and strongly affecting the SC response at the initiation site (50); (iii) the modulation of the effective SC level available for denaturation because of twist/writhe dynamics and local mechanical constraints imposed by regulatory proteins (9); and (iv) the heterogeneity of SC levels in different topological domains along the chromosome (51), in contrast to the approximation of a homogeneous level considered in this study. In particular, this heterogeneity was shown to depend on the local orientational organization of the genome because of the dynamic production of supercoils by elongating RNAPs. A recently proposed model of this mechanism, complementary to this study, explains a significant contribution to the transcriptional response to DNA relaxation even when all promoters are assumed to respond identically to SC variations (25). Therefore, integrating these two complementary factors of complexity, orientation-dependent heterogeneity of SC levels and sequence-dependent heterogeneity of promoter response, into a unified model is a natural objective for future studies.

These various complexity factors and others explain why, in the analyzed transcriptional data, the effect of the discriminator sequence emerges as a statistical feature at

the genomic scale rather than a predictive signal dictating the response of each individual promoter as observed in mutation studies. In particular, since a negative SC level favors the denaturation of G/C-rich as well as A/T-rich sequences (Fig. 1), this mechanism alone is insufficient to explain the existence of a class of relaxation-activated promoters, such as *gyrA-gyrB* (13). This behavior might be explained by more complex mechanisms involving the kinetics of promoter opening and escape by RNAP, where the stability of the open complex becomes unfavorable if it leads to abortive rather than processive transcription (10, 26), by thermodynamic competition with other structural transitions occurring at nearby sites (50), or by the effect of SC on the binding of transcription factors that are sensitive to the DNA tridimensional conformation (indirect readout) (15).

In spite of these limitations of our modeling, and based on the sequence signal observed in transcriptomic data, can we quantify the contribution of this specific mechanism in the genome-wide supercoiling response? To estimate this magnitude, we developed a genome-wide prediction of the relaxation-response based solely on the thermodynamic opening model developed above (independently from all other transcriptional effects of SC) and computed the proportion of accurate predictions among the observed differentially expressed genes (activation or repression). Compared to a null (random) model, this proportion is improved by around 10 to 15% of the responsive genes in the investigated relaxation and environmental stress assays (usually several hundred, representing a high statistical significance of predictive power; see details in Table S2 in the supplemental material). Considering the many alternate regulatory mechanisms by SC, for which no comparable estimates are available at the genomic scale (most of them lacking quantitative models), this proportion computed from a single step without parameter adjustment is quite notable. Additionally, it is likely underestimated because of many inaccurately annotated promoters (a single-nucleotide resolution is required but often not achieved) and may be reevaluated in the future based on more precise annotations. Note that because the total mRNA levels are normalized in transcriptomic data (predefined sequencing depth, erasing any global activation/repression effect), we introduced a comparable normalization step in the computation. As a result, a fraction of A/T-rich promoters appear to be activated by the DNA relaxation even if they are more difficult to open by RNAP (by competition with G/C-rich ones; see Fig. S6 and Materials and Methods).

**Simultaneous regulation by SC and ppGpp at the discriminator.** Among various further regulatory mechanisms related to this study, the alarmone ppGpp, classically associated with the stringent (starvation) response (6), deserves special attention. In contrast to many TFs, ppGpp affects the expression of a large subset of the genome by binding RNAP in combination with the transcription factor DksA (52) and modulating the stability of the open complex (29). Its repressive effect is not dependent on a strict sequence motif but rather on the presence of a C nucleotide at position  $-1$  (52). This regulatory mechanism presents many similarities to the one investigated here, and both are involved in the regulation of bacterial growth, raising the possibility of interplay between these two pathways (29, 53).

We first checked that the sequence signatures identified in this study were not due to a regulatory effect involving ppGpp rather than SC. It was observed that gyrase inhibition does not trigger any growth arrest (Fig. 2B) or signature of stringent response (17); accordingly, an analysis of the expression levels of genes involved in ppGpp synthesis (*gppA*, *spoT*, and *relA*) does not exhibit any significant response (3, 16–19). Thus, DNA relaxation does not trigger ppGpp production, and even if the two pathways are associated with a similar sequence signal in the discriminator, the observations made in this study are indeed due to a ppGpp-independent effect of SC.

We then carried out a sequence analysis of the promoters directly regulated by ppGpp through its binding to RNAP, as identified at the genomic scale in a recent study in *E. coli* (52). As expected, a strong difference in G/C% between the many promoters activated and repressed by ppGpp induction (representing 70% of  $\sigma^{70}$  promoters in total) is detected in the discriminator (Fig. S5A), similar to the pattern observed with



DNA relaxation (Fig. 3), confirming that the two pathways affect transcription at the whole-genome scale based on similar promoter sequence determinants.

While DNA relaxation does not induce ppGpp production, it was conversely shown that the induction of high levels of ppGpp by the stringent response does trigger a sharp fall in SC levels in *E. coli* (29). Thus, it is plausible that the strong sequence signature observed after ppGpp induction (Fig. S5A) actually results from the addition of two independent factors of open complex destabilization: RNAP binding by ppGpp and DNA relaxation. Interestingly, the transcriptional response to ppGpp induction was also monitored in mutant cells where it is unable to bind RNAP, inhibiting its direct regulatory activity (52). Remarkably, almost half as many genes respond as in the wild-type cells (representing 35% of  $\sigma^{70}$  promoters, although with weaker magnitudes and slightly slower response times), and these promoters exhibit a similar (albeit weaker) sequence signature at the same location (Fig. S5B). A plausible explanation is that ppGpp induction indeed triggered DNA relaxation (29), resulting in a similar but partial response compared to that of wild-type cells. This scenario remains hypothetical, as the SC levels were not directly measured in these samples; it would likely involve a posttranscriptional effect of ppGpp on gyrase activity, as frequently occurs in response to stress or metabolic signals (9). This analysis also suggests a specific effect of ppGpp for the activation of promoters with A/T%-rich discriminators (54) (compare the non and activated curves in Fig. S5); this observation might be linked to the weak difference between these two groups in several data sets involving DNA relaxation (e.g., Fig. 4A and C), although the opposite is seen in other cases (e.g., Fig. 4B and 5A and C).

Altogether, this combined analysis of transcriptomic data fully confirms the notion that the regulation by SC relaxation and ppGpp is partially redundant in their transcriptional effect but distinct; as an example, the SC dependence of *hisR* was found to be independent of *relA* in *S. enterica* (55). More precisely, SC relaxation may be considered a more fundamental form of regulation relying on the basic infrastructure of transcription, whereas ppGpp synthesis may itself trigger DNA relaxation (but not conversely). The relationship between the two pathways is further emphasized by the observation that, in the evolution experiment, the two genes most quickly and robustly affected by mutations are *topA* and *spoT* (49, 56), involved precisely in SC and ppGpp synthesis/degradation (6), respectively. Interestingly, the *spoT* mutation alone explains only a part of the observed transcriptional change (57), while similarly, the *topA* mutation alone generates only a fraction of the observed signal at the discriminator (data not shown), suggesting a synergistic action of these two mutations (49, 56). The additive selection of promoters based on the same sequence signal at the discriminator provides a plausible and natural mechanistic explanation for this feature.

Finally, in the data sets obtained with environmental stress conditions that we have analyzed (Fig. 5), the genes associated with ppGpp synthesis are partly responsive but rather in an opposite direction to the discriminator sequence signature observed (repression in heat and oxidative stress, slight activation in cold stress), and this pathway does probably not contribute significantly to the observed signal.

## MATERIALS AND METHODS

**Synthetic promoters.** Sequences 230, 329, and 313 nt upstream of the *pheP*, *peIE*, and *peID* start codons, respectively, were synthesized with mutations in the discriminator (GeneCust) and individually cloned into pUCTer-*luc* plasmids (see Table S1 in the supplemental material) upstream of a luciferase reporter gene (*luc*). *E. coli* strain MG1655 cells were then transformed with these plasmids using a standard electroporation procedure.

**Measurement of DNA relaxation response of mutant promoters *in vivo*.** *E. coli* cells carrying the plasmids with the different promoters were recovered from glycerol stock ( $-80^{\circ}\text{C}$ ) and grown overnight (about 16 h) on LB agar plates at  $37^{\circ}\text{C}$ . The obtained colonies were further transferred to liquid cultures overnight (about 16 h), with shaking at 200 rpm under selective antibiotic pressure (ampicillin at  $60\ \mu\text{g}/\text{ml}$  final concentration). LB medium was used for bacteria carrying plasmids with *pheP*-derived promoters, whereas M63 minimal medium supplemented with 0.2% glucose was used for bacteria carrying plasmids with *peIE*- and *peID*-derived promoters. Cells were washed ( $2\times$  centrifugation at 8,000 rpm), and then outgrowth cultures were performed in the same medium without antibiotics, stopped during exponential phase, and diluted for a final optical density at 600 nm ( $\text{OD}_{600}$ ) of 0.1 in a 96-well microplate. Each well

(200  $\mu$ l final volume) contained the chosen medium supplemented with D-luciferin (450  $\mu$ g/ml final). The microplate was placed in a humidity cassette and grown at 37°C until stationary phase was reached in a microplate reader (Tecan Spark). The OD<sub>600</sub> and luminescence were measured every 5 min, preceded by a 45-s shaking step (double orbital, 3.5-mm amplitude). During mid-exponential phase for *pheP* and early exponential phase for *pelE* and *pelD*, the microplate was taken out and DNA relaxation was transiently induced by injecting 5  $\mu$ l of novobiocin (50, 100, 150, and 200  $\mu$ g/ml final concentrations tested) using a multichannel pipette. Data files produced by the microplate reader were parsed using a Python home-made script, and the response to DNA relaxation was computed by comparing the luminescence values (in triplicates) of the novobiocin-shocked strain compared to the same strain injected with water (novobiocin solvent) 60 min after shock. The employed firefly luciferase has a short lifetime, between 6 min in *B. subtilis* (58) and 45 min in *E. coli* (59). Confidence intervals and *P* values were computed using Student statistics.

**Genome-wide analyses of discriminator sequences.** Transcriptomes obtained after DNA relaxation by antibiotics, inheritable supercoiling variations, or environmental stresses were collected from the literature, as were genome-wide TSS maps (Table S2). A scan for promoter motifs was conducted with bTSSfinder (60), imposing each TSS position at the experimentally determined nucleotide. Tables of detailed promoter sequences are provided in Table S3. For *E. coli*, the analysis was also tested with an alternate list of promoters (from the EcoCyc database [1]), which gave comparable results. In all expression data sets, genes were considered significantly activated/repressed under a common standard statistical selection procedure, based on a threshold of 0.05 on the adjusted *P* value, except for the evolution data (0.3; due to the otherwise low number of responsive genes; see details in reference 25). Promoters controlling several genes (operons) were considered differentially expressed if at least one gene of these genes is differentially expressed. For three data sets (heat and cold shock and *S. elongatus*), *P* values were not provided and were replaced by a threshold on log fold change values ( $\pm 0.5$ ), generating subsets of act/rep genes of sizes comparable to those in other data sets. For enterobacteria, only  $\sigma^{70}$ -dependent promoters were retained and aligned at their  $-10$  site to reduce statistical noise. Some of them also bind other  $\sigma$  factors, but  $\sigma^{70}$  is predominant in exponential phase where the analyzed samples were collected. The A/T% content was computed along 5-bp sliding windows (Fig. 3A and C). Promoters were classified according to their A/T% in a 5-nt window centered around position  $-2$  rather than the entire discriminator (of variable size), which improves the statistical analysis while not affecting the distribution of promoters significantly (Fig. S3). For *S. elongatus* and *M. pneumoniae*, where the sigma factors differ from those of enterobacteria, all promoters were retained and aligned at their TSS. As expected due to the variable size of the discriminator, the resulting A/T% signal had a poorer signal definition (Fig. S4) and exhibited small positional shifts. For *S. elongatus*, the A/T% difference was observed slightly downstream of the TSS, and we used position  $+4$  for the analysis. For *M. pneumoniae*, the A/T% peak was observed at position  $-6$ , and all positions were shifted by  $-4$  nt to impose it at the  $-10$  position. The relation between A/T% content and expression response was quantified either by linear regression (Fig. 3) or by a  $\chi^2$  test between activated and repressed promoters (Fig. 4 and 6). All error bars shown are 95% confidence intervals, except the colored areas of Fig. 3 (67% confidence intervals). In all figures, statistical significance is illustrated based on the *P* value (\*\*\*,  $P < 0.001$ ; \*\*,  $0.001 < P < 0.01$ ; \*,  $0.01 < P < 0.05$ ). Curves of Fig. 3 (A/T% profiles of promoters, linear regression) are provided in Fig. S4 for the other data sets.

**Model of transcriptional regulation by SC.** The observed correlation between promoter opening thermodynamics and expression strength (Fig. 1) is accounted for by a thermodynamic regulatory model (61):

$$k(\sigma, s) = k_0 \exp\left(\min\left(\frac{\Delta G(\sigma, s)}{k_B T}, 0\right)\right)$$

where  $k$  is the transcription rate,  $k_0$  is the basal (maximal) rate,  $s$  is the precise 14-nt sequence of the denatured region in the open complex (62), and  $k_B T$  is the Boltzmann factor. The free energy,  $\Delta G$ , is composed of two contributions, the opening penalty,  $\Delta G_{op}(\sigma, s)$  (Fig. 1B), and an additional contribution representing the opening assistance by RNAP,  $\Delta G_p^0(s)$ :

$$\Delta G(\sigma, s) = \Delta G_{op}(\sigma, s) + \Delta G_p^0(s)$$

The opening energy,  $\Delta G_{op}$ , is computed from an established coarse-grained unidimensional description of DNA twist-dependent thermodynamics (63), where the total SC level is assumed to contribute to DNA opening by RNAP (neglecting any effect of its partitioning into twist/writhe and constrained/unconstrained contributions in the thermodynamic equilibrium of open complex formation). We hypothesize that  $\Delta G_p^0(s)$  depends on the discriminator sequence, in agreement with direct measurements (32) and with the observation that the TSS position can be shifted by mutations in the discriminator (Fig. 1A) but is not affected by SC variations. At high negative SC levels, the opening penalty becomes negligible [ $\Delta G_{op}(\sigma, s) + \Delta G_p^0(s) > 0$ ] (Fig. 1B) and the maximal rate,  $k_0$ , is achieved, whereas the promoter is mostly closed when DNA is strongly relaxed.

Based on these hypotheses, the expression fold change of a promoter during an SC variation (in the regimen where it is not fully activated) depends only on and is independent of the precise (usually unknown) value of  $\Delta G_p^0(s)$ :

$$\Delta G(\sigma_0 + \Delta\sigma, s) - \Delta G(\sigma_0, s) = \Delta G_{op}(\sigma_0 + \Delta\sigma, s) - \Delta G_{op}(\sigma_0, s)$$

For the modeling of the data in Fig. 1, where absolute levels of expression (and not just fold changes) were measured, we used a single fitted value,  $\Delta G_p^0 = 3.5 k_B T \simeq 2$  kcal/mol, to avoid

overparameterization. This approximation may explain the slight discrepancies with the data (Fig. 1F and G), but the overall agreement suggests that the sequence-dependent variations of  $\Delta G_p^0(s)$  remain limited in the framework of our analysis. All following computations (for all promoters and species) were carried out with the same value of  $\Delta G_p^0$ , but since they involve expression fold changes (rather than absolute levels), the value of  $\Delta G_p^0(s)$  for each promoter has a marginal effect on the predictions.

For each promoter, the denaturation energy is computed with TwistDNA (63) using the 14-bp sequence starting from (and including) the  $-10$  hexamer, corresponding to the extent of the transcription bubble (flanked by 100-bp-long G-tracts to avoid boundary effects in the computation). The only adjustable parameter of TwistDNA is an effective salt concentration, which is calibrated on the data of Fig. 1 (21), yielding values of 1.5 mM and 3 mM for *in vitro* and *in vivo* transcription, respectively, the latter value being kept constant for all subsequent *in vivo* calculations. These low values are likely due to the strongly simplified description of the solvent (continuous distribution of monovalent ions) and DNA (unidimensional molecule) used in that software and should be considered effective parameters for the computation rather than quantitative concentrations.

Under all aforementioned approximations, it is possible to predict the quantitative regulatory effect of SC variations from their experimentally available genome-averaged value (e.g., using chloroquine-agarose gels). The validity of the computation is justified *a posteriori* by the good agreement with *in vitro* and *in vivo* expression data (Fig. 1 and 2). Note that, at the genomic scale, the SC level locally available to RNAP for the opening of a given promoter may deviate from the genome-averaged SC level because of many complicating factors beyond the simple model considered here (three-dimensional conformation of the promoter, binding of regulatory proteins and nucleoid-associated proteins, structural transitions occurring at nearby sites, etc.; see Discussion). However, because of the monotonous nature of the activation curves (Fig. 1B and E), all main results are robust when the SC levels are globally shifted by up to  $\pm 0.01$ .

**Superhelical densities.** *In vivo* SC levels used in the computations of Fig. 1 were taken from references 17, 21 (*E. coli* strains with norfloxacin), and 33 (topoisomerase mutants of *E. coli*).

Expression fold changes in response to relaxation measured in microplates with *pheP*-, *pelE*-, and *pelD*-derived promoters were reproduced (Fig. 2) with a relaxation magnitude,  $\Delta\sigma = 0.001$ , starting from a level of  $\sigma = -0.032$  in LB rich medium and  $\sigma = -0.023$  for M63+G minimal medium. This low magnitude may be partly due to the slow growth conditions in microplates but mostly to a buffering effect of the reporter system (luciferase lifetime of several to tens of minutes) and should be considered an effective value used in the modeling, as also suggested by the low repressive effect of novobiocin compared to batch cultures (34).

For the computation of the genome-wide contribution to the relaxation response (see Discussion), transcription rates from all promoters are normalized by their sum under each condition before computing fold changes, without any cutoff value (consistent with transcriptomic analysis protocols). This procedure results in the activation of a fraction of promoters (since the G/C-rich promoters represent a weaker proportion of total transcripts after the relaxation, A/T-rich promoters appear activated; see Fig. S6). Levels of SC variations associated with all investigated conditions were reviewed in the literature (Table S2), exhibiting magnitudes in the range 0.01 to 0.015, with differences due to protocols in stress/shock conditions and chloroquine-agarose gel assays. To reduce the number of adjustable parameters (considering the heterogeneity of these data), all model predictions were computed with a single initial SC level,  $\sigma = -0.045$  (a realistic value yielding the best overall agreement with observations), and a variation of  $\Delta\sigma = \pm 0.015$  (depending on the sign of the experimental response). The model predictions change only marginally when these figures are changed by less than 0.01 in either direction.

**Data availability.** See Table S2 for data availability information.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**TEXT S1**, PDF file, 0.04 MB.

**FIG S1**, PDF file, 0.1 MB.

**FIG S2**, PDF file, 0.2 MB.

**FIG S3**, PDF file, 0.1 MB.

**FIG S4**, PDF file, 0.4 MB.

**FIG S5**, PDF file, 0.1 MB.

**FIG S6**, PDF file, 0.02 MB.

**TABLE S1**, PDF file, 0.04 MB.

**TABLE S2**, PDF file, 0.1 MB.

**TABLE S3**, XLSX file, 0.3 MB.

## ACKNOWLEDGMENTS

We thank Joanna Bonci and Nicolas Paulhan for experimental contributions, the whole CRP team for helpful discussions, Georgi Muskhelishvili and Ivan Junier for their critical reading of the manuscript, and some referees for sound criticisms that helped improving the manuscript.

We declare no conflicts of interest.

## REFERENCES

- Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, Latendresse M, Muñoz-Rascado L, Ong Q, Paley S, Peralta-Gil M, Subhraveti P, Velázquez-Ramírez DA, Weaver D, Collado-Vides J, Paulsen I, Karp PD. 2017. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res* 45:D543–D550. <https://doi.org/10.1093/nar/gkw1003>.
- Brinza L, Calevro F, Charles H. 2013. Genomic analysis of the regulatory elements and links with intrinsic DNA structural properties in the shrunken genome of *Buchnera*. *BMC Genomics* 14:73. <https://doi.org/10.1186/1471-2164-14-73>.
- Junier I, Unal EB, Yus E, Lloréns-Rico V, Serrano L. 2016. Insights into the mechanisms of basal coordination of transcription using a genome-reduced bacterium. *Cell Syst* 2:391–401. <https://doi.org/10.1016/j.cels.2016.04.015>.
- Feklistov A, Sharon BD, Darst SA, Gross CA. 2014. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* 68: 357–376. <https://doi.org/10.1146/annurev-micro-092412-155737>.
- Klump S, Hwa T. 2008. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc Natl Acad Sci U S A* 105:20245–20250. <https://doi.org/10.1073/pnas.0804953105>.
- Barker MM, Gaal T, Josaitis CA, Gourse RL. 2001. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. *J Mol Biol* 305:673–688. <https://doi.org/10.1006/jmbi.2000.4327>.
- Travers A, Muskhelishvili G. 2005. DNA supercoiling—a global transcriptional regulator for enterobacterial growth? *Nat Rev Microbiol* 3:157–169. <https://doi.org/10.1038/nrmicro1088>.
- Dorman CJ, Dorman MJ. 2016. DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. *Biophys Rev* 8: 89–100. <https://doi.org/10.1007/s12551-016-0238-2>.
- Martís B S, Forquet R, Reverchon S, Nasser W, Meyer S. 2019. DNA supercoiling: an ancestral regulator of gene expression in pathogenic bacteria? *Comput Struct Biotechnol J* 17:1047–1055. <https://doi.org/10.1016/j.csbj.2019.07.013>.
- Wood DC, Lebowitz J. 1984. Effect of supercoiling on the abortive initiation kinetics of the RNA-I promoter of ColE1 plasmid DNA. *J Biol Chem* 259:11184–11187. [https://doi.org/10.1016/S0021-9258\(18\)90844-6](https://doi.org/10.1016/S0021-9258(18)90844-6).
- Borowiec JA, Gralla JD. 1985. Supercoiling response of the lac ps promoter in vitro. *J Mol Biol* 184:587–598. [https://doi.org/10.1016/0022-2836\(85\)90305-5](https://doi.org/10.1016/0022-2836(85)90305-5).
- Jovanovich SB, Lebowitz J. 1987. Estimation of the effect of coumermycin A1 on *Salmonella typhimurium* promoters by using random operon fusions. *J Bacteriol* 169:4431–4435. <https://doi.org/10.1128/jb.169.10.4431-4435.1987>.
- Menzel R, Gellert M. 1987. Modulation of transcription by DNA supercoiling: a deletion analysis of the *Escherichia coli* gyrA and gyrB promoters. *Proc Natl Acad Sci U S A* 84:4185–4189. <https://doi.org/10.1073/pnas.84.12.4185>.
- Pruss GJ, Drlica K. 1989. DNA supercoiling and prokaryotic transcription. *Cell* 56:521–523. [https://doi.org/10.1016/0092-8674\(89\)90574-6](https://doi.org/10.1016/0092-8674(89)90574-6).
- Cevost J, Vaillant C, Meyer S, Rost B. 2018. ThredDNA: predicting DNA mechanics' contribution to sequence selectivity of proteins along whole genomes. *Bioinformatics* 34:609–616. <https://doi.org/10.1093/bioinformatics/btx634>.
- Peter BJ, Arsuaga J, Breier AM, Khodursky AB, Brown PO, Cozzarelli NR. 2004. Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biol* 5:R87. <https://doi.org/10.1186/gb-2004-5-11-r87>.
- Blot N, Mavathur R, Geertz M, Travers A, Muskhelishvili G. 2006. Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep* 7:710–715. <https://doi.org/10.1038/sj.embor.7400729>.
- Jiang X, Sobetzko P, Nasser W, Reverchon S, Muskhelishvili G. 2015. Chromosomal “stress-response” domains govern the spatiotemporal expression of the bacterial virulence program. *mBio* 6:e00353-15. <https://doi.org/10.1128/mBio.00353-15>.
- Gogoleva NE, Konnova TA, Balkin AS, Plotnikov AO, Gogolev YV. 2020. Transcriptomic data of *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. 140285 treated with novobiocin. *Data Brief* 29:105297. <https://doi.org/10.1016/j.dib.2020.105297>.
- Figueroa-Bossi N, Guérin M, Rahmouni R, Leng M, Bossi L. 1998. The supercoiling sensitivity of a bacterial tRNA promoter parallels its responsiveness to stringent control. *EMBO J* 17:2359–2367. <https://doi.org/10.1093/emboj/17.8.2359>.
- Auner H, Buckle M, Deufel A, Kutateladze T, Lazarus L, Mavathur R, Muskhelishvili G, Pemberton I, Schneider R, Travers A. 2003. Mechanism of transcriptional activation by FIS: role of core promoter structure and DNA topology. *J Mol Biol* 331:331–344. [https://doi.org/10.1016/s0022-2836\(03\)00727-7](https://doi.org/10.1016/s0022-2836(03)00727-7).
- Dorman CJ. 2019. DNA supercoiling and transcription in bacteria: a two-way street. *BMC Mol Cell Biol* 20:26. <https://doi.org/10.1186/s12860-019-0211-6>.
- Liu LF, Wang JC. 1987. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A* 84:7024–7027. <https://doi.org/10.1073/pnas.84.20.7024>.
- Lilley DMJ, Chen D, Bowater RP. 1996. DNA supercoiling and transcription: topological coupling of promoters. *Q Rev Biophys* 29:203–225. <https://doi.org/10.1017/s0033583500005825>.
- El Houdaigui B, Forquet R, Hindré T, Schneider D, Nasser W, Reverchon S, Meyer S. 2019. Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Res* 47:5648–5657. <https://doi.org/10.1093/nar/gkz300>.
- Chen J, Chiu C, Gopalkrishnan S, Chen AY, Olinares PDB, Saecker RM, Winkelman JT, Maloney MF, Chait BT, Ross W, Gourse RL, Campbell EA, Darst SA. 2020. Stepwise promoter melting by bacterial RNA polymerase. *Mol Cell* 78:275–288. <https://doi.org/10.1016/j.molcel.2020.02.017>.
- Travers A. 1980. A tRNA<sup>Tyr</sup> promoter with an altered in vitro response to ppGpp. *J Mol Biol* 141:91–97. [https://doi.org/10.1016/s0022-2836\(80\)80030-1](https://doi.org/10.1016/s0022-2836(80)80030-1).
- Lamond AI. 1985. Supercoiling response of a bacterial tRNA gene. *EMBO J* 4:501–507. <https://doi.org/10.1002/j.1460-2075.1985.tb03656.x>.
- Ohlsen KL, Gralla JD. 1992. DNA melting within stable closed complexes at the *Escherichia coli* rrnB P1 promoter. *J Biol Chem* 267:19813–19818. [https://doi.org/10.1016/S0021-9258\(19\)88626-X](https://doi.org/10.1016/S0021-9258(19)88626-X).
- Schneider R, Travers A, Muskhelishvili G. 2000. The expression of the *Escherichia coli* fis gene is strongly dependent on the superhelical density of DNA. *Mol Microbiol* 38:167–175. <https://doi.org/10.1046/j.1365-2958.2000.02129.x>.
- Travers A. 1980. Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. *J Bacteriol* 141:973–976. <https://doi.org/10.1128/jb.141.2.973-976.1980>.
- Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, Gourse RL. 2006. rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell* 125:1069–1082. <https://doi.org/10.1016/j.cell.2006.04.034>.
- Mogil LS, Becker NA, Maher LJ. 2016. Supercoiling effects on short-range DNA looping in *E. coli*. *PLoS One* 11:e0165306. <https://doi.org/10.1371/journal.pone.0165306>.
- Ouafa ZA, Reverchon S, Lautier T, Muskhelishvili G, Nasser W. 2012. The nucleoid-associated proteins H-NS and FIS modulate the DNA supercoiling response of the pel genes, the major virulence factors in the plant pathogen bacterium *Dickeya dadantii*. *Nucleic Acids Res* 40:4306–4319. <https://doi.org/10.1093/nar/gks014>.
- Bellon S, Parsons JD, Wei Y, Hayakawa K, Swenson LL, Charifson PS, Lippke JA, Aldape R, Gross CH. 2004. Crystal structures of *Escherichia coli* topoisomerase IV ParE subunit (24 and 43 kilodaltons): a single residue dictates differences in novobiocin potency against topoisomerase IV and DNA gyrase. *Antimicrob Agents Chemother* 48:1856–1864. <https://doi.org/10.1128/AAC.48.5.1856-1864.2004>.
- Hsieh LS, Burger RM, Drlica K. 1991. Bacterial DNA supercoiling and [ATP]/[ADP]. Changes associated with a transition to anaerobic growth. *J Mol Biol* 219:443–450. [https://doi.org/10.1016/0022-2836\(91\)90185-9](https://doi.org/10.1016/0022-2836(91)90185-9).
- Cameron ADS, Stoebel DM, Dorman CJ. 2011. DNA supercoiling is differentially regulated by environmental factors and FIS in *Escherichia coli* and *Salmonella enterica*. *Mol Microbiol* 80:85–101. <https://doi.org/10.1111/j.1365-2958.2011.07560.x>.
- Drlica K, Snyder M. 1978. Superhelical *Escherichia coli* DNA: relaxation by coumermycin. *J Mol Biol* 120:145–154. [https://doi.org/10.1016/0022-2836\(78\)90061-x](https://doi.org/10.1016/0022-2836(78)90061-x).
- Dorman CJ, Barr GC, Ni Bhriain N, Higgins CF. 1988. DNA supercoiling and the anaerobic and growth phase regulation of tonB gene expression. *J Bacteriol* 170:2816–2826. <https://doi.org/10.1128/jb.170.6.2816-2826.1988>.
- Zechiedrich EL, Khodursky AB, Bachellier S, Schneider R, Chen D, Lilley DM, Cozzarelli NR. 2000. Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. *J Biol Chem* 275:8103–8113. <https://doi.org/10.1074/jbc.275.11.8103>.



