



HAL
open science

Identification de contraintes locales impactant l'expression des gènes chez *Escherichia coli*

Nazim Sarica

► **To cite this version:**

Nazim Sarica. Identification de contraintes locales impactant l'expression des gènes chez *Escherichia coli*. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université Paris - Saclay, 2020. Français. NNT : 2020UPASL021 . tel-04411260

HAL Id: tel-04411260

<https://hal.science/tel-04411260>

Submitted on 23 Jan 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.

Identification de contraintes locales
impactant l'expression des gènes
chez *Escherichia coli*.

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°577 : Structure et Dynamique des Systèmes Vivants

Spécialité de doctorat: voir annexe

Unité de recherche : Université Paris-Saclay, Univ Evry, CNRS, CEA, GM UMR8030

Référent : Université d'Evry Val d'Essonne

Soutenance prévue à EVRY, le 13 Octobre 2020, par

Nazim SARICA

Composition du Jury

Didier MAZEL

Professeur des Universités,
Institut Pasteur

Rapporteur

Bianca SCLAVI

Directrice de Recherches,
Sorbonne Université

Rapporteur

François KEPES

Professeur des Universités,
Retraité

Examineur

Ivan JUNIER

Chargé de Recherches,
Université Grenoble Alpes

Examineur

Stéphane DUIGOU

Chargé de Recherches,
Université Paris-Saclay

Examineur

Laurent JANNIERE

Directeur de Recherches,
CNRS

Directeur de thèse

REMERCIEMENTS

Dans cette section je cite les personnes et institutions qui m'ont permis de réaliser ce travail.

Tout d'abord, je tiens à faire un éloge de Laurent JANNIERE, mon directeur de thèse, sans qui je n'aurais pas pu réaliser ma thèse à l'iSSB. Il est et restera un exemple à mes yeux, tant au niveau scientifique qu'en tant que personne. Sa rigueur, sa dévotion et son honnêteté sont des vertus rares de nos jours.

Je tiens à remercier François KEPES, qui m'a donné l'opportunité de commencer ma carrière scientifique à l'iSSB à la fin de mes études, et m'a suivi tout au long de ma thèse malgré sa retraite. Merci d'avoir partagé vos connaissances et votre expérience par des conseils et suggestions tant pour le travail expérimental que pour l'écriture de l'article scientifique et du présent manuscrit.

Je suis reconnaissant envers Brian JESTER qui m'a confié ce projet dès mon arrivée dans l'équipe et m'a fait confiance pour le conduire pendant ces années. Dans le monde scientifique les nouvelles idées font le génie de l'innovation, les siennes sont toujours novatrices, pertinentes et nombreuses. Merci de m'avoir donné la liberté d'explorer de nouveaux aspects dans ce projet.

Je remercie Bianca SCLAVI et Didier MAZEL d'avoir accepté d'être rapporteurs pour ma thèse, ainsi que François KEPES, Ivan JUNIER et Stéphane DUIGOU d'avoir accepté de participer à mon jury de thèse en tant qu'examineurs.

Je remercie mes anciens collègues Steff HOREMANS et Morgane CHAMPLEBOUX pour leurs conseils techniques et leur effort pour rendre le lieu de travail plus joyeux.

Un grand merci à l'adorable Sylvie BOBELET pour sa bonne humeur et son aide tout au long de ma thèse. Avant même ma venue dans ce laboratoire, ses œuvres ont facilité mon arrivée et mon installation dans la région. Nos pauses déjeuners kebabs ont été marquantes et essentielles à la réussite de ma thèse.

Mes participations aux équipes iGEM 2018 et iGEM 2019 ont été une superbe expérience. Merci à Ioana POPESCU d'avoir construit ces équipes et de m'y avoir invité.

Tout ce travail n'aurait pas été possible sans motivation et appétence depuis le début de ma formation. C'est pourquoi je remercie d'abord Cyril BONTEMPS, mon premier encadrant de stage, qui reste à l'heure actuelle un des meilleurs enseignants que j'ai eu lors de ma formation. Merci d'avoir attisé la curiosité scientifique qui est désormais en moi. J'ai ensuite eu la chance de réaliser un stage dans le même laboratoire auprès de Virginie LIBANTE. Sa dévotion extraordinaire et sa motivation m'ont stupéfait, et la formation que j'ai reçue de sa part a fait de moi le chercheur que je suis aujourd'hui. Ces courts mois à tes côtés m'ont permis d'apprendre à « bien manipuler », essentiel dans le monde moléculaire. J'y ai appris à travailler méticuleusement et proprement, sans quoi je pense j'aurais dégradé une grande partie des ARN avec lesquels j'ai travaillé pendant ma thèse.

Je tiens à remercier le Ministère de l'Éducation Nationale pour la bourse qu'il m'a accordé, et le système éducatif français dans sa globalité qui permet à tous les désireux de poursuivre des études supérieures. Merci à l'Université Paris-Saclay, à l'Université d'Evry Val d'Essonne, et à l'École Doctorale SDSV. Je remercie l'iSSB et l'équipe MEGA/ASG de m'avoir permis de réaliser ma thèse dans leurs locaux.

