



**HAL**  
open science

# Régulation globale de la transcription bactérienne par le surenroulement de l'ADN

Maiwenn Pineau

► **To cite this version:**

Maiwenn Pineau. Régulation globale de la transcription bactérienne par le surenroulement de l'ADN. Sciences du Vivant [q-bio]. INSA de Lyon, 2023. Français. NNT : 2023ISAL0091 . tel-04583379

**HAL Id: tel-04583379**

**<https://theses.hal.science/tel-04583379>**

Submitted on 22 May 2024

**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 : 2023ISAL0091

**THÈSE DE DOCTORAT DE L'INSA LYON**  
membre de l'Université de Lyon

École doctorale N°341  
Évolution, Écosystème, Microbiologie, Modélisation

**Spécialité / discipline de doctorat :**  
Biomath-Bioinfo-Génomique évolutive

Soutenue publiquement le 24/11/2023, par :  
**Maiwenn PINEAU**

---

**Régulation globale de la transcription bactérienne  
par le surenroulement de l'ADN**

---

Devant le jury composé de :

Jean-Yves BOUET	Directeur de Recherche	LMGM-CBI, CNRS	Rapporteur
Ivan JUNIER	Directeur de Recherche	TIMC, CNRS	Rapporteur
Virginia LIOY	Chargée de Recherche	I2BC, CNRS	Examinatrice
Florence POPOWYCZ	Professeure	ICBMS, INSA Lyon	Examinatrice
Sam MEYER	Maître de Conférences	MAP, INSA Lyon	Directeur de thèse
Florence HOMMAIS	Professeure	MAP, UCBL	Co-directrice de thèse

## Département FEDORA – INSA Lyon - Ecoles Doctorales

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
ED 206 CHIMIE	<u>CHIMIE DE LYON</u> <a href="https://www.edchimie-lyon.fr">https://www.edchimie-lyon.fr</a> Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage <a href="mailto:secretariat@edchimie-lyon.fr">secretariat@edchimie-lyon.fr</a>	<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>
ED 341 E2M2	<u>ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION</u> <a href="http://e2m2.universite-lyon.fr">http://e2m2.universite-lyon.fr</a> Sec. : Bénédicte LANZA Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 <a href="mailto:secretariat.e2m2@univ-lyon1.fr">secretariat.e2m2@univ-lyon1.fr</a>	<b>Mme Sandrine CHARLES</b> Université Claude Bernard Lyon 1 UFR Biosciences Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69622 Villeurbanne CEDEX <a href="mailto:e2m2.codir@listes.univ-lyon1.fr">e2m2.codir@listes.univ-lyon1.fr</a>
ED 205 EDISS	<u>INTERDISCIPLINAIRE SCIENCES-SANTÉ</u> <a href="http://ediss.universite-lyon.fr">http://ediss.universite-lyon.fr</a> Sec. : Bénédicte LANZA Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 <a href="mailto:secretariat.ediss@univ-lyon1.fr">secretariat.ediss@univ-lyon1.fr</a>	<b>Mme Sylvie RICARD-BLUM</b> Laboratoire 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>
ED 34 EDML	<u>MATÉRIAUX DE LYON</u> <a href="http://ed34.universite-lyon.fr">http://ed34.universite-lyon.fr</a> Sec. : Yann DE ORDENANA Tél : 04.72.18.62.44 <a href="mailto:yann.de-ordenana@ec-lyon.fr">yann.de-ordenana@ec-lyon.fr</a>	<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>
ED 160 EEA	<u>ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE</u> <a href="https://edeea.universite-lyon.fr">https://edeea.universite-lyon.fr</a> Sec. : Philomène TRECCOURT Bâtiment Direction INSA Lyon Tél : 04.72.43.71.70 <a href="mailto:secretariat.edeea@insa-lyon.fr">secretariat.edeea@insa-lyon.fr</a>	<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>
ED 512 INFOMATHS	<u>INFORMATIQUE ET MATHÉMATIQUES</u> <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 <a href="mailto:infomaths@univ-lyon1.fr">infomaths@univ-lyon1.fr</a>	<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:direction.infomaths@listes.univ-lyon1.fr">direction.infomaths@listes.univ-lyon1.fr</a>
ED 162 MEGA	<u>MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE</u> <a href="http://edmega.universite-lyon.fr">http://edmega.universite-lyon.fr</a> Sec. : Philomène TRECCOURT Tél : 04.72.43.71.70 Bâtiment Direction INSA Lyon <a href="mailto:mega@insa-lyon.fr">mega@insa-lyon.fr</a>	<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>
ED 483 ScSo	<u>ScSo<sup>1</sup></u> <a href="https://edsciencesociales.universite-lyon.fr">https://edsciencesociales.universite-lyon.fr</a> Sec. : Mélina FAVETON Tél : 04.78.69.77.79 <a href="mailto:melina.faveton@univ-lyon2.fr">melina.faveton@univ-lyon2.fr</a>	<b>M. Bruno MILLY</b> (INSA : J.Y. TOUSSAINT) Univ. Lyon 2 Campus Berges du Rhône 18, quai Claude Bernard 69365 LYON CEDEX 07 Bureau BEL 319 <a href="mailto:bruno.milly@univ-lyon2.fr">bruno.milly@univ-lyon2.fr</a>

## Remerciements

Je souhaite sincèrement remercier l'ensemble du jury, Jean-Yves Bouet, Ivan Junier, Virginia Lioy et Florence Popowycz, pour avoir accepté d'évaluer le produit de mes 3 ans de thèse.

Je remercie aussi tous les chercheurs de mon comité de suivi individuel, Axel Cournac, Olivier Espéli, Hubert Charles et Rita Rebollo, pour les échanges très constructifs et bienveillants partagés à l'occasion de ces comités et de nos différentes rencontres. L'ensemble de leurs suggestions a été d'une grande aide pour l'avancée de mes travaux. Un grand merci à Axel pour ton accompagnement dans l'analyse des données Hi-C et à Sergio Peignier pour ton aide dans la mise en forme du package.

Pendant ma thèse, j'ai énormément appris de mes directeurs de thèses, Sam Meyer et Florence Hommais. Merci Sam de m'avoir aidée à y voir plus clair en stat, d'avoir su m'aider à percer le message des plus sombres données, d'avoir été maître dans l'art des réunions synthétiques et efficaces, et d'avoir été aussi réactif et encourageant même à l'autre bout du monde. Florence, merci pour ta gentillesse et ta patience, merci d'avoir toujours trouvé un message positif dans chaque expérience ratée, merci d'avoir veillé sur moi. Merci à vous deux d'avoir toujours pris le temps de m'accompagner à chaque étape de mon projet, même lorsque j'étais plus sur une phase expérimentale ou bioinfo et que ce n'était pas votre spécialité.

Thanks to Monika for showing me Gdansk and the surrounding areas, between two more or less successful experiences. I spent a very nice month with you, both professionally and personally!

Pendant mes 3 ans et demi à CRP, j'ai eu la chance d'avoir été entourée d'une super équipe! Chaque au revoir était bien triste (il y en a assez des cagnottes...) et quitter toute l'équipe sera un déchirement car ce sera très difficile de trouver meilleure ambiance! Merci à Sylvie et William de toujours m'avoir accueillie avec de grands sourires et d'excellents conseils malgré vos emplois du temps surchargés. Merci à Zahar d'être une oreille si attentive et de m'avoir aussi bien conseillée pour mes projets pro et perso. Merci à Cécile d'avoir gentiment accepté de partager ton calme bureau et d'avoir suivi avec autant d'intérêt mes travaux. Merci Raph d'avoir été un sacré pilier (pas que de bar) du début de mon stage jusqu'à la fin de ma thèse. Merci Amandine d'avoir été le rayon de soleil de cette équipe et la reine des apéros-punch-cacahuètes du vendredi, avec toi c'est tout simplement IMPOSSIBLE de s'ennuyer! Merci aux folles du bureau des kiwis (Camille, Jess et encore toi Amandine) pour toutes ces séances de rires, de musique de mauvais goût et de "décompression" du vendredi (désolée William et Sylvie pour le bruit...). Merci pour votre soutien de tous les jours (petite dédicace au carnet de bienveillance et... aux oreilles

de lapin) et, car il fallait bien travailler quand même, merci pour vos conseils et aides techniques!!! En ce qui concerne les rires et les comérages, on peut dire que la relève est bien assurée avec Krystal, Sara et Clara. Depuis votre arrivée, on ne peut pas franchement dire que le labo soit plus calme mais qu'est-ce que c'est agréable! Je compte sur Jess et vous pour faire perdurer les traditions de petit dej et d'apéro!

Je remercie chaleureusement tous les membres de l'UMR5240 pour les échanges lors des évènements plus ou moins studieux (séminaires, pique niques, repas divers...). Un merci tout spécial à nos voisins de couloir assidus aux petit dej et au Ponyo (Margot, Emma, Thomas, Yoan, Lilian, Sonia, Elodie...) et à tous les "jeunes" que j'ai eu le plaisir de cotoyer au travail mais aussi en dehors. Big up à Julie, Typhaine et Xavier avec qui j'ai eu plaisir à traverser ces 3 années (même le train d'Aussois n'a pas eu raison de nous...). Merci aussi à Véro, Salima et Fahis pour leur gentillesse et leur jovialité!

Il paraît qu'écrire une thèse est l'occasion de remercier tous ceux qui nous donnent envie de nous lever le matin. Vous savez que je ne suis vraiment pas du matin mais c'est vrai que quand c'est pour vous voir je suis de bien meilleure humeur! Je pense bien évidemment aux copines Fiona, Manon, Laure, Alice et Eléonore, avec qui une journée partagée se transforme toujours en souvenirs mémorables (mention spéciale pour la porte). Papa, maman, Gwenolé, Nolwenn, Marie, Florian, Ava et Timothée, la liste de la famille commence à être longue mais quel bonheur, je ne pourrais pas rêver d'une meilleure famille! Pour faire court parce qu'on n'est pas des poètes : love keur keur. D'ailleurs, j'ai bien de la chance car même ma belle-famille est au top! Merci à Tove et Eric de nous avoir accueillis avec autant de gentillesse pour le confinement et un grand nombre de week-ends. Merci aux Marie's de s'être assurées (1) que je ne perde pas le moindre gramme pendant ces années lyonnaises, (2) que je devienne encore plus addictive aux jeux de société et (3) que je ne dorme pas à la rue après notre déménagement! Gaspard, vu que je te vois tous les jours ça fonctionne moins bien pour me motiver à me lever... Par contre les jours où tu n'es pas là je n'ai vraiment aucune envie de rentrer à la maison le soir (bon il y a les chats quand même...). Merci de partager ma vie dans les meilleurs et les pires moments, de t'être assuré que je laisse toujours assez de place à ma vie perso, d'être aux petits soins quand j'en ai besoin, de me pousser/suivre dans nos projets les plus beaux et fous : nous dire OUI et nous lancer dans l'aventure Lille Bjørn (aka Hortense)!

Enfin, je remercie tous les relecteurs de ma thèse, Florence, Sam, Jess, Gaspard, Fiona et mes parents. Merci d'avoir passé autant de temps à relire ces pages, auxquelles certains d'entre vous (coucou papa et maman) ne comprenaient pas grand chose, pour relever un maximum de coquilles!

## Résumé : régulation globale de la transcription bactérienne par le surenroulement de l'ADN

L'ADN est le porteur de l'information génétique permettant le développement, le fonctionnement et la reproduction des cellules. La structure de l'ADN est très organisée, à petite échelle avec sa structure en double-hélice, mais également à plus grande échelle. En effet, chez les bactéries, le chromosome est situé dans le cytoplasme, sous une forme très compacte. Cette compaction résulte notamment du surenroulement de l'ADN (SC), c'est à dire de la déformation torsionnelle de la double-hélice de l'ADN. Cette déformation est à la fois impliquée et impactée dans les processus biologiques nécessitant la dissociation des deux brins de la double-hélice, tels que la réplication et la transcription. Selon les conditions environnementales rencontrées par les bactéries, le niveau de SC peut varier, principalement sous le contrôle de la gyrase, qui augmente le niveau de SC, et de la topoisomérase I, qui relâche l'ADN. La régulation du niveau de SC est très importante car une variation de SC impacte l'expression des gènes. En effet, il a été montré que le SC est un régulateur global, ancestral et complexe de l'expression des gènes. L'objectif de ma thèse est de caractériser la régulation globale de la transcription bactérienne par le SC.

Pour ce faire, j'ai tout d'abord étudié la réponse de la bactérie phytopathogène *Dickeya dadantii* à l'inhibition de sa topoisomérase I par un antibiotique. L'étude de la réponse transcriptionnelle de cette bactérie à une variation de SC est particulièrement intéressante car, au cours de l'infection des plantes, cette bactérie est soumise à différents stress environnementaux qui affectent son niveau de SC. Ainsi, en reproduisant des variations de SC, nous pourrions améliorer notre compréhension du processus de virulence de cette bactérie. Jusqu'à présent, chez les bactéries à Gram négatif, l'effet d'une modification du niveau de SC sur la transcription a été étudié en relâchant l'ADN, souvent en inhibant la gyrase. Nous avons obtenu le premier transcriptome d'une bactérie à Gram négatif dont la topoisomérase I est inhibée par un antibiotique. Pour cela, nous avons traité des cultures de *D. dadantii* avec un antibiotique non commercialisé, la séconéolitsine, qui inhibe la topoisomérase I puis nous avons réalisé une étude RNA-Seq afin de connaître la réponse transcriptionnelle des gènes à ce traitement. Nous avons alors mis une nouvelle fois en évidence une régulation globale et complexe de la transcription par le SC. En comparant la réponse des gènes à l'inhibition de la topoisomérase I à celle en réponse à l'inhibition de la gyrase, nous avons découvert que la réponse des gènes à une variation de SC dépend du niveau d'expression et du contexte génomique de ces gènes, de la direction de la variation de SC et du contexte physiologique de la bactérie.

Afin de mieux caractériser le lien entre SC et transcription, il est, d'une part, essentiel de pouvoir analyser un grand nombre de données de différents types et, d'autre part, de prendre en compte les variations locales du SC. C'est dans ce but que j'ai contribué au développement d'un package Python, nommé GRATIOSA, facilitant les analyses statistiques reproductibles de données expérimentales et d'annotations dans le cadre de l'étude de l'expression d'un génome bactérien selon ses propriétés spatiales. Ce package permet l'intégration de nombreuses données expérimentales dont des données de type RNA-Seq et ChIP-Seq. Les données de type ChIP-Seq peuvent notamment apporter des informations sur le niveau local du SC. En effet, il est connu depuis plusieurs décennies que le niveau de SC varie localement mais les premières données de cartographie de ce niveau de SC le long du chromosome de *Escherichia coli* n'ont été publiées que récemment. Parmi ces données se trouvent des signaux de fixation de topoisomérase I et de gyrase. En effet, l'affinité des topoisomérases dépendant du niveau de SC, ces données reflètent le niveau de SC local. En analysant ces données avec le package GRATIOSA, j'ai pu, dans un premier temps, observer l'effet de la transcription sur le niveau de SC à une échelle de 10 kb autour des unités de transcription. A cette échelle, j'ai pu faire des observations conformes au modèle du *Twin-supercoiled domain*, qui propose une accumulation de supertours négatifs et positifs, respectivement en amont et en aval de l'ARN polymérase en cours de transcription. Dans un second temps, j'ai pu observer une organisation en domaines de 100-200 kb dans lesquels le niveau de SC serait corrélé. Ces observations pourront être approfondies, en utilisant le package GRATIOSA, lorsque de nouvelles cartographies précises du SC local seront publiées.

En résumé, mes travaux de thèse ont permis de contribuer à la caractérisation de la régulation de l'expression des gènes par le SC, de partager un outil informatique facilitant les analyses du lien entre transcription et structure du chromosome, et d'améliorer nos connaissances de la structure du chromosome bactérien en confirmant le modèle du *Twin-supercoiled domain* et en apportant un nouvel aperçu d'une organisation du chromosome à l'échelle de 100 kb environ.

## **Abstract: Global regulation of bacterial transcription by DNA supercoiling**

DNA carries the genetic information necessary for the development, survival, and reproduction of cells. The DNA structure is highly organized, both at small scale with its double-helix structure and at larger scale. In bacteria, the chromosome is located in the cytoplasm in a highly compact form. This compaction is mainly a result of DNA supercoiling (SC), i.e. the torsional deformation of the DNA double helix. This deformation is affected by and involved in biological processes that require the separation of the two strands of the DNA double helix, such as replication and transcription.

The level of SC can vary depending on the environmental conditions encountered by bacteria. This level is mainly controlled by gyrase, which increases SC level, and topoisomerase I, which relaxes the DNA. The regulation of SC level is crucial because variations in SC can impact gene expression. Indeed, it has been shown that SC is a global, ancestral, and complex regulator of gene expression. The objective of my thesis is to characterize the global regulation of bacterial transcription by SC.

To achieve this, I first studied the response of the phytopathogenic bacterium *Dickeya dadantii* to the inhibition of its topoisomerase I by an antibiotic. Studying the transcriptional response of this bacterium to a variation in SC is particularly interesting because, during plant infection, this bacterium is exposed to different environmental stresses that seem to affect its SC level. Thus, by reproducing SC variations, we could enhance our understanding of the virulence process of this bacterium. So far, in Gram-negative bacteria, the effect of altering the SC level on transcription has been studied by relaxing the DNA, often by inhibiting gyrase. We obtained the first transcriptome of a Gram-negative bacterium upon inhibition of its topoisomerase I with an antibiotic. For this purpose, we treated cultures of *D. dadantii* with a non-marketed antibiotic, seconeolitsin, which inhibits topoisomerase I, and then conducted an RNA-Seq study to analyze the transcriptional response of genes to this treatment. Once again, we highlighted a global and complex regulation of transcription by SC. By comparing the gene response to topoisomerase inhibition to that of gyrase inhibition, we discovered that the gene response to SC variation depends on the gene's expression level, genomic context, the direction of SC variation, and the bacterium's physiological context.

To better characterize the interplay between SC and transcription, it is essential, on the one hand, to be able to analyze a large amount of data of different types and, on the other hand, to consider local variations in SC. For this purpose, I contributed to the development of a



Python package named GRATIOSA, which facilitates reproducible statistical analyses of experimental data and annotations in the study of a bacterial genome's expression based on its spatial properties. This package allows the integration of numerous experimental data types, including RNA-Seq and ChIP-Seq data. ChIP-Seq-type data can provide information about local SC level. It has been known for several decades that the SC level vary locally, but the first mapping data of the SC level along the *Escherichia coli* chromosome were only recently published. Among these data are topoisomerase I and gyrase binding signals. Since topoisomerase affinities depend on SC levels, these data reflect the local SC level. By analyzing this data with the GRATIOSA package, I was able to, first, observe the effect of transcription on the SC level on a scale of 10 kb around transcription units. At this scale, I made observations consistent with the Twin-supercoiled domain model, which suggests an accumulation of negative and positive supercoils, respectively upstream and downstream of the transcribing RNA polymerase. Second, I observed an organization into domains of approximately 100-200 kb where SC levels would be correlated. These observations can be further explored using the GRATIOSA package when new precise mappings of local SC become available.

In summary, my thesis work has improved the characterization of gene expression regulation by SC, a computational tool that facilitates the analysis of the link between transcription and chromosome structure, and our knowledge of bacterial chromosome structure by confirming the Twin-supercoiled domain model and providing a new perspective on chromosome organization at the scale of approximately 100 kb.

# Table des matières

<b>Remerciements</b>	<b>iii</b>
<b>Résumé</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Table des figures</b>	<b>xiii</b>
<b>Liste des tableaux</b>	<b>xv</b>
<b>Liste des abréviations</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Le génome bactérien . . . . .	1
1.1.1 Historique de l'étude des bactéries . . . . .	1
1.1.2 Les bactéries . . . . .	2
La composition des cellules bactériennes . . . . .	2
Mes modèles d'étude . . . . .	3
1.1.3 Les découvertes de l'ADN . . . . .	4
1.1.4 L'organisation du chromosome circulaire chez les bactéries . . . . .	6
1.1.5 La compaction de la chromatine . . . . .	8
La compaction du chromosome bactérien à grande échelle . . . . .	9
La compaction du chromosome bactérien à moyenne et petite échelles	11
Les différentes échelles de la compaction du chromosome eucaryote	12
1.1.6 Les acteurs de la compaction de la chromatine bactérienne . . . . .	14
Les protéines associées au nucléoïde . . . . .	14
Les protéines de maintenance structurale du chromosome . . . . .	15
Le surenroulement de l'ADN . . . . .	18
1.2 La transcription bactérienne . . . . .	20
1.2.1 L'ARN polymérase . . . . .	20
1.2.2 Les étapes de la transcription bactérienne . . . . .	22
L'initiation de la transcription . . . . .	22

	L'élongation . . . . .	23
	La terminaison . . . . .	24
1.2.3	Les modèles standards de régulation de l'initiation de la transcription . . . . .	24
	Les facteurs de transcription . . . . .	25
	Les facteurs $\sigma$ alternatifs . . . . .	27
	Les petits ligands : exemple de ppGpp . . . . .	28
1.2.4	Les techniques d'analyse du transcriptome . . . . .	28
	Les puces à ADN . . . . .	28
	Le RNA-Seq . . . . .	29
1.3	Les variations du niveau de surenroulement de l'ADN . . . . .	32
1.3.1	La régulation du surenroulement par les topoisomérases . . . . .	32
	Les topoisomérases de type I . . . . .	32
	Les topoisomérases de type II . . . . .	36
1.3.2	Les méthodes de mesure du niveau global de surenroulement de l'ADN . . . . .	39
1.3.3	L'impact des conditions environnementales . . . . .	41
1.4	La régulation de la transcription bactérienne par le surenroulement de l'ADN . . . . .	44
1.4.1	Les études <i>in vitro</i> sur plasmide . . . . .	44
1.4.2	Les études transcriptomiques <i>in vivo</i> . . . . .	45
1.4.3	Les mécanismes envisagés . . . . .	46
1.5	Étude du SC à l'échelle locale . . . . .	48
1.5.1	Impact de la transcription sur le surenroulement de l'ADN . . . . .	48
1.5.2	Les mesures du niveau local du surenroulement . . . . .	49
	La cartographie des fixations des topoisomérases . . . . .	50
	La cartographie des supertours positifs par GapR . . . . .	51
	La cartographie du surenroulement négatif par le psoralen . . . . .	52
	Récapitulatif des méthodes de mesure du SC . . . . .	55
1.6	Les objectifs de ma thèse . . . . .	57
<b>2</b>	<b>Etude de la réponse transcriptionnelle de <i>Dickeya dadantii</i> à une augmentation du niveau de surenroulement</b> . . . . .	<b>59</b>
2.1	Le genre <i>Dickeya</i> . . . . .	59
2.1.1	Le cycle infectieux des <i>Dickeya</i> spp. . . . .	59
2.1.2	Le processus infectieux de la souche <i>Dickeya dadantii</i> 3937 . . . . .	61
	La phase asymptomatique . . . . .	62
	La phase symptomatique . . . . .	63

2.1.3	Le rôle du surenroulement de l'ADN de <i>Dickeya dadantii</i> pendant l'infection . . . . .	64
2.2	L'inhibition de la topoisomérase I de <i>Dickeya dadantii</i> . . . . .	65
2.2.1	Le choix d'un antibiotique inhibiteur de la topoisomérase I . . . . .	65
	Les méthodes d'identification de nouveaux inhibiteurs de la topoI . . . . .	65
	Le choix de la séconéolitsine en inhibiteur de la topoI . . . . .	66
2.2.2	La vérification de l'efficacité de la séconéolitsine sur la topoI de <i>Dickeya dadantii in vitro</i> . . . . .	67
	La production de la topoI de <i>D. dadantii</i> . . . . .	68
	La purification de la topoI produite . . . . .	68
	La vérification de l'activité de la topoI et de son inhibition par la séconéolitsine . . . . .	70
2.3	Article . . . . .	71
2.4	L'effet du surenroulement sur la virulence de <i>Dickeya dadantii</i> . . . . .	95
<b>3</b>	<b>Etude de l'expression du génome bactérien selon son organisation spatiale avec le package GRATIOSA</b> . . . . .	<b>97</b>
3.1	Le développement du package GRATIOSA : Genome Regulation Analysis Tool Incorporating Organization and Spatial Architecture . . . . .	97
3.1.1	Présentation du package initial . . . . .	97
3.1.2	Les objectifs du package GRATIOSA . . . . .	98
3.1.3	L'implémentation de nouvelles fonctions . . . . .	100
	Les fonctions d'importation de nouvelles données . . . . .	100
	Les fonctions statistiques . . . . .	101
	Les fonctions graphiques . . . . .	102
3.1.4	La documentation du package GRATIOSA . . . . .	103
3.1.5	Article . . . . .	103
3.2	La mise au point d'une technique de cartographie du surenroulement local . . . . .	125
3.2.1	Présentation du protocole de cartographie . . . . .	125
3.2.2	Calibration du protocole de cartographie chez <i>D. dadantii</i> . . . . .	127
	Méthodologie de calibration de la concentration en TMP . . . . .	127
	Mise au point d'une méthodologie d'extraction de plasmide de <i>D. dadantii</i> . . . . .	128
	Résultat de la calibration de la quantité en TMP . . . . .	131
3.2.3	Comparaison des protocoles de cartographie . . . . .	132
3.3	Analyse des cartographies du SC local obtenues par GapR-Seq et Psora-Seq . . . . .	134
3.3.1	Matériels et méthodes . . . . .	134
	Traitement des données brutes avec des outils standards . . . . .	134

	Analyse des données pré-traitées avec le package GRATIOSA . . . . .	134
3.3.2	Résultats . . . . .	135
3.4	Etude de la topologie du chromosome de <i>D. dadantii</i> par Hi-C. . . . .	137
3.4.1	Matériels et méthodes . . . . .	137
	Préparation des échantillons . . . . .	137
	Traitement des données brutes . . . . .	137
3.4.2	Les résultats préliminaires . . . . .	139
	Les frontières et les boucles détectées . . . . .	139
	Étude du niveau d'expression des gènes proches des motifs identifiés	140
	Étude du mécanisme de transertion et de son impact sur la topologie du chromosome . . . . .	140
<b>4</b>	<b>Discussion et perspectives</b>	<b>143</b>
4.1	L'organisation du chromosome bactérien . . . . .	143
4.2	La régulation des gènes par le surenroulement de l'ADN . . . . .	146
4.3	Le rôle du surenroulement de l'ADN dans la virulence des <i>Dickeya</i> spp. . .	148
	<b>Bibliographie</b>	<b>149</b>

# Table des figures

1.1	Schématisation d'une cellule bactérienne. . . . .	2
1.2	Le genre <i>Dickeya</i> : microscopie, symptômes et distribution mondiale. . . . .	4
1.3	La structure de la double hélice de l'ADN. . . . .	5
1.4	L'organisation du génome bactérien . . . . .	7
1.5	La morphologie du nucléoïde de <i>E. coli</i> analysée par microscopie. . . . .	9
1.6	L'étude des macrodomaines de <i>E. coli</i> par Hi-C. . . . .	10
1.7	L'organisation des chromosomes bactériens et eucaryotes à différentes échelles. . . . .	13
1.8	La structuration de l'ADN par les NAPs. . . . .	14
1.9	Exemples de NAPs chez <i>E. coli</i> . . . . .	15
1.10	Structure des complexes SMC . . . . .	16
1.11	Modèle d'un mécanisme d'extrusion de boucle par un complexe SMC. . . . .	17
1.12	Le surenroulement de l'ADN . . . . .	18
1.13	Les interactions d'une ARN polymérase bactérienne avec une séquence promotrice consensus. . . . .	20
1.14	Les étapes de la transcription bactérienne. . . . .	22
1.15	Les mécanismes des facteurs de transcription. . . . .	26
1.16	Protocole expérimental d'analyse transcriptomique avec une puce à ADN. . . . .	29
1.17	Protocole d'analyse transcriptomique par RNA-Seq. . . . .	30
1.18	La régulation du surenroulement de l'ADN en réponse aux stress environnementaux. . . . .	32
1.19	Structures principales des topoisomérases de type I . . . . .	34
1.20	Le mécanisme de relâchement de l'ADN par la topoisomérase I. . . . .	35
1.21	Structures principales des topoisomérases de type II . . . . .	36
1.22	Le mécanisme de surenroulement de l'ADN par la gyrase. . . . .	37
1.23	Les méthodes d'étude du niveau de surenroulement global par électrophorèse. . . . .	40
1.24	Méthode de mesure du niveau global de SC par fluorescence. . . . .	41
1.25	Effets des chocs acides et oxydatifs sur le surenroulement de l'ADN. . . . .	42
1.26	Effet d'un choc froid sur le surenroulement de l'ADN de <i>E. coli</i> . . . . .	43
1.27	Études sur plasmides de l'effet d'une variation de surenroulement sur la transcription. . . . .	44

1.28	Effet de la transcription sur le surenroulement. . . . .	48
1.29	Le protocole de CHIP-Seq . . . . .	50
1.30	Le principe du Topo-Seq. . . . .	51
1.31	La liaison de GapR à l'ADN. . . . .	52
1.32	L'intercalation du psoralen selon le niveau de surenroulement. . . . .	53
1.33	Méthodes de cartographie du surenroulement local utilisant les propriétés d'intercalation du psoralen. . . . .	54
2.1	Le cycle de vie de <i>Dickeya</i> spp. en interaction avec les pommes de terre. . . .	60
2.2	Les étapes de l'infection d'une plante par <i>D. dadantii</i> . . . . .	61
2.3	Présentation de la séconéolitsine . . . . .	67
2.4	La production de la topoI de <i>D. dadantii</i> . . . . .	69
2.5	Tests de virulence de <i>D. dadantii</i> sur endive après traitement à la novobiocine ou la séconéolitsine . . . . .	95
3.1	Le package GRATIOSA et sa base de données. . . . .	99
3.2	Exemple de graphique obtenu avec le package GRATIOSA. . . . .	102
3.3	Protocole expérimental de la cartographie du surenroulement de l'ADN par intercalation du tri-méthyl-psoralen. . . . .	126
3.4	Comparaison par électrophorèse sur gel d'agarose des rendements de diffé- rents kits d'extraction de plasmides. . . . .	129
3.5	Comparaison par électrophorèse sur gel d'agarose des plasmides extraits selon le protocole de lavage. . . . .	130
3.6	Calibration de la concentration en TMP . . . . .	131
3.7	La mesure du SC local par Psora-seq. . . . .	132
3.8	Signaux de Psora-Seq et de GapR-Seq autour des TUs de <i>E. coli</i> . . . . .	136
3.9	Les motifs détectables sur une matrice de contact obtenue par Hi-C. . . . .	138
3.10	Extraits d'une carte de contacts de <i>D. dadantii</i> . . . . .	139
3.11	Expression des gènes selon leur proximité aux boucles et aux frontières des domaines topologiques. . . . .	140

# Liste des tableaux

1.1	Les tailles des génomes bactériens . . . . .	8
1.2	Les domaines d'interactions chromosomiques . . . . .	11
1.3	Les facteurs $\sigma$ de <i>E. coli</i> . . . . .	27
1.4	Propriétés et activités <i>in vitro</i> des différentes topoisomérases. . . . .	33
1.5	La variation du niveau de surenroulement du chromosome de <i>E. coli</i> en réponse à un stress environnemental . . . . .	42
1.6	La réponse transcriptionnelle bactérienne à une modification de niveau de surenroulement. . . . .	46
1.7	Les méthodes de mesure du SC . . . . .	56
2.1	Sélection de gènes de virulences impliqués dans différentes étapes de l'infection . . . . .	62
2.2	$IC_{50}$ de molécules inhibitrices des topoI de <i>E. coli</i> et <i>M. tuberculosis</i> . . . . .	66
3.1	Concentrations en plasmides pUC19 extraits de <i>D. dadantii</i> selon le kit et le protocole utilisés. . . . .	129
3.2	Nombre de frontières et de boucles détectées dans le chromosome de <i>D. dadantii</i> par réplikat. . . . .	139





# Liste des abréviations

<b>3C</b>	Capture de la conformation du chromosome
<b>A</b>	Adenine
<b>ADN, DNA</b>	Acide désoxyribonucléique
<b>ADNc, cDNA</b>	ADN complémentaire
<b>ANR</b>	Agence Nationale de la Recherche
<b>ARN, RNA</b>	Acide ribonucléique
<b>ARNm, mRNA</b>	ARN messenger
<b>ARNP</b>	ARN Polymérase
<b>ARNr, rRNA</b>	ARN ribosomal
<b>ARNt, tRNA</b>	ARN de transfert
<b>C</b>	Cytosine
<b>CID</b>	Chromosomal Interaction Domain
<b>CMI, MIC</b>	Concentration Minimale Inhibitrice
<b>CRP</b>	cAMP receptor protein
<b>CWDEs</b>	Enzymes dégradant les parois cellulaires
<b>DO</b>	Densité Optique
<b>DTT</b>	Dithiothréitol
<b>Dps</b>	DNA-binding protein from starved cells
<b>EDTA</b>	EthylèneDiamineTétraAcétique
<b>EtBr</b>	Bromure d'éthidium
<b>ERO</b>	Espèces Réactives de l'Oxygène
<b>FC</b>	Fold Change
<b>FEDS</b>	Fluorescent evaluation of DNA supercoiling
<b>FIS</b>	Factor for Inversion Stimulation
<b>FISH</b>	Fluorescence <i>in situ</i> hybridization
<b>FPKM</b>	Fragments Per Kilobase per Million mapped fragments
<b>G</b>	Guanine
<b>GO</b>	Gene Ontology
<b>GRATIOSA</b>	Genome Regulation Analysis Tool Incorporating Organization and Spatial Architecture
<b>Hi-C</b>	High-throughput chromatin conformation capture
<b>His</b>	Histidine
<b>H-NS</b>	Histone-like Nucleoid-Structuring protein
<b>HU</b>	Histone-like protein from <i>Escherichia coli</i> strain U93
<b>IC50</b>	Concentration Inhibitrice médiane
<b>ICBMS</b>	Institut de Chimie et Biochimie Moléculaires et Supramoléculaires
<b>IHF</b>	Integration Host Factor
<b>IPTG</b>	Isopropyl-Bêta-D-ThioGalactopyranoside
<b>kb</b>	kilobase(s)

<b>Lk</b>	Linking number
<b>Mpb</b>	Million(s) de pb
<b>NAP</b>	Protéine Associée au Nucléoïde
<b>Nb</b>	Nombre
<b>NOV</b>	Novobiocine
<b>nt</b>	Nucléotide(s)
<b>NTA</b>	Acide NitriloTriacétique
<b>PAGE</b>	Electrophorèse en Gel de PolyAcrylamide
<b>pb, bp</b>	Paire(s) de bases
<b>PCR</b>	Réaction de Polymérase en Chaîne
<b>Pels</b>	Pectate-lyases
<b>ppGpp</b>	guanosine tetraphosphate
<b>qPCR</b>	PCR quantitative
<b>rNTP</b>	Ribonucléosides triphosphates
<b>RNA-Seq</b>	Séquençage de l'ARN
<b>RPKM</b>	Reads Per Kilobase Per Million Mapped Reads
<b>SDS</b>	DodécylSulfate de Sodium
<b>SC</b>	Surenroulement de l'ADN
<b>SC+</b>	Surenroulement positif
<b>SC-</b>	Surenroulement négatif
<b>SCN</b>	Séconéolitsine
<b>SMC</b>	Maintenance structurelle du chromosome
<b>T</b>	Thymine
<b>TAD</b>	Topologically Associating Domain
<b>TEDA</b>	T5 Exonuclease-Dependent Assembly
<b>TF</b>	Facteur de Transcription
<b>TMP</b>	4,5',8-triméthylpsoralen
<b>topoI</b>	Topoisomerase I
<b>TSS</b>	Site de début de transcription
<b>TTS</b>	Site de fin de transcription
<b>TU</b>	Unité de Transcription
<b>Tw</b>	Torsion
<b>U</b>	Uracile
<b>Wr</b>	Vrillage

# Chapitre 1

## Introduction

### 1.1 Le génome bactérien

#### 1.1.1 Historique de l'étude des bactéries

Les procaryotes sont des micro-organismes unicellulaires omniprésents dans tous les écosystèmes depuis l'apparition de la vie sur Terre. Des fossiles retrouvés montrent qu'ils sont présents depuis plus de 3,5 milliards d'années [1, 2] et l'Homme s'y intéresse depuis plus de 2000 ans. Au premier siècle av. J.-C., Varron et Columelle émettaient l'hypothèse que des êtres invisibles, appelés *Animalia minuta*, provoquaient des maladies par inhalation ou ingestion. Cette hypothèse a été reprise par Roger Bacon au XIII<sup>e</sup> siècle, puis par Girolamo Fracastoro à la fin du XV<sup>e</sup> siècle. Fracastoro précisa que les maladies devaient être provoquées par des "semences" ou des "germes" se propageant entre les humains. La première observation de micro-organismes, qu'il appela *Animalcules*, fut réalisée par Antoni Van Leeuwenhoek entre les années 1674 et 1676 grâce à un microscope qu'il avait fabriqué en utilisant une lentille maintenue entre des plaques d'argent [3, 4]. Il observa ainsi des protozoaires, des bactéries et des levures [5].

Depuis, nos connaissances des micro-organismes n'ont cessé de croître notamment grâce au développement des premiers milieux de culture liquides par Louis Pasteur puis solides par Robert Koch au XIX<sup>e</sup> siècle [6, 7]. C'est à ce siècle que sont apparus les termes de *Protista* puis de *Bacteria* afin de classer les micro-organismes n'appartenant, ni au règne animal, ni au règne végétal, notamment grâce aux travaux de Christian Gottfried Ehrenberg et de Ferdinand Cohn [8]. Cette classification fut affinée au siècle suivant, lorsque, en 1957, André Lwoff établit la distinction entre virus et bactérie [9], puis lorsque, en 1962, Roger Stanier et Cornelis van Niel, définirent la bactérie comme une entité cellulaire dépourvue d'organe membrané et notamment de noyau [10].

## 1.1.2 Les bactéries

### La composition des cellules bactériennes

Les bactéries sont composées d'une membrane cellulaire qui entoure leur cytoplasme, où se trouvent notamment l'ADN bactérien et les ribosomes (Fig. 1.1). Contrairement aux eucaryotes, les bactéries ne possèdent pas de noyau. Leur ADN se situe, au sein du cytoplasme, dans une région non délimitée par une membrane, appelée le nucléoïde. À leur surface, les cellules bactériennes peuvent avoir des flagelles, structures facilitant leurs déplacements, ou des pili, aidant à leur adhésion à des surfaces [11]. Bien plus petites que les cellules eucaryotes qui mesurent généralement entre 10 et 100  $\mu\text{m}$ , les cellules bactériennes ont généralement un diamètre compris entre 0,4 et 2  $\mu\text{m}$  et une longueur de 0,5 à 5  $\mu\text{m}$  [12]. Cette petite taille complique leur isolement, leur observation et l'extraction de leur contenu.

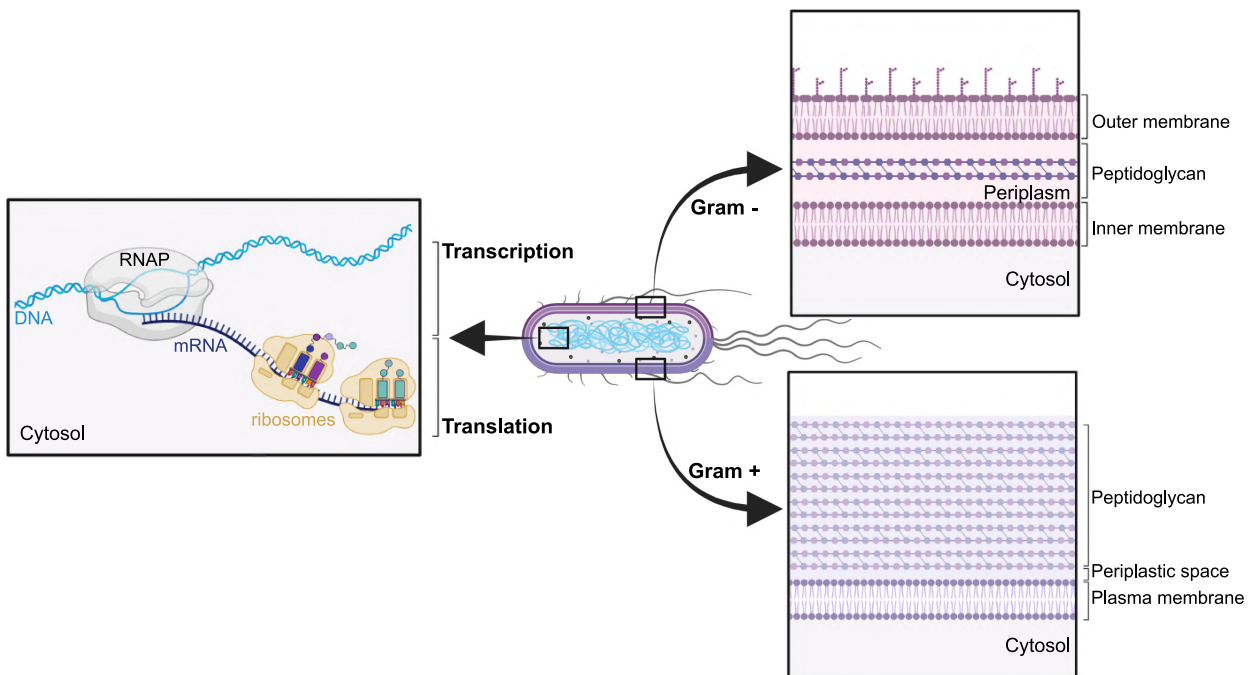


FIGURE 1.1 – Schématisation d'une cellule bactérienne. Les bactéries ne possèdent pas de noyaux, leur nucléoïde se trouve dans le cytoplasme, où les processus de transcription et de traduction sont couplés (partie gauche de la figure). Les bactéries peuvent présenter deux types de membranes : les bactéries à Gram négatif (Gram -) possèdent deux couches de membranes, séparées par une fine couche de peptidoglycane, tandis que les bactéries à Gram positif (Gram +) ont une unique membrane entourée d'un mur épais de peptidoglycane (partie droite de la figure). Schéma réalisé sur bioRender.

La membrane bactérienne est constituée de phospholipides, de protéines et de polysaccharides. On distingue principalement deux types de structures : (1) deux membranes plasmiques séparées par une fine couche de peptidoglycane, ou (2) une unique membrane plasmique entourée d'une épaisse couche de peptidoglycane. La perméabilité de la cellule

est dépendante de l'épaisseur de cette couche de peptidoglycanes [11]. Cette distinction est à l'origine de la classification des bactéries en Gram négatif (1) et Gram positif (2) selon leur réaction à un test de coloration. Ce test, nommé d'après son inventeur Hans Christian Gram, consiste à colorer les bactéries avec du violet de gentiane en présence d'une solution d'iode, puis à les décolorer avec de l'alcool. Étant plus perméables avec leur fine couche de peptidoglycanes, les bactéries à Gram négatif seront davantage décolorées que les bactéries à Gram positif. Ainsi, les bactéries à Gram négatif apparaissent roses après décoloration puis contre-coloration avec une solution de safranine, tandis que les bactéries à Gram positif restent violettes [13, 14]. Cependant, cette classification présente des limites puisque certaines bactéries ne sont pas cultivables et ne peuvent donc pas être testées [15] tandis que d'autres bactéries ne semblent correspondre à aucune de ces deux catégories. C'est notamment le cas de *Mycoplasma pneumoniae*, *Anaplasma phagocytophilum* et *Ehrlichia chaffeensis* qui ne contiennent pas de peptidoglycane [16].

### Mes modèles d'étude

Les deux modèles d'études utilisés pendant ma thèse sont les bactéries à Gram négatif *Escherichia coli* et *Dickeya dadantii*.

*E. coli* est un bacille appartenant à la famille des *Enterobacteriaceae*. Présente dans la microflore digestive des animaux à sang chaud, la souche *E. coli* K-12 a été isolée pour la première fois à partir d'un patient convalescent de la diphtérie en 1922 et est utilisée comme organisme modèle depuis les années 1940 [17]. Son génome a été séquencé pour la première fois en 1997 [18]. En raison de sa facilité de culture, de sa croissance rapide, de sa manipulation aisée et de sa stabilité génétique, *E. coli* est le modèle d'étude classique des bactéries à Gram négatif. Elle a joué un rôle clé dans la compréhension de la génétique, de l'évolution, de la biologie moléculaire et de la pathogénie [19]. *E. coli* est également une bactérie privilégiée pour les applications biotechnologiques telles que la surproduction de protéines dont elle permet un rendement élevé [20].

Le modèle d'étude plus spécifique de ma thèse est la bactérie phytopathogène *Dickeya dadantii* (Fig. 1.2A). Les bactéries du genre *Dickeya* sont des phytopathogènes nécrotrophes, c.-à-d. qui tuent leur hôte et se nourrissent de leurs tissus morts. Elles infectent un large spectre d'espèces végétales, dont certaines d'intérêt agronomique comme la pomme de terre, la banane et la tomate [21]. Cantonnées principalement aux pays tropicaux jusqu'en 2000, les échanges de marchandises et le changement climatique en font une menace croissante en Europe. Des infections ont déjà été reportées dans un grand nombre de pays européens dont la Pologne, la Grèce, la Finlande, la Bulgarie et la France (Fig. 1.2C) [22, 23, 24, 25]. Le symptôme le plus caractéristique de *Dickeya* spp. est la pourriture molle : une lésion gluante, humide, de pourriture noire qui se propage (Fig. 1.2B) [26, 27]. En raison de

leur impact sur la qualité et les rendements agricoles, et pour leur intérêt scientifique, elles figurent parmi les 10 bactéries phytopathogènes de plus grande importance scientifique et économique [28].

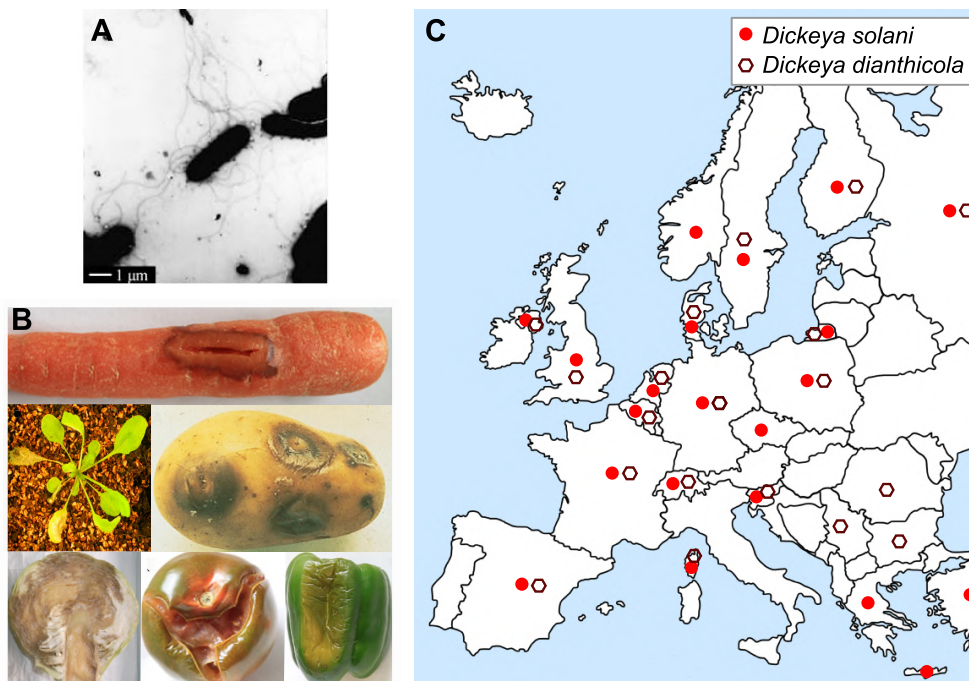


FIGURE 1.2 – (A) *Dickeya dadantii* 3937 observée au microscope électronique [29]. (B) Symptômes de pourriture molle provoqués par des bactéries du genre *Dickeya* sur différents végétaux [27]. (C) Distribution des deux espèces du genre *Dickeya* responsables des plus grandes pertes économiques actuellement en Europe. Les ronds pleins indiquent les pays dans lesquels des infections de *Dickeya solani* ont été recensées (données actualisées en octobre 2022 [30]). Les hexagones vides correspondent à *Dickeya dianthicola* (données de janvier 2023 [31]).

Les bactéries du genre *Dickeya* sont également des *Enterobacteriaceae* facilement cultivables et génétiquement manipulables. À ce jour, douze espèces du genre *Dickeya* ont été isolées, dont neuf sont des phytopathogènes de plantes (*Dickeya chrysanthemi*, *D. dadantii*, *Dickeya dianthicola*, *Dickeya fangzhongdai*, *Dickeya oryzae*, *Dickeya parazeae*, *Dickeya poaceaphila*, *Dickeya solani* et *Dickeya zae*), tandis que les trois autres ont été isolées à partir d'eau de rivières ou de lacs (*Dickeya aquatica*, *Dickeya lacustris* et *Dickeya undicola*) [21, 32]. Parmi les souches de ce genre, *Dickeya dadantii* 3937 est le principal modèle utilisé pour les études génétiques et transcriptomiques car son génome a été le premier séquencé et annoté [33, 34]. Une présentation plus détaillée du modèle *Dickeya* se trouve au Chapitre 2, en section 2.1 (à partir de la page 59).

### 1.1.3 Les découvertes de l'ADN

Comme décrit par Roger Stanier et Cornelis van Niel, les bactéries ne possèdent pas de membrane nucléaire. Contrairement à celui des eucaryotes, l'acide désoxyribonucléique

(ADN) bactérien est présent dans le cytoplasme. Il est classiquement sous la forme d'un unique chromosome circulaire. Cet ADN est le porteur de l'information génétique permettant le développement, le fonctionnement et la reproduction des organismes. Il est composé de gènes, qui peuvent être transcrits en ARN (acide ribonucléique), et de régions non transcrites. Les ARN synthétisés peuvent ensuite être traduits en protéines. Chez les bactéries, la traduction d'un ARNm peut commencer avant même que sa synthèse ne soit terminée (Fig. 1.1 en page 2) [11].

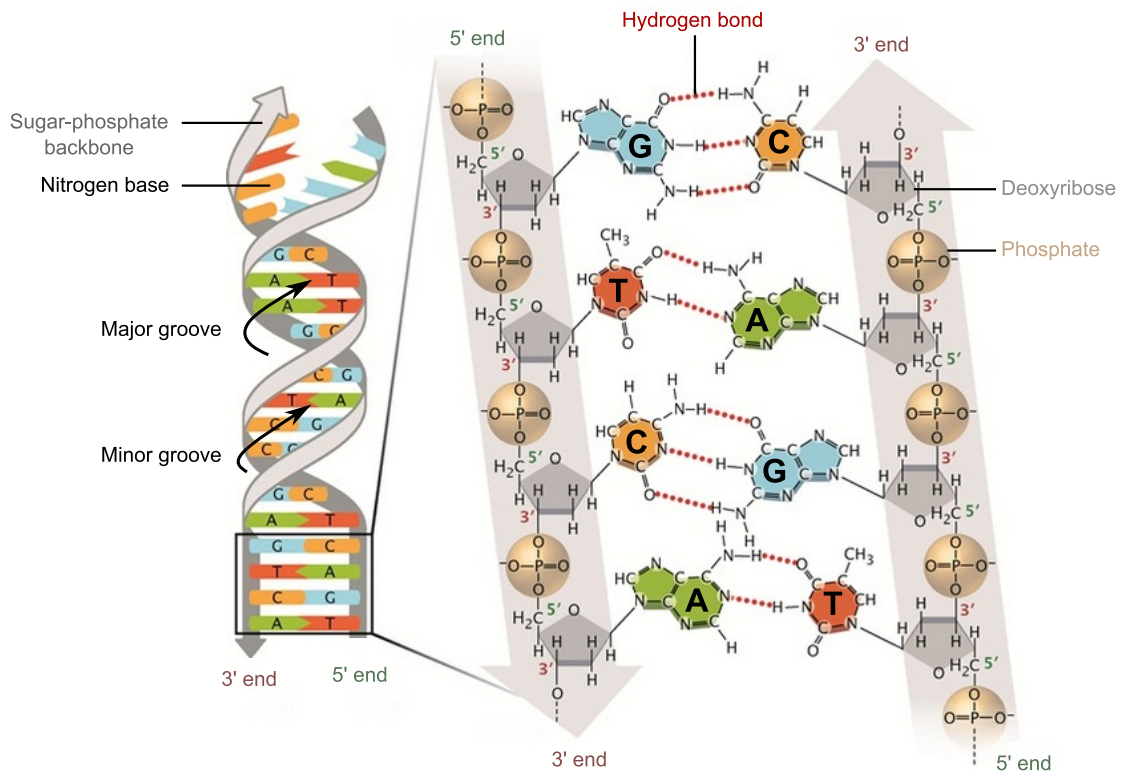


FIGURE 1.3 – La double hélice de l'ADN telle que proposée par James Watson et Francis Crick. Chacun des deux brins de la double hélice d'ADN est constitué d'un enchaînement de nucléotides. Un nucléotide est composé d'un désoxyribose, d'un groupement phosphate et d'une base azotée parmi l'adénine (A), la cytosine (C), la thymine (T) et la guanine (G). Les désoxyriboses et les groupements phosphates forment le squelette de chaque brin tandis que les bases relient les deux brins. Les bases A et T s'apparient par deux liaisons hydrogènes, les bases G et C par trois liaisons hydrogènes. L'hélice forme une alternance de sillons étroits (*minor groove*) et de sillons larges (*major groove*). Figure adaptée de [35].

L'ADN, nommé alors *nucléine*, fut isolé pour la première fois à la fin des années 1860 par Johann Friedrich Miescher [36]. En 1881, Albrecht Kossel précisa la composition de cette molécule en décrivant un acide nucléique composé de 5 bases azotées : l'adénine (A), la cytosine (C), la guanine (G), de la thymine (T) ou l'uracile (U) [37]. Au début du XX<sup>e</sup> siècle, les travaux de Phoebus Levene permirent de préciser les trois composants d'un nucléotide : une base (ATGC pour l'ADN ou AUGC pour l'ARN) mais également un désoxyribose



(dans le cas de l'ADN) ou un ribose (dans le cas de l'ARN) et un groupe phosphate [35, 38]. L'assemblage de ces nucléotides restait alors à déterminer.

En 1930, des études biochimiques menées par Erwin Chargaff révélèrent que l'ADN contient autant de G que de C et autant de A que de T. Puis, au début des années 1950, Rosalind Franklin et Maurice Wilkins proposèrent, grâce à des études de diffraction des rayons X, une structure en double hélice. Enfin, en combinant ces informations clés sur la structure de l'ADN, James Watson et Francis Cricks décrivirent l'actuel modèle de la double hélice de l'ADN : deux chaînes antiparallèles de nucléotides, appariés deux à deux par des liaisons hydrogènes, enroulées autour d'un axe central [35, 11] (Fig. 1.3). En raison de la géométrie des bases et de la torsion de la double hélice, ces chaînes forment un sillon large (sillon majeur, *major groove*) et un sillon plus étroit (sillon mineur, *minor groove*) [11] (Fig. 1.3).

#### 1.1.4 L'organisation du chromosome circulaire chez les bactéries

L'organisation du chromosome circulaire (Fig. 1.4A) est très conservée chez des bactéries même éloignées phylogénétiquement comme *E. coli* et *Bacillus subtilis* [39].

La réplication du chromosome bactérien est initiée à l'origine de réplication (*OriC* chez *E. coli*) et se déroule de manière bidirectionnelle, avec deux fourches de réplication qui se rencontrent au site *dif*. Les sites *Ori* et *dif* divisent le chromosome en deux réplichores, le réplichore gauche et le réplichore droit (Fig. 1.4A). Lors de la réplication, les fourches sont créées au niveau du site *Ori* et progressent vers le site de *dif* grâce aux hélicases qui séparent les deux brins, créant un stress torsionnel en amont des fourches de réplication. Au sein de ces fourches, un brin d'ADN, le brin direct (*leading strand*), est synthétisé de manière continue et le brin complémentaire, le brin indirect (*lagging strand*), est synthétisé de manière discontinue. La synthèse du brin direct est réalisée par l'ADN polymérase III qui ajoute des désoxyribonucléotides à l'extrémité 3' du brin en formation. La synthèse du brin indirect est plus complexe du fait du sens unique, du 3' vers le 5', de déplacement de l'ADN polymérase III sur le brin modèle. Pour ce brin indirect, la réplication est réalisée en plusieurs fragments, nommés fragments d'Okazaki, synthétisés du 3' vers le 5'. Chaque fragment est formé d'une amorce ARN synthétisée par une primase suivie d'un segment d'ADN synthétisé par l'ADN polymérase III, à partir de l'amorce ARN. L'amorce ARN est ensuite retirée par l'ADN polymérase I et remplacée par de l'ADN. Enfin, les fragments ainsi créés sont joints par une ligase afin de former le brin indirect complet. Le stress torsionnel créé en amont de la fourche de réplication se propage à travers la fourche et provoque la torsion des brins nouvellement formés autour du brin modèle, formant ainsi des caténanes qui sont décaténés à la suite de la réplication [11].

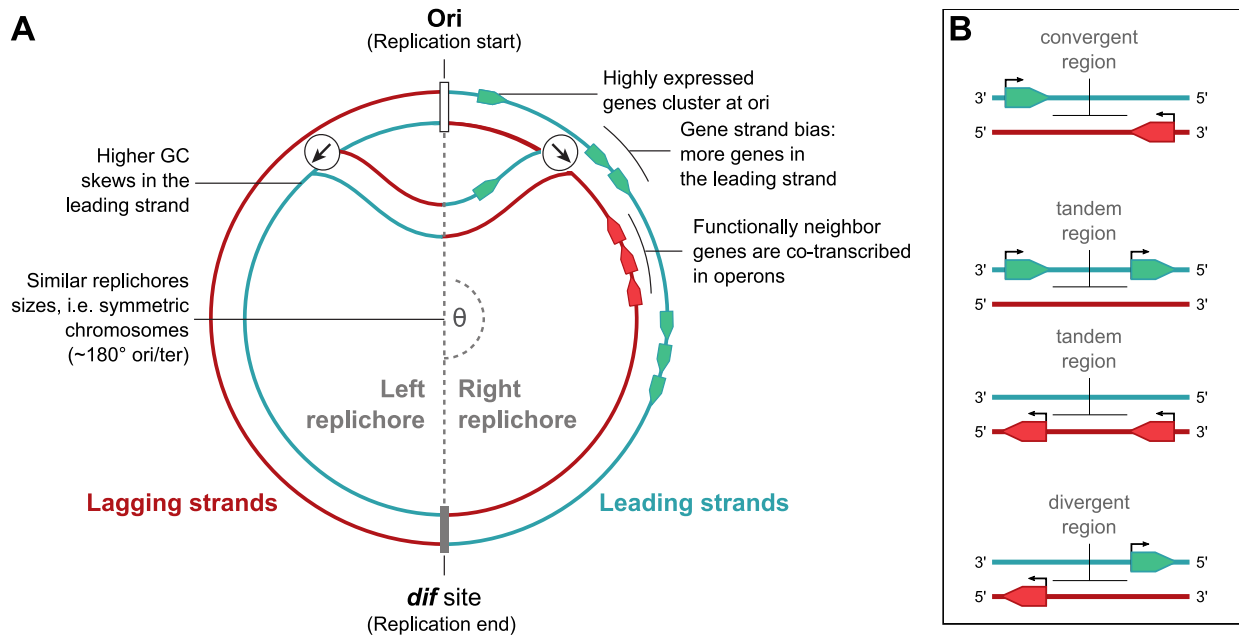


FIGURE 1.4 – L'organisation du génome bactérien. (A) Le génome bactérien est divisé en deux réplichores délimités par l'origine de réplication et par le site *dif* où les fourches de réplication se rencontrent. Lors de la réplication, le brin synthétisé en continu est appelé brin direct (*Leading strand*, vert), tandis que le brin synthétisé en discontinu est appelé brin indirect (*Lagging strand*, rouge). Le schéma, adapté de [39], présente également la répartition hétérogène des gènes sur le génome. (B) La classification des régions génomiques selon l'orientation des gènes adjacents. Entre deux gènes situés sur des brins différents, une région est considérée convergente si la transcription des gènes adjacents se fait de manière frontale et divergente si la lecture de ces gènes se fait dans la direction opposée à la région centrale. Si les gènes adjacents sont situés sur le même brin, la région entre ces gènes est tandem.

La localisation des gènes sur le chromosome circulaire bactérien n'est pas homogène. Tout d'abord, il a été remarqué que la région autour du site *Ori* est enrichie en gènes hautement exprimés pendant la croissance bactérienne (Fig. 1.4A). Cet enrichissement s'explique par la faculté des chromosomes bactériens à réaliser plusieurs cycles de réplication simultanément. En effet, en phase de croissance rapide, un nouveau cycle de réplication peut démarrer avant la fin du cycle précédent, ce qui entraîne une sur-représentation des gènes proches du site *Ori* [39]. Par ailleurs, il a été démontré que le brin direct est plus riche en gènes essentiels que le brin indirect [40]. Cette préférence s'explique par la fréquence plus élevée et les conséquences plus graves des collisions entre les ADN polymérases répliquant un brin d'ADN et les ARN polymérases transcrivant ce même brin sur le brin indirect. Sur le brin direct, la transcription se fait dans la même direction que l'avancée de la fourche de réplication. Tandis que sur le brin indirect, la transcription se fait dans la direction opposée à celle de l'avancée de la fourche de réplication, les collisions entre ADN polymérase et ARN polymérase y sont donc frontales et par conséquent plus graves que les collisions qui ont lieu sur le brin direct [41, 39]. Enfin, il a également été observé que le brin direct est souvent plus riche en G et le brin indirect plus riche en C [42, 39].

Plus localement, il a été démontré que des gènes codant pour des fonctions similaires, comme par exemple des enzymes appartenant à une même voie métabolique, ont tendance à être co-localisés sur le chromosome et peuvent former une unité polycistronique, un opéron. Un opéron est constitué d'une succession de gènes disposés sur le même brin et dont l'expression est contrôlée par le même promoteur. Cette co-localisation facilite leur expression simultanée ainsi que leur régulation commune [43].

Une région entre deux gènes peut être caractérisée selon l'orientation des gènes qui l'entourent, c'est à dire selon les brins (brin direct ou brin indirect) sur lesquels sont situés ces gènes. Pour cette caractérisation, nous utilisons les critères suivants (Fig. 1.4B) : (1) si les deux gènes voisins sont sur le même brin, la région est appelée tandem, (2) si les gènes sont situés sur des brins différents et si leur transcription se fait en direction de la région centrale, cette région est convergente, et (3) si les gènes sont sur des brins différents et si leur transcription se fait dans la direction opposée à la région centrale, cette région est divergente. Dans un complexe de 3 gènes, l'orientation du gène central par rapport à ses voisins peut également être définie avec ces mêmes critères.

### 1.1.5 La compaction de la chromatine

Les génomes bactériens sont plus petits et moins complexes que ceux des eucaryotes, avec généralement une unique copie de chaque gène (haploïdie) [11]. Un génome bactérien modèle comporte environ 5 millions de paires de bases (Mpb) et code pour environ 5000 protéines [44] mais ces nombres peuvent varier grandement d'une espèce à une autre (Tableau 1.1). Les brins d'ADN bactériens mesurent environ 1 mm (0,34 nm par pb) [11].

TABLEAU 1.1 – Les tailles des génomes bactériens

Organisme	Taille du génome	Nb de gènes codants pour des protéines	Source
<i>Nasuia deltocephalinicola</i>	0,1 Mbp	137	[45]
<i>Mycoplasma genitalium</i>	0,6 Mbp	< 500	[46]
<i>Escherichia coli</i> K-12 MG1655	4,6 Mpb	Environ 4300	[47]
<i>Dickeya dadantii</i> 3937	4,9 Mpb	Environ 4500	[33]
<i>Sorangium cellulosum</i> So0157-2	15 Mbp	Environ 11000	[48]

Il est estimé d'après un modèle biophysique qu'une molécule d'ADN non compactée représente un volume environ 1000 fois plus grand que le volume typique d'une cellule et pourtant des observations au microscope ont montré qu'elle n'occupe qu'environ 20% de la cellule [49, 50] (Fig. 1.5). Il est donc indéniable que l'ADN est compacté dans les cellules bactériennes et cette compaction a lieu dès la réplication de l'ADN, ce qui facilite notamment la distribution des molécules d'ADN synthétisées dans les cellules filles [11].

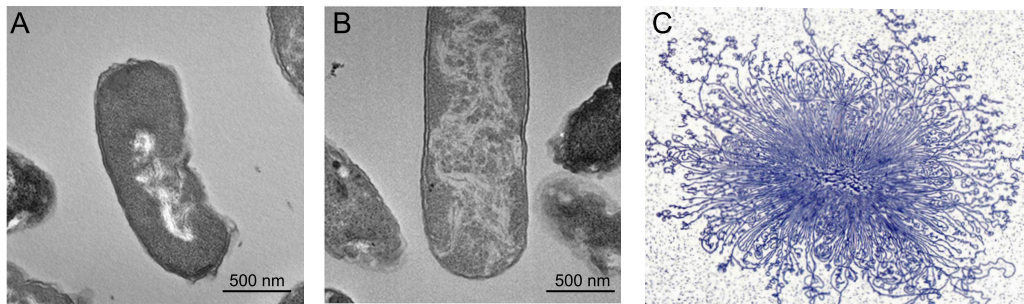


FIGURE 1.5 – Analyse par microscopie électronique en transmission de la morphologie du nucléoïde de *E. coli*. (A) Dans une souche *E. coli* sauvage, le nucléoïde (zone plus claire) est compacté au centre de la cellule. (B) Lorsque les gènes *hupA* et *hupB*, codant pour la protéine associée au nucléoïde HU, sont délétés, le nucléoïde est décondensé et semble occuper un espace plus grand (voir section 1.1.6). (C) Le nucléoïde isolé d'une bactérie lysée, coloré à l'acétate d'uranyle, représente une surface beaucoup plus grande que celle d'une cellule. Images issues de [51, 52].

### La compaction du chromosome bactérien à grande échelle : les macrodomaines identifiés chez *E. coli*

Chez les bactéries, la structuration du chromosome à grande échelle a principalement été étudiée chez *E. coli* où 4 macrodomaines, d'environ 1 Mpb, complétés par 2 régions non structurées ont été identifiés (Fig. 1.6A) [53, 54]. Les deux premiers macrodomaines ont été observés grâce à la méthode FISH (Fluorescence *in situ* hybridization). Cette méthode permet de localiser des segments d'ADN au sein d'une cellule grâce à des sondes fluorescentes spécifiques des segments d'ADN ciblés. Niki *et al.* ont ainsi observé que les segments autour du site *OriC* co-localisent et de même pour les segments autour du site *dif*. Ils ont également montré que, durant la réplication, les deux répliquats du domaine incluant *OriC* migrent vers les extrémités des cellules tandis que le domaine autour du site *dif* se positionne au centre de la cellule [53].

Deux autres macrodomaines et deux régions non structurées ont été identifiés par Valens *et al.* avec un système de recombinaison détectable qui permet d'estimer la proximité spatiale de sites distants sur le chromosome selon leur fréquence de recombinaison. Des sites proches recombinent à une fréquence plus élevée que des sites éloignés spatialement. Cette méthode permet donc de localiser des régions dans lesquelles les segments d'ADN sont spatialement proches et forment un macrodomaine. Quatre régions de ce type ont été identifiées : le macrodomaine autour du site *OriC*, le macrodomaine autour du site *dif* et deux macrodomaines (macrodomaine gauche et macrodomaine droite) situés de part et d'autre de ce dernier. D'après leur étude, les régions comprises entre le macrodomaine contenant *OriC* et les macrodomaines gauche et droite sont des régions moins structurées, dont les segments peuvent être spatialement proches des autres macrodomaines (Fig. 1.6A) [54]. Depuis, la mobilité de ces domaines a pu être observée en marquant par fluorescence des loci appartenant aux différents macrodomaines et il a été montré que

les macrodomaines ont une mobilité dans la cellule plus limitée que les régions non structurées [55].

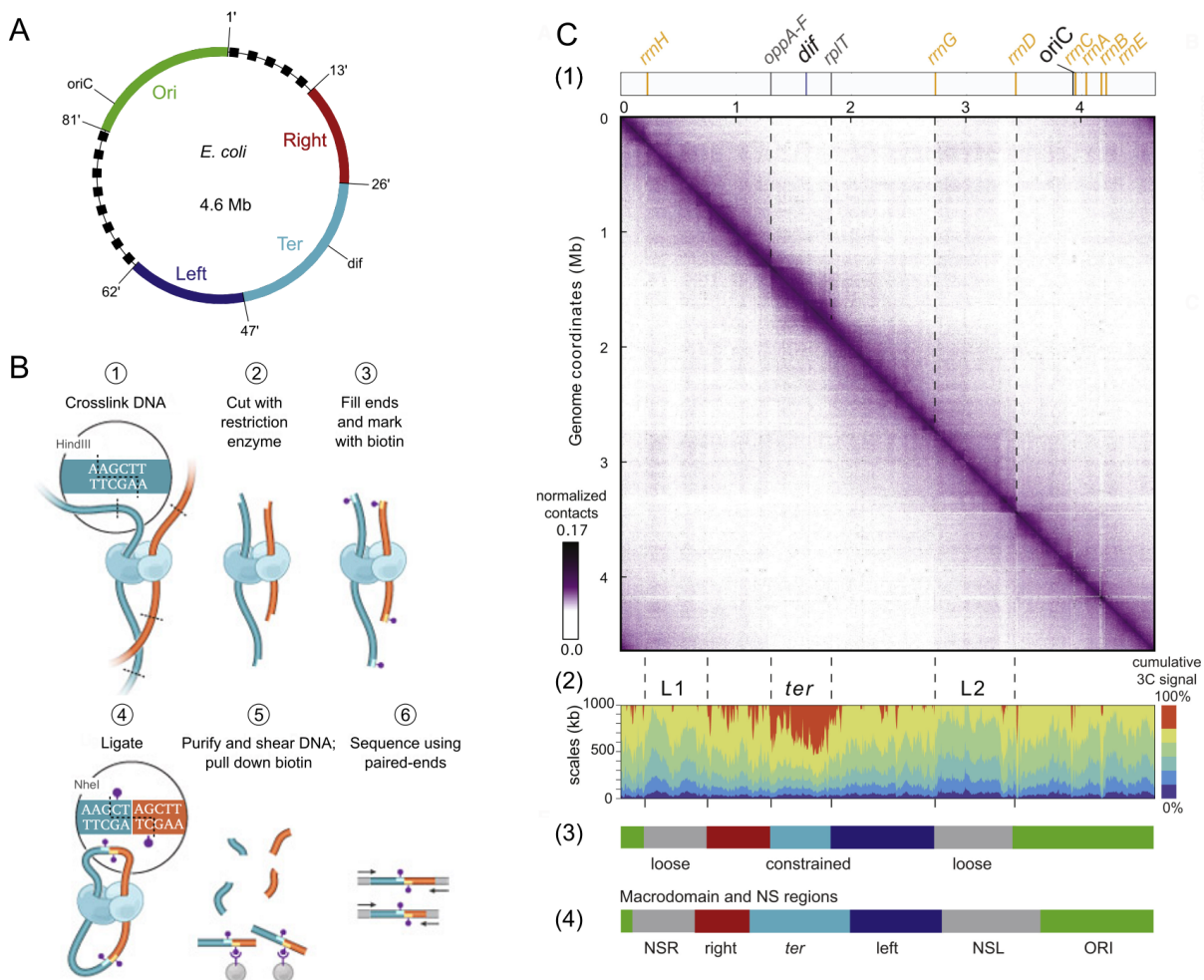


FIGURE 1.6 – L'étude des macrodomaines de *E. coli* par Hi-C. **(A)** Carte génétique des quatre macrodomaines (rouge, cyan, bleu et vert) et des deux régions moins structurées (pointillés noirs) du chromosome circulaire de *E. coli* identifiés par Valens *et al.* [54]. **(B)** Schéma du protocole Hi-C issu de [56]. (1) Les fragments d'ADN (courbes bleu et orange) interagissant au sein de la cellule sont pontés lors de l'ajout du formaldéhyde. Ces fragments sont liés par des liaisons covalentes ADN-protéines (représentées en cyan et bleu). (2) L'ADN est fragmenté par des enzymes de restriction. (3) Les extrémités des fragments sont biotinylées (points violets). (4) Les fragments de chaque paire sont ligués entre eux, avec le marquage à la biotine sur leur site de jonction. (5) L'ADN est purifié, fragmenté, puis les fragments contenant les jonctions biotinylées sont sélectionnés par affinité sur des billes de streptavidine. (6) Les fragments isolés sont séquencés afin d'identifier les loci du chromosome interagissant. **(C)** Carte des contacts chromosomiques de *E. coli* et analyse des macrodomaines réalisées par Lioy *et al.* [57]. (1) La carte normalisée des contacts présente la fréquence d'interaction entre deux positions du génome (coordonnées génomiques sur les axes x et y). Sur cette carte, plus les contacts sont fréquents, plus le croisement des positions est de couleur violette foncée. (2) Représentation en scalogramme de la fraction des contacts effectués par chaque région au sein de fenêtres de tailles croissantes. La taille de la fenêtre autour de la région considérée est présentée sur l'axe y et la fraction des contacts est indiquée par la couleur : le bleu foncé correspond aux fractions comprises entre 0 et 15%, le rouge aux fractions comprises entre 75 et 100%, les autres couleurs correspondent aux fractions intermédiaires présentées sur l'échelle à droite du scalogramme. (3) et (4) Comparaison des positionnements des quatre macrodomaines et des deux régions moins structurées sur le chromosome de *E. coli* d'après l'étude Hi-C (3) et d'après les études précédentes (4, également présenté en A). NSL et NSR correspondent aux deux régions moins structurées.

Enfin, la localisation de ces macrodomaines a pu être précisée grâce au développement des méthodes de détection des contacts chromosomiques. Les méthodes de microscopie permettent d'étudier quelques loci uniquement tandis que les techniques de capture de la conformation du chromosome (3C) permettent d'étudier l'ensemble des interactions du chromosome. Le Hi-C est une extension de ces méthodes 3C (Fig. 1.6 B et C). Le Hi-C est capable de détecter ces interactions avec une haute résolution, y compris les interactions à longue distance [56]. Pour ce faire, les cellules à analyser sont fixées par du formaldéhyde afin de ponter les loci interagissant via des liaisons covalentes ADN-protéines (Fig. 1.6 B). L'ADN est ensuite extrait et fragmenté avec des enzymes de restrictions. Grâce au pontage, les fragments interagissant restent associés par paires. Les extrémités des fragments sont marquées avec de la biotine puis liguées dans des conditions favorisant la ligation des paires de fragments pontés. Les extrémités des fragments sont ainsi liées de manière covalente par paire, avec une insertion de biotine au niveau de leurs deux sites de jonction (Fig. 1.6B, étape 4). La librairie de produits de ligatures est ensuite purifiée puis de nouveau fragmentée. Enfin, seuls les fragments contenant de la biotine sont isolés grâce à une sélection par affinité sur billes de streptavidine. Ces fragments sont séquencés afin d'identifier les loci qui étaient en contact dans la cellule. En utilisant cette méthode, Lioy *et al.* ont pu observer des régions présentant un grand nombre d'interactions au sein des régions et peu d'interactions avec les autres régions (Fig. 1.6 C). Ces régions correspondent aux macrodomaines précédemment identifiés.

### La compaction du chromosome bactérien à moyenne et petite échelles

A l'échelle d'environ  $10^4$  -  $10^5$  pb, l'ADN est structuré en domaines topologiques nommés *chromosomal interaction domains* (CIDs) [58]. Tout comme les macrodomaines, les CIDs sont caractérisés par une haute fréquence de contacts au sein d'une région chromosomique. La différence avec les macrodomaines est la taille de ces régions puisque les CIDs identifiés chez différentes bactéries grâce à la technologie Hi-C présentée précédemment (Fig. 1.6), ont une taille de l'ordre de 10 à 100kb (Tableau 1.2).

TABLEAU 1.2 – Les domaines d'interactions chromosomiques

Organisme	Nombre de CIDs identifiés	Taille des domaines	Source
<i>Escherichia coli</i>	31	40 à 300 kb	[57]
<i>Caulobacter crescentus</i>	23	30 à 420 kb	[59]
<i>Bacillus subtilis</i>	Environ 20	50 à 300 kb	[60]
<i>Mycoplasma pneumoniae</i>	44	15 à 33 kb	[61]

Les mécanismes à l'origine de la formation de ces domaines sont encore méconnus mais semblent impliquer les protéines de maintenance structurale du chromosome (SMC) (plus

de détails au paragraphe 1.1.6). Il a par ailleurs été remarqué que les frontières entre ces domaines sont enrichies en gènes fortement exprimés [57].

A plus petite échelle, la compaction de l'ADN est obtenue grâce aux rôles conjoints du surenroulement de l'ADN (SC), des protéines associées au nucléoïde (NAPs) et des protéines SMC [58, 62, 63]. Ces acteurs, détaillés dans le paragraphe 1.1.6, organisent dynamiquement le génome avec des torsions, des boucles et des raidissements, afin de faciliter les processus vitaux comme la réplication et la transcription.

Ainsi, ces différentes échelles de compaction du chromosome bactérien permettent de réduire le volume occupé par le nucléoïde dans la cellule et participent à de nombreux processus cellulaires. La disposition à grande échelle de certaines régions du chromosome facilite par exemple la réplication. L'organisation à petite échelle de l'ADN permet quant à elle de rendre certaines régions plus accessibles par les protéines comme les ARN polymérases pour faciliter la transcription des gènes situés dans ces régions par exemple (voir section 1.4).

### Les différentes échelles de la compaction du chromosome eucaryote

Dans les noyaux des cellules eucaryotes, l'ADN est également compacté à différentes échelles (Fig. 1.7A). Chaque chromosome occupe son propre territoire et forme des compartiments. Les deux compartiments principaux sont l'hétérochromatine, contenant généralement les régions les moins transcrites, et l'euchromatine, contenant les plus transcrites [64].

A moyenne échelle (Fig. 1.7Aiii), similairement aux CIDs bactériens, l'ADN est structuré en domaines topologiques, nommés *topologically associated domains* (TADs) [58, 65]. Ces TADs mesurent de l'ordre de  $10^5$  -  $10^6$  pb. Leur plus grande taille, en accord avec la plus grande taille de leur génome, a permis leur identification avant celle des CIDs bactériens.

A plus petite échelle (Fig. 1.7Ai), l'ADN est également compacté par des protéines de tailles nanométriques qui se lient à l'ADN et modifient sa géométrie [58]. Chez les eucaryotes, ces protéines bien connues sont les histones, octamères de protéines autour desquelles 145 à 147 pb peuvent s'enrouler afin de former un nucléosome [66].