41. Forquet R, Jiang X, Nasser W, Hommais F, Reverchon S, Meyer S. 2020. Mapping the complex transcriptional landscape of the phytopathogenic bacterium *Dickeya dadantii*. bioRxiv <https://doi.org/10.1101/2020.09.30.320440>.
42. Vijayan V, Zuzow R, O'Shea EK. 2009. Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proc Natl Acad Sci U S A* 106: 22564–22568. <https://doi.org/10.1073/pnas.0912673106>.
43. Yus E, Lloréns-Rico V, Martínez S, Gallo C, Eilers H, Blötz C, Stülke J, Lluich-Senar M, Serrano L. 2019. Determination of the gene regulatory network of a genome-reduced bacterium highlights alternative regulation independent of transcription factors. *Cell Syst* 9:143–158. <https://doi.org/10.1016/j.cels.2019.07.001>.
44. Drlica K. 1992. Control of bacterial DNA supercoiling. *Mol Microbiol* 6: 425–433. <https://doi.org/10.1111/j.1365-2958.1992.tb01486.x>.
45. Bartholomäus A, Fedyunin I, Feist P, Sin C, Zhang G, Valleriani A, Ignatova Z. 2016. Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philos Trans Ser A Math Phys Eng Sci* 374: 20150069. <https://doi.org/10.1098/rsta.2015.0069>.
46. Zhang Y, Burkhardt DH, Rouskin S, Li GW, Weissman JS, Gross CA. 2018. A stress response that monitors and regulates mRNA structure is central to cold shock adaptation. *Mol Cell* 70:274–286. <https://doi.org/10.1016/j.molcel.2018.02.035>.
47. Jozefczuk S, Klie S, Catchpole G, Szymanski J, Cuadros-Inostroza A, Steinhauser D, Selbig J, Willmitzer L. 2010. Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol Syst Biol* 6:364. <https://doi.org/10.1038/msb.2010.18>.
48. Tenaillon O, Barrick JE, Ribick N, Deatherage DE, Blanchard JL, Dasgupta A, Wu GC, Wielgoss S, Cruveiller S, Médigue C, Schneider D, Lenski RE. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* 536:165–170. <https://doi.org/10.1038/nature18959>.
49. Crozat E, Philippe N, Lenski RE, Geiselman J, Schneider D. 2005. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169:523–532. <https://doi.org/10.1534/genetics.104.035717>.
50. Zhabinskaya D, Benham CJ. 2012. Theoretical analysis of competing conformational transitions in superhelical DNA. *PLoS Comput Biol* 8: e1002484. <https://doi.org/10.1371/journal.pcbi.1002484>.
51. Lal A, Dhar A, Trostel A, Kouzine F, Seshasayee ASN, Adhya S. 2016. Genome scale patterns of supercoiling in a bacterial chromosome. *Nat Commun* 7:11055. <https://doi.org/10.1038/ncomms11055>.
52. Sanchez-Vazquez P, Dewey CN, Kitten N, Ross W, Gourse RL. 2019. Genome-wide effects on *Escherichia coli* transcription from ppGpp binding to its two sites on RNA polymerase. *Proc Natl Acad Sci U S A* 116: 8310–8319. <https://doi.org/10.1073/pnas.1819682116>.
53. Travers A, Muskhelishvili G. 2020. Chromosomal organization and regulation of genetic function in *Escherichia coli* integrates the DNA analog and digital information. *EcoSal Plus*. <https://doi.org/10.1128/ecosalplus.ESP-0016-2019>.
54. Gummesson B, Lovmar M, Nyström T. 2013. A proximal promoter element required for positive transcriptional control by guanosine tetraphosphate and DksA protein during the stringent response. *J Biol Chem* 288: 21055–21064. <https://doi.org/10.1074/jbc.M113.479998>.
55. O'Byrne CP, Ní Bhriain N, Dorman CJ. 1992. The DNA supercoiling-sensitive expression of the *Salmonella typhimurium* his operon requires the His attenuator and is modulated by anaerobiosis and by osmolarity. *Mol Microbiol* 6: 2467–2476. <https://doi.org/10.1111/j.1365-2958.1992.tb01423.x>.
56. Philippe N, Crozat E, Lenski RE, Schneider D. 2007. Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. *Bioessays* 29:846–860. <https://doi.org/10.1002/bies.20629>.
57. Cooper TF, Rozen DE, Lenski RE. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc Natl Acad Sci U S A* 100:1072–1077. <https://doi.org/10.1073/pnas.0334340100>.
58. Botella E, Noone D, Salzberg LI, Hokamp K, Devine SK, Fogg M, Wilkinson AJ, Devine KM. 2012. Chapter 1. High-resolution temporal analysis of global promoter activity in *Bacillus subtilis*, p 1–26. In Harwood C, Wipat A (ed), *Methods in microbiology*, vol 39. Academic Press, New York, NY. <https://doi.org/10.1016/B978-0-08-099387-4.00001-6>.
59. Thompson JF, Hayes LS, Lloyd DB. 1991. Modulation of firefly luciferase stability and impact on studies of gene regulation. *Gene* 103:171–177. [https://doi.org/10.1016/0378-1119\(91\)90270-1](https://doi.org/10.1016/0378-1119(91)90270-1).
60. Shahmuradov IA, Mohamad Razali R, Bougouffa S, Radovanovic A, Bajic VB. 2017. bTSSfinder: a novel tool for the prediction of promoters in cyanobacteria and *Escherichia coli*. *Bioinformatics* 33:334–340. <https://doi.org/10.1093/bioinformatics/btw629>.
61. Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Kuhlman T, Phillips R. 2005. Transcriptional regulation by the numbers: applications. *Curr Opin Genet Dev* 15:125–135. <https://doi.org/10.1016/j.gde.2005.02.006>.
62. Lee J, Borukhov S. 2016. Bacterial RNA polymerase-DNA interaction—the driving force of gene expression and the target for drug action. *Front Mol Biosci* 3:73. <https://doi.org/10.3389/fmolb.2016.00073>.
63. Jost D. 2013. Twist-DNA: computing base-pair and bubble opening probabilities in genomic superhelical DNA. *Bioinformatics* 29:2479–2481. <https://doi.org/10.1093/bioinformatics/btt415>.

## Chapter 5

# Quantitative contribution of the spacer length in the supercoiling-sensitivity of bacterial promoters

As stated above, SC affects transcription at multiple steps of the process beyond open-complex formation. Obtaining a complete understanding of gene regulation by SC thus requires to decipher all the mechanisms by which it modulates transcription. In this Chapter, we focus on another factor which was soon discovered to contribute to the SC response of genes, involving the orientation of -35 and -10 RNAP binding sites for closed-complex formation, depending on promoter spacer length and SC level.

### 5.1 Insights into seconeolitsine-induced Topoisomerase I inhibition in Gram-negative bacteria

The model is described extensively in the upcoming article, wherein a regulatory thermodynamic model of transcription is developed based on the step of closed-complex formation, and similarly to previous Chapter, validated at a global scale based on a combination between transcriptomic data and transcription assays conducted on mutant promoters. However, in contrast to the previous model, here, based on recent work to which I contributed [121], we use the non-marketed antibiotic seconeolitsine to study the response of mutant promoters to DNA overtwisting, in complement to novobiocin-induced DNA relaxation.

In order to characterise the effect of SC on gene expression, existing studies essentially employ well-established drugs that target DNA gyrase such as novobiocin, providing the transcriptomic response to SC variations only in one direction, that of DNA relaxation [108]. On the other hand, drugs that inhibit TopoI have never been used in Gram-negative bacteria, making the transcriptomic response to DNA overtwisting lacking. Among the possible TopoI inhibitors, one of them named seconeolitsine [145] has been tested in the Gram-positive bacterium *Streptococcus pneumoniae*, where it induces a transient SC increase at low concentrations, and, in turn, global expression changes [146]. We consequently investigated the effects of seconeolitsine on bacterial growth, chromosomal SC level and transcription in *D. dadantii*, where it shows an equal effectiveness as in *S. pneumoniae*, providing the first insights into TopoI inhibition and SC increase in Gram-negative bacteria. Among the key findings, by comparing the transcriptomic responses to DNA relaxation and overtwisting obtained with the same method (RNA-seq), it reassesses the concept of promoter "supercoiling-sensitivity" as the large majority of genes responding to one of the drugs does not respond to the other [121]. Instead of exhibiting (or not) an intrinsic property of SC-sensitivity, the response of promoters to SC variations may rather vary depending on a variety of factors, including the genomic context and the conditions of the experiment (*in vitro* vs *in vivo*, ionic composition, basal SC level related to metabolic state, etc.) [121].



I thus took advantage of this recent work to validate the predictions of the second model on mutant promoters, based on opposite SC variations (Fig. 5.1). This is presented in the following article, which is planned to be submitted to Nucleic Acids Research journal.

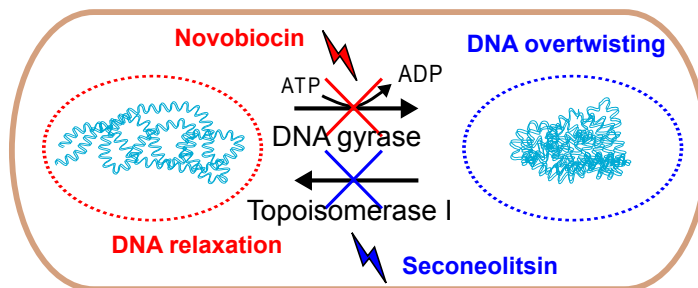


FIGURE 5.1: Induction of opposite SC variations, using the gyrase inhibitor novobiocin for DNA relaxation, and the TopoI inhibitor seconeolitsine for DNA overtwisting.

## 5.2 Article

# Quantitative contribution of the spacer length in the supercoiling-sensitivity of bacterial promoters

Raphaël Forquet, William Nasser, Sylvie Reverchon and Sam Meyer\*

Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, CNRS, UMR5240 MAP, F-69622, France

\* To whom correspondence should be addressed. Email: sam.meyer@insa-lyon.fr.

October 21, 2021

## 1 Abstract

## 2 Introduction

3 DNA supercoiling (SC), the level of over- or underwind- 36  
 4 ing of the double-helix, is a fundamental property of 37  
 5 DNA. In bacterial cells, the chromosome is maintained 38  
 6 at a negative SC level, *i.e.* in an underwound state, by 39  
 7 a finely controlled balance between the DNA relaxation 40  
 8 activity of topoisomerase I (and IV, to a lesser extent), 41  
 9 and the introduction of negative supercoils by DNA gy- 42  
 10 rase [1, 2]. This level is affected by a variety of factors, 43  
 11 including growth phase and environmental stresses [3]. 44  
 12 In addition to playing a key role in genome organisa- 45  
 13 tion [4], SC was soon discovered to affect the expression 46  
 14 of many promoters [5, 6]. More recently, through the 47  
 15 advent of transcriptomic technologies allowing genome- 48  
 16 wide expression profiling, it was shown to act as a global 49  
 17 transcriptional regulator in many bacteria, based on stud- 50  
 18 ies employing gyrase inhibitors inducing a global DNA 51  
 19 relaxation and, in turn, a global and complex response of 52  
 20 genes [3]. Fast changes in SC levels may thus play an 53  
 21 important and global role in the transcriptional response 54  
 22 of bacteria to environmental changes [1, 2]. 55

23 At the mechanistic level, SC affects the transcription 56  
 24 process at multiple steps, both indirectly through regula- 57  
 25 tory proteins, and directly by modulating the interaction 58  
 26 of RNA Polymerase (RNAP) with DNA [3]. This in- 59  
 27 cludes open-complex formation during transcription ini- 60  
 28 tiation, which is strongly facilitated by negative SC [7, 8] 61  
 29 and plays a key role in the SC response of promoters [9]. 62  
 30 But other steps of transcription also contribute to this re- 63  
 31 sponse, including closed-complex formation [10], pro- 64  
 32 moter escape [11], transcription elongation and termina- 65  
 33 tion [12]. Because of this complexity, regulatory models 66  
 34 able to predict the response of a given promoter to SC 67  
 35 variations, either quantitatively or even qualitatively, are 68  
 69  
 70

essentially lacking [3, 13], in contrast to those involv- 36  
 ing regulatory proteins [14]. Considering the widespread 37  
 relevance of this mechanism in bacterial gene expression, 38  
 obtaining such quantitative models is an important objec- 39  
 tive, both for fundamental understanding of the process 40  
 and for applications, *e.g.*, in synthetic biology [15]. The 41  
 objective of this paper is to develop such a model, fo- 42  
 cusing on the specific step of closed-complex formation 43  
 during transcription initiation. 44

The latter step involves the binding of RNAP to genes' 45  
 promoters, through specific recognition of -35 and - 46  
 10 sites by 2.4 and 4.2 regions of  $\sigma$  factors, respec- 47  
 tively [16]. The efficiency of RNAP binding on the pro- 48  
 moter depends primarily on the proximity of -35 and - 49  
 10 sequences to their respective consensus, but also on 50  
 the spacer element between them. The latter exhibits lit- 51  
 tle or no contact with the transcription machinery, except 52  
 for promoters with an extended -10 element which inter- 53  
 acts with the 3.0 region of  $\sigma$  factors [16]. Since the - 54  
 35 element is not systematically well-defined, especially 55  
 in promoters where the binding of RNAP is assisted by 56  
 regulatory proteins such as CRP [17], and even lacking 57  
 in some of them, in which case its absence is compen- 58  
 sated by an extended -10 element [18], the spacer may 59  
 be undefined for some promoters. Nevertheless, for the 60  
 others, it displays a variable length, ranging generally 61  
 from 15 to 19 nucleotides (nt) for  $\sigma$ 70-dependent pro- 62  
 moters [16, 19]. The maximal core promoter activity 63  
 is reached with spacers of 17 nt, while the addition or 64  
 subtraction of nucleotides from this optimal length re- 65  
 duces their expression by several-fold [20, 21]. While the 66  
 spacer sequence exhibits no specific requirement, punc- 67  
 tual mutations [22, 23] or modifications of its AT rich- 68  
 ness [24, 25] also affect promoter expression [15, 26], 69  
 presumably by altering its 3D conformation [27, 28, 29]. 70

In addition to altering promoter strength, the spacer 71  
 length has long been shown to strongly modulate their 72

73 SC response [20, 21]. Accordingly, most promoters of  
 74 stable RNAs in *E. coli*, strongly repressed by DNA re-  
 75 laxation, were found to contain spacers of unusual length  
 76 (16 nt) [10]. A qualitative model of this regulation mode  
 77 was proposed around 30 years ago, based on a simple ge-  
 78 ometric effect involving the helical nature of DNA [10].  
 79 Because of the latter, variable spacer lengths are associ-  
 80 ated to different relative orientations of the -35/-10 bind-  
 81 ing sites, which could be an obstacle for RNAP bind-  
 82 ing. In turn, the presence of torsional stress in the spacer  
 83 could rotate the -35/-10 binding sites toward a favourable  
 84 (or unfavourable) orientation and thus regulate the RNAP  
 85 activity (Fig. 1A), and this notion was supported by a re-  
 86 view of qualitative observations on a collection of pro-  
 87 moters [10, 30, 15]. As an additional supporting evi-  
 88 dence, the twisting of sub-optimal spacers to realign -  
 89 35/-10 sites for RNAP binding has been proposed as the  
 90 mechanistic basis of several transcription factors (TFs),  
 91 including MerR, which activates promoters with a 19-  
 92 nt spacer [31]. Yet, while the latter only occurs at spe-  
 93 cific promoters containing a MerR binding site, the ac-  
 94 tion of SC might affect the entire chromosome. Ac-  
 95 cording to this mechanism, the presence of suboptimal  
 96 spacer lengths among bacterial promoters does not lead  
 97 to a static modulation of their expression, but it allows the  
 98 cell to selectively down- or up-regulate them depending  
 99 on their length, by tuning the global SC level.

100 In this study, we propose to (1) translate the pro-  
 101 posed qualitative notion into a quantitative regulatory  
 102 model, and (2) develop a rigorous validation of its ef-  
 103 fect, in particular by analysing high-throughput expres-  
 104 sion data, which were not available when the previous  
 105 qualitative models were developed but are particularly  
 106 suited to this global regulation mode. We first present the  
 107 model, which mostly relies on knowledge-based physi-  
 108 cal parameters of DNA, and provides a range of quan-  
 109 titative and parameter-free predictions regarding the ef-  
 110 fect of superhelical variations on promoters with various  
 111 spacer lengths. Based on this framework, we re-analyse  
 112 previous quantitative *in vitro* expression data, and de-  
 113 velop *in vivo* measurements in *E. coli* based on mutant  
 114 promoters harbouring different spacer lengths, which to-  
 115 gether allow a rigorous validation of the model predic-  
 116 tions and define its application range. We then assess  
 117 the range of validity of our predictions through a statis-  
 118 tical analysis of genome-wide transcriptional responses  
 119 to SC variations, either transiently induced by gyrase-  
 120 inhibiting antibiotics, or inherited over generations in the  
 121 longest-running evolution experiment. Our results con-  
 122 sistent show that the variability of spacer lengths un-  
 123 derpins a global selectivity of promoter activity depend-

ing on the cellular SC level. They confirm and provide a  
 quantitative framework to the proposed basal regulatory  
 mechanism, which likely plays a widespread role in the  
 prokaryotic kingdom.

## Materials and Methods

### Thermodynamic model of supercoiling-dependent transcription

We assume that the formation of the closed-complex is  
 limited by an intermediate state where the spacer DNA  
 must be (un)twisted to a favourable relative orientation of  
 the -10 and -35 sites, allowing RNAP binding (Fig. 1A).  
 This deformation is treated in the elastic approximation,  
 and the associated orientational free energy depends on  
 the spacer length  $n$  and average superhelical level  $\sigma$ :

$$\Delta G_{or}(\sigma, n) = \frac{n}{2} k_{\theta} \left( \frac{\theta_P}{n} - \alpha_0 (1 + \sigma) \right)^2 \quad (1)$$

where  $k_{\theta} = 71.4 k_B T \text{ rad}^{-2}$  is the DNA sequence-  
 averaged twist stiffness [32, 33],  $k_B T$  is the Boltzmann  
 factor,  $\alpha_0 = 34^\circ$  is the average twist angle between  
 adjacent nucleotides [34, 35], and  $\theta_P$  is a global pa-  
 rameter representing the optimal twist angle between -  
 35 and -10 sites for RNAP binding. To illustrate the  
 model predictions (Fig. 1) and subsequent calculations,  
 we assumed that 17-nt spacers achieve this optimal an-  
 gle at the standard superhelical level  $\sigma = -0.06$ , *i.e.*  
 $\theta_P = 17 \times \alpha_0 \times (1 - 0.06) = 543^\circ$ , but most key predic-  
 tions are independent of its value (see next paragraphs).

In this calculation, we assume that the total superhelic-  
 ity  $\sigma$  can be utilised by RNAP for the twist deformation  
 of spacer DNA (although most of it is stored as writhe  
 at equilibrium). This unidimensional approximation of  
 the spacer DNA, already proposed [10], is based on the  
 observations that (1) the spacer DNA is too short to sig-  
 nificantly writhe (before RNAP binding), and (2) twist  
 deformations diffuse much faster than writhe [36], and  
 RNAP may thus use twist deformations from neighbour-  
 ing DNA as a reservoir for this step. Since RNAP is as-  
 sumed to impose a fixed angle  $\theta_P$ , its flexibility is ne-  
 glected in the calculation.

The total free energy associated to the transcription  
 process is assumed to contain three contributions related  
 to the effects of SC, of the spacer element, or of both  
 simultaneously:

$$\Delta G(\sigma, n, s) = \Delta G_{sc}(\sigma, s) + \Delta G_{sp}(n, s) + \Delta G_{or}(\sigma, n) \quad (2)$$

where

- $\Delta G_{or}(\sigma, n)$  is the orientational deformation energy (Eq. 1), and introduces a *coupled* dependence on  $\sigma$  and  $n$ .
- $\Delta G_{sc}(\sigma, s)$  depends on the promoter sequence  $s$ , and represents all other mechanisms of regulation by SC (e.g., promoter opening, 3D deformations, structural transitions, etc.), which are generally impossible to predict and might be stronger than the investigated one, *but are assumed to be independent of spacer length  $n$* .
- $\Delta G_{sp}(n, s)$  represents all other mechanisms of modulation of transcriptional activity by the spacer DNA, which depend on its sequence as well as its length and are equally impossible to predict (e.g., stretching, 3D conformation, sequence-specific interactions with RNAP within the closed-complex, etc), *but are assumed to be independent of  $\sigma$* .

Transcription rates are then computed using a standard thermodynamic framework [14]:

$$k(\sigma, n, s) = k_0 \exp\left(-\frac{\Delta G(\sigma, n, s)}{k_B T}\right) \quad (3)$$

where  $k$  is the transcription rate,  $k_0$  is the basal rate (which depends, e.g., on the -10/-35 sequence affinities for RNAP), and  $k_B T$  is the Boltzmann factor. To simplify the notations, we use the latter as the energy unit in the following. Based on the equations above, the model *does not predict* the general SC-dependence of a promoter *nor* the effect of the spacer on the absolute expression level (since both depend on unpredictable terms in Eq. 2), *but* it does predict how the SC-sensitivity depends on spacer length by Eq. 1. Since most of the parameters involved in the latter are known from physical measurements, it is possible to derive several quantitative and parameter-free predictions underpinning all analyses of the manuscript, as follows.

### Prediction of relative *in vitro* expression levels depending on spacer length

From *in vitro* expression data of mutant promoters (Fig. 2), we isolate the specific effect of the spacer length  $n$  by normalising each datapoint by the corresponding value obtained in the reference mutant promoter ( $n_0 = 17$ ), thereby eliminating any other regulatory effect of SC:

$$\log\left(\frac{k(\sigma, n, s)}{k(\sigma, n_0, s)}\right) = (\Delta G_{or}(\sigma, n_0) - \Delta G_{or}(\sigma, n)) + (\Delta G_{sp}(n_0, s) - \Delta G_{sp}(n, s)) \quad (4)$$

The second term is independent of  $\sigma$ , and is thus a constant for a promoter of given sequence and spacer size, hereafter quoted  $Q_{sp}(n, s)$ . Using a linear expansion in  $\Delta n/n_0 = (n - n_0)/n_0$  in the first term, the relative expression level of each spacer length simplifies to a linear dependence in  $\sigma$  (as visible in Fig. 2B without approximation):

$$\log\left(\frac{k(\sigma, n_0 + \Delta n, s)}{k(\sigma, n_0, s)}\right) \simeq Q_{sp}(n, s) - \alpha_0 k_\theta \Delta n \left[ \frac{\theta_P}{n} \left(1 - \frac{1}{2} \frac{\Delta n}{n}\right) - (1 + \sigma) \alpha_0 \right] \quad (5)$$

Crucially, the *slope* of each line in Fig. 2B is therefore defined without any adjustable parameter, and is proportional to the torsional stiffness of DNA and to  $\Delta n = n - 17$ . The intercept depends on the global parameter  $\theta_P$ , and may also depend on the spacer sequence and length due to  $Q_{sp}(n, s)$ .