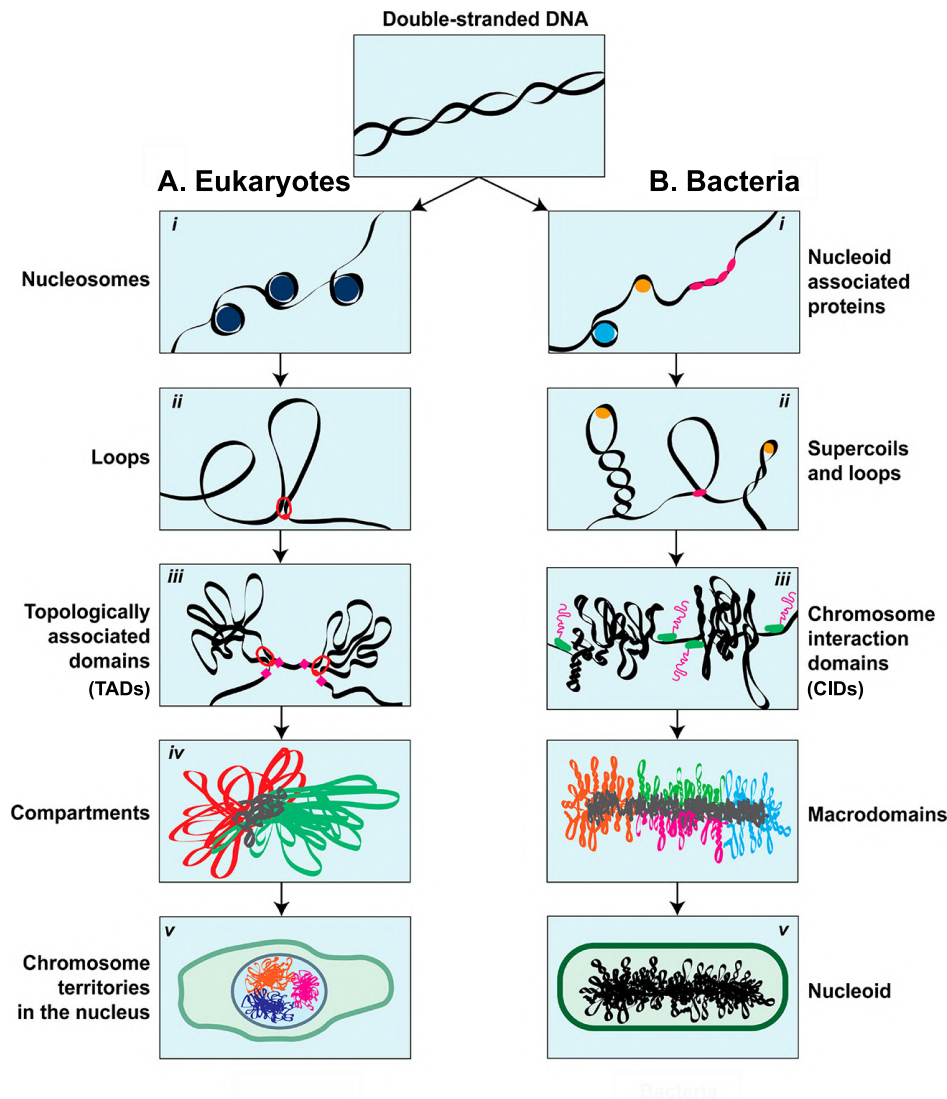


FIGURE 1.7 – L’organisation des chromosomes eucaryotes (A) et bactériens (B) à différentes échelles. (A, i) Chez les eucaryotes, l’ADN double brin est enroulé autour des histones pour former des nucléosomes. (B, i) L’ADN double brin des bactéries est structuré par les protéines associées au nucléoïde (voir section 1.1.6) qui peuvent le courber, le raidir ou former des ponts entre différents segments d’ADN. (A,ii) A plus grande échelle, l’ADN des eucaryotes forme des boucles, soit par pontage, soit par extrusion. (B, ii) L’ADN bactérien est soumis à des déformations torsionnelles, il est surenroulé négativement (voir section 1.1.6). Similairement aux eucaryotes, des boucles peuvent également être formées. (A et B, iii) Que ce soit au sein des chromosomes bactériens ou eucaryotes, des domaines caractérisés par une haute fréquence de contacts au sein du domaine ont été identifiés. (A et B, iv) Au delà des domaines d’interactions, les chromosomes eucaryotes et bactériens sont divisés en compartiments ou macrodomaines. (A, v) Les chromosomes eucaryotes occupent chacun un territoire au sein du noyau. (B, v) Le chromosome bactérien, appelé nucléoïde, n’est a priori pas compartimenté, il est condensé au centre du cytoplasme. Schéma issu de [58].



### 1.1.6 Les acteurs de la compaction de la chromatine bactérienne

#### Les protéines associées au nucléoïde

Les protéines associées au nucléoïde (NAPs) organisent la structure de l'ADN de manière dynamique, en y formant des ponts entre des brins (*bridging*), des courbures (*bending*, *wrapping*) ou à l'inverse des raidissements (*stiffening*) (Fig. 1.8) [63]. Toutes les espèces bactériennes ont au moins une NAP [67].

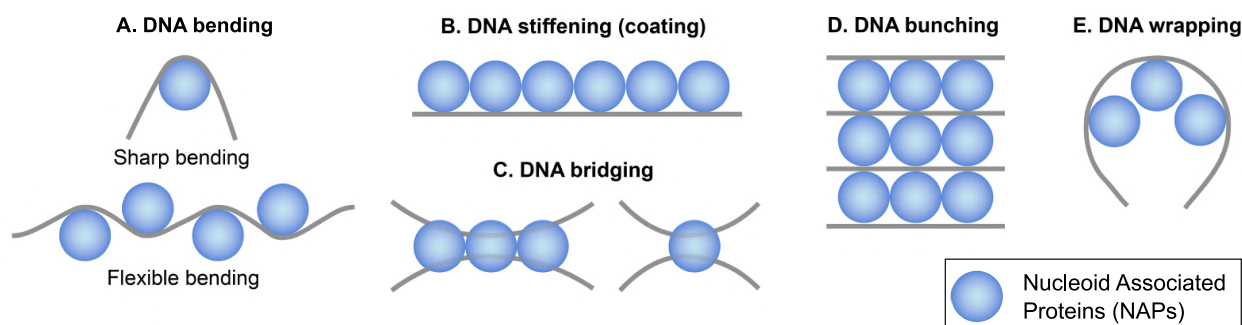


FIGURE 1.8 – La structuration de l'ADN (courbes grises) par les protéines associées au nucléoïde (NAPs, sphères bleues). Les NAPs peuvent (A) courber l'ADN, (B) le raidir, ou (C) faire des ponts en se liant simultanément à une paire de segments d'ADN, adjacents ou non. (D) La multimérisation latérale de NAPs peut permettre de rassembler différents segments d'ADN. (E) Lorsque plusieurs NAP courbant l'ADN se lient à des sites adjacents, l'ADN s'enroule autour de celles-ci. Figure issue de [68].

Au moins 12 NAPs ont été identifiées chez *E. coli*, parmi elles HU (*Histone-like protein from E. coli strain U93*), FIS (*Factor for Inversion Stimulation*), IHF (*Integration Host Factor*), H-NS (*Histone-like Nucleoid-Structuring protein*) et Dps (*DNA-binding protein from starved cells*) sont les plus étudiées [68]. L'abondance de chacune de ces NAPs dépend de la phase de croissance d'*E. coli* (Fig. 1.9A).

HU est l'une des protéines les plus abondantes chez *E. coli* et dans les souches qui en sont dépourvues, le nucléoïde est décondensé (Fig. 1.5B). A faible concentration (moins de 100 nM), HU se lie de manière stable à l'ADN en formant des courbures (Fig. 1.8A *bending*, Fig. 1.9B), tandis qu'à des concentrations plus élevées, HU forme des filaments rigides qui entraînent l'aplatissement de l'ADN (Fig. 1.8B, *stiffening*) [69]. Ces filaments peuvent ensuite se regrouper lorsque les HU se multimérisent (Fig. 1.8D *bunching*). Des filaments rigides similaires peuvent également être créés lorsque H-NS se propage le long des séquences d'ADN riches en AT, tant que la concentration de magnésium est inférieure à 2 mM. En présence de plus de 5 mM de magnésium, H-NS forme des ponts entre différents segments d'ADN, adjacents ou distants (Fig. 1.8C *bridging*, Fig. 1.9B) [70]. IHF et FIS se lient préférentiellement à des séquences spécifiques de l'ADN et y courbent la double hélice (Fig. 1.8A *bending*, Fig. 1.9B). IHF courbe l'ADN avec un angle de plus de 160° (*sharp bending*). Bien qu'absente en phase stationnaire, FIS est aussi abondante que

HU en début de phase exponentielle de croissance. Du fait de cette abondance, elle se lie également de manière non spécifique à l'ADN et peut former des ponts entre différents segments de l'ADN (Fig. 1.8C, *bridging*) [68].

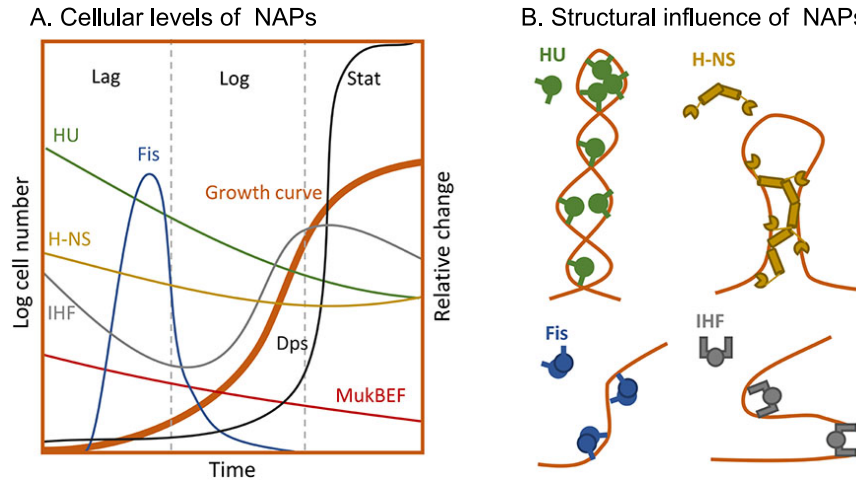


FIGURE 1.9 – (A) Niveau intracellulaire de différentes NAPs pendant les phases de croissance d'*E. coli*. (B) Impact des NAPs sur la structure de l'ADN. Figure issue de [71].

Chez *D. dadantii*, H-NS, Fis et IHF sont également présentes et régulent l'expression de divers gènes. Des expérimentations sur des souches de *D. dadantii* mutées au niveau de l'un des gènes codant pour ces NAPs ont montré que ces trois NAPs jouent un rôle important dans le processus infectieux. En effet, ces NAPs impactent l'expression de facteurs de virulence majeurs dont les gènes *pelD* et *pelE*. Ces gènes codent pour des pectinases, enzymes de dégradation de la paroi cellulaire des plantes, et leur expression est essentielle à la réussite du processus infectieux par cette bactérie (voir Chapitre 2 section 2.1.2 à partir de la page 61) [72, 73, 74, 75].

### Les protéines de maintenance structurelle du chromosome (SMC)

Chez de nombreux eucaryotes et procaryotes, des grands complexes protéiques ont été caractérisés et semblent essentiels à la ségrégation et à la compaction des chromosomes. Ces complexes se reconnaissent à leurs longues sous-unités protéiques en super-hélices (*coiled-coil*), nommées protéines de maintenance structurelle du chromosome (SMC), qui modifient la topologie de l'ADN en hydrolysant de l'ATP [76]. Chez les procaryotes, les protéines SMC s'assemblent avec des sous-unités additionnelles pour former ScpAB dans le cas de *B. subtilis* ou MukBEF dans le cas de *E. coli* (Fig. 1.10) [77]. Chez les eucaryotes, les complexes SMC sont formés à partir d'hétérodimères de différentes sous-unités SMC afin de former différents complexes ayant des fonctions distinctes : la cohésine, la condensine et le complexe Smc5/6 [78]. La cohésine participe à la formation des TADs et au maintien de l'association entre les chromatides soeurs tandis que la condensine participe à la

compaction et à l'individualisation des chromatides pour faciliter leur ségrégation. Les complexes Smc5/6 participent à la réparation des dommages de l'ADN et à la résolution des problèmes topologiques liés à la réplication de l'ADN [76].

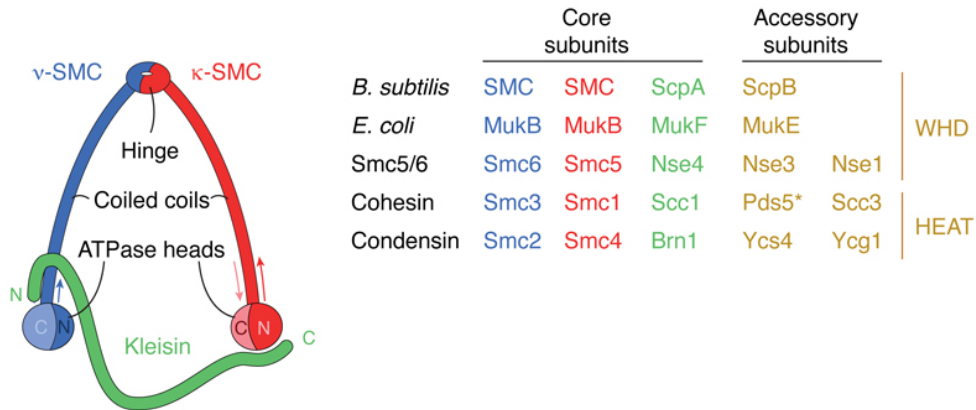


FIGURE 1.10 – Le complexe SMC forme une boucle composée de deux sous-unités SMC identiques (en bleu et en rouge) et d'une sous-unité "Kleisin" (en vert). Chacune des deux sous-unités SMC contient une longue chaîne en super-hélice (*coiled-coil*, une extrémité globulaire et une extrémité qui constitue la charnière lors de l'association des deux sous-unités). Le tableau à droite de la figure présente la composition des complexes SMC de *B. subtilis*, *E. coli* et de *S. cerevisiae* (trois dernières lignes). Figure issue de [76].

Les protéines SMC sont généralement constituées de plus de 1000 acides aminés et comportent cinq domaines [79] disposés de manière symétrique : deux domaines globulaires aux extrémités (N-terminale et C-terminale), deux longues chaînes en super-hélices (*coiled coil*) et un domaine globulaire charnière à la jonction des deux longues chaînes (Fig. 1.10). Les deux extrémités N- et C-terminales contiennent les sites de liaisons et d'hydrolyse de l'ATP et sont reliées par une sous-unité nommée "Kleisin". La structure ainsi formée a la forme d'une boucle fermée pouvant encercler l'ADN [76, 79].

Différents mécanismes d'extrusion de boucles par les complexes SMC sont toujours à l'étude. L'un d'entre eux, particulièrement étudié chez les procaryotes, est présenté en Fig. 1.11. D'après ce mécanisme, le complexe SMC, recruté par ParB, se lie à l'ADN au niveau d'un site *parS* [80]. Le complexe SMC adopte ensuite alternativement une conformation ouverte et une conformation fermée afin de réaliser l'extrusion d'une boucle d'ADN. En conformation ouverte, le complexe peut piéger une boucle d'ADN entre les deux longues chaînes du complexe, dans un espace nommé *Pro-chamber*. L'hydrolyse d'ATP provoque ensuite la déstabilisation de la conformation ouverte et sa fermeture à partir de la charnière. La boucle d'ADN piégée est alors poussée vers un espace, nommé *Meta-chamber*, formé entre le fragment "Kleisin" et les extrémités globulaires N- et C-terminales. Avec ces deux compartiments, le complexe SMC peut ainsi piéger de nouvelles boucles d'ADN dans la *Pro-chamber* puis les ajouter à la boucle accumulée dans la *Meta-chamber* [81, 76].

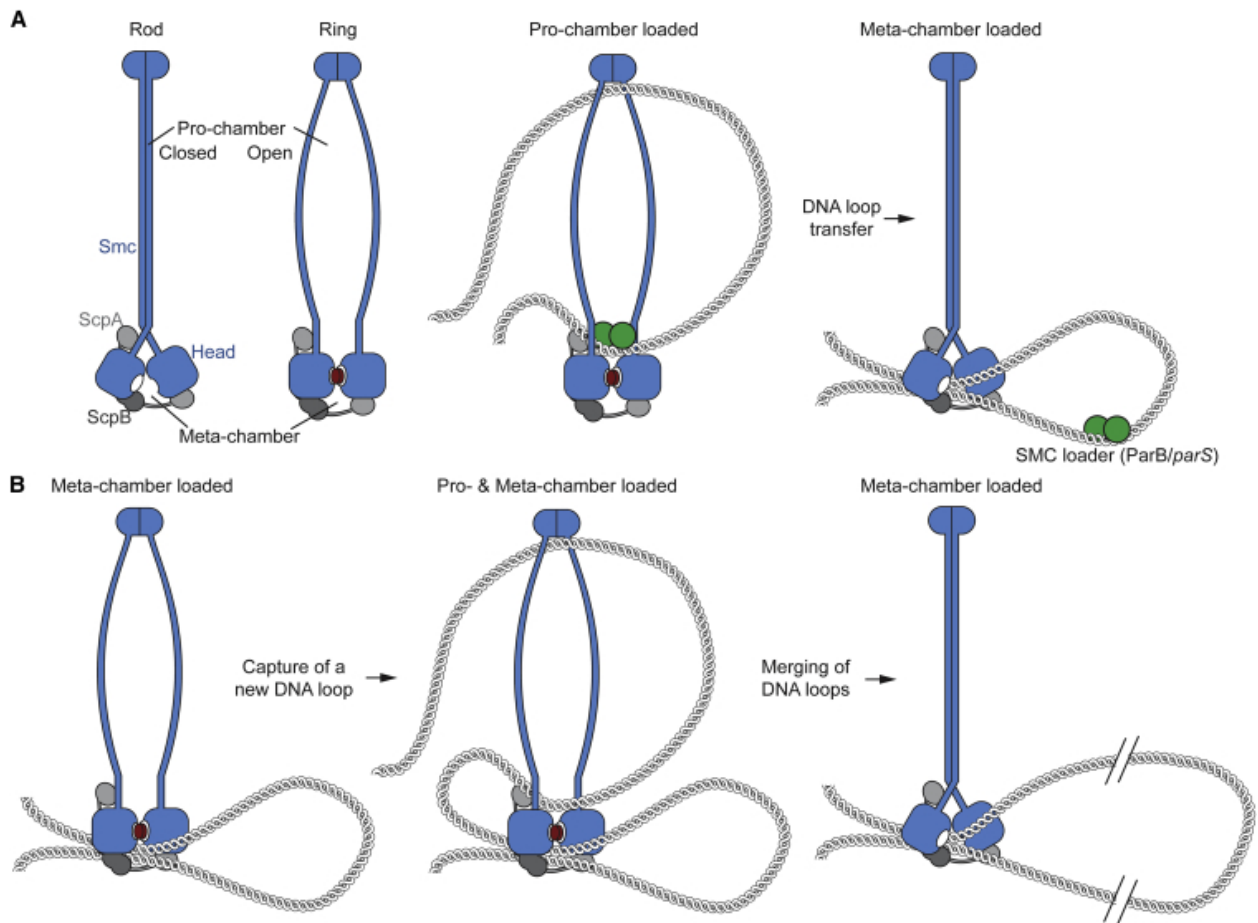


FIGURE 1.11 – Modèle proposé d'un mécanisme d'extrusion de boucle par un complexe SMC à double chambres, appliqué à *B. subtilis*. Le mécanisme peut être décomposé en deux parties : la capture d'une section du chromosome (A) et l'extrusion de la boucle (B). (A) Le complexe peut adopter une conformation fermée et une conformation ouverte. En conformation ouverte, une boucle d'ADN peut être capturée dans l'espace formé entre les deux longues chaînes (pro-chambre). Au niveau d'un site *parS*, l'association *ParB/parS* (en vert sur la figure) se lie au complexe SMC ouvert puis l'hydrolyse d'ATP provoque le changement conformationnel du complexe SMC. La pro-chambre est fermée et l'ADN est poussé dans la méta-chambre formée entre les extrémités N- et C-terminales des sous-unités SMC et le fragment "Kleisin". (B) La succession de cycles de capture de boucles d'ADN adjacentes et de changements conformationnels aboutit à l'extrusion d'une grande boucle d'ADN dans la méta-chambre. Figure issue de [81].

## Le surenroulement de l'ADN

Le surenroulement de l'ADN (SC) désigne la déformation torsionnelle de la double hélice de l'ADN, par rotation de l'axe de la molécule d'ADN (vrillage, *writhe*) ou autour de cet axe (torsion, *twist*) (Fig. 1.12) [82]. Cette déformation torsionnelle joue un rôle essentiel dans la compaction du chromosome bactérien et peut à la fois impacter et être impactée par tous les processus biologiques nécessitant la dissociation des deux brins de la double hélice de l'ADN.

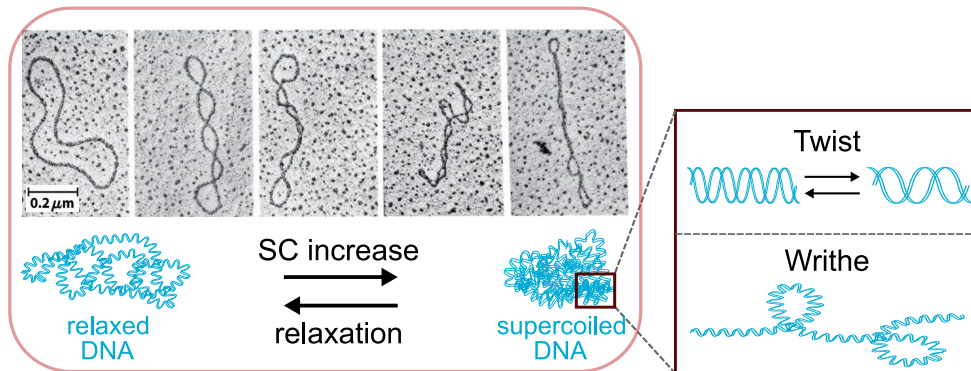


FIGURE 1.12 – Différents niveaux de surenroulement (SC) d'un plasmide observés par microscopie électronique. Le plasmide à gauche est relâché et son niveau de SC augmente sur les images de droite. A l'échelle locale, ce surenroulement prend la forme de torsions (*twist*) et de vrillages (*writhe*). Schéma adapté de [83] et images issues de [84].

L'ADN bactérien est généralement surenroulé négativement, c'est à dire que des supertours négatifs (dans le sens opposé de l'hélice) sont présents dans la double-hélice à hauteur de 1 supertour négatif par 300 pb en moyenne [11]. En comparaison à un ADN dit relâché, l'ADN négativement surenroulé a moins de tours d'hélice et donc plus de pb par tour d'hélice.

Le surenroulement peut être décrit numériquement par le nombre de liaisons (*Linking number*,  $Lk$ ), qui représente le nombre de croisements qu'un brin simple d'ADN effectue avec l'autre brin dans une molécule d'ADN circulaire. Ce nombre est la somme des torsions (nombre de tours d'hélice, voir *Twist* sur la Fig. 1.12) et des vrillages (nombre de croisements de l'hélice avec elle-même, voir *Writhe* sur la Fig. 1.12)). Tant que les brins d'ADN ne sont pas coupés, le nombre de liaison ( $Lk$ ) reste constant dans une molécule d'ADN circulaire, bien que des torsions puissent se transformer en vrillages et vice-versa [85].

Le  $Lk$  d'une molécule d'ADN circulaire relâchée est défini par  $Lk_0 = \frac{N}{\gamma}$  avec  $N$  le nombre de pb dans la molécule et  $\gamma$  le nombre de pb par tour d'hélice (10,5 pb en moyenne). Une molécule d'ADN surenroulée positivement ( $SC+$ ) a moins de pb par tour d'hélice, ce qui

se traduit par un  $Lk$  supérieur à  $Lk_0$ . À l'inverse, une molécule d'ADN surenroulée négativement (SC-) a un nombre de liaison ( $Lk$ ) inférieur à  $Lk_0$ . Ainsi, on parle de relâchement de l'ADN lorsque le  $Lk$  s'approche de  $Lk_0$ , d'augmentation du niveau de SC+ lorsqu'un  $Lk$  supérieur à  $Lk_0$  augmente et d'augmentation du niveau de SC- lorsqu'un  $Lk$  inférieur à  $Lk_0$  diminue. La mesure la plus couramment utilisée pour évaluer le surenroulement d'une molécule d'ADN est la densité de surenroulement  $\sigma = \frac{Lk-Lk_0}{Lk_0}$  [86]. Cette mesure permet de quantifier le niveau de surenroulement d'une molécule d'ADN par rapport à son état relâché. L'ADN bactérien a généralement une densité de surenroulement globale comprise entre  $\sigma = -0.03$  et  $\sigma = -0.09$  [87, 86] selon les conditions environnementales et la phase de croissance.

La théorie permet, pour un modèle donné d'ADN nu dans un solvant, de connaître le temps nécessaire au relâchement d'une molécule d'ADN surenroulée de taille connue. Dans la cellule, l'ADN n'est cependant pas nu, des barrières topologiques empêchent la propagation des supertours à longue distance, ce qui complique son étude [11].

## 1.2 La transcription bactérienne

Le processus de l'expression des gènes est composé de deux étapes : la transcription des gènes en ARN puis la traduction de certains ARN en protéines. Lors de la première étape, différents types d'ARN sont synthétisés, les ARN ribosomiaux (ARNr), les ARN messagers (ARNm) et les ARN de transfert (ARNt). La collaboration de ces trois types d'ARN permet ensuite la synthèse des protéines. En effet, les ARNm portent les informations codées par les gènes jusqu'aux ribosomes, qui constituent la machinerie de la synthèse protéique. Au sein du ribosome, la traduction de l'ARNm en protéine, est réalisée par l'intermédiaire des ARNt, qui vont associer chaque séquence de 3 nucléotides (un codon) de l'ARNm à un acide aminé spécifique. L'enchaînement de ces acides aminés constitue la protéine. Du fait de l'absence de noyau dans les cellules bactériennes, la traduction d'un ARNm peut commencer avant que sa synthèse ne soit terminée (Fig. 1.1, partie gauche) [11].

### 1.2.1 L'ARN polymérase

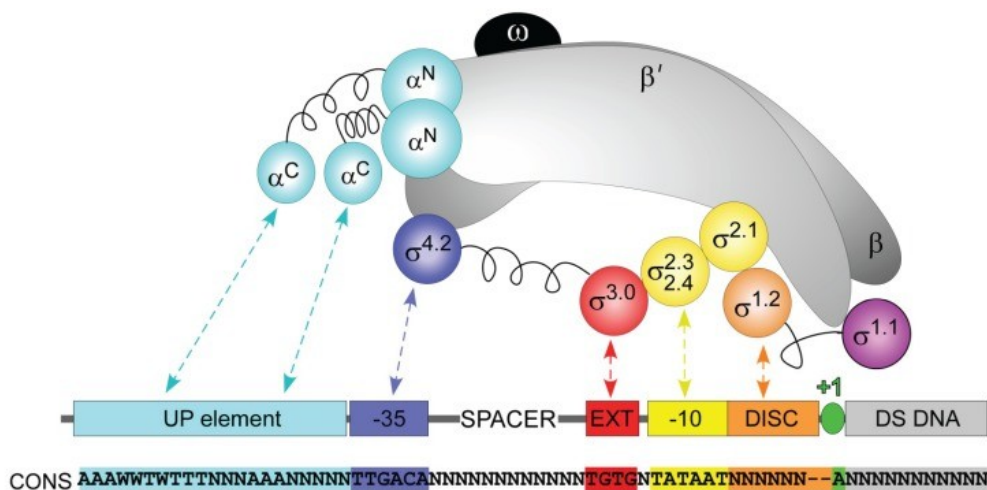


FIGURE 1.13 – Schématisation des interactions d'une ARN polymérase avec une séquence promotrice consensus. Le cœur de l'ARN polymérase est constitué de deux sous-unités  $\alpha$  identiques, de deux sous-unités très larges  $\beta$  et  $\beta'$  ainsi que d'une sous-unité  $\omega$ . A ces cinq sous-unités s'ajoute, pour former l'holoenzyme, le facteur  $\sigma$  lui même composé de plusieurs domaines interagissant avec des séquences spécifiques. La séquence CONS présentée est une séquence promotrice consensus. Elle contient les éléments UP (*upstream*), -35, -10 et -10 étendu (EXT). La région séparant l'élément -35 et le -10 étendu est appelée spacer. Celle située entre le -10 et le site de début de transcription (+1) est le discriminateur (DISC) [88].

Contrairement aux eucaryotes, les bactéries ont une unique ARN polymérase (ARNP) pour synthétiser tous leurs ARN, à l'exception des amorces d'ARN utilisées lors de la réplication. Chaque cellule bactérienne contient environ 2000 ARNP en phase lente de croissance, plus encore en phase exponentielle [11]. Chez *E. coli*, il a été déterminé que le

nombre total d'ARNP par cellule augmentait de 1500 à 11400 entre les phases de croissance lente (0,6 doublement par heure) et rapide (2,5 doublements par heure) [89].

La séquence et la composition de l'enzyme cœur de l'ARNP sont conservées chez les bactéries. Bien que les ARNP des eucaryotes et des archées aient plus de sous-unités et des séquences très différentes, leur structure basique est très similaire à celles des bactéries [90].

L'enzyme cœur de l'ARNP est composée de (1) deux sous-unités  $\alpha$  identiques, (2) deux sous-unités très larges,  $\beta$  et  $\beta'$ , qui forment une pince autour du site actif, et (3) d'une sous-unité  $\omega$ , qui aide leur assemblage [88, 11] (Fig. 1.13). Cette enzyme cœur s'associe à une sous-unité  $\sigma$  (facteur  $\sigma$ ) pour former l'holoenzyme. Ce facteur  $\sigma$ , composé de plusieurs domaines reliés de manière flexible, est chargé de la reconnaissance spécifique des promoteurs. Différents facteurs  $\sigma$  existent, 19 chez *Bacillus subtilis*, 60 chez *Streptomyces coelicolor*, 7 chez *E. coli* (voir section 1.2.3) [91, 92].

Ces facteurs  $\sigma$  reconnaissent des séquences promotrices consensus grâce à leurs différents domaines qui interagissent généralement avec les régions centrées autour des régions -35 et -10, à l'exception des facteurs de la famille de  $\sigma^{54}$  qui interagissent au niveau des positions -24 et -12 [92]. La séquence consensus la plus commune chez *E. coli* est celle reconnue par le facteur  $\sigma_{70}$ . Cette séquence type contient deux courtes régions riches en AT, les éléments -10 et -35, centrés respectivement environ 10 pb et 35 pb en amont du site de début de transcription (TSS). L'élément -10 est reconnu par les domaines  $\sigma_{2.3}$  et  $\sigma_{2.4}$ , l'élément -35 par le domaine  $\sigma_{4.2}$ . Ces deux éléments sont séparés par un spacer d'une taille d'environ 17 pb, ce qui correspond à l'écart entre les domaines  $\sigma_{2.3}$  et  $\sigma_{4.2}$  [88, 11]. Juste avant l'élément -10, l'élément -10 étendu interagit également avec le facteur  $\sigma$  (domaine  $\sigma_{3.0}$ ) et est particulièrement important pour la reconnaissance des promoteurs dont l'élément -35 est mal reconnu [93]. Entre l'élément -10 et le TSS se situe le discriminateur, une séquence de 5 à 8 pb, également liée par le facteur  $\sigma$  (domaine  $\sigma_{1.2}$ ) [94]. Le domaine  $\sigma_{1.1}$  ne semble pas interagir avec la séquence consensus mais joue un rôle essentiel puisqu'il inhibe la liaison des facteurs  $\sigma$  à l'ADN en absence d'ARNP cœur [88].

Outre les interactions directes entre le facteur  $\sigma$  et la séquence consensus promotrice présentées ci-dessus, d'autres interactions peuvent faciliter la reconnaissance et la liaison de l'ARNP au promoteur. Ces interactions peuvent impliquer des facteurs de transcription ou l'une des sous-unités de l'ARNP cœur. C'est notamment le cas de l'interaction stabilisante entre la sous-unité  $\alpha$  de l'enzyme cœur de l'ARNP avec l'élément UP (upstream) [11, 95].



## 1.2.2 Les étapes de la transcription bactérienne

Le cycle de la transcription comporte trois grandes étapes : l'initiation, l'élongation et la terminaison (Fig. 1.14).

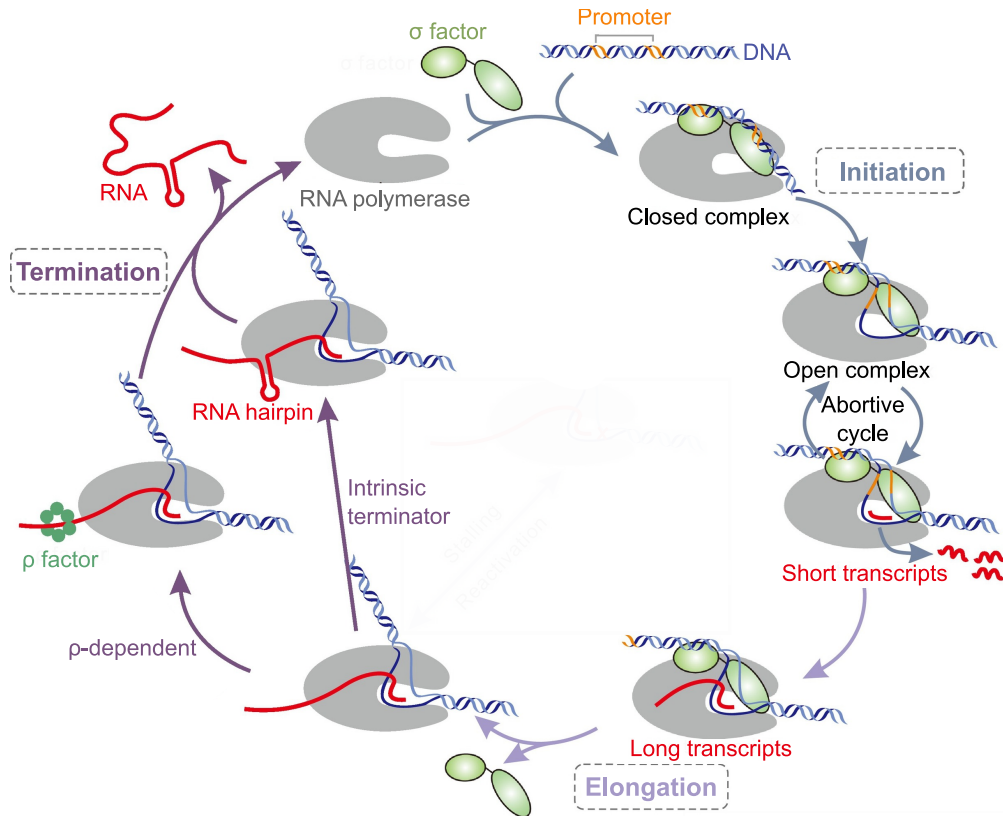


FIGURE 1.14 – Les étapes du cycle de la transcription bactérienne. Un cycle débute avec la reconnaissance du promoteur par l'ARN polymérase (ARNP) grâce au facteur  $\sigma$ . L'ARNP liée à de l'ADN double brin, au niveau d'un promoteur, forme le complexe fermé. Cette liaison provoque le changement conformationnel de l'ARNP, ce qui aboutit à la séparation des brins de l'ADN autour du site de début de transcription. Le complexe est alors ouvert. La synthèse de l'ARN en copie complémentaire de la matrice ADN débute mais les premiers ARN synthétisés ne dépassent pas les 10 nt (transcription abortive). Après quelques cycles abortifs, la synthèse des ARN complets devient possible. L'ARNP quitte le promoteur, se dissocie du facteur  $\sigma$  et la phase d'élongation débute, les ARN sont synthétisés. La phase d'élongation se termine lorsqu'un signal de terminaison de la transcription est présent sur l'ARN synthétisé. Deux types de terminaisons sont décrits chez les bactéries : la terminaison intrinsèque et la terminaison  $\rho$  dépendante. A l'issue de ce cycle, l'ARNP se dissocie de l'ADN et l'ARN est libéré. L'ARNP peut alors recommencer un cycle de transcription. Schéma adapté de [96].

### L'initiation de la transcription

L'initiation de la transcription peut être divisée en trois événements : la reconnaissance du promoteur, l'isomérisation et l'initiation de la synthèse de l'ARN.

Tout d'abord, l'ARNP se lie au facteur  $\sigma$  pour former l'holoenzyme. Ce facteur  $\sigma$  reconnaît les séquences promotrices et guide l'ARNP jusqu'à ces séquences. L'ARNP se lie alors à l'ADN double brin et forme ainsi un complexe fermé [97, 88].

La liaison de l'ARNP à l'ADN provoque un changement conformationnel de l'ARNP. La pince formée par les deux sous-unités  $\beta$  et  $\beta'$  se ferme et forme un canal, contenant le site actif de l'ARNP, autour de l'ADN et plus particulièrement autour du TSS. Ce changement conformationnel permet ensuite la séparation des deux brins de l'ADN entre les positions -11 et +2 ou +4 par l'ARNP : le complexe fermé devient ouvert [97, 11]. L'ouverture de cette région de l'ADN pour former une bulle de transcription est une étape critique car cette ouverture demande beaucoup d'énergie. La quantité d'énergie nécessaire dépend notamment du contenu en GC du discriminateur situé entre l'élément -10 et le +1 [98].

La séparation des brins a rendu les bases de l'ADN accessibles. L'ARNP peut donc débiter la synthèse de la copie complémentaire de l'ADN : l'ARN. Les ribonucléosides triphosphates (rNTPs) entrent dans le complexe et seuls les rNTPs complémentaires de la séquence matrice sont retenus. L'ARNP catalyse l'attachement du 5' phosphate d'un ribonucléotide avec le 3' hydroxyl du précédent, formant ainsi la chaîne d'ARN. Le début de cette synthèse est critique. En effet, lorsque 10 ribonucléotides environ ont été assemblés, l'un des domaines du facteur  $\sigma$  bloque le canal de sortie des ARN et provoque ainsi l'arrêt de la transcription et la libération d'un ARN de 10 nt environ. Cette initiation abortive se reproduit jusqu'à ce que ce domaine bloquant soit poussé hors du canal par un ARN synthétisé, provoquant généralement la séparation du facteur  $\sigma$  du reste de l'ARNP et le déplacement de l'ARNP au delà du promoteur [11, 88].

### L'élongation

Une fois que l'ARNP a quitté le promoteur, le complexe est stabilisé, ce qui permet à l'ARNP de poursuivre la synthèse de l'ARN, en copie complémentaire de l'ADN. Au sein de la bulle de transcription, l'ARN en cours d'élongation s'hybride sur environ 9 nt avec l'ADN matrice avant que l'ARN sorte du canal et que l'ADN se referme. L'ARNP peut transcrire de longs segments d'ADN (> 10kb), en synthétisant 30 à 100 nt par seconde mais sa vitesse n'est pas constante [97, 11]. En effet, elle peut être interrompue par divers obstacles imposés par la séquence ou la topologie de l'ADN matrice, par des lésions dans l'ADN matrice, par des protéines de liaison à l'ADN, par les machineries de réplication ou de réparation de l'ADN, par les ribosomes, par des facteurs de transcription ou encore des petites molécules comme le guanosine tetraphosphate (ppGpp) [97]. Fréquemment, l'ARNP est également interrompue à cause d'une structure secondaire tige-boucle formée dans un ARN synthétisé contenant une séquence inversée répétée. Ces pauses réduisent la

vitesse globale de la transcription mais peuvent faciliter le repliement des ARN synthétisés et des protéines en cours de traduction le long d'un ARN en cours de synthèse [11].

### La terminaison

L'élongation se termine à la rencontre d'un signal de terminaison. Deux mécanismes ont été décrits chez les bactéries : un mécanisme intrinsèque et un mécanisme dépendant d'un facteur de terminaison.

Un terminateur intrinsèque contient typiquement une séquence inversée répétée, riche en G et C, à environ 7-8 nt de l'extrémité 3' du transcrit, suivie d'une chaîne de 7 à 9 uracile. La synthèse de cette chaîne de U provoque des pauses de l'ARNP, laissant ainsi le temps à la séquence riche en G-C de se replier en structure tige-boucle [97, 99]. Cette structure déstabilise l'hybride ADN-ARN et semble conduire à la dissociation du complexe de transcription [11].

Le facteur de terminaison retrouvé dans la plupart des bactéries est le facteur Rho ( $\rho$ ). La protéine hélicase Rho forme un anneau composé de 6 sous-unités codées par le gène *rho*. D'après le modèle standard, cet anneau se lie à un site *rut* sur l'ARN, puis l'ARN passe à travers l'anneau. Rho se déplace ensuite du 5' vers le 3' et rattrape l'ARNP lorsque celle-ci fait des pauses sur des sites de terminaison Rho-dépendante. Avec son activité hélicase, Rho dissocie alors l'hybride ARN-ADN, ce qui met fin à la transcription et conduit à la dissociation de l'ARNP de son ADN matrice et de l'ARN synthétisé [99, 11]. Certains détails de ce mécanisme sont cependant remis en questions et restent à élucider [100]. Il semblerait notamment que Rho, à l'aide des co-facteurs NusA et NusG, se lie à l'ARNP au cours de l'élongation, formant un complexe de pré-terminaison. Ce complexe permettrait à Rho de rechercher en continu un signal de terminaison sur les ARN synthétisés [100].

### 1.2.3 Les modèles standards de régulation de l'initiation de la transcription

La transcription est un processus essentiel qui requière beaucoup de ressources et d'énergie. Sa régulation selon les conditions rencontrées par la bactérie est essentielle à son adaptation et à sa survie. En effet, cette régulation permet à la fois de transcrire en grande quantité les gènes utiles dans un environnement, tout en diminuant l'expression des gènes qui conduiraient à une accumulation de protéines inutiles. Cette régulation est essentielle du fait de la quantité limitée d'ARNP et d'énergie que peut y consacrer la cellule [92]. Cette régulation se fait en grande partie dès l'initiation du processus de transcription, ce qui permet de limiter les dépenses inutiles d'énergies et de ressources. De nombreux mécanismes sont impliqués dans cette régulation, en affectant l'affinité de l'ARNP pour un

promoteur ou l'activité de l'ARNP directement. Les plus étudiés sont les facteurs de transcription et les facteurs  $\sigma$  alternatifs mais d'autres mécanismes complémentaires affectent la transcription, notamment la fixation de petits ligands, le surenroulement de l'ADN (SC, Chapitre 1 partie 1.4) ou des modifications de certaines bases de l'ADN. Ce dernier mécanisme implique généralement la méthylation d'une base, fréquemment l'adénine, ce qui provoque un changement d'affinité de l'ARNP ou des facteurs de transcriptions pour le segment d'ADN modifié.

### Les facteurs de transcription

La régulation de la transcription se fait en grande partie par des protéines appelées facteurs de transcription, qui se lient de manière spécifique à l'ADN, souvent avec un motif hélice-coude-hélice ou, plus rarement, avec un motif ruban-hélice-hélice. Ces facteurs peuvent réguler négativement (répresseur) ou positivement (activateur) la transcription. Une même protéine peut être répresseur ou activateur selon les conditions et le site où elle se lie [92]. Le génome d'*E. coli* contient plus de 300 gènes qui semblent coder pour des facteurs de transcription. Certains contrôlent l'expression d'un grand nombre de gènes mais environ 1/5 d'entre eux ne contrôle qu'un promoteur unique [101, 102].

Il existe trois mécanismes généraux pour l'activation des promoteurs. Les activateurs de la classe I se lient en amont de l'élément -35 du promoteur et interagissent avec les extrémités C terminales des sous-unités  $\alpha$  de l'ARNP pour la diriger vers ce promoteur (Fig. 1.15A). CRP (cAMP receptor protein) est un exemple d'activateur de cette classe chez *E. coli*. En réponse à une concentration élevée de AMPc, le complexe AMPc-CRP se forme et active la transcription d'un grand nombre de promoteurs, y compris le promoteur *lac*, en se liant à une séquence consensus de l'ADN en amont du promoteur et en interagissant avec l'ARNP [104]. Les activateurs de la classe II se lient à une séquence chevauchant l'élément -35 et interagissent directement avec le facteur  $\sigma$  (Fig. 1.15B) afin de recruter l'ARNP associée [103]. Dans le troisième mécanisme, les activateurs se lient au niveau des éléments du promoteur, généralement les éléments -10 et -35 et modifient la conformation du promoteur afin de réorienter ces éléments pour améliorer la reconnaissance de ces promoteurs par le facteur  $\sigma$  (Fig. 1.15C) [92, 103]. Les activateurs de la famille de MerR appartiennent à cette troisième catégorie. Identifiés chez différentes organismes dont *B. subtilis*, *E. coli*, *Pseudomonas aeruginosa* et *Streptomyces lividans*, ces activateurs se lient entre les éléments -10 et -35 et provoquent la distorsion de l'ADN dans cette région. Cette distorsion permet aux éléments -10 et -35 d'être alignés de telle sorte que l'ARNP puisse s'y fixer [103, 105].

Trois principaux mécanismes de répression ont été décrits. Le premier est l'encombrement stérique, provoqué par la liaison d'un répresseur au niveau d'éléments clés du promoteur, qui empêche la liaison de l'ARNP à ce promoteur (Fig. 1.15D). Ce mécanisme est impliqué

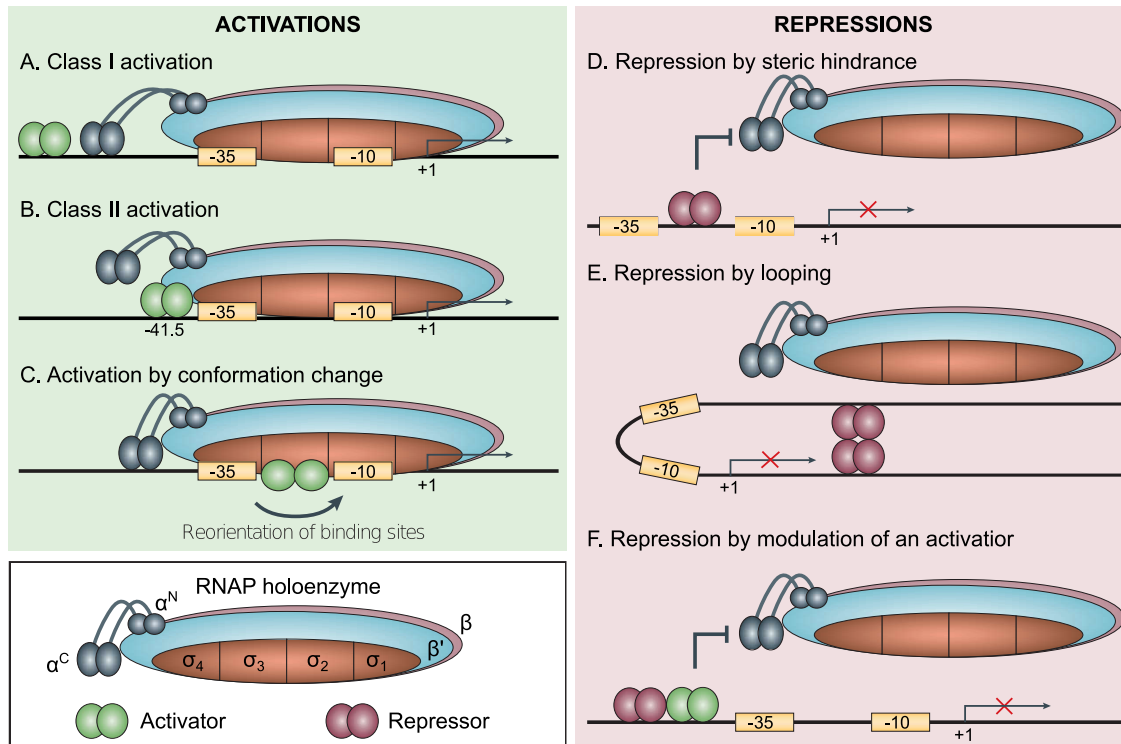


FIGURE 1.15 – Les mécanismes des facteurs de transcription, à gauche les activateurs (sphères vertes) et à droite les répresseurs (sphères rouges). Les principaux mécanismes des activateurs sont la liaison (A) en amont ou (B) au niveau du promoteur pour y guider l'ARNP, et (C) la réorientation des éléments du promoteur pour permettre une meilleure reconnaissance par l'ARNP. Les répresseurs peuvent agir (D) en se liant au promoteur pour générer un encombrement stérique qui bloque l'ARNP, (E) en se liant en multimère sur deux segments d'ADN autour du promoteur afin de courber cette région et la rendre inaccessible pour l'ARNP, ou (F) en bloquant l'activité d'un activateur. Schéma adapté de [103].

dans la régulation très étudiée du promoteur *lac*. En effet, en absence de lactose, le répresseur LacI se lie spécifiquement au site opérateur de l'opéron *lac* et empêche ainsi la fixation de l'ARNP et par conséquent l'expression de cet opéron [106]. Le second mécanisme implique les fixations de plusieurs molécules de répresseur sur des sites distants. Ces fixations peuvent courber l'ADN, provoquer la formation de boucles, et ainsi empêcher la transcription des promoteurs situés au sein de ces boucles (Fig. 1.15E) [103]. Dans le troisième mécanisme, le répresseur joue un rôle d'anti-activateur en interagissant avec un activateur pour l'empêcher d'activer la transcription (Fig. 1.15F). Ce troisième mécanisme peut être illustré avec le répresseur CytR. En absence de cytidine, CytR interagit avec CRP, or CRP ne peut interagir simultanément avec CytR et l'ARNP. Par conséquent, CytR est un anti-activateur de CRP [107]. D'autres mécanismes de répression, n'impliquant pas la fixation de l'ARNP au promoteur mais l'initiation du déplacement de l'ARNP sur l'ADN ont été décrits. Un facteur de transcription peut par exemple se lier en aval de l'ARNP, ou interagir très fortement avec l'ARNP et l'ADN, et ainsi empêcher l'ARNP de progresser le long de l'ADN.

Similairement aux facteurs de transcription, les NAPs peuvent également participer à la régulation de la transcription en induisant des modifications topologiques au niveau des promoteurs. En formant des ponts entre des fragments d'ADN distants, H-NS peut par exemple piéger des ARNP ou des séquences promotrices au sein de boucles, réprimant ainsi l'expression des gènes associés [108].

### Les facteurs $\sigma$ alternatifs

Pour pouvoir initier la transcription d'un gène, l'ARNP doit s'associer avec une sous-unité  $\sigma$  qui reconnaît une séquence promotrice, y guide l'ARNP puis facilite l'ouverture de la double hélice de l'ADN proche du TSS [103]. Différents facteurs sont en compétition pour l'association avec les ARNP (voir paragraphe 1.2.3), disponibles en quantité limitée [92]. En effet, bien qu'il y ait de l'ordre de 1000 à 10000 ARNP selon la phase de croissance dans chaque cellule [89], le nombre d'ARNP libres, c'est à dire n'étant ni impliquées dans un processus de transcription, ni liées à l'ADN de manière non-spécifique, est plus limité. D'après un modèle physique le nombre d'ARNP libre par cellule a été estimé à environ 100 en phase lente de croissance et à 1000 en phase rapide de croissance [109].

Cette compétition entre les facteurs  $\sigma$  est importante dans la régulation de la transcription car ces facteurs permettent de provoquer des changements globaux de l'expression des gènes, généralement en réponse à des stress. Il semble d'ailleurs y avoir une corrélation entre le nombre de facteurs  $\sigma$  alternatifs et la diversité des environnements rencontrés par les bactéries [103]. Chez *E. coli*, le principal facteur  $\sigma^{70}$  permet la reconnaissance et la transcription de la majorité des promoteurs mais, selon les conditions rencontrées, des facteurs  $\sigma$  alternatifs sont présents en grande quantité (Tableau 1.3).

TABLEAU 1.3 – Les facteurs  $\sigma$  de *E. coli*. Tableau adapté de [110]. Les facteurs  $\sigma$  sont généralement nommés selon leur taille, le facteur  $\sigma^{70}$  a par exemple une masse moléculaire de 70 kDa [11].

Facteur $\sigma$	Fonction
$\sigma^{70}$ / $\sigma^D$	Transcription de la majorité des gènes en phase exponentielle
$\sigma^{54}$ / $\sigma^N$	Transcription des gènes régulés par l'azote
$\sigma^{38}$ / $\sigma^S$	Expression des gènes en phase stationnaire
$\sigma^{32}$ / $\sigma^H$	Transcription des gènes de réponse aux chocs thermiques
$\sigma^{28}$ / $\sigma^F$	Expression des gènes de motilité
$\sigma^{24}$ / $\sigma^E$	Réponse aux stress extracytoplasmiques et aux températures extrêmes
$\sigma^{19}$ / $\sigma^{fecI}$	Régulation des gènes <i>fec</i> pour le transport du citrate de fer

C'est le cas du facteur  $\sigma^H$ , qui, après un choc thermique (températures élevées), permet la transcription des gènes codant pour les protéines essentielles à la réponse à ce choc. La

quantité de facteur  $\sigma^S$  varie également fortement selon les conditions de croissances et de stress. Ce facteur impacte plus de 20% des gènes d'*E. coli* et joue un rôle particulièrement important en phase stationnaire [92].

### Les petits ligands : exemple de ppGpp

L'interaction de l'ARNP avec des petits ligands permet également de moduler la transcription des gènes selon les conditions environnementales. C'est par exemple le cas de ppGpp (guanosine tetraphosphate), une alarmone présente chez un grand nombre de bactéries, qui semble pouvoir affecter positivement et négativement la transcription, en se liant proche du site actif de l'ARNP et en interagissant avec les sous-unités  $\beta$  et  $\beta'$  de l'ARNP. Synthétisée par une enzyme appelée RelA, la concentration de ppGpp augmente en réponse à des stress métaboliques, quand la disponibilité des acides aminés est réduite, notamment lorsque la croissance bactérienne est ralentie, après un stress ou en phase stationnaire [92, 103]. ppGpp est notamment capable de déstabiliser les complexes ouverts formés au niveau des promoteurs afin d'inhiber la synthèse des ARNr et des ARNt [11].

## 1.2.4 Les techniques d'analyse du transcriptome

Le transcriptome est l'ensemble des transcrits (ARN) présents dans une cellule à un instant donné. Selon les conditions spécifiques dans lesquelles les cellules se trouvent, telles que la phase de croissance et les variations de conditions environnementales, le transcriptome subit des changements, qu'il est intéressant d'étudier. Parmi les méthodes transcriptomiques couramment utilisées, le RNA-Seq est largement prédominant actuellement et a remplacé en grande partie les puces à ADN.

### Les puces à ADN

Les puces à ADN sont utilisées depuis la fin des années 90 pour analyser des séquences d'ADN et l'expression de gènes [111, 112]. Le principe des puces à ADN repose sur l'hybridation des brins d'ADN complémentaires.

La dernière génération des puces à ADN sont des surfaces solides sur lesquelles des sondes contenant des séquences d'ADN spécifiques sont positionnées. Pour analyser un échantillon d'ARN, celui-ci est rétro-transcrit en ADN complémentaire puis marqué, généralement par fluorescence. L'échantillon à analyser est ensuite incubé au contact de la puce (Fig. 1.16). Les fragments complémentaires des sondes s'hybrident de telle sorte que seuls ces fragments sont retenus lorsque la puce est lavée. Le marquage des fragments est ensuite révélé afin de connaître la quantité de fragments complémentaires de chaque

sonde. Cette quantité reflète l'expression des gènes correspondant aux séquences présentes sur les sondes [113].

Les puces à 2 canaux permettent de comparer l'expression des gènes dans deux conditions différentes en marquant différemment chaque échantillon, généralement avec les colorants fluorescents Cy3 et Cy5 (Fig. 1.16). Ces puces permettent d'étudier simultanément l'expression différentielle d'un grand nombre de gènes entre deux conditions. L'utilisation des puces à ADN nécessite de connaître la séquence de l'ensemble du génome pour créer les sondes. De plus, le signal de fond et le manque de spécificité de certaines sondes en font une méthode d'assez basses spécificité et sensibilité en comparaison aux nouvelles technologies disponibles aujourd'hui comme le RNA-Seq [114].

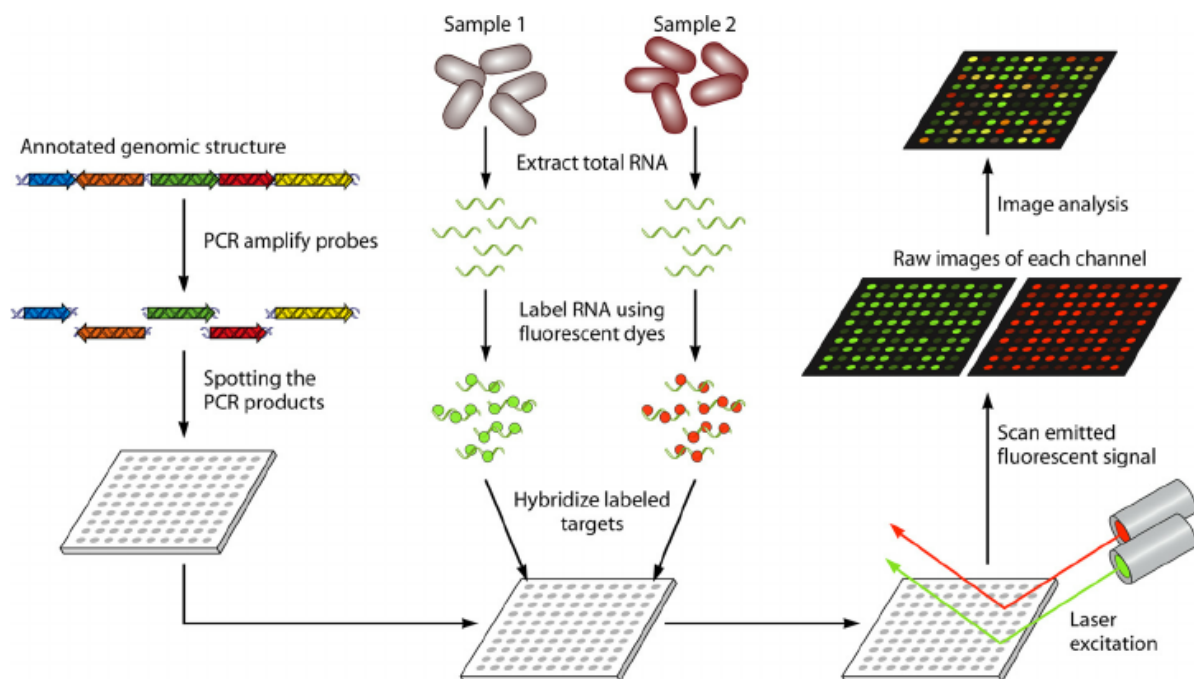


FIGURE 1.16 – Protocole expérimental d'analyse transcriptomique avec une puce à ADN. La puce à ADN est préparée en ajoutant des sondes d'ADN spécifiques d'une séquence génomique sur une surface solide. Pour comparer l'expression des gènes dans deux conditions, les ARN de chaque échantillon sont extraits puis marqués par fluorescence avec des marqueurs distincts. Les ARN marqués sont ensuite ajoutés sur la puce afin que les ARN complémentaires de chaque sonde s'y hybrident. La quantité relative d'ARN hybridés entre les deux conditions sur chaque sonde peut ensuite être estimée par fluorescence sous excitation d'un laser. Cette quantité relative reflète l'expression différentielle, entre les deux conditions testées, du gène ciblé par chaque sonde. Schéma tiré de [115].

## Le RNA-Seq

Le développement des technologies de séquençage nouvelle génération (NGS) a marqué un tournant pour les études transcriptomiques car elles permettent d'obtenir des données sur l'ensemble des transcrits en très peu de temps et avec une très bonne sensibilité.



Le RNA-Seq est aujourd'hui la méthode de choix pour l'étude des transcriptomes mais également pour la recherche de nouvelles séquences ou de nouveaux gènes [114].

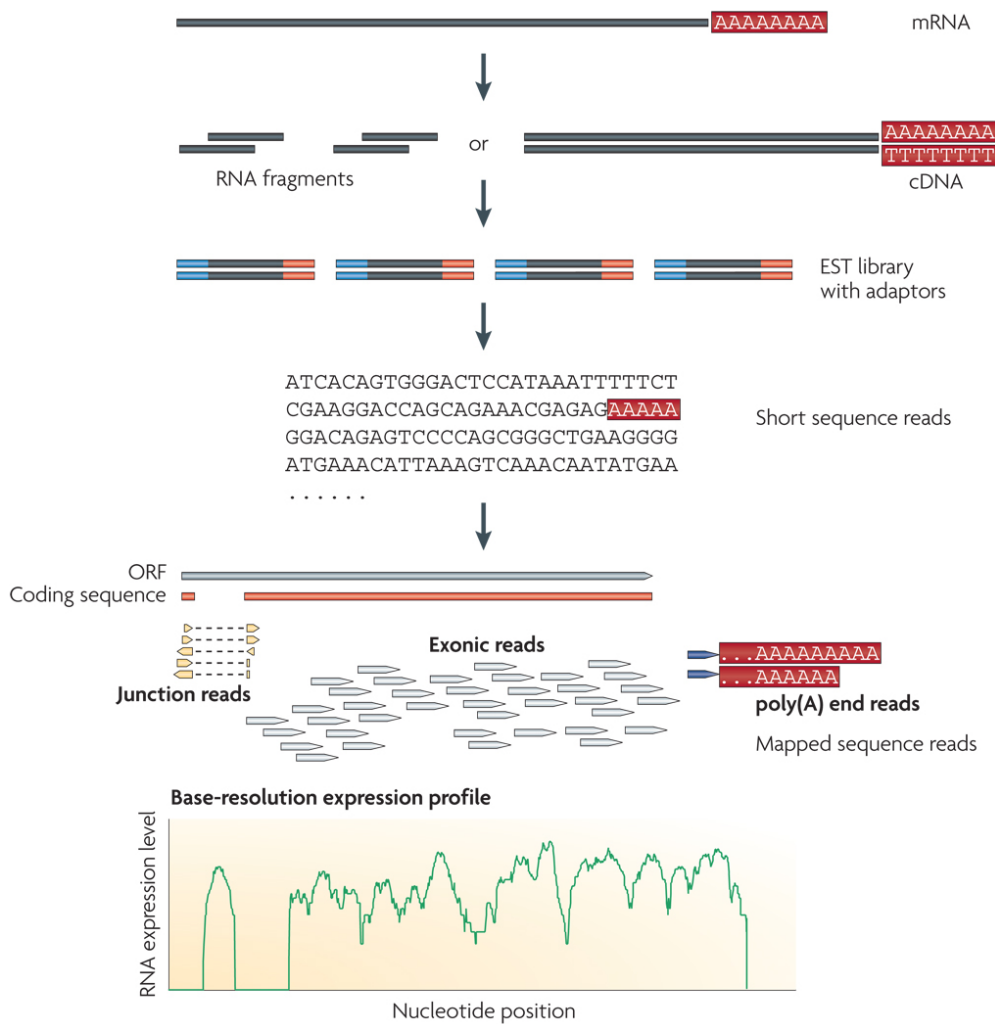


FIGURE 1.17 – Protocole d'analyse transcriptomique par RNA-Seq. Les ARN sont convertis en une banque de fragments d'ARN ou de cDNA puis des adaptateurs (bleu et orange) sont ajoutés à leurs extrémités. Un appareil de séquençage haut-débit est ensuite utilisé pour obtenir les séquences de chaque fragment. Les séquences (lectures) ainsi obtenues sont ensuite alignées sur le génome de référence. Dans cet exemple de RNA-Seq appliqué à un eucaryote, une lecture peut correspondre à un unique exon (exonic reads), à plusieurs exons successifs (junction reads) ou à la queue poly-A présente à l'extrémité de chaque ARNm (poly(A) end reads, en rouge). Schéma copié de [114].

Pour analyser un transcriptome par RNA-Seq (Fig. 1.17), les ARN doivent tout d'abord être rétrotranscrits en ADNc puis fragmentés. Selon la méthode utilisée, des adaptateurs sont ensuite ajoutés à l'une ou aux deux extrémités de chaque fragment avant de soumettre ces fragments à un séquençage haut-débit. Les séquences fournies par le séquenceur sont appelées lectures. Selon l'appareil utilisé, ces lectures peuvent être plus ou moins longues, de 50 pb à plusieurs dizaines de kb, sachant que le séquençage de longs fragments facilite ensuite l'assemblage des lectures mais peut présenter plus d'erreurs que le séquençage de

courts fragments [114]. Dans le cadre des études transcriptomiques classiques, les lectures sont alignées sur un génome de référence puis le nombre de lectures correspondant à chaque gène est dénombré pour chaque condition.

Pour pouvoir comparer les nombres de lectures entre les différents gènes et entre les différentes conditions, des normalisations sont effectuées afin de notamment prendre en compte la taille des gènes et la profondeur du séquençage. De nombreux outils informatiques ont été développés pour automatiser ces étapes d'analyse. Dans le cadre des études transcriptomiques bactériennes que j'ai effectuées pendant ma thèse, j'ai utilisé les outils standards *Bowtie 2* [116], *htseq-count* [117] et *DeSeq2* [118]. *Bowtie 2* effectue l'alignement des lectures sur le génome en identifiant la région du génome ayant la séquence la plus similaire à celle de la lecture [116]. *htseq-count* permet ensuite de dénombrer le nombre de lectures associées à chaque gène [117]. Enfin *DeSeq2* réalise la normalisation et la comparaison des nombres de lectures. *DeSeq2* génère un fichier contenant notamment, pour chaque gène, un ratio d'expression normalisé (*Fold change*, FC) et la valeur p du test de Wald ainsi que sa correction effectuée d'après la méthode de Benjamin-Hochberg (valeur p ajustée) [118]. Dans mes études transcriptomiques, j'ai considéré que les gènes ayant une valeur p ajustée inférieure à 0,05 sont différentiellement exprimés. Ceux ayant un FC supérieur à 1 (c.-à-d. un  $\log_2(FC)$  négatif) sont considérés activés tandis que ceux ayant un FC supérieur à 0 (c.-à-d. un  $\log_2(FC)$  positif) sont réprimés.

Une fois cette classification des gènes effectuée, différentes analyses peuvent être réalisées selon le cadre de l'étude. L'une des plus courantes est l'analyse de l'enrichissement fonctionnel. Elle consiste à identifier des fonctions sur-représentées au sein des gènes activés ou des gènes réprimés. Pour ce faire, une annotation fonctionnelle de l'ensemble des gènes est nécessaire. Dans mes études transcriptomiques, j'ai utilisé l'annotation Gene Ontology [119], qui associe chaque gène à un ou plusieurs processus biologique(s) (par exemple motilité et motilité dépendante des flagelles), une ou plusieurs fonction(s) moléculaire(s) (par exemple activité catalytique et activité motrice du cytosquelette) et une ou plusieurs localisation(s) (par exemple membrane et flagelle). D'autres annotations peuvent être utilisées pour effectuer d'autres analyses, notamment le brin sur lequel le gène est situé, la proximité à un site de fixation d'une enzyme particulière, la localisation au sein de domaines chromosomiques préalablement identifiés, etc.

### 1.3 Les variations du niveau de surenroulement de l'ADN

Des processus biologiques essentiels tels que la réplication et la transcription nécessitent l'ouverture de la double hélice de l'ADN. La séparation des deux brins de l'ADN provoque un stress torsionnel, qui peut diffuser le long du chromosome jusqu'à la rencontre d'une barrière topologique, par exemple une protéine fixée sur l'ADN. L'accumulation de ces stress torsionnels constitue le surenroulement de l'ADN (SC). L'ADN bactérien est généralement surenroulé négativement, à une densité  $\sigma = -0.06$  en phase exponentielle chez *E. coli* [86]. Le niveau de SC est finement régulé, selon les conditions physiologiques (phases de croissances) et environnementales (stress environnementaux) rencontrés par les bactéries (Fig. 1.18).

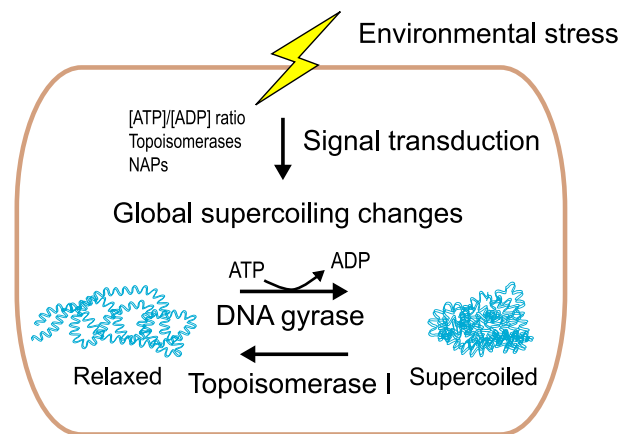


FIGURE 1.18 – La régulation du surenroulement de l'ADN en réponse aux stress environnementaux. Schéma issu de [83].

#### 1.3.1 La régulation du surenroulement par les topoisomérases

Les acteurs clés de cette régulation sont les topoisomérases. Elles sont présentes chez tous les organismes. Ces enzymes peuvent modifier le nombre de supertours en quelques étapes seulement : elles se lient à l'ADN, coupent l'un ou les deux brins, permettent le passage d'un brin d'ADN à travers la coupure puis religuent le ou les brins coupés (Fig. 1.22 et 1.20) [120, 11]. Selon le nombre de brins coupés, les topoisomérases sont classifiées en deux catégories : type I lorsqu'un unique brin est coupé, type II lorsque les deux brins sont coupés [86]. Les topoisomérases peuvent ensuite être divisées en sous-catégories (type IA, IB, IC, IIA et IIB) selon leurs propriétés structurelles et mécaniques [121] (Tableau 1.4).

##### Les topoisomérases de type I

Les topoisomérases de type I introduisent des cassures simple brin dans l'ADN. Chez les bactéries, deux topoisomérases de type I sont généralement présentes : la topoisomérase

TABLEAU 1.4 – Propriétés et activités *in vitro* des différentes topoisomérases. Tableau issu de [121].

Topoisomerase	Type	Bond formed	Proposed mechanism	ATP dependant	Mg(II) dependant	Catenation/ decatenation	Relaxation activity	Supercoiling activity
Bacterial Topo I	IA	5'	Strand passage	No	Yes	Yes <sup>a</sup>	Yes SC- <sup>b</sup>	No
Eukaryotic Topo I	IB	3'	Controlled rotation	No	Yes	Yes <sup>a</sup>	Yes	No
Topo II	IIA	5'	Strand passage	Yes	Yes	Yes	Yes	No
Topo III	IA	5'	Strand passage	No	Yes	Yes <sup>a</sup>	Yes SC- <sup>b</sup>	No
Topo IV	IIA	5'	Strand passage	Yes	Yes	Yes	Yes	No
Topo V	IC	3'	Controlled rotation	No	No	Unknown	Yes	No
Topo VI	IIB	5'	Strand passage	Yes	Yes	Yes	Yes	No
DNA gyrase	IIA	5'	Strand passage	Yes	Yes	Yes (weak)	Yes	SC- introduction
Reverse gyrase	IA	5'	Strand passage	Yes	Yes	No	Yes SC- <sup>b</sup>	SC+ introduction

<sup>a</sup> Possible only if one substrate is nicked or single stranded.<sup>b</sup> Relaxes only negatively supercoiled DNA

I (topoI) et la topoisomérase III (topoIII) [86]. Deux autres topoisomérases de type I, la topoisomérase V (topoV) et la reverse gyrase, ont également été identifiées, principalement chez les archées [121].

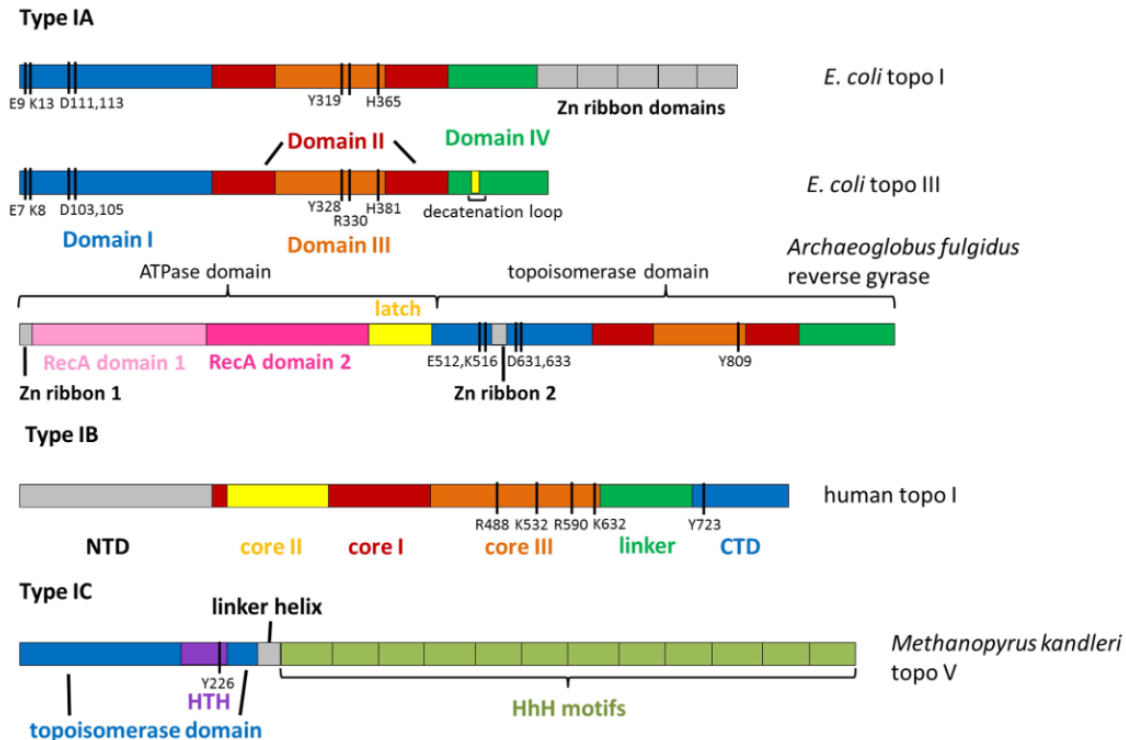


FIGURE 1.19 – Structures principales des topoisomérases de type I. Les barres verticales noires indiquent les résidus catalytiques dont la tyrosine (Y) catalytique, qui forme une liaison covalente avec le simple brin d’ADN clivé. Pour le type IB, NTD et CTD correspondent respectivement aux domaines N-terminal et C-terminal. Dans la structure du type IC, HTH indique le motif hélice-coude-hélice et HhH correspond au motif hélice-épingle-hélice. Schéma issu de [121].

La topoI, initialement nommée protéine  $\omega$ , est présente chez les procaryotes et les eucaryotes. Elle peut relâcher le surenroulement, et caténer et décaténer de l’ADN clivé. La topoI eucaryote, classifiée dans le type IB, peut relâcher le surenroulement négatif (SC-) et positif (SC+) tandis que la topoI bactérienne, de type IA, ne relâche que le SC- (Tableau 1.4) [121]. Chez *E. coli*, la topoI, codée par le gène *topA*, relâche l’ADN SC- à une vitesse d’environ 5 Lk/min, sans consommer d’énergie [122]. D’après le modèle *enzyme-bridging*, le mécanisme de la topoI, présenté en figure 1.20, est le suivant : la topoI clive un simple brin d’ADN et crée ainsi une ouverture permettant le passage du simple brin intact. Après le passage, le simple brin clivé est reliqué et la molécule a ainsi été relâchée d’un supertour négatif (Lk augmenté de 1). Pendant ce processus, la topoI adopte deux conformations : une conformation fermée dans laquelle la topoI reconnaît le brin d’ADN, puis une conformation ouverte qui lui permet de créer un pont en coupant l’un des brins et enfin de nouveau la conformation fermée afin de reliquer le brin coupé [121, 123, 124].

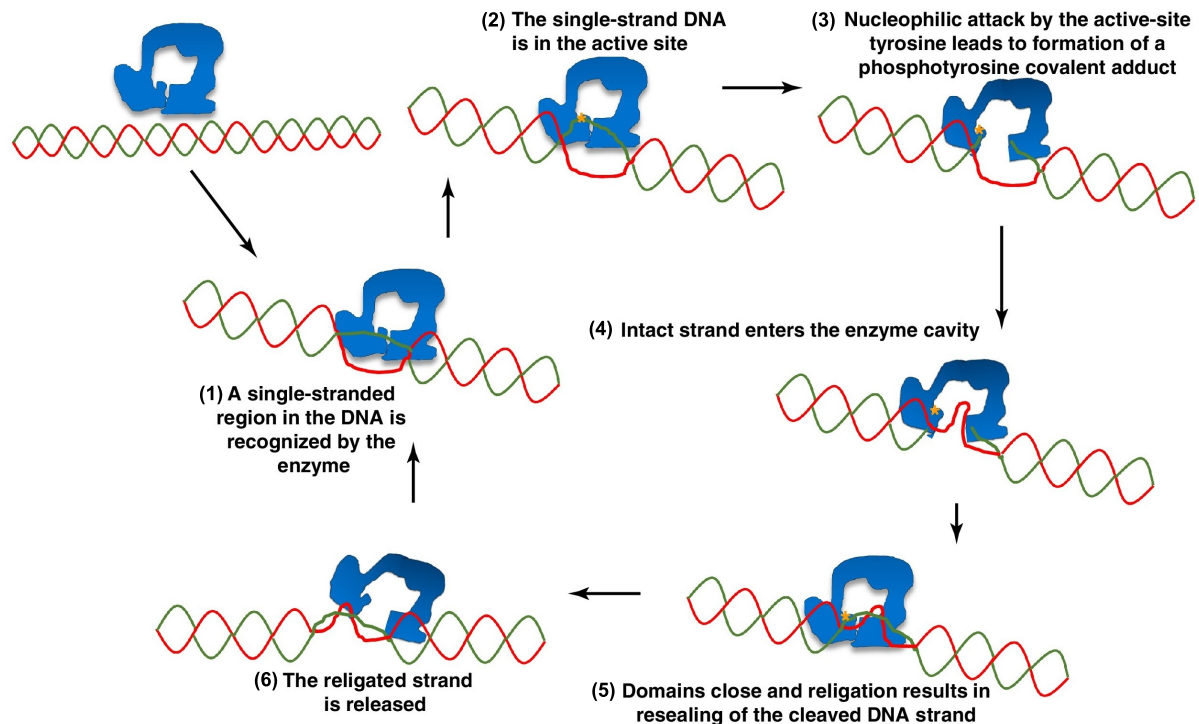


FIGURE 1.20 – Le mécanisme de relâchement de l'ADN par la topoisomérase I. (1) Une région simple brin d'une molécule d'ADN surenroulée négativement est reconnue par la topoI. (2) La topoI s'ouvre partiellement pour permettre à l'ADN de se lier proche du site actif. (3) La topoI clive le brin, via une attaque nucléophile du squelette phosphodiester de l'ADN par la tyrosine du site actif de la topoI. La topoI adopte alors une conformation ouverte, tout en formant un complexe covalent ADN - topoI, pour former un pont entre les 2 extrémités de l'ADN coupé. (4) Cette ouverture permet le passage du second brin, ce qui modifie le Lk. (5) La topoI se referme pour religuer les deux brins. (6) La topoI adopte de nouveau une conformation ouverte pour libérer le double brin relâché. Schéma issu de [123].

La topoIII appartient au type IA et peut être trouvée chez les procaryotes, eucaryotes et les archées. De structure très proche de la topoI bactérienne (Fig. 1.19), la topoIII bactérienne peut également relâcher le SC<sup>-</sup> par incrément de 1 Lk et décaténer l'ADN [121]. En comparaison à la topoI d'*E. coli*, la structure de la topoIII d'*E. coli* présente une boucle supplémentaire, connue comme la boucle de décaténation, qui semble lui permettre de décaténer efficacement des caténares multiples [121, 125]. Ainsi, contrairement à la topoI dont la fonction première semble être le relâchement de l'ADN SC<sup>-</sup>, la topoIII est principalement impliquée dans la décaténation des chromosomes, en particulier après la réplication [86].

Identifiée chez les archées, la topoV a tout d'abord été classifiée dans le type IB du fait de ses similarités avec la topoI eucaryote, notamment sa capacité à relâcher le SC<sup>-</sup> et le SC<sup>+</sup> (Tableau 1.4). Cependant, sa structure étant très différente, y compris la position de son site actif tyrosine (Fig. 1.19), elle est dorénavant classifiée dans une nouvelle classe, le type IC [121]. La topoV relâche l'ADN en coupant l'un des deux brins de l'ADN afin que celui-ci

puisse pivoter autour de l'autre brin jusqu'à religation du brin clivé. Ce mécanisme de rotation permet de modifier le Lk par des incréments supérieurs à 1 [126].

La reverse gyrase, classifiée dans le type IA, a été identifiée chez quelques eubactéries et archées. Cette enzyme peut relâcher l'ADN SC- et introduire des supertours positifs dans l'ADN, en consommant de l'ATP [121, 127].

## Les topoisomères de type II

Les topoisomères de type II coupent les deux brins de l'ADN et modifient ainsi le Lk par incrément de 2. Deux topoisomères de type II sont généralement présentes chez les bactéries : la gyrase, et la topoisomérase IV (topoIV). La topoisomérase II eucaryote (topoII) et la topoisomérase VI archée (topoVI) appartiennent également à cette famille de topoisomères [121].

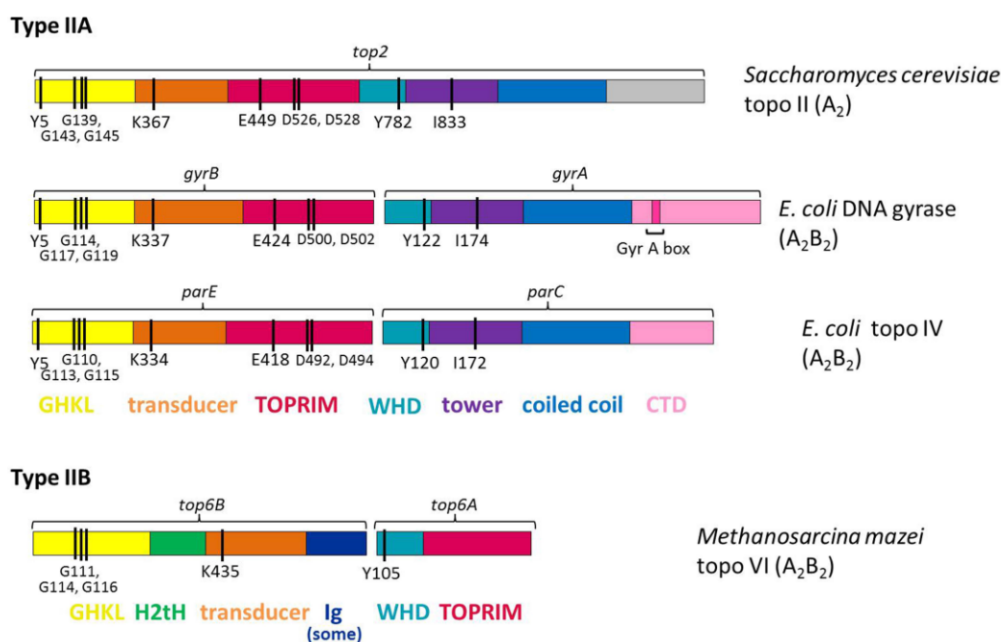


FIGURE 1.21 – Structures principales des topoisomères de type II. Les barres verticales noires indiquent les résidus catalytiques dont la tyrosine (Y) catalytique. GHKL correspond au domaine ATPase, TOPRIM au domaine topoisomérase/primase, WHD au domaine *winged-helix*, CTD au domaine C-terminal, H2tH au domaine hélice-hélice-coude-hélice et Ig à un repliement de type Immunoglobuline. Schéma issu de [121].

L'ADN gyrase est une topoisomérase de type IIA qui permet d'introduire des supertours négatifs dans une molécule d'ADN double brin mais nécessite pour cela l'hydrolyse d'ATP. La gyrase peut également, en présence d'ATP, relâcher l'ADN SC+ et décaténer l'ADN (Tableau 1.4) [121]. La gyrase est un tétramère constitué de deux sous-unités GyrA, codées par le gène *gyrA*, et de deux sous-unités GyrB, codées par *gyrB*. La sous-unité GyrA constitue la grande partie du site de liaison à l'ADN, elle permet le clivage de l'ADN et le

maintien des extrémités coupées pendant le passage d'un autre segment d'ADN double brin. La sous-unité GyrB, avec son site de liaison à l'ATP, est dédiée à la transduction d'énergie nécessaire au processus [128, 122]. Le mécanisme d'introduction des supertours négatifs dans l'ADN par la gyrase est présenté en Figure 1.22.

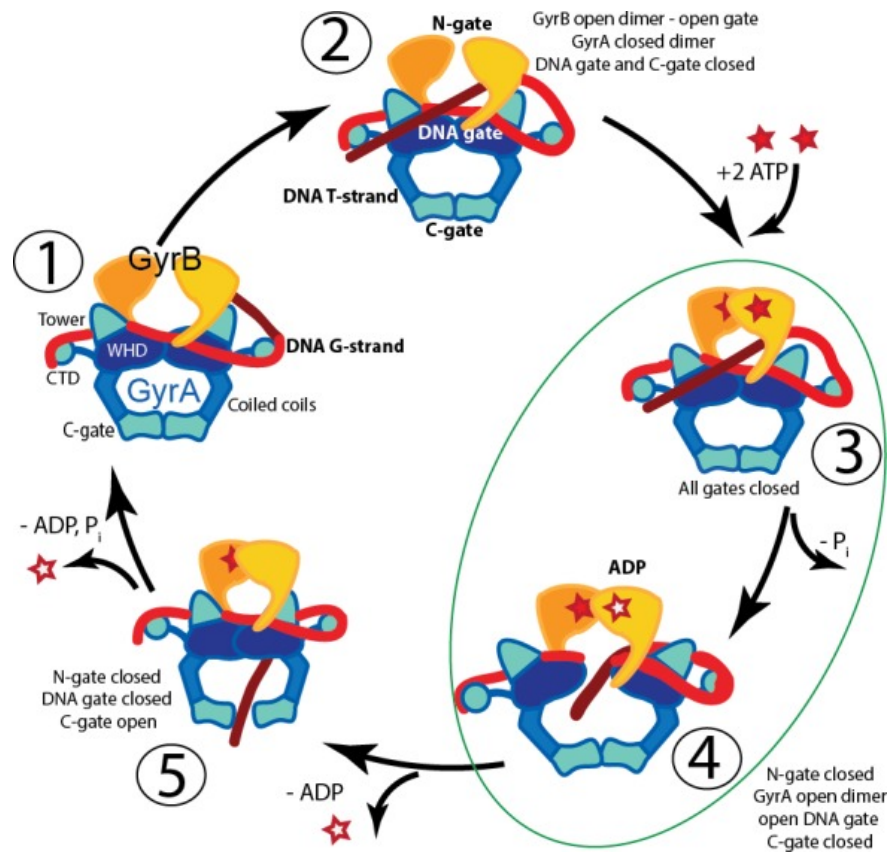


FIGURE 1.22 – Le mécanisme de surenroulement de l'ADN par la gyrase. (1) La gyrase se lie à un premier segment d'ADN double brin (en rouge) à l'interface entre les sous-unités GyrA (en bleu) et GyrB (en orange) et l'ADN double brin contourne le domaine C-terminal de GyrA. (2) Le segment du double brin ayant contourné ce domaine passe entre les deux sous-unités GyrB, qui sont alors écartées. (3) Avec la liaison de 2 ATP sur les sous-unités GyrB, ces deux sous-unités se rapprochent, piégeant ainsi les deux segments du double brin d'ADN. (4) Le premier segment d'ADN est clivé par le site actif de la gyrase, situé dans GyrA, ce qui permet au second segment d'ADN double-brin de passer à travers. Cette action consomme un ATP. (5) En consommant un second ATP, le double-brin est reliqué avec le rapprochement des sous-unités GyrA puis le brin est relâché [128].

D'après ce mécanisme, un segment d'ADN, nommé segment G, se lie et s'enroule autour du domaine N-terminal de l'une des deux sous-unités GyrA. La liaison de ce segment induit un mouvement du domaine C-terminal de la seconde sous-unité GyrA, ce qui provoque l'enroulement d'un segment d'ADN adjacent autour de ce domaine C-terminal. L'enroulement de ces deux segments d'ADN autour des domaines C-terminal et N-terminal permet de positionner un troisième segment, nommé segment T, entre les sous-unités GyrB, au croisement du segment G. Suite à ce positionnement, des molécules d'ATP se lient



aux domaines N-terminaux des deux sous-unités GyrB, ce qui provoque un changement conformationnel de l'enzyme : la "porte" formée par les deux sous-unités GyrB se ferme et piège le segment T dans le complexe. Les tyrosines catalytiques forment alors des liaisons avec le segment G, clivant ainsi le double brin, ce qui permet le passage du segment T. Le segment G est ensuite reliqué et le segment T est libéré [121, 128].

La topoIV est une enzyme de type IIA qui hydrolyse de l'ATP pour décaténer les chromosomes suite à la réplication et pour relâcher l'ADN SC- ou SC+ (Tableau 1.4) [86, 121, 129]. Chez de nombreuses bactéries dont *E. coli*, *Salmonella typhimurium* et *D. dadantii*, cette enzyme est un hétérotétramère composé de deux sous-unités codées par *parC* et de deux sous-unités codées par *parE* [130]. Les sous-unités ParC et ParE de la topoIV sont homologues aux sous-unités GyrA et GyrB mais présentent quelques différences structurales (Fig. 1.21) [121]. Chez *E. coli*, la topoIV relâche l'ADN à une vitesse de 0.8 Lk/min donc plus lentement que la topoI. Contrairement à la gyrase, elle ne permet pas d'introduire des supertours négatifs dans l'ADN. Elle est cependant 100 fois plus active dans la décaténation que la gyrase [122, 131].

La topoII appartient également au type IIA et est trouvée dans un grand nombre d'eucaryotes dont l'Homme, *S. cerevisiae* et *Drosophila melanogaster* [132, 133, 121]. Elle peut relâcher le SC+ et le SC- en présence d'ATP et de  $Mg^{2+}$ , caténer et décaténer l'ADN (Tableau 1.4). La topoII peut ainsi assurer la condensation, la maintenance de la structure et la ségrégation des chromosomes après la réplication [132, 134]. La plupart des eucaryotes supérieurs en possèdent deux isoformes (topoII $\alpha$  et topoII $\beta$ ) dont l'expression semble dépendre de l'étape du cycle cellulaire et du type de cellule [135, 136].

La topoVI semble être présente dans toutes les archées, dans certaines plantes ainsi que dans un parasite protozoaire, le *Plasmodium* [121, 137, 138, 139]. La topoVI peut décaténer l'ADN et relâcher du SC+ et du SC- [137]. Bien qu'elle comporte un domaine ATPase et un domaine topoisomérase/primase (Fig. 1.21), sa structure et sa séquence sont différentes de celles des autres topoisomérases de type IIA, d'où sa classification dans le type IIB.

Dans le cadre de mon étude sur la régulation de la transcription bactérienne par le surenroulement de l'ADN, les principales topoisomérases considérées sont la topoI, pour son activité de relâchement de l'ADN, et la gyrase, pour sa faculté à ajouter des supertours négatifs dans l'ADN. Bien que la topoIV relâche l'ADN moins efficacement que la topoI, il est important de ne pas négliger son rôle dans le processus de régulation du niveau de SC.

### 1.3.2 Les méthodes de mesure du niveau global de surenroulement de l'ADN

Les principales méthodes de mesure du niveau global de SC, i.e. du niveau de SC moyen sur l'ensemble d'une molécule d'ADN, repose sur les différences de migration de l'ADN, à travers un gradient de sucrose ou lors d'une électrophorèse en gel d'agarose. Ces différences de migrations dépendent de la structure 3D de l'ADN [140].

Les premières analyses du niveau de SC ont été effectuées par sédimentation sur un gradient de densité de sucrose. Pour cette analyse, les cellules étaient traitées radioactivement afin de marquer leur chromosome puis ces cellules étaient lysées et centrifugées afin d'en extraire les chromosomes. Les molécules d'ADN extraites étaient ensuite séparées, selon leur niveau de SC, par sédimentation sur un gradient de sucrose contenant du bromure d'éthidium (EtBr). Différentes fractions de ce gradient étaient ensuite collectées puis leur radioactivité était mesurée afin de déterminer la quantité d'ADN présente dans chaque fraction [141, 142]. Cette méthode a depuis été remplacée par des méthodes de mesure, plus simples, par électrophorèse sur gel d'agarose.

Dans un gel d'agarose, l'ADN migre également selon son niveau de SC. Une molécule d'ADN surenroulée migre plus rapidement donc plus loin sur le gel qu'une molécule d'ADN relâchée. Pour faciliter cette étude, un plasmide est électroporé dans la bactérie dont on souhaite analyser le niveau de surenroulement de son chromosome, en considérant que le niveau de surenroulement du plasmide reflète le niveau de surenroulement global du chromosome. Ce plasmide est ensuite extrait puis déposé dans le gel d'agarose. Pour améliorer la résolution des topoisomères de même poids moléculaire, cette électrophorèse est souvent réalisée en présence d'un agent intercalant. L'agent intercalant utilisé dans les premières études était l'EtBr, les études actuelles privilégient la chloroquine [143, 144, 140] (Fig. 1.23B). L'EtBr s'intercale dans l'ADN à chaque supertour et affecte ainsi le poids et la taille de la molécule, et par conséquent sa migration dans le gel. Sur ce type de gel, deux bandes successives correspondent à des topoisomères dont le Lk diffère de 1, c'est à dire d'un supertour [143]. L'intercalation d'EtBr s'est cependant révélée être trop sensible aux variations mineures des conditions expérimentales dont la quantité d'ADN déposée dans chaque puit. Cet intercalant a donc été substitué par la chloroquine, un ligand de plus faible affinité pour l'ADN [144]. Similairement à l'EtBr, la chloroquine permet la séparation distincte des topoisomères selon leur niveau de surenroulement. Cette méthode, avec une concentration constante de chloroquine, ne permet cependant pas la distinction des molécules surenroulées positivement et des molécules surenroulées négativement. Pour cette distinction, la méthode la plus fiable est l'électrophorèse 2D. Cette méthode repose sur deux migrations, une première, généralement dans un simple gel d'agarose, pour séparer les topoisomères selon leur densité puis une seconde migration perpendiculairement, en

présence d'une concentration de chloroquine permettant la séparation des topoisomères SC+ et des topoisomères SC- (Fig. 1.23C) [145].

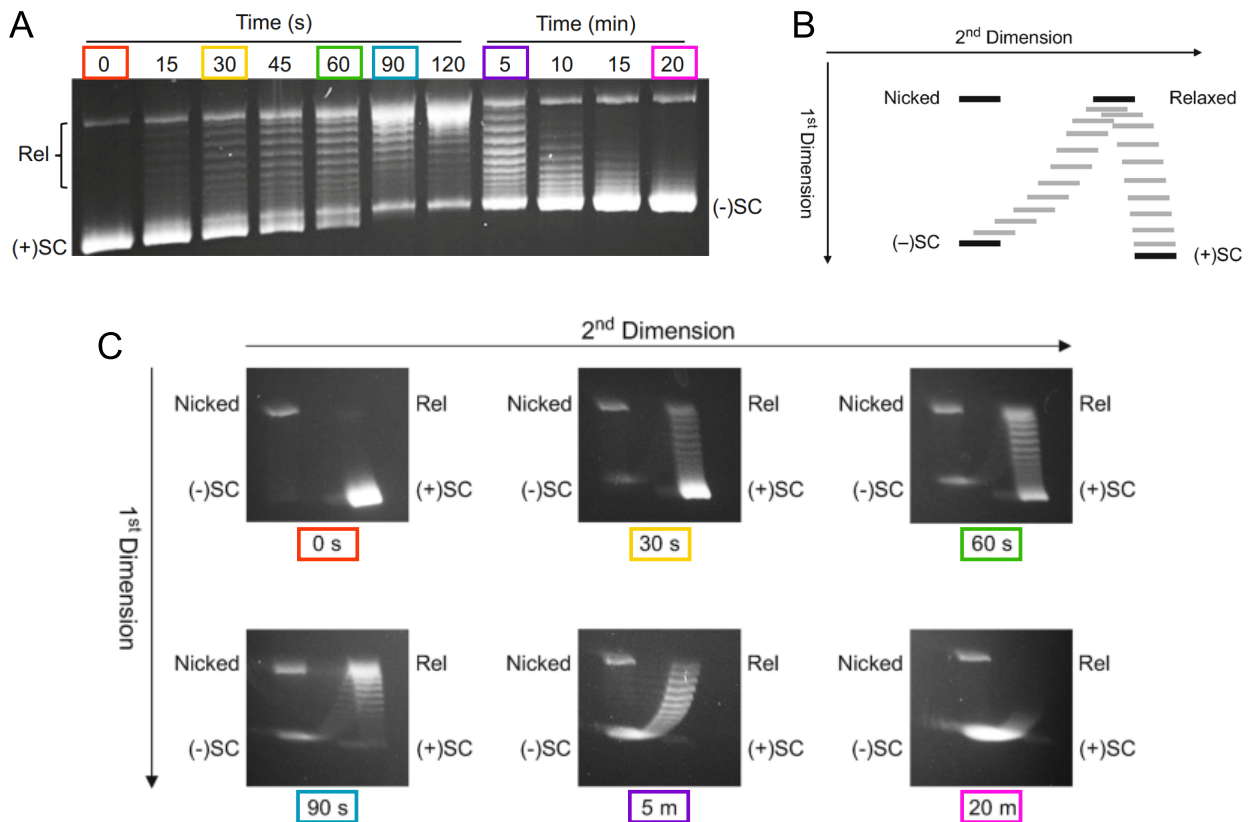


FIGURE 1.23 – Les méthodes d'étude du niveau de surenroulement global par électrophorèse. (A) Électrophorèse 1D sur gel d'agarose de plasmides surenroulés positivement traités avec de la gyrase purifiée de *Staphylococcus aureus* pendant des durées croissantes. Evolution du plasmide surenroulé positivement vers une forme surenroulée négativement selon la durée d'incubation avec la gyrase. (B) Schématisation de la migration des topoisomères selon leur niveau de surenroulement lors d'une électrophorèse à 2 dimensions, avec une migration sans chloroquine puis une migration avec ajout de chloroquine. Les molécules partiellement surenroulées négativement migrent entre la bande relâchée et la bande (-)SC. Les molécules partiellement surenroulées positivement migrent entre la bande (+)SC et la bande relâchée. (C) Électrophorèse 2D de plasmides surenroulés positivement traités avec de la gyrase purifiée de *Staphylococcus aureus* pendant des durées croissantes. Sur ces figures les positions des topoisomères relâchées sont indiquées par Rel, les topoisomères surenroulés positivement par (+)SC et les topoisomères surenroulés négativement par (-)SC. Les couleurs indiquent la correspondance entre la figure A et la figure C. Figures issues de [146].

Plus récemment, de nouvelles méthodes de mesures du niveau global de surenroulement *in vivo* utilisant la fluorescence ont été développées. L'une d'entre elles, la méthode FEDS (Fluorescent evaluation of DNA supercoiling), proposée par Duprey *et al.*, repose sur l'utilisation d'un plasmide porteur (1) d'un gène codant pour une protéine fluorescente verte associée à un promoteur dont le niveau d'expression dépend exclusivement du niveau de SC et (2) d'un gène codant pour une protéine fluorescente rouge transcrit à partir d'un promoteur constitutif, qui servira de standard interne (Fig. 1.24) [147]. Au

sein d'une cellule, l'hypothèse est faite que le niveau de SC de ce plasmide reflète le niveau de SC du chromosome. Le niveau de SC du plasmide est ensuite mesurable grâce aux fluorescences. Ainsi, cette méthode permet d'analyser le niveau de SC d'une cellule unique.

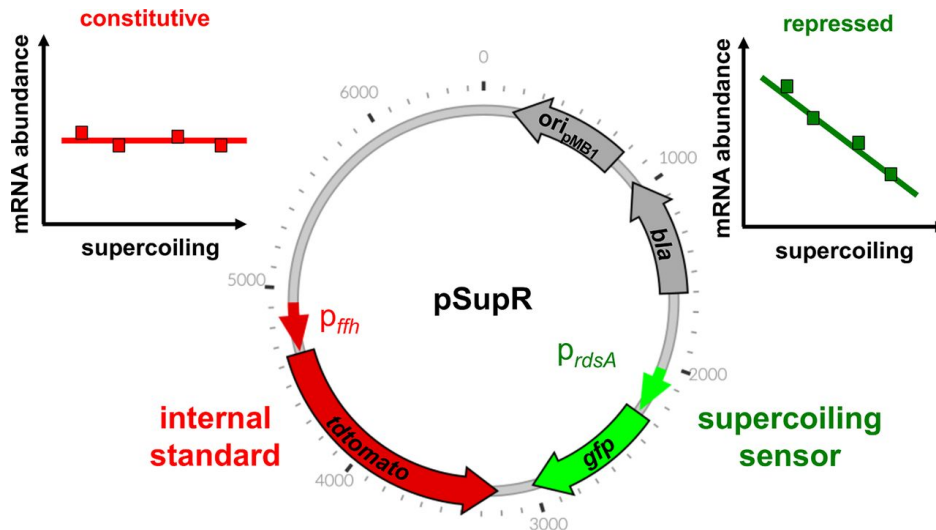


FIGURE 1.24 – Carte du plasmide rapporteur du SC : pSupR. Le plasmide comporte un promoteur constitutif ( $p_{ffh}$ ) et un promoteur exclusivement réprimé par le SC ( $p_{rdsA}$ ). Chaque promoteur permet la transcription d'un gène codant pour une protéine fluorescente, respectivement rouge et verte. Le gène *bla* apporte une résistance à l'ampicilline. Les graphiques en haut indiquent l'abondance des ARNm issus des gènes rouges et verts selon le niveau global de SC. Figure issue de [147]

### 1.3.3 L'impact des conditions environnementales

Selon les stress environnementaux rencontrés, le chromosome bactérien peut se relâcher ou à l'inverse avoir une augmentation du niveau de surenroulement, généralement passagère (Tableau 1.5). Des variations de surenroulement ont été observées en réalisant des électrophorèses de plasmides insérés dans des bactéries soumises à différents stress chez nos deux modèles d'études *E. coli* (Tableau 1.5) et *D. dadantii* (Fig. 1.25) mais également chez des espèces phylogénétiquement éloignées comme *B. subtilis* et *S. typhimurium* [83, 34, 148, 149]. Les gels présentés en Figure 1.25 montrent par exemple que l'ADN des bactéries *D. dadantii* soumises à un stress acide (provoqué par l'ajout d'acide malique) ou à un stress oxydatif (obtenu en ajoutant de l' $H_2O_2$ ) est plus relâché que l'ADN des bactéries non traitées [72].

Ces variations de niveau de surenroulement sont dues à des modifications de l'activité des topoisomérases selon les conditions environnementales. Selon les stress rencontrés, la charge énergétique de la cellule peut être modifiée, ce qui affecte l'activité de la gyrase puisque celle-ci nécessite cette énergie sous forme d'ATP. Lorsque le ratio  $[ATP]/[ADP]$

augmente, la gyrase est plus active et relâche donc le chromosome davantage [150]. D'autres mécanismes de la régulation de l'activité des topoisomérases impliquent des NAP ou la régulation au niveau transcriptionnel des gènes *topA*, *gyrA* et *gyrB* [83, 34, 72] (Tableau 1.5).

TABLEAU 1.5 – La variation du niveau de surenroulement du chromosome de *E. coli* en réponse à un stress environnemental. Les stress rencontrés provoquent des augmentations du niveau de SC-(Aug) ou des relâchements (Rel) de l'ADN. Tableau adapté de [83].

Choc	Variation du SC	Mécanisme	Source
Chaud	Rel	Modification des activités de la gyrase et de la topoI	[151]
Froid	Aug	Augmentation de l'activité de la gyrase, HU	[152]
Acide	Rel	Diminution de l'activité de la gyrase	[153]
Osmotique	Aug	Augmentation du ratio [ATP]/[ADP]	[150]
Oxydatif	Rel	Activation de l'expression de <i>topA</i> par Fis	[154]
Anaérobie	Rel	Diminution du ratio [ATP]/[ADP]	[150]

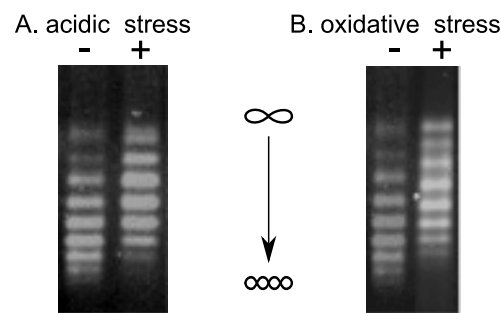


FIGURE 1.25 – Effets des chocs acide et oxydatifs sur le surenroulement de l'ADN de *D. dadantii* observés après électrophorèse sur un gel d'agarose contenant de la chloroquine. Des bactéries *D. dadantii*, préalablement électroporées avec le plasmide pUC18 ont été soumises pendant 15 min à (A) un stress acide avec l'ajout de 30 mM d'acide malique pour obtenir un pH de 4,3 et (B) un stress oxydatif provoqué par l'ajout de 500  $\mu\text{M}$  d' $\text{H}_2\text{O}_2$ . En parallèle, une culture sans stress de ces bactéries contenant pUC18 a également été effectuée. Les plasmides de chaque condition ont ensuite été extraits puis déposés sur un gel d'agarose à 1% contenant 2,5  $\mu\text{g}/\text{mL}$  de chloroquine. Étant donné que plus un plasmide est surenroulé, plus il migre loin dans le gel, et que si l'on admet que le surenroulement du plasmide reflète celui du chromosome bactérien, ces gels montrent que les deux stress testés provoquent le relâchement du chromosome de *D. dadantii*. Images issues de [72].

Dans le cas du choc froid par exemple, Mizushima *et al.* ont analysé, par des électrophorèses sur gel d'agarose en présence de chloroquine, l'effet de ce choc sur le SC de *E. coli* sauvage sans traitement, sauvage après traitement avec un inhibiteur de gyrase et dans une souche n'exprimant pas la protéine HU (Fig. 1.26) [152]. Ils ont observé que le choc froid provoque une augmentation du SC de la souche sauvage sans traitement mais que cette augmentation n'est pas visible lorsque la gyrase est inhibée par un antibiotique (Fig. 1.26A) et est

diminuée dans le mutant n'exprimant pas HU (Fig. 1.26B). D'après ces constats, Mizushima *et al.* ont conclu que la modification du niveau de SC en réponse à un choc froid dépend de l'activité de la gyrase et de la liaison de la protéine HU [152].

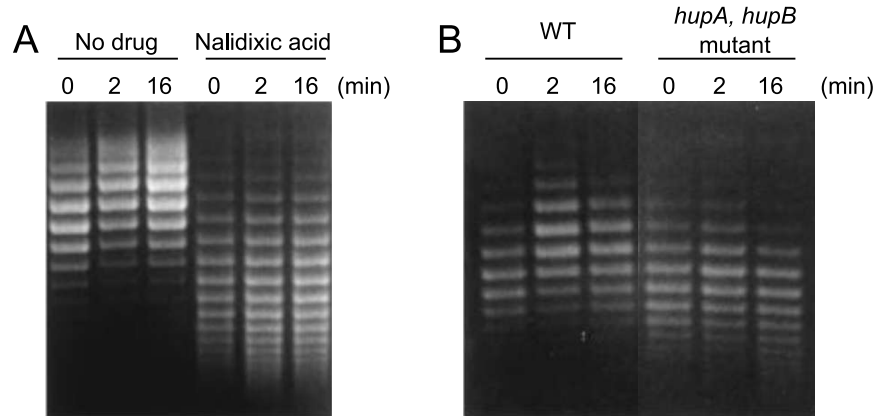


FIGURE 1.26 – Effet d'un choc froid (6°C) sur le surenroulement de l'ADN de *E. coli* (A) après un traitement avec un inhibiteur de gyrase (200 µg/mL d'acide nalidixique pendant 10 minutes) et (B) dans un mutant n'exprimant pas HU. Des bactéries *E. coli* en phase exponentielle de croissance à 42°C, préalablement électroporées avec le plasmide pUC118, ont été incubées à 6°C pendant 0, 2 et 6 minutes. Les topoisomères des plasmides extraits de chaque condition ont été séparés par électrophorèse sur un gel d'agarose à 1% en présence de 25 µg/mL de chloroquine. Gels issus de [152].

L'étude de la réponse de *E. coli* à un stress oxydatif a été réalisée par Weinstein-Fischer *et al.*. Leur première étape a été l'analyse du niveau de SC par électrophorèse sur gel d'agarose en présence de chloroquine après avoir traité les bactéries avec du peroxyde d'hydrogène [154]. Dès 2 minutes et jusqu'à 15 minutes après le traitement, ils ont ainsi observé un relâchement de l'ADN. Dans un second temps, ils ont étudié l'expression du gène *topA* (codant pour la topol) suite à ce même traitement, en évaluant la quantité d'ARNm issu du gène *topA* par une méthode d'extension d'amorces. Ils ont alors mis en avant l'induction de l'expression de *topA* provoquée par ce choc. Pour finir, ils ont réitéré cette même analyse dans des mutants des NAPs Lrp, IHF, H-NS et Fis et ont observé que l'induction de *topA* ne se produit pas en absence de Fis. En conclusion, leur étude a montré que le stress oxydatif, induit par un traitement au peroxyde d'hydrogène, active l'expression de *topA* uniquement en présence de Fis et conduit à une augmentation du niveau de SC [154].

## 1.4 La régulation de la transcription bactérienne par le surenroulement de l'ADN

Les modèles standards de la régulation de la transcription décrivent principalement les mécanismes liés aux facteurs de transcription, capables d'activer ou de réprimer l'expression d'un ou de plusieurs gènes (Section 1.2.3). Cependant, plus de la moitié des promoteurs d'*E. coli* ne sont reconnus par aucun facteur de transcription [155] et certains organismes semblent en être dépourvus [156, 157]. D'autres mécanismes de régulation complexes existent, dont le surenroulement de l'ADN, reconnu comme un régulateur global et ancestral de la transcription bactérienne [83].

### 1.4.1 Les études *in vitro* sur plasmide

Pour étudier l'effet du surenroulement sur la transcription, l'idéal serait de contrôler le niveau de surenroulement local, au niveau d'un promoteur particulier sur le chromosome, et de quantifier l'expression des gènes correspondants dans chaque condition. Une telle étude ne peut actuellement être réalisée que sur un plasmide car il n'existe pas de méthode suffisamment précise pour contrôler le niveau de surenroulement d'un chromosome à l'échelle locale.

Une gamme de plasmides avec des niveaux de surenroulement différents peut être préparée en incubant les plasmides avec de la topoI purifiée et différentes quantités de bromure d'éthidium [158, 159, 160, 72]. Ce plasmide peut être porteur du promoteur à étudier associé avec un gène rapporteur. Ainsi, la préparation d'une gamme de ce plasmide avec différents niveaux de surenroulement permettra d'étudier l'effet des variations de SC sur l'expression de ce promoteur.

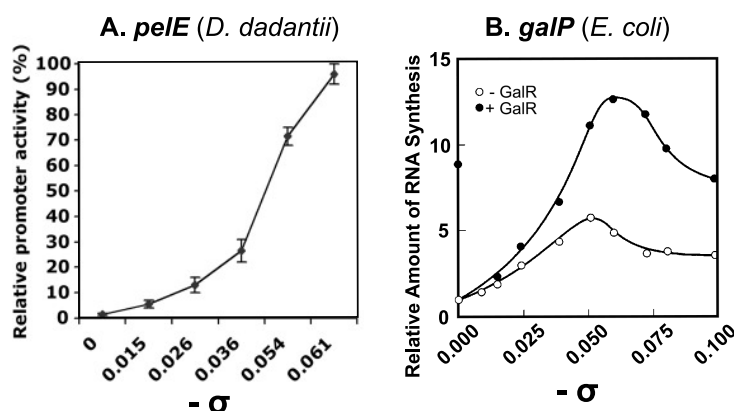


FIGURE 1.27 – Études sur plasmide de l'effet d'une variation de surenroulement sur la transcription (A) du gène *pelE*, codant pour une pectinase, chez *D. dadantii*, et (B) de l'opéron *gal*, codant pour les enzymes du métabolisme du galactose, chez *E. coli*, en présence et en absence de sa protéine régulatrice GalR. Figures issues de [72, 158]

Cette méthode a permis de montrer, chez *D. dadantii* (Fig. 1.27A) [72] et chez *E. coli* (Fig. 1.27B) [158, 159, 160], que le niveau d'expression de certains promoteurs est fortement modifié selon le niveau de surenroulement.

### 1.4.2 Les études transcriptomiques *in vivo*

S'il n'est pas possible de contrôler le niveau de SC à l'échelle locale sur un chromosome, une modification globale du niveau de SC est possible en soumettant les bactéries à des stress ou en affectant l'activité des topoisomérases. Pour réaliser une étude transcriptomique de la réponse d'une variation du niveau de SC, il est essentiel de pouvoir distinguer les effets liés à la variation du niveau de SC des autres effets. Dans le cas des bactéries soumises à des stress (choc de température, choc de pH, stress oxydatif, etc.), il est compliqué de différencier les gènes dont l'expression est affectée par le SC ou de ceux affectés par le stress. Une solution alternative pour modifier le niveau de SC est d'agir directement sur les topoisomérases, en les inhibant avec un antibiotique ou avec une mutation [161, 122, 162, 163, 62]. Le traitement avec un antibiotique présente les avantages (1) d'être peu chronophage, (2) de pouvoir être effectué sur les souches sauvages et (3) de pouvoir être utilisé à faible concentration et avec une courte durée d'incubation afin de ne pas affecter la croissance et pour permettre d'observer uniquement les effets directs de la variation de SC [83].

La très grande majorité des études réalisées dans ce but cible les topoisomérases de type II (gyrase et topoIV). Deux classes d'antibiotiques sont principalement utilisées : les aminocoumarins et les quinolones [164, 165, 166, 163, 34, 72, 167, 62, 152, 157, 168, 169]. Les coumarins, dont la novobiocine et la coumermycine, inhibent l'activité ATPase de la gyrase et, à haute concentration, de la topoisomérase IV [170, 171, 172]. Les quinolones, dont l'acide nalidixique, la ciprofloxacine et la norfloxacine, inhibent quant à elles l'activité de ligase de la gyrase et de la topoIV. Ils stabilisent l'intermédiaire enzyme-ADN et provoquent ainsi des cassures dans l'ADN, soit par dénaturation, soit par interaction avec la machinerie de réplication

Les topoisomérases étant conservées chez les bactéries [173], ces antibiotiques ont pu être utilisés pour réaliser des études transcriptionnelles chez des bactéries phylogénétiquement éloignées comme *E. coli*, *B. subtilis* et *Mycoplasma pneumoniae* (Tableau 1.6). Dans le cas de l'étude chez *D. dadantii*, des bactéries en début de phase exponentielle de croissance ou en phase de transition vers la phase stationnaire ont été traitées avec 100 µg/mL de novobiocine pendant 15 minutes. Ce traitement n'affecte pas la croissance de *D. dadantii* mais inhibe l'activité de la gyrase et provoque donc le relâchement de l'ADN. Les ARN de ces bactéries ont ensuite été extraits afin de réaliser une analyse RNA-Seq du transcriptome de *D. dadantii* après ce traitement. La comparaison des données d'expression avec et sans ce



traitement a révélé que plus de 1200 gènes sont significativement (c.-à-d. avec une valeur-p ajustée inférieure à 0,05) différentiellement exprimés en phase exponentielle et plus de 1650 en phase de transition vers la phase stationnaire. Environ la moitié (630 en phase exponentielle, 832 en phase de transition) de ces gènes sont activés. Ainsi, l'inhibition de la gyrase par la novobiocine impacte l'expression d'un grand nombre de gènes et cette régulation globale semble être liée au relâchement de l'ADN provoqué par cette inhibition [166]. De plus, les auteurs de cette étude ont remarqué que la proportion de gènes activés parmi les gènes différentiellement exprimés dépend de l'orientation de ces gènes par rapport à leurs deux voisins (voir classification présentée dans le paragraphe 1.1.4). Les gènes convergents sont plus activés que les gènes divergents [166].

Des études transcriptionnelles similaires, effectuées chez d'autres bactéries en inhibant la gyrase ou la topoI, indiquent que les variations de niveau de SC répriment ou activent significativement un nombre important de gènes (Tableau 1.6). Le SC semble être un régulateur global et ancestral de la régulation des gènes [83].

[174].

TABLEAU 1.6 – La réponse transcriptionnelle bactérienne à une modification de niveau de surenroulement provoquée par l'inhibition d'une topoisomérase par un antibiotique. Tableau adapté de [83].

Espèce	Variation du SC	Antibiotique	Nb de gènes affectés (% du génome)	Source
<i>Escherichia coli</i>	Rel	Norfloxacin	613 (15%)	[164]
	Rel	Novobiocine	1957 (48%)	[165]
<i>Dickeya dadantii</i>	Rel	Novobiocine	1249 (28%)	[166]
<i>Streptococcus pneumoniae</i>	Rel	Novobiocine	290 (14%)	[167]
	Aug	Séconéolitsine	545 (27%)	[175]
<i>Staphylococcus aureus</i>	Rel	Novobiocine	280 (11%)	[169]
<i>Bacillus subtilis</i>	Rel	Novobiocine	1075 (24%)	[168]
<i>Mycoplasma pneumoniae</i>	Rel	Novobiocine	469 (43%)	[157]

### 1.4.3 Les mécanismes envisagés

Bien qu'il n'existe actuellement pas de modèle détaillé de la régulation de la transcription par le SC, plusieurs mécanismes ont été proposés. Ces mécanismes semblent affecter simultanément différentes étapes critiques de la transcription, en particulier la reconnaissance des promoteurs par l'ARNP et l'initiation de la transcription [83].

La topologie de l'ADN permet à la fois de rapprocher des sites d'ADN et de les orienter, ce qui peut aider la fixation de facteurs de transcription et de l'ARNP. Par exemple, pour

être reconnus par les facteurs  $\sigma$ , les éléments -10 et -35 du promoteur doivent être alignés. Cet alignement dépend de la taille du spacer (15 à 20 pb, avec un optimal à 17 pb) mais également du niveau de SC. Le SC joue donc un rôle de régulateur de la transcription en modifiant l'orientation de ces éléments essentiels du promoteur [176, 177].

Un second mécanisme expliquant cette régulation est l'ouverture facilitée de la double hélice de l'ADN par l'ARNP en présence de SC négatif. La quantité d'énergie nécessaire à cette ouverture dépend du contenu en GC de la séquence promotrice, plus particulièrement du discriminateur, mais également du SC [178, 160]. En effet, le SC négatif déstabilise la double hélice et permet ainsi de diminuer fortement le coût énergétique de son ouverture [160, 179].

Il a également été montré que le SC pourrait affecter le déplacement de l'ARNP le long de l'ADN et ainsi perturber l'étape d'élongation [180] mais ce mécanisme reste à démontrer *in vivo*.

Plus indirectement, la topologie de l'ADN peut participer à la régulation de la transcription en impactant les autres mécanismes de régulation comme par exemple les facteurs de transcription et les NAPs. En effet, dans un premier temps, le niveau de SC peut affecter la liaison des facteurs de transcription et des NAPs à l'ADN. Puis, le SC peut impacter leur fonction régulatrice en modifiant l'état initial de la topologie. Dans le cas des facteurs de transcription qui activent la transcription en induisant un changement conformationnel par exemple, selon le niveau de SC, ce changement conformationnel peut ne pas apporter un alignement adéquat à la reconnaissance des éléments -35 et -10.

## 1.5 Étude du surenroulement de l'ADN à l'échelle locale

Jusqu'à présent les principales méthodes de mesure du niveau de SC permettent de quantifier le niveau global de SC sur un chromosome en utilisant un plasmide rapporteur (section 1.3.2). Il a cependant été montré que ce niveau varie localement, selon les activités locales et dynamiques de la transcription et des topoisomérases [181]. Pour approfondir l'étude du lien entre transcription et SC, il devient aujourd'hui essentiel d'en obtenir des cartographies à l'échelle locale et cela devient possible grâce à l'émergence de nouvelles méthodologies qui permettent de mesurer le niveau de SC local au sein d'un chromosome [182, 183].

### 1.5.1 Impact de la transcription sur le surenroulement de l'ADN

Du fait de la structure en double-hélice de l'ADN, il est nécessaire que le complexe de transcription (ARNP, ARN en cours de synthèse, protéines associées) tourne autour de l'ADN ou que l'ADN tourne autour de son propre axe hélicoïdal [184]. Dans ce second cas, l'ARNP agit comme un moteur topologique qui crée une distribution hétérogène du SC en amont et en aval de cette enzyme, avec davantage de supertours négatifs en amont et de supertours positifs en aval de sa position [181, 185] (Fig. 1.28). Ces supertours peuvent diffuser sur plusieurs kb [183] mais cette diffusion est limitée par la présence de barrières topologiques sur l'ADN, comme par exemple des protéines liées à l'ADN. Ces supertours négatifs et positifs s'accumulent donc respectivement en amont et en aval de l'ARNP, particulièrement autour des longs gènes ou opérons très transcrits comme les opérons ribosomaux [183]. Ce modèle a été décrit comme *Twin-supercoiled domain* par Liu et Wang en 1987 [181].

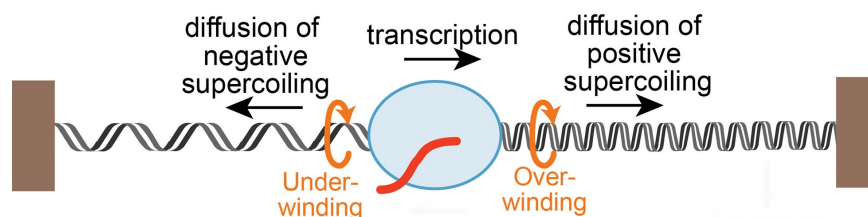


FIGURE 1.28 – Lors de la transcription, l'ARNP agit comme un moteur topologique, provoquant l'ajout de supertours négatifs en amont et positifs en aval de la région transcrite. La diffusion de ces supertours est limitée par les barrières topologiques (barres marrons) et conduit donc à une accumulation de supertours négatifs en amont et de supertours positifs en aval de la région transcrite. Schéma adapté de [186]

Comme présenté précédemment (Section 1.4), le niveau de SC affecte la transcription. Ce déséquilibre de SC peut donc, d'une part, constituer un frein à la poursuite de la transcription de ce gène et, d'autre part, affecter l'expression des gènes adjacents [187]. La

régulation de ce déséquilibre joue donc un rôle clé dans la régulation de la transcription [188]. Des études ont montré que la topoI et l'ARNP de *E. coli* peuvent être en interaction directe [189] et co-localisent *in vivo* sur le chromosome au niveau des régions promotrices chez *S. pneumoniae* [190] et *E. coli* [191]. Ces résultats semblent indiquer le rôle important des topoisomérases dans la régulation du SC dans les *Twin-supercoiled domain* [122].

Il existe donc un couplage fort entre la transcription et le SC de l'ADN [192]. L'impact *in vivo* de ce couplage est encore méconnu en raison du manque de données expérimentales sur la distribution locale du SC et pourra être davantage étudié grâce à l'émergence des nouvelles méthodes de cartographie du SC à l'échelle locale.

### 1.5.2 Les mesures du niveau local du surenroulement

Il n'existe actuellement pas de méthode permettant la mesure directe *in vivo* du niveau de SC du chromosome. Des protéines et des molécules dont l'intercalation ou la fixation sur l'ADN dépend du niveau de SC sont cependant connues et ont permis la récente mise au point de méthodes indirectes de cartographie du SC à l'échelle locale *in vivo*. Ces méthodes reposent sur des techniques proches du ChIP-Seq.

Le ChIP-Seq permet d'identifier les sites d'interaction d'une protéine sur un chromosome en fixant la protéine ciblée sur ses sites de fixation puis en séquençant uniquement les régions fixées par la protéine. Plus précisément, dans le cadre de la cartographie du SC local, les étapes suivantes sont réalisées (Fig.1.29). La protéine étudiée (GapR, topoI, gyrase) ou l'agent intercalant (psoralen biotinylé) est tout d'abord fixé sur l'ADN par pontage. L'ADN est ensuite fragmenté et les fragments liés par la protéine ou l'agent intercalant sont sélectionnés par un anticorps spécifique de la protéine ou par une molécule de haute affinité avec l'intercalant (la streptavidine dans le cas de l'intercalation du psoralen biotinylé). En parallèle, un échantillon de contrôle est également préparé en réalisant l'étape de fragmentation sans poursuivre avec la sélection. Les fragments d'ADN sont ensuite purifiés puis séquencés [193, 183, 191]. La densité de fixation le long du génome peut être obtenue à partir du séquençage grâce aux outils bioinformatiques en quelques étapes. Les séquences lues doivent être alignées sur le génome de référence avec un outil comme *Bowtie2* [116] puis la comparaison entre les alignements dans les échantillons avec sélection des fragments fixés par la protéine et dans les échantillons sans sélection peut être réalisée avec le module *bamCompare* de l'outil *deepTools* [194]. Il est également possible d'identifier les pics avec les plus hautes densités de fixation (*peak calling*) grâce à l'outil *MACS* [195].

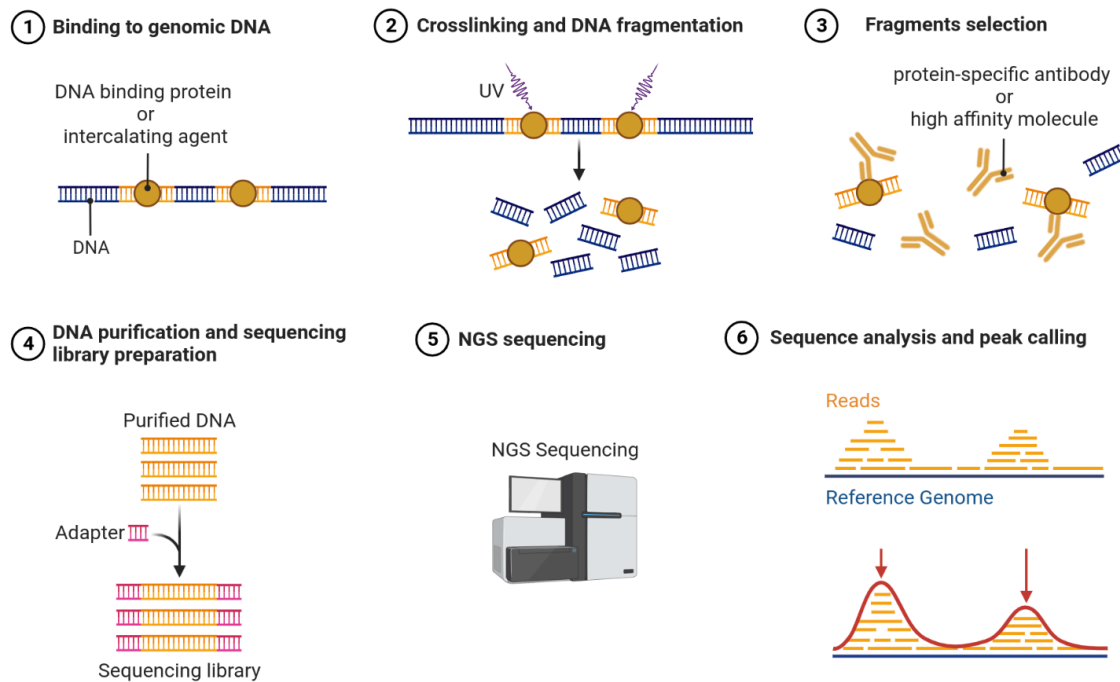


FIGURE 1.29 – Le protocole de ChIP-Seq et son adaptation dans le cadre de l'étude de l'intercalation d'une molécule dans l'ADN. Dans le but d'identifier les sites de liaison d'une protéine à l'ADN, le protocole de ChIP-Seq suit les étapes suivantes : (1) la protéine se lie à l'ADN, (2) la protéine est fixée sur l'ADN par pontage puis l'ADN est fragmenté, (3) les fragments pontés avec la protéine sont immunoprécipités à l'aide d'un anticorps spécifique de la protéine ciblée, (4) les fragments sélectionnés sont purifiés, dissociés de la protéine, et préparés pour le séquençage, (5) les fragments sont séquencés, (6) les séquences lues sont alignées sur le génome de référence afin d'observer la densité de liaison et de localiser des pics de fixation. Un protocole similaire peut être suivi pour localiser les sites d'intercalation d'une molécule, en modifiant l'étape (3). A cette étape, une molécule de haute affinité pour l'intercalant est utilisée afin de ne retenir que les fragments pontés par cet intercalant. Schéma créé à partir d'un modèle bioRender [196]

## La cartographie des fixations des topoisomérases

Les topoisomérases, en particulier la topoI et la gyrase, sont les acteurs clés de la régulation du niveau de SC. La topoI relâche l'ADN tandis que la gyrase ajoute des supertours négatifs [122]. De ce fait, la topoI se lie principalement dans les régions surenroulées négativement tandis que la gyrase se lie davantage dans les régions relâchées. Ainsi, localiser les sites ciblés par ces deux topoisomérases peut permettre d'avoir une cartographie du SC local [197, 191]. Sutormin *et al.* ont pour cela cartographié les sites de liaisons (ChIP-Seq) et les sites de coupures (Topo-Seq) des deux topoisomérases.

Le principe du Topo-Seq appliqué à la recherche des sites de clivages de la gyrase est le suivant (Fig. 1.30). Pour introduire des supertours négatifs dans l'ADN, la gyrase se lie à la double hélice puis clive ses deux brins, créant des extrémités 5' sortantes sur 4 pb. Les 2 brins clivés sont liés de manière covalente par des liaisons phosphodiesteres aux résidus de tyrosines catalytiques des sous-unités GyrA. Le complexe ainsi formé peut être

piégé par un inhibiteur comme la ciprofloxacine (quinolone). Pour localiser les sites clivés, l'ADN peut ensuite être fragmenté et les fragments d'ADN appartenant à un complexe avec une gyrase sont sélectionnés par immunoprécipitation. Les fragments sont ensuite purifiés et séquencés. Dans le cadre du Topo-Seq, le but de ce séquençage est d'identifier les extrémités 5' avec 4 pb sortantes puisque ces extrémités traduisent la présence d'un site de clivage de la gyrase [197]. Cette méthode permet donc le positionnement précis des sites de clivage.

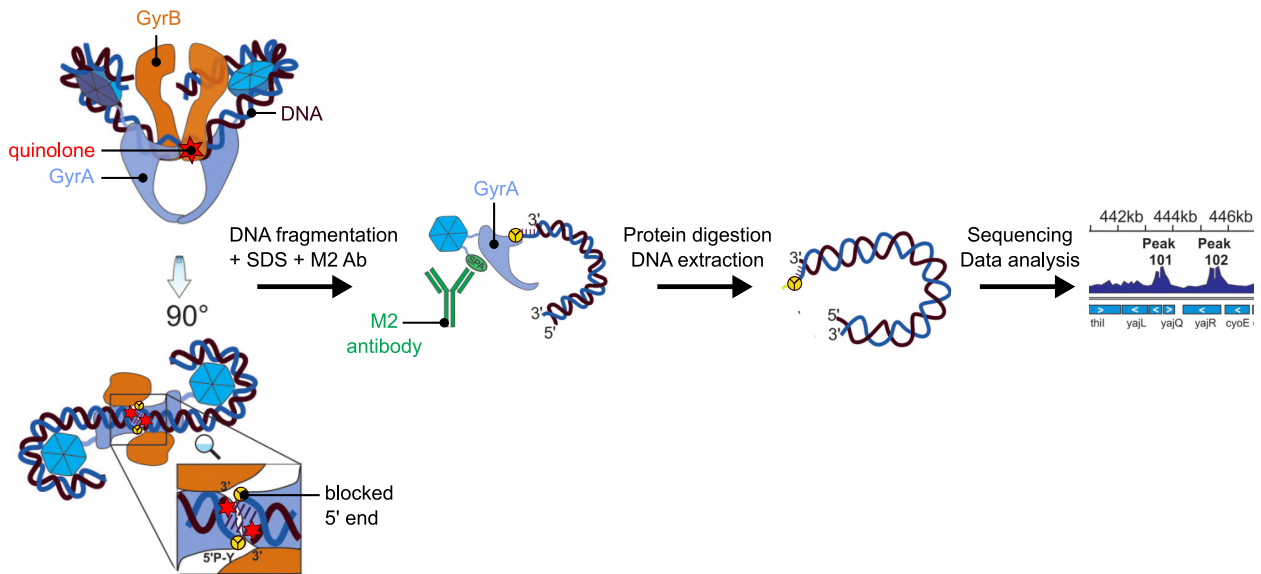


FIGURE 1.30 – Le principe du Topo-Seq repose sur le piégeage par un inhibiteur (ici de type quinolone) du complexe covalent entre la gyrase et l'ADN clivé. L'ADN est immunoprécipité afin de ne retenir que les fragments d'ADN en complexe avec une sous-unité GyrA. L'ADN est ensuite déprotéinisé et séquencé dans le but de localiser les sites de clivage. Schéma adapté de [197].

### La cartographie des supertours positifs par GapR

GapR est une protéine récemment identifiée en tant que NAP chez *Caulobacter crescentus* [199] et ayant des orthologues chez d'autres  $\alpha$ -protéobactéries [200]. Essentielle à la croissance et à la division cellulaire de cette bactérie [201], GapR semble stimuler les topoisomérases de type II (gyrase et topo IV) et jouer ainsi un rôle important lors de la réplication [193]. D'après les modèles actuels (Fig. 1.31) GapR se lie à l'ADN en conformation ouverte (dimère de dimères indépendants ou tétramère ouvert) puis subit un réarrangement pour entourer l'ADN en formant un tétramère fermé [193, 198, 202]. Les régions liées par GapR sont des régions riches en AT et surenroulées positivement. Cette dernière spécificité indique que GapR peut permettre de détecter les supertours positifs dans l'ADN. Pour ce faire, Guo *et al.* ont appliqué la technique de ChIP-Seq pour détecter les sites de liaisons de GapR sur l'ADN chez *C. crescentus* [193] puis chez *E. coli* [182].

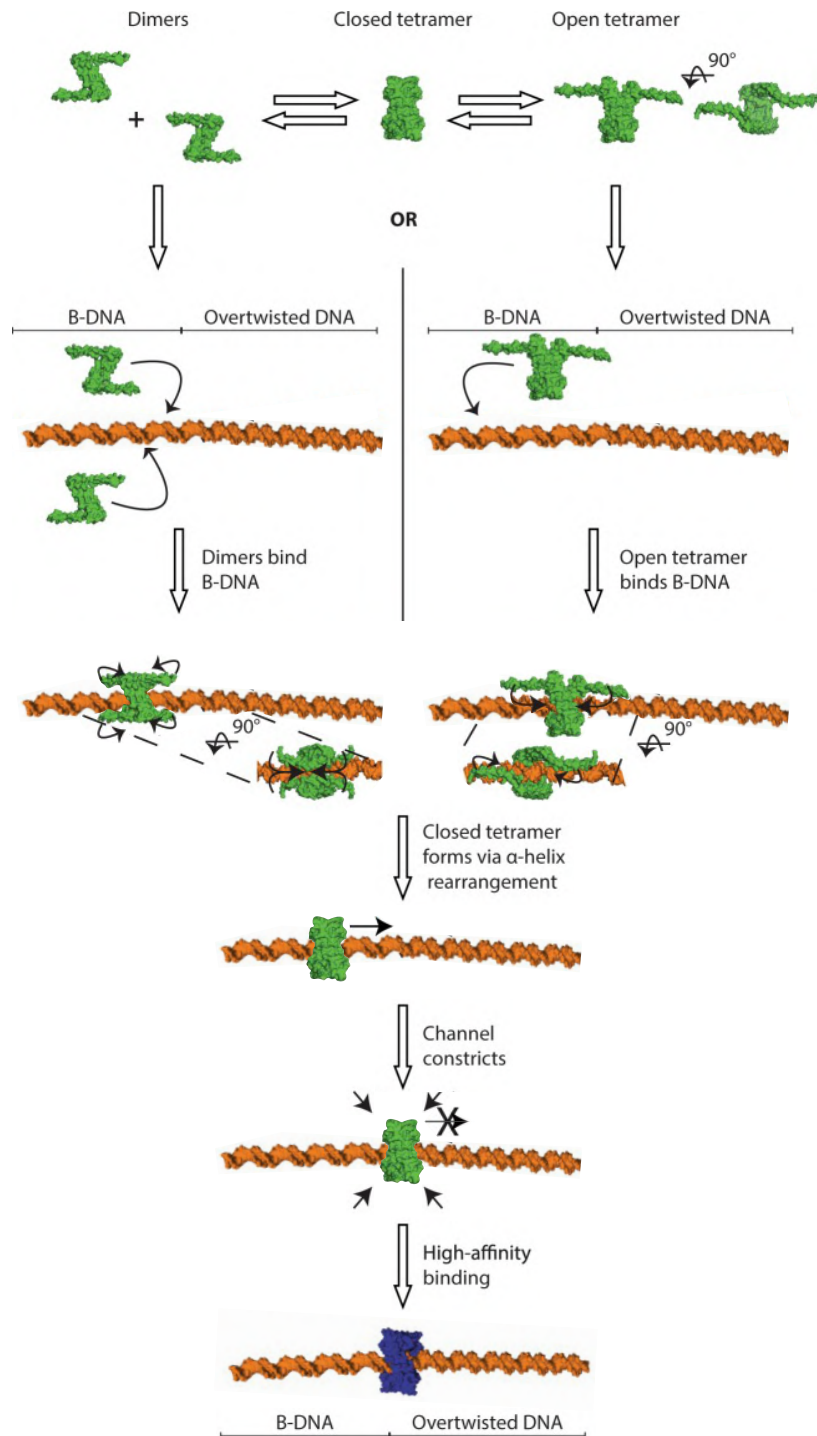


FIGURE 1.31 – Les modèles actuels de la liaison de GapR (en vert) à l'ADN (en orange). GapR s'associe à l'ADN en deux dimères (modèle à gauche) ou en tétramère ouvert (modèle à droite). Dans les deux cas, GapR s'organise ensuite en un tétramère fermé encerclant l'ADN. Dans cette conformation, GapR diffuse le long de l'ADN jusqu'à la rencontre d'une région surenroulée positivement. Cette rencontre provoque le resserrement de GapR autour de cette région et la création d'un complexe de forte affinité. Schéma adapté de [198].

### La cartographie du surenroulement négatif par le psoralen

A la fin du XX<sup>e</sup> siècle, des études ont montré que le psoralen permet d'étudier le SC d'un chromosome, qu'il soit bactérien ou eucaryote [203]. Le psoralen, et en particulier

son dérivé le 4,5',8-triméthylpsoralen (TMP), sont des molécules planes qui s'intercalent dans l'ADN, principalement dans les régions surenroulées négativement, et forment des pontages stables avec les pyrimidines, préférentiellement les thymines, après exposition à des UV à une longueur d'onde de 360 nm (Fig. 1.32) [203, 204, 205].

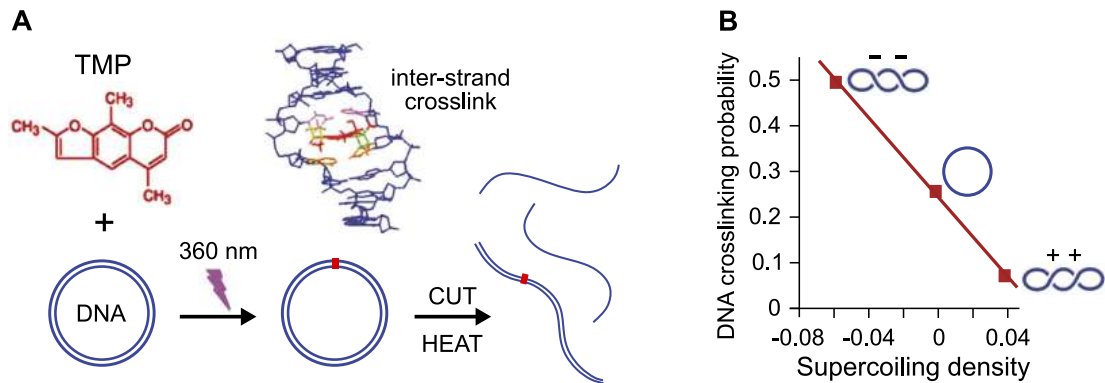


FIGURE 1.32 – Effet du surenroulement de l'ADN sur la probabilité d'intercalation du 4,5',8-triméthylpsoralen (TMP) dans l'ADN. (A) Le TMP s'intercale dans l'ADN et, sous radiation UV à 360 nm, y crée des pontages. Lorsque l'ADN est fragmenté puis dénaturé, seuls les fragments pontés restent sous forme double brin. (B) Représentation schématique de la probabilité de pontages par l'intercalation du TMP dans un ADN circulaire, ici le plasmide de 8 kb YCp50, selon son niveau de SC. Schémas issus de [206].

Plusieurs méthodes utilisant cette propriété d'intercalation spécifique du TMP dans les régions surenroulées négativement ont été mises au point pour localiser ces régions. Les premières étapes sont communes aux différentes méthodes. Tout d'abord, les cultures sont traitées avec du TMP avant d'être soumises à une radiation UV. Le TMP s'intercale dans les régions surenroulées négativement et, sous UV, y crée des pontages stables. Ensuite, l'ADN est extrait puis fragmenté (Fig. 1.32A). A cette étape, un fragment issu d'une région surenroulée sera davantage ponté que ce même fragment (séquence, contexte génomique et interactions avec les protéines identiques) dans un contexte relâché. Les étapes suivantes dépendent de la méthode utilisée (Fig. 1.33).

Pour savoir si une région spécifique est surenroulée négativement, l'ADN fragmenté peut être déposé sur un gel d'agarose dénaturant (Fig. 1.33 partie gauche). La dénaturation provoque la séparation des brins des fragments non pontés (c.-à-d. les fragments dans lesquels aucun TMP ne s'est intercalé) tandis que les fragments pontés restent sous forme double brin. Sur ce gel, les fragments simple brin migrent plus rapidement que les fragments double brin. Les fragments initialement surenroulés négativement (c.-à-d. les fragments dans lesquels le TMP s'est intercalé et a formé un pontage) migrent plus lentement que les fragments initialement plus relâchés ou surenroulés positivement. En réalisant un southern blot sur ce gel, il est ensuite possible de détecter si une séquence recherchée se retrouve davantage parmi les fragments surenroulés négativement ou parmi les autres fragments



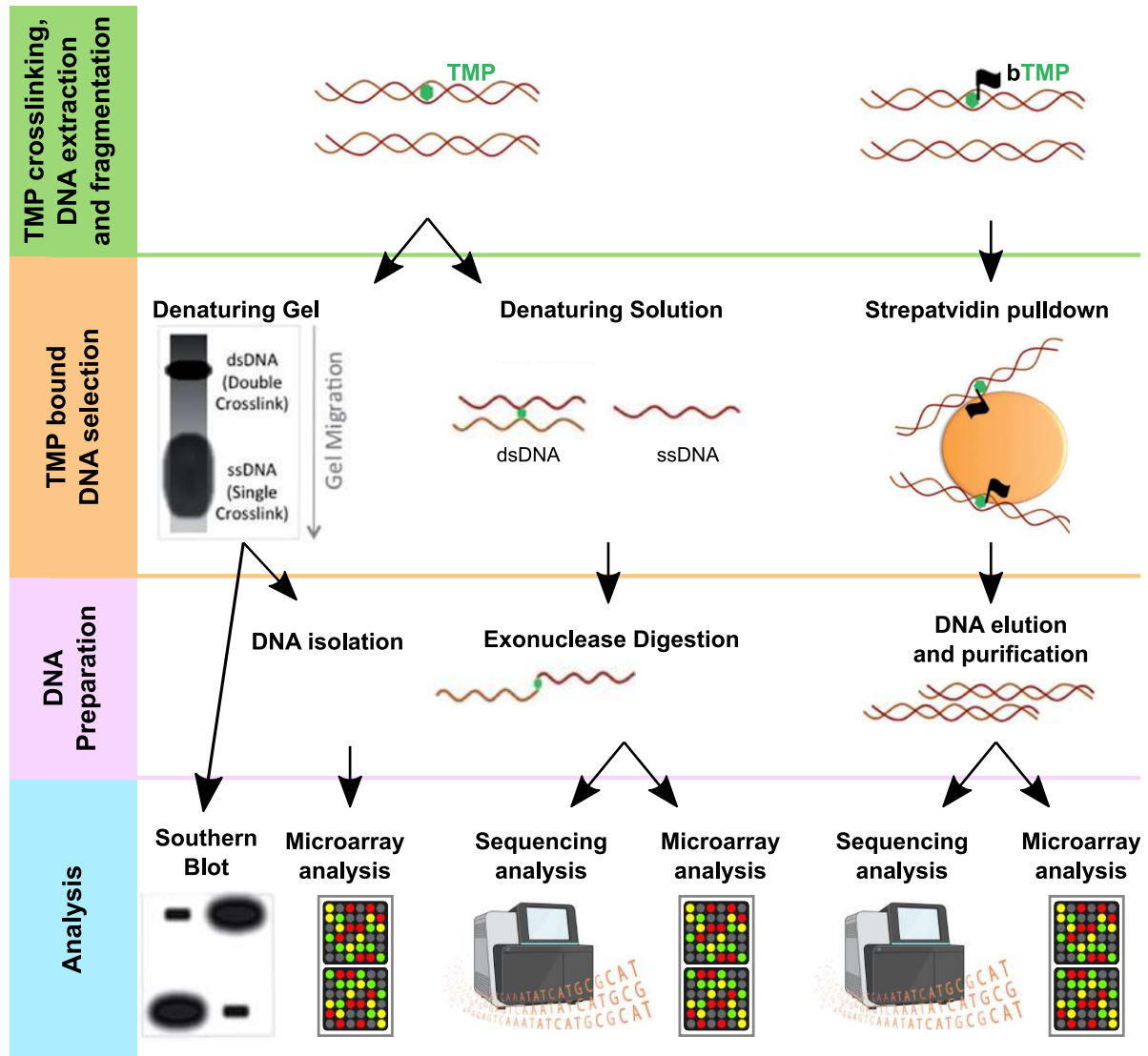


FIGURE 1.33 – Méthodes de cartographie du SC local utilisant les propriétés d’intercalation du 4,5,8-triméthylpsoralen (TMP). Trois méthodes sont présentées sur le schéma.

Dans la première (partie gauche de la figure), le TMP est utilisé en intercalant puis, après pontage, extraction de l’ADN et fragmentation, l’échantillon est déposé sur un gel dénaturant. Ce gel permet de séparer les fragments d’ADN double brin, ayant été pontés par le TMP, des fragments d’ADN simple brin. A partir de ce gel, des séquences spécifiques peuvent être recherchées parmi ces fragments en réalisant un southern blot ou en isolant l’ADN du gel pour ensuite l’analyser sur une puce à ADN. Dans la deuxième méthodologie présentée (partie centrale de la figure), les fragments sont traités dans une solution dénaturante. Suite à ce traitement, lorsque les fragments sont soumis à une digestion par des exonucléases, les fragments non pontés sont complètement digérés et seuls les fragments pontés subsistent. Les fragments restants peuvent ensuite être analysés sur puce à ADN ou par séquençage.

Dans la troisième méthode (partie droite de la figure), l’intercalant est le TMP biotinylé. Après intercalation de ce TMP biotinylé puis extraction et fragmentation de l’ADN, les fragments contenant du TMP biotinylé peuvent être retenus grâce à l’affinité de la biotine pour la streptavidine. Ces fragments retenus sont ensuite séquencés ou analysés sur puce à ADN. Schéma adapté de [140].

[207, 208]. Pour analyser plusieurs séquences simultanément, il est possible d'isoler l'ADN double brin et l'ADN simple brin à partir du gel et d'y rechercher différentes séquences à l'aide d'une puce à ADN [209].

Pour analyser l'ensemble des régions du chromosome, deux méthodes sont actuellement utilisées et se distinguent tout d'abord par l'intercalant utilisé : le TMP (utilisé dans le Chapitre 3, section 3.2.1) [210] et le TMP biotinylé (ou psoralen biotinylé dans le cas du chapitre 3, section 3.2.3 [183]). Dans le premier cas (Fig. 1.33 partie droite), après l'étape de fragmentation de l'ADN, l'ADN est dénaturé dans des solutions dénaturantes ou par choc thermique. Cette dénaturation provoque la séparation complète des deux brins constituant les fragments non pontés et la séparation partielle des fragments pontés puisque les brins appartenant à ces derniers fragments restent liés au niveau du pontage. Les fragments séparés sont ensuite digérés par des exonucléases. Ces exonucléases dégradent l'ADN simple brin du 5' vers le 3'. Ainsi, les fragments ne contenant pas de pontage sont entièrement dégradés tandis que la dégradation des fragments avec intercalation est interrompue, laissant un simple brin non dégradé de part et d'autre du site de pontage. Ces fragments restants peuvent ensuite être séquencés afin de localiser les sites de pontages et d'obtenir une mesure quantitative du SC de l'ADN. Plus une région est surenroulée, plus la fréquence de pontages provoqués par l'intercalation du TMP dans cette région est élevée, moins cette région est dégradée par les exonucléases, plus la région sera retrouvée en grande quantité lors du séquençage [210]. Dans le second cas (Fig. 1.33, partie droite), de la biotine est liée au psoralen [211] avant le traitement de la culture. Après l'extraction et la fragmentation de l'ADN, la biotine, grâce à son affinité avec la streptavidine, permet de sélectionner uniquement les fragments dans lesquels le psoralen biotinylé s'est intercalé (c.-à-d. les fragments issus des régions surenroulées négativement). Le séquençage des fragments sélectionnés permet ensuite de connaître le niveau de SC des régions [183].

### Récapitulatif des méthodes de mesure du SC

Jusqu'à présent les méthodes les plus répandues de mesures du SC étaient des mesures qui permettaient uniquement d'estimer le niveau global de SC d'un chromosome, généralement par l'intermédiaire d'un plasmide rapporteur (Tableau 1.7, partie supérieure). Actuellement, de nouvelles méthodes sont développées et appliquées afin d'obtenir des cartographies du SC local (Tableau 1.7, partie inférieure). Ces mesures permettront d'approfondir les études de l'impact du SC sur l'expression et la régulation du génome.

TABLEAU 1.7 – Les méthodes de mesure du SC

Méthode	Principe	Matrice	Niveau	Mesure	Cellule unique ?
Gradient de sucrose	Sédimentation	Chromosome	Global	SC+/- non dissociable	Non
Gel d'agarose 1D	Électrophorèse	Plasmide rapporteur	Global	SC+/- non dissociable	Non
Gel d'agarose et chloroquine 2D	Électrophorèse	Plasmide rapporteur	Global	SC+ et SC-	Non
FEDS <sup>a</sup>	Fluorescence	Plasmide rapporteur	Global	SC+ et SC-	Oui
ChIP-Seq et Topo-Seq de topoI	ChIP-Seq	Chromosome	Local	SC-	Non
ChIP-Seq et Topo-Seq de gyrase	ChIP-Seq	Chromosome	Local	SC+	Non
Psora-Seq	ChIP-Seq	Chromosome	Local	SC-	Non
GapR-Seq	ChIP-Seq	Chromosome	Local	SC+	Non

<sup>a</sup> Fluorescent Evaluation of DNA Supercoiling

## 1.6 Les objectifs de ma thèse

Le SC est un mécanisme essentiel, complexe et dynamique. Essentiel à la compaction du chromosome bactérien, il est finement régulé, principalement par les topoisomérases. L'étude du SC est complexe car sa distribution est hétérogène le long du chromosome et les variations de son niveau participent à la régulation de la transcription. Les modèles standards de la régulation de la transcription prennent principalement en compte les facteurs de transcription et très peu la thermodynamique liée au SC. Pourtant, le SC affecte un grand nombre de processus liés à la transcription, à commencer par la séparation des deux brins d'ADN dont le niveau d'énergie requis dépend du SC, mais également la liaison de l'ARNP et des facteurs de transcription. Il a déjà été montré que le SC est un régulateur global et ancestral de l'expression des gènes mais il est essentiel de poursuivre les études en ce sens, afin de montrer que toutes les variations de SC, que ce soient des augmentations du niveau de SC ou des relâchements, affectent l'expression des gènes.

Jusqu'à présent, la majorité des études de la réponse transcriptionnelle des gènes à une variation de SC s'est focalisée sur le relâchement de l'ADN provoqué par l'inhibition de la gyrase. Les gènes dont l'expression est significativement affectée par cette inhibition ont été qualifiés de gènes "sensibles au surenroulement" mais qu'en est-il de l'expression de ces gènes lorsque le niveau de SC augmente? Pour répondre à cette question, je présenterai, dans le Chapitre 2, le premier transcriptome d'une bactérie à Gram négatif dont la topoI est inhibée par un antibiotique afin d'augmenter le niveau de SC. La comparaison des données obtenues après l'inhibition de la topoI et de la gyrase permet d'approfondir nos connaissances sur la régulation transcriptionnelle par le SC et de proposer un modèle qualitatif de la réponse d'un promoteur à une variation de SC. La réalisation de cette étude chez la bactérie *D. dadantii* représente un enjeu écologique et économique puisque comprendre le rôle du SC dans le processus infectieux pourrait à terme permettre d'envisager des nouveaux moyens de lutte contre cet agent phytopathogène.

Du fait notamment de l'activité locale et dynamique des topoisomérases ainsi que de la présence de différentes barrières topologiques, le niveau du SC est hétérogène le long du chromosome et affecte donc différemment l'expression des gènes selon leur localisation. Se posent alors les questions suivantes : Comment étudier le niveau local du SC sur le chromosome? Comment prendre en compte ces variations locales dans les analyses transcriptomiques? Des nouvelles méthodes de cartographie du SC local sont actuellement développées et nécessitent des calibrations afin d'être adaptées aux différents organismes. Je présenterai au Chapitre 3 quelques étapes de calibration de la méthode de Psora-Seq chez *D. dadantii*. A terme, ces méthodes permettront l'approfondissement des études transcriptomiques de l'effet du SC sur la régulation des gènes. Pour pouvoir faciliter ce type d'analyse, nous avons développé un package bioinformatique nommé GRATIOSA.

Ce package Python, présenté au Chapitre 3, a pour but de faciliter l'étude de l'expression du génome bactérien selon l'organisation du chromosome. Il permet l'intégration de données de différents types et la réalisation de tests statistiques permettant l'analyse de ces données. La mise à jour de ce package développé depuis quelques années permet d'améliorer l'uniformisation, la reproductibilité et la traçabilité des études réalisées par les utilisateurs de ce package. Avec ce package, des données nouvelles ou déjà publiées peuvent être analysées et nous avons ainsi pu approfondir l'analyse des *Twin-supercoiled domain* chez *E. coli* à partir de données publiées. Cette analyse sera également détaillée au Chapitre 3.

Enfin, toujours dans le but d'étudier l'organisation du chromosome bactérien, un nouveau projet d'étude des interactions chromosomique a débuté chez *D. dadantii*. Les premiers résultats obtenus grâce à la méthodologie Hi-C sont également présentés dans le Chapitre 3.

Pour résumer, ma thèse a pour but d'approfondir l'étude de la régulation de la transcription par le SC chez les bactéries en :

- traitant de nouvelles données novatrices d'expression des gènes en réponse à l'inhibition de la topoI, enzyme régulatrice du niveau de SC,
- développant des outils bioinformatiques facilitant la réalisation d'analyses statistiques de l'expression et de la régulation du génome selon ses propriétés spatiales,
- analysant l'organisation du chromosome de *D. dadantii*.

## Chapitre 2

# Etude de la réponse transcriptionnelle de *Dickeya dadantii* à une augmentation du niveau de surenroulement

### 2.1 Le genre *Dickeya*

Pour étudier l'impact du SC sur la transcription, la bactérie phytopathogène *D. dadantii* est un modèle de choix. En effet, cette bactérie est un modèle bien caractérisé génétiquement [33] et facilement manipulable grâce à ses similarités avec *E. coli*. Elle est décrite par de multiples analyses transcriptomiques [34]. Il a été montré *in vitro* que les stress rencontrés par *D. dadantii* pendant l'infection des plantes affectent le niveau moyen de SC de cette bactérie et des études *in vivo* indiquent que de telles modifications du niveau de SC affectent l'expression de gènes essentiels à la réussite de cette infection [72, 26, 212]. Le SC semble donc être un mécanisme pertinent dans la recherche de moyen de lutte contre ce pathogène. Mieux comprendre le mécanisme de régulation de la transcription par le SC représente donc un enjeu écologique et économique.

#### 2.1.1 Le cycle infectieux des *Dickeya* spp.

Les bactéries *Dickeya* peuvent survivre dans des environnements divers dont le sol, les eaux souterraines, les rivières ou les mares [27]. Au sein des cultures agricoles, leur propagation peut se faire par migration dans le sol, par contact direct d'une plante blessée avec un outil contaminé ou une autre plante infectée mais également via les eaux de surface et d'irrigation [27, 213]. Les insectes, dont les pucerons, peuvent également être impliqués dans le développement de la maladie, en transportant les bactéries d'une plante malade vers une plante saine ou en blessant la plante, lorsqu'ils se nourrissent ou déposent leurs oeufs [27, 213].

Une fois l'hôte rencontré, le cycle infectieux des *Dickeya* spp. virulentes suit le rythme des cultures infectées (Fig. 2.1) [27]. Chez les pommes de terre par exemple, les bactéries du genre *Dickeya* peuvent être présentes sur les semences au printemps et peuvent se propager dans les jeunes tiges et les racines. Pendant l'été, elles se multiplient dans les espaces intercellulaires des plantes, dégradent les parois cellulaires et migrent à travers le sol pour infecter de nouvelles plantes. Pendant l'automne, après la récolte, *Dickeya* spp. peut continuer de se propager dans de nouvelles plantes en y pénétrant via les blessures causées lors de la récolte ou de la manutention. Enfin, pendant l'hiver, *Dickeya* spp. survit en association avec les résidus de culture dans le sol [27].