### Prediction of *in vivo* expression fold-changes during superhelical variations

All analysed *in vivo* data involve relative expression levels (fold-changes) after vs before a global superhelical variation  $\sigma_0 \rightarrow \sigma_0 + \Delta\sigma$  (induced by antibiotics or mutations). The predicted value simplifies to:

$$\log\text{FC}(\Delta\sigma, \sigma_0, n, s) = -\frac{n}{2} k_\theta [\alpha_0^2 (\Delta\sigma^2 + 2\Delta\sigma(1 + \sigma_0))] - [k_\theta \theta_P \alpha_0 \Delta\sigma + R_{sc}(\Delta\sigma, \sigma_0, s)] \quad (6)$$

where  $R_{sc} = \Delta G_{sc}(\sigma_0 + \Delta\sigma, s) - \Delta G_{sc}(\sigma_0, s)$  reflects all spacer-independent regulatory effects of SC on the considered promoter  $s$ . Crucially, only the first term of the equation depends on the spacer length  $n$ , and it does not depend on any unknown parameter. Thus, after linear expansion in  $\Delta\sigma$ , the *relative* effect of the superhelical variation on spacer length mutants of the same promoter  $s$  is entirely predictable and independent of  $\sigma_0$  (which is satisfactory since the absolute SC levels are not always known with precision in the analysed *in vivo* data):

$$\log\text{FC}(\Delta\sigma, n_0 + \Delta n, s) - \log\text{FC}(\Delta\sigma, n_0, s) \simeq -k_\theta \alpha_0^2 \Delta n \Delta\sigma \simeq -25 \Delta n \Delta\sigma \quad (7)$$

This dependence is shown in Supplementary Fig. S2 for a DNA relaxation  $\Delta\sigma = 0.03$ , and yields an expression ratio of around 2 between spacers differing by one nucleotide. In transcriptomic analyses, promoters are grouped by their spacer length  $n$  but differ by their overall sequence  $s$ , so that the term  $R_{sc}$  remains in the

221 form of strong statistical noise (Fig. 4 and 5), imposing  
 222 to work with the proportion of activated promoters rather  
 223 than directly with the fold-change values.

### 224 **Measurement of mutant promoters' responses to** 225 **opposite supercoiling variations *in vivo***

226 230 bp sequences upstream of *pheP* start codon were syn-  
 227 thesised with mutations in the spacer (GeneCust), and in-  
 228 dividually cloned into pUCTer-luc plasmids upstream of  
 229 a luciferase reporter gene (*luc*) (Supplementary Tab. S1).  
 230 *E. coli* strain MG1655 cells were then transformed with  
 231 these plasmids using a standard electroporation proce-  
 232 dure. The following protocol is described elsewhere [9].  
 233 Briefly, *E. coli* cells carrying the plasmids with the dif-  
 234 ferent promoters were grown at 37°C in LB medium in  
 235 a microplate reader (Tecan Spark). The  $OD_{600nm}$  and  
 236 luminescence were measured every 5 minutes to follow  
 237 bacterial growth and promoter expression, respectively.  
 238 DNA relaxation was induced by injecting 5  $\mu$ L of novo-  
 239 biocin (50, 100, 150 and 200  $\mu$ g/mL final concentrations  
 240 tested), whereas DNA overtwisting was induced by in-  
 241 jecting 2  $\mu$ L of seconeolitsine (25, 50, 75 and 100  $\mu$ M  
 242 final concentration tested). The responses to such oppo-  
 243 site DNA SC variations were then computed by compar-  
 244 ing the luminescence values (in triplicates) of the novo-  
 245 biocin or seconeolitsine-shocked strain compared to the  
 246 same strain injected with water (novobiocin solvent) or  
 247 DMSO (seconeolitsine solvent), 60' or 5' after shock, re-  
 248 spectively. The employed firefly luciferase has a lifetime  
 249 of around 45' in *E. coli* [37], and buffers the repressive  
 250 effect of novobiocin. Confidence intervals and p-values  
 251 were computed using Student's distributions. Raw data-  
 252 points are provided in Supplementary Fig. S3. We pre-  
 253 viously showed [9] that the presence of the employed  
 254 plasmids does not affect bacterial growth, and that the  
 255 expression patterns and response to novobiocin-induced  
 256 relaxation are consistent for promoters inserted either in  
 257 plasmid-borne or in chromosomal luciferase fusions, in  
 258 agreement with other studies [38, 39]. The employed  
 259 plasmids are well established as reflecting the average SC  
 260 level of the chromosome [40], and also specifically in re-  
 261 sponse to novobiocin-induced relaxation [41, 42, 43].

### 262 **Parameter fitting in mutation data**

263 In Fig. 1B and 2B, for simplicity, the curves were drawn  
 264 based on the orientational contribution (Eq. 1) only, with  
 265 the value  $\theta_P = 543^\circ$ . In Fig. 2B, the datapoints of [44]  
 266 fall on the predicted line without adjustment, suggesting  
 267  $Q_{sp} = 0$  for these promoters (see Eq. 5). The data of [45]

268 were adjusted by a value  $Q_{sp}(n, s) = 1.1 k_B T$  (Eq. 5),  
 269 corresponding to a factor 3 in expression.

270 Expression fold-changes measured in microplates with  
 271 *pheP*-derived promoters were reproduced (Fig. 3), start-  
 272 ing from a level  $\sigma = -0.06$ , with an overtwisting mag-  
 273 nitude  $\Delta\sigma = -0.02$ , and a relaxation magnitude  $\Delta\sigma =$   
 274 0.005. This lower relaxation magnitude probably partly  
 275 reflects a buffering effect of the reporter system and  
 276 should thus be considered as an effective value used in  
 277 the modelling, as further suggested by the lower repres-  
 278 sive effect of novobiocin compared to batch cultures [46].  
 279 The spacer-independent effect of SC,  $R_{sc}$  (Eq. 6), was  
 280 estimated from the repression (novobiocin) and activa-  
 281 tion (seconeolitsine) level observed with the 17-nt spacer,  
 282 with values  $R_{sc} = 0.4 k_B T$  and  $R_{sc} = -0.97 k_B T$  cor-  
 283 responding to activation factors of 0.67 and 2.1, respec-  
 284 tively (Fig. 3, 17-nt spacers). These presumably reflect  
 285 the modulation of promoter opening energy by the two  
 286 opposite superhelical variations [9].

### 287 **Genome-wide analyses of spacer responses to su-** 288 **percoiling variations**

289 Transcriptomic responses to DNA relaxation or inher-  
 290 itable SC variations were collected from the literature  
 291 (Supplementary Tab. S2). The curated map of *E. coli*  
 292 promoters with associated genes and spacer lengths was  
 293 retrieved from Ecocyc [47]. Only  $\sigma 70$ -dependent pro-  
 294 moters were retained and classified depending on their  
 295 spacer length and response to the investigated condi-  
 296 tion, under standard statistical selection procedures (ad-  
 297 justed  $P$ -value  $< 0.05$ ). For transcriptomes of *E. coli*  
 298 evolved strains, a less stringent  $P$ -value threshold (0.1)  
 299 was applied to have enough statistical power for the anal-  
 300 ysis [48] (albeit with a higher proportion of false posi-  
 301 tives). For other species, TSS maps were retrieved from  
 302 the literature [49, 50, 51, 52], and the positions of pro-  
 303 moter elements were predicted using bTSSfinder [53];  
 304 details are given in Supplementary Information. The re-  
 305 lation between promoter activation and spacer length was  
 306 quantified either by linear regression (Fig. 4A) or by a  
 307 Student's t-test between activated and repressed promot-  
 308 ers (Fig. 4B, Fig. 5BC, Supplementary Fig. S4). All com-  
 309 putations were carried using a homemade Python pack-  
 310 age. All error bars shown are 95% confidence intervals.

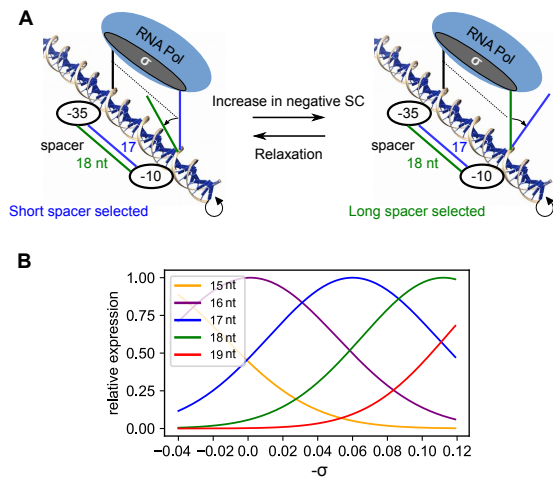


Figure 1: Quantitative modelling of the coupling between RNAP binding sites orientation, spacer length and DNA supercoiling. (A) Schematic depiction of -35/-10 alignment depending on spacer length and superhelical density. With long spacers (green), RNAP binding sites are out-of-phase at moderate SC levels, and become optimally aligned when DNA is highly negatively supercoiled, and conversely for shorter spacers (17-nt spacer in blue). (B) Relative expression levels predicted by the regulatory model for 15 to 19-nt spacers, depending on  $\sigma$ . The global parameter  $\theta_P$  was chosen such that -35 and -10 sites are optimally aligned at  $\sigma = -0.06$  for 17-nt spacers (see text). Short spacers are more expressed at relaxed levels, while long spacers are favoured at highly negative levels.

## Results

### Regulatory model of -35/-10 alignment during closed-complex formation

Following previous works [10, 30], we hypothesised that, for a simultaneous binding of the -35 and -10 hexamers by the RNAP holoenzyme during closed-complex formation, an intermediate state must be achieved where the spacer DNA is (un)twisted to a favourable orientation (Fig. 1A). We translated this notion into a quantitative regulatory model, based on a thermodynamic description of transcription [14] where this contribution can be computed from the torsional energy of spacer DNA, while RNAP is assumed to impose an optimal angle between the two sites. Treating DNA as a homogeneous polymer of known torsional stiffness, this energy depends on the spacer length and its superhelical state before RNAP binding, and the only adjustable parameter is the optimal angle of RNAP (Fig. 1A, detailed hypotheses of the model and equations are given in Materials and Methods). A difficulty in the analysis of this regulation mode is that SC affects transcriptional activity at

many other steps of the process (promoter opening, escape, elongation, etc.), with unpredictable rules but often stronger magnitude than the investigated effect (see below). The main assumption for a predictive modelling was to consider these other regulatory effects to be independent of the spacer length. Similarly, all other mechanisms by which the spacer length and sequence modulate transcriptional activity (e.g., DNA stretching, 3D deformations, specific interactions with RNAP within the closed complex, etc.) are assumed to be independent of the initial superhelical state of the promoter. Based on these simplifying assumptions, the modulation of the torsional angle between the -35 and -10 hexamers is the only mechanism of *coupled* dependence between spacer length and SC, and the quantitative contribution of the spacer length to the SC-sensitivity of the promoter can be computed without any adjustable parameter, as developed in the following.

The first prediction is the quantitative magnitude of this regulatory contribution. Geometrically, the size of the spacer (15-20 nt, in contrast to the shorter spacer of many dimeric TFs, 8-10 nt) implies that physiologically relevant superhelical variations (of around  $\pm 0.06$ ) have an orientational effect on the -10/-35 elements of the same magnitude as a variation in spacer length of one nucleotide ( $16 \times 0.06 \simeq 1$ ), and may therefore induce a quantitative regulatory effect. In turn, this effect depends on the mechanical cost of aligning an ill-oriented spacer (e.g., the 18-nt spacer in the left panel of Fig. 1A), which depends on the experimentally known value of the torsional stiffness of DNA [32]. Fig. 1B shows that, after calculation, the addition/removal of one nucleotide in the spacer corresponds to a significant regulatory factor of around 2 (e.g., compare the 18 vs 17 nt spacers at  $\sigma = -0.06$ ). While this value is milder than the "on/off" regulation of many TFs affecting specific promoters, it is sufficient for a biologically relevant effect, especially if this factor affects RNAP activity in a systematic manner. This magnitude also matches the *in vitro* observed effect of varying the spacer length (Supplementary Fig. S1), suggesting that, as hypothesised, the elasticity of DNA (rather than RNAP) is a relevant parameter in this process. But while the deformation energy has a symmetrical effect on the expression of spacers either too long or too short (same value of 16 and 18 nt spacers at  $\sigma = -0.06$ ), in contrast, SC is predicted to affect their expression in opposite directions, with short spacers being activated at relaxed levels (left) while long spacers are rather activated at highly negative levels (right).

## 381 *In vitro* validation of model predictions on mu- 382 tant promoters

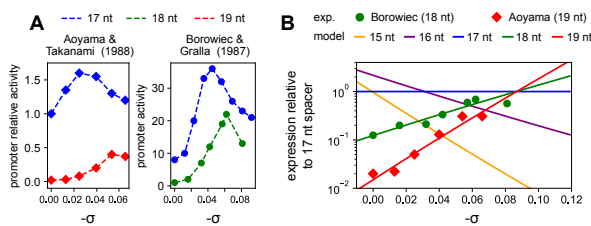


Figure 2: Comparison of *in vitro* transcription assays to model predictions. (A) Left panel: Relative activity of promoters with 17 and 19-nt spacers (blue and red, respectively), depending on SC level  $\sigma$  (data from [45]). Right panel: activity of promoters with 18-nt and 17-nt spacers (green and blue, respectively, data from [44]). (B) Transcription model predictions (solid lines) compared to the *in vitro* promoter expression data (exp.) from [45] (red dots) and [44] (green dots). Each datapoint was normalised by the corresponding value obtained in the reference mutant promoter with a 17-nt spacer, to isolate specifically the spacer length contribution to the SC response, among other possible regulatory effects (see text). The model intercept value for the 19-nt spacer was corrected by a factor 3 (see Materials and Methods).

383 The experimental validation of these predictions faces  
384 two obstacles: (1) varying SC might affect transcription  
385 through other (and possibly stronger) mechanisms, and  
386 (2) the same is true of varying the spacer length (or se-  
387 quence), both of which may hide the specific effect of  
388 the investigated mechanism. We therefore developed an  
389 analysis of previously obtained *in vitro* transcription data  
390 involving spacer length mutants of model promoters, in  
391 order to specifically distinguish this regulatory effect. In  
392 this experimental protocol, plasmid templates containing  
393 *lacP<sup>S</sup>*-derived promoters [44] and *PSC*-derived promoters  
394 [45] were prepared at well-defined superhelical densi-  
395 ties and incubated with RNAP, thus minimising the effect  
396 of external regulatory factors. Since the raw datapoints  
397 (Fig. 2A) contain all aforementioned regulatory contribu-  
398 tions (especially SC-assisted promoter opening, explain-  
399 ing the overall activation by negative SC [9]), we nor-  
400 malised each datapoint by the one obtained at the same  
401  $\sigma$ , with the same overall promoter sequence but the refer-  
402 ence 17-nt spacer length (Fig. 2B), thus eliminating these  
403 other factors and allowing a direct comparison with the  
404 model (by construction, the blue curve is flat; see detailed  
405 theoretical paragraph in Materials and Methods, Eq. 5).  
406 The prediction is a linear dependence of the datapoints  
407 (coloured lines), where the *slopes* are set by the DNA  
408 torsional stiffness value, without any adjusted param-  
409 eter, and proportional to  $(n - 17)$  (where  $n$  is the spacer

length). In particular, the slope of the 19-nt spacer line  
is twice that of the 18-nt one (both activated by negative  
is twice that of the 18-nt one (both activated by negative  
SC), and the same for 15- and 16-nt spacers with nega-  
tive values (repression by negative SC). Remarkably, the  
datapoints from two independent datasets obtained from  
different promoters (with 18- and 19-nt spacers) align  
precisely along the predicted slopes, giving a strong sup-  
port to our hypothesis that this regulatory contribution is  
dominated by the torsional elasticity of DNA. Further,  
in the study by Aoyama et al. [45], the same data were  
also collected with 16-nt spacers, but, in direct contrast  
to our predictions, these appear to be more activated at  
strong SC levels (Supplementary Fig. S5). Together with  
the previous ones, this observation suggests that our hy-  
potheses are valid for some spacer lengths, but fail for  
16-nt spacers. Possible reasons and biological implica-  
tions are developed in the Discussion, but briefly, this be-  
haviour of 16-nt spacers can be expected if they adopt a  
different conformation than 17-nt spacers in their inter-  
action with RNAP, and their differential SC-sensitivity is  
dominated not by the deformation of DNA before RNAP  
binding as we assume but, e.g., during open-complex for-  
mation. The question then arises, if this deviation is a  
feature of short spacers in general, thus strongly reducing  
the usefulness of the proposed model, or if it is a mere ex-  
ception of 16-nt spacers. We therefore ran additional *in*  
*vivo* experiments on mutant promoters containing oppo-  
site spacer lengths of 15 and 19 nucleotides.

## *In vivo* validation with opposite superhelical vari- ations

We constructed spacer length mutants of the *pheP* pro-  
moter of *E. coli*, of 17 nt (native), 15 nt (two dele-  
tions) and 19 nt (two insertions). This promoter is  
SC-sensitive [46] and not regulated by any identified  
TF [47], and is thus an interesting candidate for our reg-  
ulation mechanism based on the basal interaction with  
RNAP. Promoters were fused on plasmids in front of a  
luciferase reporter gene (Fig. 3A), and their expression  
was recorded in *E. coli* cells grown in LB rich medium  
in a microplate reader (Fig. 3B-C and Supplementary  
Fig. S3). Expression levels were computed shortly after  
treatment by the gyrase inhibitor novobiocin [3], or the  
topoisomerase I inhibitor seconeolitsine [54, 55], applied  
during exponential phase (see Materials and Methods).  
These drugs induce opposite SC variations in *E. coli*,  
DNA relaxation and overtwisting respectively [55], and  
thus provide complementary and independent tests of our  
predictions. Again, for a rigorous analysis of the data, we  
derived specific predictions on the effect of the analysed



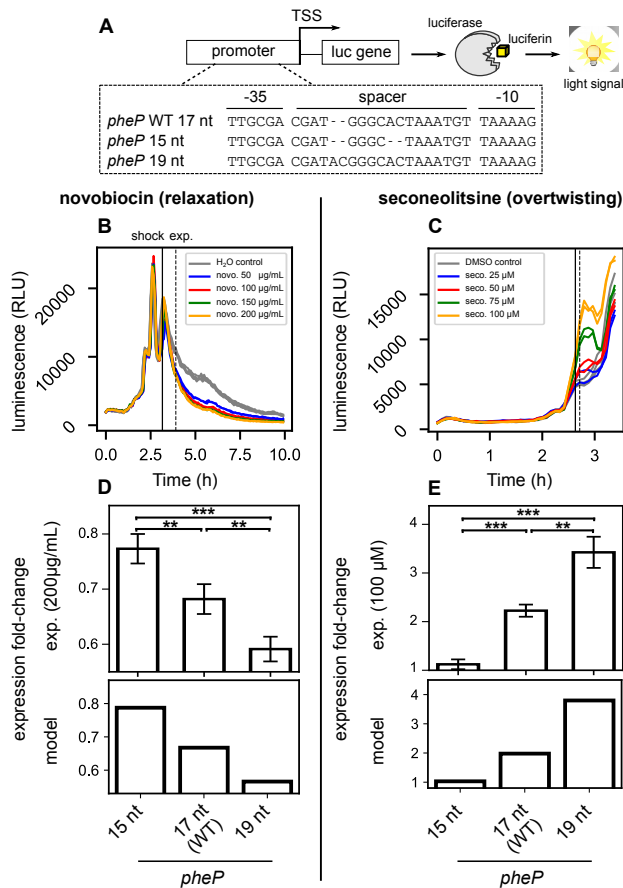


Figure 3: Responses of *pheP*-derived promoters with variable spacer length to opposite SC variations. (A) Promoter sequences were synthesised from *pheP* (*E. coli*), with mutated spacers of different lengths, controlling the expression of a *luc* gene encoding firefly luciferase, generating luminescence from luciferin substrate. (B) Promoter expression monitored in a microplate reader (bacteria carrying plasmids with *pheP* native promoter in LB medium), with a novobiocin shock applied in mid-exponential phase (time point quoted "shock") at different sublethal concentrations. (C) same as (B), with a seconeolitsine shock performed at different sublethal concentrations, which induces a stronger and more transient expression variation. (D-E) Expression fold-changes computed 60' (time point quoted "exp.") after novobiocin shock (200 µg/mL), or 5' after seconeolitsine shock (100 µM) from the experiments (upper panels), or predicted by the model (lower panels), assuming SC variations compatible with the observed expression variations levels (see Materials and Methods). See all raw data-points, including growth curves, in Supplementary Fig. S3.

459 global superhelical variations on promoters of different  
460 spacer lengths, without any parameter adjustment (see  
461 dedicated paragraph in Materials and Methods, Eq. 7).

462 The native *pheP* promoters was repressed by  
463 novobiocin-induced DNA relaxation (Fig. 3B), in agree-  
464 ment with previous data [46], and was activated by seco-

465 neolitsine (Fig. 3C); both effects increased with the ap-  
466 plied dosage. This result might be explained by the clas-  
467 sical effect of SC-assisted promoter opening [9], as in the  
468 *in vitro* data above; the relatively modest (but highly sig-  
469 nificant) repression levels in response to novobiocin are  
470 due to a buffering effect of the reporter system (see Mat-  
471 erials and Methods). Next, comparing the regulatory ef-  
472 fect of the shocks on the mutants vs native promoter, we  
473 observed that novobiocin represses the 19-nt-spacer pro-  
474 moter with much stronger magnitude, whereas the 15-  
475 nt promoter was almost insensitive to DNA relaxation  
476 (Fig. 3F), as expected. Conversely, under seconeolitsine-  
477 induced DNA overtwisting, the activation fold-change  
478 was strongest for the 19-nt promoter, and very low for the  
479 15-nt promoter (Fig. 3G). These four independent obser-  
480 vations (differential effect of either superhelical variation  
481 on a shorter or longer spacer) are fully consistent with  
482 model predictions, and the quantitative values were re-  
483 produced by fitting the magnitude of the effective super-  
484 helical variations induced by the drugs (Fig. 3D-E, one  
485 fitted parameter for each shock, see Materials and Meth-  
486 ods). These results confirm, after the *in vitro* data above,  
487 that 19-nt spacers are favoured by strong negative SC lev-  
488 els, and show that 15-nt spacers are rather favoured by  
489 DNA relaxation. The opposite behaviour of 16-nt spacers  
490 noted above thus seems to be an abnormal case where our  
491 hypotheses break down, and these promoters will there-  
492 fore be disregarded in the upcoming analyses (see Dis-  
493 cussion). For all other spacer lengths, since the proposed  
494 mechanism relies on the basal interaction between RNAP  
495 and promoter elements, we then wished to enlarge the  
496 analysis to the entire genome of *E. coli*, in order to test  
497 its validity and relevance at the global scale.

### 498 Global effect of the spacer in promoter SC- 499 sensitivity

500 We first looked at the variability of spacer lengths among  
501 *E. coli*  $\sigma 70$ -dependent promoters, based on an available  
502 curated promoter map [47]. We focused on  $\sigma 70$  since  
503 it is predominant in exponential phase where the anal-  
504 ysed samples were collected. The most frequent spacer  
505 length is 17-nt (27% total promoters), and most other are  
506 distributed between 15 and 19-nt (78% total promoters),  
507 with the remaining below 15-nt and above 19-nt (22% to-  
508 tal promoters) (Supplementary Fig. S6). In the analyses  
509 below, we focus on 15 to 19-nt spacers, which are likely  
510 less affected by annotation errors, and for which the mag-  
511 nitude of physiologically relevant SC variations are suf-  
512 ficient to induce a significant regulatory effect. Based  
513 on the analyses above, we hypothesised that promoters

with short spacers would be more activated by DNA relaxation than those harbouring long ones (Supplementary Fig. S2), and conversely for DNA overtwisting. However, in contrast to the mutation studies above conducted on the same promoters, genome-wide analyses involve the comparison of promoters that differ by many additional factors beyond their spacer, including different genomic contexts, surrounding sequences, bindings of transcriptional regulators, etc. We thus searched for a statistical relation between promoter selectivity during SC variations and spacer length, rather than a prediction valid for all analysed promoters (see Materials and Methods).

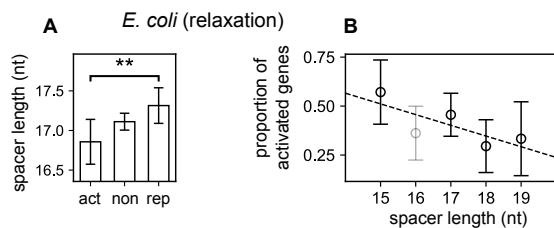


Figure 4: Genome-wide relation between spacer length and promoter selectivity during DNA relaxation in *E. coli*. **(A)** Mean comparison of spacer lengths for  $\sigma 70$ -dependent promoters activated (act), not significantly affected (non) or repressed (rep) by norfloxacin-induced DNA relaxation (LZ54 vs LZ41 strains [46]). As expected from our modelling for a DNA relaxation, activated promoters have significantly shorter spacers compared to repressed ones ( $P = 0.007$ ). **(B)** Proportion of activated promoters among those responsive to DNA relaxation, depending on their spacer length (linear regression  $P = 0.07$ ). As observed from *in vitro* data, 16-nt spacers (in grey) do not follow the model and were excluded from statistical analyses (see Discussion).

In *E. coli*, the transcriptomic response to DNA relaxation was obtained with DNA microarrays [46], after a norfloxacin shock in two alternate topoisomerase mutant strains [56]. The analysis shows that promoters activated by DNA relaxation indeed harbour significantly shorter spacers than repressed ones (Fig. 4A,  $P = 0.007$ ), and correspondingly, classifying the promoters based on their spacer length (Fig. 4B) exhibits a clear decreasing tendency (correlation  $P = 0.07$ ). The relatively high level of noise is due to the heterogeneity of promoter sequences within each group, and also likely to a fraction of inaccurately annotated promoters, since a single-nucleotide resolution in the definition of -10 and -35 hexamers is required for an accurate analysis, but not always achieved. In particular, as stated in the Introduction, the -35 element is not systematically well-defined, and even lacking in some promoters [18]. As expected from the observations above, the 16-nt spacers again deviate from the model predictions and were excluded from statisti-

cal analyses (see Discussion for a functional analysis of these promoters). For all others, these results demonstrate that the variability of spacer length is used by bacterial cells for the global selectivity of promoters in response to DNA relaxation.