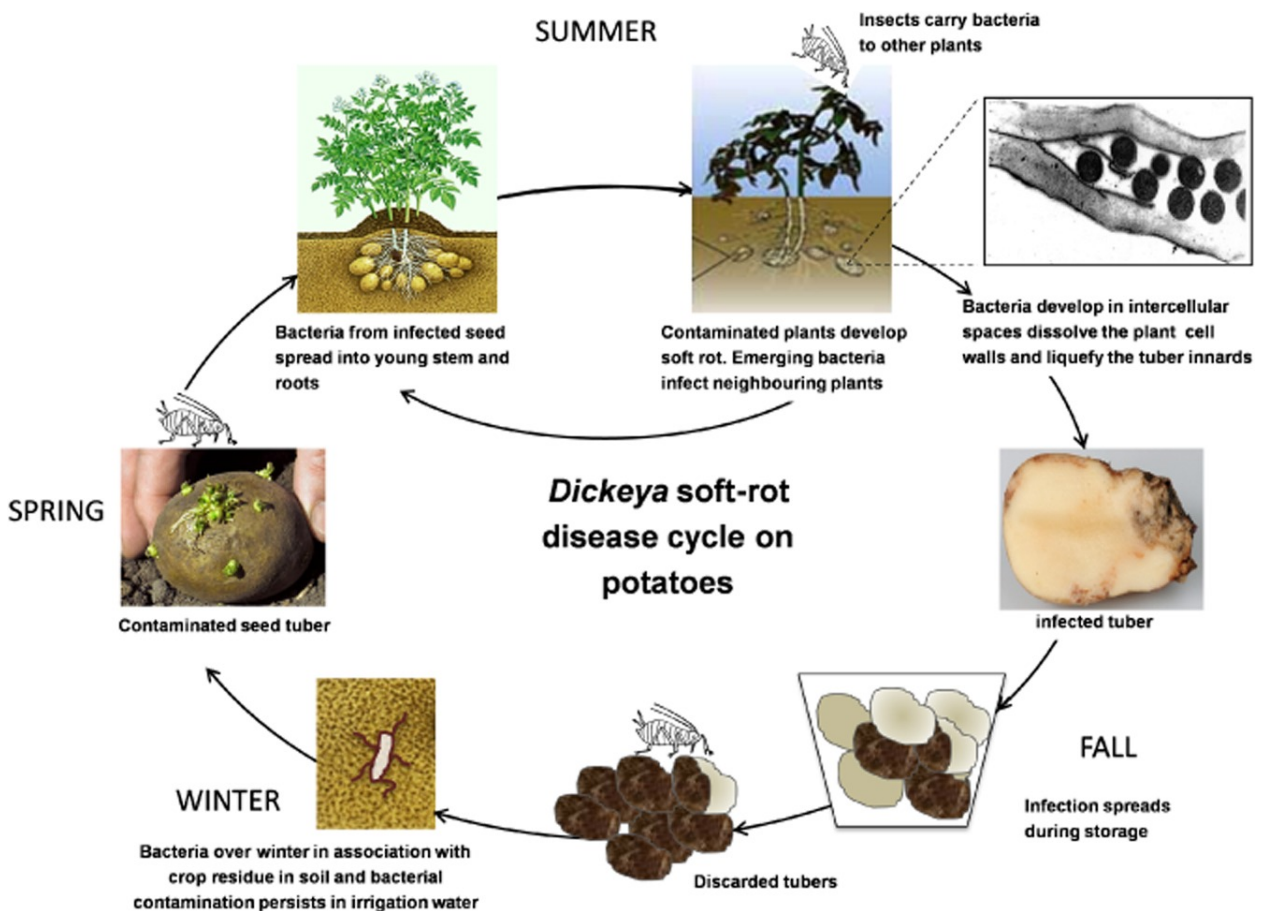


FIGURE 2.1 – Le cycle de vie de *Dickeya* spp. en interaction avec les pommes de terre. Au printemps, les bactéries du genre *Dickeya* spp. peuvent fréquemment être trouvées sur les tubercules de semences. Ces bactéries se propagent dans les jeunes plants en été et peuvent migrer à travers le sol ou l'eau pour infecter de nouvelles plantes. Les symptômes de pourriture molle apparaissent lorsque les bactéries dégradent les parois cellulaires des pommes de terre. Lors de la manipulation et du stockage des pommes de terre, les bactéries peuvent se propager à de nouveaux tubercules par le biais de toutes les blessures infligées lors de ces opérations, ainsi que par celles causées par les insectes lors de leur alimentation ou de leur ponte. En effet, les insectes jouent un rôle dans la propagation de l'infection en transportant les bactéries d'une pomme de terre malade à une pomme de terre saine. Pendant l'hiver, les bactéries survivent en association avec les résidus de culture dans le sol et peuvent continuer à se propager à travers les eaux de surface et d'irrigation. Schéma issu de [27].

### 2.1.2 Le processus infectieux de la souche *Dickeya dadantii* 3937

*D. dadantii* fait partie des neuf espèces phytopathogènes du genre *Dickeya* identifiées chez des plantes (voir Chapitre 1, paragraphe 1.1.2 en page 4). En conditions favorables (température de 30°C environ, humidité élevée, faible disponibilité en oxygène), *D. dadantii* peut infecter une culture de végétaux en quelques jours. Cette infection se décompose en une première phase asymptomatique lors de laquelle la bactérie pénètre et colonise les tissus de la plante puis une seconde phase symptomatique pendant laquelle l'infection se généralise, provoquant la macération des tissus et conduisant à la mort de la plante (Fig. 2.2) [27].

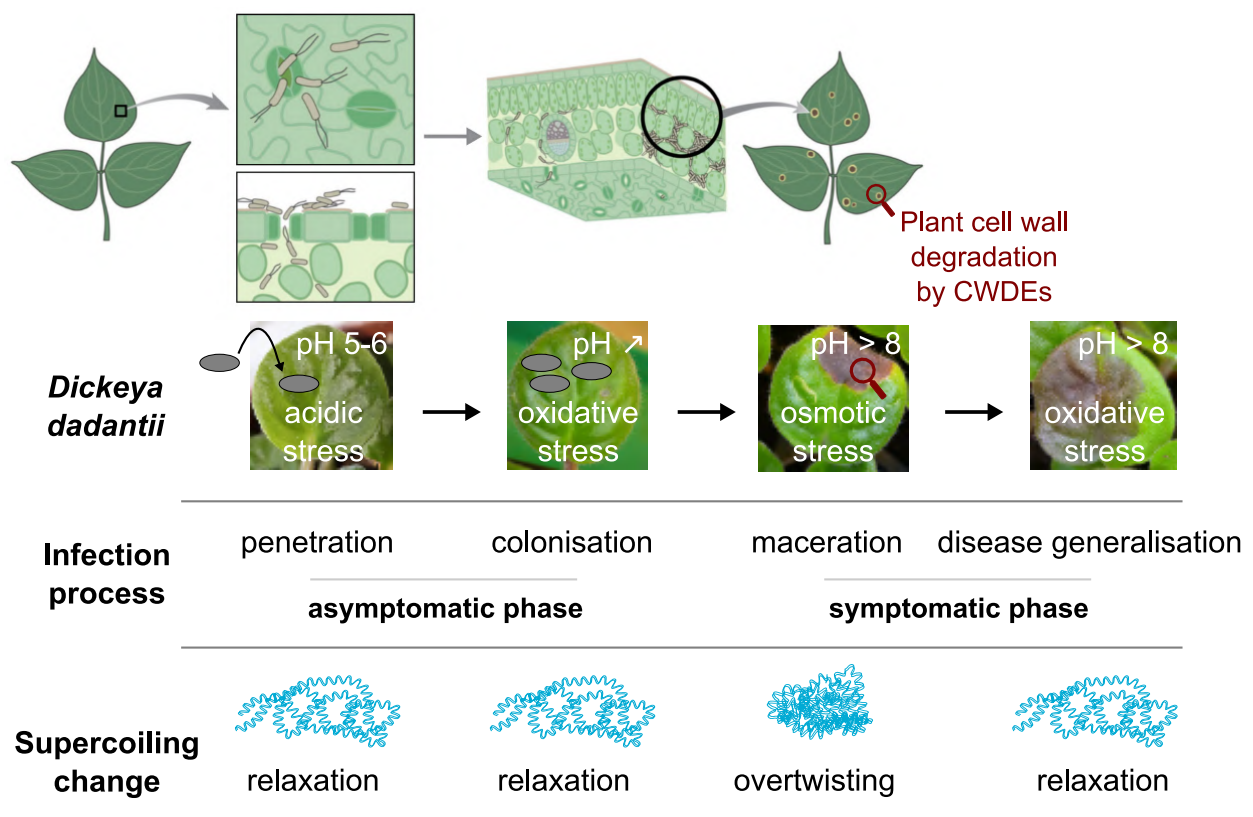


FIGURE 2.2 – Les étapes de l'infection d'une plante par *D. dadantii*. Le processus infectieux comporte une phase asymptomatique et une phase symptomatique. Pendant la première phase, *D. dadantii* pénètre dans la plante puis colonise les tissus. La phase symptomatique se caractérise par l'apparition des symptômes de la pourriture molle. Ces symptômes sont dus aux CWDEs, enzymes qui dégradent les parois cellulaires des végétaux. Pendant ce processus d'infection, les bactéries sont soumises à différents stress qui affectent leur SC. Schéma adapté de [214, 215].

Un grand nombre de gènes de virulence, dont une sélection est présentée en Tableau 2.1, sont impliqués dans ce processus.



TABLEAU 2.1 – Sélection de gènes de virulences impliqués dans différentes étapes de l'infection. Extrait du tableau publié par [27].

Gène(s)	Enzyme(s)	Rôle pendant l'infection	Source
<i>bcsABCD</i>	Synthèse de microfibrilles de cellulose	Adhésion à la surface de la plante	[216]
<i>wza-wzb-wzc</i>	Synthèse d'exopolysaccharides	Adhésion à la surface de la plante	[217]
<i>rhlA</i>	Synthèse de biosurfactants	Colonisation de la surface de la plante	[218]
<i>motAB-cheAWDRBY</i>	Flagelles et chimiotaxie	Colonisation de la la surface de la plante	[219]
<i>cfa</i>	Synthèse d'acides gras cyclopropane	Résistance au stress acide	[27]
<i>asr</i>	Protéine périplasmique	Résistance au stress acide	[27]
<i>sodA-C</i>	Superoxyde dismutase	Résistance au stress oxydatif	[220]
<i>katG-E</i>	Catalase-peroxydase	Résistance au stress oxydatif	[217]
<i>acsF-A</i>	Synthèse d'achromobactine	Capture du fer	[221]
<i>cbrABCF</i>	Système de transport de l'achromobactine	Capture du fer	[221]
<i>fct-cbsCEBAP</i>	Synthèse et récepteur de chrysobactine	Capture du fer	[221]
<i>pelABCDEL</i>	Pectinases	Dégradation de la pectine des parois cellulaires	[222]
<i>celZ</i>	Cellulase	Dégradation de la cellulase des parois cellulaires	[223]
<i>outC-O</i>	Système de sécretion de type II	Secrétion des pectinases et de la cellulase	[224]
<i>prtABCG</i>	Proteases	Dégradation des parois cellulaires	[225]
<i>prtDEF</i>	Système de sécretion de type I	Secrétion des protéases	[225]

### La phase asymptomatique

La phase asymptomatique se divise en deux parties, l'adhésion de la bactérie à la plante afin de pénétrer dans ses tissus puis la colonisation de l'apoplaste.

*D. dadantii* doit tout d'abord adhérer à la surface des plantes et produit dans ce but des fibrilles de cellulose, qui lui permettent de former des agrégats à la surface de la plante, ainsi que des biosurfactants, qui augmentent la mouillabilité de cette surface [216, 218, 27]. Une fois sur cette surface, *D. dadantii* se déplace à l'aide de ses flagelles vers des points d'entrées dans la plante (stomates, sites de blessures...) qu'elle peut notamment localiser

grâce à des signaux chimiques tels que l'acide jasmonique, produit par les tissus blessés [219]. Une fois dans les tissus, *D. dadantii* utilise les nutriments disponibles (sucres et autres oligosaccharides présents dans l'apoplaste) pour se multiplier [26]. Face à cette attaque, la plante déclenche une réponse de défense en produisant des métabolites secondaires (alcaloïdes, phytoalexines) et des molécules oxydantes, réduisant ainsi la disponibilité en oxygène. Pour survivre à ce stress oxydatif, *D. dadantii* produit des enzymes qui neutralisent les espèces réactives de l'oxygène et réparent les protéines oxydées, tandis que les pompes à efflux permettent l'exportation des substances toxiques [226, 220, 227, 26]. Par ailleurs, la bactérie doit faire face à des modifications de pH de son environnement car la région dans laquelle elle se trouve, l'apoplaste, est un environnement acide. Pour y survivre, la bactérie doit mettre en place des changements métaboliques, tels que l'augmentation de l'activité des transporteurs de protons et la réduction de la production de flagelles. La plante, en se défendant, contribue à l'alcalinisation de l'apoplaste et cela provoque une diminution de la solubilité du fer, déjà peu présent dans l'apoplaste. Le fer étant un élément essentiel pour ses activités, *D. dadantii* répond à cette faible disponibilité en produisant des sidérophores (chrysobactine, achromobactine), qui lui permettent de capter le fer nécessaire à ses réactions enzymatiques [228, 27, 221, 34].

### La phase symptomatique

Lorsque tous les nutriments de l'apoplaste sont consommés et si la densité bactérienne est suffisamment élevée, la phase symptomatique commence avec l'initialisation de la synthèse d'enzymes dégradant les parois cellulaires des plantes (CWDEs). Ce groupe d'enzymes est composé d'une cellulase, de protéases et principalement de pectinases, qui sont à l'origine des symptômes de pourriture molle [229, 223]. *D. dadantii* possède 21 gènes codant pour les pectinases (dont les gènes *pelD* et *pelE*, voir Tableau 2.1), qui vont dégrader la pectine, un polymère de plusieurs polysaccharides présent dans les parois cellulaires des plantes [26, 230]. La pectine assure la cohésion et la rigidité des cellules végétales. Sa dégradation provoque la désorganisation des tissus et leur macération [231, 26].

La production massive et rapide de pectinases (Pels) avant que la plante puisse se défendre est nécessaire pour la réussite de l'infection par *D. dadantii*, l'initiation de l'expression des gènes *pels* représente donc une étape clé de cette infection [232, 73]. Cette expression est finement régulée par un système complexe, comprenant des NAP, dont H-NS, IHF et FIS [72], ainsi que plusieurs régulateurs connus qui répondent à différents signaux environnementaux tels que la présence de pectine (KdgR) [233], la carence en fer (Fur) [234], la température (PecT) [235] et le choc acide (MfbR) [236]. Enfin, lors de la macération des tissus végétaux, les cellules végétales sont totalement détruites, ce qui libère des

composants vacuolaires et cytoplasmiques dans l'apoplaste, exposant ainsi les bactéries à un stress osmotique [26].

### 2.1.3 Le rôle du surenroulement de l'ADN de *Dickeya dadantii* pendant l'infection

Comme détaillé précédemment, *D. dadantii* est confrontée, pendant l'infection, à un stress acide puis un stress oxydatif et enfin un stress osmotique (Fig. 2.2). Pour la réussite du processus infectieux, la bactérie doit détecter et répondre à ces stress grâce à une expression coordonnée de ses gènes. Le SC semble impliqué dans cette coordination car, d'une part ces stress provoquent des modifications du niveau global du SC (voir Fig. 2.2 et la Fig. 1.25 du chapitre précédent) et, d'autre part ces modifications affectent l'expression des gènes [212, 83, 72].

Il a été montré que le relâchement de l'ADN, obtenu en inhibant la gyrase par un antibiotique, la novobiocine, affecte l'expression d'environ 30% des gènes de *D. dadantii* [166]. Parmi ces gènes affectés se trouvent des facteurs clés de la virulence de *D. dadantii* dont *pelE* et *pelD*, *prtB*, et *celZ* qui codent respectivement pour des pectinases, une protéase et une cellulase (Tableau 2.1). *pelE* fait partie des 60 gènes les plus réprimés par ce traitement, que ce soit dans les bactéries en phase exponentielle de croissance ou en phase de transition vers la phase stationnaire [166].

L'impact de la variation opposée de SC, c.-à-d. d'une augmentation du SC, sur l'expression des gènes n'a cependant pas été étudiée jusqu'à présent chez cette bactérie.

## 2.2 L'inhibition de la topoisomérase I de *Dickeya dadantii*

Pour mieux comprendre le mécanisme de régulation de la transcription par le surenroulement, il était nécessaire de compléter les données transcriptomiques de *D. dadantii* [166] par de nouvelles données obtenues cette fois en augmentant le niveau de SC. Cette augmentation de SC peut être provoquée par des stress, comme un stress osmotique [34], cependant une telle perturbation active de nombreuses voies métaboliques de réponses au stress, indépendamment de la régulation par le SC. Pour pouvoir étudier spécifiquement le mécanisme de régulation par le SC, il est préférable d'agir directement sur les topoisomérases. Nous avons donc choisi d'inhiber l'activité de relâchement de la topoI de *D. dadantii* à l'aide d'un antibiotique, en réalisant un traitement de durée suffisamment courte et avec une concentration d'antibiotique suffisamment faible pour affecter le SC sans impacter la croissance bactérienne, et en limitant l'impact sur les voies métaboliques indépendantes du SC.

### 2.2.1 Le choix d'un antibiotique inhibiteur de la topoisomérase I

L'inhibition de la topoI bactérienne est un axe de recherche pour la lutte des maladies infectieuses humaines [237, 238]. Des recherches sont menées pour identifier des antibiotiques ciblant la topoI bactérienne sans affecter les topoisomérases humaines. Ces recherches ont permis d'identifier de potentiels inhibiteurs de topoI bactériennes qui pourraient être utilisés pour inhiber la topoI de *D. dadantii* (Tableau 2.2).

#### Les méthodes d'identification de nouveaux inhibiteurs de la topoI

Une méthode fréquemment utilisée pour identifier des inhibiteurs d'une cible dont la structure est connue est le criblage *in silico* de banques chimiques. Les composés naturels ou synthétiques de la banque sont testés par *docking* ou par dynamique moléculaire pour vérifier que l'interaction entre ce composé et la cible est possible. Les structures cristallographiques des topoI de *E. coli* et *M. tuberculosis* étant connues, cette méthode a permis de mettre en avant plusieurs inhibiteurs potentiels de ces topoI [239, 240, 241].

L'inhibition de l'activité de la topoI peut ensuite être testée avec des essais biochimiques *in vitro*. La méthode standard est l'observation des différents topoisomères de plasmides, traités avec la topoI ciblée et l'inhibiteur, séparés par électrophorèse sur gel d'agarose [242, 237]. Comme présenté précédemment (Chapitre 1), plus le plasmide est surenroulé, plus il migre rapidement dans le gel d'agarose. Un plasmide traité avec la topoI est plus relâché et migre donc moins loin que le plasmide surenroulé. Si un inhibiteur de topoI est ajouté lors de ce traitement, l'activité de relâchement de la topoI sera limitée donc ce plasmide migrera plus loin que le plasmide relâché par une topoI active [243].

Une fois l'inhibition de la topoI confirmée, il est utile de caractériser les propriétés de cette inhibition *in vitro* en déterminant la concentration d'antibiotique nécessaire pour réduire l'activité enzymatique de la topoI de 50% (IC<sub>50</sub>, concentration inhibitrice médiane) [244, 239, 242].

### Le choix de la séconéolitsine en inhibiteur de la topoI

Pendant les 15 dernières années, ces méthodes ont permis d'identifier une quinzaine d'antibiotiques inhibiteurs des topoI de *E. coli* et *M. tuberculosis* (Tableau 2.2).

TABLEAU 2.2 – IC<sub>50</sub> de molécules inhibitrices des topoI de *E. coli* et *M. tuberculosis*. Tableau issu de [237].

Molécule	IC <sub>50</sub> topoI de <i>E. coli</i>	IC <sub>50</sub> topoI de <i>M. tuberculosis</i>	Source
Bis-benzimidazole DPA 154	6,6 µM	Non déterminée	[245]
Bis-benzimidazole PPEF	9,4 µM	Non déterminée	[246]
Imipramine	Pas d'inhibition à 25 µM	< 0,1 µM	[240]
Norclomipramine	Pas d'inhibition à 25 µM	< 0,1 µM	[240]
Polyamine 2471-12	7,5 µM	7,5 µM	[247]
Polyamine 2471-24	10 µM	7,5 µM	[247]
Gold(III) macrocycle 10	5 µM	10 µM	[244]
Gold(III) chelate 14	1,3 µM	1,3 µM	[244]
Fluoroquinophenoxazine 11a	0,48 µM	0,98 µM	[248]
Fluoroquinophenoxazine 11g	0,48 µM	0,24 µM	[248, 249]
VCC891909	Non déterminée	< 7,5 µM	[241]
Seconeolitsine	Non déterminée	5,6 µM	[242]
N-méthyl-seconeolitsine	Non déterminée	8,4 µM	[242]
Piperidine amide 7	15,6 – 31,3 µM	2 µM	[239]
NSC76027	2,2 µM	4 µM	[250]

Parmi les antibiotiques identifiés, la séconéolitsine nous a paru être l'antibiotique adéquat pour notre étude transcriptomique de la réponse bactérienne à l'inhibition de la topoI chez *D. dadantii*. Ce choix s'explique d'une part par la faisabilité de la synthèse de la séconéolitsine par un procédé publié [124]. Cette synthèse s'effectue en quelques étapes (Fig. 2.3A) à partir de boldine, un alcaloïde commercialisé issu de la feuille de l'arbre boldo. Nos collaborateurs de l'Institut de Chimie et Biochimie Moléculaires et Supramoléculaires (ICBMS) ont reproduit cette synthèse (Fig. 2.3).

D'autre part, le choix de cet antibiotique a été conforté par les résultats d'une étude similaire effectuée chez *Streptococcus pneumoniae* [175]. Cette étude a montré que, chez cette

bactérie à Gram positif, le traitement par la séconéolitsine augmente le niveau global de SC et affecte l'expression de 10% de ses gènes.

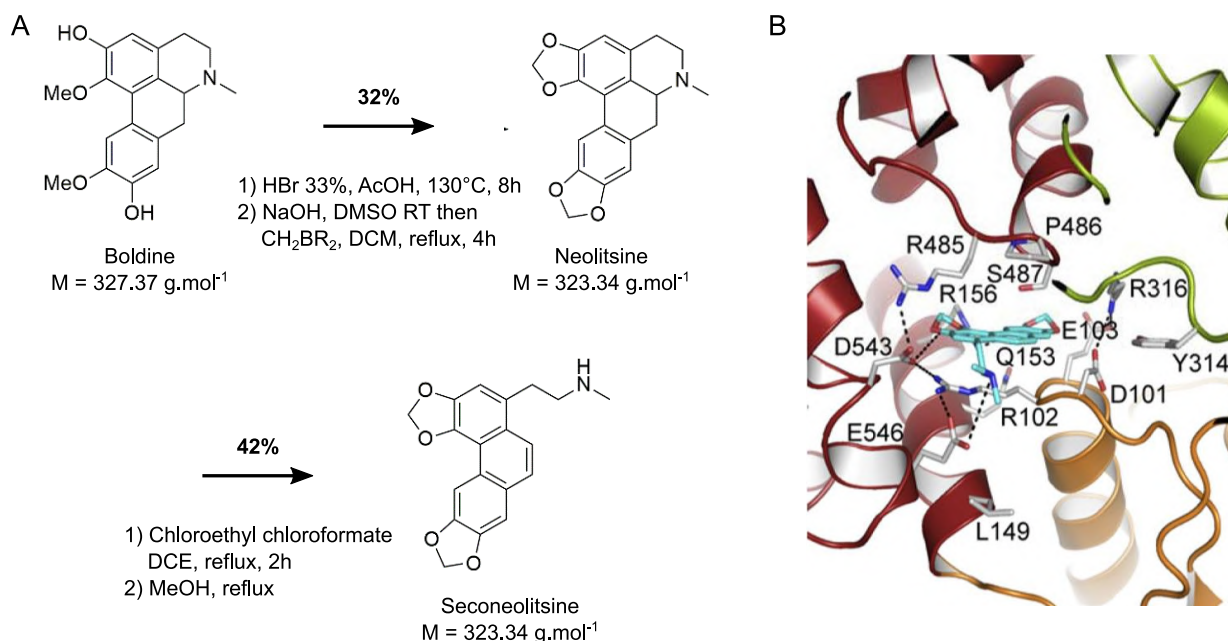


FIGURE 2.3 – (A) Synthèse de la séconéolitsine à partir de la boldine : étapes et rendements obtenus par nos collaborateurs de l'ICBMS. (B) interactions entre le site de liaison aux nucléotides de la topoI et la N-méthyl-séconéolitsine. Figure issue de [251]

Enfin, d'après le docking de la séconéolitsine sur la topoI de *S. pneumoniae*, la séconéolitsine interagit fortement avec le site de liaison aux nucléotides de la topoI [251]. Dans ce site de liaison, la séconéolitsine se place dans la même orientation et entre les mêmes positions que la N-méthyl-séconéolitsine dont l'interaction est présentée en Fig. 2.3B. Cette interaction empêche le changement de conformation de la topoI et inhibe ainsi son activité de relâchement de l'ADN. Les résidus de la topoI de *S. pneumoniae* en interaction avec la séconéolitsine sont très conservés dans les topoI de *E. coli* et de *D. dadantii*. Ainsi, il est fortement probable que les topoI de ces deux dernières bactéries puissent également être inhibées par la séconéolitsine.

### 2.2.2 La vérification de l'efficacité de la séconéolitsine sur la topoI de *Dickeya dadantii* *in vitro*

Pour vérifier l'effet inhibiteur de la séconéolitsine sur la topoI d'une bactérie à Gram négatif, et plus particulièrement sur celle de *D. dadantii*, nous avons souhaité réaliser des essais *in vitro* de relâchement de plasmides surenroulés. Certaines topoisomérases purifiées, comme la topoisomérases IV, la gyrase et la topoisomérase I de *E. coli*, sont commercialisées. Ce n'est pas le cas de celles de *D. dadantii* dont l'étude est trop peu

répandue pour justifier cette commercialisation. Pour pouvoir vérifier l'efficacité de la séconéolitsine sur la topoI de *D. dadantii in vitro*, j'ai donc dû procéder à sa surproduction puis à sa purification (Fig. 2.4A).

### La production de la topoI de *D. dadantii*

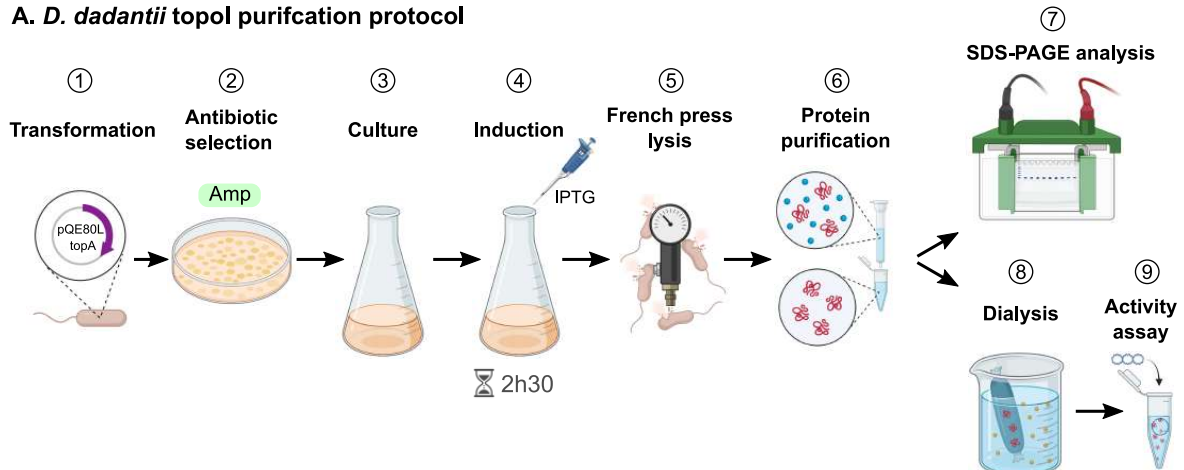
Pour surexprimer la topoI de *D. dadantii* j'ai utilisé le vecteur de surexpression pQE80L (Fig. 2.4B). Ce plasmide, outre ses gènes de résistance à des antibiotiques, porte le répresseur *lacI* ainsi que l'opérateur *lacO*. Cet opérateur est situé au niveau du promoteur T5, en amont du site multiple de clonage. En absence d'inducteur analogue au lactose, comme l'IPTG, le répresseur codé par *lacI* forme un complexe avec l'opérateur *lacO* et empêche la transcription des gènes dépendant du promoteur T5. En présence d'IPTG, le répresseur se lie préférentiellement à l'IPTG, ce qui change la conformation du répresseur et empêche sa liaison avec l'opérateur, rendant le promoteur T5 de nouveau accessible pour l'ARN polymérase [252, 253]. Ainsi, une induction à l'IPTG est nécessaire pour activer l'expression du gène inséré à la suite du promoteur T5 dans ce plasmide. Par ailleurs, ce vecteur permet d'ajouter une queue de six histidines (6xHis) à l'extrémité N-terminale de la protéine produite, ce qui permettra la purification de la protéine produite avec une résine sélective.

J'ai effectué l'insertion du gène *topA* de *D. dadantii* dans ce plasmide par clonage TEDA [254]. Pour cela, j'ai indépendamment amplifié par PCR le plasmide pQE80L et le gène *topA* de manière à ce que chacun de ces fragments porte une séquence complémentaire permettant leur appariement. Ces fragments peuvent, en présence de T5 exonucléase, s'hybrider et sont reliés *in vivo* après transformation de ce plasmide dans les bactéries.

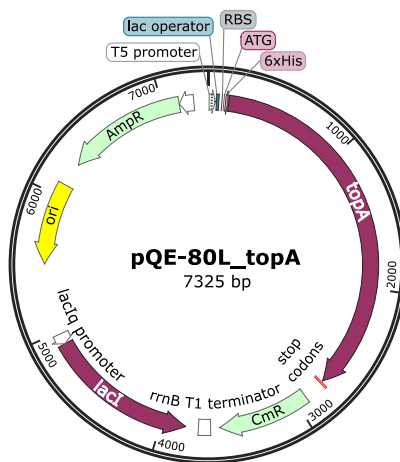
Pour avoir une surexpression importante de la topoI de *D. dadantii*, nous avons choisi de réaliser cette production dans la souche *E. coli* NM522. J'ai donc transformé ce plasmide contenant *topA* dans la souche *E. coli* NM522 (Fig. 2.4A étape 1) par électroporation. J'ai ensuite sélectionné les bactéries ayant incorporé le plasmide en les cultivant sur un milieu gélosé sélectif contenant de l'ampicilline (Fig. 2.4A étape 2). La production de topoI est ensuite réalisée en milieu riche liquide, supplémenté de 100 µg/mL d'ampicilline, avec une induction à 1mM d'IPTG effectuée pendant 2h30 lorsque la culture atteint une  $DO_{600nm}$  de 0,6 (Fig. 2.4A étapes 3 et 4). Les bactéries sont ensuite lysées par lyse enzymatique et mécanique à l'aide de lysozyme et d'une presse de French (Fig. 2.4A étape 4).

### La purification de la topoI produite

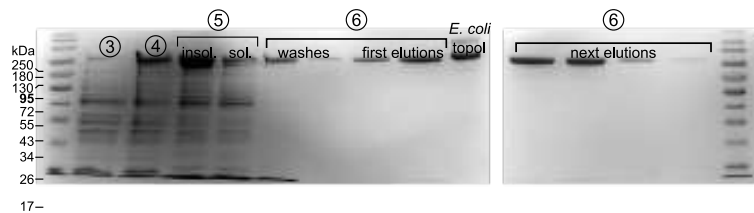
La purification de la topoI produite est facilitée par la présence de la queue 6xHis apportée par le plasmide pQE80L lors de l'expression du gène inséré. En effet, cette queue permet de réaliser une purification par affinité avec une résine Nickel-NTA (acide nitrilotriacétique).

A. *D. dadantii* *topoI* purification protocol

## B. Plasmid construction (for step ①)



## C. SDS-PAGE analysis (step ⑦)



## D. Activity assay (step ⑨)

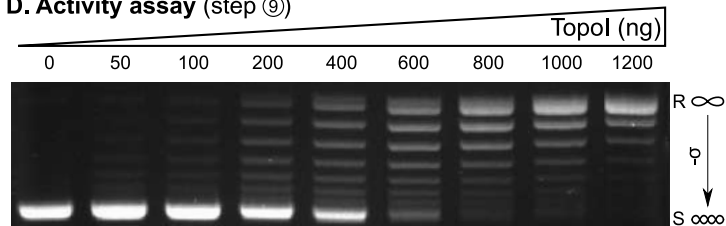


FIGURE 2.4 – (A) Protocole de la surproduction et de la purification de la *topoI* de *D. dadantii*, schématisé avec BioRender. (1) Transformation du plasmide porteur du gène *topA* de *D. dadantii*, détaillé en B, dans la souche *E. coli* NM522. (2) Sélection des souches porteuses du plasmide grâce à la résistance à l'ampicilline apportée par le plasmide. (3) Culture des souches en milieu riche LB jusqu'à  $DO_{600nm}$  0,6. (4) Induction de la production de la *topoI* par ajout de 1 mM d'IPTG. (5) Lyse des cellules à la presse de French, dans un tampon contenant de la lysozyme. (6) Purification des protéines (étape non détaillée sur le schéma). (7) Vérification de la qualité de la purification sur gel SDS-Page, présentée en C. (8) Dialyse des éluions les plus pures et concentrées. (9) Vérification de l'activité de la *topoI* purifiée, présentée en D. (B) Construction du plasmide pQE80L avec l'insertion du gène *topA*. Ce plasmide porte des résistances à l'ampicilline (AmpR) et au chloramphénicol (CmR) ainsi que l'opérateur et le répresseur lac, qui permettent de conditionner l'expression du gène inséré à la présence d'IPTG. (C) Gel SDS-PAGE permettant l'analyse des protéines présentes dans la culture avant (3) et après (4) l'induction de la surexpression de la *topoI* par l'ajout d'IPTG, des protéines insolubles (notées insol.) et solubles (notées sol.) après la lyse (5) et des protéines présentes dans les fractions de lavage puis d'éluion lors de la purification de la protéine (6). Ce gel valide la surexpression de la *topoI*, sa pureté après purification et sa concentration importante dans les fractions 2, 3 et 4 de l'éluion. (D) Vérification de l'activité de la *topoI* de *D. dadantii* purifiée. 0,5  $\mu$ g de plasmides pUC18 surenroulés ont été incubés avec la masse indiquée de *topoI* pendant 15 minutes à 27°C. R et S indiquent respectivement les formes relâchées et surenroulées du plasmide.



Le lysat obtenu à l'étape précédente, riche en topoI marquées par poly-histidines, est mis au contact de cette résine avant d'être purifié sur une colonne de polypropylène. Deux lavages sont effectués pour éliminer les impuretés puis la topoI pure est éluée à l'aide d'imidazole dont l'affinité pour la résine est supérieure à celle de la topoI marquée (Fig. 2.4A étape 5). L'analyse du profil protéique des éluions sur gel SDS-PAGE (Fig. 2.4C), confirme la pureté de ces éluions et la forte concentration en topoI dans les fractions d'éluion 2, 3 et 4. Les fractions 3 et 4 sont dialysées dans un tampon contenant (1) du Tris-HCl et du NaCl pour respectivement équilibrer le pH et la concentration en sels afin de garantir la solubilité et la conservation de l'activité de la protéine, (2) de l'EDTA pour inhiber les éventuelles protéases, (3) du DTT pour protéger de l'oxydation et (4) du glycérol pour permettre une cryoconservation de la protéine à  $-20^{\circ}\text{C}$ . Deux dialyses sont réalisées successivement dans de grands volumes de tampon afin de minimiser la concentration résiduelle des composants présents dans le tampon d'éluion, dont l'imidazole. L'ensemble des tampons utilisés est décrit dans notre article [255]. Les tampons de lyse et d'éluion sont issus de [256]. La réalisation d'un dosage de Bradford de la solution de topoI m'a ensuite permis d'estimer sa concentration à 5 mg/mL.

### La vérification de l'activité de la topoI et de son inhibition par la séconéolitsine

L'activité de la topoI a été contrôlée *in vitro* à l'aide de plasmides pUC18 commercialisés sous forme surenroulée. Ces plasmides ont été traités avec des quantités croissantes de topoI puis déposés sur un gel d'agarose (Fig. 2.4D). Soumises à un voltage moyen (70V pendant 3,5 h), les topoisomères de ce plasmide migrent selon leur niveau de surenroulement, permettant ainsi de vérifier que, plus la quantité de topoI utilisée dans le traitement du plasmide est élevée, plus le plasmide est relâché. J'ai également appliqué ce protocole à la topoI commerciale de *E. coli* afin de vérifier son activité et d'évaluer la quantité d'enzymes nécessaire au relâchement d'environ 50% des topoisomères des plasmides (IC50). J'ai ainsi confirmé que les topoI de *D. dadantii* et de *E. coli* étaient bien actives, avec des IC50 de 4,6  $\mu\text{M}$  pour *D. dadantii* et 6,7  $\mu\text{M}$  pour *E. coli*.

L'inhibition de la topoI par la séconéolitsine a ensuite pu être observée *in vitro* en traitant les plasmides surenroulés avec la quantité de topoI permettant ce relâchement de 50% puis en y ajoutant différentes quantités de séconéolitsine (0 à 20  $\mu\text{M}$ ). La migration de ces plasmides sur gel d'agarose a confirmé qu'en présence de séconéolitsine, la topoI semble inhibée. En effet, la quantité croissante de séconéolitsine (jusqu'à 7,5  $\mu\text{M}$  chez *E. coli* et jusqu'à 15  $\mu\text{M}$  chez *D. dadantii*) permet d'obtenir une augmentation progressive de la proportion de topoisomères sous forme surenroulée [255].

## 2.3 Article

La confirmation de l'inhibition de la topoI de *D. dadantii* par la séconéolitsine permet d'approfondir l'étude du lien entre SC et transcription. En effet, la séconéolitsine, en inhibant la topoI, permet d'augmenter le niveau global du SC. J'ai donc pu analyser le transcriptome de *D. dadantii* face à cette modification de SC en réalisant une étude RNA-Seq. Les résultats sont présentés dans l'article suivant publiés dans Nucleic Acids Research. Les tableaux supplémentaires sont disponibles en ligne [255].

Dans cet article, nous avons démontré *in vitro* que la séconéolitsine inhibe l'activité de la topoI de *D. dadantii* et que cet antibiotique provoque une augmentation du niveau de SC *in vivo*. Nous avons obtenu le premier transcriptome d'une bactérie à Gram-négatif à l'inhibition de la topoI. Ce transcriptome a été obtenu par RNA-Seq en traitant des cultures de *D. dadantii*, en phase exponentielle de croissance ou en phase de transition vers la phase stationnaire, avec une concentration sub létale de séconéolitsine. Environ 10% des gènes ont été affectés dans chacune des phases de croissance mais peu d'entre eux répondent dans les deux phases. La réponse d'un gène dépend donc de la phase de croissance mais également du sens de la variation de SC puisque les gènes différentiellement exprimés après le choc à la séconéolitsine sont différents de ceux affectés par le choc à la novobiocine, inhibiteur de gyrase. Nous avons ensuite caractérisé les gènes différentiellement exprimés selon leur niveau d'expression, leur localisation et leur orientation. Nous avons alors pu conclure que la notion de "gène sensible au surenroulement" est une caractéristique complexe puisque cette sensibilité dépend de l'activité locale et dynamique des topoisomérases, de l'état métabolique des cellules et du contexte génomique. Dans cet article, nous proposons un modèle qualitatif de l'expression d'un gène selon le niveau de SC, basé sur des données *in vitro* d'expression de promoteurs. Ce modèle est cohérent avec les observations obtenues après les inhibitions de la gyrase et de la topoI.

Published online 11 August 2022

Nucleic Acids Research, 2022, Vol. 50, No. 16 9149–9161  
<https://doi.org/10.1093/nar/gkac679>

# What is a supercoiling-sensitive gene? Insights from topoisomerase I inhibition in the Gram-negative bacterium *Dickeya dadantii*

Maiwenn Pineau<sup>1</sup>, Shiny Martis B.<sup>1</sup>, Raphaël Forquet<sup>1</sup>, Jessica Baude<sup>1</sup>, Camille Villard<sup>1</sup>, Lucie Grand<sup>2</sup>, Florence Popowycz<sup>2</sup>, Laurent Soullère<sup>2</sup>, Florence Hommais<sup>1</sup>, William Nasser<sup>1</sup>, Sylvie Reverchon<sup>1</sup> and Sam Meyer<sup>1,\*</sup>

<sup>1</sup>Université de Lyon, INSA Lyon, Université Claude Bernard Lyon 1, CNRS UMR5240, Laboratoire de Microbiologie, Adaptation et Pathogénie, 69621 Villeurbanne, France and <sup>2</sup>Université de Lyon, INSA Lyon, Université Claude Bernard Lyon 1, CPE Lyon, CNRS UMR 5246, Institut de Chimie et Biochimie Moléculaires et Supramoléculaires, 69622 Villeurbanne, France

Received April 09, 2021; Revised July 04, 2022; Editorial Decision July 15, 2022; Accepted July 27, 2022

## ABSTRACT

DNA supercoiling is an essential mechanism of bacterial chromosome compaction, whose level is mainly regulated by topoisomerase I and DNA gyrase. Inhibiting either of these enzymes with antibiotics leads to global supercoiling modifications and subsequent changes in global gene expression. In previous studies, genes responding to DNA relaxation induced by DNA gyrase inhibition were categorised as ‘supercoiling-sensitive’. Here, we studied the opposite variation of DNA supercoiling in the phytopathogen *Dickeya dadantii* using the non-marketed antibiotic seconeolitsine. We showed that the drug is active against topoisomerase I from this species, and analysed the first transcriptomic response of a Gram-negative bacterium to topoisomerase I inhibition. We find that the responding genes essentially differ from those observed after DNA relaxation, and further depend on the growth phase. We characterised these genes at the functional level, and also detected distinct patterns in terms of expression level, spatial and orientational organisation along the chromosome. Altogether, these results highlight that the supercoiling-sensitivity is a complex feature, which depends on the action of specific topoisomerases, on the physiological conditions, and on their genomic context. Based on previous *in vitro* expression data of several promoters, we propose a qualitative model of SC-dependent regulation that accounts for many of the contrasting transcriptomic features observed after DNA gyrase or topoisomerase I inhibition.

## INTRODUCTION

DNA supercoiling (SC) is the product of torsional stress ubiquitously experienced by the double-helix in all kingdoms of life. In bacteria, the chromosome is maintained in a steady-state level of negative SC by the interplay of nucleoid associated proteins (NAPs) and the activity of topoisomerases. The DNA gyrase (belonging to class II topoisomerases) introduces negative supercoils by ATP-dependent double-strand cleavage, whereas topoisomerase I (topoI, class IA) removes excessive negative supercoils through ATP-independent single-strand cleavage, and topoisomerase IV (topoIV, class II) through ATP-dependent double-strand cleavage (1–3). The activity of these topoisomerases is finely controlled by cells according to a homeostasis mechanism (4), and this balance plays a key role in many cellular functions, and in particular in the expression of the genome, which is our focus in this study.

The presence of torsional stress in the DNA template is known to affect the transcription process at several successive steps: by modulating the binding of transcriptional regulators and RNA Polymerase (RNAP) itself, the formation and stability of the open complex (5), promoter clearance (6), elongation and termination (7,8). As a result, SC acts as a global transcriptional regulator (9,10), although the precise underlying mechanisms remain controversial. Early studies demonstrated a strong regulatory action of SC on the promoters of stable RNAs in *Salmonella enterica* and *Escherichia coli* (5,11), pointing to a role in growth control (10) consistent with the close relationship between SC and the cellular metabolism (12). But other promoters were found to be equally affected (7,13), which was then confirmed and broadened by high-throughput transcriptomic methods (14–16). In analogy to the ‘regulons’ of transcriptional factors, these promoters were of-

\*To whom correspondence should be addressed. Tel: +33 4 72 43 85 16; Email: sam.meyer@insa-lyon.fr

ten termed ‘supercoiling-sensitive’, although that notion remains poorly defined, considering the lack of clearly identified sequence determinants (17,18), and the variability in the response of many promoters to SC alterations depending on their context and the experimental protocol of the assay. For example, the *lacP* promoter of *E. coli* is strongly repressed by DNA relaxation *in vitro* (7), but is unaffected *in vivo* (14,15); the proportion of genes activated by DNA relaxation in *S. enterica* varied between 70% in a random fusion assay (19) and 27% in a RNA-Seq transcriptome (20).

*In vivo*, these responses to SC variations were obtained by two distinct methods (21). The expression level can be measured in topoisomerase mutant strains, which usually exhibit a different SC level than the parental strain (21,22); however, the difference in promoters’ expression then reflects not only the direct regulatory effect of SC, but also that of the resulting global change in transcriptional regulatory activity in the mutant strain, and these two contributions are difficult to distinguish. To avoid this issue, it is often preferred to use a wild-type strain, and induce a rapid SC variation by applying topoisomerase inhibiting antibiotics (8,21). Commonly used drugs belong to the coumarin family (coumermycin, novobiocin), inhibiting the ATPase activity of DNA gyrase (and topoIV), and the quinolone family (norfloxacin, ciprofloxacin, oxolinic acid) inhibiting the ligase activity of DNA gyrase and topoIV (1,3,23). These drugs induce a sudden DNA relaxation, whose effect on gene expression can then be measured. The main shortcoming is that they also trigger SC-independent stress-response pathways in the cell. In order to characterise specifically the effect of SC on transcriptional regulation, it is thus desirable to compare the expression patterns obtained with different methods (15). In this respect, a major limitation of existing studies is that, since DNA gyrase is the primary target of all these drugs in Gram-negative bacteria, the transcriptomic response was analysed only in one direction, DNA relaxation, introducing a strong bias in the analysis of the SC-sensitivity of promoters.

The opposite variation could also be induced, but only by applying quinolones on engineered strains harbouring mutations in a gyrase gene, where only the relaxing activity of topoIV is inhibited by the drug (2,24). In wild-type cells, topoI seemed a particularly suitable drug target (25), both in clinical research as it is the only enzyme of type IA topoisomerases family in many pathogenic species, but also as a way to study the effect of SC in transcriptional regulation, since this enzyme plays a direct role in the handling of torsional stress associated with transcription (26), while topoIV is predominantly involved in replication (27). In its catalytic cycle, topoI binds a stretch of single-stranded DNA, cleaves it and undergoes a conformational change to an open conformation, allowing the complementary DNA strand to pass the gate, followed by the religation of the DNA backbone with a gain of one linking number (28–31). In recent years, many compounds were shown to act as topoI inhibitors with unequal effectiveness as antimicrobial agents (25). In particular, one of them named seconeolitsine was shown to be effective against *Streptococcus pneumoniae* and *Mycobacterium tuberculosis* topoI, presumably by interacting with its nucleotide binding site, preventing the topoI conformational change and thus inhibiting DNA

binding (31). When applied *in vivo* at low concentration, this drug induces a transient increase in negative SC associated with a global change in the transcriptional landscape (32).

Here, we show this drug to be equally effective in Gram-negative bacteria, and we use it to report the first transcriptomic response to topoI inhibition and resulting increase in negative SC in Gram-negative bacteria, using the phytopathogen *Dickeya dadantii* as a model. The latter contains the same set of topoisomerases as *E. coli* with a strong sequence homology, and generally, has a strong proximity to the enterobacterial models *E. coli* and *S. enterica*. Interestingly, SC was shown to be an important regulator of its key virulence genes (16,33), and SC-affecting environmental signals are influential in its infection process, in particular osmolarity variations resulting in an increase in negative SC (16,33). Deciphering the mechanisms of SC-related transcriptional regulation in that species is thus important for our understanding of the mechanisms of virulence, as well as transcription as a general process.

In the following, we first demonstrate the inhibitory effect of seconeolitsine on *D. dadantii* (as well as *E. coli*) topoisomerase I, and its antibacterial action against that species. We then show that a seconeolitsine shock at low concentration quickly increases the cellular negative SC level. We analyse the effect of this shock on the expression of the genome, and in particular, we illustrate the relationship between gene expression strength and spatial gene organisation and the response to topoI inhibition by seconeolitsine. By comparing this response with that of the DNA gyrase inhibitor novobiocin, we propose a qualitative model explaining many notable features possibly involved in defining the supercoiling-sensitive property of promoters.

## MATERIALS AND METHODS

### Seconeolitsine synthesis

Seconeolitsine was synthesised in 13% yield starting from boldine, following the protocol described in the original patent (34). In the first step, a reaction of demethylation was conducted in acidic conditions followed by a reaction with dibromomethane. The intermediate neolitsine was then reacted with chloroethyl chloroformate in dichloroethane followed by aromatization and ring opening achieved in refluxing methanol. The product was characterised by high resolution electrospray ionization mass spectrometry (ESI-HRMS) ([M + H]<sup>+</sup>: computed for C<sub>19</sub>H<sub>18</sub>NO<sub>4</sub>: 324.1230; found 324.1220) and its purity was validated by nuclear magnetic resonance (NMR) (Supplementary Figure S1).

### Protein expression and purification

*D. dadantii* 3937 *topA* gene was amplified and cloned into pQE80L plasmid using the TEDA method (35) to over-produce N-terminally 6xhis-tagged topoisomerase I. *E. coli* NM522 carrying the expression plasmid were grown at 37°C in LB medium until OD<sub>600nm</sub> reached 0.6. Protein expression was then induced by adjusting the final concentration of the culture at 1 mM IPTG. After 2.5 h of induction, the cells were harvested by centrifugation, resuspended in a cold lysis buffer (20 mM NaH<sub>2</sub>PO<sub>4</sub>, 0.5 M NaCl, 20 mM imidazole, 2.5 mM TCEP (tris(2-carboxyethyl)phosphine),

1 mg ml<sup>-1</sup> lysozyme, pH 7.4) (36) and disrupted through a French pressure cell press. After clarification of the obtained lysate by a 15 min centrifugation at 15 000 rpm, the supernatants were mixed with Sigma HIS-Select Nickel Affinity Gel (at a ratio of 3:1) equilibrated in lysis buffer before being added into a polypropylene column (Qiagen). After extensive washing with a cold lysis buffer, the bound topoI was eluted with a cold elution buffer (20 mM, NaH<sub>2</sub>PO<sub>4</sub>, 0.5 M NaCl, 500 mM imidazole, 2.5 mM TCEP, pH 7.4) (36). Dialysis desalination was performed overnight with a first dialysis buffer (50 mM Tris-HCl, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, pH 7.5) and 6 h with a storage buffer (50 mM Tris-HCl, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 50% glycerol, 0.1% Triton X-100, pH 7.5). The purity of topoI was assessed by SDS-PAGE and the concentration of the purified samples was measured with the Bradford protein assay (37). Comparisons with *E. coli* topoI were made with a commercial topoI (NEB).

#### ***In vitro* analysis of topoisomerase I, topoisomerase IV and DNA gyrase inhibition by seconeolitsine**

*D. dadantii* topoI concentration required to relax 50% of pUC18 topoisomers was determined after 15 min of incubation at 37°C in rCutSmart Buffer (NEB). For the inhibition assays, topoI was first preincubated with seconeolitsine and rCutSmart Buffer at 4°C for 10 min. This mix was then incubated with pUC18 at 37°C for 15 min. All reaction products were analysed by electrophoresis on 1.2% agarose gel at 70 V for 3.5 h. The IC<sub>50</sub> was defined as the concentration that reduces topoI relaxing activity by 50% (using the three most migrated bands together as a marker of the most negatively supercoiled topoisomers). Topoisomerase IV and DNA gyrase inhibition by seconeolitsine were assessed with the Inspiralis *E. coli* Topoisomerase IV Relaxation Kit and *E. coli* Gyrase Supercoiling Assay Kit, following manufacturer's instructions.

#### **Minimum inhibitory concentration (MIC) and survival rate in solid medium**

LB Agar plates containing seconeolitsine dissolved in DMSO (50 mM stock solution) and IPTG (100 mM stock solution) were prepared to have seconeolitsine final concentrations between 0 and 750 μM and IPTG final concentrations of 0 or 0.1 mM. *D. dadantii* 3937, *E. coli* NM522 and *E. coli* NM522 carrying pQE80L::topA plasmids were grown at 30°C (*D. dadantii*) or 37°C (*E. coli*) until OD<sub>600 nm</sub> = 0.3. Cultures were then serial-diluted and placed on prepared plates. After 20 h of incubation at 30°C, colonies were counted. The survival rate was calculated as the ratio between the number of colonies observed on plates with or without seconeolitsine. The MIC was defined as the lowest seconeolitsine concentration without visible growth on the LB plates.

#### **Seconeolitsine inhibitory action in liquid cultures**

*D. dadantii* 3937 were grown at 30°C in microplates containing Luria-Broth medium and increasing concentrations of

seconeolitsine dissolved in DMSO (5 or 10 mM stock solution, keeping the final volume of DMSO below 4%). Optical densities were recorded every 5 min using an automatic microplate reader (Tecan Spark), and growth curves were fitted to a Gompertz equation to estimate growth rates and time lags (38).

#### **Bacterial cultures for seconeolitsine shock**

*D. dadantii* 3937 were grown at 30°C in M63 supplemented with sucrose at 0.2% (wt/vol) until the exponential (OD<sub>600nm</sub> = 0.2) or transition to stationary phases (OD<sub>600nm</sub> = 1.1). Cells were then shocked with seconeolitsine dissolved in DMSO at 50 μM during 5 min (RT-qPCR experiments) or 15 min (RT-qPCR and RNA-Seq experiments). An additional control was performed with pure DMSO for RT-qPCR experiments.

#### **Topoisomer separation in chloroquine-agarose gels**

The topoisomer distribution was analysed as previously described (39). Reporter plasmids pUC18 were transformed into *D. dadantii* 3937. Fifteen minutes after the shock, plasmids were extracted with the Qiaprep Spin Miniprep kit and migrated on a 1% agarose gel containing 2.5 μg ml<sup>-1</sup> chloroquine at 2.5 V cm<sup>-1</sup> for 16 h. Under these conditions, more negatively supercoiled migrate faster in the gel. Chloroquine gels were subjected to densitometric analysis using Image Lab 6.0 software (Biorad). Distributions of topoisomers were normalised and quantified in each lane independently.

#### **RNA extraction**

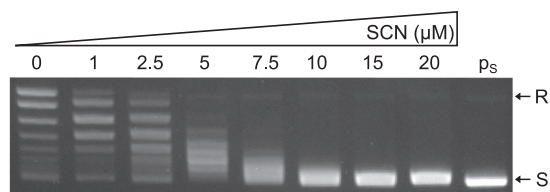
Total RNAs were extracted either with the frozen-phenol method (40) (RNA-Seq experiments) or with the Qiagen RNeasy Plus Mini Kit, including a bacterial lyse with a lysozyme solution at 1 mg ml<sup>-1</sup> and the optional DNase treatment (RT-qPCR experiments). The absence of genomic DNA contamination was further verified by PCR amplification with the Lucigen EconoTaq PLUS GREEN and *ryhB* primers (Tab. S1), following manufacturer's instructions. When necessary, an additional DNase treatment was performed using the NEB DNase I to ensure RNA purity. Extracted RNAs were quantified using a ND-1000 NanoDrop spectrophotometer. RNA quality was checked by agarose gel electrophoresis.

#### **Quantitative real time PCR**

1 μg of total RNAs were reverse transcribed using the Thermo Scientific RevertAid First Strand cDNA Synthesis Kit. Reaction mixes were incubated at 25°C for 5 min, 42°C for 60 min and 70°C for 5 min.

The quantitative PCR was carried out using the Thermo Scientific Maxima SYBR Green/ROX qPCR Master Mix with the LC480 Lightcycler from Roche and the primers listed in Tab. S1. The following thermal cycling reactions were executed: (i) an initial denaturation step at 95°C for 10 min, (ii) 45 amplification cycles at 95°C for 15 s, 58°C for 30 s and 72°C for 40 s. The housekeeping gene *rpoA* was used as a normalizer for the gene expression ratios. The uniqueness of the amplification product is verified with the melting curve.

9152 *Nucleic Acids Research*, 2022, Vol. 50, No. 16



**Figure 1.** Inhibition of topoI *in vitro* relaxing activity by seconeolitsine. The specified amount of seconeolitsine was pre-incubated for 10 min with 800 ng topoisomerase I, then 0.5  $\mu$ g of supercoiled pUC18 plasmids (p) was added and incubated for 15 min. The IC<sub>50</sub> was estimated at a value of 4  $\mu$ M (see Materials and Methods).

### RNA sequencing

All samples were collected in two biological replicates (eight samples in total). Steps of ribosomal RNA depletion, cDNA library preparation and high-throughput sequencing were carried by the MGX Montpellier GenomiX platform, using the Illumina TruSeq stranded mRNA sample preparation kit and HiSeq2500 sequencing providing 50-nt single-end reads. The sequenced reads were deposited in ArrayExpress under accession number E-MTAB-10134. They were mapped on the reference genome of *D. dadantii* 3937 (NCBI NC\_014500.1) with Bowtie2 and counted with htseq-count. Gene differential expression analysis was performed with DESeq2 with a threshold of 0.05 on the adjusted *P*-value.

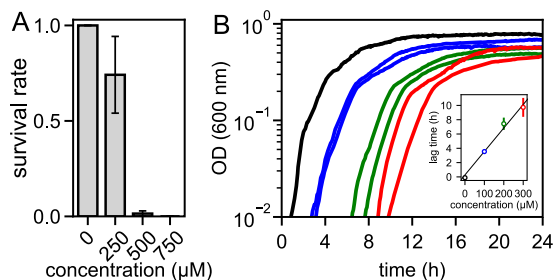
### Statistics and data analysis

All statistical analyses and graphs were made with a home-made Python code. Error bars are 95% confidence intervals. Proportions of activated genes among differentially expressed genes were compared with  $\chi^2$ -tests. Stars indicate the level of significance based on the *P*-value ( $^{***}P < 0.001$ ;  $^{**}0.001 < P < 0.01$ ;  $^{*}0.01 < P < 0.05$ ). The orientation of a gene is defined relative to the orientation of its neighbours (either convergent, divergent or tandem). Functional enrichment was analysed using the Gene Ontology classification (41). Only functions corresponding to at least four *D. dadantii* genes were considered. Chromosomal domains were previously defined in (16).

## RESULTS

### Seconeolitsine inhibits *D. dadantii* topoisomerase I *in vitro*

The comparison of *topA* sequences from enterobacteria *D. dadantii* and *E. coli* with those of *M. tuberculosis* and *S. pneumoniae* showed that the topoI residues bound by seconeolitsine were mostly conserved in the former (Supplementary Figure S2), suggesting that the inhibitory activity of the drug might be also effective in enterobacteria. To test this hypothesis, we synthesised seconeolitsine following the protocol described in the original patent (34), and the purity of the product was validated by NMR (Supplementary Figure S1). Its inhibitory activity against *D. dadantii* topoI was evaluated by adding increasing concentrations of the drug to a solution of purified enzymes, resulting in a progressive reduction of their relaxing activity with an IC<sub>50</sub> in



**Figure 2.** Antibiotic effect of seconeolitsine on *D. dadantii* 3937. (A) Survival rate in the presence of increasing amounts of seconeolitsine in solid medium. Each bar indicates the proportion of growing colonies with the specified amount of seconeolitsine in comparison to plates without seconeolitsine, with a 95% statistical confidence interval (see Materials and Methods). (B) Growth curves in the presence of increasing amounts of seconeolitsine in liquid medium. The linear increase of the lag time with drug concentration, obtained from a quantitative analysis of growth curves (see Materials and Methods) is shown in the inset.

the micromolar range (4  $\mu$ M, Figure 1, Supplementary Figure S3 and Materials and Methods), slightly lower than that observed with *M. tuberculosis* topoI (42). We also observed an inhibitory effect on purified topoI from *E. coli* (Supplementary Figure S3C), suggesting that seconeolitsine might be effective against topoI from a broader variety of bacterial species. The estimated IC<sub>50</sub> value was higher in *E. coli* (around 7  $\mu$ M) than in *D. dadantii*, but this difference may be affected by experimental differences between these assays (different initial topoisomer distributions, Supplementary Figure S3).

We ran additional tests to find if seconeolitsine might affect DNA topology through other mechanisms. Inhibition assays were carried with topoisomerase IV and DNA gyrase from *E. coli* (Supplementary Figure S4). We observed an inhibitory activity on topoisomerase IV, but only at several-fold higher concentrations than with topoI (with the pBR322 plasmids employed in that assay), whereas almost no effect was observed with the DNA gyrase. Additionally, supercoiled plasmids exhibited no variation in migration distance even with very high concentration of the drug (data not shown), suggesting that the drug does not intercalate into DNA.

### High concentrations of seconeolitsine impede *D. dadantii* growth

We then investigated the antibacterial activity of the drug, by analysing its effect on *D. dadantii* growth. In solid medium, we observed a progressive reduction in bacterial growth, with a minimal inhibitory concentration (MIC) of around 500  $\mu$ M (Figure 2A). In liquid cultures in microplates (Figure 2B), we observed that the drug increasingly impedes growth, with a lag time proportional to the applied dose in the 100–300  $\mu$ M concentration range. Altogether, the antibacterial effect of the drug occurs at much higher concentrations in *D. dadantii* than *M. tuberculosis* (MIC of 500  $\mu$ M versus 16  $\mu$ M). Since the *in vitro* IC<sub>50</sub> values are comparable for the topoI enzymes from the two species, this strong difference presumably arises from cellular properties (in particular the membrane structures), re-

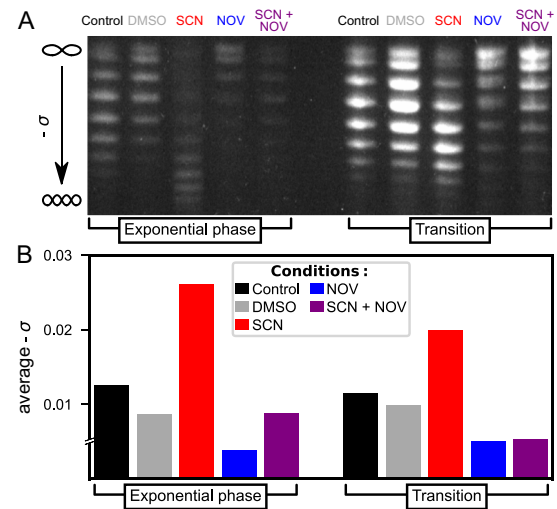
sulting in a different bioavailability of the drug molecules in the cells. As a comparison, *E. coli* cells were inhibited by lower concentrations of seconeolitsine than *D. dadantii*, with a MIC of around 250  $\mu\text{M}$  (Supplementary Figure S5), whereas the growth of the Gram-positive bacterium *Bacillus subtilis* is impeded already at concentrations around 20  $\mu\text{M}$ , comparable to those of *S. pneumoniae* (Supplementary Figure S6).

The *in vitro* data above suggested that topoI is likely the primary target of seconeolitsine in *D. dadantii* cells, but since topoIV was also inhibited at higher drug concentration (Supplementary Figure S4), we ran several tests to confirm it. We checked by microscopy that *D. dadantii* cells grown at a partially inhibitory concentration of seconeolitsine do not exhibit any filamentation, a phenotype typical of topoIV inhibition (due to SlmA-induced lack of DNA segregation, data not shown). We also verified that several genes (*cysJ/N*, *rhsA*) specifically induced in response to topoIV inhibition (O. Espéli, pers. comm.) were unaffected or repressed by seconeolitsine treatment (see transcriptome below). Conversely, we analysed the effect of overexpressing topoI on cell survival, in a medium containing the drug at a partially inhibitory concentration (Supplementary Figure S7). While the survival rate is around 30% in absence of the inducer, it is significantly higher (73%,  $P = 0.0015$ ) when topoI is overexpressed, suggesting that at least a significant fraction of the seconeolitsine molecules are indeed targeted to topoI. Altogether, we conclude that topoI inhibition is presumably the dominant mechanism of action of the drug *in vivo*, although other mechanisms such as an effect on topoIV at high concentration cannot be excluded (see Discussion).

### Seconeolitsine shock increases DNA superhelicity in *D. dadantii* cells

Based on the previous observations and in line with previous studies (32), we anticipated that a seconeolitsine shock at sublethal concentration might induce a rapid increase in negative SC by transiently inhibiting the activity of topoI in *D. dadantii* cells. Indeed, a concentration of 50  $\mu\text{M}$  induced a significant shift in the distribution of topoisomers of the pUC18 plasmid extracted 15 min after the shock (Figure 3A, this time delay was previously chosen to monitor the impact of novobiocin in *D. dadantii*). This concentration was used in all further experiments, because at the same time, it was sufficiently weak to avoid any observable effect on the growth of exponentially growing cells (Supplementary Figure S8), thus minimising general physiological side-effects of the shock versus the direct regulatory effect of DNA SC that we investigate. Note that in *S. pneumoniae*, relatively higher concentrations were used in transcription experiments (up to 8  $\mu\text{M}$ ,  $0.5 \times \text{MIC}$ ).

The distribution of topoisomers is entirely resolved in the untreated and seconeolitsine-treated samples, and thus allows an unambiguous quantification of the observed profiles. In the treated cells, the average negative SC level is increased by  $\Delta\sigma = -0.014$  in exponential phase and  $\Delta\sigma = -0.009$  at the transition to stationary phase (quantified topoisomer distributions are available in Supplementary Figure S9). The weaker effect observed at the latter stage was ex-



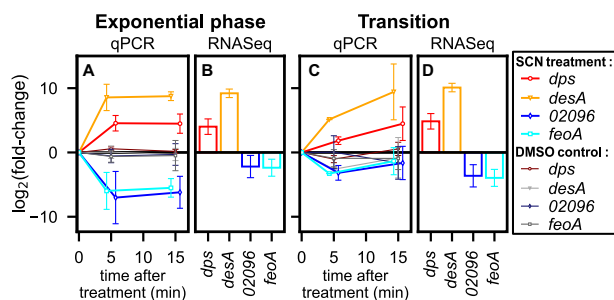
**Figure 3.** A seconeolitsine (SCN) shock at 50  $\mu\text{M}$  concentration induces an increase of negative SC level in cellular DNA after 15 min. Conversely, a novobiocin (NOV) shock at 100  $\mu\text{g ml}^{-1}$  induces DNA relaxation. (A) Agarose-chloroquine gels of pUC18 plasmids isolated from *D. dadantii* 3937 cells. At the employed concentration of chloroquine, the downward migration increases with SC level, and the SC increase induced by seconeolitsine can be fully resolved. (B) Average negative SC level computed from the quantification of the topoisomer distribution. Note that the most relaxed fraction of the topoisomer distribution in presence of novobiocin was not fully resolved, preventing an exact estimation of the relaxation magnitude in these samples, hence the discontinuity indicated in the y-axis.

pected since both DNA gyrase and topoI are more active in the exponential phase (8,10,43). We checked that this increase is absent when only DMSO (used as solvent for seconeolitsine) is applied. In both phases, the sharp increase in negative SC induced by seconeolitsine is in clear opposition to the relaxed levels measured after novobiocin treatment (33), as we expected based on the opposite activity of topoI versus DNA gyrase. Accordingly, in a control sample where both drugs are added simultaneously (rightwards lanes), the plasmids reach an intermediate superhelical level.

Since there are no previous studies of topoI inhibition in *D. dadantii*, these data cannot be directly compared to previously published data; however, the shift in topoisomer distributions observed after seconeolitsine treatment is qualitatively similar to that observed after an osmotic shock (33), which is also known to increase negative SC in *E. coli*, *S. typhimurium* and several other species (8). An additional experiment shows a similar effect in *E. coli* cells, albeit with a stronger effect of seconeolitsine at this concentration of 50  $\mu\text{M}$  (Supplementary Figure S10).

### Transcriptional response of selected promoters

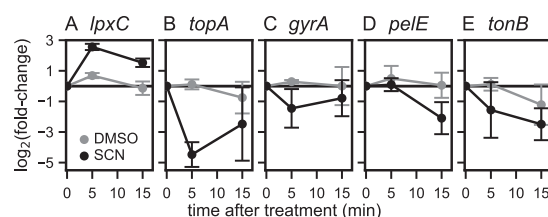
We expected the global increase in negative SC level to affect the expression of many genes of the *D. dadantii* chromosome, and therefore analysed the transcriptional effect of the seconeolitsine shock using RNA-Seq, with a qRT-PCR validation of selected genes. We first illustrate the kinetics of the transcriptional response of four genes strongly responsive to seconeolitsine: the *dps* gene encoding the NAP Dps,



**Figure 4.** Kinetics of promoter activation (*dps*, *desA*) or repression (*Dda3937\_02096*, *feoA*) by seconeolitsine (SCN) shock. Gene expression levels were measured in exponential phase (A and B) and at the transition to stationary phase (in C and D), either by qRT-PCR (5 and 15 min post-shock, coloured markers and thick lines in A and C) or by RNA-Seq (after 15 min incubation with seconeolitsine, B and D). Control datapoints obtained after incubation with the same volume of pure DMSO solvent are indicated as thin lines, and exhibit no detectable effect. All error bars shown indicate 95% confidence intervals, obtained with three biological replicates (qRT-PCR) or from RNA-Seq analysis.

which is possibly the most abundant DNA-binding protein in stationary phase (44) and condenses the chromosome under conditions of resource scarcity or stress; the *desA* gene involved in efflux systems; a gene of unknown function (accession number *Dda3937\_02096*); and *feoA* involved in iron transport. In the exponential phase (Figure 4A), these genes react very quickly (5 min) and in opposite manners. The response measured by RNA-Seq after 15 min (B) was entirely consistent with that measured by qRT-PCR (A); in the latter, we confirmed that DMSO triggers no detectable transcriptional response (thin lines), indicating that seconeolitsine is indeed the active molecule. Similar effects were observed at the transition to stationary phase (Figure 4C and D). The functions of these strongly responsive genes suggest that they are part of a mechanism of drug-response by the bacteria. On the other hand, SC modulates the expression of many genes in a global but usually milder manner (45), as can be observed in Figure 5 with genes expected to respond specifically to SC variations.

The *lpxC* gene illustrates some difficulties encountered when analysing SC-controlled regulation. That gene was previously identified as particularly stable in the presence of various changes of environmental conditions, and because of this apparent lack of regulation, was considered as a suitable internal normalizer for qRT-PCR experiments with *D. dadantii* (46). However, *lpxC* was later found to be sensitive to DNA relaxation by novobiocin (16,47). Similarly here, its expression is increased by seconeolitsine treatment in both growth phases, as observed in both qRT-PCR experiments (Figure 5A) and RNA-Seq data (Supplementary Tab. S2), the former being either calibrated by concentration gradient or using *rpoA* as internal normalizer. This example shows that SC may affect the expression of a large class of promoters, possibly even those lacking a direct dependence on transcription factors since it modulates the direct interaction of RNAP with promoter DNA (8). The expression of *rpoA*, on the other hand, appeared stable in the investigated conditions and it was used as a normalizer for all qRT-PCR data presented.



**Figure 5.** Effect of seconeolitsine shock on various genes expected to respond to variations of SC: (A) *lpxC*; (B) *topA*; (C) *gyrA*; (D) *pelE*; (E) *tonB*. The expression was measured by qPCR 5 and 15 min after the shock in the exponential phase (black dots). Control datapoints treated with the DMSO solvent are shown (grey dots).

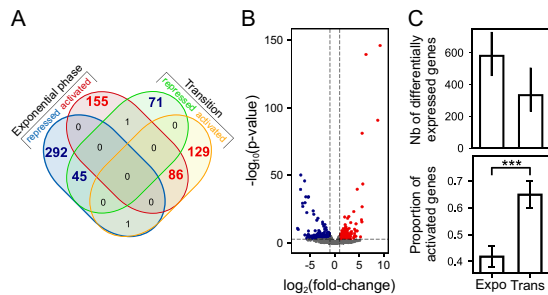
We then investigated the response of topoisomerase genes. The *topA* gene was found to be repressed by the shock after 5 min (Figure 5B), in agreement with observations in *S. pneumoniae* (32), but this effect was already reduced after 15 min, and accordingly, was not detected in the less sensitive RNA-Seq data after the same time delay. By analogy with *S. pneumoniae*, this behaviour might reflect a rapid kinetics of SC homeostasis (32). Note however that the repression of the *topA* gene in both species contradicts the behaviour expected for a simple homeostasis mechanism, which would instead lead to an activation, just like *gyrA/B* genes are activated by DNA gyrase inhibition (4). And in a different study involving oxolinic acid in mutant *E. coli* cells (48), the *topA* promoter was indeed activated by an increase in negative SC, suggesting that its response is possibly more condition-dependent than that of *gyrA/B* genes. Since the basal SC level was more relaxed in those *E. coli* mutants and the magnitude of SC variation was weaker, a possible explanation is that the very high negative SC level reached after seconeolitsine treatment might exceed the dynamic range of the homeostatic response of the *topA* promoter.

Among other topoisomerases, we observed a slight repression of *gyrA* expression (Figure 5C, not significant in the RNA-Seq data) as well as a possible activation of DNA gyrase inhibitors (the *Dda3937\_01484* gene, associated to this function by sequence homology, was found significantly activated in the RNA-Seq data, but not confirmed by qRT-PCR). No effect on topoIV genes (*parC/E*) and topoIII (*topB*) was detected. Altogether, the regulatory mechanisms of SC homeostasis in response to topoI inhibition remain to be fully characterised, and might thus involve a rapid reduction of DNA gyrase activity in addition to changes in topoI expression.

We looked at the *pelE* gene, which encodes a major virulence factor of *D. dadantii* responsible for plant cell wall degrading activity, and is strongly repressed by novobiocin (33). *pelE* was repressed by seconeolitsine in exponential phase (Figure 5D), and not significantly affected at the transition where topoI activity is weaker. The fact that this gene is repressed by both novobiocin (relaxation) and seconeolitsine (increase in negative SC) suggests that the expression is optimal at the natural SC level, consistent with the tight regulation of this level in the cell (22).

Finally, we investigated the *tonB* gene, involved in iron siderophores and vitamin B<sub>12</sub> transport at the cell mem-





**Figure 6.** Global response of the *D. dadantii* genomic expression to a seconeolitsine shock. (A) Venn diagram of significantly activated or repressed genes in the two growth phases, with a threshold of 0.05 on the adjusted p-value. The red numbers refer to genes activated by seconeolitsine (in either phase), and the blue numbers to repressed genes. These numbers vary by around 30% when the threshold is changed by a factor 2. (B) Volcano-plot showing the genomic response to a seconeolitsine shock in exponential growth. Red dots and blue dots correspond to activated and repressed genes, respectively ( $p$ -value threshold of 0.05 and  $\log_2(\text{fold-change})$  threshold of 1 are indicated as dashed lines). Unaffected genes are shown in grey. (C) Top: Total number of differentially expressed genes (among 4260 genes in total) with an adjusted  $p$ -value threshold of 0.05 (bars indicate variations of this number with thresholds of 0.025 and 0.1). Bottom: proportion of activated genes among them (with 95% statistical confidence intervals).

brane. This gene was previously found to be repressed by an increase in negative SC induced by anaerobiosis in both *E. coli* and *S. enterica*, and this repression was relieved by a novobiocin treatment restoring a SC level close to the physiological one (49). Similarly here, we observed a strong repression of *tonB* expression by seconeolitsine in both qRT-PCR experiments (Figure 5E) and RNA-Seq data (Supplementary Tab. S3) in the exponential phase, giving further support to the repressive effect of strongly negative SC levels on that promoter.

### Global transcriptional effect of seconeolitsine shock

A comparison of the response of several other genes confirms that qRT-PCR and RNA-Seq results are well correlated (Supplementary Figure S11), leading us to analyse the transcriptomic results at the global scale. While most previous studies of DNA gyrase inhibition in Gram-negative bacteria were carried in the exponential phase only (14–16,20), we have measured the response to seconeolitsine treatment in the two stages of growth, in each case 15 min after the shock. The lists of differentially expressed genes in either growth phase are given in Supplementary Tables S2 and S3.

The distributions of affected genes are provided in Figure 6. The shock has a significant impact on around 13% of the genome in the exponential phase, and 7% at the transition to stationary phase. Only a small minority of the genes respond significantly in both phases, in which case the response goes in the same direction, whereas many genes respond significantly only in one phase (e.g. *tonB*). This behaviour was not unexpected, since the chromosome conformation (including SC level) and topoisomerase activities are quite different in these two phases (10,43). Among differentially expressed genes, the proportion of activated vs re-

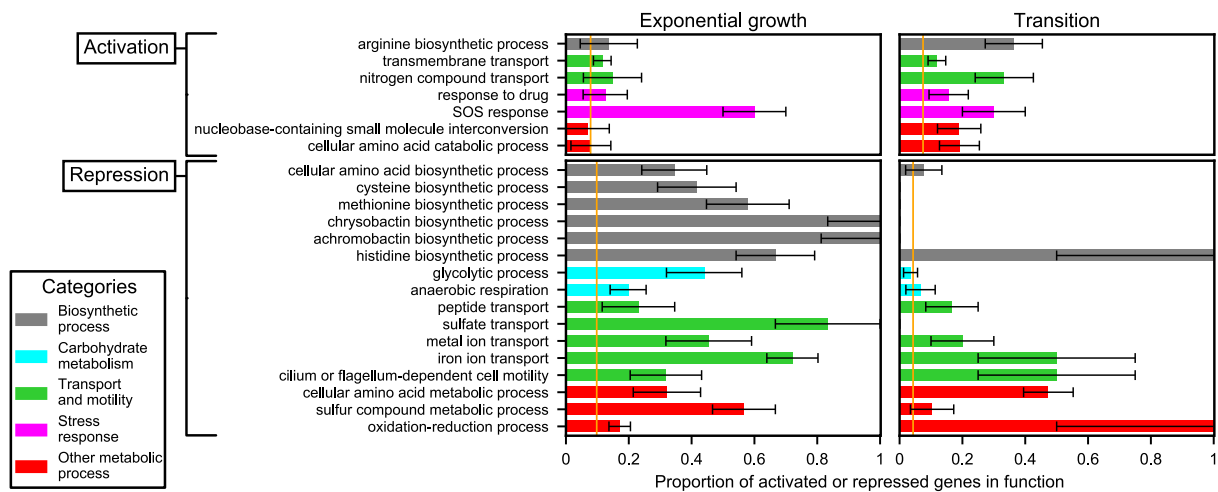
pressed ones is considerably higher at the transition than in the exponential phase (Figure 6C).

In previous studies, DNA relaxation was shown to regulate the expression of the genome in a functionally scattered way, with limited enrichment in specific regulatory pathways (45). We therefore analysed if the same is true of the seconeolitsine shock (Figure 7). Indeed, relatively few Gene Ontology (GO) categories exhibit a strong systematic response, and they belong to very diverse functional groups. Expectedly, the most present pathways are related to (i) metabolism and biosynthesis, as already observed during DNA relaxation (14), which are affected differently in the two phases (see grey, blue and red groups in Figure 7); and (ii) transport end efflux systems, which may, in part, participate in the cellular response to the drug, and are mostly affected similarly in the two phases (green group in Figure 7). We also noted a strong activation of the iron metabolism pathway. But importantly, these enriched functions comprise <40% of the total number of differentially expressed genes, showing that most genes are regulated separately rather than within their entire functional category. We now look in more detail at spatial organisational features of the global pattern of expression.

### Spatial organisation of promoters sensitive to seconeolitsine shock

We started by representing the large-scale distribution of regions enriched in activated or repressed genes along the chromosome (Figure 8). Strikingly, whereas these regions are almost identical in the two investigated growth phases during a novobiocin shock (Figure 8B), they are essentially different during a seconeolitsine shock (Figure 8A), suggesting that, while the large-scale distribution of DNA gyrase activity is similar in the two growth phases, that of *topoI* is growth phase-dependent.

Previous analyses of *D. dadantii* transcriptomes led to the definition of eleven domains of coherent stress-response, termed CODOs (16,49), which harbour distinct DNA physical properties, are differentially regulated by NAPs and novobiocin, and respond coherently to various stress signals encountered during plant infection. These domains are indicated in Figure 8 (black boundaries between the wheels), and in many cases, coincide with patterns of *topoI* activation/repression. As an example, domain 7 (bottom left) harbouring several virulence genes (type VI secretion systems, flagella and chemotaxis operons) is repressed by *topoI* inhibition at the transition to stationary phase. Interestingly, the same effect is observed when the bacteria are subjected to an osmotic shock at this stage of growth (16,49), which also triggers an increase in negative SC (16,33), and mimics the physiological conditions encountered at the beginning of the maceration phase of plant infection (50). Other domains are repressed in exponential phase (domain 10), or activated either in exponential phase (domain 4) or at the transition (domain 2), this latter again consistent with the effect of an osmotic shock, which down-regulates catabolic activity in general and specific stress-responsive genes in domain 2 in particular (16). In summary, although the physical nature and the mechanisms underlying the emergence of these domains remain to be clar-



**Figure 7.** Functional enrichment analysis of activated (top) or repressed (bottom) genes, during a shock in exponential (left) or transition to stationary phase (right). Each bar indicates the proportion of differentially expressed genes in the considered function (with a 95% statistical confidence interval), which can be compared to the genomic average (orange vertical lines): the considered function is enriched if the confidence interval does not cross the orange line. Colours indicate the repartition in broad functional groups.

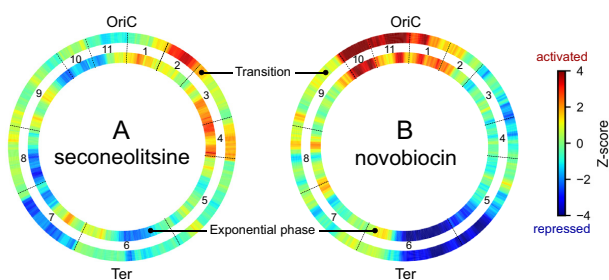
ified, the transcriptional effect of seconeolitsine gives further support to the notion that they reflect an architectural ordering of the chromosome involving SC and affecting its expression, in line with comparable observations in *S. pneumoniae* (51).

A notable feature of the large-scale expression pattern (Figure 8) is that, while the DNA gyrase inhibition pattern is characterised by a clear ori/ter vertical asymmetry (B), the topoI inhibition pattern rather displays an approximate left/right replichore asymmetry (A), reminding the asymmetry in topoI occupancy observed in *S. pneumoniae* (52). However, a statistical comparison of the proportions of activated genes did not exhibit any global difference between the left and right replichores, suggesting that this difference is rather localised in specific regions. Rather, we did find a higher proportion of activated genes on the lagging vs lead-

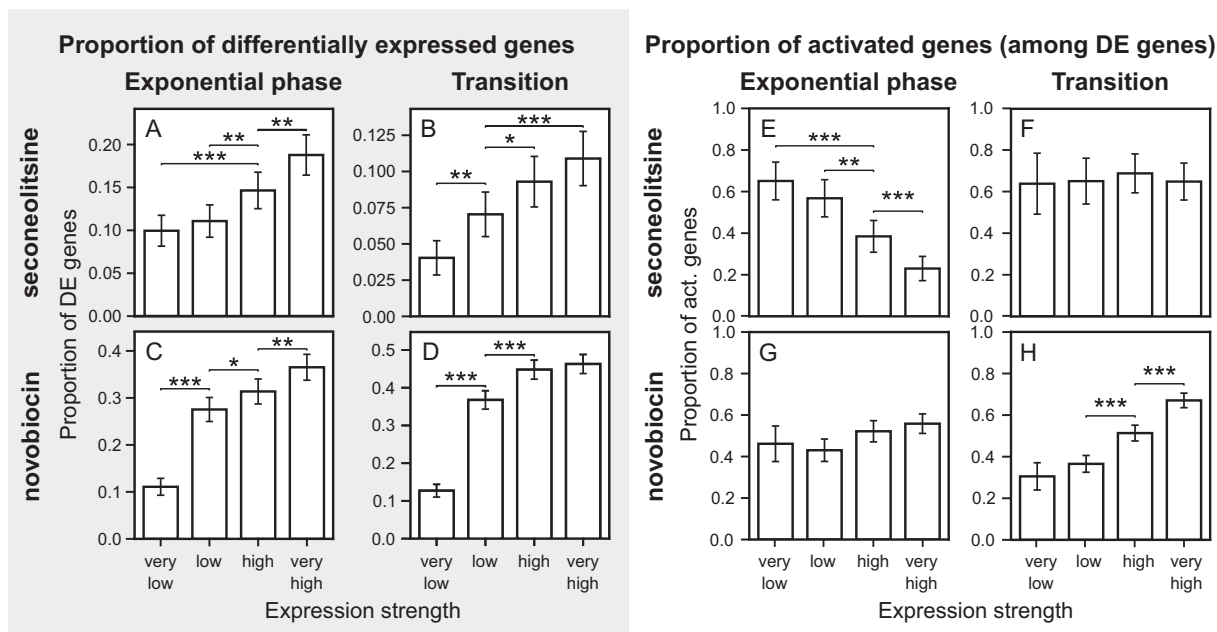
ing strand on transition to stationary phase (Supplementary Figure S12), suggesting that topoI is more important to dissipate negative supercoils on the leading strand (considering both replichores, i.e. with RNAP and DNA polymerase translocating either in the same or in opposite directions). A putative explanation is that the topological constraints might be weaker on the lagging strand where replication does not proceed continuously, but since this difference is not observed in exponential phase where replication is more active, it is likely that the topological constraints induced by the latter are then efficiently handled by topoIV.

#### Topoisomerase I inhibition hinders the expression of strong promoters

Since the leading strand is known to be enriched in highly expressed genes, we looked for a relationship between expression strength and response to seconeolitsine (as well as novobiocin) treatment. TopoI is known to colocalise with RNAP and possibly release negative supercoils generated in its wake during elongation at strongly expressed promoters (26,52), whereas conversely, the DNA gyrase is thought to be recruited downstream of the elongating RNAP (53). Indeed, we found a very strong and progressive increase in the proportion of differentially expressed genes depending on their expression level, for both treatments and in all conditions (Figure 9, left panel), with a four-fold difference between the first and last quartiles. This observation confirms that topoI, as well as DNA gyrase, do not only modulate the SC level of the chromosome at the global scale, but have a strong local and dynamical activity during the transcription process at most operons (and not only at a few highly expressed ones) (26,52). We next looked at the direction of the transcriptional effect of each treatment, and in contrast, found a strong variability (Figure 9, right panel). Highly expressed genes are particularly hampered by topoI inhibition in the exponential phase, from 23% activated genes (in



**Figure 8.** Distribution of genomic regions enriched in activated (red) or repressed (blue) genes, in exponential phase (internal wheels) or transition to stationary phase (external wheels), during topoI inhibition by seconeolitsine (A) or gyrase inhibition by novobiocin (B). The colours represent the statistical significance of the proportion of activated over repressed genes in sliding 500-kb windows (Z-score > 2 or < -2, respectively); if the number of differentially expressed genes in the window is low, the Z-score remains close to 0 and appears in green. Eleven domains of coherent expression (CODOs) previously identified (16) are indicated.



**Figure 9.** Proportion of differentially expressed genes (A–D) and of activated genes (among differentially expressed genes) (E–H) depending on expression strength, in the exponential phase (A, C, E, G) or at the transition to stationary phase (B, D, F, H), after seconeolitsine (A, B, E, F) or novobiocin (C, D, G, H) treatment. Error bars indicate 95% confidence intervals. Genes were separated into quartiles based on their average number of reads across samples.

the lowest quartile) up to 65% in the highest quartile, but not at the transition to stationary phase, where the proportion is constant. In contrast (and surprisingly), DNA gyrase inhibition favours highly expressed genes at the transition to stationary phase, whereas the proportion is constant in the exponential phase. Possible reasons include a globally weaker transcription level at the transition (reducing DNA gyrase requirement), or a weaker inhibitory effect of the transcription-induced negative supercoils in the latter phase where the global SC level is more relaxed (see Discussion).

### Role of neighbouring gene orientation

We then investigated a possible relation between neighbouring gene orientations and the response to seconeolitsine. Such a relationship was expected for the same reason as the previous observation, since RNAP-generated supercoils accumulate not only behind actively transcribed genes, but more specifically between divergent operons (26,47,52,54). The orientation of a gene is here defined by the coding DNA strands of its two neighbours relative to it (in the case of tandem genes, the two neighbours belong to the same strand, which can either be the same as the considered gene or the opposite one). Figure 10 shows that the expected dependence is indeed observed in both growth phases, with genes located between divergent neighbours being significantly more repressed by topoI inhibition compared to convergent ones. This observation is coherent with the observation of a high level of topoI binding in the intergenic region between divergent genes in *Mycobacterium tuberculosis* (54), *E. coli* (26) and *S. pneumoniae* (52). This strong

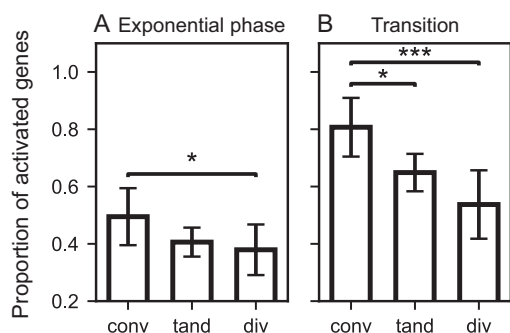
localised activity is probably required to relieve the accumulating negative supercoils, and topoI inhibition thus results in a strong repression of these genes. A similar effect of gene orientation had been already observed following novobiocin treatment (Supplementary Figure S13), highlighting the tight relationship between topoisomerase activity and the genomic organisation due to transcription-induced supercoils (47).

## DISCUSSION

### Effect of seconeolitsine on *D. dadantii*

We have collected the first transcriptomic response to a transient increase in negative SC after inhibition of topoI in a Gram-negative bacterium. Surprisingly, while the latter enzyme is inhibited *in vitro* at a similar micromolar-range concentration as topoI from Gram-positive bacteria, (i) the antibacterial effect occurs only at considerably higher concentration than in the latter (several hundred versus 20  $\mu\text{M}$ ), and (ii) a strong increase in negative SC is detected at a much lower concentration of 50  $\mu\text{M}$ , without significant effect on bacterial growth. The latter feature is not specific to seconeolitsine, since the same is observed (in the opposite superhelical direction) with the DNA gyrase inhibitor novobiocin (at 100  $\mu\text{g ml}^{-1}$  concentration), suggesting that the chromosome is able to handle a broad range of dynamical SC variations without deleterious effects on the cell.

The observed differences between *in vitro* and *in vivo* concentrations, as well as those between Gram-positive and Gram-negative bacteria, may be explained by several factors. We noticed that the solubility of seconeolitsine is sensitive to the physico-chemical conditions, and it is therefore



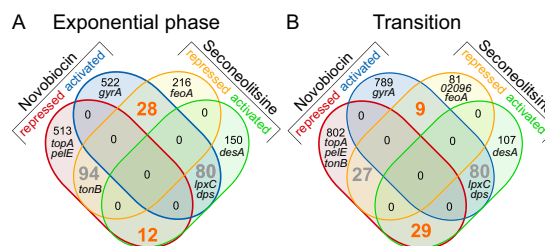
**Figure 10.** Gene orientation-dependent transcriptional response to seconeolitsine. The proportion of activated genes (among differentially expressed ones) is significantly higher among those located between convergent (conv) than divergent (div) neighbours, both in exponential phase (A) and at the transition to stationary phase (B). Tandem (tand) genes exhibit intermediate values. Number of differentially expressed genes (conv, tand, div): 57, 205 and 67 in exponential phase, and 97, 362 and 116 at the transition, respectively.

possible that the availability of the drug is affected by the growth medium (especially at high concentrations); however, this is probably a secondary effect, since the strong superhelical effect *in vivo* suggests that the drug efficiently enters the cell already at 50  $\mu$ M. The most likely explanation is that (i) the membrane of *D. dadantii* cells is a stronger obstacle to the drug molecules than that of Gram-positive bacteria; and (ii) at high concentration, the drug molecules are efficiently expelled by *D. dadantii* when they become toxic, as suggested by the strong activation of efflux and stress-response systems.

The *in vitro* analysis showed that seconeolitsine inhibits topoIV, albeit at a higher concentration than topoI (Supplementary Figure S4). We cannot exclude that a part of the observed increase in negative SC is due to the former, as well as an effect of the drug on topoisomerase III (although no effect was detected on the latter's expression), and other indirect effects of the shock such as a modification of DNA gyrase activity due to the stress response of the cell. However, we did not observe any signature of all these mechanisms, whereas the overexpression of topoI had a clear effect on the drug effect on *D. dadantii* growth, and many of the observed transcriptomic features are compatible with known properties of the topoI binding landscape (see below). In the transcriptomic response obtained after drug treatment, we thus assume that the inhibition of topoI is the dominant factor.

#### The supercoiling-sensitivity of promoters is condition-dependent

Since all previous analyses in Gram-negative species involved the opposite variation, DNA relaxation induced by DNA gyrase inhibition, we wished to compare these complementary responses, in order to refine our understanding of the notion often referred to as the 'supercoiling-sensitivity' of promoters. Figure 11 shows that, among genes responding to one of the drugs, the large majority does not respond to the other: genes appearing as sensitive to DNA relaxation are therefore essentially different



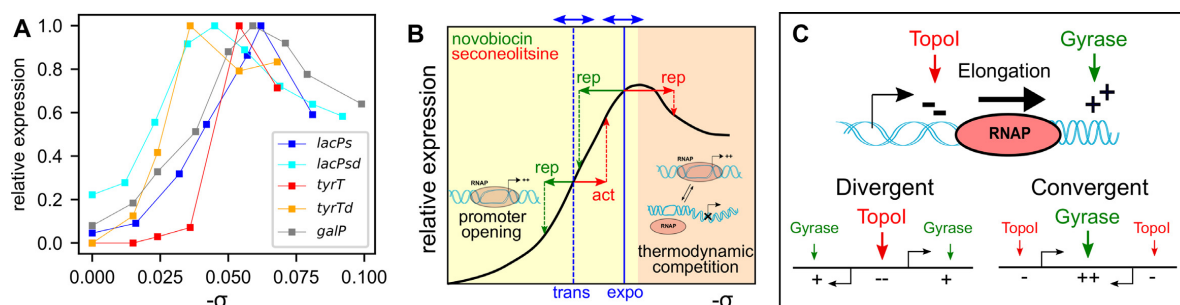
**Figure 11.** Venn diagrams of genomic response to novobiocin and seconeolitsine, in either exponential phase (A) or transition to stationary phase (B). The number of genes responding in opposite directions to the two drugs are indicated in orange, and those in the same direction in grey. Selected genes are indicated in their respective categories. 02096 is the gene with accession number *Dda3937.02096*.

from those sensitive to an increase of SC. This observation is possibly affected by the limited sensitivity of the RNA-Seq experiment, where some genes confirmed by qRT-PCR (*pelE*, *gyrA*) fell below the threshold of statistical significance. Among the genes responding to both drugs, most of them do in the same direction, including some belonging to stress-response functions of the cells possibly via SC-independent regulatory pathways (such as *dps* or *desA*) but also some likely directly regulated by SC (such as *pelE*, *topA* or *tonB*). Finally, a remarkably low number of genes respond in opposite directions to the two drugs, as would yet be naively expected from promoters exhibiting an intrinsic and general property of supercoiling-sensitivity. Note that the latter proportions of similar vs opposed responses to the two drugs were comparable in *S. pneumoniae* cells in exponential phase (Supplementary Figure S14).

These observations, together with others made in this study, highlight the complexity of the SC-related regulation of transcription. The response of a given promoter depends on global parameters related to the physiology of the cell (growth phase, metabolic state, ...) but also to more localised and dynamic factors (local activity of topoisomerase enzymes, mechanical effects of local transcription and replication, binding of nucleoid-associated proteins and regulators, ...), explaining the lack of predictive models of this form of regulation. We now discuss these two contributions successively.

#### A qualitative model for the response of bacterial promoters to global variations of DNA supercoiling

In order to eliminate the local parameters influencing SC-dependent transcriptional regulation and focus on the most global features, it is useful to introduce *in vitro* transcription data, where genes are expressed on plasmids at controlled superhelical levels in absence of any regulatory proteins, and where the former contribution is minimal. Figure 12A recapitulates several available datasets of this kind (7,11,55) obtained with a broad sampling of SC levels comprising typical physiological levels, either in standard conditions (from -0.04 to -0.06), upon DNA gyrase inhibition (lower negative SC levels) and upon topoI inhibition (higher negative SC levels). The employed promoters belong to different promoter families, either from stable RNAs (*tyrT*) and mutant



**Figure 12.** (A) Regulation of various bacterial promoters by SC *in vitro*. The employed native promoters encode either stable (*tyrT*) or messenger RNAs (*galP*), whereas *lacPs* is a mutant promoter derived from *lacP*. *tyrTd* and *lacPsd* are mutant versions of *tyrT* and *lacPs*, respectively (7,11,55). (B) Qualitative regulatory model summarising the data of A (black solid line) with putative mechanisms: promoter DNA opening for open-complex formation (yellow background) (17) and thermodynamic opening competition (orange background, see text) (56). Physiological SC levels valid for many bacteria in exponential or transition to stationary phase are indicated in blue, with double arrows symbolising limited precision and species-dependent variability (57). The expected regulatory effect of an antibiotic shock in either phase is indicated in green. (C) Model of orientation-dependent binding of topoisomerases, and subsequent transcriptional regulation by topoI inhibition (adapted from (61)).

promoters derived thereof (*tyrTd*), or promoters of protein-encoding genes (*galP*) or derivatives of *lacP* (*lacPs*, *lacPsd*). In spite of conspicuous differences between these curves, a similar pattern is clearly and repeatedly observable: the expression is very low on an entirely relaxed DNA template, then increases drastically and monotonously until reaching maximal expression at a (promoter-dependent) optimal SC level close to the physiological level in exponential phase ( $\approx -0.06$ ), then decreases at higher SC levels. This behaviour is schematised in Figure 12B, where the horizontal axis is voluntarily left without quantitative values. The two background colours highlight the two regulation regimes with putative associated mechanisms: the initial activation curve is likely due to the SC-induced reduction of DNA opening free energy during open-complex formation, which occurs preferentially at the highly AT-rich region starting at the  $-10$  promoter element where the transcription bubble is formed (17); the decrease is more complex and either due to the opening of secondary sites competing with the  $-10$  element (56), or to a reduction in processive initiation due to an excessive stability of the open-complex resulting in more abortive transcripts (6). Accordingly, while a modification of the AT-richness of the promoter sequence downstream of the  $-10$  element (in *lacPsd* and *tyrTd* compared to *lacPs* and *tyrT*, respectively) clearly shifts the activation curve horizontally in a predictable manner (17), the second part of the curve is more variable, and the position of the maximum differs significantly from one promoter to the other (7).

Based on this empirical model, the regulatory effect of a SC variation *in vivo* (e.g. due to a topoisomerase inhibitor) is then expected to depend both on the initial global level in the cell, and on the direction and magnitude of the change. Approximate values of the average SC level in exponential or transition to stationary phase are indicated in blue, with exact values varying between species (57). During a relaxation shock, in either phase, the expression rate is predicted to be shifted leftwards to a lower level, explaining the comparable pattern of expression observed with novobiocin in both phases (Figure 8B). During a seconeolitsine shock, on the other hand, the inhibition of topoI (rightwards shift) is expected to induce the expression rate of most promoters at

the transition to stationary phase, whereas in the exponential phase, the SC level is already close to the maximum of the curve, and the shock should thus reduce the expression level of many of them.

Although very simplified, this analysis from *in vitro* data might contribute to several notable observations that we have made from our data: (i) many promoters, such as *lpxC*, *tonB* and *dps*, respond to novobiocin and seconeolitsine in the same direction (Figure 11), suggesting a non-monotonous SC-activation curve; (ii) the expression pattern associated to seconeolitsine is more condition-dependent than that of novobiocin (Figure 8); (iii) seconeolitsine mostly represses promoters in the exponential phase, and activates them at the transition to stationary phase (Figure 6C).

#### Role of topoisomerase I in resolving transcription-induced supercoils

While the global SC level affects the expression of the entire genome, our analysis also highlighted the importance of several local parameters in the promoters' response to topoI inhibition. Previous studies showed an effect of large-scale features related to DNA replication (left/right replichore, leading/lagging strand) (52), but these had quite limited impact in our data. In contrast, we found two predominant features at the kilobase-scale, promoter strength (Figure 9) and local gene orientation (Figure 10), both pointing to a role of topoI in the handling of supercoils generated during transcription elongation (Figure 12C), following the model of Liu and Wang (58). These two observed features are entirely consistent with genomic distributions of topoisomerases observed by ChIP-Seq in several species (26,52,54), as well as early studies on the specific role of topoI in the handling of negative supercoils at divergent promoters (59). But interestingly, while those ChIP-Seq data highlight the (one-sided) effect of transcription on SC distributions, the analysed transcriptomes also reflect the reciprocal relation, i.e. that SC acts as a regulatory factor, and transcription and SC are thus involved in a double-sided and nonlinear coupling (47). To our knowledge, this study is the first to highlight the strong influ-

9160 *Nucleic Acids Research*, 2022, Vol. 50, No. 16

ence of a promoter's strength on its sensitivity to topoi and DNA gyrase inhibition at the genomic scale (Figure 9, left panel). But while this result was expected based on previous ChIP-Seq data (26,52,54), the *direction* of the resulting regulatory effect (activation or repression, Figure 9 right panel) is quite surprising, and reflects the complexity of this coupling. For example, while topoi inhibition expectedly disfavours highly expressed genes (especially in the exponential phase), the latter are favoured by DNA gyrase inhibition at the transition to stationary phase. In a similar manner, while divergent genes are expectedly disfavoured by topoi inhibition (Figure 10), presumably due to the accumulation of negative supercoils in the central region, the same is surprisingly observed after DNA gyrase inhibition (Supplementary Figure S13), even though positive supercoils then presumably accumulate between convergent genes. Such effects may not be predictable from a simple static model like that proposed on Figure 12B, since they result from an intrinsically dynamic interplay between transcription elongation, diffusion of supercoils, and recruitment of new RNAP enzymes at nearby promoters. A recent unidimensional stochastic model of this process was able to reproduce the counter-intuitive effect of DNA gyrase inhibition on convergent genes (47); in those simulations, the latter behaviour arose because the positive supercoils generated by nearby genes were sufficient to partly repress these promoters already when DNA gyrase was fully active. However, not only does this explanation require experimental support, but the transcription-supercoiling coupling is likely highly dependent on more subtle 3D parameters, in particular the partition of local SC into constrained and unconstrained fractions, and into twist and writhe. These contributions are affected very differently by the two main considered topoisomerase enzymes, since DNA gyrase introduces supercoils by crossing two distal loci coming into close spatial proximity, i.e. predominantly introduces writhe (30), whereas topoi cleaves a single strand of negatively supercoiled DNA, i.e. predominantly removes an excess of negative twist (1). SC distributions are also strongly affected by the DNA sequence and by the recruitment of nucleoid-associated proteins, most of which induce distortions into DNA and displace the equilibrium between twist and writhe in favour of the latter. Altogether, a better understanding of this regulation will thus significantly benefit from a detailed and high-resolution mapping of the distribution of local SC levels along the chromosome (60).

#### DATA AVAILABILITY

The sequenced reads were deposited in ArrayExpress, under accession number E-MTAB-10134.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

#### ACKNOWLEDGEMENTS

We thank Georgi Muskhelishvili for his suggestions and critical reading of the manuscript, Olivier Espéli for useful advice, and Florelle Deboudard for technical help.

#### FUNDING

INSA Lyon [BQR 2016]; Agence Nationale de la Recherche [ANR-18-CE45-0006-01 to S.M.]. Funding for open access charge: Agence Nationale de la Recherche [ANR-18-CE45-0006-01]; INSA Lyon [BQR].

*Conflict of interest statement.* None declared.

#### REFERENCES

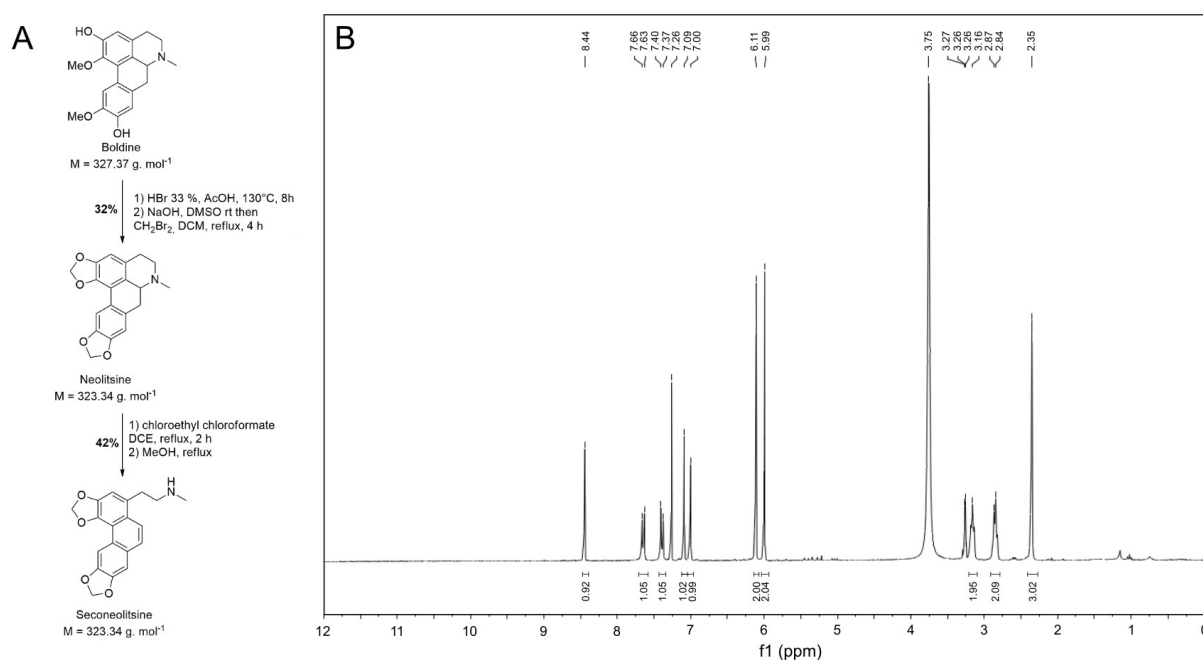
- Maxwell, A. and Gellert, M. (1986) Mechanistic aspects of DNA topoisomerases. In: Anfinsen, C.B., Edsall, J.T. and Richards, F.M. (eds). *Advances in Protein Chemistry*. Academic Press, Vol. 38, pp. 69–107.
- Zechiedrich, E.L., Khodursky, A.B., Bachellier, S., Schneider, R., Chen, D., Lilley, D.M. and Cozzarelli, N.R. (2000) Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. *J. Biol. Chem.*, **275**, 8103–8113.
- Chen, S.H., Chan, N.-L. and Hsieh, T. (2013) New mechanistic and functional insights into DNA topoisomerases. *Annu. Rev. Biochem.*, **82**, 139–170.
- Menzel, R. and Gellert, M. (1983) Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. *Cell*, **34**, 105–113.
- Ohlsen, K.L. and Gralla, J.D. (1992) Interrelated effects of DNA supercoiling, ppGpp, and low salt on melting within the *Escherichia coli* ribosomal RNA *rrnB* P1 promoter. *Mol. Microbiol.*, **6**, 2243–2251.
- Wood, D.C. and 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.
- Lim, H.M., Lewis, D.E.A., Lee, H.J., Liu, M. and Adhya, S. (2003) Effect of varying the supercoiling of DNA on transcription and its regulation. *Biochemistry*, **42**, 10718–10725.
- Martis, B.S., Forquet, R., Reverchon, S., Nasser, W. and Meyer, S. (2019) DNA supercoiling: an ancestral regulator of gene expression in pathogenic bacteria? *Comput. Struct. Biotechnol. J.*, **17**, 1047–1055.
- Pruss, G.J. and Drlica, K. (1989) DNA supercoiling and prokaryotic transcription. *Cell*, **56**, 521–523.
- Travers, A. and Muskhelishvili, G. (2005) DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nat. Rev. Microbiol.*, **3**, 157–169.
- 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. *J. Mol. Biol.*, **331**, 331–344.
- Hsieh, L.S., Rouviere-Yaniv, J. and Drlica, K. (1991) Bacterial DNA supercoiling and [ATP]/[ADP] ratio: changes associated with salt shock. *J. Bacteriol.*, **173**, 3914–3917.
- Borowiec, J.A. and Gralla, J.D. (1985) Supercoiling response of the *lac* promoter in vitro. *J. Mol. Biol.*, **184**, 587–598.
- Blot, N., Mavathur, R., Geertz, M., Travers, A. and Muskhelishvili, G. (2006) Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep.*, **7**, 710–715.
- Peter, B.J., Arsuaaga, J., Breier, A.M., Khodursky, A.B., Brown, P.O. and Cozzarelli, N.R. (2004) Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biol.*, **5**, R87.
- Jiang, X., Sobetzko, P., Nasser, W., Reverchon, S. and Muskhelishvili, G. (2015) Chromosomal 'stress-response' domains govern the spatiotemporal expression of the bacterial virulence program. *Mbio*, **6**, e00353-15
- Forquet, R., Pineau, M., Nasser, W., Reverchon, S. and Meyer, S. (2021) Role of the discriminator sequence in the supercoiling sensitivity of bacterial promoters. *Msystems*, **6**, e0097821.
- Forquet, R., Nasser, W., Reverchon, S. and Meyer, S. (2022) Quantitative contribution of the spacer length in the supercoiling-sensitivity of bacterial promoters. *Nucleic Acids Res.*, **50**, 7287–7297
- Jovanovich, S.B. and Lebowitz, J. (1987) Estimation of the effect of coumermycin A1 on *Salmonella typhimurium* promoters by using random operon fusions. *J. Bacteriol.*, **169**, 4431–4435.

20. Gogoleva, N.E., Konnova, T.A., Balkin, A.S., Plotnikov, A.O. and Gogolev, Y.V. (2020) Transcriptomic data of *Salmonella enterica* subsp. *enterica* serovar *Typhimurium* str. 14028S treated with novobiocin. *Data Brief*, **29**, 105297.
21. Drlica, K. (1992) Control of bacterial DNA supercoiling. *Mol. Microbiol.*, **6**, 425–433.
22. Pruss, G.J., Manes, S.H. and Drlica, K. (1982) *Escherichia coli* DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. *Cell*, **31**, 35–42.
23. Champoux, J.J. (2001) DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.*, **70**, 369–413.
24. Franco, R.J. and Drlica, K. (1989) Gyrase inhibitors can increase *gyrA* expression and DNA supercoiling. *J. Bacteriol.*, **171**, 6573–6579.
25. Seddek, A., Annamalai, T. and Tse-Dinh, Y.-C. (2021) Type IA Topoisomerases as Targets for Infectious Disease Treatments. *Microorganisms*, **9**, 86.
26. Sutormin, D., Galivondzhyan, A., Musharova, O., Travin, D., Rusanova, A., Obratsova, K., Borukhov, S. and Severinov, K. (2021) Interaction between transcribing RNA polymerase and topoisomerase I prevents R-loop formation in *E. coli*. bioRxiv doi: <https://doi.org/10.1101/2021.10.26.465782>, 26 October 2021, preprint: not peer reviewed.
27. Cheng, B., Zhu, C.-X., Ji, C., Ahumada, A. and Tse-Dinh, Y.-C. (2003) Direct interaction between *Escherichia coli* RNA polymerase and the zinc ribbon domains of DNA topoisomerase I. *J. Biol. Chem.*, **278**, 30705–30710.
28. Spakman, D., Bakx, J.A.M., Biebricher, A.S., Peterman, E.J.G., Wuite, G.J.L. and King, G.A. (2021) Unravelling the mechanisms of type IA topoisomerases using single-molecule approaches. *Nucleic Acids Res.*, **49**, 5470–5492.
29. Nagaraja, V., Godbole, A.A., Henderson, S.R. and Maxwell, A. (2017) DNA topoisomerase I and DNA gyrase as targets for TB therapy. *Drug Discov. Today*, **22**, 510–518.
30. Bush, N.G., Evans-Roberts, K. and Maxwell, A. (2015) DNA topoisomerases. *EcoSal Plus*, **6**, <https://doi.org/10.1128/ecosalplus.ESP-0010-2014>.
31. García, M.T., Blázquez, M.A., Ferrándiz, M.J., Sanz, M.J., Silva-Martín, N., Hermoso, J.A. and de la Campa, A.G. (2011) New alkaloid antibiotics that target the DNA topoisomerase I of *Streptococcus pneumoniae*. *J. Biol. Chem.*, **286**, 6402–6413.
32. Ferrándiz, M.-J., Martín-Galiano, A.J., Arnanz, C., Camacho-Soguero, I., Tirado-Vélez, J.-M. and de la Campa, A.G. (2016) 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.*, **44**, 7292–7303.
33. Ouafa, Z.-A., Reverchon, S., Lautier, T., Muskhelishvili, G. and 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.
34. Campa, A.G.D.L., Ferrer, M.A.B. and Esteban, M.T.G. (2012) Use of seconolitsin and n-methyl-seconolitsin a for the manufacture of medicines. Patent ES2363074B1.
35. Xia, Y., Li, K., Li, J., Wang, T., Gu, L. and Xun, L. (2019) T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis. *Nucleic Acids Res.*, **47**, e15.
36. Tan, K., Zhou, Q., Cheng, B., Zhang, Z., Joachimiak, A. and Tse-Dinh, Y.-C. (2015) Structural basis for suppression of hypernegative DNA supercoiling by *E. coli* topoisomerase I. *Nucleic Acids Res.*, **43**, 11031–11046.
37. Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, **72**, 248–254.
38. Zwietering, M.H., Jongenburger, I., Rombouts, F.M. and van 't Riet, K. (1990) Modeling of the bacterial growth curve. *Appl. Environ. Microbiol.*, **56**, 1875–1881.
39. Reverchon, S., Meyer, S., Forquet, R., Hommais, F., Muskhelishvili, G. and Nasser, W. (2021) The nucleoid-associated protein IHF acts as a 'transcriptional domainin' protein coordinating the bacterial virulence traits with global transcription. *Nucleic Acids Res.*, **49**, 776–790.
40. Maes, M. and Messens, E. (1992) Phenol as grinding material in RNA preparations. *Nucleic Acids Res.*, **20**, 4374.
41. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. et al. (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
42. García, M.T., Carreño, D., Tirado-Vélez, J.M., Ferrándiz, M.J., Rodrigues, L., Gracia, B., Amblar, M., Ainsa, J.A. and de la Campa, A.G. (2018) Boldine-Derived alkaloids inhibit the activity of DNA topoisomerase I and growth of mycobacterium tuberculosis. *Front. Microbiol.*, **9**, 1659.
43. Balke, V.L. and Gralla, J.D. (1987) Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *J. Bacteriol.*, **169**, 4499–4506.
44. Verma, S.C., Qian, Z. and Adhya, S.L. (2019) Architecture of the *Escherichia coli* nucleoid. *PLoS Genet.*, **15**, e1008456.
45. Travers, A. and Muskhelishvili, G. (2020) Chromosomal organization and regulation of genetic function in *Escherichia coli* integrates the DNA analog and digital information. *EcoSal Plus*, **9**, <https://doi.org/10.1128/ecosalplus.ESP-0016-2019>.
46. Hommais, F., Zghidi-Abouzid, O., Oger-Desfeux, C., Pineau-Chapelle, E., Van Gijsegem, F., Nasser, W. and Reverchon, S. (2011) *lpxC* and *yafS* are the most suitable internal controls to normalize real time RT-qPCR expression in the phytopathogenic bacteria *dickeya dadantii*. *PLoS One*, **6**, e20269.
47. El Houdaigui, B., Forquet, R., Hindré, T., Schneider, D., Nasser, W., Reverchon, S. and Meyer, S. (2019) Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Res.*, **47**, 5648–5657.
48. Tse-Dinh, Y.C. and Beran, R.K. (1988) Multiple promoters for transcription of the *Escherichia coli* DNA topoisomerase I gene and their regulation by DNA supercoiling. *J. Mol. Biol.*, **202**, 735–742.
49. Muskhelishvili, G., Forquet, R., Reverchon, S., Meyer, S. and Nasser, W. (2019) Coherent domains of transcription coordinate gene expression during bacterial growth and adaptation. *Microorganisms*, **7**, 694.
50. Leonard, S., Hommais, F., Nasser, W. and Reverchon, S. (2017) Plant-phytopathogen interactions: bacterial responses to environmental and plant stimuli. *Environ. Microbiol.*, **19**, 1689–1716.
51. de la Campa, A.G., Ferrándiz, M.J., Martín-Galiano, A.J., García, M.T. and Tirado-Vélez, J.M. (2017) The transcriptome of *Streptococcus pneumoniae* induced by local and global changes in supercoiling. *Front. Microbiol.*, **8**, 1447.
52. Ferrándiz, M.-J., Hernández, P. and Campa, A.G. de la (2021) Genome-wide proximity between RNA polymerase and DNA topoisomerase I supports transcription in *Streptococcus pneumoniae*. *PLoS Genet.*, **17**, e1009542.
53. Sutormin, D., Rubanova, N., Logacheva, M., Ghilarov, D. and Severinov, K. (2019) Single-nucleotide-resolution mapping of DNA gyrase cleavage sites across the *Escherichia coli* genome. *Nucleic Acids Res.*, **47**, 1373–1388.
54. Ahmed, W., Sala, C., Hegde, S.R., Jha, R.K., Cole, S.T. and Nagaraja, V. (2017) Transcription facilitated genome-wide recruitment of topoisomerase I and DNA gyrase. *PLoS Genet.*, **13**, e1006754.
55. Borowiec, J.A. and Gralla, J.D. (1987) All three elements of the *lac* promoter mediate its transcriptional response to DNA supercoiling. *J. Mol. Biol.*, **195**, 89–97.
56. Hatfield, G.W. and Benham, C.J. (2002) DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu. Rev. Genet.*, **36**, 175–203.
57. Rovinskiy, N.S., Agbleke, A.A., Chesnokova, O.N. and Higgins, N.P. (2019) Supercoil levels in *E. coli* and *Salmonella* chromosomes are regulated by the C-terminal 35–38 Amino acids of GyrA. *Microorganisms*, **7**, 81.
58. Liu, L.F. and Wang, J.C. (1987) Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci.*, **84**, 7024–7027.
59. Chen, D., Bowater, R., Dorman, C.J. and Lilley, D.M. (1992) Activity of a plasmid-borne *leu-500* promoter depends on the transcription and translation of an adjacent gene. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 8784–8788.
60. Lal, A., Dhar, A., Trostel, A., Kouzine, F., Seshasayee, A.S.N. and Adhya, S. (2016) Genome scale patterns of supercoiling in a bacterial chromosome. *Nat. Commun.*, **7**, 11055.
61. Lilley, D.M.J., Chen, D. and Bowater, R.P. (1996) DNA supercoiling and transcription: topological coupling of promoters. *Q. Rev. Biophys.*, **29**, 203–225.

What is a supercoiling-sensitive gene?  
Insights from topoisomerase I inhibition  
in the Gram-negative bacterium *Dickeya dadantii*

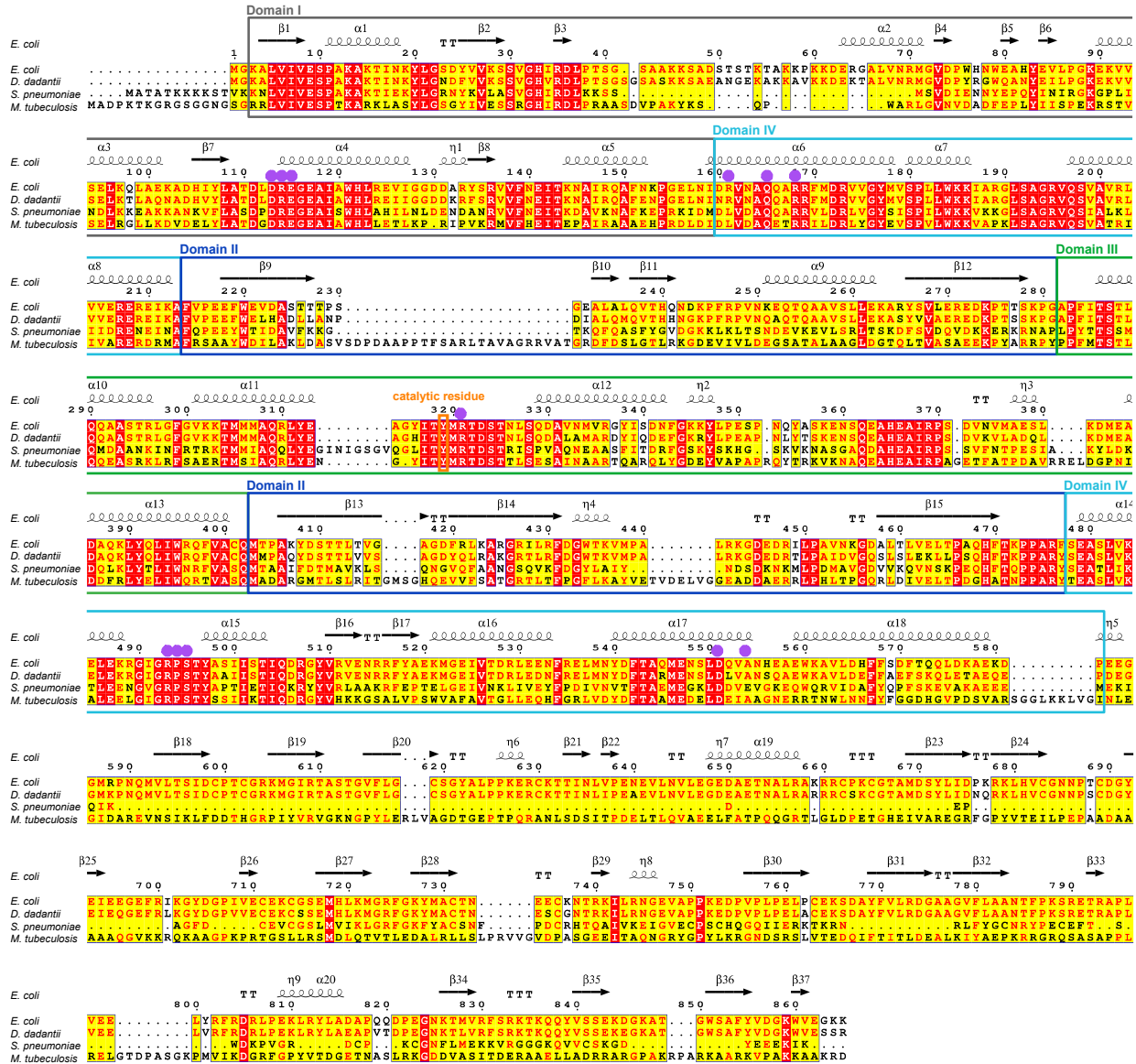
## Supplementary Information

Maïwenn Pineau, Shiny Martis B., Raphaël Forquet, Jessica Baude, Camille Villard, Lucie Grand, Florence Popowycz, Laurent Soullère, Florence Hommais, William Nasser, Sylvie Reverchon and Sam Meyer

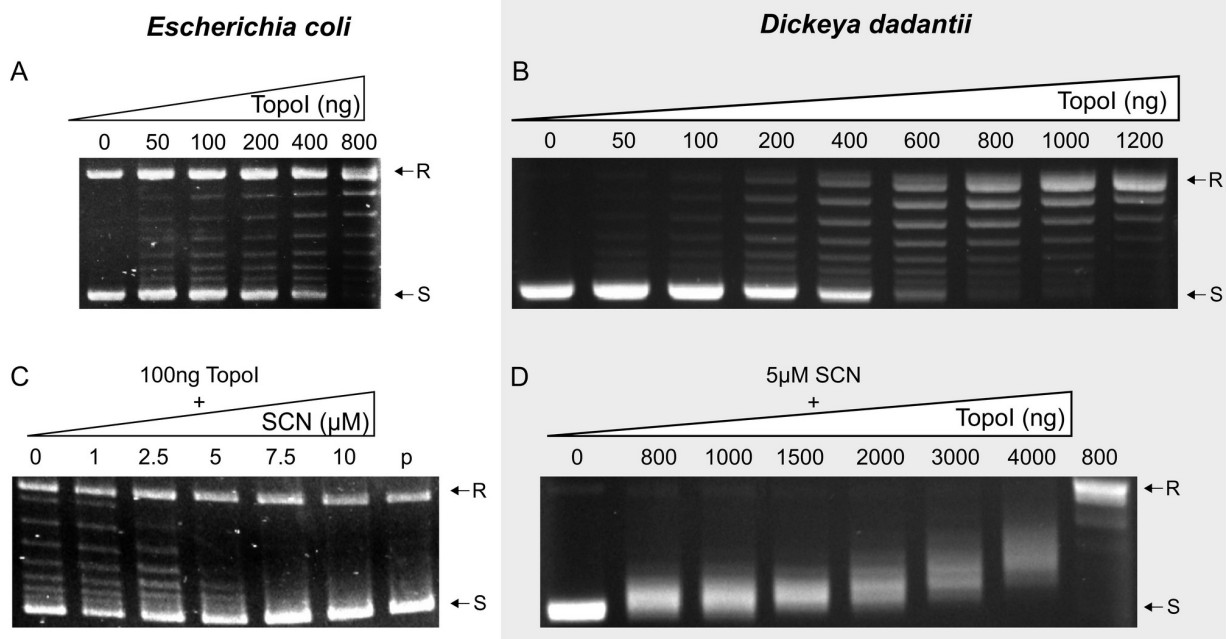


**Fig. S1:** (A) Synthesis of seconeolitsine from boldine. (B) Purity of the product measurement: spectrum of the compound by proton nuclear magnetic resonance (<sup>1</sup>H NMR) (300MHz in CDCl<sub>3</sub> / CD<sub>3</sub>OD).

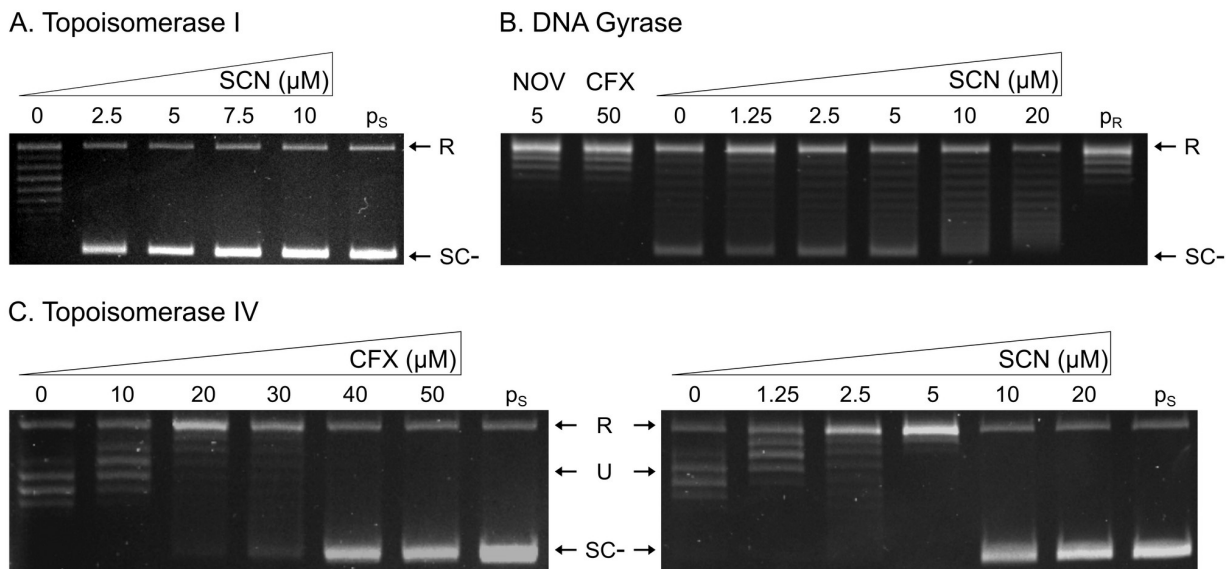




**Fig. S2:** Alignment of *E. coli*, *D. dadantii*, *S. pneumoniae* and *M. tuberculosi* topoisomerase I sequences obtained using ENDscript (1). Identities are highlighted in red, similarities in yellow. Secondary structure elements of *E. coli* topoisomerase I are shown above the sequences ( $\alpha$ -helices by squiggles,  $\beta$ -strands by arrows, strict  $\alpha$ -turns by TTT letters and  $\beta$ -turns by TT letters). The catalytic tyrosine residue is shown by an orange box (Tyr-319 for *E. coli*). Residues indicated with purple dots are part of the nucleotide-binding site and might interact with seconeoltsine. Grey, dark blue, green and light blue boxes indicate previously described topoisomerase I domains (2).

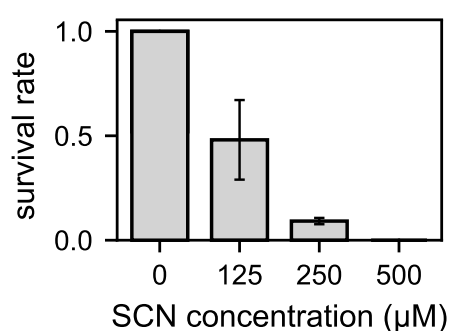


**Fig. S3:** Activity of purified topoisomerase I from *E. coli* (A) and *D. dadantii* (B). 0.5 μg of pUC18 plasmid was incubated with the indicated mass of topoisomerase I for 15 min at 37°C. R corresponds to pUC18 relaxed form, S to the highly negatively supercoiled form. Inhibition of *E. coli* (C) and *D. dadantii* (D) topoisomerase I by seconeolitsine. The indicated amount of seconeolitsine was first incubated 10 min with 100 ng *E. coli* topoisomerase I (C) or the indicated amount of *D. dadantii* topoisomerase I (D) at 4°C. 0.5 μg of pUC18 plasmid (p) was then added and the mix was incubated for 15 min at 37°C. In (D), the lane at the right of the gel contained no seconeolitsine and 800 ng topoI. The IC<sub>50</sub> of seconeolitsine for *E. coli* topoisomerase I was estimated around 7 μM (see Materials and Methods).

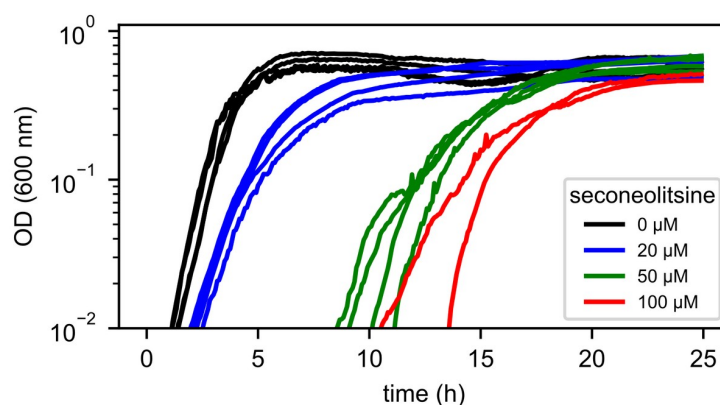


**Fig. S4:** Activity of purified *D. dadantii* topoisomerase I (A), *E. coli* DNA gyrase (B) and *E. coli* topoisomerase IV (C) inhibited by novobiocin (NOV), ciprofloxacin (CFX, a gyrase and topoIV inhibitor used as control) or seconeolitsine (SCN). Agarose gels of supercoiled (p<sub>S</sub>) or relaxed (p<sub>R</sub>)

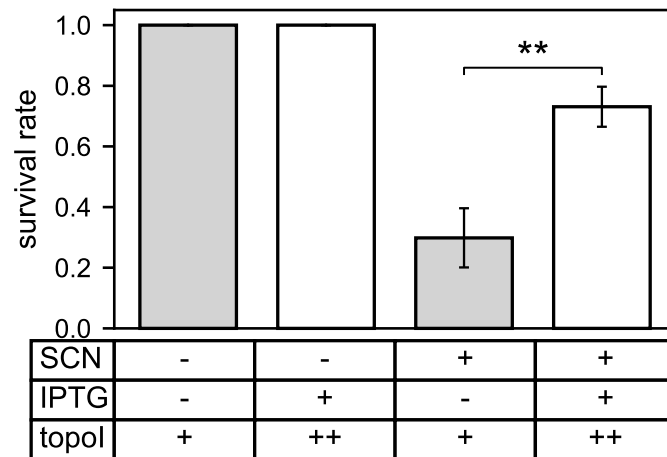
pBR322 plasmids were treated following manufacturer's instructions (Inspiralis kits), with the indicated amount of inhibitor. S indicates the negatively supercoiled form, R the relaxed form, and U the (unknown) form obtained after topoIV treatment without inhibitor. With the employed pBR322 plasmids, topoI is fully inhibited by less than 2.5  $\mu\text{M}$  seconeolitsine. DNA gyrase is almost unaffected (slight reduction of the most supercoiled topoisomers at the highest concentrations of seconeolitsine). TopoIV is inhibited by seconeolitsine with an IC<sub>50</sub> between 5 and 10  $\mu\text{M}$ ; note that the U form migrates as weakly positively supercoiled plasmids, so that a partial topoIV inhibition (by low concentrations of CFX or SCN) results in more relaxed plasmids (remaining at the top of the gel), whereas a full inhibition by either drug results in highly negatively supercoiled plasmids (bottom of the gel, like the untreated plasmids). The IC<sub>50</sub> of topoIV is around three-fold higher than the IC<sub>50</sub> of topoI.



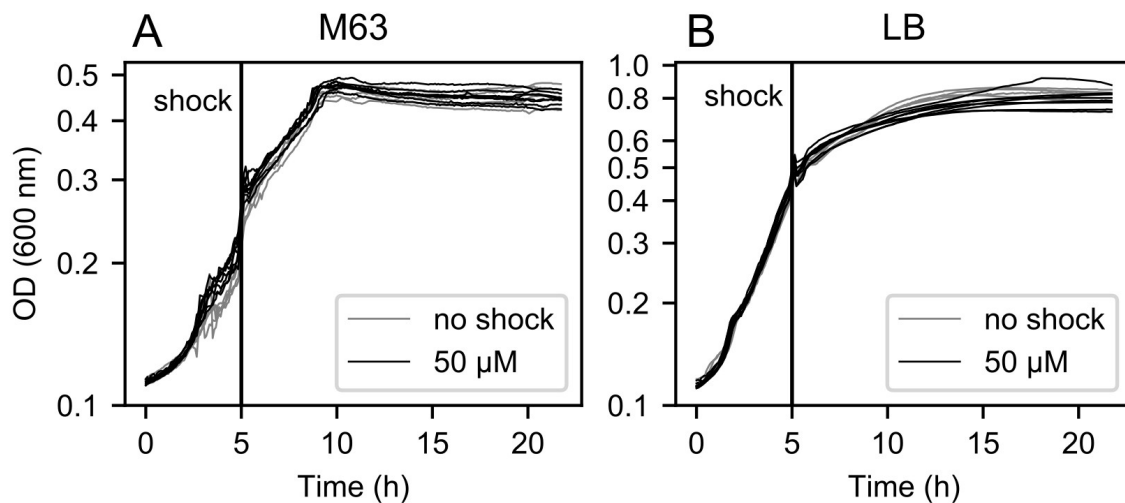
**Fig. S5:** Inhibition of the growth of *E. coli* NM522 cells by seconeolitsine. Each bar indicates the proportion of growing colonies on plates with the specified amount of seconeolitsine compared to plates without seconeolitsine, with a 95% statistical confidence interval.



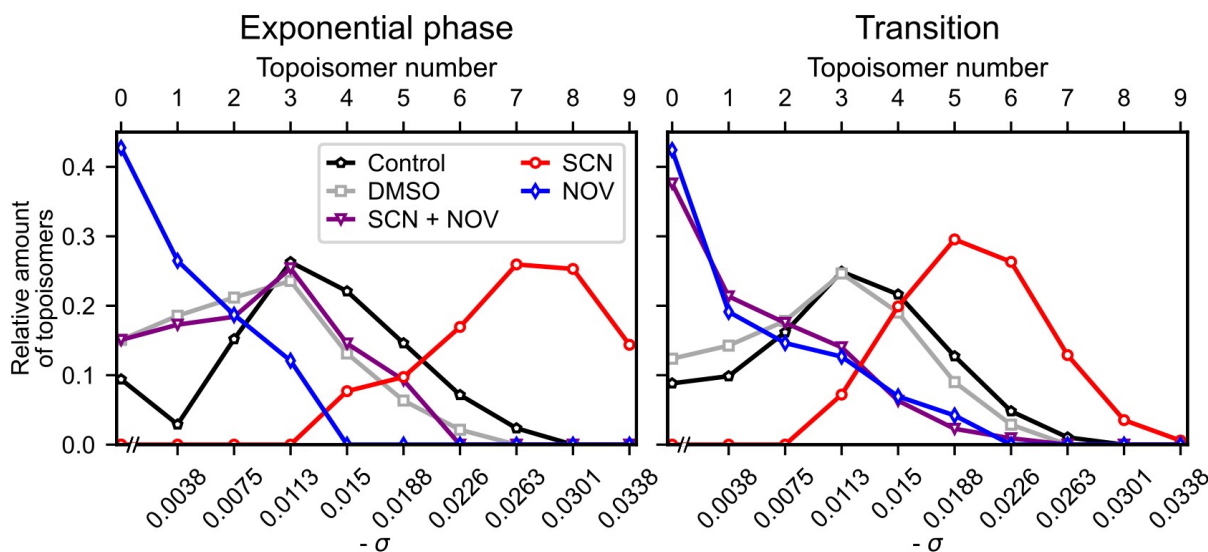
**Fig. S6:** Growth curves of the Gram-positive bacterium *B. subtilis* in presence of increasing amounts of seconeolitsine, solvated in a constant (5% vol) volume of DMSO.



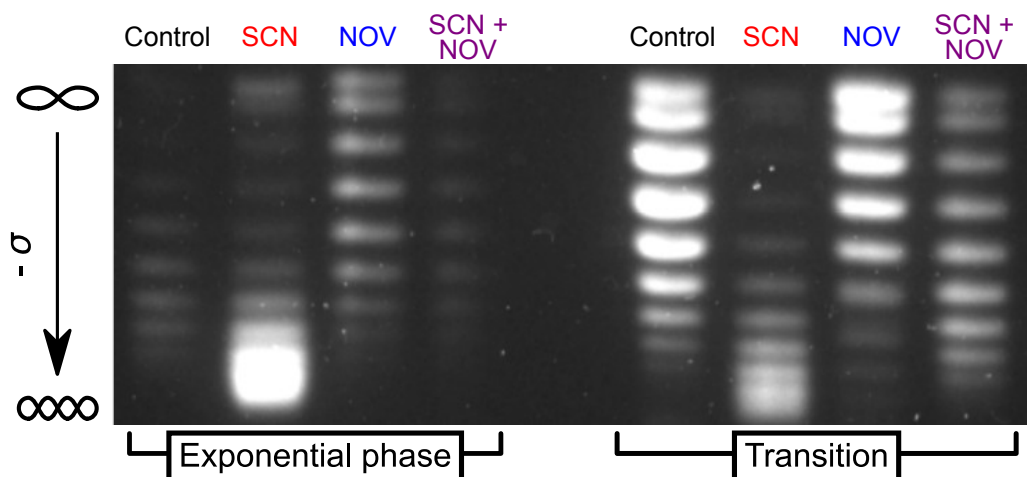
**Fig. S7:** Effect of topoisomerase I expression level on survival rate in presence of seconeolitsine. The *D. dadantii* topoisomerase I gene under the control of an IPTG-inducible promoter was inserted into pQE80L and transformed into *E. coli* NM522 cells. The survival rate is defined as the proportion of growing colonies observed on plates containing 125  $\mu$ M seconeolitsine vs without seconeolitsine. In the absence of IPTG, *D. dadantii* topoisomerase I is only expressed at a basal level and the survival rate on plates containing seconeolitsine is around 30%. In the presence of 0.5 mM IPTG, *D. dadantii* topoisomerase I is overexpressed and the survival rate is significantly higher (73%,  $P=0.0015$ ).



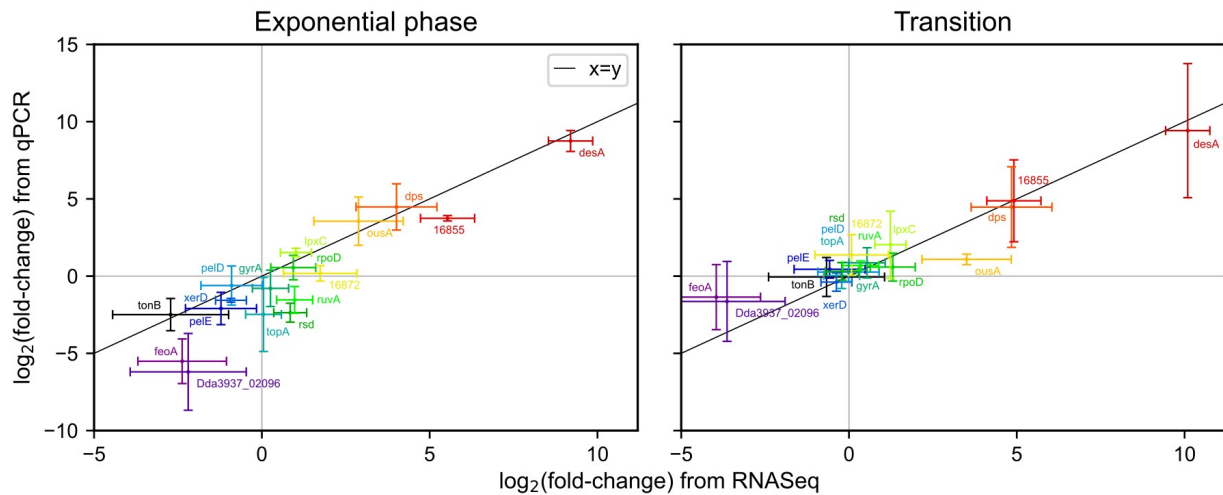
**Fig. S8:** Growth curves of *D. dadantii* treated with a seconeolitsine shock at 50  $\mu$ M during exponential growth (vertical line), exhibiting no significant effect (multiple black and grey lines indicate replicates of treated and untreated samples, respectively, and are almost indistinguishable). The chosen concentration thus minimises the metabolic effect of the shock, while inducing a significant increase in SC (Fig. 3). The experiment was carried in M63+glucose (A) and LB (B) culture media.



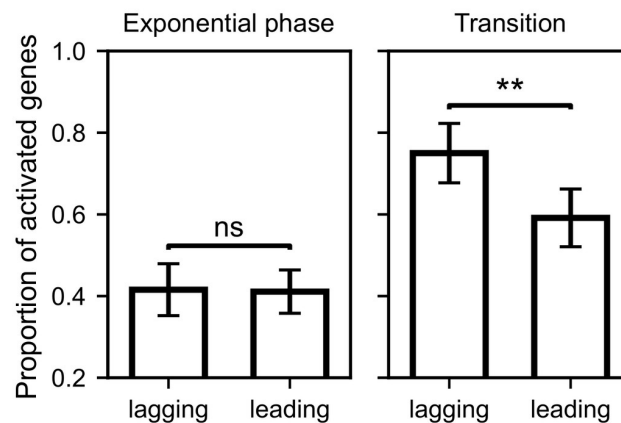
**Fig. S9:** Quantification of topoisomer distribution of pUC18 plasmids isolated from *D. dadantii*. Cells were treated with seconeolitsine (50  $\mu\text{M}$ ), novobiocin (100  $\mu\text{g.ml}^{-1}$ ) or both, in exponential (left) and transition to stationary phase (right). The most relaxed fractions were not fully resolved on the chloroquine gel, hence the discontinuity indicated in the x-axis.



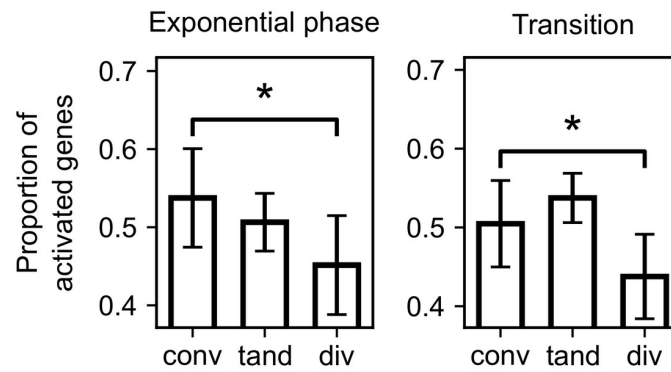
**Fig. S10:** Agarose-chloroquine gels of pUC18 plasmids isolated from *E. coli* NM522 cells in the exponential phase (left) or at the transition to stationary phase (right), 15 min after antibiotic shocks (seconeolitsine 50  $\mu\text{M}$ , novobiocin 100  $\mu\text{g.ml}^{-1}$ , combination of both, from left to right). The downward migration increases with the negative SC level. The results are qualitatively similar to those obtained in *D. dadantii*, with a stronger effect of seconeolitsine.



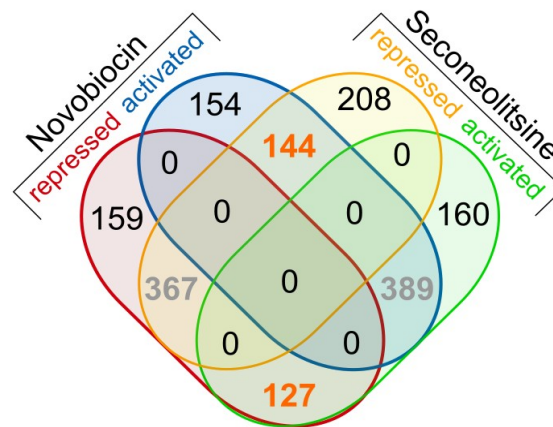
**Fig. S11:** Comparison of gene responses obtained with qRT-PCR and RNA-Seq to seconeolitsine treatment in exponential (left) and transition to stationary phase (right). Pearson's correlation coefficients reached 0.904 with p-value of  $6.23e^{-7}$  (exponential phase) and 0.917 with p-value of  $2.32e^{-7}$  (transition).



**Fig. S12:** Proportion of activated genes (among differentially expressed ones) on the lagging vs leading replicative strand in exponential phase and at the transition to stationary phase ( $P=0.0015$ ).



**Fig. S13:** Proportion of activated genes (among differentially expressed ones) on convergent, tandem and divergent genes following novobiocin treatment in *D. dadantii* in exponential phase ( $P=0.030$ ) and at the transition to stationary phase ( $P=0.044$ ) (3).



**Fig. S14:** Venn diagram of the response of *S. pneumoniae* genes to novobiocin and seconeolitsine treatments in exponential phase, computed from the data of (4). The number of genes responding in opposite directions to the two drugs are indicated in orange, and those in the same direction in gray.

**Tab. S1:** List of primers used in PCR.

Gene / plasmid name	Accession number	Forward primer	Reverse primer
<i>ryhB</i>	Dda3937_04678	GACTGAGAATTCGCGTTCCGGGAAAGC CAAA	CATAGCATGCTAAAAAAAAGCCAGCAC CCGAGC
<i>tonB</i>	Dda3937_00565	GACGCCAGAACCGGAAATAC	CGTTTCCTTCGGTGGTTCC
<i>ousA</i>	Dda3937_00791	GCAACACGTGGAGAAACTGG	CTTGGCATGTAGGTCAGCAG
<i>ruvA</i>	Dda3937_01621	CGGCGTCGGATATGAAGTTC	CACGCCGTTCACTTTGATCA
<i>rsd</i>	Dda3937_00236	ACGCTGGATGAATTTTGCCA	ATTTGCTGGGTGTTGCTACG
<i>dps</i>	Dda3937_02884	ACACGAGATGCTGGATACCT	CGTTGTGGATATCAAGCGGG
<i>xerD</i>	Dda3937_02296	GCTGCCGAAAGATCTGACG	TGTTGCCCTTCCCTATCACG
<i>desA</i>	Dda3937_00786	CCATCAGCCAGCAAACCTAC	GCGAAGGTCAGATCCAGTTG
	Dda3937_01473	GGCGGAACAGAACCTTAACC	CGCTTTCACGTACGGATACC
	Dda3937_01484	AACGGAACAAGATCAGCTGC	AGGTCATACAGCTCGTTCCA
<i>lpxC</i>	Dda3937_02232	ACGGTAAGCGCATTCTGGTA	CATCCAGCAGCAGGTAGACA
<i>pelD</i>	Dda3937_03372	TTGTGGAAGGTAACGCGCAGTTTG	ATGGCAAATTCACCAACGGCTCTC
<i>pelE</i>	Dda3937_03371	AGCGAATTCAAAGCAGCACT	GGCGTTTCGATGTACAGGTT
<i>topA</i>	Dda3937_00587	GAGCAAGAAAAGCGCTGAAG	TTGCGAGATAGACGTGATCG
<i>rpoD</i>	Dda3937_03534	CCGACCGCTATTGATGAAAT	CACAGTTCGACGCACAGTTT
<i>gyrA</i>	Dda3937_01774	GGTAAATATCACCCGCATGG	TCTCCAGATCGGAAAGCAGT
<i>feoA</i>	Dda3937_02095	GGGGTT TCAGAAAGGATCGG	AGAATCAAATCTGCGCTTCG
	Dda3937_02096	TCAAACGCTGCT TGATCTCC	TTGCCCATCTCCAGTTG AT
<i>rpoA</i>	Dda3937_01515	AAACCGCGCCTGGTAGATA	CCTTTCAGGTTGAGCAGGAT
<i>topA</i>	Dda3937_00587	CACCATCACCATCACGGATCCATGGGTA AAGCTCTCGTTATCGTC	AAGCTCAGCTAATTAAGCTTCTCGAGTT AGCGGCTGCTTCCACCCAC
pQE80L		AAGCTTAATTAGCTGAGCTTGGACTCC	GGATCCGTGATGGTGATGGTGATGCGAT CC

**Tab. S2:** Lists of *D. dadantii* genes responding to seconeolitsine treatment in exponential and transition to stationary phase.

**Tab. S3:** Lists of *D. dadantii* genes (locus tags and gene names or ASAP identifiers (5)) significantly affected by seconeolitsine and novobiocin treatment.



## References

1. Robert,X. and Gouet,P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, 42, W320-324.
2. García,M.T., Blázquez,M.A., Ferrándiz,M.J., Sanz,M.J., Silva-Martín,N., Hermoso,J.A. and de la Campa,A.G. (2011) New alkaloid antibiotics that target the DNA topoisomerase I of *Streptococcus pneumoniae*. *J. Biol. Chem.*, 286, 6402–6413.
3. El Houdaigui,B., Forquet,R., Hindré,T., Schneider,D., Nasser,W., Reverchon,S. and Meyer,S. (2019) Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Res.*, 47, 5648–5657.
4. Ferrándiz,M.-J., Martín-Galiano,A.J., Arnanz,C., Camacho-Soguero,I., Tirado-Vélez,J.-M. and de la Campa,A.G. (2016) 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.*, 44, 7292–7303.
5. Glasner,J.D., Liss,P., Plunkett,G., Darling,A., Prasad,T., Rusch,M., Byrnes,A., Gilson,M., Biehl,B., Blattner,F.R., *et al.* (2003) ASAP, a systematic annotation package for community analysis of genomes. *Nucleic Acids Res.*, 31, 147–151.

## 2.4 L'effet du surenroulement sur la virulence de *Dickeya dadantii*

Le SC de *D. dadantii* semble jouer un rôle de senseur des conditions environnementales et de coordinateur de l'expression des gènes de virulence pendant l'infection. En particulier, les gènes qui codent pour les CWDEs sont affectés. Les gènes *pelE* et *prtB* sont à la fois réprimés par la novobiocine et la séconéolitsine tandis que *celZ* est réprimé par la novobiocine mais activé par la séconéolitsine [72, 255]. Afin d'évaluer l'effet de ces antibiotiques sur la virulence, j'ai infecté des feuilles d'endives par des cultures de *D. dadantii* traitées ou non avec l'un de ces antibiotiques en concentration sublétales, en suivant le protocole présenté en Figure 2.5A.

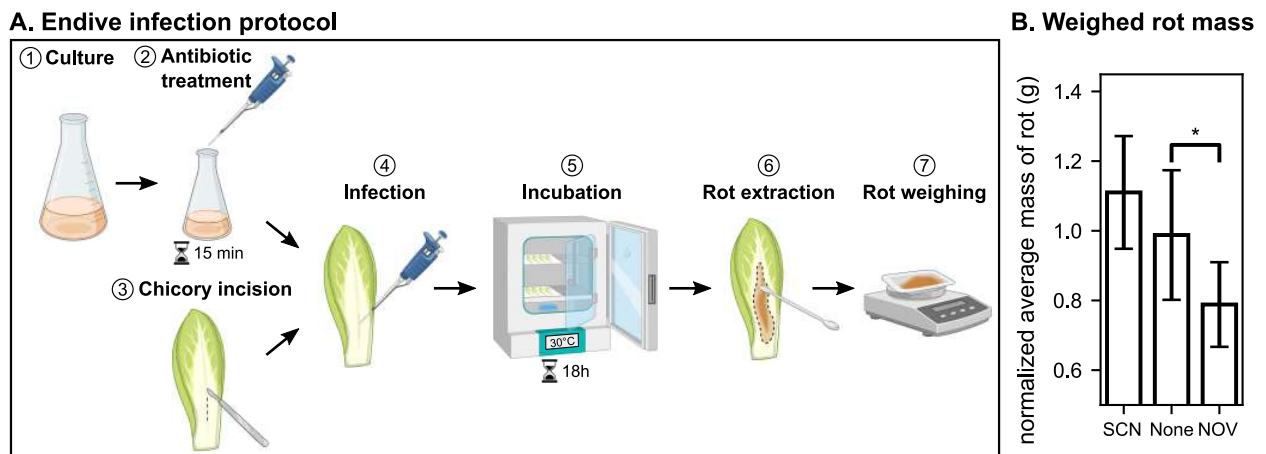


FIGURE 2.5 – (A) Protocole de l'infection de feuilles d'endive par *D. dadantii*, schématisé avec BioRender. (1) Culture des bactéries en milieu riche LB jusqu'à  $DO_{600nm}$  0,25. (2) Ajout facultatif d'un antibiotique (séconéolitsine ou novobiocine) dans la culture pendant 15 minutes. (3) Incision des feuilles d'endives sur 2 cm avec un scalpel, sans les transpercer. (4) Ajout de 5  $\mu$ L de culture dans chaque incision. (5) Incubation des feuilles infectées à 30°C pendant 18h dans des bacs dont les fonds sont tapissés de feuilles de papier absorbant humidifiées et en recouvrant les bacs de film étirable. Un bac d'eau est également présent dans l'étuve. (6) Prélèvement de la pourriture molle en grattant la feuille avec une spatule. (7) Pesée de la pourriture molle. (B) Masse moyenne de pourriture par feuille d'endive infectée par *D. dadantii* après traitement à la séconéolitsine (SCN), à la novobiocine (NOV) ou sans traitement (None). La barre d'erreur représente l'intervalle de confiance à 95%.

J'ai effectué 3 réplicats biologiques de 5 feuilles d'endives (soit 15 feuilles d'endives au total) pour chaque condition (avec bactéries sans traitement, avec bactéries et novobiocine, avec bactéries et séconéolitsine, sans bactérie mais avec novobiocine et sans bactérie mais avec séconéolitsine). A partir des masses de pourriture de chaque feuille, j'ai pu calculer la masse moyenne pour chaque condition puis j'ai normalisé cette quantité par la moyenne obtenue sans antibiotique. J'ai ainsi pu noter (Fig. 2.5B) que la masse de pourriture molle obtenue est significativement diminuée par le traitement avec la novobiocine. Ce traitement semble

donc affecter la virulence, ce qui est cohérent avec la répression des gènes de virulence, dont les gènes *pel* qui codent pour les pectinases (principales enzymes de dégradation des parois cellulaires végétales), par cet antibiotique. Le traitement à la séconéolitsine ne semble quant à lui pas avoir d'effet significatif sur la masse de pourriture récoltée. Cette absence d'effet peut cependant être due aux problèmes de solubilité de la séconéolitsine. En effet, bien que la séconéolitsine soit soluble dans le DMSO, elle précipite lorsque le DMSO est trop dilué ou en cas de baisse de température. Il est donc possible que les conditions de ce protocole d'infection ne permettent pas d'évaluer l'effet de l'augmentation du niveau de surenroulement sur la virulence.

En conclusion, ce test sur endives nous a permis de confirmer que le relâchement de l'ADN provoqué par l'inhibition de la gyrase par la novobiocine affecte la virulence de *D. dadantii*. L'effet de l'inhibition de la topoI sur la virulence reste quant à lui à démontrer en adaptant le protocole aux contraintes de solubilité de la séconéolitsine.

## Chapitre 3

# Etude de l'expression du génome bactérien selon son organisation spatiale avec le package GRATIOSA

### 3.1 Le développement du package GRATIOSA : Genome Regulation Analysis Tool Incorporating Organization and Spatial Architecture

Depuis plusieurs années, mon laboratoire réalise et publie des analyses transcriptomiques et génomiques [98, 177, 255]. Ces analyses ont été effectuées en utilisant des données omiques extraites de la littérature, de bases de données publiques et de données obtenues par notre équipe et nos collaborateurs. Ces données sont ensuite traitées avec un package développé en langage Python au sein de mon équipe. L'objectif de ce package est de pouvoir confronter des données de différents types dont des données transcriptomiques et diverses annotations des chromosomes bactériens. Dans le cadre de ma thèse, j'ai simplifié, complété et documenté ce package afin qu'il puisse aider un plus grand nombre d'utilisateurs dans leurs analyses statistiques de l'expression du génome bactérien selon son organisation spatiale.

#### 3.1.1 Présentation du package initial

A mon arrivée dans l'équipe, de nombreuses données d'expression ainsi que des annotations génomiques et fonctionnelles avaient été rassemblées au sein d'une base de données pour différents organismes dont *E. coli*, *D. dadantii*, *B. subtilis* et *S. pneumoniae*. Des programmes avaient également été développés pour pouvoir importer puis étudier ces données. Ces programmes sont à l'origine du package GRATIOSA (Genome Regulation

Analysis Tool Incorporating Organization and Spatial Architecture). Ils sont organisés selon le paradigme de programmation orientée objet. Ce paradigme est basé sur les concepts de classes et d'objets. Un objet représente une instance concrète d'une classe et possède des attributs et des méthodes. Les attributs sont des variables associées à cet objet et définies dans la classe. Les méthodes sont des fonctions définies dans la classe et qui peuvent notamment permettre d'initialiser ou de modifier les attributs de l'objet. Par exemple, le génome de *D. dadantii* est un objet de la classe *Genome*. Cet objet a notamment comme attributs : *seq* la séquence et *length* la taille du génome. Ces deux attributs sont initialisés lorsque la méthode *load\_seq* est appelée sur l'objet *Genome*.