### Analysis of transcriptomic data in various species

Since the investigated mechanism relies on highly conserved molecular actors, RNAP and topoisomerases, it might affect a particularly broad range of bacterial species, and we therefore wished to extend this analysis to other organisms. But while the transcriptomic response to DNA relaxation induced by gyrase inhibitors has been recorded in several species [3] (Supplementary Tab. S2), curated and accurate promoter maps are generally lacking. We therefore based our analysis on available maps of transcription start sites (TSS) obtained from specifically designed transcriptomic data (Supplementary Tab. S2), followed by a scan for promoter motifs [53]. We thus obtained a list of putative promoters with associated  $\sigma$  factors and associated spacer lengths for two other enterobacteria, *Salmonella enterica* and the phytopathogen *Dickeya dadantii*, and at a drastically larger evolutionary distance, for the cyanobacterium *Synechococcus elongatus* and the small tenericute *Mycoplasma pneumoniae*. However, we noted that promoter prediction programs perform poorly in the detection of -35 elements: using the *E. coli* promoter map as a benchmark dataset, we found that the predicted -35 position deviated from the annotated one in around 50% promoters. In other species, this inaccuracy presumably resulted in a much higher level of statistical noise than for the annotated *E. coli* promoters above. In spite of this obstacle, a difference of spacer length in the same direction as in *E. coli* was observed in all investigated species (Supplementary Fig. S4), albeit with weaker magnitudes and levels of statistical significance. Altogether, while improvements in promoter definition are clearly required for a solid conclusion, this systematic observation suggests that the variability of spacer length might indeed underpin a selective activation and repression of promoters by global SC variations throughout the prokaryotic kingdom.

### Inheritable selection of promoters based on the spacer length

We finally investigated if the present mechanism could be involved not only in transient DNA relaxation re-

592 sponses induced by antibiotic shocks, but also in inheri-  
 593 table variations of global gene expression in the longest-  
 594 running evolution experiment [57, 58]. Indeed, in this  
 595 experiment involving the growth of *E. coli* cells in a  
 596 daily refreshed minimal medium, point mutations affect-  
 597 ing the SC level were quickly and naturally selected, as  
 598 they provided substantial fitness gains [57] (Fig. 5A). A  
 599 first mutation (in *topA*) was fixed before 2,000 genera-  
 600 tions, and a second mutation (in *fis*) before 20,000 gen-  
 601 erations, both leading to an inheritable increase of nega-  
 602 tive SC (Fig. 5A). Based on our modelling and the previ-  
 603 ous observations, we therefore expected promoters with  
 604 a long spacer to experience enhanced expression in the  
 605 evolved strains compared to the ancestor. Such a rela-  
 606 tion is indeed observed, both after 2,000 generations  
 607 (Fig. 5B,  $P = 0.04$ ) and 20,000 generations (Fig. 5C,  
 608  $P = 0.026$ ). The signal is significant but slightly weaker  
 609 than that observed with antibiotics (Fig. 4); this may be  
 610 explained by the inheritable (rather than transient) nature  
 611 of the SC variation, which induces an adaptive response  
 612 of the cells via other regulatory pathways. Again, these  
 613 results suggest that promoters of different spacer lengths  
 614 respond differently to SC variations due to, e.g., muta-  
 615 tions in topoisomerase genes that are observed even be-  
 616 tween closely related species [59].

## 617 Discussion

618 While the spacer length and sequence are known to mod-  
 619 ulate transcriptional activity, we wished to quantitatively  
 620 model and test the long-proposed idea [44, 45, 10] that  
 621 SC plays a specific role in this process through a sim-  
 622 ple and basal orientational effect during closed-complex  
 623 formation. This analysis is complicated by the simultane-  
 624 ous occurrence of other regulatory actions of SC, in par-  
 625 ticular in promoter opening, but the investigated effect  
 626 still emerged as a predictable quantitative signal, both  
 627 in specifically designed mutant promoter assays and as  
 628 a statistical tendency in whole-genome data. The model  
 629 and the latter results altogether suggest that this mech-  
 630 anism has a widespread relevance in bacterial transcrip-  
 631 tion, although more detailed and comprehensive analyses  
 632 will be required to confirm it in various species.

### 633 Limitations of the regulatory model: the case of 634 16-nt promoters

635 The model was based on the hypothesis that the investi-  
 636 gated physical mechanism could be decoupled from both  
 637 other regulatory effects of SC (assumed to be independ-  
 638 ent of the spacer length) and other modulating effects

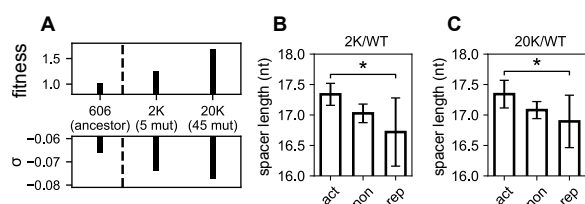


Figure 5: Contribution of the spacer to the global selectivity of promoters by inheritable increase of negative supercoiling in *E. coli*. (A) Reproduced from [48]. Strains from the longest-running evolution experiment [58]: 606 (ancestral genetic background), 2K (clone isolated from population at 2,000 generations), 20K (at 20,000 generations). Evolved strains exhibit higher chromosomal SC density  $\sigma$  compared to the ancestor, due to the natural acquisition of two point mutations: one in *topA* before 2,000 generations (among the five observed), and one in *fis* before 20,000 generations (among the 45 observed). Those mutations are associated to fitness gains through global expression changes [48]. (B) Mean comparison of spacer lengths for promoters activated (act), not significantly affected (non) or repressed (rep) in the 2K evolved strain compared to the ancestor. As expected from our modelling for an increase of negative SC, activated promoters have longer spacers compared to repressed ones ( $P = 0.04$ ). (C) Same for the 20K evolved strain compared to the ancestor, where the same difference is observed ( $P = 0.026$ ). The wider confidence intervals for repressed promoters result from a higher activation among the responsive genes in the investigated conditions.

of the spacer length and sequence (assumed to be inde- 639  
 640 pendent of SC). These hypotheses, already proposed and  
 641 supported by a collection of qualitative observations on  
 642 individual promoters [10], are validated here *a posteri-* 642  
 643 *ori* by the good agreement between model predictions  
 644 and analysed data of different kinds. A notable excep- 644  
 645 tion, however, are promoters involving 16-nt long spac- 645  
 646 ers, for which all observations converge to an opposite 646  
 647 behaviour, being more repressed by DNA relaxation than 647  
 648 17-nt spacers (Supplementary Fig. S5, Fig. 4B). This be- 648  
 649 haviour does not necessarily imply that the orientational 649  
 650 effect proposed here is absent, but possibly that it is over- 650  
 651 come by a stronger opposite regulatory effect of SC at 651  
 652 a later stage of transcription, in particular during open- 652  
 653 complex formation where different elements of the pro- 653  
 654 moter make extensive and complex contacts with RNAP 654  
 655 and the destabilisation of the double helix by SC has a 655  
 656 strong influence [60, 16, 9].

A well-studied class of promoters involving 16-nt 657  
 658 spacers are those encoding stable RNAs in *E. coli*, sub- 658  
 659 ject to stringent control. They are also characterised by 659  
 660 a G/C-rich "discriminator" inducing unusual interactions 660  
 661 with RNAP, in particular a strong repression by the desta-

662 bilisation of the transcription bubble by ppGpp, but also  
 663 by DNA relaxation occurring in the cell upon transition  
 664 to stationary phase [61, 62, 63]. Both effects were at-  
 665 tributed to the unusual kinetics of promoter opening and  
 666 escape due to the discriminator sequence [64, 60, 65],  
 667 but the latter sensitivity to DNA relaxation is consistent  
 668 with our observations based on the effect of the 16-nt  
 669 long spacer (Supplementary Fig. S5, Fig. 4), and both  
 670 properties might thus contribute in their behaviour. We  
 671 further tested if the unusual discriminator content of pro-  
 672 moters containing 16-nt long spacers was also present in  
 673 the promoters of protein-encoding genes. Indeed, com-  
 674 pared to all other such promoters, those containing 16-  
 675 nt spacers exhibit significantly G/C-richer discriminators  
 676 (Supplementary Fig. S7,  $P < 0.001$ ). This observa-  
 677 tion further suggests a tight relation between both proper-  
 678 ties of these promoters (spacer length and discrim-  
 679 inator sequence), which might underpin a specific path-  
 680 way in open-complex formation, beyond the range of our  
 681 model focused on the closed-complex. This notion is fur-  
 682 ther emphasised by the observation that the reactivity of  
 683 spacer DNA to potassium permanganate or DMS within  
 684 the open-complex depends on the spacer length or se-  
 685 quence, indeed suggesting an effect on its conformation  
 686 at that later stage of transcription [28, 66].

### 687 **Additional factors influencing the relative orien-** 688 **tation of -10/-35 elements**

689 While we focused on the effect of SC, the (un)twisting  
 690 of the spacer DNA for RNAP binding has been pro-  
 691 posed as the mechanistic basis for the regulatory action of  
 692 several TFs, including MerR, which regulates the (*mer*)  
 693 operon encoding components of the mercury (Hg) resis-  
 694 tance system [31]. The well-studied *mer* promoter in-  
 695 cludes the MerR binding site overlapping -35 and -10 el-  
 696 ements which are separated by a 19-nt spacer, a length  
 697 essential for normal activation by MerR [67]. In the ab-  
 698 sence of Hg(II) salts, MerR binds to *mer* promoter in its  
 699 repressor conformation. In the presence of Hg(II), the lat-  
 700 ter binds to MerR, causing a conformational change and,  
 701 in turn, the untwisting of DNA to realign -35/-10 sites  
 702 for effective open-complex formation [31]. Negative SC  
 703 results in DNA unwinding similar to that of MerR activa-  
 704 tion, and a global increase in SC indeed facilitates MerR-  
 705 mediated activation, impedes MerR-mediated repression,  
 706 and conversely for DNA relaxation [68]. The same  
 707 mechanism presumably applies to other metal-dependent  
 708 regulators [31, 69]. Finally, the same mechanism may oc-  
 709 cur in Mu bacteriophage, where protein C-mediated ac-  
 710 tivation of the *mom* gene (encoding a DNA modification

function) requires a 19-nt spacer between -35/-10 sites,  
 and involves a twisting deformation to realigning the lat-  
 ter for RNAP recruitment [70].

### **Effect of spacer sequence on -35/-10 alignment by RNA Polymerase**

Apart from the limitations mentioned above, in the  
 present model, we only considered how SC modulates  
 the relative torsional orientation of -10 and -35 el-  
 ements depending on the spacer length, and neglected  
 any effect of its sequence. However, the latter has been  
 shown to affect transcriptional activity of various pro-  
 moters [26, 22, 23, 71, 24, 72, 25]. Focusing specifi-  
 cally on the torsional orientation between the -35 and -10  
 sites (sensitive to SC), the spacer sequence might mod-  
 ulate two parameters that were considered as constant in  
 the equations: the average twist angle between succes-  
 sive basepairs ( $\alpha_0$  in Eq. 1) and the torsional stiffness  
 ( $k_\theta$ ), whose values were estimated for all basepair step  
 sequences from a collection of crystallographic struc-  
 tures [73, 32]. We implemented these values [33] to es-  
 timate the magnitude of the resulting adjustment of the  
 regulatory effect of SC, for all  $\sigma 70$ -dependent promot-  
 ers of *E. coli* (Supplementary Fig. S8, with details of  
 the calculation in Supplementary Information). Overall,  
 the maximal span of the sequence contribution remains  
 weaker than the gain/loss of one nucleotide in the spacer  
 (Supplementary Fig. S8A), confirming our hypothesis  
 that the effect of the spacer sequence is weaker than that  
 of its length. We then asked if this modulation results  
 from the sequence-induced heterogeneity of the twist an-  
 gle (i.e., the DNA structure) or stiffness (i.e., elasticity),  
 or both. We therefore modified the computation to im-  
 pose a sequence-averaged value to either of these two pa-  
 rameters (Supplementary Fig. S8B-C). We observed that  
 the effect of the sequence on DNA stiffness alone has  
 almost no regulatory effect (Supplementary Fig. S8B),  
 meaning that the sequence contributes mostly by mod-  
 ulating the total twist angle of the spacer, although as  
 noted, this modulation remains modest compared to that  
 induced by changing the spacer length.

### **Acknowledgements**

We thank the whole CRP team for helpful discussions,  
 and Nicolas Paulhan for experimental contributions.

## References

- [1] Travers, A. and Muskhelishvili, G. (February, 2005) DNA supercoiling - a global transcriptional regulator for enterobacterial growth?. *Nature Reviews. Microbiology*, **3**(2), 157–169.
- [2] Dorman, C. J. and Dorman, M. J. (November, 2016) DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. *Biophysical Reviews*, **8**(Suppl 1), 89–100.
- [3] Martis B., S., Forquet, R., Reverchon, S., Nasser, W., and Meyer, S. (January, 2019) DNA Supercoiling: an Ancestral Regulator of Gene Expression in Pathogenic Bacteria?. *Computational and Structural Biotechnology Journal*, **17**, 1047–1055.
- [4] Verma, S. C., Qian, Z., and Adhya, S. L. (December, 2019) Architecture of the Escherichia coli nucleoid. *PLoS Genetics*, **15**(12).
- [5] Jovanovich, S. B. and Lebowitz, J. (1987) Estimation of the effect of coumermycin A1 on Salmonella typhimurium promoters by using random operon fusions. *Journal of Bacteriology*, **169**(10), 4431–4435.
- [6] Pruss, G. J. and Drlica, K. (1989) DNA supercoiling and prokaryotic transcription. *Cell*, **56**(4), 521–523.
- [7] Figueroa-Bossi, N., Guérin, M., Rahmouni, R., Leng, M., and Bossi, L. (1998) The supercoiling sensitivity of a bacterial tRNA promoter parallels its responsiveness to stringent control.. *The EMBO Journal*, **17**(8), 2359–2367.
- [8] Auner, H., Buckle, M., Deufel, A., Kutateladze, T., Lazarus, L., Mavathur, R., Muskhelishvili, G., Pemberton, I., Schneider, R., and Travers, A. (2003) Mechanism of transcriptional activation by FIS: role of core promoter structure and DNA topology. *Journal of Molecular Biology*, **331**(2), 331–344.
- [9] Forquet, R., Pineau, M., Nasser, W., Reverchon, S., and Meyer, S. (August, 2021) Role of the Discriminator Sequence in the Supercoiling Sensitivity of Bacterial Promoters. *mSystems*, **6**(4), e0097821.
- [10] Wang, J. Y. and Syvanen, M. (July, 1992) DNA twist as a transcriptional sensor for environmental changes. *Molecular Microbiology*, **6**(14), 1861–1866.
- [11] Wood, D. C. and Lebowitz, J. (1984) Effect of supercoiling on the abortive initiation kinetics of the RNA-I promoter of ColE1 plasmid DNA. *The Journal of Biological Chemistry*, **259**(18), 11184–11187.
- [12] Dorman, C. J. (2019) DNA supercoiling and transcription in bacteria: a two-way street. *BMC molecular and cell biology*, **20**(1), 26.
- [13] Forquet, R., Pineau, M., Nasser, W., Reverchon, S., and Meyer, S. (May, 2021) The discriminator sequence is a primary determinant in the supercoiling response of bacterial promoters. *bioRxiv*, p. 2020.10.01.322149 Publisher: Cold Spring Harbor Laboratory Section: New Results.
- [14] Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (April, 2005) Transcriptional regulation by the numbers: applications. *Current Opinion in Genetics & Development*, **15**(2), 125–135.
- [15] Klein, C. A., Teufel, M., Weile, C. J., and Sobetzko, P., More than a Spacer: The Bacterial Promoter Spacer Modulates Promoter Strength and Timing by Length, Sequence, TG-Motifs and DNA Supercoiling Sensitivity. Technical report, (October, 2021).
- [16] Ruff, E. F., Record, M. T., and Artsimovitch, I. (May, 2015) Initial Events in Bacterial Transcription Initiation. *Biomolecules*, **5**(2), 1035–1062.
- [17] Soberón-Chávez, G., Alcaraz, L. D., Morales, E., Ponce-Soto, G. Y., and Servín-González, L. (2017) The Transcriptional Regulators of the CRP Family Regulate Different Essential Bacterial Functions and Can Be Inherited Vertically and Horizontally. *Frontiers in Microbiology*, **8**, 959.
- [18] Kumar, A., Malloch, R. A., Fujita, N., Smillie, D. A., Ishihama, A., and Hayward, R. S. (July, 1993) The Minus 35-Recognition Region of Escherichia coli Sigma 70 is Inessential for Initiation of Transcription at an "Extended Minus 10" Promoter. *Journal of Molecular Biology*, **232**(2), 406–418.
- [19] Mitchell, J. E., Zheng, D., Busby, S. J. W., and Minchin, S. D. (August, 2003) Identification and analysis of 'extended -10' promoters in Escherichia coli. *Nucleic Acids Research*, **31**(16), 4689–4695.

- 842 [20] Aoyama, T., Takanami, M., Ohtsuka, E., Taniyama,  
843 Y., Marumoto, R., Sato, H., and Ikehara, M.  
844 (September, 1983) Essential structure of E. coli pro-  
845 moter: effect of spacer length between the two con-  
846 sensus sequences on promoter function.. *Nucleic  
847 Acids Research*, **11**(17), 5855–5864.
- 848 [21] Mulligan, M. E., Brosius, J., and McClure, W. R.  
849 (March, 1985) Characterization in vitro of the effect  
850 of spacer length on the activity of Escherichia coli  
851 RNA polymerase at the TAC promoter.. *Journal of  
852 Biological Chemistry*, **260**(6), 3529–3538.
- 853 [22] Chan, B. and Busby, S. (December, 1989) Recogni-  
854 tion of nucleotide sequences at the Escherichia coli  
855 galactose operon P1 promoter by RNA polymerase.  
856 *Gene*, **84**(2), 227–236.
- 857 [23] Thouvenot, B., Charpentier, B., and Branlant, C.  
858 (October, 2004) The strong efficiency of the Es-  
859 cherichia coli gapA P1 promoter depends on a com-  
860 plex combination of functional determinants. *The  
861 Biochemical Journal*, **383**(Pt 2), 371–382.
- 862 [24] Liu, M., Tolstorukov, M., Zhurkin, V., Garges,  
863 S., and Adhya, S. (May, 2004) A mutant spacer  
864 sequence between -35 and -10 elements makes  
865 the Plac promoter hyperactive and cAMP receptor  
866 protein-independent. *Proceedings of the National  
867 Academy of Sciences of the United States of Amer-  
868 ica*, **101**(18), 6911–6916.
- 869 [25] Hook-Barnard, I. G. and Hinton, D. M. (January,  
870 2009) The promoter spacer influences transcription  
871 initiation via sigma70 region 1.1 of Escherichia  
872 coli RNA polymerase. *Proceedings of the National  
873 Academy of Sciences of the United States of Amer-  
874 ica*, **106**(3), 737–742.
- 875 [26] Hook-Barnard, I. G. and Hinton, D. M. (2007)  
876 Transcription initiation by mix and match elements:  
877 flexibility for polymerase binding to bacterial pro-  
878 moters. *Gene Regulation and Systems Biology*, **1**,  
879 275–293.
- 880 [27] Singh, S. S., Typas, A., Hengge, R., and Grainger,  
881 D. C. (July, 2011) Escherichia coli 70 senses se-  
882 quence and conformation of the promoter spacer re-  
883 gion. *Nucleic Acids Research*, **39**(12), 5109–5118.
- 884 [28] Auble, D. T. and deHaseth, P. L. (August, 1988)  
885 Promoter recognition by Escherichia coli RNA  
886 polymerase. Influence of DNA structure in the  
spacer separating the -10 and -35 regions. *Journal  
of Molecular Biology*, **202**(3), 471–482.
- 889 [29] Warne, S. E. and deHaseth, P. L. (June, 1993) Pro-  
890 moter recognition by Escherichia coli RNA poly-  
891 merase. Effects of single base pair deletions and in-  
892 sertions in the spacer DNA separating the -10 and -  
893 35 regions are dependent on spacer DNA sequence.  
894 *Biochemistry*, **32**(24), 6134–6140.
- 895 [30] Unniraman, S. and Nagaraja, V. (December, 2001)  
896 Axial distortion as a sensor of supercoil changes: a  
897 molecular model for the homeostatic regulation of  
898 DNA gyrase. *Journal of Genetics*, **80**(3), 119–124.
- 899 [31] Brown, N. L., Stoyanov, J. V., Kidd, S. P., and  
900 Hobman, J. L. (June, 2003) The MerR family of  
901 transcriptional regulators. *FEMS Microbiology Re-  
902 views*, **27**(2-3), 145–163.
- 903 [32] Xu, F. and Olson, W. K. (June, 2010) DNA ar-  
904 chitecture, deformability, and nucleosome position-  
905 ing. *Journal of Biomolecular Structure & Dynam-  
906 ics*, **27**(6), 725–739.
- 907 [33] Cevost, J., Vaillant, C., and Meyer, S. (Febru-  
908 ary, 2018) ThreaDNA: predicting DNA mechan-  
909 ics' contribution to sequence selectivity of proteins  
910 along whole genomes. *Bioinformatics*, **34**(4), 609–  
911 616.
- 912 [34] Watson, J. D. and Crick, F. H. C. (April, 1953)  
913 Molecular Structure of Nucleic Acids: A Structure  
914 for Deoxyribose Nucleic Acid. *Nature*, **171**(4356),  
915 737–738.
- 916 [35] Harteis, S. and Schneider, S. (July, 2014) Mak-  
917 ing the Bend: DNA Tertiary Structure and Protein-  
918 DNA Interactions. *International Journal of Molec-  
919 ular Sciences*, **15**(7), 12335–12363.
- 920 [36] Joyeux, M. and Junier, I. (September, 2020) Re-  
921 quirements for DNA-Bridging Proteins to Act as  
922 Topological Barriers of the Bacterial Genome. *Bio-  
923 physical Journal*, **119**(6), 1215–1225.
- 924 [37] Thompson, J. F., Hayes, L. S., and Lloyd, D. B.  
925 (July, 1991) Modulation of firefly luciferase stabil-  
926 ity and impact on studies of gene regulation. *Gene*,  
927 **103**(2), 171–177.
- 928 [38] Dorman, C. J., Barr, G. C., Ni Bhriain, N., and  
929 Higgins, C. F. (1988) DNA supercoiling and the  
930 anaerobic and growth phase regulation of tonB gene

- 931 expression. *Journal of Bacteriology*, **170**(6), 2816–  
932 2826.
- 933 [39] Schneider, R., Travers, A., and Muskhelishvili, G.  
934 (2000) The expression of the *Escherichia coli* *fis*  
935 gene is strongly dependent on the superhelical den-  
936 sity of DNA. *Molecular Microbiology*, **38**(1), 167–  
937 175.
- 938 [40] Hsieh, L.-S., Burger, R. M., and Drlica, K. (1991)  
939 Bacterial DNA supercoiling and [ATP]/[ADP].  
940 Changes associated with a transition to anaerobic  
941 growth. *Journal of Molecular Biology*, **219**(3), 443–  
942 450.
- 943 [41] Cameron, A. D. S., Stoebel, D. M., and Dorman,  
944 C. J. (2011) DNA supercoiling is differentially reg-  
945 ulated by environmental factors and FIS in *Es-*  
946 *cherichia coli* and *Salmonella enterica*. *Molecular*  
947 *Microbiology*, **80**(1), 85–101.
- 948 [42] Ouafa, Z.-A., Reverchon, S., Lautier, T., Muskhelishvili, G., and Nasser, W. (May, 2012) The nucleoid-associated proteins H-NS and FIS modulate the DNA supercoiling response of the *pel* genes, the major virulence factors in the plant pathogen bacterium *Dickeya dadantii*. *Nucleic Acids Research*, **40**(10), 4306–4319.
- 955 [43] Drlica, K. and Snyder, M. (1978) Superhelical *Es-*  
956 *cherichia coli* DNA: Relaxation by coumermycin.  
957 *Journal of Molecular Biology*, **120**(2), 145–154.
- 958 [44] Borowiec, J. A. and Gralla, J. D. (May, 1987) All  
959 three elements of the *lac ps* promoter mediate its  
960 transcriptional response to DNA supercoiling. *Journal*  
961 *of Molecular Biology*, **195**(1), 89–97.
- 962 [45] Aoyama, T. and Takanami, M. (March, 1988) Su-  
963 percoiling response of *E. coli* promoters with differ-  
964 ent spacer lengths. *Biochimica Et Biophysica Acta*,  
965 **949**(3), 311–317.
- 966 [46] Blot, N., Mavathur, R., Geertz, M., Travers, A., and  
967 Muskhelishvili, G. (July, 2006) Homeostatic regu-  
968 lation of supercoiling sensitivity coordinates tran-  
969 scription of the bacterial genome. *EMBO Reports*,  
970 **7**(7), 710–715.
- 971 [47] Keseler, I. M., Mackie, A., Santos-Zavaleta, A.,  
972 Billington, R., Bonavides-Martínez, C., Caspi, R.,  
973 Fulcher, C., Gama-Castro, S., Kothari, A., Krum-  
974 menacker, M., Latendresse, M., Muñoz-Rascado,  
975 L., Ong, Q., Paley, S., Peralta-Gil, M., Subhraveti,  
P., Velázquez-Ramírez, D. A., Weaver, D., Collado-  
Vides, J., Paulsen, I., and Karp, P. D. (January,  
2017) The EcoCyc database: reflecting new knowl-  
edge about *Escherichia coli* K-12. *Nucleic Acids*  
*Research*, **45**(Database issue), D543–D550.
- [48] El Houdaigui, B., Forquet, R., Hindré, T., Schnei-  
der, D., Nasser, W., Reverchon, S., and Meyer, S.  
(June, 2019) Bacterial genome architecture shapes  
global transcriptional regulation by DNA supercoil-  
ing. *Nucleic Acids Research*, **47**(11), 5648–5657.
- [49] Webber, M. A., Ricci, V., Whitehead, R., Patel, M.,  
Fookes, M., Ivens, A., and Piddock, L. J. V. (July,  
2013) Clinically relevant mutant DNA gyrase alters  
supercoiling, changes the transcriptome, and con-  
fers multidrug resistance. *mBio*, **4**(4).
- [50] Forquet, R., Jiang, X., Nasser, W., Hommais, F.,  
Reverchon, S., and Meyer, S. (October, 2021)  
Mapping the complex transcriptional landscape of  
the phytopathogenic bacterium *Dickeya dadantii*. p.  
2020.09.30.320440.
- [51] Junier, I., Unal, E. B., Yus, E., Lloréns-Rico, V.,  
and Serrano, L. (2016) Insights into the Mecha-  
nisms of Basal Coordination of Transcription Using  
a Genome-Reduced Bacterium. *Cell Systems*, **2**(6),  
391–401.
- [52] Vijayan, V., Jain, I. H., and O’Shea, E. K. (2011) A  
high resolution map of a cyanobacterial transcrip-  
tome. *Genome Biology*, **12**(5), R47.
- [53] Shahmuradov, I. A., Mohamad Razali, R.,  
Bougouffa, S., Radovanovic, A., and Bajic, V. B.  
(February, 2017) bTSSfinder: a novel tool for the  
prediction of promoters in cyanobacteria and *Es-*  
*cherichia coli*. *Bioinformatics*, **33**(3), 334–340.
- [54] Ferrándiz, M.-J., Martín-Galiano, A. J., Arnanz,  
C., Camacho-Soguero, I., Tirado-Vélez, J.-M., and  
de la Campa, A. G. (September, 2016) An in-  
crease in negative supercoiling in bacteria reveals  
topology-reacting gene clusters and a homeostatic  
response mediated by the DNA topoisomerase I  
gene. *Nucleic Acids Research*, **44**(15), 7292–7303.
- [55] Pineau, M., B, S. M., Forquet, R., Baude, J.,  
Grand, L., Popowycz, F., Soulère, L., Hommais, F.,  
Nasser, W., Reverchon, S., and Meyer, S. (April,  
2021) What is a supercoiling-sensitive gene? In-  
sights from topoisomerase I inhibition in the Gram-  
negative bacterium *Dickeya dadantii*. *bioRxiv*, p.