La version initiale du package, programmée en Python 2, comportait les classes *Genome*, *Gene*, *TSS*, *TTS* et *TU*. Dès cette version, le package permettait de charger les données transcriptomiques, obtenues par RNA-Seq ou avec des puces à ADN, ainsi qu'un grand nombre d'annotations sur les organismes dont les positions des gènes, les séquences des promoteurs et les facteurs  $\sigma$  associés. Ainsi, cette version du package permettait, en chargeant les données en attributs des objets, de faciliter la comparaison des données pour répondre à des questions telles que : les gènes les plus exprimés en phase exponentielle ont-ils une séquence promotrice particulièrement riche en AT? Les gènes différentiellement exprimés lors d'un choc antibiotique sont-ils majoritairement des gènes divergents, convergents ou tandem (classification décrite au Chapitre 1, section 1.1.4)? Cette comparaison était facilitée par la structuration apportée par le package puisque un objet *Gene* a notamment comme attributs une valeur d'expression par condition, une séquence promotrice et une orientation. Cette version du package ne contenait cependant ni test statistique, ni fonction graphique. Les analyses étaient donc effectuées avec des scripts, non incorporés au package, écrits selon l'étude à réaliser.

### 3.1.2 Les objectifs du package GRATIOSA

Dans le but d'étudier l'expression du génome bactérien selon son organisation spatiale, le package GRATIOSA permet de combiner un grand nombre de données (Fig. 3.1, partie gauche) apportant des informations sur :

- l'organisation génomique : annotations des gènes, positions des promoteurs, coordonnées des unités de transcription (TUs)...
- l'organisation structurelle du chromosome : données de type ChIP-Seq comme les données de cartographie du SC présentées au Chapitre 1 partie 1.5.2, informations locales obtenues avec diverses méthodologies comme par exemple les positions des frontières entre les domaines topologiques identifiées par Hi-C,
- l'expression du génome dans différentes conditions.

Du fait de l'hétérogénéité de ces données, ces analyses sont généralement réalisées avec des scripts d'analyse propres à chaque utilisateur. Avec GRATIOSA, nous proposons un package qui facilite la réalisation d'analyses statistiques et quantitatives reproductibles, en automatisant l'importation, la combinaison des données et l'analyse statistique en générant des fichiers et figures présentant les résultats de cette analyse.

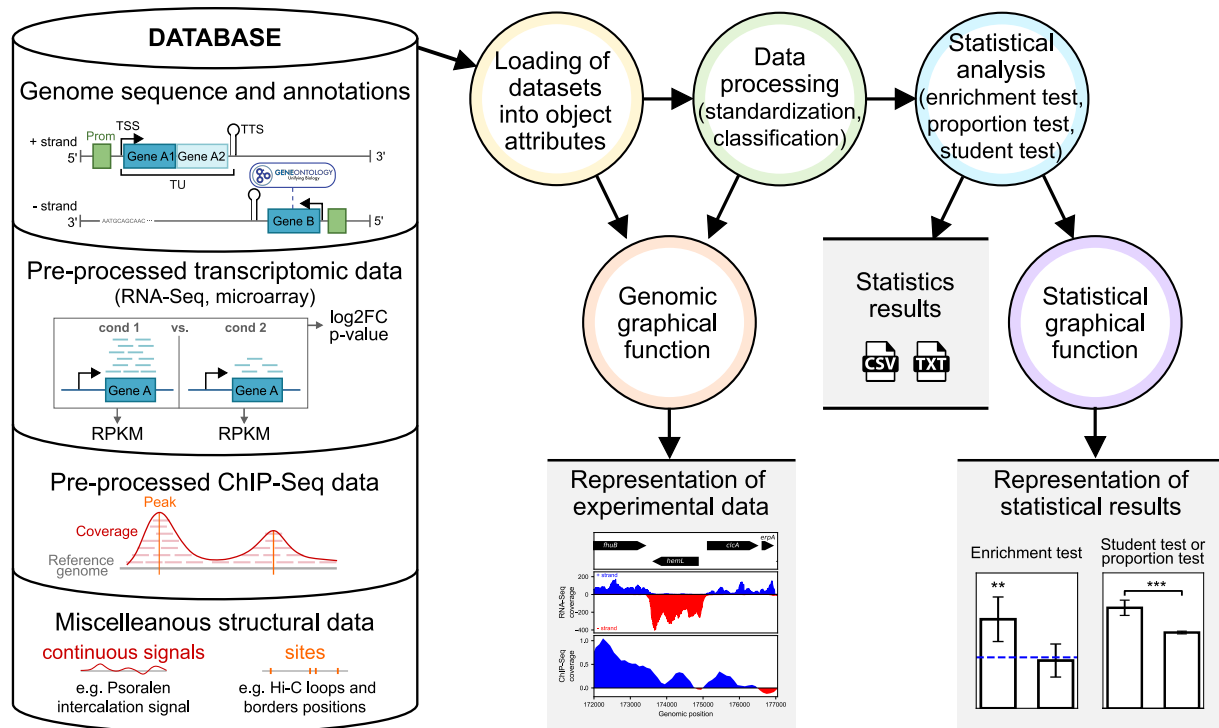


FIGURE 3.1 – Le package GRATIOSA et sa base de données. La partie gauche de la figure présente les différents types de données pouvant être analysés avec le package. La partie droite présente les étapes principales et les sorties fournies par le package. Figure de l'article en cours de soumission.

Pour ce faire, le package GRATIOSA effectue les étapes suivantes (Fig. 3.1) :

— **Importation des données en tant qu'attribut des différents objets.**

Généralement, les attributs chargés ont l'un des formats suivants :

- (1) une valeur ou une caractéristique par position génomique,
- (2) une valeur ou une caractéristique par gène,
- (3) une liste de positions génomiques

— **Traitement des données**

A partir des attributs chargés, de nouveaux attributs peuvent être créés. Par exemple, à partir de données d'expression dans deux conditions différentes, des expressions différentielles peuvent être déterminées. De même, les signaux chargés peuvent être lissés. Par ailleurs, les données quantitatives peuvent être catégorisées afin de les analyser en tant que données qualitatives.

#### — Analyse statistique

L'implémentation de tests statistiques dans le package GRATIOA permet de garantir la reproductibilité et la traçabilité des analyses réalisées grâce à la génération de fichiers de sortie standards contenant les résultats des analyses.

#### — Représentations graphiques

Pour guider l'analyse, réaliser des contrôles intermédiaires et illustrer les résultats, il est utile de pouvoir visualiser graphiquement les données et les résultats.

Certaines de ces étapes dont les tests statistiques et les fonctions graphiques n'étaient pas présentes dans le package initial et je les ai donc implémentées dans le package GRATIOA. J'ai également profité de cette mise à jour du package pour effectuer le passage à Python 3 et pour modifier ou éliminer les fonctions obsolètes. J'ai pour cela réalisé un schéma général de l'ensemble des fonctionnalités du package afin d'identifier les dépendances et fonctions essentielles. J'ai également homogénéisé le format des fichiers d'entrée et de sortie.

### 3.1.3 L'implémentation de nouvelles fonctions

Dans le but d'avoir un package complet, permettant l'uniformisation et la reproductibilité des analyses, de l'importation des données jusqu'à la réalisation des tests statistiques (Fig. 3.1), j'ai implémenté plusieurs nouvelles fonctions.

#### Les fonctions d'importation de nouvelles données

Dans la version initiale, les données d'expression et les données transcriptomiques étaient chargées en attributs de l'objet *Genome* associé à l'organisme étudié. Lorsque j'ai décidé d'ajouter de nouvelles données expérimentales, dont les données CHIP-Seq et d'autres données structurales, j'ai choisi de créer une classe par grand type de données (Fig. 3.1, partie gauche). Le package GRATIOA contient par conséquent les nouvelles classes *Transcriptome*, *Chipseq* et *HiC*. Les objets de la classe *Transcriptome* peuvent notamment avoir des données d'expression différentielle en attributs. La classe *Chipseq* permet l'importation de données de couvertures (signaux continus) ou de positions (sites) qui peuvent par exemple correspondre aux sites de fixations d'une protéine sur l'ADN. D'autres données structurales peuvent être ajoutées, que ce soient des signaux continus ou des listes de positions discrètes, notamment dans la classe *HiC*.

Grâce au format standard appliqué aux données dès leur importation, les données peuvent facilement être traitées. Par exemple, les signaux ajoutés en attributs des objets *Chipseq* sont des listes avec une valeur par position génomique. Un tel signal peut facilement être moyenné, sur des fenêtres glissantes ou non, à l'aide des méthodes associées à la classe *Chipseq*, et former ainsi un nouvel attribut. Il est tout de même important de noter que

ces fonctions ont pour but d'intégrer les données dans un format facilitant leur analyse; elles n'effectuent pas les pré-traitements standards tels que l'alignement sur le génome, le calcul de l'expression différentielle ou la détermination de pics (*peak calling*).

### Les fonctions statistiques

Pour faciliter les analyses statistiques, j'ai tout d'abord implémenté des fonctions intermédiaires fréquemment utilisées dans le cadre de nos analyses. La fonction la plus utilisée est la classification. Cette fonction permet de convertir des données quantitatives en données qualitatives. Dans le cas de données d'expression par exemple, cette fonction permet d'obtenir la liste des gènes correspondants à chaque décile d'expression. La classification peut être réalisée selon des seuils définis ou en catégories de tailles choisies.

Trois tests statistiques peuvent ensuite être réalisés avec le package GRATIOSA :

— **Le test d'enrichissement**

Le test d'enrichissement permet de contrôler la sur-représentation d'une caractéristique au sein d'une catégorie. Ce test permet par exemple de rechercher un enrichissement fonctionnel significatif dans une liste de gènes différentiellement exprimés. En d'autres termes, dans cet exemple, la proportion de gènes associés à une caractéristique fonctionnelle parmi les gènes différentiellement exprimés est comparée à cette même proportion parmi la totalité des gènes. Si la proportion parmi les gènes différentiellement exprimés est significativement supérieure à celle parmi l'ensemble des gènes, la fonction est considérée comme enrichie parmi les gènes différentiellement exprimés. Autrement dit, le nombre de gènes associés à cette fonction parmi les gènes différentiellement exprimés est supérieur au nombre qui serait obtenu avec une répartition aléatoire. Il est à noter que ce test évalue uniquement la sur-représentation et non la sous-représentation d'une caractéristique.

— **Le test de proportion**

Comme son nom l'indique, le test de proportion permet de tester la différence entre des proportions. Par exemple, il permet de comparer les proportions de gènes différentiellement exprimés parmi les gènes convergents et les gènes divergents.

— **Le test de Student**

Le test de Student permet de tester la significativité de la différence entre les moyennes de différentes catégories. Il peut par exemple permettre de comparer le niveau moyen d'expression des gènes convergents et des gènes divergents.

J'ai implémenté l'ensemble de ces tests en utilisant les modules Python standards tels que `scipy.stats` et `statsmodels.stats`. Afin de permettre la reproductibilité et le suivi des tests, chaque test effectué génère un fichier résultat contenant *a minima* les proportions ou les moyennes testées, l'intervalle de confiance à 95% et la valeur-p associée à chaque test.



### Les fonctions graphiques

Les résultats des tests statistiques peuvent ensuite être visualisés sous forme de diagrammes en barres grâce aux fonctions graphiques incluses dans le package (Fig. 3.1, encart en bas à droite). Ces diagrammes représentent les proportions ou les moyennes, les intervalles de confiance à 95% et les résultats des tests statistiques sous forme d'étoile de significativité (\* pour les valeurs- $p < 0,05$ ; \*\* pour les valeurs- $p < 0,01$ ; et \*\*\* pour les valeurs- $p < 0,001$ ). Pour les tests d'enrichissement, une barre horizontale indique également la proportion attendue. Dans le cadre d'une étude d'enrichissement fonctionnel, cette barre correspond à la proportion des gènes annotés pour cette fonction parmi l'ensemble du génome.

Outre ces diagrammes en barres, j'ai implémenté d'autres fonctions graphiques dans le package afin de visualiser des signaux dans leur contexte génomique (Fig. 3.2). Ces fonctions permettent de tracer une couverture de signal ChIP-Seq ou une couverture RNA-Seq. La couverture RNA-Seq correspond au nombre de lectures alignées sur chacun des brins. Cette fonction graphique peut être utilisée avec ou sans l'annotation des gènes et peut permettre de visualiser l'ensemble du génome ou une région précise.

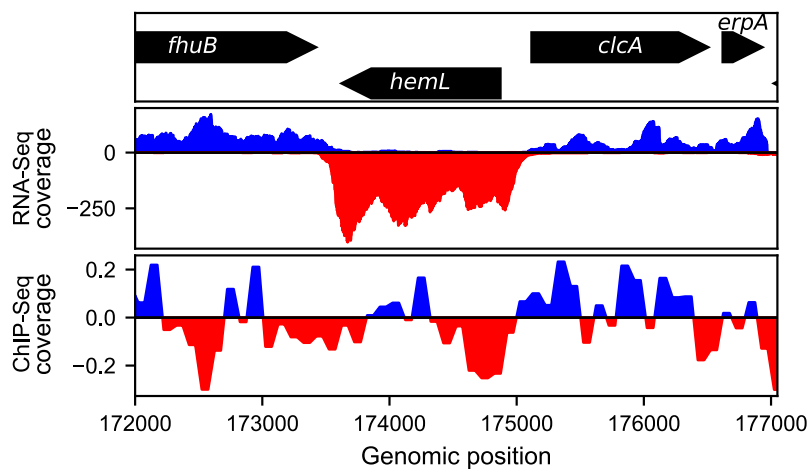


FIGURE 3.2 – Visualisation d'une couverture RNA-Seq et d'une couverture ChIP-Seq dans leur contexte génomique. Figure créée avec le package GRATIOSA à partir de données d'expression en croissance exponentielle et d'un signal de Psora-Seq chez *E. coli*. La couverture RNA-Seq correspond au nombre de lectures alignées sur chaque brin. Le nombre de lectures alignées sur le brin direct est représenté en bleu au dessus de l'axe des abscisses, le nombre de lectures alignées sur le brin indirect est en rouge sous l'axe des abscisses. Le graphique ChIP-Seq représente le  $\log_2$  du ratio entre les signaux avec et sans sélection. Les valeurs négatives du  $\log_2$  de ce ratio sont représentées en rouge, les positives en bleu.

### 3.1.4 La documentation du package GRATIOSA

La version initiale du package était très peu documentée. Elle n'était utilisée que par quelques utilisateurs, au sein de notre équipe, et développée selon les besoins du moment, sans réelle documentation. Dans l'objectif de pouvoir diffuser ce package, au sein du laboratoire mais également publiquement, il était essentiel de palier ces lacunes de documentation. J'ai donc documenté chaque méthode et chaque fonction afin d'expliciter :

- son contexte, son objectif et les dépendances aux autres méthodes ou fonctions grâce à une description de quelques lignes,
- la liste des arguments obligatoires et optionnels en précisant le format et le contenu attendus pour chaque argument,
- les sorties produites par la fonction, les attributs créés ou modifiés par la méthode, en précisant les formats,
- les pré-requis et précautions à prendre lors de l'appel à cette méthode ou fonction,
- un exemple d'utilisation intégrant si nécessaire le chargement de modules et la réalisation des étapes précédant l'appel à cette fonction.

J'ai également ajouté une description pour chaque module et chaque classe. Une fois ces éléments ajoutés en commentaires (*docstring*) de chaque classe, module, fonction et méthode, j'ai utilisé l'outil Sphinx pour générer une documentation générale. Cette documentation est accessible à l'adresse suivante : <https://gratiosa.rtf.io>.

En parallèle et dans le cadre des applications illustrées dans l'article ci-après, j'ai également rédigé des exemples d'utilisation du package GRATIOSA au format Jupyter Notebook afin de faciliter la prise en main du package. Ces exemples seront publiés avec l'article et sont dès à présent disponibles sur GitHub : <https://github.com/sammeyer2017/GRATIOSA>.

### 3.1.5 Article

Nous espérons publier ce package prochainement en illustrant son utilisation sur un cas d'étude de l'organisation du chromosome bactérien. Pour cette étude, nous avons choisi d'analyser les données de ChIP-Seq des topoisomérases de *E. coli* présentées au Chapitre 1, partie 1.5.2 et publiées par Sutormin et al. [197, 191]. Notre analyse a mis en évidence une organisation du chromosome à deux échelles distinctes.

A l'échelle de 10 kb environ, la fixation de la topoI et de la gyrase est fortement affectée par la transcription. Nous avons observé un profil cohérent avec le modèle du *Twin-supercoiled domain* (voir Chapitre 1, 1.5.1). En amont des TUs fortement exprimées, nous constatons un enrichissement de la fixation de topoI, suggérant une accumulation de SC<sup>-</sup>. En aval de ces TUs, nous observons un enrichissement de la fixation de gyrase, pouvant être dû à une accumulation de SC<sup>+</sup>.

Nous avons également remarqué la présence d'une autocorrélation du signal de gyrase à plus grande échelle (100 kb) environ. La présence de cette autocorrélation pourrait indiquer l'existence d'une organisation chromosomique à cette échelle. Il est d'ailleurs intéressant de noter que cette échelle correspond à la taille des CIDs identifiés chez *E. coli* [57] (voir Chapitre 1, section 1.1.5).

# Spatial analysis of bacterial transcriptome and chromosome structural data with GRATIOSA

Maïwenn Pineau, Raphaël Forquet, Sylvie Reverchon, William Nasser,  
Florence Hommais, Sam Meyer<sup>1</sup>

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

<sup>1</sup>To whom correspondence should be addressed. Tel: +33 4 72 43 85 16; Email: [sam.meyer@insa-lyon.fr](mailto:sam.meyer@insa-lyon.fr)

---

## Abstract

Bacterial chromosome compaction via DNA supercoiling (SC) impacts DNA accessibility, affecting replication and gene expression. High-throughput data like RNASeq and ChIP-Seq offer insight on the interplay between genomic spatial organization and expression. We introduce GRATIOSA, a Python package simplifying data analysis by providing a unified framework for importing, combining various data types, and facilitating subsequent quantitative and statistical analysis.

We exemplify GRATIOSA using recent *Escherichia coli* expression data and ChIP-Seq data of topoisomerase I and DNA Gyrase. According to the twin-supercoiled domain model, local transcription influences SC, with an accumulation of negative (resp. positive) supercoils behind (resp. ahead of) the elongation complex. To manage these mechanical constraints, key regulators of supercoiling, topoisomerase I and DNA Gyrase, must be recruited.

Our analysis of topoisomerases ChIP-Seq confirms the *in vivo* relevance of the twin-supercoiled model, impacting the top 40% highly expressed genes, with increasing topoisomerase recruitment correlating with higher expression levels. Spatial analysis reveals that topoisomerase I recruitment is mainly driven by transcription, characterized by a 10 kb length-scale. In contrast, DNA Gyrase exhibits dual contributions: transcription at the 10 kb scale and a larger 100-200 kb scale, reflecting a chromosomal organization consistent with previous *E. coli* 3D conformation maps.

---

## Introduction

### Analyzing the relation between chromosomal structure and transcriptional regulation

DNA encodes the genetic information of all organisms in the form of a sequence of nucleotides, but it is also an elastic polymer. In bacterial cells, the chromosome is highly compacted through various factors, including DNA supercoiling (SC) and Nucleoid-Associated Proteins (NAPs), which together organize its structure dynamically through looping, bending, and twisting of the DNA<sup>1,2</sup>. In turn, the conformation of the chromosome affects DNA accessibility and therefore its replication as well as gene expression. The structural arrangement of DNA within the cell thus represents an “analog” information source crucial for the regulation of gene expression, which complements the “digital” information encoded by the genetic sequence<sup>3,4</sup>.

Standard models of bacterial transcriptional regulation are based on the “digital” action of transcription factors (TFs), which recognize specific sequences in a distinct subset of promoters<sup>5</sup>. In contrast, models of the “analog” regulation mode are mostly lacking, even though it affects the expression of the genome in a global (albeit smoother) manner<sup>3,4</sup>.

High-throughput data such as RNA-seq, ChIP-seq and Hi-C data provide invaluable new information on the relationship between gene expression and chromosomal organization<sup>6-8</sup>. However, the analysis methodology should then differ from that inherited from pure TF-related regulation. For example, in the latter, a gene is entirely controlled by its own promoter. Accordingly, most softwares currently used in the analysis of differential expression (e.g., DESeq2<sup>9</sup>) and inference of regulatory networks consider each gene independently of its spatial location on the genome. Yet numerous observations suggest that nearby genes are co-expressed even when they are unrelated in classical transcriptional regulatory networks<sup>10</sup>, presumably because they belong to the same topological (or chromosomal interaction) domain, and this feature is evolutionarily conserved. Addressing this spatial form of regulation thus requires new, global, quantitative and statistical methodologies of analysis.

Since the chromosome conformation is organized by proteins (NAPs, topoisomerases, and others), this analysis strongly benefits from the combination of expression data (RNA-Seq) with structural data such as ChIP-Seq and Hi-C. Some existing Python packages, such as GenPipes<sup>11</sup>, pyrpipe<sup>12</sup> and

SnakeChunks<sup>13</sup>, propose an integrated software solution for the analysis of RNA-Seq and ChIP-Seq data; they intend to facilitate the implementation of standard analysis methods within a unified framework (downloading of raw sequencing reads, alignment, differential expression, peak calling, and Gene Ontology enrichment tests), but do not address the spatial effects mentioned above. One of the difficulties is that this “analog” regulation remains a loosely defined notion in terms of mechanisms, and as a result, each study addresses its own precise questions: a few examples include the co-expression between adjacent genes that cannot be explained by usual regulatory mechanisms but can be related to NAP binding<sup>10</sup>, the quantitative and double-sided relationship between topoisomerase I (topoI) binding and gene expression (and orientation) at a scale of a few kilobases<sup>14,15</sup>, and the interplay of 3D conformation, local SC and gene expression at the megabase scale<sup>16</sup>.

With the rapid accumulation of high-throughput structural data, such analyses are growing in number, and are usually carried using handmade analysis scripts, with two major drawbacks: (1) significant efforts are required for importing and comparing data of various types and formats, and (2) reproducibility is difficult to ensure. Here, we address this issue by presenting GRATIOOSA, a new Python package intended to facilitate such quantitative and spatial analyses of expression/structural data. It provides a unified framework for the automatic import and combination of many usual data types, and an integrated and flexible framework for the subsequent quantitative and statistical analyses of these data along bacterial genomes.

### GRATIOOSA: Genome Regulation Analysis Tool Incorporating Organization and Spatial Architecture

GRATIOOSA does not replace usual RNA-Seq or ChIP-Seq pre-analysis tools (read quality check, mapping, differential expression or peak calling, possibly using the integrated packages mentioned above, or others), but rather facilitates the following steps which previously required manual processing, i.e., the integration, comparison and combined statistical analyses of these data. It includes (1) the integration of various data types and formats in a unified framework, allowing direct comparison; (2) an implementation of spatial statistical analyses of the relationship between transcriptional regulation and chromosome organization, greatly reducing the required code length; and (3) an improvement in reproducibility, in particular across different species in order to identify common regulatory mechanisms.

GRATIOOSA is written in Python, and is targeted

to computational biologists. The automated import of data and standardized statistical tests and procedures are combined with the flexibility of Python for developing custom subsequent analyses, depending on the particular regulatory mechanism under investigation. It is primarily designed to treat expression data (RNA-Seq or microarrays), and ChIP-Seq data, but can be used with any type of continuous signals along the genome (other binding signals, mutation rates, binding prediction profiles...) or lists of discrete features (annotated or predicted protein binding sites, Hi-C topological domain borders, ...).

In this article, we present the structure of the software, and illustrate its features through a specific example of “analog” regulation: the quantitative interplay between gene expression and topoisomerase activity along the *Escherichia coli* genome.

### Case study: Exploration of *Escherichia coli* twin-supercoiled domains using topoisomerase ChIP-seq data

SC, i.e., torsion-induced deformation of the DNA double-helix, besides its essential role in the bacterial chromosome compaction, is recognized as a global transcriptional regulator<sup>17-19</sup>. DNA torsional stress modulates the activity of the transcriptional machinery during the initiation, elongation, and termination of transcription<sup>17,20</sup>, in particular by altering the DNA accessibility for the RNA Polymerase (RNAP) and transcriptional regulators. The superhelical level is essentially controlled globally by the action of topoisomerases, mostly topoI and gyrase<sup>21,22</sup>, but is also strongly affected by local transcription, according to the twin-supercoiled domain<sup>23</sup>, leading to the accumulation of negative (resp. positive) supercoils behind (resp. ahead of) the elongation complex, and the local recruitment of topoisomerase to handle these mechanical constraints. This complex, double-sided and dynamical interplay has long been postulated to underpin a global and ancestral regulator of bacterial transcription<sup>17,19,24</sup>, but its precise and quantitative impact *in vivo* remained fuzzy because of a lack of experimental knowledge of the distribution of supercoils in bacterial genomes. Recently obtained ChIP-Seq of topoisomerases<sup>6,15,25</sup> provide invaluable information to better understand the link between transcription and local DNA SC<sup>16</sup>. Which and how many genes are affected by these topological constraints *in vivo*? What is their spatial extension? What is their contribution in global gene regulation? After presenting the GRATIOOSA software, we use it to provide some insights into these questions.

## Materials and Methods

### Framework

GRATIOSA is written in several Python source files, each defining one of the main object classes (Genome.py, Gene.py, GO.py, Transcriptome.py, Chipseq.py, ...) with their attributes (e.g., a Genome object contains Genes and a genomic sequence) and methods (e.g., load\_annotation, load\_seq, ...). In addition, some miscellaneous functions for each class are defined in a separate source file (e.g., intermediate functions for reading a gff file are defined in useful\_functions\_genome.py). Finally, three additional program files contain intermediate functions for statistical tests (stat\_analysis.py), and the graphical representation of genomic data (plot\_genome.py) or of statistical results (plot\_stat\_analysis.py). All methods and functions rely on the packages NumPy<sup>26</sup>, Matplotlib<sup>27</sup>, pandas<sup>28</sup>, pysam<sup>29</sup>, SciPy<sup>30</sup>, and Statsmodels<sup>31</sup>.

Three main types of interdependent objects can be distinguished: the Genome object, the elementary objects (Gene, TU, TTS, TTS objects), and the objects containing experimental data (Transcriptome, ChIP-Seq, Hi-C objects) (Supplementary Fig. S1). An elementary object represents a single gene or a single TU (Transcription Unit), while the Genome object contains the annotation of all elementary objects (i.e., all genes, all TUs). Similarly, the elementary object may have attributes such as expression level or average ChIP-Seq signal for its segment, but the overall expression levels and genomic signals are attributes of the Transcriptome and ChIP-Seq objects, respectively.

### Experimental data pre-processing

GRATIOSA can import experimental data formatted either as continuous signals (i.e., one value per genomic position), discrete sites (i.e., list of genomic positions), or gene-wise values (i.e., one value per gene). The data files follow usual formats generated with standard external tools such as DESeq2, MACS, or deepTools (Table 1).

Functional enrichment analysis using Gene Ontology (GO) terms is also performed using standard description files, namely an annotation file for the analyzed species and a file describing all GO terms (obo, available from the GO website<sup>32</sup>). The annotation file (typically generated by software such as Blast2GO) contains at least two columns, with each row relating one gene with one GO (with multiple lines for each). The obo file can be automatically downloaded with the package.

**Table 1:** Summary table of the data types analyzed by the package (mostly tabular files). The structure (rows and columns) and typical software used to obtain them are shown.

Data type	Rows	Columns	Standard format	Commonly used packages
RNA-Seq coverage	pos.	coverage	bedgraph	Bedtools <sup>33</sup>
RNA-Seq expression	genes	locus, expression level (RPKM or FPKM)	tab	Cufflinks <sup>34</sup> , HTSeq <sup>35</sup>
RNA-Seq differential expression	genes	locus, fold-change, p-value	csv	limma <sup>36</sup> , DESeq2 <sup>9</sup>
Microarray differential expression	genes	locus, fold-change, p-value	csv	limma <sup>36</sup>
ChIP-Seq coverage	pos.	bin start, bin end, coverage	bedgraph	deepTools <sup>37</sup>
ChIP-Seq peaks	sites	position, value	bed	MACS <sup>38</sup>
Hi-C loops positions	sites	first bin position, second bin position	tsv	chromosight <sup>39</sup>
Hi-C borders positions	sites	first bin position	tsv	chromosight <sup>39</sup>

### Database organization

To be analyzed with GRATIOSA, data must be organized in a database with a fixed structure: each directory represents one organism (with a reference genome) and contains multiple subdirectories (Supplementary Fig. S2). The *annotation* subdirectory contains the annotation, usually in gff3 format. All other subdirectories have the same structure, each containing experimental data files and an info file (See Experimental data pre-processing). *chipseq* and *hic* subdirectories are further subdivided, to distinguish between two types of data (peaks and signals in the case of ChIP-Seq, loops and borders in the case of Hi-C), each following the same structure as other subdirectories (see Supplementary Fig. S2).

### Data importation

Data importation is automatic from the annotation files (sequence, annotation). For experimental data present in multiple files (e.g., experimental conditions for gene expression), the import function uses an “info” file, i.e., a tab-separated text file containing the list of data files that will be loaded for analysis (Supplementary Fig. S3). This method enables the automatic selection and import of data from various sources or softwares, whether in standard or custom formats, by specifying some descriptors of each data file. For example, in the case of the *expression* directory describing genome-wide expression levels in various conditions (microarray intensities or typically for RNA-Seq data, RPKM values), the info file contains a list of rows (one per experimental

condition to be loaded), each indicating the condition name, associated file name, column index for gene names/tags, column index for expression data, whether the data is in log format or not, and the separator used in the file.

Depending on the type of data to be analyzed, the corresponding object needs to be initialized first (here Transcriptome, Supplementary Fig. S3, gray part), and then data import is achieved using the appropriate method (e.g., `load_expression` to import expression data onto the Transcriptome object, see Supplementary Fig. S3, yellow part). Intermediate functions (e.g., `useful_function_Genome.read_seq`) are used to read data files, and might be called by the user in unusual cases (e.g., files with very custom formats).

### Data processing

Once data are loaded (typically as Python lists, NumPy arrays, or Pandas dataframes), new attributes can be either added manually or computed from existing ones (Supplementary Fig. S3, green part). For example, ChIP-Seq signals can be binned, smoothed, or averaged. For binning, the user defines a bin size and method (by default, average signal within the bin); for smoothing, the size of the sliding window; for averaging, the list of different signals to be used (usually replicates). Raw or computed signals can be automatically related to Gene objects by computing the average signal over the gene region. Similarly, various attributes (such as orientation, expression level or fold-change, protein occupancy) can be added to Gene objects based on imported Genome, Transcriptome or Chipseq attributes (Supplementary Fig. S1, attributes written in italics). The orientation of a gene or intergenic region is determined by the direction of the neighboring genes (tandem, divergent, convergent). Before performing statistical analyses, quantitative data can be classified into qualitative categories (Supplementary Fig. S3, green part), based on user-defined thresholds or on quantiles (of equal or custom sizes).

### Statistical analysis and outputs

The package offers three standard statistical tests (Supplementary Table S1, Supplementary Fig. S3, blue part): an enrichment test (hypergeometric, with false discovery rate correction), particularly useful for functional enrichment analysis; a proportion test (based on the z-score, equivalent to a  $\chi^2$  test), for example for comparing the proportion of differentially expressed genes in different sets of genes; and a standard Student's t-test for quantitative data, for example for comparing the average expression level or fold-change among different sets of genes. These tests are implemented using

common Python modules, and their results are saved in a csv file.

### Graphical functions

Statistical results can be visualized as bar plots generated using the `matplotlib.pyplot.bar` module, with error bars representing 95% confidence intervals (Supplementary Fig. S3, purple part), and optionally, an annotation of these bar plots with brackets and stars according to their significance level (\* for p-value < 0.05, \*\* for p-value < 0.01, and \*\*\* for p-value < 0.001).

Other graphical functions are implemented in the package to visualize signals, such as ChIP-Seq and RNA-Seq coverages within their genomic context (Fig. 1). RNA-Seq coverage corresponds to the number of reads for each genomic position, on each strand. Depending on the size of the visualization window (from a few genes to the entire genomes), genes on each strand can be annotated.

### Processing of *Escherichia coli* topoisomerase and expression signals

The *Escherichia coli* W3110 topoisomerase signals analyzed in this article were previously published by Sutormin et al., with FPKM values of annotated transcription units (TUs)<sup>6,15</sup>. The mapping of gyrase cleavage sites was obtained using a method called Topo-Seq; we used the datasets obtained with ciprofloxacin as gyrase poison (GEO:GSE117186)<sup>6</sup>. Topoisomerase I binding sites were identified with ChIP-Seq (GEO:GSE181915)<sup>15</sup>, with or without rifampicin treatment (an inhibitor of transcription). For each sample, the coverage was first normalized by its mean, and the log<sub>2</sub> ratios between immunoprecipitation (IP) and control (mock) coverage were computed. The log<sub>2</sub> ratios were then normalized (by their mean) and the resulting values were saved in bedgraph format compatible with GRATIOSA. The replicates of topoI and gyrase signals were averaged for each genomic position (using the `Chipseq.load_signals_average` method), representing the log<sub>2</sub> occupancy by each protein relative to the control data. The signals were then separated into macro-scale and micro-scale contributions (see Results): the macro-scale contribution was obtained by smoothing each signal along 200-kb sliding windows (with `useful_functions_Chipseq.smoothing`); the micro-scale contribution was obtained by subtracting the macro-scale contribution from the raw signal. For expression analyses, we generated approximate expression signals (per genomic position) from the FPKM (per TU) expression values provided<sup>15</sup>.

Pearson correlation coefficients for signal autocorrelation were calculated using the

scipy.stats.pearsonr function by comparing each signal with the same signal shifted in genomic position. Correlation coefficients were computed using the same SciPy function to topoisomerase signals and the expression signal, after the smoothing of these signals at different scales.

## Results

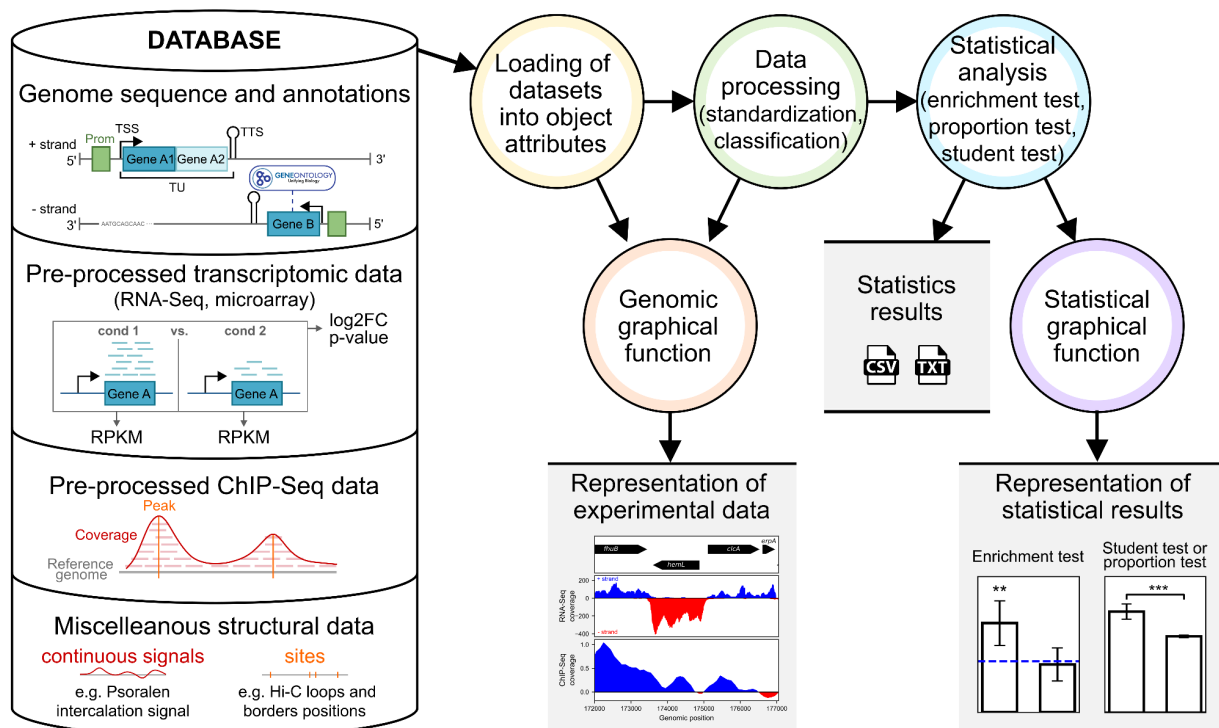
### The GRATIOSA package

GRATIOSA facilitates the integration and comparison of bacterial genomic data types using standardized statistical analyses, and is particularly designed to investigate quantitatively the spatial relationship between transcriptional regulation and chromosome organization. Two basic types of experimental data can be analyzed with the package: expression data (transcriptomics) and structural data (ChIP-Seq or related technologies), and both can be represented either in a continuous (RNA-Seq or ChIP-Seq coverage) or in a discrete manner (gene expression levels, positions of ChIP-Seq peaks). An overview of the package is shown in Fig. 1, and more detailed technical information is given in Materials and Methods. The simplest method to import data with GRATIOSA is to organize the data files (in usual formats) following a predefined structure, classified by organism and data type (Supplementary Fig. S2). They can be accompanied by an information file that enables an automatic selection and importation of these data (e.g.,

a subset of experimental conditions for a particular analysis of expression data). This basic structure can be used to import many other and/or custom types of data, e.g., predicted protein binding (discrete) sites or (continuous) signals, miscellaneous annotated features, positions of Hi-C domain borders, etc.

The GRATIOSA package is typically imported into a Jupyter notebook, and data analysis is executed as Python commands, in three major steps (Fig. 1). Firstly, different objects (*Genome*, *Transcriptome*, *Chipseq*, etc, see Supplementary Fig. S3, gray) are initialized, depending on the user's needs. The required data are then loaded as object attributes (Supplementary Figs. S1 and S3, yellow).

During this second step, the loaded attributes are processed by the software (Fig. 1) and become easily manipulable (mainly as NumPy arrays or Pandas objects), either manually (for custom computations) or using the tools implemented in the package (described in Supplementary Figure S3, green). Among the latter, attributes associated with genomic positions can be scaled to genes (for example, the average ChIP-Seq signal is computed for each gene). Signals can be smoothed, among other options. Genes can be classified according to their orientation, distance, or length (based on the annotation). Thanks to the easily manipulable formats, the user can perform custom additional processing to modify or add attributes, even using manually imported external data. A verification and



**Fig. 1:** Graphical overview of the GRATIOSA package. The left section (database) presents the different types of data processed by the package. The right section shows the main steps and outputs of the package.



graphical exploration of the signals can be performed at the end of the first or second steps using a graphical function that plots signals with the annotated genome. Finally, to prepare for statistical analysis, quantitative data (expression level, protein binding strength, etc) can also be classified, using thresholds or class sizes, e.g., in expression deciles.

The last step is the statistical analysis (Fig. 1, Supplementary Fig. S3, blue) with enrichment or proportion tests (for qualitative attributes) or Student's t-tests (for quantitative comparisons), as described in Materials and Methods. Test results are exported as tables (in csv format) and can be visualized as annotated bar plots created with graphical functions included in the package (Supplementary Fig. S3, purple).

These steps are illustrated in the following Case Study, for which GRATIOSA allowed a drastic reduction in code writing. The associated codes have been deposited in GitHub: <https://github.com/sammeyer2017/GRATIOSA>.

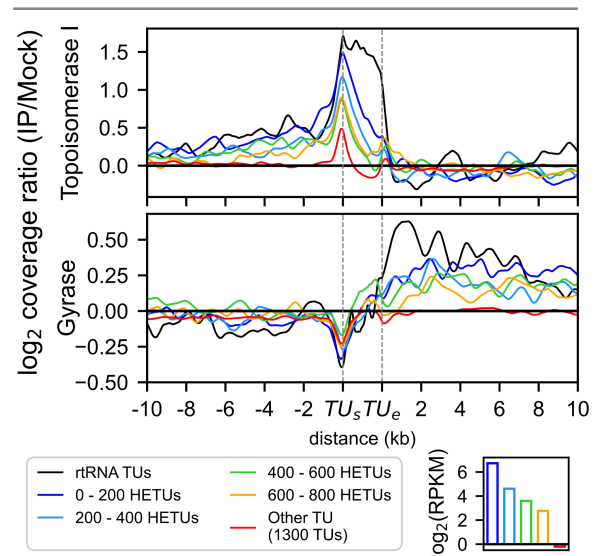
### Case study : Exploration of *Escherichia coli* twin-supercoiled domains using topoisomerases ChIP-seq data

Since topoI preferentially binds to negatively supercoiled regions (where DNA forms transient denaturation bubbles), and DNA gyrase to positive supercoiled regions (forming plectonemes), studying their genome-wide binding and cleavage distributions could provide an insight into the *in vivo* presence, amplitude, and extension of twin-supercoiled domains. These signals have been obtained in *E. coli* by Sutormin *et al.*, using ChIP-seq and Topo-Seq methods respectively<sup>6,15</sup>. These authors found enrichments of topoI upstream of the transcribed region and of gyrase downstream for the 200 most highly expressed transcription units (HETUs) out of approximately 2100 annotated TUs in *E. coli*<sup>15</sup>, in agreement with the twin-supercoiled domain model and thus presumably reflecting an asymmetric buildup of topological deformations. We wondered if this behavior is specific to that minority of genes, or if it affects a larger fraction of the genome, to what distance, and wished to analyze it in more detail. All data import and most classification steps could be achieved with few lines of code with GRATIOSA, allowing the user to focus on the choice and implementation of the specific analyses and plots shown in the following.

We first wished to visualize the twin-domains in more detail. We classified all TUs by expression level and analyzed the average gyrase and topoI coverages (using log<sub>2</sub>-normalized ratios of IP-signals versus mock, with an average level of 0)

from 10 kb before TU start to 10 kb after TU end (Fig. 2).

The topoI binding signal (Fig. 2, top) exhibits a clear gradual increase over 10 kb upstream of the TU start, with a pronounced peak at the TU start. Following the TU start, the signal quickly decreases and becomes close to zero by the end of the TU (except for stable RNAs in black). Within the TU, higher expression levels are associated with higher topoI binding density, consistent with the known colocalization and interactions between topoI and RNAP<sup>15,40,41</sup>. Within stable RNA (tRNA and rRNA) operons (black curve), topoI binding appears relatively flat, potentially indicating an ongoing topoI recruitment throughout the TU, in contrast to all other categories (even strongly expressed). This observation suggests a specific extensive requirement for topoI for those operons. Upstream of the transcription start site, the density of topoisomerase I binding is also markedly dependent on TU expression levels. This enrichment is likely not due to the interaction of topoI with RNA Polymerase, but reflects the negative supercoils generated during elongation, in clear agreement with the twin-supercoiled domain model<sup>23</sup>. *In vivo*, based on the ChIP-Seq data, this



**Fig. 2:** TopoI binding (top) and gyrase cleavage (bottom) signals around TUs in *E. coli* (ChIP-Seq data from<sup>6,15</sup>). Shown signals are normalized log<sub>2</sub> ratios between coverage after immunoprecipitation (IP) and control coverage (mock). TUs were classified based on their expression level, with 200 TUs in each class of highly expressed TUs (around one decile each), one class with all other protein-coding TUs (red), and one class with strongly expressed TUs encoding stable RNAs (tRNA or mRNA, black). By construction, both signals have a genomic average of 0. The bottom right panel shows the average expression level of each expression class, except for stable RNAs that have the highest expression level (but where FPKM do not quantitatively reflect the expression level). TU<sub>s</sub> and TU<sub>e</sub> indicate the start and the end of the TU.

phenomenon affects not only 200 HETUs, but at least the 800 most highly expressed TUs (approximately 40% of the genome). It extends remarkably far: up to 10 kb upstream of the TU for the top 400 most highly expressed TUs, up to 8 kb for the next 200 TUs, and up to 6 kb for the following 200 TUs. Thus, the apparent extension of the twin-supercoiled domain is dependent on the TU expression level. In the 1300 (~60%) weakly expressed TUs (red), no enrichment is observed. After the end of the TU (downstream), the topoI binding is slightly weaker than the genomic average (negative levels for all expression classes), as would be expected from the twin-domain model if positive supercoils inhibit topoI binding.

The gyrase signal (Fig. 2, bottom) exhibits significant differences compared to that of topoI. As expected again, the 10 kb upstream of TUs are less cleaved by gyrase than the genomic average, possibly due to the negative supercoils generated by transcription, with a particularly strong depletion at the beginning of the TU. Within the TU, we observe no enrichment in gyrase binding, suggesting that the enzymes are not required in the immediate vicinity of RNAP (in contrast to topoI). After the TU, the signal increases sharply, then slowly decreases but remains strongly enriched after 10 kb behind the TU, in all classes of the 40% highly expressed TUs, and thus extends further and more uniformly than the topoI enrichment before the TU. This enrichment is clearly increasing with expression strength, whereas the 60% weakly expressed TUs exhibit no enrichment at all. Note that the large fluctuations observed may be attributed to the effect of neighboring TUs, alternative TUs, internal promoters, or sequence biases affecting topoisomerase binding (especially for rRNAs that are present in lower numbers); yet the regularity of several observations suggest that

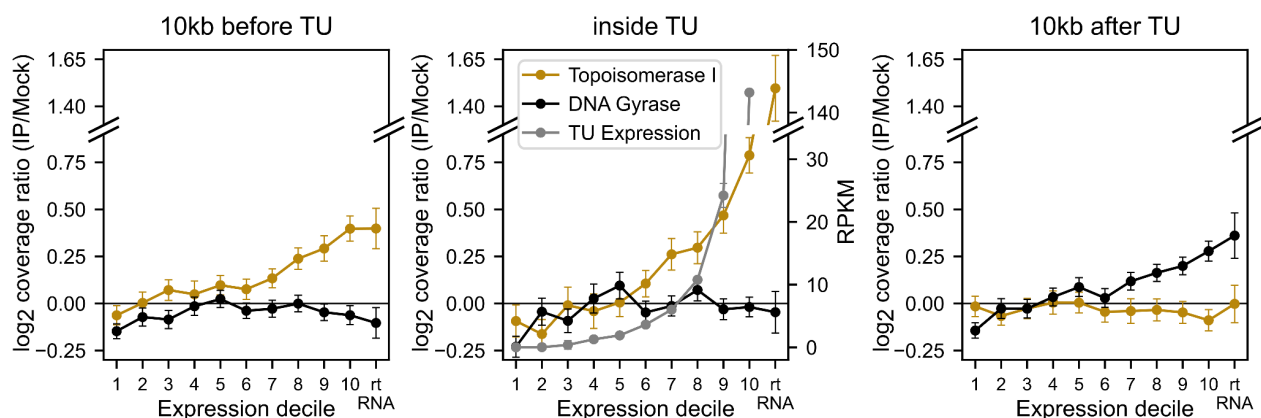
the expression of the central TU is the main factor influencing the shape of these curves.

### Role of gene expression strength and orientation in shaping local supercoiling distributions

The correlation between expression level and twin-supercoiled domain intensity can be further investigated by examining the total binding signals around all TUs, separated in expression deciles, in the 10 kb upstream of the TUs (Fig. 3, left panel), within the TUs (Fig. 3, middle panel, expression levels are indicated in gray), and in the 10 kb downstream (Fig. 3, right panel).

This representation confirms the previous observations, with additional information. Within TUs, the average density of topoI (Fig. 3, gold curve middle panel) exhibits a moderate increase in the first deciles of TU expression levels, which becomes much more pronounced in the last 4 deciles (i.e., the 40% most highly expressed TUs). In the 10 kb region upstream of the TUs (Fig. 3, gold curve left panel), the expression level dependence is very low in the six first deciles, whereas topoI binding steadily increases in the last four deciles. After the TUs (Fig. 3, gold curve right panel), as observed in Fig. 2, the topoI binding signal is weakly negative for all expression levels, suggesting a weak accumulation of positive supercoils.

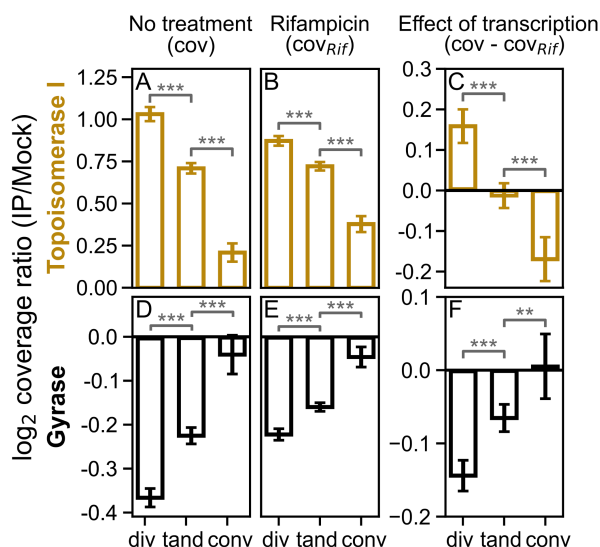
As for gyrase (Fig. 3, black curves), its recruitment seems constant for all TUs, even the most expressed ones, before and within the TU. Before the TUs, the signal is slightly negative, suggesting a presence of negative supercoils. After the TUs, a gradual enrichment of gyrase cleavage sites is observed with a clear increase, once again, in the last 4 deciles, and quite limited before.



**Fig. 3:** Average topoI binding (gold) and gyrase cleavage (black) signals in the 10 kb before (left), inside (middle), and in the 10 kb after (right) TUs, depending on the expression level. TUs are classified in expression deciles, and the average expression level (FPKM) of each decile is shown in gray in the middle panel; operons of stable RNAs are classified separately. Mind the breaks in the vertical axes. Error bars indicate 95% confidence intervals

Together with those of Fig. 2, these results (1) systematically confirm the *in vivo* relevance of the twin-supercoiled model; (2) consistently suggest that these topological constraints mostly affect the 40% most highly expressed genes, with a clear and steady increase in the recruitment of topoisomerases depending on the expression level, whereas the 60% weaker genes are less affected, presumably because they generate moderate mechanical distortions that might be (at least partly) dissipated without local recruitment of topoisomerases; (3) within TUs, *topoI* is recruited in the wake of the elongating RNA Polymerase<sup>15</sup>, particularly close to the promoter, but gyrase is not enriched.

Previous studies have shown that, because of the asymmetry of transcription-induced supercoils, local gene orientation is important for the transcription-SC coupling<sup>14,15,25,42</sup>. The package has built-in functions to classify intergenic regions depending on the orientation of their neighbors (divergent, convergent, tandem), and quantitatively compare how they recruit topoisomerases. We observed (Fig. 4) an enrichment of *topoI* and an under-representation of gyrase between divergent genes compared to convergent genes, as already noted by Sutormin *et al.* following expectations of the twin-supercoiled domain model<sup>15</sup>. By comparing the signals from these regions with and without rifampicin, an inhibitor of transcription, we observed that this difference was still present, to a lesser extent, after the treatment (Fig. 4 B and E). This difference can be explained by differences in base sequence, especially since divergent regions are more enriched in AT-rich promoters (Supplementary Fig. S4), i.e., binding preferences independent of transcriptional activity. It might also reflect an incomplete dissipation of supercoils generated by previous transcriptional activity (before rifampicin treatment) or an incomplete transcriptional inhibition by the drug. By normalizing the untreated coverage (Fig. 4 A and D) with the coverage after rifampicin treatment (Fig. 4 B and E), we erase out any static effect, and confirm a differential and opposite recruitment of topoisomerases, specifically induced by transcriptional activity depending on the local gene orientation (Fig. 4 C and F). The enrichment in *topoI* is also symmetrical in *div/conv* compared to the tandem regions (close to 0, i.e., the genomic average), whereas gyrase exhibits no enrichment in convergent regions (close to genomic average) and a depletion in *div* and tandem region; this difference is difficult to interpret (since intergenic regions contain only a small fraction of the genome), but it suggests that in the short scale of most intergenic regions (<200 bp), the buildup of positive supercoils between convergent genes is not sufficient to strongly attract gyrase, whereas the inhibitory action

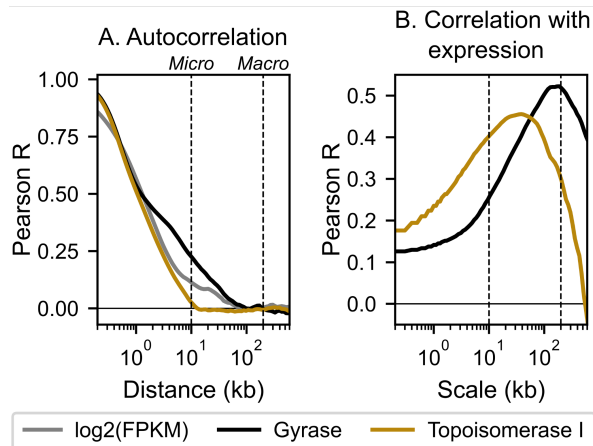


**Fig. 4:** *TopoI* binding sites (gold) and gyrase cleavage sites (black) density in intergenic regions, depending on the orientation of surrounding genes. Each bar represents the average normalized log<sub>2</sub> ratios between coverage after immunoprecipitation (IP) and control coverage (mock) for a specific orientation, calculated after binning intergenic positions into 100 bp bins. Signals A and D were obtained without treatment, B and E were obtained after treatment with rifampicin, a transcription inhibitor, and C and F show the difference between untreated and rifampicin-treated samples. Error bars are 95% confidence intervals.

of negative supercoils (upstream of TUs) efficiently reduces its binding.

#### Extension of twin-domains and multiscale interplay between transcription and chromosome topology

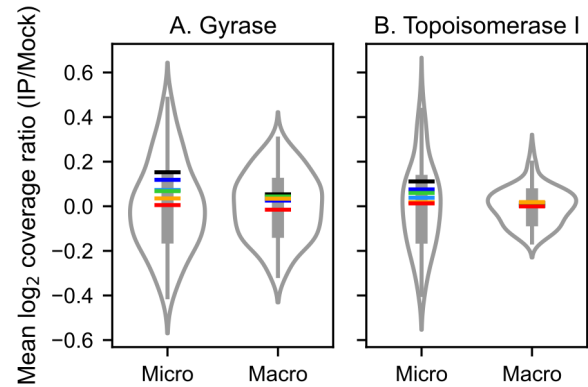
Fig. 2 showed that topoisomerase recruitment was affected at least up to 10 kb before and after TUs, close to the most commonly accepted size for topological domains in *E. coli*<sup>43</sup>. We now look at the extension of topoisomerase binding correlations, possibly reflecting these domains. The binding autocorrelation curves are shown in Fig. 5A. Unsurprisingly, the correlation is strong, for both topoisomerases and for gene expression patterns, below one kilobase, the typical gene length, as expected if short-scale binding variations are related to transcription (and close to the spatial resolution of ChIP-Seq data). The correlation reduces with distance, but in a very different manner for the two enzymes: the characteristic length-scale of the *topoI* binding distribution is 10 kb, above which it is entirely decorrelated; in contrast, at that scale, the correlation of gyrase binding is still quite strong, and vanishes only a full order of magnitude above, after 100 kb. These two separate length-scales were already described in earlier works, and correspond to two different levels of genome organization:



**Fig. 5:** (A) Pearson coefficients for the autocorrelation of gyrase cleavage signals (in black), topoI binding signals (gold), or expression signals (gray) as a function of shift distance. (B) Pearson coefficients for the correlation between gyrase (black) or topoI (gold) signals and expression signal smoothed at increasing scales.

(1) the 10-kb (hereafter quoted “micro”) scale is that of transcription units and their immediate neighboring interactions, possibly within transient topological domains<sup>43</sup>, leading to strong and evolutionarily conserved co-expression<sup>10</sup>; (2) the larger 100-200 kb scale (“macro”) is that of the 31 chromatin interaction domains (CIDs, ranging from 60 to 300 kb, with an average of 150 kb) observed in 3D conformation maps in *E. coli*<sup>44,45</sup>. The curve of gene expression (gray) confirms this separation, with the correlation dropping almost entirely at the micro-scale, but a small (~10%) contribution is still observed up to the macro-scale, possibly induced by the large-scale organization of the chromosome. Accordingly, the correlation of topoisomerase binding with expression is maximal when they are smoothed at length-scales of a few dozen kb (topoI) vs 200 kb (gyrase).

This observation suggests that topoisomerases are recruited based on two (partly) separate contributions in terms of mechanisms and length-scale: (1) micro-scale transcription and (2) macro-scale 3D organization of the chromosome. To test this hypothesis explicitly, we used a signal filtering method to compute how each of these two length-scales contributes to the binding distributions (by smoothing over the macro-scale of 200-kb, vs subtracting the smoothed signal and retaining only the micro-scale variations). The result is shown on Fig. 6. At the micro-scale, as expected from Figs. 2-3, the two enzymes exhibit a quite similar behavior, involving variations of comparable magnitude (vertical standard deviations) directly related to the local gene expression strength (the colored dots follow the expression strength), suggesting that transcription is indeed the major driver of topoisomerase recruitment at that scale. In

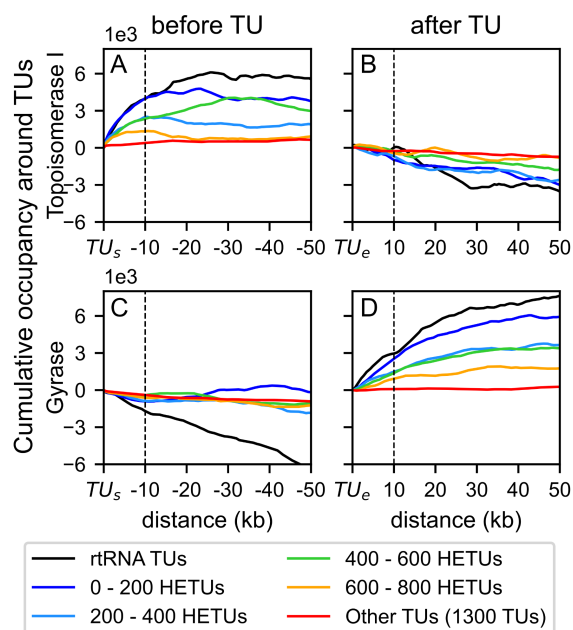


**Fig. 6:** Distributions of gyrase cleavage (A) and topoisomerase I binding (B) magnitudes across macro-scale regions of the chromosome (200 kb, corresponding to large-scale 3D structuring of the chromosome) versus micro-scale variations (10 kb, dominated by transcription) within these regions. The contributions were computed by signal filtering (see Materials and Methods) and binned at 50 kb. Colored lines indicate the average signal after (gyrase) or before (topoI) different classes of TUs (same colors as in Fig. 2).

contrast, the contribution of the macro-scale chromosome organization differs significantly. Gyrase binding is quite different among the different macro-regions of the chromosome (vertical standard deviation), contributing to almost as much variability as the micro-scale. In contrast, the recruitment of topoI is relatively uniform at the macro-scale, showing that micro-scale transcription alone, rather than the large-scale organization of the chromosome, determines the recruitment of that enzyme.

Since the micro-scale variations of topoisomerase binding seem dominated by transcription, to what distance does this influence transcription, to what distance does this influence extend? We addressed this question in Fig. 7, by computing the cumulative binding (area under the binding profile) between the TU start/end and a remote position up to a distance of 50 kb: a flat curve thus indicates a local binding intensity similar to the genomic average (and thus no influence of the TUs under consideration), whereas an increase (resp. decrease) indicates a stronger (resp. weaker) binding, possibly due to transcription.

As already observed, topoI binding increases with expression level in the first 10 kb before the TU, and the extension of the transcription-induced negative supercoils also tends to increase (Fig. 7A). For stable RNAs (black curves), the signal reaches 25-30 kb, consistent with the size identified around the ribosomal operons by psoralen binding<sup>8</sup>. After the TU (Fig. 7B), the decrease (weak topoI binding) mostly occurs in the first 10-20 kb, and increases with the expression level, although the slope remains slightly negative on a longer scale.



**Fig. 7:** Extension of the transcriptional-induced topological constraints. The top and bottom panels show cumulative signals (i.e., total binding from the TU) for TopoI binding and gyrase cleavage, respectively, in the 50 kb before the TU start (left panels) and after the TU end (right panels). TUs were classified based on their expression level (each expression category is represented by a different color, as in Fig. 2). The plotted signals correspond to the average area under the curve of the micro-scale signal (i.e., after subtracting the macro-scale contribution, especially for gyrase), from the TU start or end site to increasing distances (up to 50 kb).

For gyrase, the signal is flat before the TU start (Fig. 7C), except for stable RNAs (this observation may also reflect the small distance between them). After the TU (Fig. 7D), a strong enrichment in gyrase cleavage sites is observed, and increases with the expression level; but in contrast to topoI, it propagates way beyond the 10 kb region, from around 20 kb for moderately expressed TUs (orange) to around 40 kb for strongly expressed ones (dark blue) and even beyond for stable RNAs. Thus, the positive supercoils induced by transcription may extend as far as several dozen kilobases. We assume here that this extension reflects the influence of the central TU predominantly, rather than that of neighboring TUs, since the latter have similar levels regardless of the class of the former (Supplementary Fig. S5) and thus unlikely explain the ordered intensities of the binding profiles.

## Discussion

The interplay between gene expression and chromosome structure is a complex topic, which requires dedicated conceptual and technical tools. While classical regulatory networks are often analyzed in a qualitative way (does a TF bind at a

promoter?), a whole layer of transcriptional regulation can presumably be only understood through quantitative analysis (does it bind more or less?), considering spatial (1D or 3D) relations between genes, and combining different actors (architectural proteins, SC, expression, ...) acting together through statistical analyses. GRATIOSA aims to facilitate such analyses, by automating the formatting and import of various data into a uniform framework, and providing a range of standard tests and visualization tools. By definition, this form of regulation escapes standard models and thus requires some level of custom analysis, restricting the software to users with some proficiency in Python programming. But as these skills are becoming more widespread, the gain that GRATIOSA offers in programming time and reproducibility should help make these analyses accessible even to biologists that can only invest a part of their time in computational analyses. This is illustrated in the example of topoisomerase distribution, where the basic steps of the analysis are automated, and the programming effort can be dedicated to the design of specific statistical tests and graphs (see notebooks in Supplementary Files) to get insights into this question..

35 years after the proposal of the twin-domain model of transcription<sup>23</sup>, our understanding of the role, magnitude, and consequences of this phenomenon in bacterial regulation is still fragmentary. In particular, how many genes are subject to strong topological constraints *in vivo*, how these affect their expression in physiological conditions, and how far the expression of a given gene modulates that of its (many) neighbors, remain largely unknown. Here, we analyzed population-averaged, high-throughput data to provide quantitative insights into these questions. The reader should keep in mind that we have only access to statistical correlations, and all observed features may be subject to biases due to factors that we did not consider explicitly, e.g., differences in sequences in/around TUs of different expression strengths, presence of proteins, etc. This is especially true for operons of stable RNAs that are present in low numbers and are presumably regulated by many specific mechanisms). Any mechanistic inference from these analyses thus remains subject to future experimental tests.

The analyzed data consistently confirm the widespread relevance of the twin-domain model in bacterial genomes (Fig. 2), and not only in a small minority of highly expressed genes as already demonstrated before<sup>15</sup>. The phenomenon is detected specifically in the 40% highly expressed fraction of the genome, and exhibits a clear and increasing magnitude with the expression strength (Fig. 3). While the 60% remaining genes might also induce

local supercoils (especially if the sensitivity of the experimental method is a limiting factor in our analysis), this observation suggests that the low expression level of around half *E. coli* genes can be sustained without local recruitment of topoisomerases. The local orientation of neighboring genes strongly enhances the effect. Analyzing the specific effect of transcription (Fig. 4 C-F) suggests that topoI is not specifically required for an average expression level at tandem genes, but is strongly recruited between divergent promoters, as observed long ago in *S. enterica*<sup>46</sup>, in addition to its strong activity at strongly expressed genes. In contrast, gyrase is not specifically attracted by the expression of convergent genes (Fig. 4F), suggesting that, in these analyzed experimental conditions, accumulating positive supercoils do not usually prevent transcription elongation even in convergent regions. This observation matches a recent result<sup>47</sup> for a single gene, and with the observation that gyrase inhibition does not usually inhibit but rather activates convergent genes<sup>14</sup>. RNAP stalling because of positive supercoils may thus be limited to a few highly expressed genes *in vivo* (at least in the exponential phase where the baseline superhelical level is strongly negative). At the promoters of tandem and especially divergent genes, gyrase binding is reduced because of the negative supercoils present there.

What is the extension of twin-supercoiled domains, and the relative magnitude of transcription vs other mechanisms (especially the larger 3D organization of the chromosome) in shaping the local distribution of supercoils? This question is important, as theoretical models suggested that the propagation of supercoils might induce collective expression modes between neighboring genes in the densely packed bacterial genomes<sup>48</sup>. The spatial analysis (Figs. 5-6) show that topoI recruitment is strongly dominated by transcription with a characteristic length-scale of 10 kb; whereas gyrase contains two different contributions of comparable magnitude, one from transcription (at a comparable scale), and one an order of magnitude larger (100-200 kb), as identified in 3D conformational maps (“CIDs”), and presumably induced by a complex interplay of architectural proteins, transcription and replication-related signals<sup>45</sup>. Thus, the CID structuring of the chromosome presumably has a significant impact on gyrase binding, but not on topoI. This may be related to differences in sequence or structure between CIDs (since gyrase binds plectonemic DNA and recognizes specific binding sequence, whereas topoI binds denatured DNA), as all of these signals seem somewhat correlated with CID borders (Supplementary Fig. S6); but understanding this relation will require a more precise analysis, rather than the coarse

averaging at a fixed macro-scale of 200 kb as we have considered here for our analysis focusing on transcription.

At the micro-scale, the separation of the 40% strongest TUs into expression classes clearly suggests that transcription is the dominant factor in the recruitment of gyrase and topoI, following a strong linear (rather than 3D) ordering (Fig. 7). TopoI is strongly recruited along transcription units (especially near the TSS) and upstream of them, up to a distance of a few to 10-20 kb, increasing with the expression level. Gyrase binds downstream of TUs up to 30-40 kb, not only in highly expressed operons, but in a large fraction of the genome. These spatial scales are slightly larger, but of the same order of magnitude as those previously proposed for topological domains<sup>43</sup>. The precise values however remain subject to further investigation because of the intrinsic limitations of the employed statistical methods: (1) the separation into two scales and mechanisms is partly artificial, if they overlap for small CIDs and/or those shaped by transcription itself; (2) the employed expression classes may differ not only in the strength of the central TU (used in their definition), but in the properties of neighboring regions where we investigate topoisomerase binding. We checked that the expression levels of these neighboring TUs do not follow the same order as the central ones (Supplementary Fig. S6), but some correlation might remain and contribute to the observed extension of twin-supercoiled domains. The analyzed data are among the first to provide relatively direct evidence that in the genomic context, these extend at least to 10 kb in both directions, for a substantial fraction of the genome. Given the density of bacterial genomes, this length signifies that most gene share some local topological regulatory interaction with a dozen of neighbors, presumably resulting in a complex collective interplay previously hypothesized<sup>10,48</sup>, which may contribute to the observed correlation of gene expression up to 10 kb apart (versus 2-3 kb long TUs). Ultimately, this issue may only be solved by directly observing the topoisomerase binding distributions after inducing a specific gene in well-controlled conditions (rather than correlations); but we then still expect the expression of neighboring genes to change after the induction, and to contribute to the resulting topological landscape.

## Code availability

GRATIOSA is an open-access Python package published to PyPI. The package is also accessible on GitHub with the scripts required to reproduce our analyses:

<https://github.com/sammeyer2017/GRATIOSA>.

Documentation is available at: <https://gratiosa.rtfid.io/>.

## References

- Dame, R. T. & Tark-Dame, M. Bacterial chromatin: converging views at different scales. *Curr. Opin. Cell Biol.* **40**, 60–65 (2016).
- Browning, D. F., Grainger, D. C. & Busby, S. J. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr. Opin. Microbiol.* **13**, 773–780 (2010).
- Travers, A. & Muskhelishvili, G. Chromosomal Organization and Regulation of Genetic Function in Escherichia coli Integrates the DNA Analog and Digital Information. *EcoSal Plus* **9**, (2020).
- Marr, C., Geertz, M., Hütt, M.-T. & Muskhelishvili, G. Dissecting the logical types of network control in gene expression profiles. *BMC Syst. Biol.* **2**, 18 (2008).
- Bintu, L. *et al.* Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* **15**, 116–124 (2005).
- 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.* **47**, 1373–1388 (2019).
- Guo, M. S., Kawamura, R., Littlehale, M. L., Marko, J. F. & Laub, M. T. High-resolution, genome-wide mapping of positive supercoiling in chromosomes. *eLife* **10**, e67236 (2021).
- Visser, B. J. *et al.* Psoralen mapping reveals a bacterial genome supercoiling landscape dominated by transcription. *Nucleic Acids Res.* **50**, 4436–4449 (2022).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Junier, I., Unal, E. B., Yus, E., Lloréns-Rico, V. & Serrano, L. Insights into the Mechanisms of Basal Coordination of Transcription Using a Genome-Reduced Bacterium. *Cell Syst.* **2**, 391–401 (2016).
- Bourgey, M. *et al.* GenPipes: an open-source framework for distributed and scalable genomic analyses. *GigaScience* **8**, giz037 (2019).
- Singh, U., Li, J., Seetharam, A. & Wurtele, E. S. pyrpipe: a Python package for RNA-Seq workflows. *NAR Genomics Bioinforma.* **3**, lqab049 (2021).
- Rioualen, C., Charbonnier-Khamvongsa, L., Collado-Vides, J. & van Helden, J. Integrating Bacterial ChIP-seq and RNA-seq Data With SnakeChunks. *Curr. Protoc. Bioinforma.* (2019).
- El Houdaigui, B. *et al.* Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Res.* **47**, 5648–5657 (2019).
- Sutormin, D. *et al.* Interaction Between Transcribing RNA Polymerase and Topoisomerase I Prevents R-loop Formation in E. coli. 2021.10.26.465782 Preprint at <https://doi.org/10.1101/2021.10.26.465782> (2021).
- Fu, Z., Guo, M. S., Zhou, W. & Xiao, J. Differential Roles of positive and negative supercoiling in compacting the E. coli genome. 2022.12.30.522362 Preprint at <https://doi.org/10.1101/2022.12.30.522362> (2022).
- Martis B, S., Forquet, R., Reverchon, S., Nasser, W. & Meyer, S. DNA Supercoiling: an Ancestral Regulator of Gene Expression in Pathogenic Bacteria? *Comput. Struct. Biotechnol. J.* **17**, 1047–1055 (2019).
- Dillon, S. C. & Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8**, 185–195 (2010).
- Travers, A. & Muskhelishvili, G. DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nat. Rev. Microbiol.* **3**, 157–169 (2005).
- Lim, H. M., Lewis, D. E. A., Lee, H. J., Liu, M. & Adhya, S. Effect of varying the supercoiling of DNA on transcription and its regulation. *Biochemistry* **42**, 10718–10725 (2003).
- Maxwell, A. & Gellert, M. Mechanistic aspects of DNA topoisomerases. *Adv. Protein Chem.* **38**, 69–107 (1986).
- Champoux, J. J. DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* **70**, 369–413 (2001).
- Liu, L. F. & Wang, J. C. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 7024–7027 (1987).
- Dorman, C. J. DNA supercoiling and transcription in bacteria: a two-way street. *BMC Mol. Cell Biol.* **20**, (2019).
- Ahmed, W. *et al.* Transcription facilitated genome-wide recruitment of topoisomerase I and DNA gyrase. *PLoS Genet.* **13**, e1006754 (2017).
- Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
- Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

28. team, T. pandas development. pandas-dev/pandas: Pandas. (2023) doi:10.5281/zenodo.8239932.
29. Pysam.pysam-developers. <https://github.com/pysam-developers/pysam> (2023)
30. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
31. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. in 92–96 (2010). doi:10.25080/Majora-92bf1922-011.
32. Download ontology. *Gene Ontology Resource* <http://geneontology.org/docs/download-ontology/>.
33. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinforma.* **47**, 11.12.1-34 (2014).
34. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
35. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **31**, 166–169 (2015).
36. Wettenhall, J. M. & Smyth, G. K. limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinforma. Oxf. Engl.* **20**, 3705–3706 (2004).
37. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-165 (2016).
38. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
39. Matthey-Doret, C. *et al.* Computer vision for pattern detection in chromosome contact maps. *Nat. Commun.* **11**, 5795 (2020).
40. Ferrándiz, M.-J., Hernández, P. & Campa, A. G. de la. Genome-wide proximity between RNA polymerase and DNA topoisomerase I supports transcription in *Streptococcus pneumoniae*. *PLOS Genet.* **17**, e1009542 (2021).
41. Cheng, B., Zhu, C.-X., Ji, C., Ahumada, A. & Tse-Dinh, Y.-C. Direct interaction between *Escherichia coli* RNA polymerase and the zinc ribbon domains of DNA topoisomerase I. *J. Biol. Chem.* **278**, 30705–30710 (2003).
42. Pineau, M. *et al.* What is a supercoiling-sensitive gene? Insights from topoisomerase I inhibition in the Gram-negative bacterium *Dickeya dadantii*. *Nucleic Acids Res.* **50**, 9149–9161 (2022).
43. Postow, L., Hardy, C. D., Arsuaga, J. & Cozzarelli, N. R. Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev.* **18**, 1766–1779 (2004).
44. Liou, V. S. *et al.* Multiscale Structuring of the *E. coli* Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell* **172**, 771-783.e18 (2018).
45. Dame, R. T., Rashid, F.-Z. M. & Grainger, D. C. Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat. Rev. Genet.* **21**, 227–242 (2020).
46. Chen, D., Bowater, R., Dorman, C. J. & Lilley, D. M. Activity of a plasmid-borne leu-500 promoter depends on the transcription and translation of an adjacent gene. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 8784–8788 (1992).
47. Boulas, I. *et al.* Assessing in vivo the impact of gene context on transcription through DNA supercoiling. *Nucleic Acids Res.* gkad688 (2023) doi:10.1093/nar/gkad688.
48. Grohens, T., Meyer, S. & Beslon, G. Emergence of Supercoiling-Mediated Regulatory Networks through Bacterial Chromosome Organization. 2022.09.23.509185 Preprint at <https://doi.org/10.1101/2022.09.23.509185> (2023).



## Supplementary Information

# Integrated computational analysis of bacterial transcriptome and chromosome structural data with GRATIOSA

Maïwenn Pineau, Raphaël Forquet, Sylvie Reverchon, William Nasser,  
Florence Hommais and Sam Meyer\*

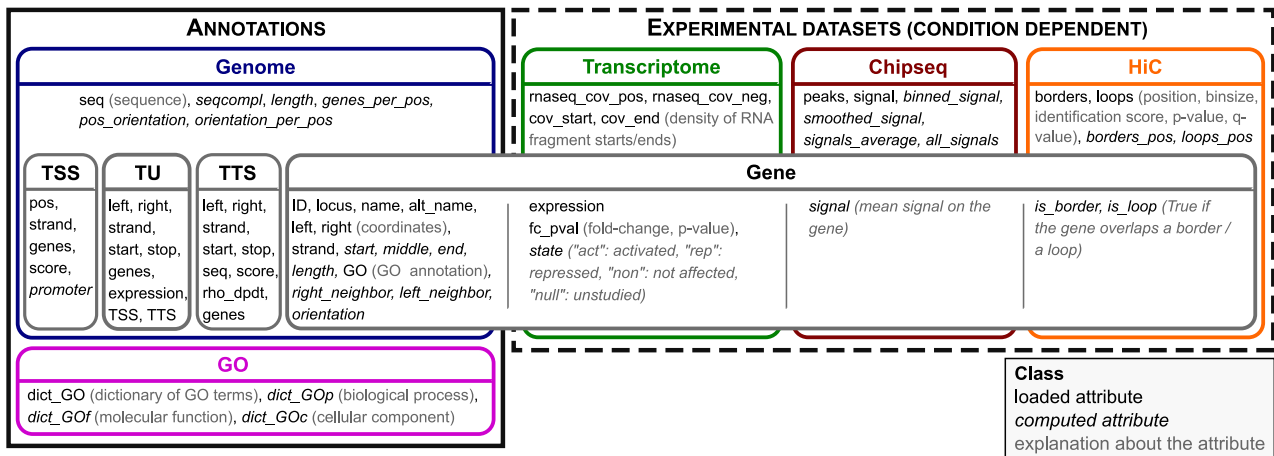
Université de Lyon, INSA Lyon, Université Claude Bernard Lyon 1, CNRS UMR5240, Laboratoire de Microbiologie, Adaptation et  
Pathogénie, 69621 Villeurbanne, France

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

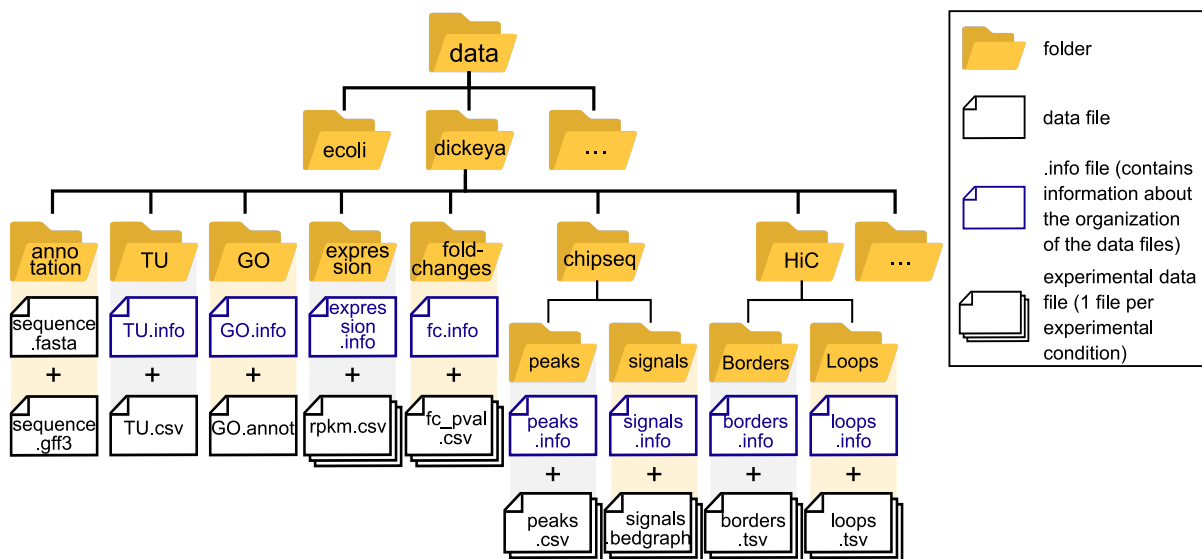
### **This PDF file includes:**

- Supplementary Figures S1 to S6
- Supplementary Table S1

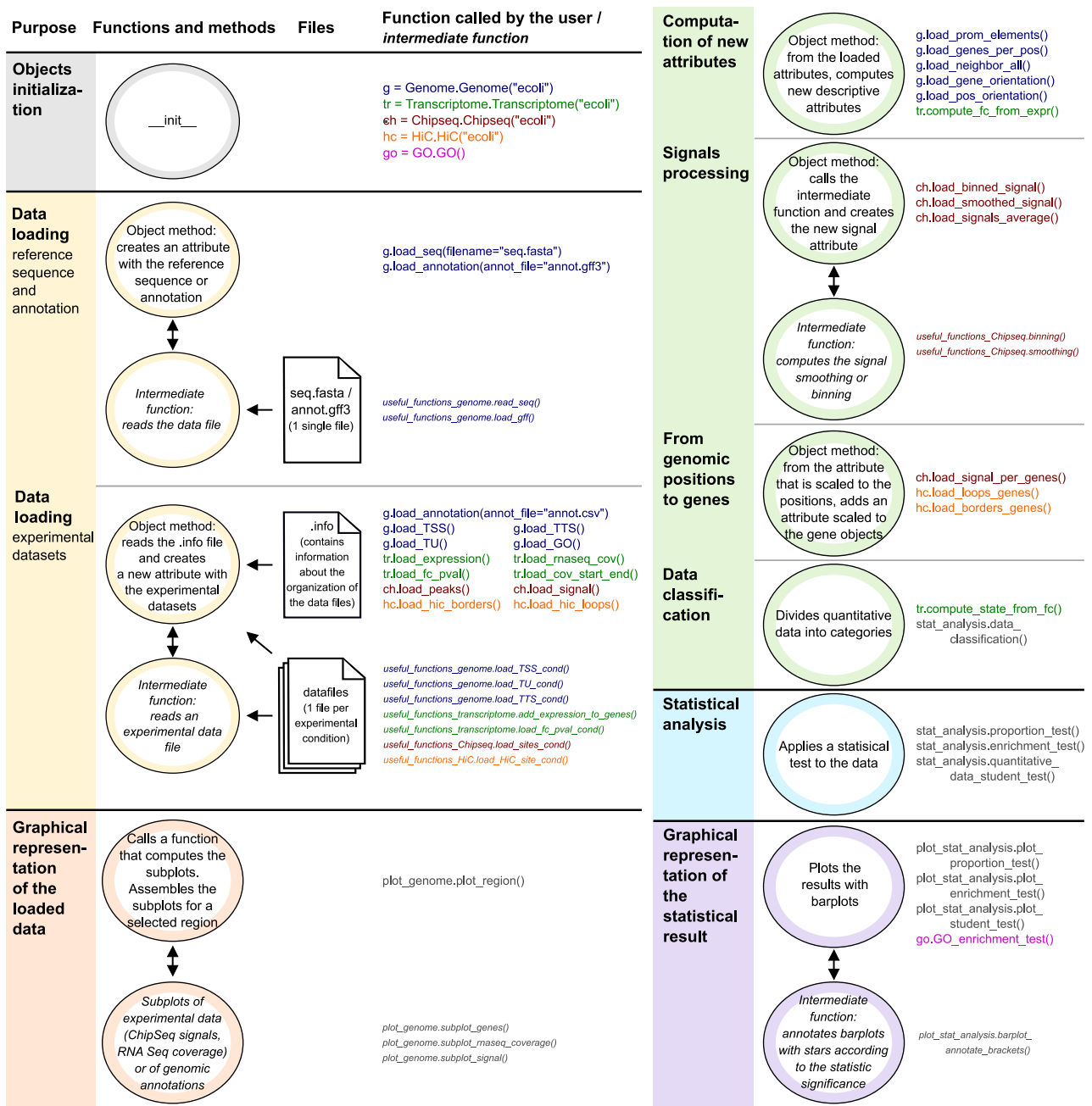
Supplementary figures



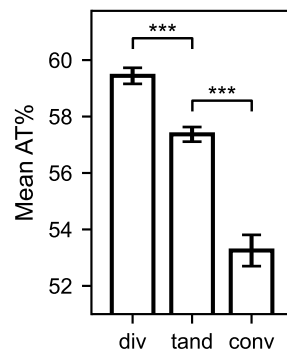
**Supplementary Figure S1:** List of attributes that can be loaded or computed for all classes. The Gene object is an elementary object containing the data of a single gene. Loading the data of all genes is achieved using the Genome objects (for annotations) or either Transcriptome, ChIP-Seq, or Hi-C (for experimental datasets). The same applies to TSS, TU, and TTS objects, whose attributes can be loaded and computed through the Genome class.



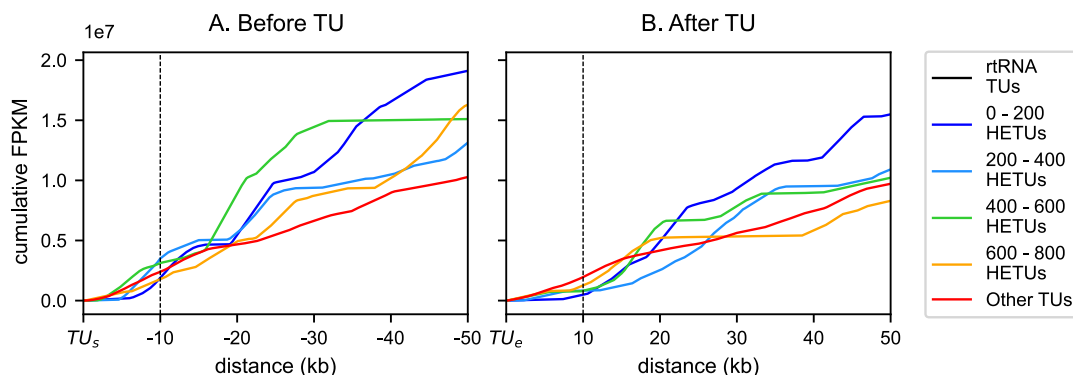
**Supplementary Figure S2:** Structure of the database, containing one directory per organism and one sub-directory for each data type. Each type of experimental data is associated with an info file which allows for the automatic import of a selection of data, of diverse formats.



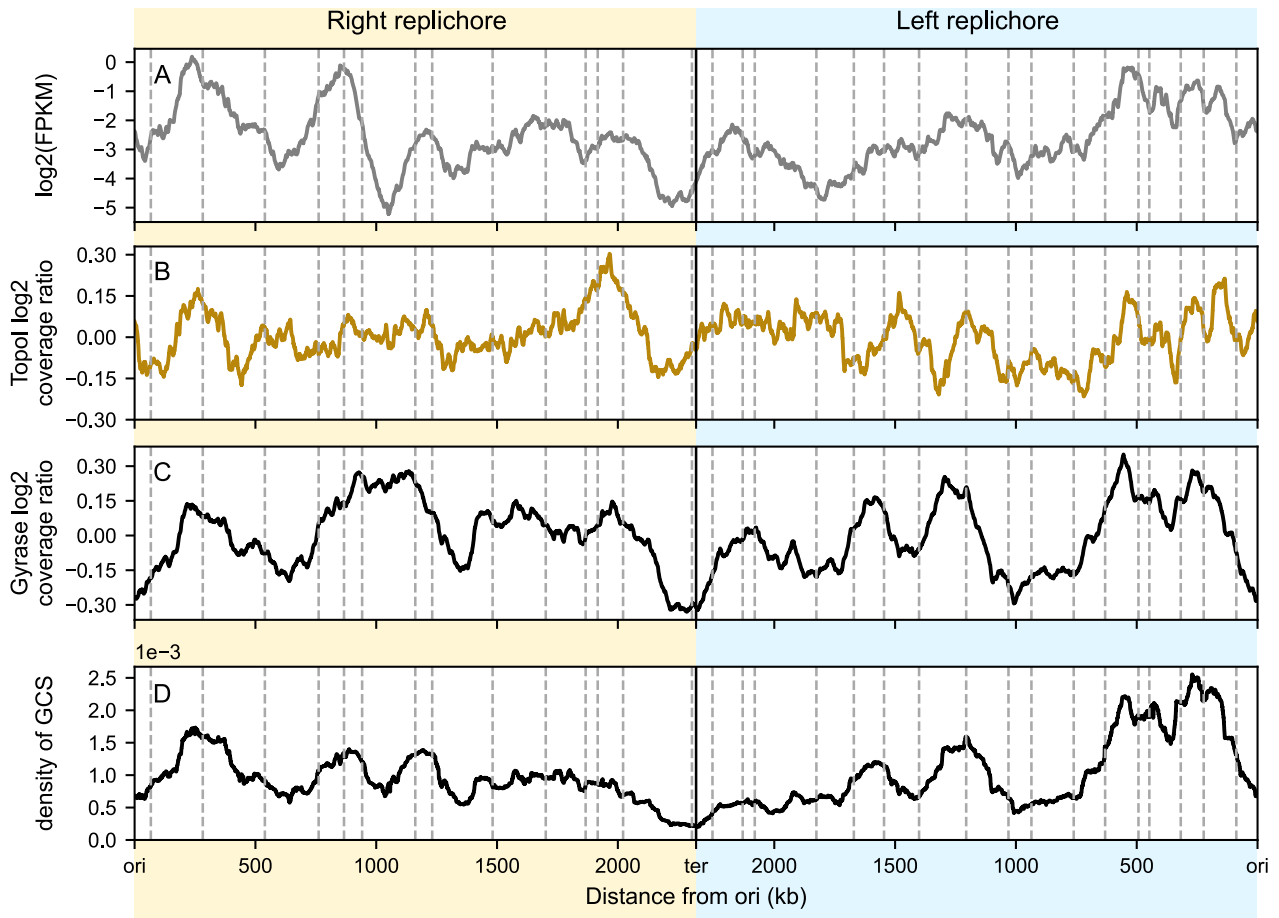
**Supplementary Figure S3:** Standard pipeline for an analysis using the GRATIOSA package. The pipeline consists of four steps (with colors matching those used in Fig 1): objects initialization (gray), data loading as object attributes (yellow), computation of new attributes (green), and statistical analysis (blue). It is also possible to visualize the loaded data (orange) and the statistical results (purple). These steps are performed using the methods associated with each object (Genome, Transcriptome, ChIP-Seq...). Each function call is shown in the color of the associated class in Fig S2 (all functions are listed). Function calls written in italics are intermediate functions called within the package, and, in usual cases, not by the user.



**Supplementary Figure S4:** Average AT content of the intergenic regions by orientation. Error bars are 95% confidence intervals



**Supplementary Figure S5:** Cumulative FPKM (expression level) of the neighboring TUs in the 50 kb before the TU start (**A**) and after the TU end (**B**): the expression of the neighbors does not increase with that of the central TU. TUs were classified based on their expression level, excluding TUs containing stable RNAs (shown in black). The top 200 expressed genes are referred to as 0-200 HETUs, the next 200 most highly expressed genes as 200-400 HETUs, and so on. A FPKM value was associated to each position based on the FPKM values of annotated TUs. If a given position corresponds to multiple TUs, the average FPKM of those TUs is used. The cumulative FPKM represents the sum of FPKM values for each position, from the TU start or TU end to increasing distances (up to 50 kb).



**Supplementary Figure S6:** Distribution of expression (A), topoisomerase I binding (B), gyrase binding (C), and gyrase cleavage sites (D) signals along the chromosome. The expression signal corresponds to the  $\log_2(\text{FPKM})$ . Topoisomerase I and gyrase signals are the  $\log_2$  ratios between coverage after immunoprecipitation and control coverage. The density of gyrase cleavage sites (GCS) identified by Sutormin et al. [1] is very similar to the gyrase coverage profile. The yellow and blue backgrounds correspond, respectively, to the right and left replichores. Gray dashed lines indicate the approximate positions of the CIDs' borders, as identified by Liroy et al. in the MG1655 strain [2]. The chromosomal inversion occurring between positions 3425 and 4215 kb in the W3110 strain has been accounted for in this figure.

## Supplementary table

**Supplementary Table S1:** Summary table of the statistical tests performed by the package. This table presents the expected inputs, employed modules, and outputs.

Statistical test	Inputs	Dependencies	Outputs
<b>Enrichment test</b> test of the over-representation of a feature in a category	<ul style="list-style-type: none"> <li>classification of all items into categories</li> <li>dictionary containing the feature associated with each item</li> <li>choice of a targeted feature</li> </ul>	<ul style="list-style-type: none"> <li><code>scipy.stats.hypergeom.interval</code></li> <li><code>scipy.stats.hypergeom.sf</code></li> <li><code>statsmodels.stats.multitest.fdr correction</code></li> </ul>	For each category: <ul style="list-style-type: none"> <li>Number of targeted items (i.e., items associated with the chosen feature) that belong to the category</li> <li>Expected number of targeted items that should belong to the category according to the category size and the total number of elements having that feature</li> <li>Proportion of items in the category among all the items associated with the chosen feature.</li> <li>95% confidence interval around the proportion</li> <li>p-value and adjusted p-value of the enrichment test</li> </ul>
<b>Proportion test</b> test of the difference in the proportion of elements associated with a feature between categories	<ul style="list-style-type: none"> <li>classification of all items into categories</li> <li>dictionary containing the feature associated with each item</li> <li>choice of a targeted feature</li> </ul>	<ul style="list-style-type: none"> <li><code>proportion.proportion_confint</code></li> <li><code>statsmodels.stats.proportion.proportions_ztest</code></li> </ul>	For each category: <ul style="list-style-type: none"> <li>Proportion of the category items that have the targeted feature</li> <li>95% confidence interval around the proportion</li> <li>P-value of the proportion test (one- or two-tailed test depending on the user's choice)</li> </ul>
<b>Student test</b> test of the difference between the means of different categories	<ul style="list-style-type: none"> <li>list of values classified into categories</li> </ul>	<ul style="list-style-type: none"> <li><code>scipy.stats.norm.interval</code></li> <li><code>scipy.stats.ttest_ind</code></li> </ul>	For each category: <ul style="list-style-type: none"> <li>Mean of the values</li> <li>95% confidence interval around the mean</li> <li>P-value of the student test (one- or two-tailed test depending on the user's choice)</li> </ul>

**References**

- [1] Sutormin, D., Rubanova, N., Logacheva, M., Ghilarov, D., and Severinov, K. Single-nucleotide-resolution mapping of DNA gyrase cleavage sites across the Escherichia coli genome. **47**(3), 1373–1388.
- [2] Lioy, V. S., Cournac, A., Marbouty, M., Duigou, S., Mozziconacci, J., Espéli, O., Boccard, F., and Koszul, R. Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins. **172**(4), 771–783.e18 Publisher: Elsevier.

## 3.2 La mise au point d'une technique de cartographie du surenroulement local

Les résultats de l'analyse des ChIP-Seq des topoisomérases apportent des informations intéressantes sur l'organisation du chromosome de *E. coli* et montrent que l'étude du SC à l'échelle locale peut permettre de caractériser l'organisation du chromosome bactérien. Afin d'étudier l'organisation de son chromosome et dans le but d'approfondir les analyses de la régulation de l'expression des gènes selon le niveau de SC local, nous avons souhaité réaliser une cartographie du SC de *D. dadantii* à l'échelle locale.

A partir d'une méthodologie testée sur les eucaryotes, l'équipe de Monika Glinkowska (Département de génétique moléculaire bactérienne, Université de Gdansk, Pologne) développe une méthode permettant de caractériser la distribution spatiale du SC le long d'un chromosome bactérien (Fig. 3.3). Cette technique repose sur la sensibilité de l'intercalation du tri-méthyl-psoralen (TMP) au niveau local de SC négatif : le TMP s'intercale davantage dans les régions où le niveau de SC négatif est élevé (Chapitre 1, section 1.5.2). Lorsqu'un double brin d'ADN avec du TMP intercalé est soumis aux UV, cette intercalation prend la forme d'un pontage, c'est à dire que les deux brins sont liés de manière covalente à la molécule de TMP [203, 204, 205].

### 3.2.1 Présentation du protocole de cartographie

Je suis allée travailler avec Monika Glinkowska pendant le mois d'octobre 2021 afin de prendre connaissance de cette technique, pour pouvoir à terme l'appliquer au génome de *D. dadantii*. J'y ai découvert et adapté le protocole suivant (Fig. 3.3).

#### — Culture des bactéries et traitement avec le TMP

Les bactéries sont mises en culture dans des erlenmeyers dans des bains chauffants liquides (37°C pour *E. coli*, 30°C pour *D. dadantii*), sous agitation. Lorsque la DO souhaitée est atteinte, du TMP est ajouté dans la culture. Le TMP est sensible à la lumière, par conséquent il est essentiel que les cultures soient dans l'obscurité, par exemple en entourant les erlenmeyers de papier aluminium.

#### — Pontage de l'ADN avec le TMP, sous UV

Après 10 minutes d'incubation du TMP dans la culture, le contenu des erlenmeyers est versé dans des boîtes de Pétri, en mettant 12 mL de culture par boîte. Les boîtes sont ensuite soumises aux UV pendant 4 min à 360 nm et 0,120 J.cm<sup>-2</sup> afin de provoquer la formation de pontages entre les brins d'ADN via le TMP.

#### — Extraction de l'ADN

Directement après la fin du traitement UV, les cultures sont versées dans des falcons



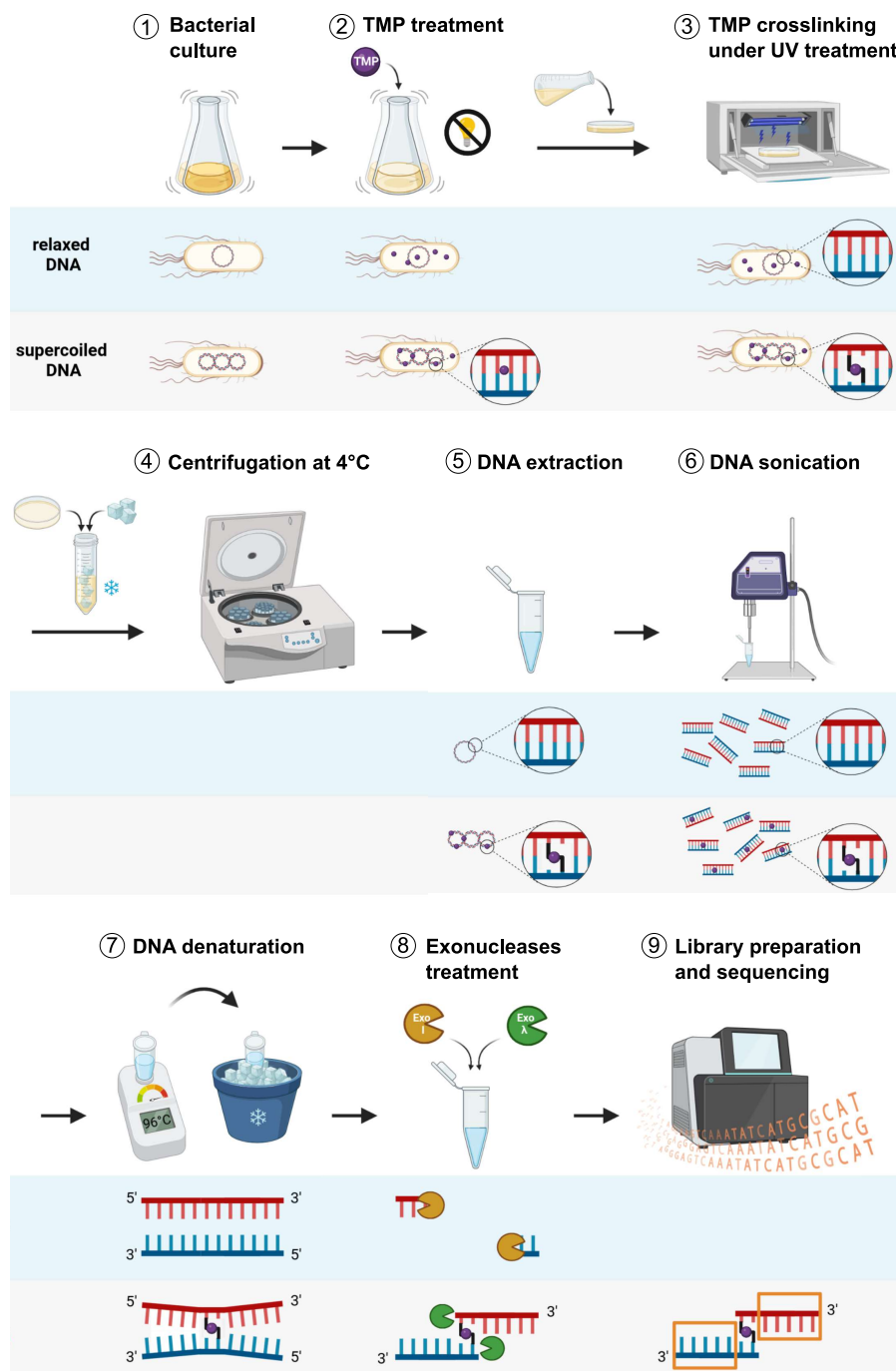


FIGURE 3.3 – Technique expérimentale de cartographie du niveau local du SC et illustration de l'effet sur un ADN relâché (fond bleu) et sur un ADN surenroulé (fond gris). (1) Culture bactérienne jusqu'à la phase de croissance désirée et choc (facultatif). (2) Ajout du TMP dans l'obscurité. Intercalation du TMP dans les zones surenroulées négativement. (3) Traitement des cultures aux UV, pontage du TMP intercalé. (4) Centrifugation des cultures à froid. (5) Extraction de l'ADN bactérien en utilisant du phénol-chloroforme. Seul l'ADN surenroulé contient un grand nombre de pontages. (6) Fragmentation de l'ADN par sonication. (7) Dénaturation de l'ADN non-crosslinké par choc de température. (8) Traitement aux exonucleases : clivage des simples brins d'ADN par l'exonucléase I (3' vers 5') et digestion de l'ADN du 5' vers 3' par l'exonucléase λ jusqu'au site de pontage. (6) Séquençage des fragments non digérés. Schéma créé avec bioRender.

plongés dans de la glace, afin d'interrompre immédiatement toute réaction. Les cultures sont ensuite centrifugées à 4°C dans le but d'obtenir un culot bactérien. L'ADN est extrait des culots bactériens en utilisant du phénol-chloroforme [257].

— **Fragmentation de l'ADN**

L'ADN est ensuite fragmenté par un sonicateur afin d'obtenir des fragments d'environ 100 à 300 pb.

— **Dénaturation des fragments d'ADN et dégradation des simples brins**

Afin de séparer les paires de bases non pontées, l'ADN est dénaturé par traitement thermique. Les fragments ne contenant pas de pontage sont sous forme simple brin et peuvent être complètement dégradés par les exonucléases I. Les fragments contenant un pontage sont liés par ce pontage mais le reste du fragment est dissocié ce qui permet aux extrémités 5' d'être dégradées par des exonucléases  $\lambda$  jusqu'au site de pontage.

— **Séquençage des fragments d'ADN**

Les extrémités non dégradées, c'est à dire les queues 3' des fragments pontés, peuvent ensuite être séquencées afin de localiser les sites de pontages.

### 3.2.2 Calibration du protocole de cartographie chez *D. dadantii*

Pour appliquer ce protocole chez *D. dadantii*, il était nécessaire de réaliser des calibrations. La principale calibration sur laquelle j'ai travaillé est la concentration en TMP.

#### Méthodologie de calibration de la concentration en TMP

Cette concentration est un paramètre clé puisque l'utilisation d'une trop grande concentration de TMP aboutira à un pontage non spécifique du SC tandis qu'une trop faible concentration de TMP ne permettra d'identifier qu'une faible proportion des régions surenroulées négativement.

L'étude de l'ADN génomique complet étant laborieuse, j'ai effectué cette calibration à l'aide de plasmides en considérant que leur niveau de SC reflète celui du chromosome. Pour obtenir une concentration en TMP convenable pour l'étude du SC du chromosome, l'objectif était de quantifier le TMP nécessaire à l'obtention de pontages dans la moitié des plasmides. Pour ce faire, j'ai électroporé le plasmide pUC18 dans *D. dadantii*. Les plasmides étant peu stables chez *D. dadantii*, il était nécessaire de réaliser cette électroporation avant chaque culture. J'ai ensuite suivi les étapes 1 à 4 du protocole présenté en figure 3.3, en traitant les cultures avec différentes concentrations de TMP (0,5 à 4  $\mu$ g de TMP par mL de culture). Pour observer les pontages des plasmides, il était ensuite nécessaire de les extraire. Pour évaluer le taux de plasmides pontés, les plasmides extraits ont été digérés par EcoRI afin de les linéariser puis je les ai soumis à un choc thermique afin de les dénaturer.

Suite à ce traitement, les plasmides non pontés sont sous forme de simple brin tandis que les plasmides pontés, grâce à leur liaison créée via le TMP, parviennent à s'hybrider de nouveau et forment un double brin linéaire. Ces deux formes distinctes peuvent être différenciées par migration sur gel d'agarose, où l'ADN simple brin, de plus faible poids moléculaire, migre plus rapidement donc plus loin que l'ADN double brin.

### Mise au point d'une méthodologie d'extraction de plasmide de *D. dadantii*

Pour extraire les plasmides de *D. dadantii*, j'ai comparé les kits Plasmide Mini de A&A Biotechnology, Miniprep GenElute Plasmid de Sigma-Aldrich et QIAprep Miniprep de Qiagen. Pour chaque extraction, j'ai utilisé 24 mL de culture de *D. dadantii* en phase exponentielle. D'après les indications du fournisseur, le protocole d'extraction de plasmides à partir d'un culot bactérien avec le kit QIAprep Miniprep de Qiagen repose sur les étapes suivantes.

#### — Préparation de la solution bactérienne

Les cultures bactériennes sont centrifugées puis le culot bactérien est resuspendu dans un tampon contenant de la RNaseA. La RNaseA permet l'élimination des contaminants ARN.

#### — Lyse des bactéries

Les bactéries sont lysées dans du NaOH et du SDS. Le SDS dénature la membrane cellulaire, ce qui conduit à la libération du contenu cellulaire. Les conditions alcalines dénaturent l'ADN et les protéines. La durée du traitement avec ce tampon est donc critique afin de ne pas dénaturer irréversiblement le plasmide. L'arrêt de ce traitement, la neutralisation du lysat, est obtenu avec un tampon riche en sels.

#### — Isolement des plasmides dans un tampon riche en sels

La haute concentration en sels provoque la précipitation des protéines dénaturées, de l'ADN chromosomique, des débris cellulaire et du SDS tandis que les plasmides restent en solution. Une centrifugation permet de séparer ce précipitât de la solution contenant les plasmides.

#### — Purification des plasmides

La solution est ensuite lavée sur une membrane de silice qui absorbe sélectivement l'ADN plasmidique en solution dans un tampon à haute concentration de sels. Les tampons utilisés pour le lavage permettent d'entraîner les ARN, protéines cellulaires, nucléases et autres métabolites à travers la membrane, sans être retenus. Une fois le lavage effectué et la membrane séchée, un tampon d'élution faible en ions permet d'éluer les plasmides retenus dans la membrane.

J'ai appliqué ce protocole ainsi que celui du kit A&A Biotechnology afin d'extraire des plasmides pUC19 de *D. dadantii* (Fig. 3.4).

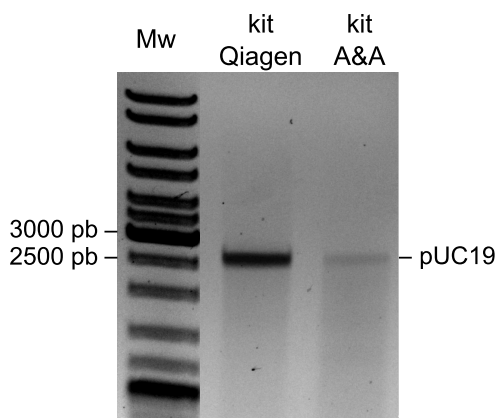


FIGURE 3.4 – Électrophorèse sur gel d'agarose 1% de plasmides pUC19 extraits de *D. dadantii* avec les kits QIAprep Miniprep de Qiagen et Plasmide Mini de A&A Biotechnology. Le volume déposé dans chaque puit étant identique, ce gel montre la supériorité du rendement obtenu avec le kit Qiagen en comparaison à celui obtenu avec le kit A&A Biotechnology.

Le rendement obtenu avec le kit Qiagen étant supérieur à celui obtenu avec le kit A&A Biotechnology, j'ai ensuite souhaité l'optimiser, tout en testant un troisième kit : le kit Miniprep GenElute Plasmid de Sigma-Aldrich. J'ai notamment testé en doublant les volumes de tampons, en ajoutant 4 minutes d'incubation après l'ajout du tampon de lyse et en ajoutant de l'EDTA ou de la RNaseA à la fin de l'extraction (Tableau 3.1).

TABLEAU 3.1 – Concentrations en plasmides pUC19 extraits de *D. dadantii* selon le kit et le protocole utilisés.

Kit utilisé	Volume tampons	Temps incubation	Traitement ajouté	Concentration plasmide (ng/μL)
Qiagen	1x	10-20 s		215
Qiagen	2x	10-20 s		56
Qiagen	1x	4 min		97
Qiagen	2x	4 min		72
Qiagen	1x	10-20 s	EDTA	119
Qiagen	1x	10-20 s	RNaseA	67
Sigma	1x	10-20 s		74
Sigma	2x	10-20 s		28
Sigma	1x	4 min		46
Sigma	2x	4 min		28
Sigma	1x	10-20 s	EDTA	60
Sigma	1x	10-20 s	RNaseA	60

Les concentrations obtenues (Tableau 3.1) indiquent que le kit Qiagen permet d'obtenir les meilleurs rendements, en particulier en respectant les volumes de tampons indiqués dans le protocole et en limitant le temps d'incubation avec le tampon de lyse.

Une fois cette extraction réalisée, l'étape suivante du protocole de calibration, c'est à dire la digestion des plasmides par EcoRI, peut être réalisée. En réalisant cette digestion, j'ai découvert que mes plasmides se dégradait fortement. Cette dégradation étant probablement due à la présence de nucléases, j'ai dû réaliser de nouveaux tests sur le protocole d'extraction des plasmides afin d'éliminer ces nucléases. Pour cela, j'ai testé 3 alternatives. La première est l'ajout d'un traitement avec du mélange chloroforme-alcool phénolique-alcool isoamylique à la suite de l'élution. La seconde est l'ajout d'une étape de nettoyage avec le tampon fourni (PB) à la suite du premier nettoyage recommandé dans le protocole. La troisième est l'ajout de ce même nettoyage en utilisant une nouvelle colonne de purification, de préférence juste avant la réalisation de la digestion. Cette dernière méthode a permis d'optimiser l'élimination des nucléases et ainsi la pureté de l'extraction (Fig. 3.5).

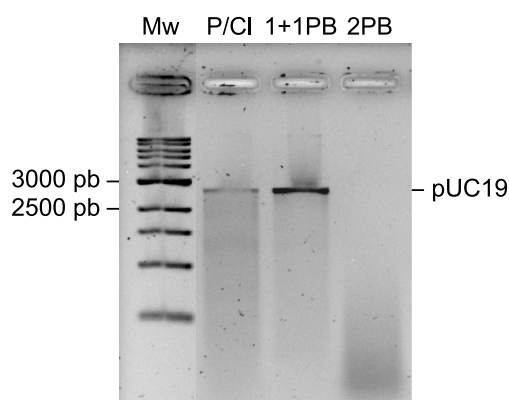


FIGURE 3.5 – Électrophorèse sur gel d'agarose 1% de plasmides pUC19 extraits de *D. dadantii* avec le kit QIAprep Miniprep de Qiagen puis digérés par EcoRI selon le protocole de lavage effectué à la fin de l'extraction. P/Cl indique les plasmides purifiés au phénol-chloroforme. 1+1PB indique les plasmides lavés une seconde fois avec le tampon de lavage PB sur une nouvelle colonne, juste avant la digestion. 2PB indique les plasmides lavés deux fois à la suite avec le tampon PB, sans changer de colonne. Le volume déposé dans chaque puit étant identique, ce gel montre que le traitement avec deux colonnes et deux lavages (1+1PB) permet d'améliorer la qualité de l'extraction.

En résumé, afin d'obtenir des concentrations et des puretés suffisantes lors des extractions d'un plasmide à haut nombre de copies chez *D. dadantii* en phase exponentielle, 24 mL de culture doivent être centrifugés puis le kit QIAprep Miniprep de Qiagen peut être utilisé. Afin de minimiser les dégradations de l'ADN, il semble souhaitable de réaliser les modifications suivantes par rapport au protocole du fournisseur : (1) effectuer les centrifugations à 4°C, (2) limiter le nombre d'extractions simultanées à deux extractions afin de diminuer les temps d'incubation et (3) réitérer les étapes de lavage, en utilisant une nouvelle colonne de purification, juste avant l'utilisation du plasmide extrait.

### Résultat de la calibration de la quantité en TMP

Une fois le protocole d'extraction des plasmides de *D. dadantii* mis au point, j'ai réalisé la calibration de la concentration en TMP permettant l'étude du SC local en suivant la méthodologie présentée précédemment. J'ai donc traité des cultures bactériennes avec différentes quantités de TMP puis j'ai extrait, linéarisé par digestion par EcoRI et enfin dénaturé les plasmides. Suite à cette dénaturation, seuls les plasmides pontés par le TMP sont sous forme double brin tandis que les autres plasmides sont sous forme simple brin. Comme on peut le voir sur la Figure 3.6, jusqu'à 4 µg de TMP par mL de culture, plus la concentration en TMP utilisée pour le traitement est élevée, plus la proportion de plasmides sous forme double brin est élevée. Au delà de 4 µg/mL, la proportion de plasmides sous forme double brin ne semble plus augmenter, indiquant une possible saturation en TMP. Il est essentiel de se placer en dessous de ce seuil de saturation afin d'obtenir une gamme d'intercalation selon le niveau de SC local le long du chromosome. Une concentration de 1.5 ou 2 µg/mL semble permettre cette étude.

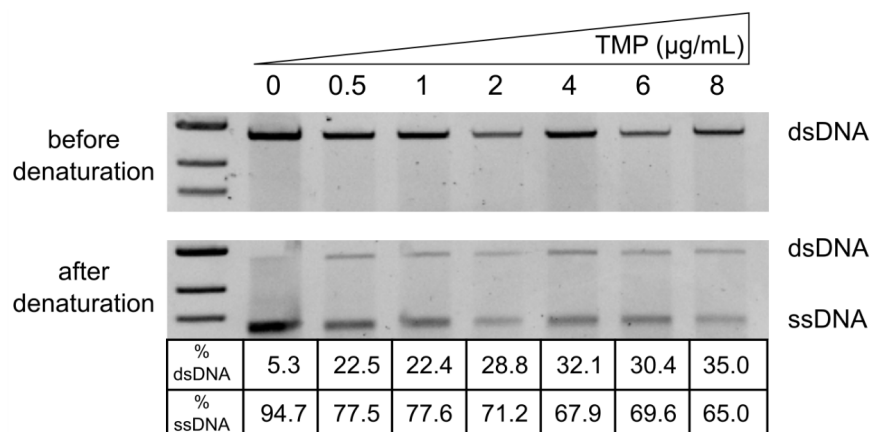


FIGURE 3.6 – Électrophorèse sur gel d'agarose 1% de plasmides pUC19 extraits de *D. dadantii* après le traitement des cultures avec différentes concentrations de TMP. Les plasmides extraits ont été digérés puis déposés sur gel avant (gel supérieur) et après (gel inférieur) leur dénaturation par choc thermique. Les plasmides sous forme simple brin (ssDNA) migrent plus rapidement donc plus loin que les plasmides sous forme double brin (dsDNA). Le tableau indique la proportion de plasmides sous chaque forme après dénaturation.

### 3.2.3 Comparaison des protocoles de cartographie

En mars 2022, Visser *et al.* ont publié la cartographie du SC local de *E. coli* obtenue grâce à une méthode reposant également sur l'intercalation du psoralen dans les régions d'ADN surenroulées négativement. Leur protocole, bien que reposant sur le même intercalant que la méthode de M. Glinkowska, diffère dans un grand nombre d'étapes puisque les sites pontés par cet intercalant sont localisés à l'aide du marquage à la biotine du psoralen (Fig. 3.7), tandis que, dans le protocole de M. Glinkowska, ces sites sont identifiés grâce à l'étape de digestion par les exonucléases.

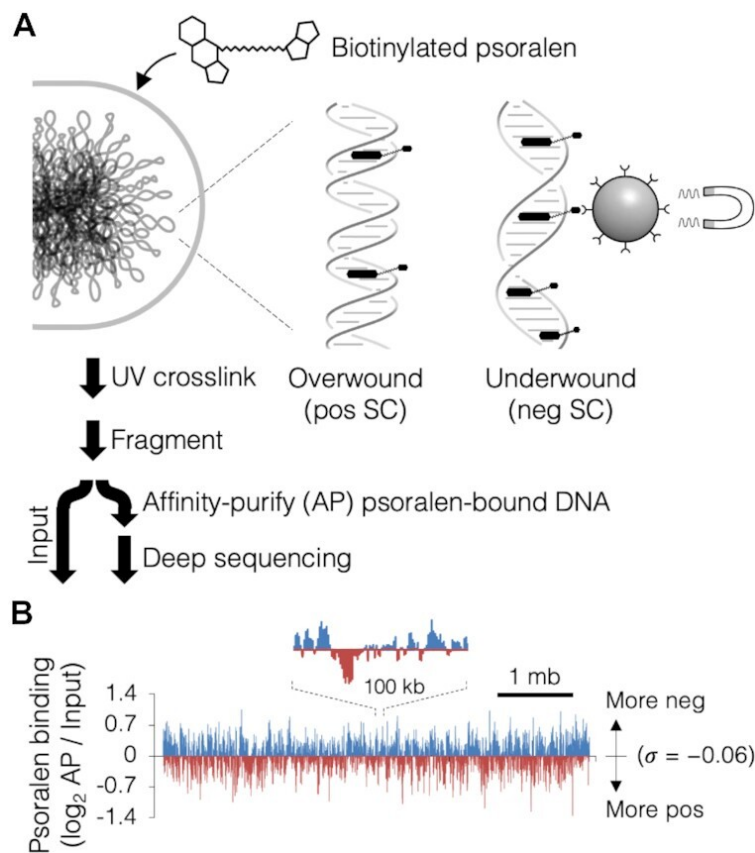


FIGURE 3.7 – La mesure du SC local par Psora-Seq. (A) Schématisation du protocole mis au point par Visser *et al.*. Les cultures bactériennes sont traitées avec du psoralen biotinylé qui se lie préférentiellement dans les régions surenroulées négativement. Les cultures sont ensuite soumises aux UV afin de créer des pontages entre les brins d'ADN via le TMP. L'ADN est ensuite extrait puis fragmenté. Les fragments contenant du TMP biotinylé peuvent être sélectionnés grâce à leur affinité pour la streptavidine. Les fragments sélectionnés ainsi qu'une banque de fragments non sélectionnés sont ensuite séquencés afin de comparer les échantillons avec sélection (AP) et sans sélection (input). (B) Profil de liaison du psoralen le long du génome de *E. coli* en phase exponentielle. Ce profil est obtenu en dénombrant les lectures par kb dans l'échantillon avec sélection (AP) et dans l'échantillon sans sélection (input) puis en calculant le  $\log_2$  du ratio entre ces deux dénombrements. Figure issue de [183].

Dans leur protocole, Visser *et al.* ont fragmenté l'ADN par sonication pour obtenir des fragments de 500 pb environ puis ils ont réalisé une purification par affinité des fragments

avec intercalation de psoralen biotinylé sur des billes magnétiques couplées à de la streptavidine. La biotine ayant une grande affinité pour la streptavidine, seuls les fragments pontés sont retenus par ces billes. Après lavage, les fragments retenus sont dissociés des billes dans une solution de formamide déionisé chauffée à 95°C. Ce dernier traitement a également été effectué sur une solution contenant l'ensemble des fragments (sans la sélection par affinité) afin d'avoir un contrôle (input). L'ensemble des échantillons ainsi préparés ont ensuite été séquencés [183].

La cartographie publiée par Visser *et al.* a permis de partager une première méthode et des premières données ouvrant la porte à l'approfondissement des nombreuses données de transcriptomique disponibles. Plutôt que de poursuivre la calibration de la méthode de M. Glinkowska pour pouvoir l'appliquer sur *D. dadantii*, nous avons donc décidé d'exploiter dans un premier temps les données publiées par Visser *et al.*. Cette analyse nous permettra notamment d'évaluer la pertinence de cette méthodologie afin de choisir à terme la méthode la plus adéquate pour la cartographie du SC local de *D. dadantii*. D'un point de vue technique, le principal avantage apparent de la méthodologie de Visser *et al.* semble être l'amélioration de la sélection des fragments pontés grâce à la biotine. L'inconvénient par rapport à la méthode proposée par M. Glinkowska est la résolution car les sites précis de liaisons du TMP au sein des fragments de 500 pb n'est pas identifiée. De plus, il est important de noter que l'intensité du signal détecté avec la méthode de Visser *et al.* est faible. Pas moins de 6 réplicats ont été nécessaires pour obtenir un signal distinguable du bruit de fond.



### 3.3 Analyse des cartographies du SC local obtenues par GapR-Seq et Psora-Seq

La topoI et la gyrase ayant respectivement plus d'affinité pour les régions SC- et SC+, l'étude du profil de fixation de ces topoisomérases avec le package GRATIOSA nous a permis de réaliser des premières analyses du SC local (voir la section 3.1.5). La fiabilité de ces données est cependant limitée par les autres paramètres pouvant influencer l'affinité des topoisomérases. Il est donc utile de comparer ces cartographies du SC à celles obtenues par d'autres méthodes. Pour ce faire, j'ai complété mon analyse du niveau de SC local du chromosome de *E. coli* avec les données de Psora-Seq et de GapR-Seq publiées respectivement par Visser *et al.* [183] et Guo *et al.* [182]. La méthode GapR-Seq a été présentée au Chapitre 1, section 1.5.2 (à partir de la page 51).

#### 3.3.1 Matériels et méthodes

##### Traitement des données brutes avec des outils standards

Les données de Psora-Seq sont disponibles sur *NCBI Sequence Read Archive*, avec la référence PRJNA805241. J'ai utilisé les réplicats SRR17974445, SRR17974447, SRR17974449, SRR17974481 et SRR17974483 pour les fragments sans sélection (*mock*) associés aux réplicats SRR17974444, SRR17974446, SRR17974448, SRR17974480 et SRR17974482 pour les fragments sélectionnés par affinité à la streptavidine (*pull-down*) (voir protocole présenté au paragraphe 3.2.3, en page 132). Les données GapR-Seq sont accessibles sur *NCBI Gene Expression Omnibus*, avec la référence GSE152882. J'ai utilisé les réplicats SRR12057816 et SRR12057817 en *mock* et SRR12057818, SRR12057819 et SRR12057820 en *pull-down*. Après avoir vérifié la qualité des données avec l'outil *FastQC*, j'ai effectué l'alignement des lectures sur le génome de référence de *E. coli* MG1655 (NC\_000913) avec *Bowtie2* [116]. Les alignements sont ensuite convertis en couverture avec *bamCoverage*, en choisissant la normalisation par RPKM [258]. Enfin, les  $\log_2$  des ratios entre les couvertures *mock* et *pull-down* sont calculés après avoir normalisé chacune des deux couvertures par leur couverture moyenne.

Les données d'expression utilisées ont été publiées avec les données de GapR-Seq et ont pour référence SRR12057814 [182]. Ces données ont été alignées sur le même génome de référence avec *tophat2* puis les FPKM ont été calculés avec *Cufflinks*.

##### Analyse des données pré-traitées avec le package GRATIOSA

Pour faciliter l'analyse des données de cartographie du SC local obtenues par GapR-Seq et Psora-Seq, j'ai utilisé le package GRATIOSA. J'ai pour cela créé un objet *Genome* auquel j'ai

attribué la séquence, un objet *Transcriptome* auquel j'ai attribué les données d'expression et un objet *Chipseq* sur lequel j'ai chargé les données de GapR-Seq et de Psora-Seq.

Une fois les données chargées en attributs des différents objets, j'ai aisément pu appliquer différentes fonctions du package afin de les analyser simultanément. Dans un premier temps, j'ai utilisé les données d'expression pour classer les TUs, à l'aide de la fonction de classification. Dans un second temps, j'ai utilisé les méthodes associées à la classe *Chipseq* pour traiter les signaux. J'ai tout d'abord lissé les signaux sur des fenêtres glissantes de 500 b. J'ai ensuite calculé, pour chaque position, la moyenne des signaux lissés des différents réplicats. En parallèle, j'ai calculé, à partir de la séquence, le contenu en AT le long du génome et j'ai également traité ce contenu comme un signal, en le lissant sur des fenêtres glissantes de 500 b également.

### 3.3.2 Résultats

L'analyse des signaux de fixation des topoI et de gyrase autour des TUs a mis en avant la fixation préférentielle de la topoI en amont et de la gyrase en aval des TUs (voir article en section 3.1.5). Cette fixation, conforme au modèle du *Twin supercoiled-domain*, semble indiquer un niveau local davantage SC- en amont et SC+ en aval des TUs, particulièrement lorsque ces TUs sont très exprimées. Pour reproduire cette analyse, j'ai tracé les signaux de GapR-Seq et de Psora-Seq de 10 kb en amont du site de début de transcription jusqu'à 10 kb en aval du site de fin de transcription, en classifiant les TUs selon leur niveau d'expression (Fig. 3.8).

Le signal obtenu avec le Psora-Seq (Fig. 3.8, partie haute) montre une différence entre le niveau de SC en amont et en aval des 800 TUs les plus exprimées. En effet, peu de psoralen semble s'être intercalé dans les 10 kb en aval des sites de fin de transcription de ces gènes, ce qui indique une région moins surenroulée négativement. Le plus marquant sur ces courbes sont les pics de psoralen au niveau des sites de début et de fin de transcription. Ces pics ne semblent cependant pas être dus à la transcription mais plutôt à la préférence de l'intercalation du psoralen dans les régions riches en AT puisque ces pics coïncident parfaitement avec le signal du contenu en AT (Fig. 3.8, partie basse), ce qui est cohérent avec le pontage préférentiel du psoralen avec les thymines (voir Chapitre 1, section 1.5.2).

Le signal obtenu par GapR-Seq (Fig. 3.8, partie intermédiaire), est plus hétérogène selon la catégorie d'expression. Pour les 200 TUs les plus exprimées (courbe bleue), une nette diminution est visible sur les 10 kb en amont des TUs. Cette faible présence indique un enrichissement en SC- en amont de ces TUs. Sur cette même courbe, un enrichissement en GapR est visible sur environ 2 kb en aval du TU. Cet enrichissement est visible sur une plus longue distance (7 kb environ) pour les 200 TUs classifiées à la suite (courbe bleue claire). Ces enrichissements pourraient indiquer une région riche en SC+ en aval des 400

TUs les plus exprimées. Le profil de SC local obtenu par GapR-Seq semble cependant également être biaisé du fait de l'encombrement stérique provoqué par l'ARNP. En effet, au sein de la TU, il semblerait que plus le niveau d'expression de cette TU est élevé, plus le signal de GapR est faible. Cette tendance pourrait signifier une diminution en SC+ au sein des TUs les plus exprimées mais cette déplétion peut également être due à la présence fréquente de l'ARNP dans cette région. Il est probable que GapR ne parvienne pas à se fixer dans une région encombrée par l'ARNP ou que la fixation de l'ARNP conduise à la dissociation de GapR de l'ADN.

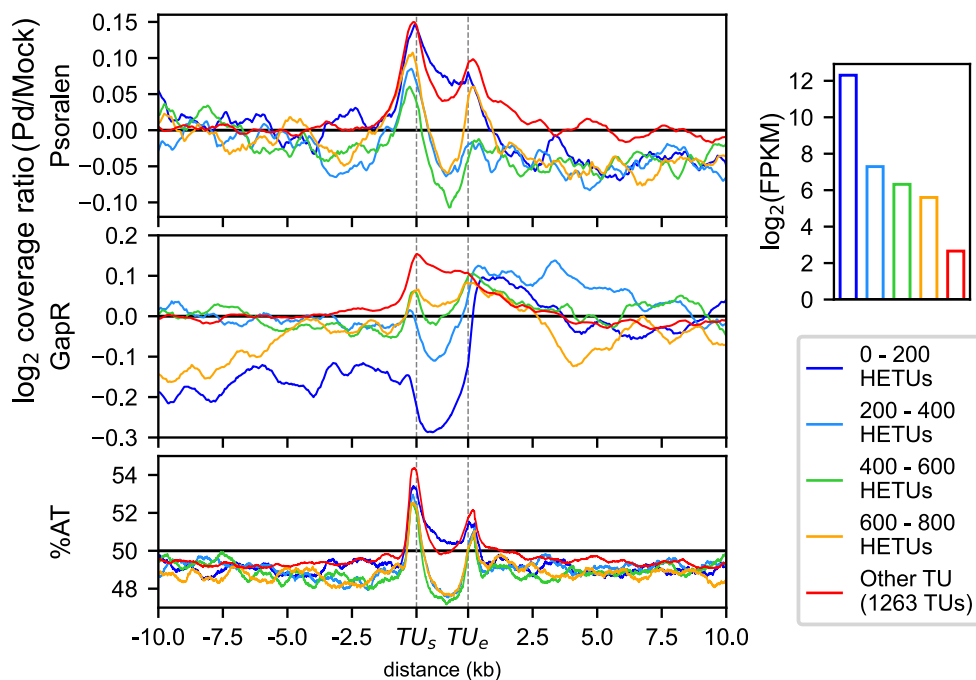


FIGURE 3.8 – Signaux de Psora-Seq [183], GapR-Seq [182] et du contenu en AT autour des TUs de *E. coli*. Les signaux de Psora-Seq et GapR-Seq sont les  $\log_2$  des ratios entre les couvertures obtenues après sélections des fragments (*Pull-down*, Pd) et des couvertures sans sélection (*Mock*). Les TUs ont été classifiées selon leur niveau d'expression, avec 200 TUs par catégorie pour les TUs les plus exprimées (*Highly Expressed Transcription Units*, HETUs) et avec une catégorie (rouge) contenant tous les autres TUs. Le diagramme à barres en haut à droite indique le niveau moyen d'expression de chaque catégorie.  $TU_s$  et  $TU_e$  indiquent les sites de début et de fin de transcription

En conclusion, les différences de fixation du psoralen et de GapR autour des TUs les plus exprimées indiquent qu'il semble y avoir davantage de SC- en amont et de SC+ en aval de ces TUs. Un déséquilibre de SC semble donc être créé lors du processus de transcription, conformément au modèle de *Twin-supercoiled domain*. Il semble difficile d'approfondir davantage l'analyse de ces signaux à cause des biais liés à la séquence et à l'encombrement stérique.

## 3.4 Etude de la topologie du chromosome de *D. dadantii* par Hi-C.

Pour étudier les interactions au sein du chromosome, le Hi-C est la méthode permettant d'obtenir la meilleure résolution. Nous avons donc décidé d'utiliser cette méthodologie sur *D. dadantii* pour obtenir la première carte de contact d'une bactérie phytopathogène.

### 3.4.1 Matériels et méthodes

#### Préparation des échantillons

Afin d'appliquer la méthode Hi-C (présentée au Chapitre 1, section 1.1.5, page 11), une culture de *D. dadantii* en milieu minimum M63 supplémenté de 0,2% de saccharose en phase exponentielle a été centrifugée, lysée avec le Precellys Evolution puis traitée avec le kit Arima-HiC Kit. Pour la fragmentation du chromosome, les enzymes HinfI, Sau3AI ont été utilisés pour les 3 premiers réplicats puis une enzyme supplémentaire, MluCI, a été ajoutée pour les 2 derniers réplicats. Les échantillons ont ensuite été purifiés avec le kit AMPure XP Beckman. Les étapes de préparation au séquençage (fragmentation supplémentaire, amplification PCR, préparation des libraires avec le kit Colibri) et le séquençage Illumina ont été réalisés par l'entreprise Seqalis.

#### Traitement des données brutes

Pour le traitement des données brutes, j'ai utilisé deux pipelines développés par Matthey-Doret *et al.* : *hicstuff* [259] et *chromosight* [260].

Le pipeline *hicstuff* (1) effectue l'alignement des lectures sur le génome, en faisant appel à *Bowtie2*, (2) identifie les paires de lectures, (3) associe chaque lecture à un fragment de restriction en prenant en compte les enzymes de restriction utilisées, (4) peut filtrer les paires problématiques issues de religations ou d'erreur de digestion par exemple, et (5) génère une matrice en comptant le nombre d'occurrences de chaque paire de fragments de restriction [259]. *hicstuff* propose également une normalisation afin de notamment prendre en compte la longueur des fragments de restriction et le contenu en AT. La résolution de la matrice générée peut également être modifiée en choisissant une taille de *bin*.

A partir de ces matrices, le pipeline *chromosight* peut être utilisé. Ce pipeline permet de détecter des motifs sur les matrices, notamment les boucles (*loops*) et les frontières entre les domaines topologiques (*borders*) tels que présentés sur la Figure 3.9. Les coordonnées d'une boucle (encadrée en vert sur la Figure 3.9) sont les positions des deux fragments en contact à la base de la boucle. La boucle apparaît comme un point de forte intensité

sur la matrice de contact (Fig. 3.9 A et C). Une frontière a une unique position, celle de la séparation entre deux domaines topologiques (i.e. deux régions présentant un grand nombre d'interactions au sein de la région). Sur la matrice, un domaine topologique prend la forme d'un carré de forte intensité autour de la diagonale centrale (voir TAD sur la Fig. 3.9 A). La frontière est donc identifiable à la jonction de deux carrés de fortes intensités (Fig. 3.9 A et D). Ainsi, pour détecter les boucles et les frontières, le pipeline compare chaque position de la matrice à des motifs de référence représentant distinctement une boucle ou une frontière. Le pipeline produit alors un coefficient de corrélation représentant la similarité entre le motif de référence et le motif observé sur la matrice.

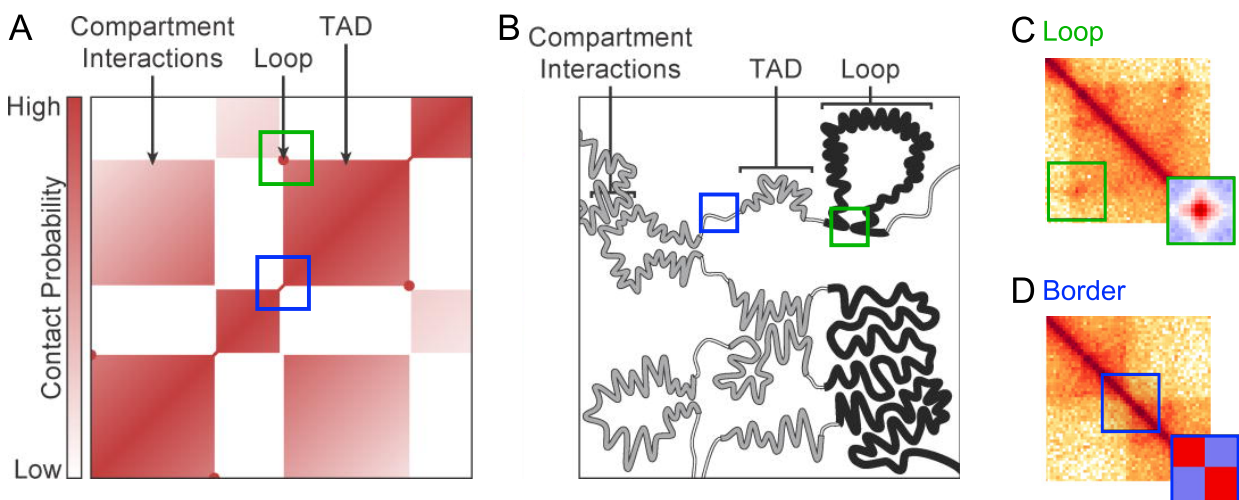


FIGURE 3.9 – Les motifs détectables sur une matrice de contact obtenue par Hi-C, en particulier les boucles (*loops*) et les frontières entre domaines topologiques (*borders*) respectivement encadrés en vert et en bleu. (A) Schématisation d'une carte de contact. La probabilité de contact de chaque paire de positions est représentée avec l'intensité de la couleur rouge. Plus la probabilité est élevée, plus le rouge est intense. Les domaines topologiques TADs sont des régions riches en contacts, ce qui se traduit par un carré rouge intense le long de la diagonale. Les boucles (*Loop*) sont des pics d'intensité. Les carrés de grande intensité distants de la diagonale représentent des interactions liées à la compartimentation du chromosome. (B) Schématisation du repliement du chromosome avec les domaines topologiques, les boucles et les interactions liées à la compartimentation. (C) et (D) Extraits de matrice de contacts représentant une boucle (C) ou une frontière (D) et les motifs modèles recherchés par l'outil *Chromosight*. Figures issues de [261, 260]

A partir des données de séquençage obtenues pour *D. dadantii*, j'ai pu générer une carte de contacts avec une résolution de 2 kb pour chacun des 5 répliquats. J'ai ensuite détecté les boucles et les frontières avec un seuil de corrélation de Pearson avec le motif de référence de 0,5 pour les frontières et 0,3 pour les boucles.

Les fichiers générés par *Chromosight* sont au format .tsv et contiennent une ligne par boucle ou par frontière. Les différentes colonnes représentent les coordonnées (deux colonnes dans le cas des boucles, une colonne dans le cas des frontières) ainsi que des scores de

corrélation et des valeurs-p. Ces fichiers peuvent être importés avec le package GRATIOSA afin d'intégrer ces coordonnées dans les analyses proposées par le package.

### 3.4.2 Les résultats préliminaires

#### Les frontières et les boucles détectées

*Chromosight* a identifié un grand nombre de boucles et de frontières (Tableau 3.2, Fig. 3.10 A et B) dans chacun des réplicats. Afin d'améliorer la précision de ces détections, j'ai comparé les positions trouvées dans chacun de ces réplicats et je n'ai conservé que celles communes à minimum trois réplicats, en acceptant un écart de 2 kb. J'ai ainsi sélectionné 52 frontières et 81 boucles.

TABLEAU 3.2 – Nombre de frontières et de boucles détectées dans le chromosome de *D. dadantii* par réplicat.

Réplicat	Nombre de frontières	Nombre de boucles
Jan21.1	53	106
Avr21.1	60	129
Avr21.2	44	71
Nov22.1	65	171
Nov22.2	67	192

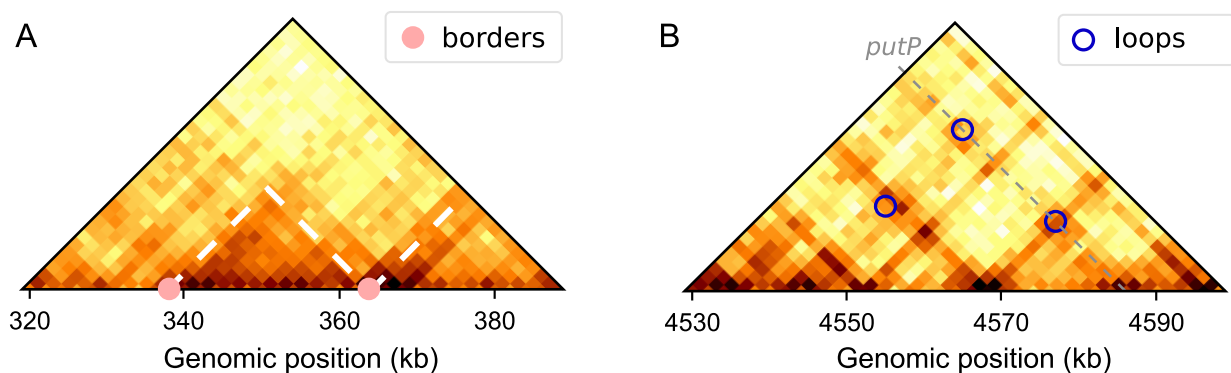


FIGURE 3.10 – Extraits d'une carte de contact de *D. dadantii* dans laquelle des frontières (A) ou des boucles (B) ont été détectées par *Chromosight*. Les points roses indiquent les positions des frontières et les lignes pointillées blanches mettent en avant les domaines. Les cercles bleus indiquent les boucles et la ligne grise pointillée indique les interactions du gène *putP*, qui code pour un transporteur localisé dans la membrane interne des cellules.

### Étude du niveau d'expression des gènes proches des motifs identifiés

Des études précédemment réalisées chez *Caulobacter crescentus*, *B. subtilis* et *E. coli* ont mis en avant une corrélation entre la forte expression des gènes et la présence de frontières entre les domaines topologiques [59, 60, 57]. A l'aide du package GRATIOSA et de données d'expression précédemment publiées [34], j'ai analysé le niveau d'expression des gènes situés au niveau des frontières et des boucles identifiées chez *D. dadantii*. J'ai alors observé que les frontières entre les domaines de *D. dadantii* sont également enrichies en gènes fortement exprimés et que cet enrichissement est encore plus important au niveau des boucles (Fig. 3.11). Ces observations suggèrent que les contraintes topologiques associées à l'expression des gènes peuvent être liées à la formation de structures 3D ou, à l'inverse, que ces structures peuvent impacter l'expression des gènes.

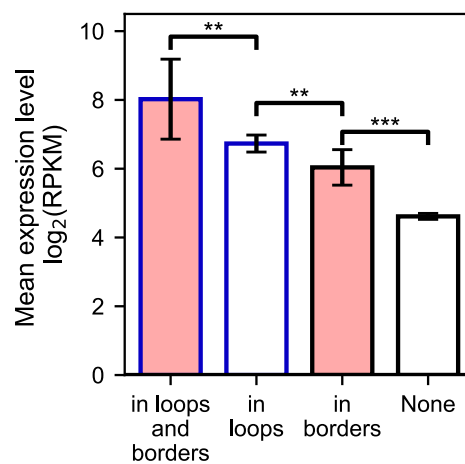


FIGURE 3.11 – Niveau moyen d'expression des gènes selon leur proximité à l'un, l'autre, les deux ou aucun des motifs analysés (frontière en rose, boucle en bleu). Les barres d'erreurs représentent l'intervalle de confiance à 95%.

### Étude du mécanisme de transertion et de son impact sur la topologie du chromosome

Une hypothèse de longue date postule que les gènes très exprimés codant pour des protéines membranaires sont localisés près de la membrane, permettant ainsi un couplage efficace transcription-traduction-insertion dans la membrane (Fig. 3.12) [262].

Ce mécanisme, nommé transertion, a été testé expérimentalement chez *E. coli*, en induisant l'expression de protéines membranaires et en observant le positionnement des gènes par microscopie fluorescente. Libby *et al.* ont ainsi observé le repositionnement de ces gènes proches de la membrane [263]. Une telle contrainte mécanique pourrait contribuer à l'organisation et à la régulation du chromosome bactérien.

*D. dadantii* est une bactérie phytopathogène qui interagit fortement avec son environnement extérieur et possède donc de multiples systèmes de sécrétion, notamment pour la sécrétion

des facteurs de virulence [27]. Cette bactérie est donc un modèle pertinent pour l'étude du mécanisme de transertion et de son impact sur la conformation du chromosome.

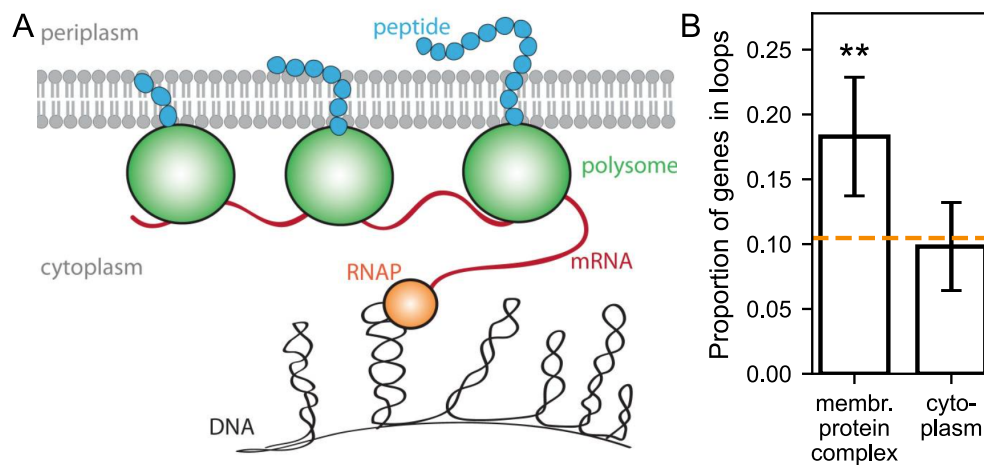


FIGURE 3.12 – (A) Schématisation de l'hypothèse de la transertion : couplage de la transcription, de la traduction et de l'insertion des protéines dans la membrane. Schéma issu de [264].

(B) Enrichissement en gènes proches des boucles parmi les gènes associés au terme GO *membrane protein complex*. Le terme GO *cytoplasm* est ici un contrôle négatif. Les proportions au sein des catégories GO peuvent être comparées à la moyenne génomique indiquée avec la ligne pointillée orange. Cette ligne indique la proportion de gènes du génome qui sont considérés comme proches des boucles. Les barres d'erreur indiquent l'intervalle de confiance à 95%. Figure générée avec le package GRATIOSA.

Nous avons donc souhaité tester si les gènes codant pour des protéines membranaires étaient impliqués dans la formation des boucles. Pour ce faire, nous avons utilisé l'annotation, fournie par Gene Ontology (GO) [119], de la localisation cellulaire de l'activité des protéines codées par chaque gène (*cellular component*). J'ai alors constaté un enrichissement significatif en gènes situés à proximité des boucles (i.e. situés proches de l'une des deux extrémités en contact de la boucle) parmi les gènes codant pour des protéines appartenant à des complexe protéiques membranaires (Fig. 3.12B). Cet enrichissement suggère que le processus de transertion pourrait contribuer à l'organisation du chromosome en créant des barrières topologiques qui perturbent la diffusion des supertours et induisent la formation de boucles. En revanche, je n'ai pas constaté d'enrichissement pour les gènes situés à proximité des frontières entre les domaines topologiques.





## Chapitre 4

# Discussion et perspectives

Le SC est un mécanisme omniprésent dans le chromosome. Il impacte sa structure, à l'échelle de la paire de base et jusqu'à la conformation globale du chromosome. Depuis plus de 50 ans, des simulations biophysiques ainsi que des expériences *in vitro* et *in vivo*, sont réalisées afin de mieux caractériser l'organisation du chromosome et le rôle de cette structure dans les processus biologiques tels que la réplication et la transcription [265, 266, 163, 62]. Malgré ces années d'étude, de nombreuses questions persistent. Pendant ma thèse, j'ai tâché d'apporter quelques réponses à certaines d'entre elles. Comment le SC est-il réparti le long du chromosome? Quels sont les gènes affectés par les variations de SC? Quel rôle joue le SC dans la virulence de la bactérie phytopathogène *D. dadantii*?

### 4.1 L'organisation du chromosome bactérien

Les analyses à différentes échelles du chromosome bactérien ont précédemment mis en évidence son organisation en macrodomaines d'environ 1 Mpb [53, 54], en CIDs d'environ  $10^4 - 10^5$  pb [57, 60] et sa compaction à plus petite échelle par l'action conjointe du SC, des NAPs et des protéines SMC [145, 68, 77] (voir Chapitre 1 section 1.1.5, à partir de la page 8).

Jusqu'à récemment, les méthodes de mesure du SC permettaient uniquement d'avoir une mesure globale du niveau de SC [146] alors qu'il était connu que ce niveau est hétérogène le long du chromosome [181]. Le récent développement de méthodes de cartographie du SC local [182, 183] est donc une étape clé pour l'étude de l'organisation du chromosome par le SC. Dans le Chapitre 3 j'ai exploité ces nouvelles méthodologies et données publiées chez *E.coli* [182, 183, 197, 191] en utilisant le package Python GRATIOSA que nous avons développé pour faciliter les analyses statistiques de l'organisation du chromosome bactérien et de son lien avec la transcription. Les données de ChIP-Seq de la topoI et de la gyrase [197, 191] ainsi que les données de cartographie du SC local obtenues par GapR-Seq [182] et Psora-Seq [183], nous ont permis d'observer les *Twin-supercoiled domains* [181], avec un enrichissement en SC- et en SC+ à une échelle de 10 kb

respectivement en amont et en aval des TUs les plus transcrites. Malgré ces observations encourageantes, il est important de noter que les méthodes actuelles de cartographie du SC local ne reflètent qu'une vision statique d'un groupe de cellules. Au sein de chaque cellule, le SC est dynamique, avec des torsions qui peuvent se transformer en vrillages et *vice-versa* [85]. Ces vrillages et torsions ne sont pas détectés séparément avec les méthodes actuelles de cartographie. Le Psora-Seq par exemple indique uniquement le niveau de torsions [183]. Il sera donc essentiel de combiner les cartographies obtenues avec différentes méthodes afin d'obtenir une cartographie plus précise du SC local, même si celle-ci restera une vue d'ensemble d'un groupe de cellules. En ce qui concerne le Psora-Seq, il est dès à présent possible d'améliorer la méthodologie en ajoutant un contrôle supplémentaire avec de l'ADN fragmenté avant l'intercalation du psoralen. Ce contrôle permettrait d'évaluer le taux d'intercalation de psoralen dépendant uniquement de la séquence, dont son contenu en AT, et non du niveau de SC.

Une autre piste enthousiasmante ouverte avec les cartographies du SC local est l'étude du lien entre SC et NAPs. Les NAPs sont présentes en abondance dans la cellule et impactent la structure du chromosome mais leur rôle sur le SC est encore mal caractérisé. Des études avec des molécules d'ADN uniques tenues par des pinces magnétiques ont mis en évidence l'impact de certaines NAPs, dont H-NS et HU, sur le SC [267, 268]. Une méthode de simulation développée par Dahlke *et al.* a également indiqué que les courbures induites par la fixation de FIS favorisent énergétiquement la formation de supertours [269]. Les NAPs semblent donc pouvoir affecter significativement le niveau local de SC et il semble probable que le niveau de SC affecte également l'affinité des NAPs. Cette corrélation pourra être étudiée *in vivo* en combinant les cartographies du SC local avec les données ChIP-Seq de NAPs déjà disponibles, dont celles obtenues pour H-NS, HU et Fis chez *E. coli* [270, 271]. Ces analyses pourront aisément être réalisées à l'aide du package GRATIOSA. En important les positions des pics de fixation des NAPs, il sera par exemple possible d'analyser le niveau de SC local autour de ces pics afin d'obtenir des informations sur l'impact du SC local sur la fixation des NAPs ou *vice-versa*.

En analysant les données de ChIP-Seq de gyrase nous avons également observé une organisation à plus grande échelle avec une corrélation du signal jusqu'à 100-200 kb. Cette corrélation, de taille cohérente avec celle des domaines identifiés par Liroy *et al.* [57] avec la technologie Hi-C, pourrait correspondre à une échelle de compaction peu étudiée jusqu'à présent et qu'il serait intéressant d'approfondir. Pour combiner l'ensemble des informations apportées par les cartes Hi-C avec les nouvelles données de ChIP-Seq, afin notamment d'étudier cette échelle d'organisation du chromosome, il serait utile de faire évoluer le package GRATIOSA. En effet, ce package ne permet actuellement d'intégrer que des données 1D, c'est à dire des données par position génomique. L'intégration de

données 3D serait une évolution pertinente de ce package afin de prendre en compte des données d'interactions spatiales entre des régions distinctes du chromosome, telles que celles fournies dans les cartes de contacts obtenues par Hi-C. Pour ce faire, il serait utile dans un premier temps de créer un nouvel attribut dans la classe *hic* qui indiquerait, pour chaque position, les données quantitatives sur les interactions avec chaque autre position du chromosome. Cet attribut correspondrait à la matrice de contacts Hi-C. La normalisation de ces matrices ne serait pas intégrée au package car elle peut être réalisée avec des outils aboutis tels que *hicstuff* [259], cependant le package pourrait permettre de combiner les réplicats de ces matrices en calculant par exemple, pour chaque interaction, la moyenne entre les réplicats. Le package pourra également déterminer la valeur moyenne d'interaction entre les gènes annotés. En combinant ces données avec des données de type ChIP-Seq, il serait également possible d'étudier les interactions spatiales entre les régions ciblées par une même ou par plusieurs protéines étudiées par ChIP-Seq. Ce type d'analyse permettra de combiner les données de ChIP-Seq des topoisomérases avec les données de Hi-C et pourra également permettre de confronter le profil de fixation des NAPs à ces mêmes données de Hi-C.

L'intégration des données de Hi-C dans le package GRATIOSA serait d'autant plus pertinente que de nombreuses données de Hi-C ont été collectées par mon équipe chez *D. dadantii* afin d'identifier les interactions au sein de son chromosome lorsque la bactérie est soumise à différents stress, dont ceux rencontrés lors du processus d'infection des plantes. L'analyse préliminaire de ces données de Hi-C a révélé que le niveau d'expression des gènes est lié à la formation de structures 3D dont les domaines et les boucles. De plus, nous avons remarqué que les gènes codant pour des protéines membranaires sont davantage situés au niveau des extrémités des boucles formées par le chromosome. Cette observation pourrait être liée au positionnement des gènes très exprimés codant pour des protéines membranaires près de la membrane (mécanisme de transertion [262]). Ce positionnement des gènes a été observé par microscopie chez *E. coli*, en induisant l'expression de protéines membranaires [263]. Il a également été montré, avec une construction sur plasmide, que l'ancrage à la membrane constitue une barrière de diffusion des supertours et peut ainsi affecter l'expression des gènes autour de cette barrière [272]. Ainsi, il semble possible que le mécanisme de transertion, avec le positionnement des gènes codant pour les protéines membranaires, provoque une accumulation de SC au niveau de cet ancrage et que cette accumulation conduise à la création de structures observables avec la méthodologie Hi-C. Afin de tester l'hypothèse selon laquelle l'ancrage des gènes à la membrane provoque l'apparition de ces structures, des constructions génétiques seront réalisées par mon équipe pour ancrer artificiellement des régions du chromosome à la membrane. La technique Hi-C sera ensuite appliquée à ces bactéries génétiquement modifiées afin d'observer les éventuelles structures formées dans le chromosome suite à ces ancrages. Il sera pertinent

de réaliser différentes constructions avec des ancrages permettant de tester l'impact du niveau d'expression et de l'orientation des gènes situés autour de ces points d'ancrage. Ces expérimentations seront financées par l'Agence Nationale de la Recherche (ANR) qui a accepté de soutenir ce projet suite à la présentation de nos résultats préliminaires.

Ainsi, grâce au développement de nouvelles méthodes donc le Hi-C et les techniques de cartographie du SC local, nos connaissances sur l'organisation du chromosome se précisent et se préciseront encore davantage dans les prochaines années. Mes analyses des cartographies avec le package GRATIOSA ont notamment permis de confirmer la présence *in vivo* des *Twin-supercoiled domain* dont l'intensité semble dépendre du niveau d'expression des TUs. Il est donc certain que la transcription impacte le niveau local de SC et il a également été montré que le SC impacte la transcription. Ce second point était au coeur du Chapitre 2 de ma thèse.

## 4.2 La régulation des gènes par le surenroulement de l'ADN

De précédentes études *in vivo* ont démontré que le SC est un régulateur global de l'expression des gènes [161, 122, 162, 163, 62, 83], cependant ce mécanisme de régulation est encore mal compris et la quasi absence de données sur la réponse des gènes à une augmentation du niveau de SC fait partie des limites à cette compréhension. J'ai donc analysé la réponse transcriptionnelle de *D. dadantii* à une augmentation du niveau global de SC, obtenue grâce à un traitement novateur avec un inhibiteur de topol : la séconéolitsine (Chapitre 2). En obtenant le premier transcriptome d'une bactérie à Gram négatif dont la topol est inhibée par un antibiotique, nous rendons possible la comparaison des réponses transcriptomnelles au relâchement de l'ADN et à l'augmentation du niveau de SC. Notre étude a permis de prouver que, quelle que soit la direction de cette variation, une modification du niveau de SC affecte l'expression des gènes de manière globale et complexe et qu'il est donc essentiel de prendre en compte ce mécanisme dans les études et les modèles de l'expression du génome bactérien. Dans notre article publié dans *Nucleic Acids Research* [255], nous avons remis en question la notion de gène "sensible au SC". En effet, le SC régule un grand nombre de gènes mais la réponse d'un gène dépend notamment de la direction de la variation de SC, du contexte génomique et du niveau d'expression du gène ainsi que de l'état physiologique de la bactérie. Ainsi, l'expression d'un gène peut être affectée par le relâchement de l'ADN mais pas par une augmentation de SC.

Un mécanisme qu'il serait intéressant d'approfondir suite à cette étude est l'impact de l'orientation des gènes sur le SC local et sur la transcription de ces gènes. Il a été montré que des supertours négatifs s'accumulent en aval des gènes activement transcrits [191],

conformément au *Twin-supercoiled domain* [181] et en particulier entre les opérons divergents [273]. Ce mécanisme, initialement identifié sur le promoteur *leu* de *Salmonella enterica* [274], semble être répandu dans le génome de *D. dadantii* puisque nous avons observé que les gènes situés entre des gènes divergents sont significativement plus réprimés que ceux situés entre des gènes convergents. Nous nous y attendions puisque la topoI est nécessaire au retrait de cet excès de supertours négatifs. Étonnamment, le même profil de répression a été observé suite à l'inhibition de la gyrase : les gènes situés entre des gènes divergents sont davantage réprimés que ceux situés entre des gènes convergents. Ces observations permettent d'affirmer que l'activité locale des topoisomérases dépend de l'orientation des gènes, mais le profil d'expression obtenu avec l'inhibition de la gyrase soulève des questions sur les mécanismes de régulation des supertours positifs en absence de gyrase et sur l'impact réel de l'accumulation de ces supertours sur la transcription. Un profil similaire avait été observé chez *E. coli* [62, 166] et il serait intéressant d'analyser les réponses des gènes observées chez d'autres organismes suite à l'inhibition de leur gyrase [275, 276] afin d'observer si ce profil est conservé. Le package GRATIOSA sera un outil adéquat pour transposer cette même analyse chez différents organismes. En effet, les fichiers utilisés dans la BDD associée au package sont des fichiers standards d'annotation ou de données expérimentales. Ainsi, un utilisateur souhaitant étudier un nouvel organisme peut aisément trouver les fichiers d'annotation au bon format, importer ses données expérimentales puis reproduire les analyses, en utilisant les mêmes scripts. La reproductibilité et la traçabilité de ces analyses sont des points forts de ce package.

Maintenant que des cartographies du SC local sont disponibles, les intégrer dans les études transcriptomiques, à l'aide du package GRATIOSA, permettra de préciser grandement les analyses de la régulation de la transcription par le SC. Lorsque des méthodes de cartographie complètes seront disponibles, il sera pertinent de les appliquer dans différentes conditions de stress et chez différents organismes, dont *D. dadantii*, afin de pouvoir étudier avec précision les variations locales du niveau de SC provoquées par ces stress et l'impact sur la transcription de ces variations. En réalisant une cartographie du SC local après traitement à la séconéolitsine et à la novobiocine, il sera possible de savoir si les variations de SC observées globalement affectent le chromosome de manière homogène ou seulement certaines régions. Dans ce second cas, il sera intéressant de savoir si les régions dont le niveau de SC est particulièrement affecté correspondent aux localisations des gènes différentiellement exprimés lors de ces traitements. En combinant l'ensemble des données concernant l'impact du SC sur l'expression des gènes il sera peut être alors possible de proposer un mécanisme de régulation transcriptionnelle par le SC bien plus précis que le modèle qualitatif proposé dans l'article présenté au Chapitre 2.