- 2021.04.09.439150 Publisher: Cold Spring Harbor Laboratory Section: New Results.
- [56] Zechiedrich, E. L., Khodursky, A. B., Bachellier, S., Schneider, R., Chen, D., Lilley, D. M., and Cozzarelli, N. R. (2000) Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. *The Journal of Biological Chemistry*, **275**(11), 8103–8113.
- [57] Crozat, E., Philippe, N., Lenski, R. E., Geiselman, J., and Schneider, D. (February, 2005) Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics*, **169**(2), 523–532.
- [58] Tenailon, O., Barrick, J. E., Ribbeck, N., Deatherage, D. E., Blanchard, J. L., Dasgupta, A., Wu, G. C., Wielgoss, S., Cruveiller, S., Médigue, C., Schneider, D., and Lenski, R. E. (August, 2016) Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature*, **536**(7615), 165–170.
- [59] Rovinskiy, N. S., Agbleke, A. A., Chesnokova, O. N., and Higgins, N. P. (March, 2019) Supercoil Levels in *E. coli* and *Salmonella* Chromosomes Are Regulated by the C-Terminal 35–38 Amino Acids of GyrA. *Microorganisms*, **7**(3).
- [60] Ohlsen, K. L. and Gralla, J. D. (1992) DNA melting within stable closed complexes at the *Escherichia coli* rrnB P1 promoter. *The Journal of Biological Chemistry*, **267**(28), 19813–19818.
- [61] Schneider, D. A., Ross, W., and Gourse, R. L. (April, 2003) Control of rRNA expression in *Escherichia coli*. *Current Opinion in Microbiology*, **6**(2), 151–156.
- [62] Dennis, P. P., Ehrenberg, M., and Bremer, H. (December, 2004) Control of rRNA Synthesis in *Escherichia coli*: a Systems Biology Approach. *Microbiology and Molecular Biology Reviews*, **68**(4), 639–668 Publisher: American Society for Microbiology.
- [63] Hori, H., Tomikawa, C., Hirata, A., Toh, Y., Tomita, K., Ueda, T., and Watanabe, K. (2014) Transfer RNA Synthesis and Regulation. In *eLS* American Cancer Society.
- [64] Lamond, A. I. (1985) Supercoiling response of a bacterial tRNA gene.. *The EMBO Journal*, **4**(2), 501–507.
- [65] Winkelman, J. T., Chandrangsu, P., Ross, W., and Gourse, R. L. (March, 2016) Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of *E. coli* ribosomal RNA promoters. *Proceedings of the National Academy of Sciences*, **113**(13), E1787–E1795.
- [66] Newlands, J. T., Ross, W., Gosink, K. K., and Gourse, R. L. (August, 1991) Factor-independent activation of *Escherichia coli* rRNA transcription: II. Characterization of complexes of rrnB P1 promoters containing or lacking the upstream activator region with *Escherichia coli* RNA polymerase. *Journal of Molecular Biology*, **220**(3), 569–583.
- [67] Parkhill, J. and Brown, N. L. (September, 1990) Site-specific insertion and deletion mutants in the mer promoter-operator region of Tn501; the nineteen base-pair spacer is essential for normal induction of the promoter by MerR.. *Nucleic Acids Research*, **18**(17), 5157–5162.
- [68] Condee, C. W. and Summers, A. O. (December, 1992) A mer-lux transcriptional fusion for real-time examination of in vivo gene expression kinetics and promoter response to altered superhelicity. *Journal of Bacteriology*, **174**(24), 8094–8101.
- [69] Heldwein, E. E. and Brennan, R. G. (January, 2001) Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature*, **409**(6818), 378–382.
- [70] Basak, S. and Nagaraja, V. (December, 2001) DNA unwinding mechanism for the transcriptional activation of momP1 promoter by the transactivator protein C of bacteriophage Mu. *The Journal of Biological Chemistry*, **276**(50), 46941–46945.
- [71] Mellies, J., Brems, R., and Villarejo, M. (June, 1994) The *Escherichia coli* proU promoter element and its contribution to osmotically signaled transcription activation.. *Journal of Bacteriology*, **176**(12), 3638–3645.
- [72] Repoila, F. and Gottesman, S. (November, 2003) Temperature Sensing by the dsrA Promoter. *Journal of Bacteriology*, **185**(22), 6609–6614.
- [73] Calladine, C. R. (October, 1982) Mechanics of sequence-dependent stacking of bases in B-DNA. *Journal of Molecular Biology*, **161**(2), 343–352.

## Supplementary Information for

### Quantitative contribution of the spacer length in the supercoiling-sensitivity of bacterial promoters

Raphaël Forquet, William Nasser, Sylvie Reverchon and Sam Meyer\*

Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, CNRS, UMR5240 MAP, F-69622, France

\* To whom correspondence should be addressed. Email: sam.meyer@insa-lyon.fr.

October 21, 2021

**This PDF file includes:**

- Supplementary text
- Supplementary figures S1 to S8
- Supplementary tables S1 to S2
- Supplementary information references

## Supplementary text

### Genome-wide analyses of spacer responses to supercoiling variations in distant bacteria

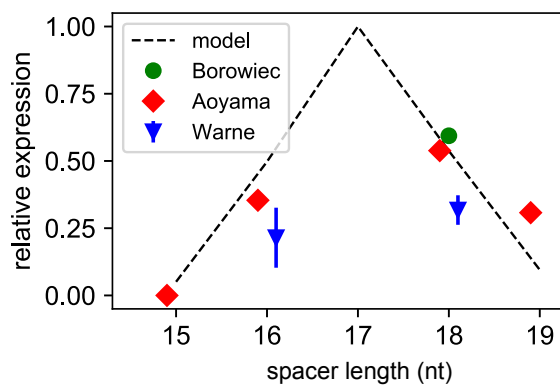
Transcriptomic data were collected from the literature (Supplementary Tab. S2). Since curated promoter maps were not available, genome-wide TSS maps and associated genes were retrieved from literature (Supplementary Tab. S2), and a scan for promoter motifs upstream of TSSs was conducted with bTSSfinder [1]. Only  $\sigma 70$ -dependent promoters were retained for all organisms, except for the cyanobacterium *S. elongatus*, for which only  $\sigma A$ -dependent promoters were kept (primary  $\sigma$  factor), and classified depending on their spacer length and response to the investigated condition. The thresholds for statistical selection procedures are indicated in Supplementary Tab. S2, and were adjusted to generate subsets of act/rep genes of sizes comparable among the different data sets, while having enough statistical power for the analysis. The relation between promoter activation and spacer length was then quantified by a Student's t-test between activated and repressed promoters (Supplementary Fig. S4), such as in Fig. 4 and 5. All error bars shown are 95% confidence intervals.

### Modelling of the effect of spacer sequence on -35/-10 alignment by RNA Polymerase

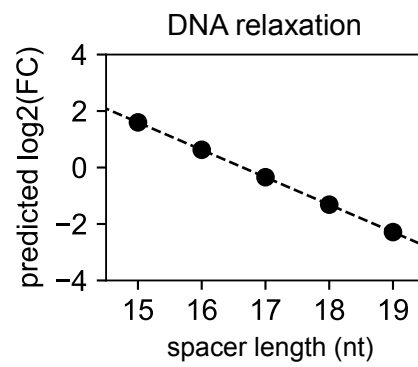
To estimate the contribution of the spacer sequence to the total angle between -35/-10 binding sites and to the spacer stiffness, the twisting deformation energy was computed as follow:

$$\Delta G(\sigma, s) = \frac{1}{2} k_{\theta}(s) \left( \theta_P - \theta_O(s) - \alpha_0 \frac{k_{\theta}}{k_{\theta}(s)} \sigma \right)^2$$

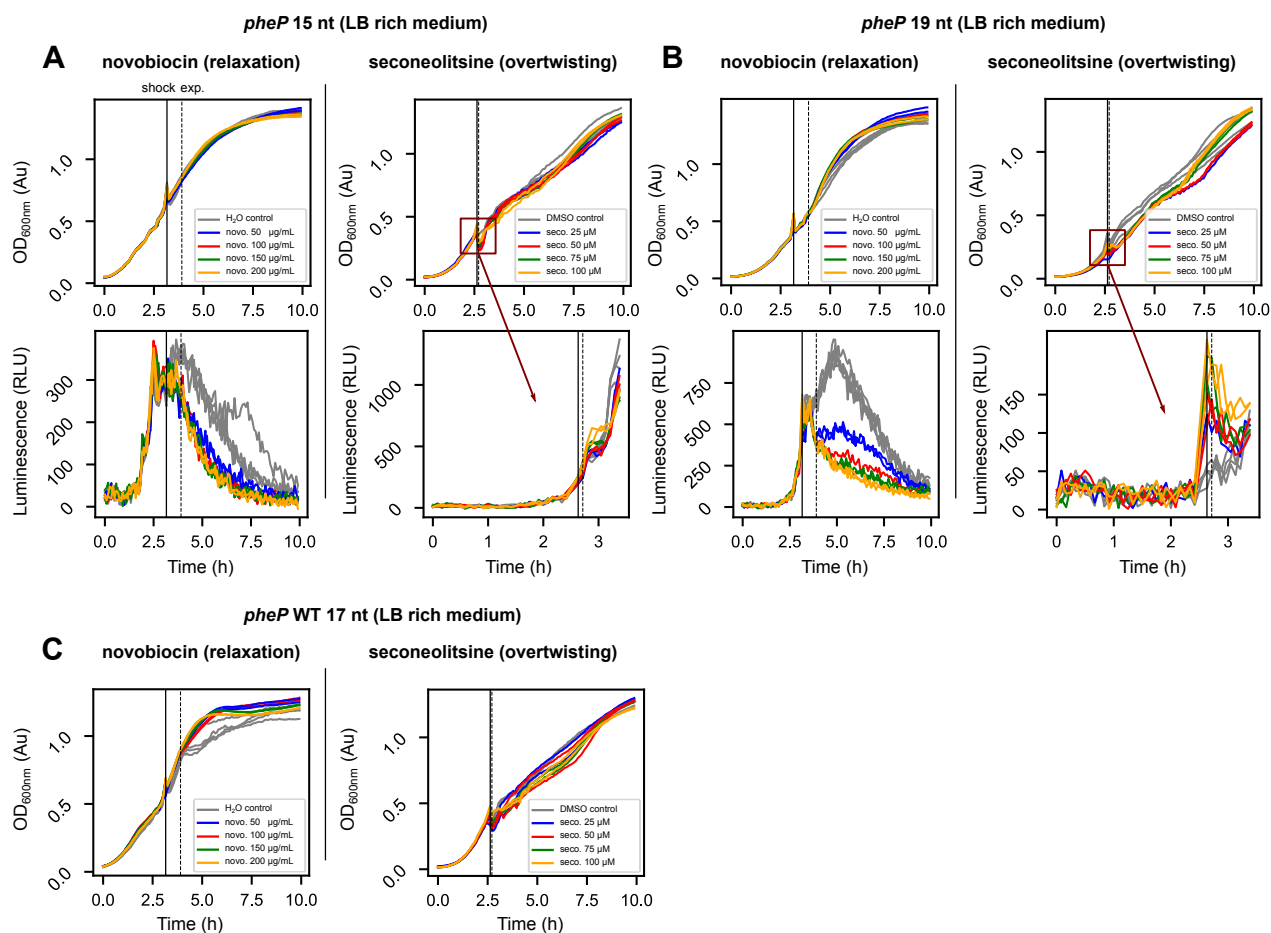
where  $\sigma$  is the supercoiling level,  $s$  is the promoter spacer sequence,  $k_{\theta}(s)$  is the DNA spacer twist stiffness estimated with ThreaDNA [2],  $\theta_P = 543$  is the optimal angle between -35 and -10 sites for RNAP binding as previously determined,  $\theta_O(s)$  is the effective angle estimated with ThreaDNA [2] which depends on base composition,  $\alpha_0 = 34^\circ$  is the average twist angle between adjacent nucleotides. Then, for each *E. coli*  $\sigma 70$ -dependent promoter, the response to DNA relaxation was predicted starting from a level  $\sigma = -0.06$ , with a relaxation magnitude  $\Delta\sigma = 0.03$  (Supplementary Fig. S8).



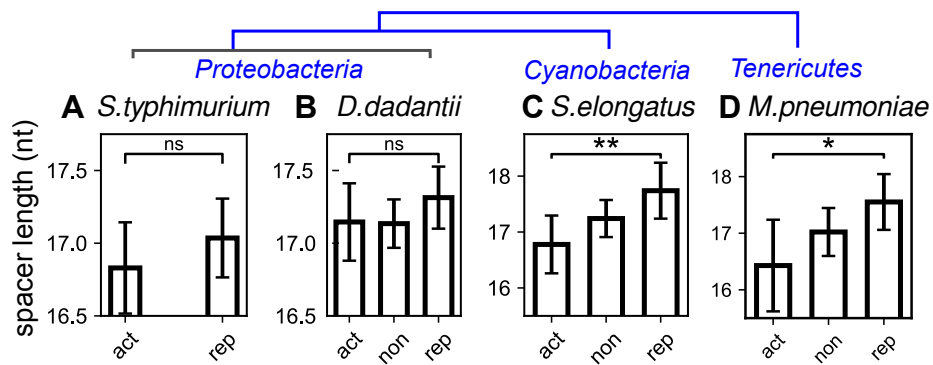
**Supplementary Figure S1:** Predicted relative promoter expression levels depending on spacer length (dashed lines), for a supercoiling level  $\sigma = -0.06$ , and comparison to *in vitro* promoter expression data (dots) from [3] (green), [4] (red), and [5] (blue). Error bars are shown for [5] data due to the usage of different promoters for which no SC level was measured.



**Supplementary Figure S2:** Predicted expression fold-changes from  $\sigma = -0.06$  to  $\sigma = -0.03$  for 15 to 19-nt spacers.

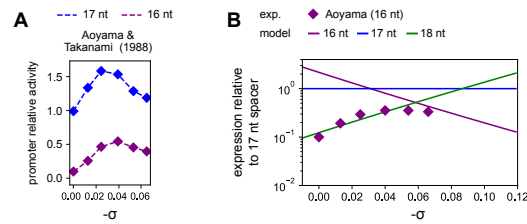


**Supplementary Figure S3:** Bacterial growth and promoter expression monitored in a microplate reader (raw data). **(A)** *E. coli* bacteria carrying plasmids with *pheP* 15-nt spacer promoter in LB rich medium. **(B)** *pheP* 19-nt spacer promoter. **(C)** *pheP* native 17-nt spacer promoter. For all promoters, upper panels display bacterial growth measured by  $OD_{600nm}$ , lower panels the expression levels measured by luminescence, left panels the experiments conducted with novobiocin, and right panels the experiments conducted with seconeolitsine. "Shock" and "exp." refer to the time points used for the antibiotic shock, and for the computation of expression fold-changes, respectively. Due to the stronger and more transient expression variation induced by seconeolitsine shock, a zoom is performed for luminescence. The  $OD_{600nm}$  discrepancies at shock time or after several hours are optical artefacts due to the opening of the recorder and/or the formation of sediments disrupting the measurements (the luminescence does not vary correspondingly among replicates). The luminescence curves of *pheP* native 17-nt spacer promoter are provided in Fig. 3.

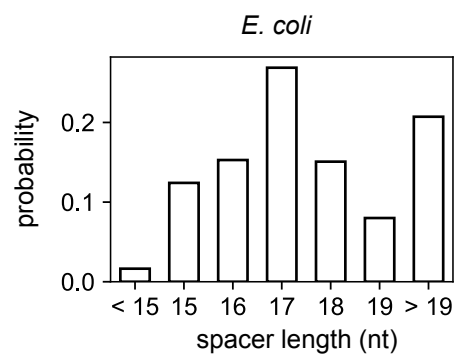


**Supplementary Figure S4:** Statistical relation between spacer length and promoter's response to DNA relaxation (act: activated, non: not significantly affected, rep: repressed) in distant bacterial species (see Supplementary Tab. S2 for experimental conditions and references): **(A)** *Salmonella typhimurium* ( $P = 0.169$ ), **(B)** *Dickeya dadantii* ( $P = 0.170$ ), **(C)** *Synechococcus elongatus* ( $P = 0.007$ ), **(D)** *Mycoplasma pneumoniae* ( $P = 0.017$ ). In proteobacteria and *M. pneumoniae*, only  $\sigma 70$  promoters were considered, whereas only  $\sigma A$  promoters were considered for *S. elongatus*. A schematic phylogeny is depicted above.

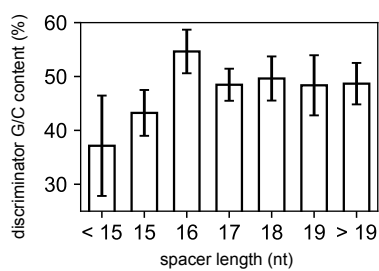




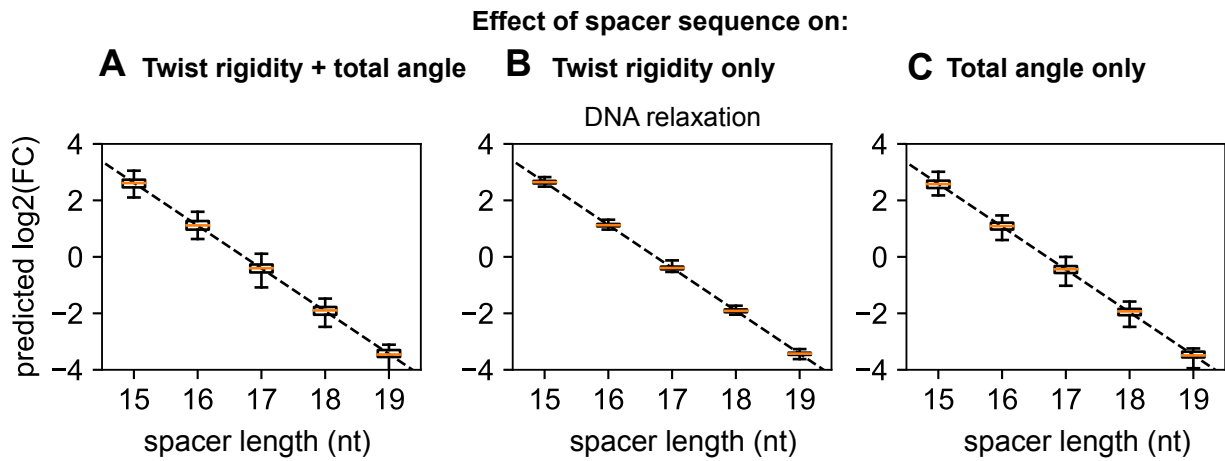
**Supplementary Figure S5:** Comparison of *in vitro* transcription assays for 16-nt spacers to model predictions. **(A)** Activity of promoters with 17 and 16-nt spacers (blue and purple, respectively) relative to that of the 17-nt spacer in the relaxed state, depending on supercoiling level  $\sigma$  (data from [4]). **(B)** Transcription model predictions (solid lines) compared to the *in vitro* promoter expression data (exp.). Each datapoint was normalised by the corresponding value obtained in the reference mutant promoter with a 17-nt spacer, to isolate specifically the spacer length contribution to the SC response, among other possible regulatory effects (see text).



**Supplementary Figure S6:** Distribution of spacer lengths at *E. coli*  $\sigma_{70}$ -dependent promoters in the Ecocyc database [6].



**Supplementary Figure S7:** Analysis of the coupling between promoter spacer length and discriminator G/C-content. Promoters with 16-nt spacers exhibit G/C-richer discriminators compared to other promoters ( $P$ -value  $< 0.001$ ), 15-nt and  $< 15$ -nt spacers exhibit A/T-richer discriminators ( $P$ -value = 0.015 and 0.004, respectively), whereas 17, 18, 19 and  $> 19$ -nt spacers have similar, not significantly different discriminator A/T-contents.



**Supplementary Figure S8:** Boxplot of predicted relaxation fold-changes from  $\sigma = -0.06$  to  $\sigma = -0.03$  for spacer sequences of *E. coli*  $\sigma 70$ -dependent promoters from the Ecocyc database [6]. The box extends from the first quartile to the third quartile values of the data, with an orange line at the median, and with whiskers extending from each end of the box to the extreme values. **(A)** With sequence-dependent spacer twist rigidity and total angle between -35/-10 RNAP binding sites. The maximal effect magnitude of the sequence is similar to that of a gain or loss of one nucleotide in the spacer. **(B)** With sequence-dependent spacer twist rigidity and sequence-averaged total angle. **(C)** With sequence-dependent total angle and sequence-averaged spacer twist rigidity. In the model, the effect of spacer sequence is mainly driven by modifications of the total angle between -35/-10 sites in the model.

Promoter	Sequence
> <i>pheP</i> _WT_17nt	CTCGAGTCAGAGGTGATGAGCCGGATTGCCGCGCCGATGATTGGCGGCATGATCACCGCACCTTTGCTGTCGCTGTTTATT ATCCCGCGCGGTATAAGCTGATGTGGCTGCACCGACATCGGGTACGGAAATAAAAGCAGGATACCCCGTTTAACCGTGTG GATTGTGTC TTGCGACGATGGGCTAAATGT TAAAAGGTGCCCTCAACAAAAAAGACACACAGGGGAAAGGCGGATCC
> <i>pheP</i> _15nt	CTCGAGTCAGAGGTGATGAGCCGGATTGCCGCGCCGATGATTGGCGGCATGATCACCGCACCTTTGCTGTCGCTGTTTATT ATCCCGCGCGGTATAAGCTGATGTGGCTGCACCGACATCGGGTACGGAAATAAAAGCAGGATACCCCGTTTAACCGTGTG GATTGTGTC TTGCGACGATGGGCTAAATGT TAAAAGGTGCCCTCAACAAAAAAGACACACAGGGGAAAGGCGGATCC
> <i>pheP</i> _19nt	CTCGAGTCAGAGGTGATGAGCCGGATTGCCGCGCCGATGATTGGCGGCATGATCACCGCACCTTTGCTGTCGCTGTTTATT ATCCCGCGCGGTATAAGCTGATGTGGCTGCACCGACATCGGGTACGGAAATAAAAGCAGGATACCCCGTTTAACCGTGTG GATTGTGTC TTGCGACGATACGGGCTAAATGT TAAAAGGTGCCCTCAACAAAAAAGACACACAGGGGAAAGGCGGATCC
	XhoI restriction site, -35 element, spacer, -10 element, TSS, BglII restriction site

Plasmid	Description	Origin
pUCTer- <i>luc</i>	High-copy-number vector (pUC18 derivative) containing a multiple cloning site upstream of the <i>luc</i> reporter gene, followed by a <i>rnnB</i> terminator and a <i>cat</i> gene conferring chloramphenicol resistance.	Laboratory collection

**Supplementary Table S1:** List of *pheP*-derived synthetic promoter sequences with mutated spacers, and pUCTer-*luc* plasmid used in this study.

Species	Condition	Threshold adj. P-value	Threshold log <sub>2</sub> (FC)	Supercoiling variation	Transcription start sites reference
<i>Escherichia coli</i>	norfloxacin [7]	0.05	0	relaxation	[6]
	inheritable supercoiling variation [8]	0.1	0	overtwisting	[6]
<i>Salmonella typhimurium</i>	<i>gyrA</i> mutant [9]	NA	0.4	relaxation	[10]
<i>Dickeya dadantii</i>	novobiocin [11]	0.05	0	relaxation	[12]
<i>Synechococcus elongatus</i>	supercoiling correlation* [13]	NA	0.4	relaxation/ overtwisting	[14]
<i>Mycoplasma pneumoniae</i>	novobiocin [15]	0.1	0	relaxation	[15]

**Supplementary Table S2:** Compilation of investigated species, conditions and references. The thresholds for statistical selection procedures are indicated, and were adjusted to generate subsets of activated/repressed genes of sizes comparable among the different data sets. For a threshold of 0.4 on the log<sub>2</sub>(fold-change), genes are considered activated for a log<sub>2</sub>(FC) > 0.4, repressed for a log<sub>2</sub>(FC) < -0.4, and not significantly affected for a log<sub>2</sub>(FC) comprised between -0.4 and 0.4. The correlation\* condition from *S. elongatus* corresponds to the phasing of gene expression in the SC circadian oscillation and provides an indirect proxy of gene response to SC relaxation [13].