### 4.3 Le rôle du surenroulement de l'ADN dans la virulence des *Dickeya* spp.

*D. dadantii* est une bactérie pertinente pour l'étude de l'effet du SC sur la régulation de la transcription puisque ce mécanisme semble être impliqué dans la régulation de la virulence de cette bactérie phytopathogène. Notre étude avec l'inhibition de la topoI a confirmé que les variations de SC affectent l'expression des enzymes de dégradation des parois cellulaires de végétaux, qui provoquent les symptômes de la pourriture molle. Ces enzymes sont également présentes chez les autres espèces du genre *Dickeya*, notamment *D. solani* et *D. dianthicola* [277, 278] qui menacent actuellement les cultures agricoles européennes [30, 31]. Ces différentes espèces phytopathogènes sont, similairement à *D. dadantii*, exposées à des stress environnementaux pendant le processus infectieux. Il est vraisemblable que le SC joue un rôle de senseur de ces stress et régule l'expression des facteurs de virulence selon le contexte environnemental. Comment le SC de ces différentes bactéries varie-t-il pendant les étapes de l'infection ? Quel effet ces variations de SC ont-elles sur l'induction des facteurs de virulence ? Pour étudier ce lien entre stress environnementaux, modification du SC et induction des facteurs de virulence, il serait intéressant, d'une part, de mesurer le niveau de SC des bactéries *in planta* lors des différentes étapes de l'infection et, d'autre part, de réaliser des tests de virulence des différentes espèces du genre *Dickeya* en réponse à des modifications de leur niveau de SC. Pour réaliser les tests en réponse à une augmentation du niveau de SC, il sera cependant nécessaire de trouver un substitut à la séconéolitsine car ses difficultés de solubilité ne permettent pas de l'utiliser dans des conditions reproduisant le contexte de l'infection. En effet, la séconéolitsine précipite lors des baisses de température et des dilutions en milieu aqueux. Des recherches sont actuellement en cours au sein de l'ICBMS afin d'identifier un composant, proche de la séconéolitsine, ayant la même faculté d'inhibition de la topoI et une solubilité améliorée. Ce nouveau composant pourra permettre de réaliser les tests de virulence avec les différentes bactéries afin d'étudier l'impact des variations de SC sur leur virulence. Mieux comprendre les mécanismes de régulation de la virulence de ces pathogènes pourrait, à terme, permettre d'élaborer de nouveaux moyens de lutte contre ces menaces croissantes.

# Bibliographie

1. WACEY D, KILBURN M, SAUNDERS M, CLIFF J et BRASIER M. Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nature Geoscience*. 2011 21; 4:698-702. DOI : 10.1038/ngeo1238
2. DODD MS, PAPINEAU D, GRENE T, SLACK JF, RITTNER M, PIRAJNO F, O'NEIL J et LITTLE CTS. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature*. 2017; 543:60-4. DOI : 10.1038/nature21377
3. ANANTHANARAYAN R. Ananthanarayan and Paniker's Textbook of Microbiology. Orient Blackswan, 2006. 672 p.
4. SASTRY AS et K SB. Essentials of Medical Microbiology. JP Medical Ltd, 2018 31. 716 p.
5. PORTER JR. Antony van Leeuwenhoek : tercentenary of his discovery of bacteria. *Bacteriological Reviews*. 1976; 40:260-9
6. BONNET M, LAGIER JC, RAOULT D et KHELAIPIA S. Bacterial culture through selective and non-selective conditions : the evolution of culture media in clinical microbiology. *New Microbes and New Infections*. 2020; 34:100622. DOI : 10.1016/j.nmni.2019.100622
7. PRESCOTT LM, WILLEY JM, SHERWOOD LM et WOOLVERTON CJ. Microbiologie. De Boeck Supérieur, 2018 3. 1122 p.
8. DREWS G. The roots of microbiology and the influence of Ferdinand Cohn on microbiology of the 19th century. *FEMS microbiology reviews*. 2000; 24:225-49. DOI : 10.1111/j.1574-6976.2000.tb00540.x
9. LWOFF A. The concept of virus. *Journal of General Microbiology*. 1957; 17:239-53. DOI : 10.1099/00221287-17-2-239
10. STANIER RY et VAN NIEL CB. The concept of a bacterium. *Archiv Fur Mikrobiologie*. 1962; 42:17-35. DOI : 10.1007/BF00425185
11. HENKIN TM et PETERS JE. Snyder and Champness Molecular Genetics of Bacteria. John Wiley & Sons, 2020 27. 642 p.
12. MARSHALL WF, YOUNG KD, SWAFFER M, WOOD E, NURSE P, KIMURA A, FRANKEL J, WALLINGFORD J, WALBOT V, QU X et ROEDER AH. What determines cell size? *BMC Biology*. 2012 14; 10:101. DOI : 10.1186/1741-7007-10-101
13. COICO R. Gram staining. *Current Protocols in Immunology*. 2001; Appendix 3:A.3O.1-A.3O.2. DOI : 10.1002/0471142735.ima03os23
14. BEVERIDGE TJ et DAVIES JA. Cellular responses of *Bacillus subtilis* and *Escherichia coli* to the Gram stain. *Journal of Bacteriology*. 1983; 156:846-58



15. SCHLEIFER KH. Classification of Bacteria and Archaea : Past, present and future. *Systematic and Applied Microbiology*. 2009 1; 32:533-42. DOI : 10.1016/j.syapm.2009.09.002
16. OTTEN C, BRILLI M, VOLLMER W, VIOLLIER PH et SALJE J. Peptidoglycan in obligate intracellular bacteria. *Molecular Microbiology*. 2018; 107:142-63. DOI : 10.1111/mmi.13880
17. BROWNING DF, HOBMAN JL et BUSBY SJW. Laboratory strains of *Escherichia coli* K-12 : things are seldom what they seem. *Microbial Genomics*. 2023 6; 9:mgen000922. DOI : 10.1099/mgen.0.000922
18. BLATTNER FR, PLUNKETT G, BLOCH CA, PERNA NT, BURLAND V, RILEY M, COLLADOVIDES J, GLASNER JD, RODE CK, MAYHEW GF, GREGOR J, DAVIS NW, KIRKPATRICK HA, GOEDEN MA, ROSE DJ, MAU B et SHAO Y. The complete genome sequence of *Escherichia coli* K-12. *Science (New York, NY)*. 1997 5; 277:1453-62. DOI : 10.1126/science.277.5331.1453
19. COBO-SIMÓN M, HART R et OCHMAN H. *Escherichia Coli* : What Is and Which Are? *Molecular Biology and Evolution*. 2023 4; 40:msac273. DOI : 10.1093/molbev/msac273
20. ROSANO GL, MORALES ES et CECCARELLI EA. New tools for recombinant protein production in *Escherichia coli* : A 5-year update. *Protein Science : A Publication of the Protein Society*. 2019; 28:1412-22. DOI : 10.1002/pro.3668
21. HUGOUVIEUX-COTTE-PATTAT N, CONDEMINÉ G, GUEGUEN E et SHEVCHIK VE. *Dickeya* Plant Pathogens. *eLS*. John Wiley & Sons, Ltd, 2020 :1-10. DOI : 10.1002/9780470015902.a0028932
22. LAURILA J, HANNUKALA A, NYKYRI J, PASANEN M, HÉLIAS V, GARLANT L et PIRHONEN M. Symptoms and yield reduction caused by *Dickeya* spp. strains isolated from potato and river water in Finland. *European Journal of Plant Pathology*. 2010 1; 126:249-62. DOI : 10.1007/s10658-009-9537-9
23. SARRIS P, TRANTAS E, PAGOULATOU M, STAVROU D, VERVERIDIS F et GOUMAS D. First report of potato blackleg caused by biovar 3 *Dickeya* sp. (*Pectobacterium chrysanthemi*) in Greece. *New Disease Reports*. 2011; 24:21-1. DOI : 10.5197/j.2044-0588.2011.024.021
24. CZAJKOWSKI R, PÉROMBELON MCM, VEEN JA van et WOLF JM van der. Control of blackleg and tuber soft rot of potato caused by *Pectobacterium* and *Dickeya* species : a review. *Plant Pathology*. 2011; 60:999-1013. DOI : 10.1111/j.1365-3059.2011.02470.x
25. HEALTH (PLH) EPoP. Scientific Opinion on the risk of *Dickeya dianthicola* for the EU territory with identification and evaluation of risk reduction options. *EFSA Journal*. 2013; 11:3072. DOI : 10.2903/j.efsa.2013.3072
26. REVERCHON S, MUSKHELISVILI G et NASSER W. Virulence Program of a Bacterial Plant Pathogen : The *Dickeya* Model. *Progress in Molecular Biology and Translational Science*. 2016; 142:51-92. DOI : 10.1016/bs.pmbts.2016.05.005
27. REVERCHON S et NASSER W. *Dickeya* ecology, environment sensing and regulation of virulence programme. *Environmental Microbiology Reports*. 2013; 5:622-36. DOI : 10.1111/1758-2229.12073
28. TOTH IK, WOLF JM van der, SADDLER G, LOJKOWSKA E, HÉLIAS V, PIRHONEN M, TSROR (LAHKIM) L et ELPHINSTONE JG. *Dickeya* species : an emerging problem for potato production in Europe. *Plant Pathology*. 2011; 60:385-99. DOI : 10.1111/j.1365-3059.2011.02427.x

29. JOKO T, HIRATA H et TSUYUMU S. Sugar transporter (MfsX) of the major facilitator super-family is required for flagella-mediated pathogenesis in *Dickeya dadantii* 3937. *Journal of General Plant Pathology*. 2007 1 ; 73:266-73. DOI : 10.1007/s10327-007-0018-8
30. OSDAGHI E. *Dickeya solani* (black leg disease of potato). *CABI Compendium*. 2022 4; CABI Compendium:120278. DOI : 10.1079/cabicompendium.120278
31. *Dickeya dianthicola* (ERWICD)[World distribution] | EPPO Global Database
32. HUGOUVIEUX-COTTE-PATTAT N, PÉDRON J et VAN GIJSEGEM F. Insight into biodiversity of the recently rearranged genus *Dickeya*. *Frontiers in Plant Science*. 2023 20; 14:1168480. DOI : 10.3389/fpls.2023.1168480
33. GLASNER JD, YANG CH, REVERCHON S, HUGOUVIEUX-COTTE-PATTAT N, CONDEMINÉ 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 et PERNA NT. Genome Sequence of the Plant-Pathogenic Bacterium *Dickeya dadantii* 3937. *Journal of Bacteriology*. 2011 ; 193:2076-7. DOI : 10.1128/JB.01513-10
34. JIANG X, SOBETZKO P, NASSER W, REVERCHON S et MUSKHELISHVILI G. Chromosomal "stress-response" domains govern the spatiotemporal expression of the bacterial virulence program. *mBio*. 2015 28; 6:e00353-00315. DOI : 10.1128/mBio.00353-15
35. PRAY L. Discovery of DNA Double Helix : Watson and Crick. *Nature Education*. 2008
36. DAHM R. Friedrich Miescher and the discovery of DNA. *Developmental Biology*. 2005 15; 278:274-88. DOI : 10.1016/j.ydbio.2004.11.028
37. KOSSEL (CHIMISTE).) A. Untersuchungen über die Nucleine und ihre Spaltungsprodukte. Karl J. Trübner, 1881. book
38. LEVENE P. THE STRUCTURE OF YEAST NUCLEIC ACID IV. AMMONIA HYDROLYSIS. *Journal of Biological Chemistry*. 1919 1
39. ROCHA E. The Organization of the Bacterial Genome. *Annual review of genetics*. 2008 1; 42:211-33. DOI : 10.1146/annurev.genet.42.110807.091653
40. ROCHA EPC et DANCHIN A. Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Research*. 2003 15; 31:6570-7. DOI : 10.1093/nar/gkg859
41. BREWER BJ. When polymerases collide : replication and the transcriptional organization of the *E. coli* chromosome. *Cell*. 1988 3; 53:679-86. DOI : 10.1016/0092-8674(88)90086-4
42. LOBRY JR. Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular Biology and Evolution*. 1996; 13:660-5. DOI : 10.1093/oxfordjournals.molbev.a025626
43. ZHENG Y, SZUSTAKOWSKI JD, FORTNOW L, ROBERTS RJ et KASIF S. Computational identification of operons in microbial genomes. *Genome Research*. 2002; 12:1221-30. DOI : 10.1101/gr.200602
44. LAND M, HAUSER L, JUN SR, NOOKAEW I, LEUZE MR, AHN TH, KARPINETS T, LUND O, KORA G, WASSENAAR T, POUDEL S et USSERY DW. Insights from 20 years of bacterial

- genome sequencing. *Functional & Integrative Genomics*. 2015; 15:141-61. DOI : 10.1007/s10142-015-0433-4
45. BENNETT GM et MORAN NA. Small, smaller, smallest : the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. *Genome Biology and Evolution*. 2013; 5:1675-88. DOI : 10.1093/gbe/evt118
46. FRASER CM, GOCAYNE JD, WHITE O, ADAMS MD, CLAYTON RA, FLEISCHMANN RD, BULT CJ, KERLAVAGE AR, SUTTON G, KELLEY JM, FRITCHMAN RD, WEIDMAN JF, SMALL KV, SANDUSKY M, FUHRMANN J, NGUYEN D, UTTERBACK TR, SAUDEK DM, PHILLIPS CA, MERRICK JM, TOMB JF, DOUGHERTY BA, BOTT KF, HU PC, LUCIER TS, PETERSON SN, SMITH HO, HUTCHISON CA et VENTER JC. The minimal gene complement of *Mycoplasma genitalium*. *Science (New York, NY)*. 1995 20; 270:397-403. DOI : 10.1126/science.270.5235.397
47. TOUCHON M, HOEDE C, TENAILLON O, BARBE V, BAERISWYL S, BIDET P, BINGEN E, BONACORSI S, BOUCHIER C, BOUVET O, CALTEAU A, CHIAPELLO H, CLERMONT O, CRUVEILLER S, DANCHIN A, DIARD M, DOSSAT C, KAROU ME, FRAPY E, GARRY L, GHIGO JM, GILLES AM, JOHNSON J, LE BOUGUÉNEC C, LESCAT M, MANGENOT S, MARTINEZ-JÉHANNE V, MATIC I, NASSIF X, OZTAS S, PETIT MA, PICHON C, ROUY Z, RUF CS, SCHNEIDER D, TOURET J, VACHERIE B, VALLENET D, MÉDIGUE C, ROCHA EPC et DENAMUR E. Organised Genome Dynamics in the *Escherichia coli* Species Results in Highly Diverse Adaptive Paths. *PLoS Genetics*. 2009 23; 5:e1000344. DOI : 10.1371/journal.pgen.1000344
48. HAN K, LI Zf, PENG R, ZHU Lp, ZHOU T, WANG Lg, LI Sg, ZHANG Xb, HU W, WU Zh, QIN N et LI Yz. Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Scientific Reports*. 2013 1; 3:2101. DOI : 10.1038/srep02101
49. SPAHN C, ENDESFELDER U et HEILEMANN M. Super-resolution imaging of *Escherichia coli* nucleoids reveals highly structured and asymmetric segregation during fast growth. *Journal of Structural Biology*. 2014 1; 185:243-9. DOI : 10.1016/j.jsb.2014.01.007
50. JOYEUX M. Compaction of bacterial genomic DNA : clarifying the concepts. *Journal of Physics : Condensed Matter*. 2015; 27:383001. DOI : 10.1088/0953-8984/27/38/383001
51. MACVANIN M, EDGAR R, CUI F, TROSTEL A, ZHURKIN V et ADHYA S. Noncoding RNAs Binding to the Nucleoid Protein HU in *Escherichia coli*. *Journal of Bacteriology*. 2012 26; 194:6046-55. DOI : 10.1128/jb.00961-12
52. PALADE GE et JAMIESON J. CIL :41048, *Diplococcus pneumoniae*. CIL. Dataset. 2012
53. NIKI H, YAMAICHI Y et HIRAGA S. Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes & Development*. 2000 15; 14:212-23
54. VALENS M, PENAUD S, ROSSIGNOL M, CORNET F et BOCCARD F. Macrodomain organization of the *Escherichia coli* chromosome. *The EMBO Journal*. 2004 27; 23:4330-41. DOI : 10.1038/sj.emboj.7600434
55. ESPELI O, MERCIER R et BOCCARD F. DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Molecular Microbiology*. 2008; 68:1418-27. DOI : 10.1111/j.1365-2958.2008.06239.x

56. BERKUM N, LIEBERMAN-AIDEN E, WILLIAMS L, IMAKAEV M, GNIRKE A, MIRNY L, DEKKER J et LANDER E. Hi-C : A Method to Study the Three-dimensional Architecture of Genomes. *Journal of visualized experiments : JoVE*. 2010 6; 39. DOI : 10.3791/1869
57. LIOY VS, COURNAC A, MARBOUTY M, DUIGOU S, MOZZICONACCI J, ESPÉLI O, BOCCARD F et KOSZUL R. Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell*. 2018 8; 172:771-783.e18. DOI : 10.1016/j.cell.2017.12.027
58. BIRNIE A et DEKKER C. Genome-in-a-Box : Building a Chromosome from the Bottom Up. *ACS Nano*. 2021 26; 15:111-24. DOI : 10.1021/acsnano.0c07397
59. LE TBK, IMAKAEV MV, MIRNY LA et LAUB MT. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science (New York, NY)*. 2013 8; 342:731-4. DOI : 10.1126/science.1242059
60. MARBOUTY M, LE GALL A, CATTONI DI, COURNAC A, KOH A, FICHE JB, MOZZICONACCI J, MURRAY H, KOSZUL R et NOLLMANN M. Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Molecular Cell*. 2015 20; 59:588-602. DOI : 10.1016/j.molcel.2015.07.020
61. TRUSSART M, YUS E, MARTINEZ S, BAÙ D, TAHARA YO, PENGOT T, WIDJAJA M, KRETSCHMER S, SWOGER J, DJORDJEVIC S, TURNBULL L, WHITCHURCH C, MIYATA M, MARTI-RENOM MA, LLUCH-SENAR M et SERRANO L. Defined chromosome structure in the genome-reduced bacterium *Mycoplasma pneumoniae*. *Nature Communications*. 2017 8; 8:14665. DOI : 10.1038/ncomms14665
62. BLOT N, MAVATHUR R, GEERTZ M, TRAVERS A et MUSKHELISHVILI G. Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO reports*. 2006; 7:710-5. DOI : 10.1038/sj.embor.7400729
63. DAME RT et TARK-DAME M. Bacterial chromatin : converging views at different scales. *Current Opinion in Cell Biology. Cell nucleus* 2016 1; 40:60-5. DOI : 10.1016/j.ceb.2016.02.015
64. STEENSEL B van et FURLONG EEM. The role of transcription in shaping the spatial organization of the genome. *Nature Reviews Molecular Cell Biology*. 2019; 20:327-37. DOI : 10.1038/s41580-019-0114-6
65. SEXTON T, YAFFE E, KENIGSBERG E, BANTIGNIES F, LEBLANC B, HOICHMAN M, PARRINELLO H, TANAY A et CAVALLI G. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*. 2012 3; 148:458-72. DOI : 10.1016/j.cell.2012.01.010
66. LUGER K, MÄDER AW, RICHMOND RK, SARGENT DF et RICHMOND TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 1997; 389:251-60. DOI : 10.1038/38444
67. DORMAN CJ. Function of nucleoid-associated proteins in chromosome structuring and transcriptional regulation. *Journal of Molecular Microbiology and Biotechnology*. 2014; 24:316-31. DOI : 10.1159/000368850
68. VERMA SC, QIAN Z et ADHYA SL. Architecture of the *Escherichia coli* nucleoid. *PLoS genetics*. 2019; 15:e1008456. DOI : 10.1371/journal.pgen.1008456

69. NOORT J van, VERBRUGGE S, GOOSEN N, DEKKER C et DAME RT. Dual architectural roles of HU : Formation of flexible hinges and rigid filaments. *Proceedings of the National Academy of Sciences*. 2004 4; 101:6969-74. DOI : 10.1073/pnas.0308230101
70. LIU Y, CHEN H, KENNEY LJ et YAN J. A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes & Development*. 2010 15; 24:339-44. DOI : 10.1101/gad.1883510
71. KROGH T, MØLLER-JENSEN J et KALETA C. Impact of Chromosomal Architecture on the Function and Evolution of Bacterial Genomes. *Frontiers in Microbiology*. 2018 27; 9. DOI : 10.3389/fmicb.2018.02019
72. OUAFA ZA, REVERCHON S, LAUTIER T, MUSKHELISHVILI G et 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*. *Nucleic Acids Research*. 2012; 40:4306-19. DOI : 10.1093/nar/gks014
73. ZGHIDI-ABOUZID O, HÉRAULT E, RIMSKY S, REVERCHON S, NASSER W et BUCKLE M. Regulation of pel genes, major virulence factors in the plant pathogen bacterium *Dickeya dadantii*, is mediated by cooperative binding of the nucleoid-associated protein H-NS. *Research in Microbiology*. 2016; 167:247-53. DOI : 10.1016/j.resmic.2016.02.001
74. LAUTIER T et NASSER W. The DNA nucleoid-associated protein Fis co-ordinates the expression of the main virulence genes in the phytopathogenic bacterium *Erwinia chrysanthemi*. *Molecular Microbiology*. 2007; 66:1474-90. DOI : 10.1111/j.1365-2958.2007.06012.x
75. REVERCHON S, MEYER S, FORQUET R, HOMMAIS F, MUSKHELISHVILI G et NASSER W. The nucleoid-associated protein IHF acts as a 'transcriptional domainin' protein coordinating the bacterial virulence traits with global transcription. *Nucleic Acids Research*. 2021 25; 49:776-90. DOI : 10.1093/nar/gkaa1227
76. HASSLER M, SHALTIEL IA et HAERING CH. Towards a unified Model of SMC Complex Function. *Current biology : CB*. 2018 5; 28:R1266-R1281. DOI : 10.1016/j.cub.2018.08.034
77. GRUBER S. MukBEF on the march : taking over chromosome organization in bacteria? *Molecular Microbiology*. 2011; 81:855-9. DOI : 10.1111/j.1365-2958.2011.07764.x
78. UHLMANN F. SMC complexes : from DNA to chromosomes. *Nature Reviews Molecular Cell Biology*. 2016; 17:399-412. DOI : 10.1038/nrm.2016.30
79. SOPPA J. Prokaryotic structural maintenance of chromosomes (SMC) proteins : distribution, phylogeny, and comparison with MukBs and additional prokaryotic and eukaryotic coiled-coil proteins. *Gene*. 2001 31; 278:253-64. DOI : 10.1016/S0378-1119(01)00733-8
80. GRUBER S et ERRINGTON J. Recruitment of condensin to replication origin regions by ParB/SpoOJ promotes chromosome segregation in *B. subtilis*. *Cell*. 2009 15; 137:685-96. DOI : 10.1016/j.cell.2009.02.035
81. DIEBOLD-DURAND ML, LEE H, RUIZ AVILA LB, NOH H, SHIN HC, IM H, BOCK FP, BÜRMAN F, DURAND A, BASFELD A, HAM S, BASQUIN J, OH BH et GRUBER S. Structure of Full-Length SMC and Rearrangements Required for Chromosome Organization. *Molecular Cell*. 2017 20; 67:334-347.e5. DOI : 10.1016/j.molcel.2017.06.010

82. PRUSS GJ et DRLICA K. DNA supercoiling and prokaryotic transcription. *Cell*. 1989 24; 56:521-3. DOI : 10.1016/0092-8674(89)90574-6
83. MARTIS B. S, FORQUET R, REVERCHON S, NASSER W et MEYER S. DNA Supercoiling : an Ancestral Regulator of Gene Expression in Pathogenic Bacteria? *Computational and Structural Biotechnology Journal*. 2019 1; 17:1047-55. DOI : 10.1016/j.csbj.2019.07.013
84. BOYLE J. Lehninger principles of biochemistry (4th ed.) : Nelson, D., and Cox, M. *Biochemistry and Molecular Biology Education*. 2005; 33:74-5. DOI : 10.1002/bmb.2005.494033010419
85. MIRKIN S. DNA topology : Fundamentals. *Encyclopedia Life Sci*. 2002 1. DOI : 10.1038/npg.els.0001038
86. HIGGINS NP et VOLOGODSKII AV. Topological Behavior of Plasmid DNA. *Microbiology Spectrum*. 2015 20; 3. DOI : 10.1128/microbiolspec.PLAS-0036-2014
87. DORMAN CJ. Dna Supercoiling and Bacterial Gene Expression. *Science Progress*. 2006 1; 89:151-66. DOI : 10.3184/003685006783238317
88. RUFF EF, RECORD MT et ARTSIMOVITCH I. Initial events in bacterial transcription initiation. *Biomolecules*. 2015 27; 5:1035-62. DOI : 10.3390/biom5021035
89. BREMER H et DENNIS P. Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. *E Coli Salmonella Cell Mol Biol*. 1996 1; 2
90. DECKER KB et HINTON DM. Transcription regulation at the core : similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annual Review of Microbiology*. 2013; 67:113-39. DOI : 10.1146/annurev-micro-092412-155756
91. MITTENHUBER G. An inventory of genes encoding RNA polymerase sigma factors in 31 completely sequenced eubacterial genomes. *Journal of Molecular Microbiology and Biotechnology*. 2002; 4:77-91
92. BERVOETS I et CHARLIER D. Diversity, versatility and complexity of bacterial gene regulation mechanisms : opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiology Reviews*. 2019 4; 43:304-39. DOI : 10.1093/femsre/fuz001
93. MITCHELL JE, ZHENG D, BUSBY SJW et MINCHIN SD. Identification and analysis of 'extended -10' promoters in *Escherichia coli*. *Nucleic Acids Research*. 2003 15; 31:4689-95. DOI : 10.1093/nar/gkg694
94. HAUGEN SP, BERKMEN MB, ROSS W, GAAL T, WARD C et GOURSE RL. rRNA promoter regulation by nonoptimal binding of sigma region 1.2 : an additional recognition element for RNA polymerase. *Cell*. 2006 16; 125:1069-82. DOI : 10.1016/j.cell.2006.04.034
95. RAO L, ROSS W, APPLEMAN JA, GAAL T, LEIRMO S, SCHLAX PJ, RECORD MT et GOURSE RL. Factor independent activation of *rrnB* P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. *Journal of Molecular Biology*. 1994 4; 235:1421-35. DOI : 10.1006/jmbi.1994.1098
96. HU Y et LIU B. Roles of zinc-binding domain of bacterial RNA polymerase in transcription. *Trends in Biochemical Sciences*. 2022 30; 47. DOI : 10.1016/j.tibs.2022.03.007
97. LEE J et BORUKHOV S. Bacterial RNA Polymerase-DNA Interaction—The Driving Force of Gene Expression and the Target for Drug Action. *Frontiers in Molecular Biosciences*. 2016; 3. DOI : 10.3389/fmolb.2016.00073

98. FORQUET R, PINEAU M, NASSER W, REVERCHON S et MEYER S. Role of the Discriminator Sequence in the Supercoiling Sensitivity of Bacterial Promoters. *mSystems*. 2021 31; 6:e0097821. DOI : 10.1128/mSystems.00978-21
99. RAY-SONI A, BELLECOURT MJ et LANDICK R. Mechanisms of Bacterial Transcription Termination : All Good Things Must End. *Annual Review of Biochemistry*. 2016 2; 85:319-47. DOI : 10.1146/annurev-biochem-060815-014844
100. HAO Z, SVETLOV V et NUDLER E. Rho-dependent transcription termination : a revisionist view. *Transcription*. 12:171-81. DOI : 10.1080/21541264.2021.1991773
101. MADAN BABU M et TEICHMANN SA. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Research*. 2003 15; 31:1234-44
102. MARTÍNEZ-ANTONIO A et COLLADO-VIDES J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology*. 2003; 6:482-9. DOI : 10.1016/j.mib.2003.09.002
103. BROWNING DF et BUSBY SJW. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*. 2004; 2:57-65. DOI : 10.1038/nrmicro787
104. EBRIGHT RH. Transcription activation at Class I CAP-dependent promoters. *Molecular Microbiology*. 1993; 8:797-802. DOI : 10.1111/j.1365-2958.1993.tb01626.x
105. BROWN NL, STOYANOV JV, KIDD SP et HOBMAN JL. The MerR family of transcriptional regulators. *FEMS microbiology reviews*. 2003; 27:145-63. DOI : 10.1016/S0168-6445(03)00051-2
106. MÜLLER-HILL B. *The Lac Operon : A Short History of a Genetic Paradigm*. Walter de Gruyter, 1996. 224 p.
107. SHIN M, KANG S, HYUN SJ, FUJITA N, ISHIHAMA A, VALENTIN-HANSEN P et CHOY HE. Repression of deoP2 in *Escherichia coli* by CytR : conversion of a transcription activator into a repressor. *The EMBO Journal*. 2001 1; 20:5392-9. DOI : 10.1093/emboj/20.19.5392
108. DAME RT, WYMAN C, WURM R, WAGNER R et GOOSEN N. Structural Basis for H-NS-mediated Trapping of RNA Polymerase in the Open Initiation Complex at the *rrnB* P1 \*. *Journal of Biological Chemistry*. 2002 18; 277:2146-50. DOI : 10.1074/jbc.C100603200
109. KLUMPP S et HWA T. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proceedings of the National Academy of Sciences*. 2008 23; 105:20245-50. DOI : 10.1073/pnas.0804953105
110. BACUN-DRUZINA V, BUTORAC A, MRVČIĆ J, DRAGICEVIC T et STEHLIK-TOMAS V. Bacterial Stationary-Phase Evolution. *Food Technology and Biotechnology*. 2011 1; 49
111. PEASE AC, SOLAS D, SULLIVAN EJ, CRONIN MT, HOLMES CP et FODOR SP. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proceedings of the National Academy of Sciences of the United States of America*. 1994 24; 91:5022-6. DOI : 10.1073/pnas.91.11.5022
112. SCHENA M, SHALON D, DAVIS RW et BROWN PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science (New York, NY)*. 1995 20; 270:467-70. DOI : 10.1126/science.270.5235.467

113. SCHENA M, SHALON D, HELLER R, CHAI A, BROWN PO et DAVIS RW. Parallel human genome analysis : microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences of the United States of America*. 1996 1; 93:10614-9. DOI : 10.1073/pnas.93.20.10614
114. Z W, M G et M S. RNA-Seq : a revolutionary tool for transcriptomics. *Nature reviews Genetics*. 2009; 10. DOI : 10.1038/nrg2484
115. MILLER M et TANG YW. Basic Concepts of Microarrays and Potential Applications in Clinical Microbiology. *Clinical microbiology reviews*. 2009 1; 22:611-33. DOI : 10.1128/CMR.00019-09
116. LANGMEAD B et SALZBERG SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012; 9:357-9. DOI : 10.1038/nmeth.1923
117. ANDERS S, PYL PT et HUBER W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*. 2015 15; 31:166-9. DOI : 10.1093/bioinformatics/btu638
118. LOVE MI, HUBER W et ANDERS S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014 5; 15:550. DOI : 10.1186/s13059-014-0550-8
119. GENE ONTOLOGY CONSORTIUM. The Gene Ontology resource : enriching a Gold mine. *Nucleic Acids Research*. 2021 8; 49(D1):D325-D334. DOI : 10.1093/nar/gkaa1113
120. WANG JC. DNA topoisomerases. *Annual Review of Biochemistry*. 1996; 65:635-92. DOI : 10.1146/annurev.bi.65.070196.003223
121. BUSH NG, EVANS-ROBERTS K et MAXWELL A. DNA Topoisomerases. *EcoSal Plus*. 2015 17; 6. DOI : 10.1128/ecosalplus.ESP-0010-2014
122. ZECHIEDRICH EL, KHODURSKY AB, BACHELLIER S, SCHNEIDER R, CHEN D, LILLEY DMJ et COZZARELLI NR. Roles of Topoisomerases in Maintaining Steady-state DNA Supercoiling in *Escherichia coli* \*. *Journal of Biological Chemistry*. 2000 17; 275:8103-13. DOI : 10.1074/jbc.275.11.8103
123. NAGARAJA V, GODBOLE AA, HENDERSON SR et MAXWELL A. DNA topoisomerase I and DNA gyrase as targets for TB therapy. *Drug Discovery Today. SI :Tuberculosis* 2017 1; 22:510-8. DOI : 10.1016/j.drudis.2016.11.006
124. GONZÁLEZ DE LA CAMPA A, GARCÍA ESTEBAN MT et BLÁZQUEZ FERRER MA. Use of Seconeolitsine and N-Methyl-Seconeolitsine for the Manufacture of Medicaments. *Brev. CARLOS III UNIVERSITAT DE VALENCIA CIBER DE ENFERMEDADES RESPIRATORIAS CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS I de salud*. 2011 23
125. LI Z, MONDRAGÓN A, HIASA H, MARIANS KJ et DIGATE RJ. Identification of a unique domain essential for *Escherichia coli* DNA topoisomerase III-catalysed decatenation of replication intermediates. *Molecular Microbiology*. 2000; 35:888-95. DOI : 10.1046/j.1365-2958.2000.01763.x
126. TANEJA B, SCHNURR B, SLESAREV A, MARKO JF et MONDRAGÓN A. Topoisomerase V relaxes supercoiled DNA by a constrained swiveling mechanism. *Proceedings of the National Academy of Sciences of the United States of America*. 2007 11; 104:14670-5. DOI : 10.1073/pnas.0701989104



127. KIKUCHI A et ASAI K. Reverse gyrase—a topoisomerase which introduces positive superhelical turns into DNA. *Nature*. 1984 21 ; 309:677-81. DOI : 10.1038/309677a0
128. SOCZEK KM, GRANT T, ROSENTHAL PB et MONDRAGÓN A. CryoEM structures of open dimers of gyrase A in complex with DNA illuminate mechanism of strand passage. *eLife*. 7:e41215. DOI : 10.7554/eLife.41215
129. LEVINE C, HIASA H et MARIANS KJ. DNA gyrase and topoisomerase IV : biochemical activities, physiological roles during chromosome replication, and drug sensitivities. *Biochimica Et Biophysica Acta*. 1998 1 ; 1400:29-43. DOI : 10.1016/s0167-4781(98)00126-2
130. PENG H et MARIANS KJ. Escherichia coli topoisomerase IV. Purification, characterization, subunit structure, and subunit interactions. *The Journal of Biological Chemistry*. 1993 15 ; 268:24481-90
131. ZECHIEDRICH EL et COZZARELLI NR. Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in Escherichia coli. *Genes & Development*. 1995 15 ; 9:2859-69. DOI : 10.1101/gad.9.22.2859
132. BAUER DLV, MARIE R, RASMUSSEN KH, KRISTENSEN A et MIR KU. DNA catenation maintains structure of human metaphase chromosomes. *Nucleic Acids Research*. 2012 ; 40:11428-34. DOI : 10.1093/nar/gks931
133. HSIEH T et BRUTLAG D. ATP-dependent DNA topoisomerase from *D. melanogaster* reversibly catenates duplex DNA rings. *Cell*. 1980 ; 21:115-25. DOI : 10.1016/0092-8674(80)90119-1
134. DINARDO S, VOELKEL K et STERNGLANZ R. DNA topoisomerase II mutant of *Saccharomyces cerevisiae* : topoisomerase II is required for segregation of daughter molecules at the termination of DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*. 1984 ; 81:2616-20. DOI : 10.1073/pnas.81.9.2616
135. CAPRANICO G, TINELLI S, AUSTIN CA, FISHER ML et ZUNINO F. Different patterns of gene expression of topoisomerase II isoforms in differentiated tissues during murine development. *Biochimica Et Biophysica Acta*. 1992 17 ; 1132:43-8. DOI : 10.1016/0167-4781(92)90050-a
136. TURLEY H, COMLEY M, HOULBROOK S, NOZAKI N, KIKUCHI A, HICKSON ID, GATTER K et HARRIS AL. The distribution and expression of the two isoforms of DNA topoisomerase II in normal and neoplastic human tissues. *British Journal of Cancer*. 1997 ; 75:1340-6. DOI : 10.1038/bjc.1997.227
137. BERGERAT A, GADELLE D et FORTERRE P. Purification of a DNA topoisomerase II from the hyperthermophilic archaeon *Sulfolobus shibatae*. A thermostable enzyme with both bacterial and eucaryal features. *The Journal of Biological Chemistry*. 1994 4 ; 269:27663-9
138. HARTUNG F et PUCHTA H. Molecular characterisation of two paralogous SPO11 homologues in *Arabidopsis thaliana*. *Nucleic Acids Research*. 2000 1 ; 28:1548-54. DOI : 10.1093/nar/28.7.1548
139. ARAVIND L, IYER LM, WELLEMS TE et MILLER LH. Plasmodium biology : genomic gleanings. *Cell*. 2003 26 ; 115:771-85. DOI : 10.1016/s0092-8674(03)01023-7
140. CORLESS S et GILBERT N. Investigating DNA supercoiling in eukaryotic genomes. *Briefings in functional genomics*. 2017 1 ; 16:379-89. DOI : 10.1093/bfpgp/elix007

141. DRLICA K et WORCEL A. Conformational transitions in the Escherichia coli chromosome : Analysis by viscometry and sedimentation. *Journal of Molecular Biology*. 1975 25 ; 98:393-411. DOI : 10.1016/S0022-2836(75)80126-4
142. DRLICA K et SNYDER M. Superhelical Escherichia coli DNA : Relaxation by coumermycin. 1978
143. KELLER W. Determination of the number of superhelical turns in simian virus 40 DNA by gel electrophoresis. *Proceedings of the National Academy of Sciences*. 1975 ; 72:4876-80. DOI : 10.1073/pnas.72.12.4876
144. SHURE M, PULLEYBLANK DE et VINOGRAD J. The problems of eukaryotic and prokaryotic DNA packaging and in vivo conformation posed by superhelix density heterogeneity. *Nucleic Acids Research*. 1977 ; 4:1183-205. DOI : 10.1093/nar/4.5.1183
145. MIRKIN SM. DNA Topology : Fundamentals. *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd, 2001. DOI : 10.1038/npg.e1s.0001038
146. GIBSON EG, OVIATT AA et OSHEROFF N. Two-Dimensional Gel Electrophoresis to Resolve DNA Topoisomers. *DNA Electrophoresis : Methods and Protocols*. Methods in Molecular Biology. New York, NY : Springer US, 2020 :15-24. DOI : 10.1007/978-1-0716-0323-9\_2
147. DUPREY A et GROISMAN EA. FEDS : a Novel Fluorescence-Based High-Throughput Method for Measuring DNA Supercoiling In Vivo. *mBio*. 2020 28 ; 11:10.1128/mbio.01053-20. DOI : 10.1128/mbio.01053-20
148. KRISPIN O et ALLMANSBERGER R. Changes in DNA supertwist as a response of Bacillus subtilis towards different kinds of stress. *FEMS microbiology letters*. 1995 15 ; 134:129-35. DOI : 10.1111/j.1574-6968.1995.tb07926.x
149. HIGGINS CF, DORMAN CJ, STIRLING DA, WADDELL L, BOOTH IR, MAY G et BREMER E. A physiological role for DNA supercoiling in the osmotic regulation of gene expression in S. typhimurium and E. coli. *Cell*. 1988 26 ; 52:569-84. DOI : 10.1016/0092-8674(88)90470-9
150. HSIEH LS, ROUVIERE-YANIV J et DRLICA K. Bacterial DNA supercoiling and [ATP]/[ADP] ratio : changes associated with salt shock. *Journal of Bacteriology*. 1991 ; 173:3914-7. DOI : 10.1128/jb.173.12.3914-3917.1991
151. OGATA Y, MIZUSHIMA T, KATAOKA K, MIKI T et SEKIMIZU K. Identification of DNA topoisomerases involved in immediate and transient DNA relaxation induced by heat shock in Escherichia coli. *Molecular & general genetics : MGG*. 1994 1 ; 244:451-5. DOI : 10.1007/bf00583895
152. MIZUSHIMA T, KATAOKA K, OGATA Y, INOUE R et SEKIMIZU K. Increase in negative supercoiling of plasmid DNA in Escherichia coli exposed to cold shock. *Molecular Microbiology*. 1997 ; 23:381-6. DOI : 10.1046/j.1365-2958.1997.2181582.x
153. QUINN HJ, CAMERON ADS et 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 genetics*. 2014 ; 10:e1004215. DOI : 10.1371/journal.pgen.1004215

154. WEINSTEIN-FISCHER D, ELGRABLY-WEISS M et ALTUVIA S. Escherichia coli response to hydrogen peroxide : a role for DNA supercoiling, topoisomerase I and Fis. *Molecular Microbiology*. 2000 ; 35:1413-20. DOI : 10.1046/j.1365-2958.2000.01805.x
155. KESELER IM, MACKIE A, SANTOS-ZAVALA A, BILLINGTON R, BONAVIDES-MARTÍNEZ C, CASPI R, FULCHER C, GAMA-CASTRO S, KOTHARI A, KRUMMENACKER M, LATENDRESSE M, MUÑIZ-RASCADO L, ONG Q, PALEY S, PERALTA-GIL M, SUBHRAVETI P, VELÁZQUEZ-RAMÍREZ DA, WEAVER D, COLLADO-VIDES J, PAULSEN I et KARP PD. The EcoCyc database : reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Research*. 2017 4; 45(D1):D543-D550. DOI : 10.1093/nar/gkw1003
156. BRINZA L, CALEVRO F et CHARLES H. Genomic analysis of the regulatory elements and links with intrinsic DNA structural properties in the shrunken genome of Buchnera. *BMC genomics*. 2013 1; 14:73. DOI : 10.1186/1471-2164-14-73
157. JUNIER I, UNAL EB, YUS E, LLORÉNS-RICO V et SERRANO L. Insights into the Mechanisms of Basal Coordination of Transcription Using a Genome-Reduced Bacterium. *Cell Systems*. 2016; 2:391-401. DOI : 10.1016/j.cels.2016.04.015
158. LIM HM, LEWIS DEA, LEE HJ, LIU M et ADHYA S. Effect of varying the supercoiling of DNA on transcription and its regulation. *Biochemistry*. 2003 16; 42:10718-25. DOI : 10.1021/bi030110t
159. BOROWIEC JA et GRALLA JD. All three elements of the lac ps promoter mediate its transcriptional response to DNA supercoiling. *Journal of Molecular Biology*. 1987 5; 195:89-97. DOI : 10.1016/0022-2836(87)90329-9
160. AUNER H, BUCKLE M, DEUFEL A, KUTATELADZE T, LAZARUS L, MAVATHUR R, MUSKHELISHVILI G, PEMBERTON I, SCHNEIDER R et TRAVERS A. Mechanism of transcriptional activation by FIS : role of core promoter structure and DNA topology. *Journal of Molecular Biology*. 2003 8; 331:331-44. DOI : 10.1016/s0022-2836(03)00727-7
161. DRLICA K. Control of bacterial DNA supercoiling. *Molecular Microbiology*. 1992; 6:425-33. DOI : 10.1111/j.1365-2958.1992.tb01486.x
162. PRUSS GJ, FRANCO RJ, CHEVALIER SG, MANES SH et DRLICA K. Effects of DNA gyrase inhibitors in Escherichia coli topoisomerase I mutants. *Journal of Bacteriology*. 1986; 168:276-82
163. PETER BJ, ARSUAGA J, BREIER AM, KHODURSKY AB, BROWN PO et COZZARELLI NR. Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. *Genome Biology*. 2004 1; 5:R87. DOI : 10.1186/gb-2004-5-11-r87
164. JEONG KS, XIE Y, HIASA H et KHODURSKY AB. Analysis of Pleiotropic Transcriptional Profiles : A Case Study of DNA Gyrase Inhibition. *PLOS Genetics*. 2006 29; 2:e152. DOI : 10.1371/journal.pgen.0020152
165. CHEUNG KJ, BADARINARAYANA V, SELINGER DW, JANSE D et CHURCH GM. A Microarray-Based Antibiotic Screen Identifies a Regulatory Role for Supercoiling in the Osmotic Stress Response of Escherichia coli. *Genome Research*. 2003 1; 13:206-15. DOI : 10.1101/gr.401003

166. EL HOUDAIGUI B, FORQUET R, HINDRÉ T, SCHNEIDER D, NASSER W, REVERCHON S et MEYER S. Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Research*. 2019 20; 47:5648-57. DOI : 10.1093/nar/gkz300
167. FERRÁNDIZ MJ, MARTÍN-GALIANO AJ, SCHVARTZMAN JB et CAMPA AG de la. The genome of *Streptococcus pneumoniae* is organized in topology-reacting gene clusters. *Nucleic Acids Research*. 2010; 38:3570-81. DOI : 10.1093/nar/gkq106
168. SIOUD M, BOUDABOUS A et CEKAITE L. Transcriptional responses of *Bacillus subtilis* and *thuringiensis* to antibiotics and anti-tumour drugs. *International Journal of Molecular Medicine*. 2009; 23:33-9
169. SCHRÖDER W, BERNHARDT J, MARINCOLA G, KLEIN-HITPASS L, HERBIG A, KRUPP G, NIESELT K et WOLZ C. Altering gene expression by aminocoumarins : the role of DNA supercoiling in *Staphylococcus aureus*. *BMC Genomics*. 2014 16; 15:291. DOI : 10.1186/1471-2164-15-291
170. GORMLEY NA, ORPHANIDES G, MEYER A, CULLIS PM et MAXWELL A. The interaction of coumarin antibiotics with fragments of DNA gyrase B protein. *Biochemistry*. 1996 16; 35:5083-92. DOI : 10.1021/bi952888n
171. GELLERT M, O'DEA MH, ITOH T et TOMIZAWA J. Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase. *Proceedings of the National Academy of Sciences of the United States of America*. 1976; 73:4474-8. DOI : 10.1073/pnas.73.12.4474
172. HARDY CD et COZZARELLI NR. Alteration of *Escherichia coli* Topoisomerase IV to Novobiocin Resistance. *Antimicrobial Agents and Chemotherapy*. 2003 1; 47:941-7. DOI : 10.1128/AAC.47.3.941-947.2003
173. FORTERRE P et GADELLE D. Phylogenomics of DNA topoisomerases : their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Research*. 2009; 37:679-92. DOI : 10.1093/nar/gkp032
174. HIGGINS PG, FLUIT AC et SCHMITZ FJ. Fluoroquinolones : structure and target sites. *Current Drug Targets*. 2003; 4:181-90. DOI : 10.2174/1389450033346920
175. FERRÁNDIZ MJ, MARTÍN-GALIANO AJ, ARNAZ C, CAMACHO-SOQUERO I, TIRADO-VÉLEZ JM et CAMPA AG de la. 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 Research*. 2016 6; 44:7292-303. DOI : 10.1093/nar/gkw602
176. WANG JY et SYVANEN M. DNA twist as a transcriptional sensor for environmental changes. *Molecular Microbiology*. 1992; 6:1861-6. DOI : 10.1111/j.1365-2958.1992.tb01358.x
177. FORQUET R, NASSER W, REVERCHON S et MEYER S. Quantitative contribution of the spacer length in the supercoiling-sensitivity of bacterial promoters. *Nucleic Acids Research*. 2022 22; 50:7287-97. DOI : 10.1093/nar/gkac579
178. PEMBERTON IK, MUSKHELISHVILI G, TRAVERS AA et BUCKLE M. The G+C-rich discriminator region of the *tyrT* promoter antagonises the formation of stable preinitiation complexes. *Journal of Molecular Biology*. 2000 16; 299:859-64. DOI : 10.1006/jmbi.2000.3780
179. BORUKHOV S et NUDLER E. RNA polymerase : the vehicle of transcription. *Trends in Microbiology*. 2008; 16:126-34. DOI : 10.1016/j.tim.2007.12.006

180. MA J, BAI L et WANG MD. Transcription under torsion. *Science (New York, NY)*. 2013 28; 340:1580-3. DOI : 10.1126/science.1235441
181. LIU LF et WANG JC. Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 1987; 84:7024-7. DOI : 10.1073/pnas.84.20.7024
182. GUO MS, KAWAMURA R, LITTLEHALE ML, MARKO JF et LAUB MT. High-resolution, genome-wide mapping of positive supercoiling in chromosomes. *eLife*. 2021 19; 10:e67236. DOI : 10.7554/eLife.67236
183. VISSER BJ, SHARMA S, CHEN PJ, MCMULLIN AB, BATES ML et BATES D. Psoralen mapping reveals a bacterial genome supercoiling landscape dominated by transcription. *Nucleic Acids Research*. 2022 6; 50:4436-49. DOI : 10.1093/nar/gkac244
184. WU HY, SHYY SH, WANG JC et LIU LF. Transcription generates positively and negatively supercoiled domains in the template. *Cell*. 1988 6; 53:433-40. DOI : 10.1016/0092-8674(88)90163-8
185. MA J et WANG MD. RNA polymerase is a powerful torsional motor. *Cell Cycle (Georgetown, Tex)*. 2014; 13:337-8. DOI : 10.4161/cc.27508
186. SHEN BA et LANDICK R. Transcription of Bacterial Chromatin. *Journal of Molecular Biology. RNA polymerase reaches 60* 2019 20; 431:4040-66. DOI : 10.1016/j.jmb.2019.05.041
187. MEYER S et BESLON G. Torsion-Mediated Interaction between Adjacent Genes. *PLoS Computational Biology*. 2014 4; 10. DOI : 10.1371/journal.pcbi.1003785
188. DORMAN CJ. DNA supercoiling and transcription in bacteria : a two-way street. *BMC Molecular and Cell Biology*. 2019 18; 20. DOI : 10.1186/s12860-019-0211-6
189. CHENG B, ZHU CX, JI C, AHUMADA A et TSE-DINH YC. Direct interaction between *Escherichia coli* RNA polymerase and the zinc ribbon domains of DNA topoisomerase I. *The Journal of Biological Chemistry*. 2003 15; 278:30705-10. DOI : 10.1074/jbc.M303403200
190. FERRÁNDIZ MJ, HERNÁNDEZ P et CAMPA AGdl. Genome-wide proximity between RNA polymerase and DNA topoisomerase I supports transcription in *Streptococcus pneumoniae*. *PLOS Genetics*. 2021 30; 17:e1009542. DOI : 10.1371/journal.pgen.1009542
191. SUTORMIN D, GALIVONDZHYAN A, MUSHAROVA O, TRAVIN D, RUSANOVA A, OBRAZTSOVA K, BORUKHOV S et SEVERINOV K. Interaction Between Transcribing RNA Polymerase and Topoisomerase I Prevents R-loop Formation in *E. coli*. *bioRxiv*, 2021 26:2021.10.26.465782. DOI : 10.1101/2021.10.26.465782
192. DORMAN CJ, BARR GC, NI BHRIAIN N et HIGGINS CF. DNA supercoiling and the anaerobic and growth phase regulation of *tonB* gene expression. *Journal of Bacteriology*. 1988; 170:2816-26. DOI : 10.1128/jb.170.6.2816-2826.1988
193. GUO MS, HAAKONSEN DL, ZENG W, SCHUMACHER MA et LAUB MT. A bacterial chromosome structuring protein binds overtwisted DNA to stimulate type II topoisomerases and enable DNA replication. *Cell*. 2018 4; 175:583-597.e23. DOI : 10.1016/j.cell.2018.08.029
194. RAMÍREZ F, DÜNDAR F, DIEHL S, GRÜNING BA et MANKE T. deepTools : a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*. 2014; 42(Web Server issue):W187-191. DOI : 10.1093/nar/gku365

195. ZHANG Y, LIU T, MEYER CA, EECKHOUTE J, JOHNSON DS, BERNSTEIN BE, NUSBAUM C, MYERS RM, BROWN M, LI W et LIU XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biology*. 2008 ; 9:R137. DOI : 10.1186/gb-2008-9-9-r137
196. ONA S. ChIP Sequencing BioRender
197. SUTORMIN D, RUBANOVA N, LOGACHEVA M, GHILAROV D et SEVERINOV K. Single-nucleotide-resolution mapping of DNA gyrase cleavage sites across the Escherichia coli genome. *Nucleic Acids Research*. *Nucleic Acids Res* 2019 20; 47:1373-88. DOI : 10.1093/nar/gky1222
198. TARRY MJ, HARMEL C, TAYLOR JA, MARCZYNSKI GT et SCHMEING TM. Structures of GapR reveal a central channel which could accommodate B-DNA. *Scientific Reports*. 2019 13; 9:16679. DOI : 10.1038/s41598-019-52964-2
199. TAYLOR JA, PANIS G, VIOLLIER PH et MARCZYNSKI GT. A novel nucleoid-associated protein coordinates chromosome replication and chromosome partition. *Nucleic Acids Research*. 2017 6; 45:8916-29. DOI : 10.1093/nar/gkx596
200. LOURENÇO RF, SAURABH S, HERRMANN J, WAKATSUKI S et SHAPIRO L. The Nucleoid-Associated Protein GapR Uses Conserved Structural Elements To Oligomerize and Bind DNA. *mBio*. 2020 9; 11:e00448-20. DOI : 10.1128/mBio.00448-20
201. RICCI DP, MELFI MD, LASKER K, DILL DL, MCADAMS HH et SHAPIRO L. Cell cycle progression in Caulobacter requires a nucleoid-associated protein with high AT sequence recognition. *Proceedings of the National Academy of Sciences of the United States of America*. 2016 4; 113:E5952-E5961. DOI : 10.1073/pnas.1612579113
202. HUANG Q, DUAN B, QU Z, FAN S et XIA B. The DNA Recognition Motif of GapR Has an Intrinsic DNA Binding Preference towards AT-rich DNA. *Molecules*. 2021 24; 26:5776. DOI : 10.3390/molecules26195776
203. CECH T et PARDUE ML. Cross-linking of DNA with trimethylpsoralen is a probe for chromatin structure. *Cell*. 1977; 11:631-40. DOI : 10.1016/0092-8674(77)90080-0
204. YOAKUM GH et COLE RS. Cross-linking and relaxation of supercoiled DNA by psoralen and light. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*. 1978 21; 521:529-46. DOI : 10.1016/0005-2787(78)90295-2
205. SINDEN RR, CARLSON JO et PETTIJOHN DE. Torsional tension in the DNA double helix measured with trimethylpsoralen in living E. coli cells : analogous measurements in insect and human cells. *Cell*. 1980; 21:773-83. DOI : 10.1016/0092-8674(80)90440-7
206. BERMÚDEZ I, GARCÍA-MARTÍNEZ J, PÉREZ-ORTÍN JE et ROCA J. A method for genome-wide analysis of DNA helical tension by means of psoralen-DNA photobinding. *Nucleic Acids Research*. 2010; 38:e182. DOI : 10.1093/nar/gkq687
207. JUPE ER, SINDEN RR et CARTWRIGHT IL. Stably maintained microdomain of localized unrestrained supercoiling at a Drosophila heat shock gene locus. *The EMBO Journal*. 1993; 12:1067-75
208. MATSUMOTO K et HIROSE S. Visualization of unconstrained negative supercoils of DNA on polytene chromosomes of Drosophila. *Journal of Cell Science*. 2004 1; 117:3797-805. DOI : 10.1242/jcs.01225

209. KOUZINE F, GUPTA A, BARANELLO L, WOJTOWICZ D, BENAÏSSA K, LIU J, PRZYTYCKA TM et LEVENS D. Transcription dependent dynamic supercoiling is a short-range genomic force. *Nature structural & molecular biology*. 2013; 20:396-403. DOI : 10.1038/nsmb.2517
210. TEVES SS et HENIKOFF S. Transcription-generated torsional stress destabilizes nucleosomes. *Nature Structural & Molecular Biology*. 2014; 21:88-94. DOI : 10.1038/nsmb.2723
211. SAFFRAN WA, WELSH JT, KNOBLER RM, GASPARRO FP, CANTOR CR et EDELSON RL. Preparation and characterization of biotinylated psoralen. *Nucleic Acids Research*. 1988 11; 16:7221-31. DOI : 10.1093/nar/16.15.7221
212. DORMAN CJ. DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria. *Infection and Immunity*. 1991; 59:745-9
213. GRENIER AM, DUPORT G, PAGÈS S, CONDEMINÉ G et RAHBÉ Y. The Phytopathogen *Dickeya dadantii* (*Erwinia chrysanthemi* 3937) Is a Pathogen of the Pea Aphid. *Applied and Environmental Microbiology*. 2006; 72:1956-65. DOI : 10.1128/AEM.72.3.1956-1965.2006
214. MELOTTO M, UNDERWOOD W et HE SY. Role of stomata in plant innate immunity and foliar bacterial diseases. *Annual Review of Phytopathology*. 2008; 46:101-22. DOI : 10.1146/annurev.phyto.121107.104959
215. FORQUET R. Modelling transcriptional regulation by DNA supercoiling in bacteria. Thèse de doct. Université de Lyon, 2021 17
216. JAHN CE, SELIMI DA, BARAK JD et CHARKOWSKI AO. The *Dickeya dadantii* biofilm matrix consists of cellulose nanofibres, and is an emergent property dependent upon the type III secretion system and the cellulose synthesis operon. *Microbiology (Reading, England)*. 2011; 157(Pt 10):2733-44. DOI : 10.1099/mic.0.051003-0
217. CONDEMINÉ G, CASTILLO A, PASSERI F et ENARD C. The PecT repressor coregulates synthesis of exopolysaccharides and virulence factors in *Erwinia chrysanthemi*. *Molecular plant-microbe interactions : MPMI*. 1999; 12:45-52. DOI : 10.1094/MPMI.1999.12.1.45
218. HOMMAIS F, OGER-DESFEUX C, VAN GIJSEGEN F, CASTANG S, LIGORI S, EXPERT D, NASSER W et REVERCHON S. PecS Is a Global Regulator of the Symptomatic Phase in the Phytopathogenic Bacterium *Erwinia chrysanthemi* 3937. *Journal of Bacteriology*. 2008; 190:7508-22. DOI : 10.1128/JB.00553-08
219. ANTÚNEZ-LAMAS M, CABRERA-ORDÓÑEZ E, LÓPEZ-SOLANILLA E, RAPOSO R, TRELLES-SALAZAR O, RODRÍGUEZ-MORENO A et RODRÍGUEZ-PALENZUELA P. Role of motility and chemotaxis in the pathogenesis of *Dickeya dadantii* 3937 (ex *Erwinia chrysanthemi* 3937). *Microbiology (Reading, England)*. 2009; 155(Pt 2):434-42. DOI : 10.1099/mic.0.022244-0
220. SANTOS R, FRANZA T, LAPORTE ML, SAUVAGE C, TOUATI D et EXPERT D. Essential role of superoxide dismutase on the pathogenicity of *Erwinia chrysanthemi* strain 3937. *Molecular plant-microbe interactions : MPMI*. 2001; 14:758-67. DOI : 10.1094/MPMI.2001.14.6.758
221. FRANZA T et EXPERT D. Role of iron homeostasis in the virulence of phytopathogenic bacteria : an 'à la carte' menu - PMC. *Molecular Plant Pathology*. 2013
222. ROBERT-BAUDOY J, NASSER W, CONDEMINÉ G, REVERCHON S, SHEVCHIK V et HUGOUVIEUX-COTTE-PATTAT N. Pectic Enzymes of *Erwinia Chrysanthemi*, Regulation and Role in Pathogenesis. 2000 1. 221 p.

223. PY B, BORTOLI-GERMAN I, HAIECH J, CHIPPAUX M et BARRAS F. Cellulase EGZ of *Erwinia chrysanthemi* : structural organization and importance of His98 and Glu133 residues for catalysis. *Protein Engineering*. 1991 ; 4:325-33. DOI : 10.1093/protein/4.3.325
224. CONDEMINE G, DOREL C, HUGOUVIEUX-COTTE-PATTAT N et ROBERT-BAUDOY J. Some of the out genes involved in the secretion of pectate lyases in *Erwinia chrysanthemi* are regulated by *kdgR*. *Molecular Microbiology*. 1992 ; 6:3199-211. DOI : 10.1111/j.1365-2958.1992.tb01775.x
225. LÉTOFFÉ S, DELEPELAIRE P et WANDERSMAN C. Protease secretion by *Erwinia chrysanthemi* : the specific secretion functions are analogous to those of *Escherichia coli* alpha-haemolysin. *The EMBO journal*. 1990 ; 9:1375-82. DOI : 10.1002/j.1460-2075.1990.tb08252.x
226. LAMB C et DIXON RA. THE OXIDATIVE BURST IN PLANT DISEASE RESISTANCE. *Annual Review of Plant Physiology and Plant Molecular Biology*. 1997 ; 48:251-75. DOI : 10.1146/annurev.arplant.48.1.251
227. HASSOUNI ME, CHAMBOST JP, EXPERT D, VAN GIJSEGEN F et BARRAS F. The minimal gene set member *msrA*, encoding peptide methionine sulfoxide reductase, is a virulence determinant of the plant pathogen *Erwinia chrysanthemi*. *Proceedings of the National Academy of Sciences of the United States of America*. 1999 2 ; 96:887-92. DOI : 10.1073/pnas.96.3.887
228. GRIGNON C et SENTENAC H. pH and Ionic Conditions in the Apoplast. *Annual Review of Plant Physiology and Plant Molecular Biology*. 1991 ; 42:103-28. DOI : 10.1146/annurev.pp.42.060191.000535
229. HUGOUVIEUX-COTTE-PATTAT N, DOMINGUEZ H et ROBERT-BAUDOY J. Environmental conditions affect transcription of the pectinase genes of *Erwinia chrysanthemi* 3937. *Journal of Bacteriology*. 1992 ; 174:7807-18. DOI : 10.1128/jb.174.23.7807-7818.1992
230. COLLMER A et KEEN NT. The Role of Pectic Enzymes in Plant Pathogenesis. *Annual Review of Phytopathology*. 1986 ; 24:383-409. DOI : 10.1146/annurev.py.24.090186.002123
231. MOHNEN D. Pectin structure and biosynthesis. *Current Opinion in Plant Biology*. 2008 ; 11:266-77. DOI : 10.1016/j.pbi.2008.03.006
232. KEPSEU WD, SEPULCHRE JA, REVERCHON S et NASSER W. Toward a Quantitative Modeling of the Synthesis of the Pectate Lyases, Essential Virulence Factors in *Dickeya dadantii*. *The Journal of Biological Chemistry*. 2010 10 ; 285:28565-76. DOI : 10.1074/jbc.M110.114710
233. REVERCHON S, NASSER W et ROBERT-BAUDOY J. Characterization of *kdgR*, a gene of *Erwinia chrysanthemi* that regulates pectin degradation. *Molecular Microbiology*. 1991 ; 5:2203-16. DOI : 10.1111/j.1365-2958.1991.tb02150.x
234. FRANZA T. Role of iron homeostasis in the virulence of phytopathogenic bacteria : an 'à la carte' menu - PubMed
235. HÉRAULT E, REVERCHON S et 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*. *Environmental Microbiology*. 2014 ; 16:734-45. DOI : 10.1111/1462-2920.12198



236. REVERCHON S, VAN GIJSEGEN F, EFFANTIN G, ZGHIDI-ABOUZID O et NASSER W. Systematic targeted mutagenesis of the MarR/SlyA family members of *Dickeya dadantii* 3937 reveals a role for MfbR in the modulation of virulence gene expression in response to acidic pH. *Molecular Microbiology*. 2010; 78:1018-37. DOI : 10.1111/j.1365-2958.2010.07388.x
237. SEDDEK A, ANNAMALAI T et TSE-DINH YC. Type IA Topoisomerases as Targets for Infectious Disease Treatments. *Microorganisms*. 2021 1; 9:86. DOI : 10.3390/microorganisms9010086
238. TIRADO-VÉLEZ JM, CARREÑO D, SEVILLANO D, ALOU L, YUSTE J et CAMPA AG de la. Seconeolitsine, the Novel Inhibitor of DNA Topoisomerase I, Protects against Invasive Pneumococcal Disease Caused by Fluoroquinolone-Resistant Strains. *Antibiotics (Basel, Switzerland)*. 2021 13; 10:573. DOI : 10.3390/antibiotics10050573
239. SANDHAUS S, CHAPAGAIN PP et TSE-DINH YC. Discovery of novel bacterial topoisomerase I inhibitors by use of in silico docking and in vitro assays. *Scientific Reports*. 2018 23; 8:1437. DOI : 10.1038/s41598-018-19944-4
240. GODBOLE AA, AHMED W, BHAT RS, BRADLEY EK, EKINS S et NAGARAJA V. Targeting *Mycobacterium tuberculosis* topoisomerase I by small-molecule inhibitors. *Antimicrobial Agents and Chemotherapy*. 2015; 59:1549-57. DOI : 10.1128/AAC.04516-14
241. TEMESSZENTANDRÁSI-AMBRUS C, TÓTH S, VERMA R, BÁNHEGYI P, SZABADKAI I, BASKA F, SZÁNTAI-KIS C, HARTKOORN RC, LINGERFELT MA, SARKADI B, SZAKÁCS G, ÓRFI L, NAGARAJA V, EKINS S et TELBISZ Á. Characterization of new, efficient *Mycobacterium tuberculosis* topoisomerase-I inhibitors and their interaction with human ABC multidrug transporters. *PloS One*. 2018; 13:e0202749. DOI : 10.1371/journal.pone.0202749
242. GARCÍA MT, CARREÑO D, TIRADO-VÉLEZ JM, FERRÁNDIZ MJ, RODRIGUES L, GRACIA B, AMBLAR M, AINSA JA et CAMPA AG de la. Boldine-Derived Alkaloids Inhibit the Activity of DNA Topoisomerase I and Growth of *Mycobacterium tuberculosis*. *Frontiers in Microbiology*. 2018 24; 9:1659. DOI : 10.3389/fmicb.2018.01659
243. LEVENE SD. Analysis of DNA topoisomers, knots, and catenanes by agarose gel electrophoresis. *Methods in Molecular Biology (Clifton, NJ)*. 2009; 582:11-25. DOI : 10.1007/978-1-60761-340-4\_2
244. GUPTA R, RODRIGUES FELIX C, AKERMAN MP, AKERMAN KJ, SLABBER CA, WANG W, ADAMS J, SHAW LN, TSE-DINH YC, MUNRO OQ et ROHDE KH. Evidence for Inhibition of Topoisomerase 1A by Gold(III) Macrocycles and Chelates Targeting *Mycobacterium tuberculosis* and *Mycobacterium abscessus*. *Antimicrobial Agents and Chemotherapy*. 2018 26; 62:e01696-17. DOI : 10.1128/AAC.01696-17
245. RANJAN N, FULCRAND G, KING A, BROWN J, JIANG X, LENG F et ARYA DP. Selective Inhibition of Bacterial Topoisomerase I by alkynyl-bisbenzimidazoles. *MedChemComm*. 2014 1; 5:816-25. DOI : 10.1039/C4MD00140K
246. NIMESH H, SUR S, SINHA D, YADAV P, ANAND P, BAJAJ P, VIRDI JS et TANDON V. Synthesis and biological evaluation of novel bisbenzimidazoles as *Escherichia coli* topoisomerase IA inhibitors and potential antibacterial agents. *Journal of Medicinal Chemistry*. 2014 26; 57:5238-57. DOI : 10.1021/jm5003028

247. SANDHAUS S, ANNAMALAI T, WELMAKER G, HOUGHTEN RA, PAZ C, GARCIA PK, ANDRES A, NARULA G, RODRIGUES FELIX C, GEDEN S, NETHERTON M, GUPTA R, ROHDE KH, GIULIANOTTI MA et TSE-DINH YC. Small-Molecule Inhibitors Targeting Topoisomerase I as Novel Antituberculosis Agents. *Antimicrobial Agents and Chemotherapy*. 2016; 60:4028-36. DOI : 10.1128/AAC.00288-16
248. YU X, ZHANG M, ANNAMALAI T, BANSOD P, NARULA G, TSE-DINH YC et SUN D. Synthesis, evaluation, and CoMFA study of fluoroquinophenoxazine derivatives as bacterial topoisomerase IA inhibitors. *European Journal of Medicinal Chemistry*. 2017 5; 125:515-27. DOI : 10.1016/j.ejmech.2016.09.053
249. GARCIA PK, ANNAMALAI T, WANG W, BELL RS, LE D, MARTIN PANCORBO P, SIKANDAR S, SEDDEK A, YU X, SUN D, UHLEMANN AC, TIWARI PB, LENG F et TSE-DINH YC. Mechanism and resistance for antimycobacterial activity of a fluoroquinophenoxazine compound. *PLoS One*. 2019; 14:e0207733. DOI : 10.1371/journal.pone.0207733
250. TIWARI PB, CHAPAGAIN PP, SEDDEK A, ANNAMALAI T, ÜREN A et TSE-DINH YC. Covalent Complex of DNA and Bacterial Topoisomerase : Implications in Antibacterial Drug Development. *ChemMedChem*. 2020 3; 15:623-31. DOI : 10.1002/cmdc.201900721
251. GARCÍA MT, BLÁZQUEZ MA, FERRÁNDIZ MJ, SANZ MJ, SILVA-MARTÍN N, HERMOSO JA et CAMPA AG de la. New Alkaloid Antibiotics That Target the DNA Topoisomerase I of *Streptococcus pneumoniae*. *The Journal of Biological Chemistry*. 2011 25; 286:6402-13. DOI : 10.1074/jbc.M110.148148
252. LEWIS M. The lac repressor. *Comptes Rendus Biologies. Retour sur l'operon lac* 2005 1; 328:521-48. DOI : 10.1016/j.crvi.2005.04.004
253. KERCHER MA, LU P et LEWIS M. Lac repressor—operator complex. *Current Opinion in Structural Biology*. 1997 1; 7:76-85. DOI : 10.1016/S0959-440X(97)80010-3
254. XIA Y, LI K, LI J, WANG T, GU L et XUN L. T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis. *Nucleic Acids Research*. 2019 20; 47:e15. DOI : 10.1093/nar/gky1169
255. PINEAU M, MARTIS B S, FORQUET R, BAUDE J, VILLARD C, GRAND L, POPOWYCZ F, SOULÈRE L, HOMMAIS F, NASSER W, REVERCHON S et MEYER S. What is a supercoiling-sensitive gene? Insights from topoisomerase I inhibition in the Gram-negative bacterium *Dickeya dadantii*. *Nucleic Acids Research*. 2022 9; 50:9149-61. DOI : 10.1093/nar/gkac679
256. TAN K, ZHOU Q, CHENG B, ZHANG Z, JOACHIMIAK A et TSE-DINH YC. Structural basis for suppression of hypernegative DNA supercoiling by *E. coli* topoisomerase I. *Nucleic Acids Research*. 2015 15; 43:11031-46. DOI : 10.1093/nar/gkv1073
257. CHENG HR et JIANG N. Extremely rapid extraction of DNA from bacteria and yeasts. *Biotechnology Letters*. 2006; 28:55-9. DOI : 10.1007/s10529-005-4688-z
258. RAMÍREZ F, RYAN DP, GRÜNING B, BHARDWAJ V, KILPERT F, RICHTER AS, HEYNE S, DÜNDAR F et MANKE T. deepTools2 : a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*. 2016 8; 44(W1):W160-165. DOI : 10.1093/nar/gkw257

259. MATTHEY-DORET C, BAUDRY L, AMAURY, AXELCOURNAC, REMI-MONTAGNE, GUIGLIELMONI N, FOUTEL-RODIER T et SCOLARI VF. `koszullab/hicstuff` : More flexible view subcommand. 2020 5. DOI : 10.5281/zenodo.4066351
260. MATTHEY-DORET C, BAUDRY L, BREUER A, MONTAGNE R, GUIGLIELMONI N, SCOLARI V, JEAN E, CAMPEAS A, CHANUT PH, ORIOL E, MÉOT A, POLITIS L, VIGOUROUX A, MOREAU P, KOSZUL R et COURNAC A. Computer vision for pattern detection in chromosome contact maps. *Nature Communications*. 2020 16; 11:5795. DOI : 10.1038/s41467-020-19562-7
261. EAGEN KP. Principles of Chromosome Architecture Revealed by Hi-C. *Trends in Biochemical Sciences*. 2018 1; 43:469-78. DOI : 10.1016/j.tibs.2018.03.006
262. NORRIS V et MADSEN MS. Autocatalytic gene expression occurs via transertion and membrane domain formation and underlies differentiation in bacteria : a model. *Journal of Molecular Biology*. 1995 10; 253:739-48. DOI : 10.1006/jmbi.1995.0587
263. LIBBY EA, ROGGIANI M et GOULIAN M. Membrane protein expression triggers chromosomal locus repositioning in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*. 2012 8; 109:7445-50. DOI : 10.1073/pnas.1109479109
264. BAKSHI S, CHOI H, MONDAL J et WEISSHAAR JC. Time-dependent Effects of Transcription- and Translation-halting Drugs on the Spatial Distributions of the *E. coli* Chromosome and Ribosomes. *Molecular microbiology*. 2014; 94:871-87. DOI : 10.1111/mmi.12805
265. PETER BJ, ULLSPERGER C, HIASA H, MARIANS KJ et COZZARELLI NR. The structure of supercoiled intermediates in DNA replication. *Cell*. 1998 18; 94:819-27. DOI : 10.1016/s0092-8674(00)81740-7
266. VOLOGODSKII A. DNA supercoiling helps to unlink sister duplexes after replication. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*. 2010; 32:9-12. DOI : 10.1002/bies.200900143
267. LIM CJ, KENNEY LJ et YAN J. Single-molecule studies on the mechanical interplay between DNA supercoiling and H-NS DNA architectural properties. *Nucleic Acids Research*. 2014 1; 42:8369-78. DOI : 10.1093/nar/gku566
268. SCHNURR B, VORGAS C et STAVANS J. Compaction and supercoiling of single, long dna molecules by hu protein. *Biophysical Reviews and Letters*. 2006; 01:29-44. DOI : 10.1142/S1793048006000021
269. DAHLKE K et SING CE. Influence of Nucleoid-Associated Proteins on DNA Supercoiling. *The Journal of Physical Chemistry B*. 2019 5; 123:10152-62. DOI : 10.1021/acs.jpcc.9b07436
270. KAHRAMANOGLU C, SESHASAYEE ASN, PRIETO AI, IBBERSON D, SCHMIDT S, ZIMMERMANN J, BENES V, FRASER GM et LUSCOMBE NM. Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Research*. 2011; 39:2073-91. DOI : 10.1093/nar/gkq934
271. PRIETO AI, KAHRAMANOGLU C, ALI RM, FRASER GM, SESHASAYEE ASN et LUSCOMBE NM. Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12. *Nucleic Acids Research*. 2012; 40:3524-37. DOI : 10.1093/nar/gkr1236

272. LILLEY DMJ, CHEN D et BOWATER RP. DNA supercoiling and transcription : topological coupling of promoters. *Quarterly Reviews of Biophysics*. 1996; 29:203-25. DOI : 10.1017/S0033583500005825
273. AHMED W, SALA C, HEGDE SR, JHA RK, COLE ST et NAGARAJA V. Transcription facilitated genome-wide recruitment of topoisomerase I and DNA gyrase. *PLoS genetics*. 2017; 13:e1006754. DOI : 10.1371/journal.pgen.1006754
274. CHEN D, BOWATER R, DORMAN CJ et LILLEY DM. Activity of a plasmid-borne leu-500 promoter depends on the transcription and translation of an adjacent gene. *Proceedings of the National Academy of Sciences of the United States of America*. 1992 15; 89:8784-8. DOI : 10.1073/pnas.89.18.8784
275. SZAFRAN MJ, GONGEROWSKA M, MAŁECKI T, ELLIOT M et JAKIMOWICZ D. Transcriptional Response of *Streptomyces coelicolor* to Rapid Chromosome Relaxation or Long-Term Supercoiling Imbalance. *Frontiers in Microbiology*. 2019; 10. DOI : 10.3389/fmicb.2019.01605
276. GMUENDER H, KURATLI K, DI PADOVA K, GRAY CP, KECK W et EVERS S. Gene Expression Changes Triggered by Exposure of *Haemophilus influenzae* to Novobiocin or Ciprofloxacin : Combined Transcription and Translation Analysis. *Genome Research*. 2001; 11:28-42
277. KHAYI S, BLIN P, CHONG TM, ROBIC K, CHAN KG et FAURE D. Complete Genome Sequences of the Plant Pathogens *Dickeya solani* RNS 08.23.3.1.A and *Dickeya dianthicola* RNS04.9. *Genome Announcements*. 2018 25; 6:10.1128/genomea.01447-17. DOI : 10.1128/genomea.01447-17
278. MA X, PERNA NT, GLASNER JD, HAO J, JOHNSON S, NASARUDDIN AS, CHARKOWSKI AO, WU S, FEI Z, PERRY KL, STODGHILL P et SWINGLE B. Complete Genome Sequence of *Dickeya dianthicola* ME23, a Pathogen Causing Blackleg and Soft Rot Diseases of Potato. *Microbiology Resource Announcements*. 2019 14; 8:e01526-18. DOI : 10.1128/MRA.01526-18





## FOLIO ADMINISTRATIF

### THESE DE L'INSA LYON, MEMBRE DE L'UNIVERSITE DE LYON

NOM : PINEAU

DATE de SOUTENANCE : 24/11/2023

(avec précision du nom de jeune fille, le cas échéant)

Prénoms : Maïwenn, Maëliiss, Cécile

TITRE : Régulation globale de la transcription bactérienne par le surenroulement de l'ADN

NATURE : Doctorat

Numéro d'ordre : 2023ISAL0091

Ecole doctorale : E2M2 Ecosystèmes Evolution Modélisation Microbiologie

Spécialité : Biomath-Bioinfo-Génomique évolutive

#### RESUME :

Chez les bactéries, le chromosome est situé dans le cytoplasme, sous une forme très compacte. Cette compaction résulte notamment du surenroulement de l'ADN (SC), c'est à dire de la déformation torsionnelle de la double-hélice de l'ADN. Selon les conditions environnementales rencontrées par les bactéries, le niveau de SC peut varier. Ce niveau est principalement contrôlé par la gyrase, qui augmente le niveau de SC, et la topoisomérase I, qui relâche l'ADN. La régulation du niveau de SC est très importante car le SC est un régulateur de l'expression des gènes. L'objectif de ma thèse est de caractériser la régulation globale de la transcription bactérienne par le SC.

Nous avons obtenu le premier transcriptome d'une bactérie à Gram négatif dont la topoisomérase I est inhibée par un antibiotique. Nous avons mis une nouvelle fois en évidence une régulation globale et complexe de la transcription par le SC et nous avons découvert que la réponse des gènes à une variation de SC dépend du niveau d'expression et du contexte génomique des gènes, de la direction de la variation de SC et du contexte physiologique de la bactérie.

J'ai ensuite développé un package Python facilitant les analyses statistiques reproductibles de données expérimentales (ChIP-Seq, RNA-Seq...) et d'annotations dans le cadre de l'étude de l'expression d'un génome bactérien selon ses propriétés spatiales. Avec ce package j'ai pu exploiter des données publiées sur la fixation de la topoisomérase I et de la gyrase le long du chromosome. En analysant ces données qui donnent un aperçu du niveau de SC local, j'ai pu, d'une part, étudier l'effet de la transcription sur le niveau de SC à une échelle de 10 kb autour des unités de transcription et, d'autre part, observer une organisation méconnue du chromosome en domaines de 100-200 kb environ.

MOTS-CLÉS : Surenroulement de l'ADN, régulation de la transcription, RNA-Seq, ChIP-Seq, package bioinformatique,

Laboratoire de recherche : Microbiologie, Adaptation et Pathogénie (MAP), UMR5240

Directeur de thèse : Sam MEYER

Co-directrice de thèse : Florence HOMMAIS

Président de jury :

Composition du jury :

Jean-Yves BOUET  
Ivan JUNIER  
Virginia LIOY  
Florence POPOWYCZ  
Sam MEYER  
Florence HOMMAIS

Directeur de recherche  
Directeur de recherche  
Chargée de recherche  
Professeure  
Maître de conférences  
Professeure

LMGM-CBI, CNRS  
TIMC, CNRS  
I2BC, CNRS  
ICBMS, INSA Lyon  
MAP, INSA Lyon  
MAP. UCBL

Rapporteur  
Rapporteur  
Examinatrice  
Examinatrice  
Directeur de thèse  
Co-directrice de thèse