## References

- [1] Shahmuradov, I. A., Mohamad Razali, R., Bougouffa, S., Radovanovic, A., and Bajic, V. B. (February, 2017) bTSSfinder: a novel tool for the prediction of promoters in cyanobacteria and *Escherichia coli*. *Bioinformatics*, **33**(3), 334–340.
- [2] Cevost, J., Vaillant, C., and Meyer, S. (February, 2018) ThreaDNA: predicting DNA mechanics' contribution to sequence selectivity of proteins along whole genomes. *Bioinformatics*, **34**(4), 609–616.
- [3] Borowiec, J. A. and Gralla, J. D. (May, 1987) All three elements of the lac ps promoter mediate its transcriptional response to DNA supercoiling. *Journal of Molecular Biology*, **195**(1), 89–97.
- [4] Aoyama, T. and Takanami, M. (March, 1988) Supercoiling response of *E. coli* promoters with different spacer lengths. *Biochimica Et Biophysica Acta*, **949**(3), 311–317.
- [5] Warne, S. E. and deHaseth, P. L. (June, 1993) Promoter recognition by *Escherichia coli* RNA polymerase. Effects of single base pair deletions and insertions in the spacer DNA separating the -10 and -35 regions are dependent on spacer DNA sequence. *Biochemistry*, **32**(24), 6134–6140.
- [6] Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., Latendresse, M., Muñoz-Rascado, L., Ong, Q., Paley, S., Peralta-Gil, M., Subhraveti, P., Velázquez-Ramírez, D. A., Weaver, D., Collado-Vides, J., Paulsen, I., and Karp, P. D. (January, 2017) The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Research*, **45**(Database issue), D543–D550.
- [7] Blot, N., Mavathur, R., Geertz, M., Travers, A., and Muskhelishvili, G. (July, 2006) Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Reports*, **7**(7), 710–715.
- [8] El Houdaigui, B., Forquet, R., Hindré, T., Schneider, D., Nasser, W., Reverchon, S., and Meyer, S. (June, 2019) Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Research*, **47**(11), 5648–5657.
- [9] Webber, M. A., Ricci, V., Whitehead, R., Patel, M., Fookes, M., Ivens, A., and Piddock, L. J. V. (July, 2013) Clinically relevant mutant DNA gyrase alters supercoiling, changes the transcriptome, and confers multidrug resistance. *mBio*, **4**(4).
- [10] Kröger, C., Dillon, S. C., Cameron, A. D. S., Papenfort, K., Sivasankaran, S. K., Hokamp, K., Chao, Y., Sittka, A., Hébrard, M., Händler, K., Colgan, A., Leekitcharoenphon, P., Langridge, G. C., Lohan, A. J., Loftus, B., Lucchini, S., Ussery, D. W., Dorman, C. J., Thomson, N. R., Vogel, J., and Hinton, J. C. D. (May, 2012) The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(20), E1277–E1286.
- [11] Jiang, X., Sobetzko, P., Nasser, W., Reverchon, S., and Muskhelishvili, G. (July, 2015) Chromosomal “Stress-Response” Domains Govern the Spatiotemporal Expression of the Bacterial Virulence Program. *mBio*, **6**(3), e00353–15.
- [12] Forquet, R., Jiang, X., Nasser, W., Hommais, F., Reverchon, S., and Meyer, S. (October, 2021) Mapping the complex transcriptional landscape of the phytopathogenic bacterium *Dickeya dadantii*. p. 2020.09.30.320440.
- [13] Vijayan, V., Zuzow, R., and O’Shea, E. K. (December, 2009) Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(52), 22564–22568.
- [14] Vijayan, V., Jain, I. H., and O’Shea, E. K. (2011) A high resolution map of a cyanobacterial transcriptome. *Genome Biology*, **12**(5), R47.
- [15] Junier, I., Unal, E. B., Yus, E., Lloréns-Rico, V., and Serrano, L. (2016) Insights into the Mechanisms of Basal Coordination of Transcription Using a Genome-Reduced Bacterium. *Cell Systems*, **2**(6), 391–401.





## Chapter 6

# Discussion and perspectives

To achieve the main thesis objectives, *i.e.* the modelling and validation of transcriptional regulation by SC at the genome-scale, but also the characterisation of *D. dadantii* transcriptome, I have first implemented a database and Python programming tools. The latter have been proven very effective to perform various tasks, including the statistical analyses conducted for the side projects presented in Chapter 2, *i.e.* the validation of the stochastic regulatory model of the transcription-supercoiling coupling [115], the investigation of the phenotypic and transcriptional effects of IHF mutation on *D. dadantii* virulence [122], and the transcriptome-based study of TopoI inhibition by the non-marketed antibiotic seconeolitsine in *D. dadantii* [121]. In the future, the database must be kept updated with new data, such that the reusable Python programs contribute to other projects within the team and beyond.

### 6.1 Future studies on *D. dadantii* transcriptome

In Chapter 3, we have mapped the complex transcriptional landscape of *D. dadantii*, based on a combination of independent transcriptomic datasets, including one of the first reported applications of long-read Nanopore native RNA-seq in prokaryotes. Beyond serving as a community resource to help elucidating the regulation of *D. dadantii* gene expression, it also provides insights into basal rules of coordination of transcription that might be valid for other bacteria. This comprises the extent of transcription beyond the scale of operons, transcriptional read-through at terminators, and the organisation of genes into noncontiguous transcription units and excludons. Among the future directions of research related to this work, an online browser could be developed for a genome-wide visualisation of the identified transcription



infection. To that end, dual RNA-seq may be performed, which is a promising method allowing to analyse gene expression simultaneously in both the pathogen and the host, based on deep sequencing [148].

Finally, since this work focused on protein-encoding genes, it could be combined with the *D. dadantii* non-coding RNA landscape identified recently [149], for a comprehensive picture of gene regulation in this phytopathogen. Indeed, while most investigations on *D. dadantii* virulence have relied on transcriptional regulation by TFs [7] and chromosome dynamics [119], recent studies highlight the importance of post-transcriptional regulation by non-coding RNAs in the control of gene expression. For instance, the inactivation of Hfq and ProQ RNA chaperones, *i.e.* RNA-binding proteins that facilitate RNA-RNA interactions [150], generates avirulent bacteria [151], demonstrating the key role of non-coding RNAs as regulators of *D. dadantii* pathogenicity, by targeting virulence genes and more.

## 6.2 Limitations of our modelling and future directions

In Chapters 4 and 5, we have developed and validated two independent thermodynamic models of transcriptional regulation by SC, which may contribute to further progress in understanding and predicting gene response to SC variations. By modulating RNAP-promoter interaction independently from additional regulatory protein, and based on fundamental properties of DNA, SC may be an ubiquitous and global regulation mode, as suggested by the genome-scale validation of model predictions in distant bacteria in terms of phylogeny and lifestyle. In the following, we discuss the limitations of our modelling approach in more detail, and propose some perspectives for future work, including model refinement.

First, these models were kept voluntarily as simple as possible, with only few adjustable parameters, since they focus on one of the multiple steps by which SC affects transcription, neglecting the possible contribution of other regulatory mechanisms to the SC response of promoters. This includes the modulation of regulatory protein binding at promoters such as H-NS and CRP [108, 152], or the thermodynamic competition with other structural transitions (denaturation, cruciform, exclusion, G-quadruplex, Z-DNA...) occurring at nearby sites depending on the SC level and affecting the transcription rate of promoters [86] (Fig. 6.2). This explains why at the genome-scale, the

effect of the discriminator sequence or the spacer length emerge as a statistical feature rather than a predictive signal dictating the response of each promoter, in contrast to mutation studies. As a key advantage of their simplicity, these models provide mechanistic insights into transcriptional regulation by SC, that may be eventually integrated into an unified regulatory model. The possible interplay between these two mechanisms independently modelled is discussed in Chapter 5, even if their respective contribution to the SC response of promoters should be evaluated quantitatively in the future.

Second, we consider a homogeneous level of SC along the chromosome, with promoters responding distinctly to SC variations depending on their discriminator G/C-content or spacer length. Since the total SC level has been shown to be heterogeneously distributed in different topological domains along the chromosome [104], in particular due to the dynamic production of supercoils by elongating RNAPs [126], this also explains discrepancies between model predictions and experimental data. Indeed, as previously presented in Chapter 2, the relative distance and orientation between adjacent genes explain a significant contribution to the SC response of promoters, even when they are assumed to respond identically [115]. Integrating this orientation-dependent heterogeneity of SC levels and the sequence-dependent heterogeneity of promoter response into an unified regulatory model is therefore a natural objective for future studies. This may be achieved by the high-resolution mapping of local SC through the incorporation of psoralen into DNA, which preferentially binds underwound regions of the double-helix. This was recently performed in *E. coli* [104], albeit with a low-resolution, preventing the quantification of local SC changes at the scale of gene promoters. A project is thus currently running in the team in collaboration with the University of Gdansk in Poland, for a high-resolution mapping of negative SC with psoralen. As a complementary approach, a method named GapR-seq has been recently developed for generating high-resolution maps of positive SC, which is based on the chromatin immunoprecipitation of GapR, a protein that preferentially binds overwound regions of the double-helix [153]. A combination of these high-resolution maps of local SC may thus characterise quantitatively the SC variations occurring at each gene promoter under given conditions, *e.g.* in response to DNA relaxation/overtwisting, in contrast to the homogeneous and average SC levels considered in our models. Eventually, this may allow a more precise confrontation between experimental data and model predictions, and thus to refine the models of transcriptional regulation by SC based on the sequence-specific response of promoters.

Third, these models are based on an unidimensional description of DNA, where the total SC level is assumed to be present in the form of twist, neglecting its partition into writhe which may also influence the SC response of promoters by modulating DNA conformation and, *e.g.* regulatory protein binding at promoters [108, 152]. It also neglects its partition into constrained/unconstrained superhelicity, which was estimated about a half of the total in *E. coli* [98]. It is thus assumed that the total SC level can be used by RNAP to twist/untwist the double-helix, which may be the case if (i) the associated writhe is converted to twist, since the latter equilibrates much more rapidly [97], and (ii) the potential topological barriers constraining SC get unbound during the process of transcription initiation. However, this is impossible to predict since it depends on unknown kinetic constants, DNA sequence, protein-binding pattern, and a variety of other factors, which are currently inaccessible experimentally with usual chloroquine gels, precluding the establishment of more precise models. These approximations neglect notably the formation of R-loops, which have received increasing attention in recent years. R-loops are frequent and ubiquitous three-stranded nucleic acid structures consisting of an RNA:DNA hybrid and a displaced single-stranded DNA [154, 155] (Fig. 6.2). They form in particular during transcription, upon hybridisation of the nascent RNA with the complementary template DNA strand, upstream the elongating RNAP where negative SC is generated (Fig. 6.2). Indeed, beyond the contribution of DNA sequence which favours the formation of co-transcriptional R-loops in G-rich regions of transcripts [154], it is largely driven by SC [156]. R-loops transiently absorb vast amounts of negative SC generated by elongating RNAPs, which energetically favour their formation, relaxing DNA in turn (Fig. 6.2). Beyond their key role in relieving the superhelical stress generated by transcription, the sequestration of negative SC by R-loop formation may regulate nearby promoters by (i) preventing the facilitation of DNA melting favoured by DNA unwinding, (ii) hindering the formation of alternative DNA structures that compete for negative SC, and (iii) abolishing the long-range contacts favoured by plectonemic supercoils, and possibly protein-DNA interactions (Fig. 6.2). Conversely, the DNA relaxation induced by R-loop formation may give rise to other DNA structures or protein-DNA interactions [154].

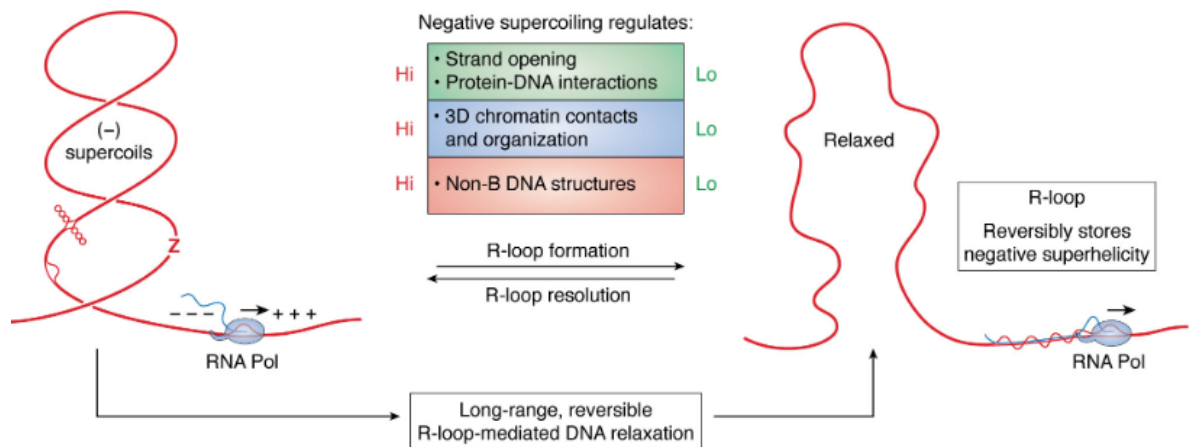


FIGURE 6.2: Copied from [154]: co-transcriptional R-loop formation and biological roles. Negative SC is generated upstream elongating RNAPs, possibly driving the formation of alternative DNA structures (bubble DNA, cruciform and Z-DNA depicted), increased chromatin contacts and protein-DNA interactions. R-loops may form reversibly by absorbing negative SC, relaxing DNA in turn. Hi: high, Lo: low.

Finally, since (i) SC is a fundamental property of DNA present in all living organisms, and (ii) the investigated mechanisms rely on universal properties of DNA and basal RNAP-promoter interaction, the results presented in this thesis may be extrapolated in part and with caution to other kingdoms of life.

### 6.3 Applications of our modelling in eukaryotes

In eukaryotes, DNA is packaged by wrapping around histones to form structures called nucleosomes that constrain SC [157]. Free DNA is found to be relaxed at the global scale, although several experiments demonstrated the presence of important levels of SC in different topological domains from 10 to 1000 kb, where it plays a role in chromatin structure and nucleosome remodelling [102, 158, 159]. Most SC is generated by RNAP during transcription [160, 161], due to the absence of DNA gyrase introducing negative supercoils into DNA. They however contain topoisomerases that solve the topological problems associated with DNA transactions and relax the double-helix [90, 92, 162]. Similarly to prokaryotes, SC induces structural changes of DNA through twist/writhe modifications, including transitions to alternative DNA structures with regulatory potential [86, 102, 163]. Although less studied than in prokaryotes, SC also influences key steps of transcription, as underwound DNA promotes the binding of transcription complex, promoter melting and RNAP catalytic activity [102].



There are three different types of RNAP in eukaryotes depending on the number and type of sub-units they contain, as well as the type of RNAs they transcribe. They show similarities with bacterial RNAP in overall structure, in particular RNAP II, which transcribes mRNAs and some small regulatory RNAs [164]. As a consequence, prokaryotic and eukaryotic promoters also share common structural features (Fig. 6.3A), suggesting that mechanisms similar to the ones investigated may also be valid for eukaryotic promoters. Like bacterial RNAP which associates with a  $\sigma$  factor to initiate transcription, RNAP II cannot act alone and requires a set of initiation factors (TFIIB, IID, IIE, IIF and IIH) to form a large multiprotein-DNA complex (more details in [165, 166]). The best characterised core promoter elements, the TATA box and the Inr (initiator element), can act synergistically when separated by 25 to 30-bp, in which case transcription is enhanced by the cooperative binding of the TFIID complex to the two elements, or independently if separated by more than 30-bp [165]. It is consequently conceivable that SC, by modifying DNA helicity, may modulate the physical distance and orientation between these two elements, possibly promoting or preventing TFIID binding to the two elements, similarly to the mechanism investigated in Chapter 5 involving the orientation between -35 and -10 elements in bacteria. Besides, negative SC facilitates promoter opening also in eukaryotes, which is normally ensured by TFIIF containing an ATP-consuming helicase subunit [166]. For instance, the yeast *CUP1* promoter can be transcribed in a relaxed plasmid only if a set of TFs (including TFIIF) is present, whereas in the presence of negative SC, it can be transcribed in absence of any TF [167] (Fig. 6.3B). SC-dependent promoter opening could thus be an important regulator of transcription also in eukaryotes [102]. Regarding the mechanism investigated in Chapter 4 of SC-assisted promoter opening depending on discriminator G/C-content, a similar mechanism might also occur in eukaryotes, depending on the sequence of the Inr element which is part of the transcription bubble (Fig. 6.3A).

Finally, since transcription generates SC in virtually every organism, there is also a transcription-supercoiling coupling in eukaryotes for which a quantitative and systematic model was proposed [168]. It may account for the co-regulation of neighbour genes [102], with a key role of relative orientation between adjacent genes, as observed in prokaryotes [115].



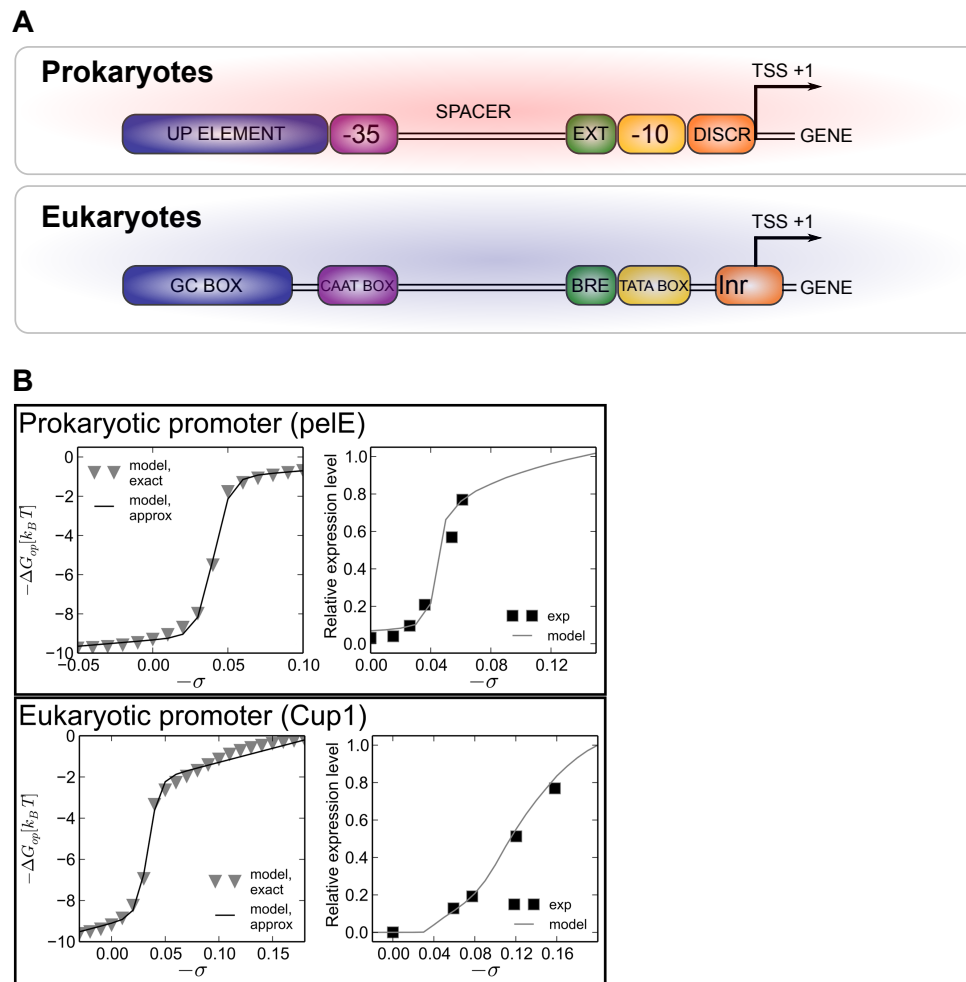


FIGURE 6.3: Common structural features between prokaryotic and RNAP II eukaryotic promoters, together with their SC-dependent opening profiles and expression levels. **(A)** Common structural features between prokaryotic and RNAP II eukaryotic promoters (different distance scales used). Additional information is provided in Fig. 1.2 for prokaryotic promoters, and in [165] for RNAP II eukaryotic ones. The promoter of the latter is composed of optional proximal control elements, the GC box around position -100 (relative to the TSS), and the CAAT box around position -80. The core promoter is composed of the BRE element around position -35, the TATA box around position -25, and the Inr around the TSS. **(B)** Copied from [168]): SC-dependent opening profiles and expression levels of (upper panel) *pelE* bacterial promoter from *D. dadantii*, and (lower panel) *CUP1* yeast promoter from *S. cerevisiae*.

Eukaryotic organelles possess their own genomes, and are thought to have been acquired through endosymbiotic events with bacterial ancestors [3]. It is thus not surprising that DNA gyrase, although absent from most eukaryotes, was found in parasites [169], and in the mitochondria and chloroplasts of plants [170–172] and algae [173]. In the red alga *Cyanidioschyzon merolae*, the replication of the chloroplast genome, but also that of the mitochondrial and nuclear genomes, is prevented by the gyrase inhibitor nalidixic acid [173, 174]. Similar observations have been made in plants, including *Arabidopsis thaliana* and *Nicotiana benthamiana*, based on studies employing various gyrase inhibitors or mutations [170–172, 175]. Altogether, these studies demonstrate that DNA gyrase plays an essential role in chloroplasts and, to a lower extent, in mitochondria, by ensuring DNA replication and partitioning, but also by maintaining negative SC and relaxing positive SC, although its exact functions in plant organelles and physiology remain to be elucidated. Besides, the ability of the plant *Nicotiana benthamiana* gyrase subunits to be substituted for their *E. coli* counterparts suggests that these organelles have retained bacterial-like mechanisms of SC regulation [170]. Thus, it is conceivable that within these eukaryotic organelles, fluctuations in DNA gyrase activity may occur, inducing global SC variations, and, in turn, global transcriptional changes through different regulatory mechanisms.

## 6.4 Applications of our modelling in archaea

Concerning archaea, the latter share a number of common features with bacteria, including the unicellularity, an asexual reproduction mode, the absence of organelles or internal membrane-bound structures such as a nucleus, the presence of a single and small circular genome, and a dense gene organisation into operons [176]. Not all archaea encode histone proteins and SMC condensin complexes, but all of them contain NAPs involved in genome compaction and organisation, but also in transcription regulation [177]. Histone and NAPs likely act as topological barriers to SC diffusion in archaea, following the same principles as in eukaryotes and bacteria [92]. There are also significant differences between bacteria and archaea, notably regarding their cell wall and membrane composition, as well as their central metabolism, since most archaea thrive in extreme environments [176]. Archaea are closer to eukaryotes regarding DNA replication [178] and transcription [177]. The single archaeal RNAP is homolog to the eukaryotic RNAP II, and both are

consequently dependent on similar promoter structures [177]. The extrapolation above of the investigated mechanisms in bacteria to eukaryotes is consequently also relevant for archaea, wherein DNA replication and transcription likewise generate SC, although not formally demonstrated.

Concerning their topoisomerases (reviewed extensively in [92, 179]), archaea have a class I, ATP-dependent TopoVI instead of a TopoI, which relaxes both positive and negative supercoils, while exhibiting a  $Mg^{2+}$ -dependent decatenase activity. They also have a class I TopoIII displaying an ATP-independent yet  $Mg^{2+}$ -dependent decatenase activity. Some of them have a DNA gyrase, while all hyperthermophilic archaea, *i.e.* that thrive in extreme temperatures ( $> 60^{\circ}C$ ), have a class I reverse gyrase that relaxes negative supercoils and introduces positive supercoils in DNA in an  $Mg^{2+}$  / ATP-dependent manner. As a consequence, the global SC level of the chromosome is found to be relaxed or positive in hyperthermophilic archaea, presumably to stabilise DNA and avoid its melting favoured by their high growth temperatures [180, 181]. Nevertheless, since positive SC is detrimental to DNA replication and transcription, these archaea must find a balance between DNA stability and the capacity to perform these essential processes. This may explain the presence of both gyrase and reverse gyrase in some hyperthermophilic archaea, contributing to the complex picture of genome topology in these organisms [179, 180].

There is also a balance between topoisomerase activity in archaea, similarly to the balance between TopoI and DNA gyrase activity in bacteria to maintain a negative chromosomal SC level. For mesophilic archaea, *i.e.* that thrive in temperatures between 20 to  $40^{\circ}C$ , chromosome is found to be negatively supercoiled at the global scale, while this level is even more negative in extreme halophiles, *i.e.* that thrive in high-salt concentrations, which may be explain by the fact that the latter tend to stabilise DNA [179]. Interestingly, global SC variations are also observed in response to environmental changes in archaea, similarly to bacteria. For instance, during heat and cold shock, transient and opposite SC changes are observed, in the same direction (DNA relaxation and overtwisting, respectively) in both mesophilic bacteria and hyperthermophilic archaea, despite different overall chromosomal SC levels (negative vs null to positive, respectively) [181] (Fig. 6.4). This suggests that the ability to modulate global SC level in response to environmental changes may be an universal phenomenon by which bacteria, and possibly archaea, adjust their global gene expression programs to adapt to new physiological

conditions.

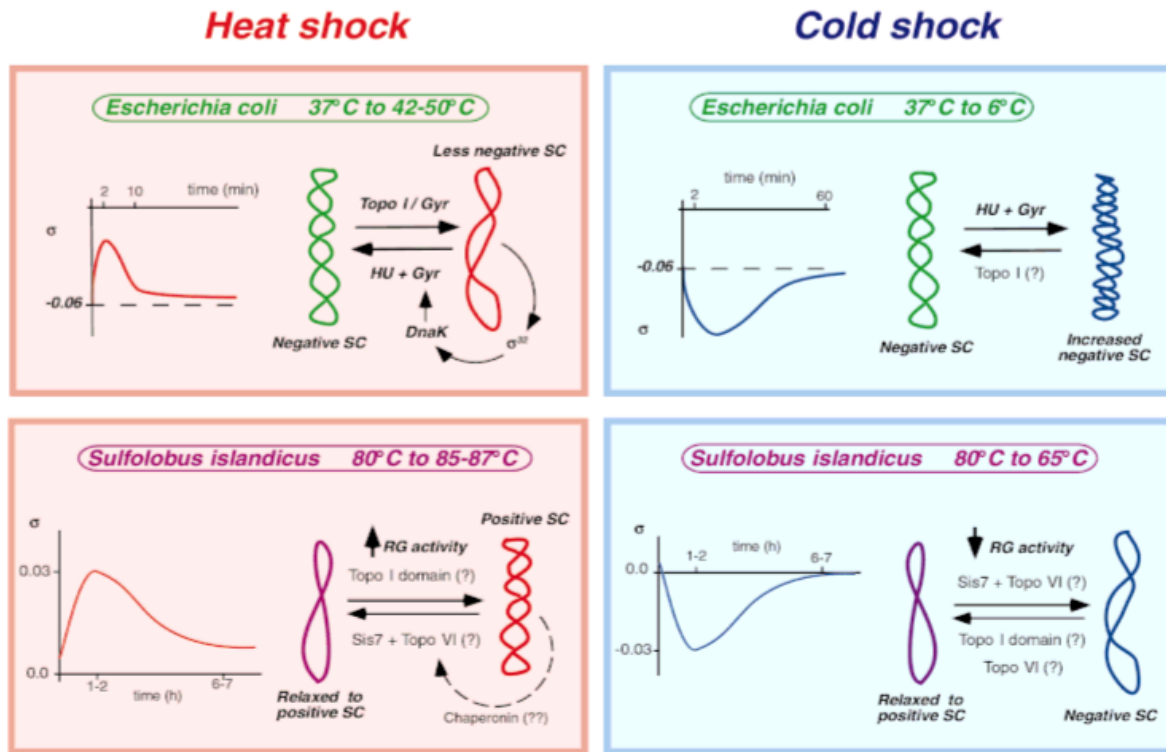


FIGURE 6.4: Copied from [181]: effects of heat and cold shock on plasmid topology in bacteria (*E. coli*) and in hyperthermophilic archaea (*Sulfolobus islandicus*). The time of exposure to the shock temperature and the SC variations associated are indicated. A schematic interpretation of the topological changes and the proteins involved in their regulation is depicted. Gyr: DNA gyrase, RG: reverse DNA gyrase.



# Bibliography

- [1] Thomas Cavalier-Smith, Martin Brasier, and T. Martin Embley. "Introduction: how and when did microbes change the world?" In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 361.1470 (2006), pp. 845–850.
- [2] Heinrich D. Holland. "The oxygenation of the atmosphere and oceans". In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 361.1470 (2006), pp. 903–915.
- [3] John M. Archibald. "Endosymbiosis and Eukaryotic Cell Evolution". In: *Current Biology* 25.19 (2015), R911–R921.
- [4] Jen Nguyen, Juanita Lara-Gutiérrez, and Roman Stocker. "Environmental fluctuations and their effects on microbial communities, populations and individuals". In: *FEMS Microbiology Reviews* fuaa068 (2020).
- [5] Patricia M. Schulte. "What is environmental stress? Insights from fish living in a variable environment". In: *Journal of Experimental Biology* 217.1 (2014), pp. 23–34.
- [6] Ary A. Hoffmann and Peter A. Parsons. *Evolutionary genetics and environmental stress*. Oxford science publications. Oxford ; New York: Oxford University Press, 1991.
- [7] Sylvie Reverchon and William Nasser. "Dickeya ecology, environment sensing and regulation of virulence programme". In: *Environmental Microbiology Reports* 5.5 (2013), pp. 622–636.
- [8] Aaron David Goldman and Laura F. Landweber. "What Is a Genome?" In: *PLOS Genetics* 12.7 (2016), e1006181.
- [9] Joshua Lederberg. "Plasmid (1952–1997)". In: *Plasmid* 39.1 (1998), pp. 1–9.
- [10] Paul C. Kirchberger, Marian L. Schmidt, and Howard Ochman. "The Ingenuity of Bacterial Genomes". In: *Annual Review of Microbiology* 74.1 (2020), pp. 815–834.
- [11] Louis-Marie Bobay and Howard Ochman. "The Evolution of Bacterial Genome Architecture". In: *Frontiers in Genetics* 8 (2017), p. 72.

- [12] Geoffrey M. Cooper. "The Complexity of Eukaryotic Genomes". In: *The Cell: A Molecular Approach*. 2nd edition (2000).
- [13] Chih-Horng Kuo, Nancy A. Moran, and Howard Ochman. "The consequences of genetic drift for bacterial genome complexity". In: *Genome Research* 19.8 (2009), pp. 1450–1454.
- [14] Sherwood Casjens. "The diverse and dynamic structure of bacterial genomes". In: *Annual Review of Genetics* 32 (1998), pp. 339–377.
- [15] Subhash C. Verma, Zhong Qian, and Sankar L. Adhya. "Architecture of the Escherichia coli nucleoid". In: *PLOS Genetics* 15.12 (2019), e1008456.
- [16] Marie Touchon, Claire Hoede, Olivier Tenaillon, Valérie Barbe, Simon Baeriswyl, Philippe Bidet, Edouard Bingen, Stéphane Bonacorsi, Christiane Bouchier, Odile Bouvet, Alexandra Calteau, Hélène Chiapello, Olivier Clermont, Stéphane Cruveiller, Antoine Danchin, Médéric Diard, Carole Dossat, Meriem El Karoui, Eric Frapy, Louis Garry, Jean Marc Ghigo, Anne Marie Gilles, James Johnson, Chantal Le Bouguéneq, Mathilde Lescat, Sophie Mangenot, Vanessa Martinez-Jéhanne, Ivan Matic, Xavier Nassif, Sophie Oztas, Marie Agnès Petit, Christophe Pichon, Zoé Rouy, Claude Saint Ruf, Dominique Schneider, Jérôme Turret, Benoit Vacherie, David Vallenet, Claudine Médigue, Eduardo P. C. Rocha, and Erick Denamur. "Organised Genome Dynamics in the Escherichia coli Species Results in Highly Diverse Adaptive Paths". In: *PLOS Genetics* 5.1 (2009), e1000344.
- [17] Katelyn McGary and Evgeny Nudler. "RNA polymerase and the ribosome: the close relationship". In: *Current Opinion in Microbiology* 16.2 (2013), pp. 112–117.
- [18] Emily F. Ruff, M. Thomas Record, and Irina Artsimovitch. "Initial Events in Bacterial Transcription Initiation". In: *Biomolecules* 5.2 (2015), pp. 1035–1062.
- [19] Finn Werner and Dina Grohmann. "Evolution of multisubunit RNA polymerases in the three domains of life". In: *Nature Reviews Microbiology* 9.2 (2011), pp. 85–98.
- [20] John D. Helmann. "Where to begin? Sigma factors and the selectivity of transcription initiation in bacteria". In: *Molecular Microbiology* 112.2 (2019), pp. 335–347.
- [21] Douglas F. Browning and Stephen J. W. Busby. "Local and global regulation of transcription initiation in bacteria". In: *Nature Reviews Microbiology* 14.10 (2016), pp. 638–650.



- [22] Indra Bervoets and Daniel Charlier. “Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology”. In: *FEMS Microbiology Reviews* 43.3 (2019), pp. 304–339.
- [23] Andrey Revyakin, Chenyu Liu, Richard H. Ebright, and Terence R. Strick. “Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching”. In: *Science* 314.5802 (2006), pp. 1139–1143.
- [24] Ananya Ray-Soni, Michael J. Bellecourt, and Robert Landick. “Mechanisms of Bacterial Transcription Termination: All Good Things Must End”. In: *Annual Review of Biochemistry* 85.1 (2016), pp. 319–347.
- [25] Sharmistha Banerjee, Jisha Chalissery, Irfan Bandey, and Ranjan Sen. “Rho-dependent Transcription Termination: More Questions than Answers”. In: *Journal of microbiology (Seoul, Korea)* 44.1 (2006), pp. 11–22.
- [26] Marina V. Rodnina. “Translation in Prokaryotes”. In: *Cold Spring Harbor Perspectives in Biology* 10.9 (2018), a032664.
- [27] Citlalli Mejía-Almonte, Stephen J. W. Busby, Joseph T. Wade, Jacques van Helden, Adam P. Arkin, Gary D. Stormo, Karen Eilbeck, Bernhard O. Palsson, James E. Galagan, and Julio Collado-Vides. “Redefining fundamental concepts of transcription initiation in bacteria”. In: *Nature Reviews Genetics* (2020), pp. 1–16.
- [28] Flora Picard, Clémentine Dressaire, Laurence Girbal, and Muriel Coccagn-Bousquet. “Examination of post-transcriptional regulations in prokaryotes by integrative biology”. In: *Comptes Rendus Biologies. Integrative Biology and Modelling in Agronomical Sciences* 332.11 (2009), pp. 958–973.
- [29] Elke Van Assche, Sandra Van Puyvelde, Jos Vanderleyden, and Hans P. Steenackers. “RNA-binding proteins involved in post-transcriptional regulation in bacteria”. In: *Frontiers in Microbiology* 6 (2015), p. 141.
- [30] Boris Macek, Karl Forchhammer, Julie Hardouin, Eilika Weber-Ban, Christophe Grangeasse, and Ivan Mijakovic. “Protein post-translational modifications in bacteria”. In: *Nature Reviews Microbiology* 17.11 (2019), pp. 651–664.
- [31] François Jacob and Jacques Monod. “Genetic regulatory mechanisms in the synthesis of proteins”. In: *Journal of Molecular Biology* 3 (1961), pp. 318–356.
- [32] D. Charlier, G. Weyens, M. Roovers, J. Piette, C. Bocquet, A. Piérard, and N. Glansdorff. “Molecular interactions in the control region of

- the *carAB* operon encoding *Escherichia coli* carbamoylphosphate synthetase". In: *Journal of Molecular Biology* 204.4 (1988), pp. 867–877.
- [33] Szabolcs Semsey, Mark Geanakopoulos, Dale E.A. Lewis, and Sankar Adhya. "Operator-bound GalR dimers close DNA loops by direct interaction: tetramerization and inducer binding". In: *The EMBO Journal* 21.16 (2002), pp. 4349–4356.
- [34] Maria Monsalve, Belén Calles, Mario Mencia, Fernando Rojo, and Margarita Salas. "Binding of phage 29 protein p4 to the early A2c promoter: recruitment of a repressor by the RNA polymerase". Edited by M. Gottesman. In: *Journal of Molecular Biology* 283.3 (1998), pp. 559–569.
- [35] Paulami Rudra, Ranjit Kumar Prajapati, Rajdeep Banerjee, Shreya Sengupta, and Jayanta Mukhopadhyay. "Novel mechanism of gene regulation: the protein Rv1222 of *Mycobacterium tuberculosis* inhibits transcription by anchoring the RNA polymerase onto DNA". In: *Nucleic Acids Research* 43.12 (2015), pp. 5855–5867.
- [36] Poul Valentin-Hansen, Lotte Sogaard-Andersen, and Henrik Pedersen. "A flexible partnership: the CytR anti-activator and the cAMP–CRP activator protein, comrades in transcription control". In: *Molecular Microbiology* 20.3 (1996), pp. 461–466.
- [37] Brian Benoff, Huanwang Yang, Catherine L. Lawson, Gary Parkinson, Jinsong Liu, Erich Blatter, Yon W. Ebright, Helen M. Berman, and Richard H. Ebright. "Structural basis of transcription activation: the CAP- $\alpha$  CTD-DNA complex". In: *Science (New York, N.Y.)* 297.5586 (2002), pp. 1562–1566.
- [38] E. Fic, P. Bonarek, A. Gorecki, S. Kedracka-Krok, J. Mikolajczak, A. Polit, M. Tworzydło, M. Dziedzicka-Wasylewska, and Z. Wasylewski. "cAMP receptor protein from *Escherichia coli* as a model of signal transduction in proteins—a review". In: *Journal of Molecular Microbiology and Biotechnology* 17.1 (2009), pp. 1–11.
- [39] Douglas F. Browning, Jeffrey A. Cole, and Stephen J. W. Busby. "Regulation by nucleoid-associated proteins at the *Escherichia coli* *nir* operon promoter". In: *Journal of Bacteriology* 190.21 (2008), pp. 7258–7267.
- [40] Sivaramesh Wigneshweraraj, Daniel Bose, Patricia C. Burrows, Nicolas Joly, Jörg Schumacher, Mathieu Rappas, Tillmann Pape, Xiaodong Zhang, Peter Stockley, Konstantin Severinov, and Martin Buck. "Modus operandi of the bacterial RNA polymerase containing the sigma54

- promoter-specificity factor". In: *Molecular Microbiology* 68.3 (2008), pp. 538–546.
- [41] A. J. Whitmarsh and R. J. Davis. "Regulation of transcription factor function by phosphorylation". In: *Cellular and molecular life sciences: CMLS* 57.8-9 (2000), pp. 1172–1183.
- [42] Regina Brigelius-Flohé and Leopold Flohé. "Basic Principles and Emerging Concepts in the Redox Control of Transcription Factors". In: *Antioxidants & Redox Signaling* 15.8 (2011), pp. 2335–2381.
- [43] Agnes Ullmann. "Escherichia coli Lactose Operon". In: *eLS*. American Cancer Society, 2009.
- [44] Fernando Rojo. "Repression of Transcription Initiation in Bacteria". In: *Journal of Bacteriology* 181.10 (1999), pp. 2987–2991.
- [45] T. Philip Malan, Annie Kolb, Henri Buc, and William R. McClure. "Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the P1 promoter". In: *Journal of Molecular Biology* 180.4 (1984), pp. 881–909.
- [46] Ingrid M. Keseler, Amanda Mackie, Alberto Santos-Zavaleta, Richard Billington, César Bonavides-Martínez, Ron Caspi, Carol Fulcher, Socorro Gama-Castro, Anamika Kothari, Markus Krummenacker, Mario Latendresse, Luis Muñoz-Rascado, Quang Ong, Suzanne Paley, Martin Peralta-Gil, Pallavi Subhraveti, David A. Velázquez-Ramírez, Daniel Weaver, Julio Collado-Vides, Ian Paulsen, and Peter D. Karp. "The EcoCyc database: reflecting new knowledge about Escherichia coli K-12". In: *Nucleic Acids Research* 45.Database issue (2017), pp. 543–550.
- [47] Lilia Brinza, Federica Calevro, and Hubert Charles. "Genomic analysis of the regulatory elements and links with intrinsic DNA structural properties in the shrunken genome of Buchnera". In: *BMC Genomics* 14.1 (2013), p. 73.
- [48] Ivan Junier, E. Besray Unal, Eva Yus, Verónica Lloréns-Rico, and Luis Serrano. "Insights into the Mechanisms of Basal Coordination of Transcription Using a Genome-Reduced Bacterium". In: *Cell Systems* 2.6 (2016), pp. 391–401.
- [49] Xindan Wang, Paula Montero Llopis, and David Z. Rudner. "Organization and segregation of bacterial chromosomes". In: *Nature Reviews Genetics* 14.3 (2013), pp. 191–203.
- [50] Virginia S. Lioy, Ivan Junier, and Frédéric Boccard. "Multiscale Dynamic Structuring of Bacterial Chromosomes". In: *Annual Review of Microbiology* 75 (2021), pp. 541–561.

- [51] Shane C. Dillon and Charles J. Dorman. "Bacterial nucleoid-associated proteins, nucleoid structure and gene expression". In: *Nature Reviews Microbiology* 8.3 (2010), pp. 185–195.
- [52] Tiago N. Cordeiro, Holger Schmidt, Cristina Madrid, Antonio Juárez, Pau Bernadó, Christian Griesinger, Jesús García, and Miquel Pons. "Indirect DNA Readout by an H-NS Related Protein: Structure of the DNA Complex of the C-Terminal Domain of Ler". In: *PLOS Pathogens* 7.11 (2011), e1002380.
- [53] Andrew Travers. "DNA–protein interactions: IHF - the master bend". In: *Current Biology* 7.4 (1997), R252–R254.
- [54] Sabrina Harteis and Sabine Schneider. "Making the Bend: DNA Tertiary Structure and Protein-DNA Interactions". In: *International Journal of Molecular Sciences* 15.7 (2014), pp. 12335–12363.
- [55] Joanna Hołówka and Jolanta Zakrzewska-Czerwińska. "Nucleoid Associated Proteins: The Small Organizers That Help to Cope With Stress". In: *Frontiers in Microbiology* 11 (2020), p. 590.
- [56] S. S. Antipov, M. N. Tutukina, E. V. Preobrazhenskaya, F. A. Kondrashov, M. V. Patrushev, S. V. Toshchakov, I. Dominova, U. S. Shvyreva, V. V. Vrublevskaya, O. S. Morenkov, N. A. Sukharicheva, V. V. Panyukov, and O. N. Ozoline. "The nucleoid protein Dps binds genomic DNA of *Escherichia coli* in a non-random manner". In: *PLOS ONE* 12.8 (2017), e0182800.
- [57] Martijn S. Luijsterburg, Maarten C. Noom, Gijs J. L. Wuite, and Remus Th Dame. "The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective". In: *Journal of Structural Biology* 156.2 (2006), pp. 262–272.
- [58] Charles J Dorman, Maria A Schumacher, Matthew J Bush, Richard G Brennan, and Mark J Buttner. "When is a transcription factor a NAP?" In: *Current Opinion in Microbiology. Cell Regulation* 55 (2020), pp. 26–33.
- [59] Kui Han, Zhi-feng Li, Ran Peng, Li-ping Zhu, Tao Zhou, Lu-guang Wang, Shu-guang Li, Xiao-bo Zhang, Wei Hu, Zhi-hong Wu, Nan Qin, and Yue-zhong Li. "Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu". In: *Scientific Reports* 3 (2013), p. 2101.
- [60] Hans Bremer and Patrick P. Dennis. "Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates". In: *EcoSal Plus* 3.1 (2008).

- [61] Dipankar Chatterji, Nobuyuki Fujita, and Akira Ishihama. "The mediator for stringent control, ppGpp, binds to the  $\sigma$ -subunit of Escherichia coli RNA polymerase". In: *Genes to Cells* 3.5 (1998), pp. 279–287.
- [62] Anna Perederina, Vladimir Svetlov, Marina N. Vassilyeva, Tahir H. Tahirov, Shigeyuki Yokoyama, Irina Artsimovitch, and Dmitry G. Vassilyev. "Regulation through the secondary channel–structural framework for ppGpp-DksA synergism during transcription". In: *Cell* 118.3 (2004), pp. 297–309.
- [63] Vasili Hauryliuk, Gemma C. Atkinson, Katsuhiko S. Murakami, Tanel Tenson, and Kenn Gerdes. "Recent functional insights into the role of (p)ppGpp in bacterial physiology". In: *Nature reviews. Microbiology* 13.5 (2015), pp. 298–309.
- [64] Lisa U. Magnusson, Anne Farewell, and Thomas Nyström. "ppGpp: a global regulator in Escherichia coli". In: *Trends in Microbiology* 13.5 (2005), pp. 236–242.
- [65] Patricia Sanchez-Vazquez, Colin N. Dewey, Nicole Kitten, Wilma Ross, and Richard L. Gourse. "Genome-wide effects on Escherichia coli transcription from ppGpp binding to its two sites on RNA polymerase". In: *Proceedings of the National Academy of Sciences of the United States of America* 116.17 (2019), pp. 8310–8319.
- [66] Josep Casadesús and David Low. "Epigenetic Gene Regulation in the Bacterial World". In: *Microbiology and Molecular Biology Reviews* 70.3 (2006), pp. 830–856.
- [67] Shanil P. Haugen, Wilma Ross, and Richard L. Gourse. "Advances in bacterial promoter recognition and its control by factors that do not bind DNA". In: *Nature Reviews Microbiology* 6.7 (2008), pp. 507–519.
- [68] Hidde de Jong. "Modeling and simulation of genetic regulatory systems: a literature review". In: *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 9.1 (2002), pp. 67–103.
- [69] Ahmet Ay and David N. Arnosti. "Mathematical modeling of gene expression: a guide for the perplexed biologist". In: *Critical reviews in biochemistry and molecular biology* 46.2 (2011), pp. 137–151.
- [70] Michael Hecker, Sandro Lambeck, Susanne Toepfer, Eugene van Someren, and Reinhard Guthke. "Gene regulatory network inference: data integration in dynamic models—a review". In: *Bio Systems* 96.1 (2009), pp. 86–103.

- [71] Michael C. Mackey, Moisés Santillán, and Necmettin Yildirim. “Modeling operon dynamics: the tryptophan and lactose operons as paradigms”. In: *Comptes Rendus Biologies* 327.3 (2004), pp. 211–224.
- [72] Xiao-jiang Feng, Sara Hooshangi, David Chen, Genyuan Li, Ron Weiss, and Herschel Rabitz. “Optimizing Genetic Circuits by Global Sensitivity Analysis”. In: *Biophysical Journal* 87.4 (2004), pp. 2195–2202.
- [73] Bin Huang, Dongya Jia, Jingchen Feng, Herbert Levine, José N. Onuchic, and Mingyang Lu. “RACIPE: a computational tool for modeling gene regulatory circuits using randomization”. In: *BMC Systems Biology* 12.1 (2018), p. 74.
- [74] Hernan G. Garcia, Jane Kondev, Nigel Orme, Julie A. Theriot, and Rob Phillips. “Thermodynamics of Biological Processes”. In: *Methods in Enzymology* 492 (2011), pp. 27–59.
- [75] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terence Hwa, Jané Kondev, and Rob Phillips. “Transcriptional regulation by the numbers: models”. In: *Current opinion in genetics & development* 15.2 (2005), pp. 116–124.
- [76] Rob Phillips, Nathan M. Belliveau, Griffin Chure, Hernan G. Garcia, Manuel Razo-Mejia, and Clarissa Scholes. “Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression”. In: *Annual Review of Biophysics* 48.1 (2019), pp. 121–163.
- [77] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terence Hwa, Jané Kondev, Thomas Kuhlman, and Rob Phillips. “Transcriptional regulation by the numbers: applications”. In: *Current opinion in genetics & development* 15.2 (2005), pp. 125–135.
- [78] Thomas Kuhlman, Zhongge Zhang, Milton H. Saier, and Terence Hwa. “Combinatorial transcriptional control of the lactose operon of *Escherichia coli*”. In: *Proceedings of the National Academy of Sciences* 104.14 (2007), pp. 6043–6048.
- [79] Madeline A. Shea and Gary K. Ackers. “The OR control system of bacteriophage lambda: A physical-chemical model for gene regulation”. In: *Journal of Molecular Biology* 181.2 (1985), pp. 211–230.
- [80] Gary K. Ackers, Alexander D. Johnson, and Madeline A. Shea. “Quantitative model for gene regulation by lambda phage repressor”. In: *Proceedings of the National Academy of Sciences* 79.4 (1982), pp. 1129–1133.

- [81] Muir Morrison, Manuel Razo-Mejia, and Rob Phillips. “Reconciling kinetic and thermodynamic models of bacterial transcription”. In: *PLoS Computational Biology* 17.1 (2021), e1008572.
- [82] James D. Watson and Francis H. C. Crick. “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid”. In: *Nature* 171.4356 (1953), pp. 737–738.
- [83] James D. Watson and Francis H. C. Crick. “Genetical Implications of the Structure of Deoxyribonucleic Acid”. In: *Nature* 171.4361 (1953), pp. 964–967.
- [84] Fedor Kouzine, David Levens, and Laura Baranello. “DNA topology and transcription”. In: *Nucleus* 5.3 (2014), pp. 195–202.
- [85] Sergei M. Mirkin. “DNA Topology: Fundamentals”. In: *eLS*. American Cancer Society, 2001.
- [86] Dina Zhabinskaya and Craig J. Benham. “Theoretical Analysis of Competing Conformational Transitions in Superhelical DNA”. In: *PLoS Computational Biology* 8.4 (2012).
- [87] Vladimir N. Potaman and Richard R. Sinden. *DNA: Alternative Conformations and Biology*. Landes Bioscience, 2013.
- [88] N. Patrick Higgins and Alexander V. Vologodskii. “Topological Behavior of Plasmid DNA”. In: *Microbiology spectrum* 3.2 (2015).
- [89] Nitu Kumari and Sathees C. Raghavan. “G-quadruplex DNA structures and their relevance in radioprotection”. In: *Biochimica et Biophysica Acta (BBA) - General Subjects* 1865.5 (2021), p. 129857.
- [90] Patrick Forterre, Simonetta Gribaldo, Danièle Gadelle, and Marie-Claude Serre. “Origin and evolution of DNA topoisomerases”. In: *Biochimie* 89.4 (2007), pp. 427–446.
- [91] Shannon J. McKie, Keir C. Neuman, and Anthony Maxwell. “DNA topoisomerases: Advances in understanding of cellular roles and multi-protein complexes via structure-function analysis”. In: *BioEssays* 43.4 (2021), p. 2000286.
- [92] Alexandre Duprey and Eduardo A. Groisman. “The regulation of DNA supercoiling across evolution”. In: *Protein Science: A Publication of the Protein Society* (2021).
- [93] Ksenia Terekhova, Kathryn H. Gunn, John F. Marko, and Alfonso Mondragón. “Bacterial topoisomerase I and topoisomerase III relax supercoiled DNA via distinct pathways”. In: *Nucleic Acids Research* 40.20 (2012), pp. 10432–10440.



- [94] Xindan Wang, Rodrigo Reyes-Lamothe, and David J. Sherratt. "Modulation of *Escherichia coli* sister chromosome cohesion by topoisomerase IV". In: *Genes & Development* 22.17 (2008), pp. 2426–2433.
- [95] Lisa Postow, Christine D. Hardy, Javier Arsuaga, and Nicholas R. Cozzarelli. "Topological domain structure of the *Escherichia coli* chromosome". In: *Genes & Development* 18.14 (2004), pp. 1766–1779.
- [96] Tung B. K. Le, Maxim V. Imakaev, Leonid A. Mirny, and Michael T. Laub. "High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome". In: *Science* 342.6159 (2013), pp. 731–734.
- [97] Marc Joyeux and Ivan Junier. "Requirements for DNA-Bridging Proteins to Act as Topological Barriers of the Bacterial Genome". In: *Biophysical Journal* 119.6 (2020), pp. 1215–1225.
- [98] James B. Bliska and Nicholas R. Cozzarelli. "Use of site-specific recombination as a probe of DNA structure and metabolism in vivo". In: *Journal of Molecular Biology* 194.2 (1987), pp. 205–218.
- [99] Yingting Liu, Zhi-Chun Hua, and Fenfei Leng. "DNA supercoiling measurement in bacteria". In: *Methods in molecular biology (Clifton, N.J.)* 1703 (2018), pp. 63–73.
- [100] Nikolay S. Rovinskiy, Andrews A. Agbleke, Olga N. Chesnokova, and N. Patrick Higgins. "Supercoil Levels in *E. coli* and *Salmonella* Chromosomes Are Regulated by the C-Terminal 35–38 Amino Acids of GyrA". In: *Microorganisms* 7.3 (2019).
- [101] Alexandre Duprey and Eduardo A. Groisman. "FEDS: a Novel Fluorescence-Based High-Throughput Method for Measuring DNA Supercoiling In Vivo". In: *mBio* 11.4 (2020), e01053–20.
- [102] Samuel Corless and Nick Gilbert. "Effects of DNA supercoiling on chromatin architecture". In: *Biophysical Reviews* 8.3 (2016), pp. 245–258.
- [103] Elizabeth G. Gibson, Alexandria A. Oviatt, and Neil Osheroff. "Two-Dimensional Gel Electrophoresis to Resolve DNA Topoisomers". In: *DNA Electrophoresis: Methods and Protocols*. Methods in Molecular Biology. New York, NY: Springer US, 2020, pp. 15–24.
- [104] Avantika Lal, Amlanjyoti Dhar, Andrei Trostel, Fedor Kouzine, Aswin S. N. Seshasayee, and Sankar Adhya. "Genome scale patterns of supercoiling in a bacterial chromosome". In: *Nature Communications* 7.1 (2016), p. 11055.
- [105] Karl Drlica. "Control of bacterial DNA supercoiling". In: *Molecular Microbiology* 6.4 (1992), pp. 425–433.

- [106] Andrew Travers and Georgi Muskhelishvili. "DNA supercoiling - a global transcriptional regulator for enterobacterial growth?" In: *Nature Reviews. Microbiology* 3.2 (2005), pp. 157–169.
- [107] Charles J. Dorman and Matthew J. Dorman. "DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression". In: *Biophysical Reviews* 8.Suppl 1 (2016), pp. 89–100.
- [108] Shiny Martis B., Raphaël Forquet, Sylvie Reverchon, William Nasser, and Sam Meyer. "DNA Supercoiling: an Ancestral Regulator of Gene Expression in Pathogenic Bacteria?" In: *Computational and Structural Biotechnology Journal* 17 (2019), pp. 1047–1055.
- [109] Nicolas Blot, Ramesh Mavathur, Marcel Geertz, Andrew Travers, and Georgi Muskhelishvili. "Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome". In: *EMBO Reports* 7.7 (2006), pp. 710–715.
- [110] Tyrrell Conway, James P. Creecy, Scott M. Maddox, Joe E. Grissom, Trevor L. Conkle, Tyler M. Shadid, Jun Teramoto, Phillip San Miguel, Tomohiro Shimada, Akira Ishihama, Hirotada Mori, and Barry L. Wanner. "Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing". In: *mBio* 5.4 (2014), pp. 01442–01414.
- [111] Yan Zhang, David H. Burkhardt, Silvi Rouskin, Gene-Wei Li, Jonathan S. Weissman, and Carol A. Gross. "A Stress Response that Monitors and Regulates mRNA Structure Is Central to Cold Shock Adaptation". In: *Molecular Cell* 70.2 (2018), 274–286.e7.
- [112] Szymon Jozefczuk, Sebastian Klie, Gareth Catchpole, Jędrzej Szymaniński, Alvaro Cuadros-Inostroza, Dirk Steinhauser, Joachim Selbig, and Lothar Willmitzer. "Metabolomic and transcriptomic stress response of *Escherichia coli*". In: *Molecular Systems Biology* 6 (2010), p. 364.
- [113] Alexander Bartholomäus, Ivan Fedyunin, Peter Feist, Celine Sin, Gong Zhang, Angelo Valleriani, and Zoya Ignatova. "Bacteria differently regulate mRNA abundance to specifically respond to various stresses". In: *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 374.2063 (2016).
- [114] Geetha Kannan, Jessica C. Wilks, Devon M. Fitzgerald, Brian D. Jones, Sandra S. Bondurant, and Joan L. Slonczewski. "Rapid acid treatment of *Escherichia coli*: transcriptomic response and recovery". In: *BMC microbiology* 8 (2008), p. 37.

- [115] Bilal El Houdaigui, Raphaël Forquet, Thomas Hindré, Dominique Schneider, William Nasser, Sylvie Reverchon, and Sam Meyer. “Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling”. In: *Nucleic Acids Research* 47.11 (2019), pp. 5648–5657.
- [116] Natalia E. Gogoleva, Tatiana A. Konnova, Alexander S. Balkin, Andrey O. Plotnikov, and Yuri V. Gogolev. “Transcriptomic data of *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. 14028S treated with novobiocin”. In: *Data in Brief* 29 (2020).
- [117] Carsten Kröger, Shane C. Dillon, Andrew D. S. Cameron, Kai Papenfort, Sathesh K. Sivasankaran, Karsten Hokamp, Yanjie Chao, Alexandra Sittka, Magali Hébrard, Kristian Händler, Aoife Colgan, Pimplapas Leekitcharoenphon, Gemma C. Langridge, Amanda J. Lohan, Brendan Loftus, Sacha Lucchini, David W. Ussery, Charles J. Dorman, Nicholas R. Thomson, Jörg Vogel, and Jay C. D. Hinton. “The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium”. In: *Proceedings of the National Academy of Sciences of the United States of America* 109.20 (2012), pp. 1277–1286.
- [118] Mark A. Webber, Vito Ricci, Rebekah Whitehead, Meha Patel, Maria Fookes, Alasdair Ivens, and Laura J. V. Piddock. “Clinically relevant mutant DNA gyrase alters supercoiling, changes the transcriptome, and confers multidrug resistance”. In: *mBio* 4.4 (2013), e00273–13.
- [119] Xuejiao Jiang, Patrick Sobetzko, William Nasser, Sylvie Reverchon, and Georgi Muskhelishvili. “Chromosomal “Stress-Response” Domains Govern the Spatiotemporal Expression of the Bacterial Virulence Program”. In: *mBio* 6.3 (2015), e00353–15.
- [120] Raphael Forquet, Xuejiao Jiang, William Nasser, Florence Hommais, Sylvie Reverchon, and Sam Meyer. “Mapping the complex transcriptional landscape of the phytopathogenic bacterium *Dickeya dadantii*”. In: (2021), p. 2020.09.30.320440.
- [121] Maïwenn Pineau, Shiny Martis B, Raphaël Forquet, Jessica Baude, Lucie Grand, Florence Popowycz, Laurent Soulère, Florence Hommais, William Nasser, Sylvie Reverchon, and Sam Meyer. “What is a supercoiling-sensitive gene? Insights from topoisomerase I inhibition in the Gram-negative bacterium *Dickeya dadantii*”. In: *bioRxiv* (2021), p. 2021.04.09.439150.
- [122] Sylvie Reverchon, Sam Meyer, Raphaël Forquet, Florence Hommais, Georgi Muskhelishvili, and William Nasser. “The nucleoid-associated

- protein IHF acts as a ‘transcriptional domainin’ protein coordinating the bacterial virulence traits with global transcription”. In: *Nucleic Acids Research* 49.2 (2021), pp. 776–790.
- [123] Vikram Vijayan, Rick Zuzow, and Erin K. O’Shea. “Oscillations in supercoiling drive circadian gene expression in cyanobacteria”. In: *Proceedings of the National Academy of Sciences of the United States of America* 106.52 (2009), pp. 22564–22568.
- [124] Vikram Vijayan, Isha H. Jain, and Erin K. O’Shea. “A high resolution map of a cyanobacterial transcriptome”. In: *Genome Biology* 12.5 (2011), p. 47.
- [125] Ilham Ayub Shahmuradov, Rozaimi Mohamad Razali, Salim Bougouffa, Aleksandar Radovanovic, and Vladimir B Bajic. “bTSSfinder: a novel tool for the prediction of promoters in cyanobacteria and *Escherichia coli*”. In: *Bioinformatics* 33.3 (2017), pp. 334–340.
- [126] Leroy F. Liu and James C. Wang. “Supercoiling of the DNA template during transcription.” In: *Proceedings of the National Academy of Sciences of the United States of America* 84.20 (1987), pp. 7024–7027.
- [127] Charles J. Dorman. “DNA supercoiling and transcription in bacteria: a two-way street”. In: *BMC Molecular and Cell Biology* 20 (2019), p. 26.
- [128] Laurent Moulin, A. Rachid Rahmouni, and Frédéric Boccard. “Topological insulators inhibit diffusion of transcription-induced positive supercoils in the chromosome of *Escherichia coli*”. In: *Molecular Microbiology* 55.2 (2005), pp. 601–610.
- [129] Michael L. Opel and G. Wesley Hatfield. “DNA supercoiling-dependent transcriptional coupling between the divergently transcribed promoters of the *ilvYC* operon of *Escherichia coli* is proportional to promoter strengths and transcript lengths”. In: *Molecular Microbiology* 39.1 (2001), pp. 191–198.
- [130] Michael W. Mangan, Sacha Lucchini, Vittoria Danino, Tadhg O. Cróinín, Jay C. D. Hinton, and Charles J. Dorman. “The integration host factor (IHF) integrates stationary-phase and virulence gene expression in *Salmonella enterica* serovar Typhimurium”. In: *Molecular Microbiology* 59.6 (2006), pp. 1831–1847.
- [131] Georgi Muskhelishvili, Raphaël Forquet, Sylvie Reverchon, Sam Meyer, and William Nasser. “Coherent Domains of Transcription Coordinate Gene Expression During Bacterial Growth and Adaptation”. In: *Microorganisms* 7.12 (2019), p. 694.

- [132] Nicole Hugouvieux-Cotte-Pattat, Guy Condemine, Erwan Gueguen, and Vladimir E. Shevchik. "Dickeya Plant Pathogens". In: *eLS*. American Cancer Society, 2020, pp. 1–10.
- [133] S. Reverchon, G. Muskhelishvili, and W. Nasser. "Virulence Program of a Bacterial Plant Pathogen: The Dickeya Model". In: *Progress in Molecular Biology and Translational Science* 142 (2016), pp. 51–92.
- [134] Anne-Marie Grenier, Gabrielle Duport, Sylvie Pagès, Guy Condemine, and Yvan Rahbé. "The Phytopathogen *Dickeya dadantii* (*Erwinia chrysanthemi* 3937) Is a Pathogen of the Pea Aphid". In: *Applied and Environmental Microbiology* 72.3 (2006), pp. 1956–1965.
- [135] Jeremy D. Glasner, Ching-Hong Yang, Sylvie Reverchon, Nicole Hugouvieux-Cotte-Pattat, Guy Condemine, Jean-Pierre Bohin, Frédérique Van Gijsegem, Shihui Yang, Thierry Franza, Dominique Expert, Guy Plunkett, Michael J. San Francisco, Amy O. Charkowski, Béatrice Py, Kenneth Bell, Lise Rauscher, Pablo Rodriguez-Palenzuela, Ariane Tous-saint, Maria C. Holeva, Sheng Yang He, Vanessa Douet, Martine Boc-cara, Carlos Blanco, Ian Toth, Bradley D. Anderson, Bryan S. Biehl, Bob Mau, Sarah M. Flynn, Frédéric Barras, Magdalen Lindeberg, Paul R. J. Birch, Shinji Tsuyumu, Xiangyang Shi, Michael Hibbing, Mee-Ngan Yap, Mathilde Carpentier, Elie Dassa, Masahiro Umehara, Ji-hyun F. Kim, Michael Rusch, Pritin Soni, George F. Mayhew, Derrick E. Fouts, Steven R. Gill, Frederick R. Blattner, Noel T. Keen, and Nicole T. Perna. "Genome Sequence of the Plant-Pathogenic Bacterium *Dickeya dadantii* 3937". In: *Journal of Bacteriology* 193.8 (2011), pp. 2076–2077.
- [136] Zghidi-Abouzid Ouafa, Sylvie Reverchon, Thomas Lautier, Georgi Muskhelishvili, and William Nasser. "The nucleoid-associated proteins H-NS and FIS modulate the DNA supercoiling response of the *pel* genes, the major virulence factors in the plant pathogen bacterium *Dickeya dadantii*". In: *Nucleic Acids Research* 40.10 (2012), pp. 4306–4319.
- [137] C Grignon and H Sentenac. "pH and Ionic Conditions in the Apoplast". In: *Annual Review of Plant Physiology and Plant Molecular Biology* 42.1 (1991), pp. 103–128.
- [138] Chris Lamb and Richard A. Dixon. "The Oxidative Burst in Plant Disease Resistance". In: *Annual Review of Plant Physiology and Plant Molecular Biology* 48.1 (1997), pp. 251–275.
- [139] Olivier Tenaillon, Jeffrey E. Barrick, Noah Ribeck, Daniel E. Deather-age, Jeffrey L. Blanchard, Aurko Dasgupta, Gabriel C. Wu, Sébastien

- Wielgoss, Stéphane Cruveiller, Claudine Médigue, Dominique Schneider, and Richard E. Lenski. "Tempo and mode of genome evolution in a 50,000-generation experiment". In: *Nature* 536.7615 (2016), pp. 165–170.
- [140] Estelle Crozat, Nadège Philippe, Richard E. Lenski, Johannes Geiselman, and Dominique Schneider. "Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection". In: *Genetics* 169.2 (2005), pp. 523–532.
- [141] Richard E. Lenski. "Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations". In: *The ISME Journal* 11.10 (2017), pp. 2181–2194.
- [142] Santiago F. Elena and Richard E. Lenski. "Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation". In: *Nature Reviews. Genetics* 4.6 (2003), pp. 457–469.
- [143] Steven Bellon, Jonathan D. Parsons, Yunyi Wei, Koto Hayakawa, Lora L. Swenson, Paul S. Charifson, Judith A. Lippke, Robert Aldape, and Christian H. Gross. "Crystal Structures of *Escherichia coli* Topoisomerase IV ParE Subunit (24 and 43 Kilodaltons): a Single Residue Dictates Differences in Novobiocin Potency against Topoisomerase IV and DNA Gyrase". In: *Antimicrobial Agents and Chemotherapy* 48.5 (2004), pp. 1856–1864.
- [144] Raphaël Forquet, Maïwenn Pineau, William Nasser, Sylvie Reverchon, and Sam Meyer. "Role of the Discriminator Sequence in the Supercoiling Sensitivity of Bacterial Promoters". In: *mSystems* 6.4 (2021), e0097821.
- [145] María Teresa García, María Amparo Blázquez, María José Ferrándiz, María Jesús Sanz, Noella Silva-Martín, Juan A. Hermoso, and Adela G. de la Campa. "New alkaloid antibiotics that target the DNA topoisomerase I of *Streptococcus pneumoniae*". In: *The Journal of Biological Chemistry* 286.8 (2011), pp. 6402–6413.
- [146] María-José Ferrándiz, Antonio J. Martín-Galiano, Cristina Arnanz, Isabel Camacho-Soguero, José-Manuel Tirado-Vélez, and Adela G. de la Campa. "An increase in negative supercoiling in bacteria reveals topology-reacting gene clusters and a homeostatic response mediated by the DNA topoisomerase I gene". In: *Nucleic Acids Research* 44.15 (2016), pp. 7292–7303.

- [147] Daniel Ryan, Laura Jenniches, Sarah Reichardt, Lars Barquist, and Alexander J. Westermann. "A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe *Bacteroides thetaiotaomicron*". In: *Nature Communications* 11.1 (2020), p. 3557.
- [148] Alexander J. Westermann, Stanislaw A. Gorski, and Jörg Vogel. "Dual RNA-seq of pathogen and host". In: *Nature Reviews Microbiology* 10.9 (2012), pp. 618–630.
- [149] Simon Leonard, Sam Meyer, Stephan Lacour, William Nasser, Florence Hommais, and Sylvie Reverchon. "APER0: a genome-wide approach for identifying bacterial small RNAs from RNA-Seq data". In: *Nucleic Acids Research* 47.15 (2019), e88.
- [150] Lukas Rajkowitsch, Doris Chen, Sabine Stampfl, Katharina Semrad, Christina Waldsich, Oliver Mayer, Michael F. Jantsch, Robert Konrat, Udo Bläsi, and Renée Schroeder. "RNA chaperones, RNA annealers and RNA helicases". In: *RNA biology* 4.3 (2007), pp. 118–130.
- [151] Simon Leonard, Camille Villard, William Nasser, Sylvie Reverchon, and Florence Hommais. "RNA Chaperones Hfq and ProQ Play a Key Role in the Virulence of the Plant Pathogenic Bacterium *Dickeya dadantii*". In: *Frontiers in Microbiology* 12 (2021), p. 687484.
- [152] Jasmin Cevost, Cédric Vaillant, and Sam Meyer. "ThreaDNA: predicting DNA mechanics' contribution to sequence selectivity of proteins along whole genomes". In: *Bioinformatics* 34.4 (2018), pp. 609–616.
- [153] Monica S Guo, Ryo Kawamura, Megan L Littlehale, John F Marko, and Michael T Laub. "High-resolution, genome-wide mapping of positive supercoiling in chromosomes". In: *eLife* 10 (2021), e67236.
- [154] Frederic Chedin and Craig J. Benham. "Emerging roles for R-loop structures in the management of topological stress". In: *The Journal of Biological Chemistry* 295.14 (2020), pp. 4684–4695.
- [155] José M. Santos-Pereira and Andrés Aguilera. "R loops: new modulators of genome dynamics and function". In: *Nature Reviews. Genetics* 16.10 (2015), pp. 583–597.
- [156] Marc Drolet. "Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology". In: *Molecular Microbiology* 59.3 (2006), pp. 723–730.
- [157] Christophe Lavelle. "Pack, unpack, bend, twist, pull, push: the physical side of gene expression". In: *Current Opinion in Genetics & Development* 25 (2014), pp. 74–84.



- [158] Nick Gilbert and James Allan. "Supercoiling in DNA and chromatin". In: *Current Opinion in Genetics & Development* 25.100 (2014), pp. 15–21.
- [159] Catherine Naughton, Nicolaos Avlonitis, Samuel Corless, James G. Prendergast, Ioulia K. Mati, Paul P. Eijk, Scott L. Cockroft, Mark Bradley, Bauke Ylstra, and Nick Gilbert. "Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures". In: *Nature structural & molecular biology* 20.3 (2013), pp. 387–395.
- [160] Jie Ma and Michelle D. Wang. "RNA polymerase is a powerful torsional motor". In: *Cell Cycle (Georgetown, Tex.)* 13.3 (2014), pp. 337–338.
- [161] Fedor Kouzine, Ashutosh Gupta, Laura Baranello, Damian Wojtowicz, Khadija Benaissa, Juhong Liu, Teresa M. Przytycka, and David Levens. "Transcription dependent dynamic supercoiling is a short-range genomic force". In: *Nature structural & molecular biology* 20.3 (2013), pp. 396–403.
- [162] Yves Pommier, Yilun Sun, Shar-yin N. Huang, and John L. Nitiss. "Roles of eukaryotic topoisomerases in transcription, replication and genomic stability". In: *Nature Reviews Molecular Cell Biology* 17.11 (2016), pp. 703–721.
- [163] Václav Brázda, Rob C. Laister, Eva B. Jagelská, and Cheryl Arrow-smith. "Cruciform structures are a common DNA feature important for regulating biological processes". In: *BMC molecular biology* 12 (2011), p. 33.
- [164] Richard H. Ebright. "RNA Polymerase: Structural Similarities Between Bacterial RNA Polymerase and Eukaryotic RNA Polymerase II". In: *Journal of Molecular Biology* 304.5 (2000), pp. 687–698.
- [165] Stephen T. Smale and James T. Kadonaga. "The RNA polymerase II core promoter". In: *Annual Review of Biochemistry* 72 (2003), pp. 449–479.
- [166] D. B. Nikolov and S. K. Burley. "RNA polymerase II transcription initiation: A structural view". In: *Proceedings of the National Academy of Sciences* 94.1 (1997), pp. 15–22.
- [167] Benoit P. Leblanc, Craig J. Benham, and David J. Clark. "An initiation element in the yeast CUP1 promoter is recognized by RNA polymerase II in the absence of TATA box-binding protein if the DNA is negatively supercoiled". In: *Proceedings of the National Academy of Sciences of the United States of America* 97.20 (2000), pp. 10745–10750.

- [168] Sam Meyer and Guillaume Beslon. "Torsion-Mediated Interaction between Adjacent Genes". In: *PLoS Computational Biology* 10.9 (2014), e1003785.
- [169] Mohd Ashraf Dar, Atul Sharma, Neelima Mondal, and Suman Kumar Dhar. "Molecular Cloning of Apicoplast-Targeted Plasmodium falciparum DNA Gyrase Genes: Unique Intrinsic ATPase Activity and ATP-Independent Dimerization of PfGyrB Subunit". In: *Eukaryotic Cell* 6.3 (2007), pp. 398–412.
- [170] Hye Sun Cho, Sang Sook Lee, Kwang Dong Kim, Inhwan Hwang, Jong-Seok Lim, Youn-Il Park, and Hyun-Sook Pai. "DNA Gyrase Is Involved in Chloroplast Nucleoid Partitioning". In: *The Plant Cell* 16.10 (2004), pp. 2665–2682.
- [171] Melisa K. Wall, Lesley A. Mitchenall, and Anthony Maxwell. "Arabidopsis thaliana DNA gyrase is targeted to chloroplasts and mitochondria". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.20 (2004), pp. 7821–7826.
- [172] Katherine M. Evans-Roberts, Lesley A. Mitchenall, Melisa K. Wall, Julie Leroux, Joshua S. Mylne, and Anthony Maxwell. "DNA Gyrase Is the Target for the Quinolone Drug Ciprofloxacin in Arabidopsis thaliana". In: *The Journal of Biological Chemistry* 291.7 (2016), pp. 3136–3144.
- [173] Roichi Itoh, Hidenori Takahashi, Kyoko Toda, Haruko Kuroiwa, and Tsuneyoshi Kuroiwa. "DNA gyrase involvement in chloroplast-nucleoid division in *Cyanidioschyzon merolae*". In: *European Journal of Cell Biology* 73.3 (1997), pp. 252–258.
- [174] Yuki Kobayashi, Yu Kanasaki, Ayumi Tanaka, Haruko Kuroiwa, Tsuneyoshi Kuroiwa, and Kan Tanaka. "Tetrapyrrole signal as a cell-cycle coordinator from organelle to nuclear DNA replication in plant cells". In: *Proceedings of the National Academy of Sciences* 106.3 (2009), pp. 803–807.
- [175] Takashi Moriyama and Naoki Sato. "Enzymes involved in organellar DNA replication in photosynthetic eukaryotes". In: *Frontiers in Plant Science* 5 (2014), p. 480.
- [176] Wolfram Zillig. "Comparative biochemistry of Archaea and Bacteria". In: *Current Opinion in Genetics & Development* 1.4 (1991), pp. 544–551.
- [177] Breanna R. Wenck and Thomas J. Santangelo. "Archaeal transcription". In: *Transcription* 11.5 (2020), pp. 199–210.

- [178] Elizabeth R. Barry and Stephen D. Bell. “DNA Replication in the Archaea”. In: *Microbiology and Molecular Biology Reviews* 70.4 (2006), pp. 876–887.
- [179] Florence Garnier, Mohea Couturier, H el ene D ebat, and Marc Nadal. “Archaea: A Gold Mine for Topoisomerase Diversity”. In: *Frontiers in Microbiology* 12 (2021), p. 1083.
- [180] Anna Valenti, Giuseppe Perugino, Mos  Rossi, and Maria Ciaramella. “Positive supercoiling in thermophiles and mesophiles: of the good and evil”. In: *Biochemical Society Transactions* 39.1 (2011), pp. 58–63.
- [181] Purificaci n L pez-Garc a and Patrick Forterre. “DNA topology and the thermal stress response, a tale from mesophiles and hyperthermophiles”. In: *BioEssays* 22.8 (2000), pp. 738–746.





## FOLIO ADMINISTRATIF

### THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : FORQUET

DATE de SOUTENANCE : 17/12/2021

Prénoms : Raphaël

TITRE : Modelling transcriptional regulation by DNA supercoiling in bacteria

NATURE : Doctorat

Numéro d'ordre : 2021LYSEI102

Ecole doctorale : E2M2 Ecosystèmes Evolution Modélisation Microbiologie

Spécialité : Biomath-Bioinfo-Génomique évolutive

#### RESUME :

Bacteria are exposed to environmental fluctuations, to which they respond by quick and global changes in gene expression. Usual models of transcriptional regulation are centred on transcription factors which recognise specific sequences in genes' promoters, but disregard the important role of DNA supercoiling (SC), an ubiquitous property of the double-helix resulting from torsional stress. SC acts as a global and ubiquitous regulator in response to environmental changes, as suggested by many recent transcriptomics studies. The objective of the present thesis is to develop quantitative models of this regulation mode, by identifying the promoter-sequence determinants of gene SC-sensitivity, based on RNA Polymerase-DNA interaction and independently from additional regulatory proteins. To this end, we combine (i) the analysis of available transcriptomic data under conditions of SC variations, (ii) transcription assays on mutant promoters, and (iii) a thermodynamic modelling of transcription at the genome-scale. We first characterise the transcriptome of *D. dadantii*, a phytopathogen in which extensive data have been accumulated regarding the role of SC during plant infection, defining its transcription units and promoters. We then present two models explaining how global SC variations can selectively activate/repress promoters, depending on (i) the G/C-content of their discriminator for SC-assisted promoter opening, and (ii) their spacer length for the SC-dependent orientation between -35 and -10 elements affecting RNA Polymerase binding. Transcription assays are conducted on mutant promoters, and quantitatively confirm the predictions of the models. The universality of these mechanisms is demonstrated by analysing transcriptomes of distant bacteria under conditions of SC variations. Altogether, these results show that SC, based on the fundamental properties of DNA, constitutes an ubiquitous regulation mode in the prokaryotic kingdom.

MOTS-CLÉS : Transcriptional regulation, DNA supercoiling, quantitative thermodynamic modelling, discriminator, spacer, phytopathogen, transcription unit

Laboratoire (s) de recherche : Microbiology Adaptation and Pathogenesis (MAP)

Directeur de thèse: REVERCHON Sylvie

Co-directeur de thèse : MEYER Sam

Président de jury :

Composition du jury :

Junier, Ivan	Chargé de recherche	TIMC	Rapporteur
Strick, Terence	Professeur	IBENS	Rapporteur
Muskhelishvili, Georgi	Professeur	Agricultural University of Georgia	Examinateur
Reverchon, Sylvie	Professeure	INSA Lyon	Directrice de thèse
Meyer, Sam	Maître de conférences	INSA Lyon	Co-directeur de thèse