De manière globale, je tiens à remercier certains de mes amis et proches qui m'ont moralement soutenu tout au long de cette aventure. Pour conclure, je tiens à remercier divers

acteurs non-présents dans le labo mais qui ont eu un grand impact dans ma vie. D'abord, l'équipe Géofoot pour toutes ces sessions les mardi midi, ce fut un réel plaisir et je vous demande encore une fois pardon pour toutes les passes que j'ai ratées. Ensuite, mon kiné M. JULIEN, qui m'a permis de retrouver les terrains de football 5 mois seulement après mon opération du genou. Et enfin, celui qui est désormais mon ami, Hamid, gérant du restaurant rapide aux terrasses à Evry, chez qui j'ai bu plus de cafés que je n'ai mangé de frites.

« It must be a strange world not being a scientist, going through life not knowing – or maybe not caring – about where the air came from, and where the stars at night came from, or how far they are from us. I want to know » - Michio Kaku

TABLE DES FIGURES

Figure 1 : Représentation de la structure des bases de l'ADN.....	5
Figure 2 : Différentes représentations de la double hélice ADN (12).....	6
Figure 3 : Reconnaissance du promoteur bactérien durant l'initiation de la transcription (45).....	13
Figure 4 : Spécificité des éléments du promoteur (68).....	17
Figure 5 : Composants d'un promoteur bactérien.....	18
Figure 6 : Représentation d'un ARNm naissant contenant un terminateur de transcription.....	21
Figure 7 : Activation de la transcription par un changement conformationnel de l'ADN (127).....	25
Figure 8 : Mécanismes de répression de l'initiation de la transcription (136).....	27
Figure 9 : Organisation d'un opéron contenant 3 gènes.....	34
Figure 10 : Organisation du nucléoïde d'E.coli (176).....	35
Figure 11 : Chromosome d'E. coli organisé en microdomaines.....	38
Figure 12 : Courbures, embobinages, et pontages de l'ADN par des protéines architecturales (19).....	40
Figure 13 : Représentation de l'opéron lactose.....	44
Figure 14 : Représentation schématique des régions opératrices du phage lambda (230).....	45
Figure 15 : La liaison de NAP et le surenroulement influencent la structure du nucléoïde (18).....	46
Figure 16 : Structure tertiaire de l'ADN d'un plasmide de 5 kpb observée par microscopie électronique (236)...	48
Figure 17 : Représentation du surenroulement dans un ADN circulaire (237).....	48
Figure 18 : Liens entre 2 courbes fermées.....	48
Figure 19 : Représentation des géométries adoptées par un ADN surenroulé (243).....	49
Figure 20 : Modèle « Twin-supercoiled domain » (260).....	51
Figure 21 : Un cycle de décaténation par une Topoisomérase I (296).....	56
Figure 22 : Représentation de la construction introduite dans le génome par mutagenèse par transposition....	61
Figure 23 : Transcription du gène emGFP dans la cassette BCJ932 dans 209 positions le long du chromosome d'E. coli.....	61
Figure 24 : Construction de la souche exprimant lambda-cl sous contrôle du promoteur P _{rha}	65
Figure 25 : Gel de migration de Western Blot montrant l'expression de lambda-cl.....	65
Figure 26 : Représentation du gène emGFP sous contrôle du promoteur muté.....	66
Figure 27 : Représentation de la construction 932 et ses sites génomiques d'insertion.....	68
Figure 28 : Impact de la formation d'une boucle d'ADN sur l'expression du gène emGFP placé sous contrôle du promoteur p1 dans les régions Ori et Ter.....	69
Figure 29 : Représentation de la construction BCJ927.....	71
Figure 30 : Impact de la formation d'une boucle d'ADN sur l'expression du gène emGFP sous contrôle du promoteur fort p3 dans les régions Ori et Ter.....	71
Figure 31 : Représentation des constructions contenant le gène emGFP et le GRS.....	74
Figure 32 : Impact de la formation d'une boucle d'ADN sur l'expression du gène emGFP sous contrôle du promoteur p3 dans les régions Ori et Ter en présence du GRS ou d'un fragment d'ADN neutre.....	75
Figure 33 : Représentation des constructions arborant 2 gènes à l'intérieur de la boucle d'ADN.....	77
Figure 34 : Niveaux d'expression du gène emGFP dans les 3 configurations et lorsque le gène mCherry est faiblement transcrit.....	78
Figure 35 : Niveaux d'expression du gène emGFP dans les 3 configurations et lorsque le gène mCherry est fortement transcrit.....	78
Figure 36 : Comparaison des niveaux d'expression du gène emGFP dans les 3 configurations, en fonction du niveau de transcription du gène mCherry.....	79
Figure 37 : Niveaux d'expression du gène mCherry en fonction du promoteur et de l'orientation relative par rapport au gène emGFP.....	80
Figure 38 : Représentation des constructions à 2 gènes arborant le GRS.....	81
Figure 39 : Niveaux d'expression du gène emGFP dans les boucles à 2 gènes avec GRS.....	82
Figure 40 : Niveaux d'expression du gène mCherry dans les boucles à 2 gènes avec GRS.....	83
Figure 41 : Représentation des différentes conformation adoptées par une région ADN en fonction de la liaison de Lambda-cl.....	84
Figure 42 : Représentation des 3 mutants de lambda-cl.....	85
Figure 43 : Expression de emGFP (en vert) et mCherry (en rouge) organisés en tandem en fonction des mutants de Lambda-cl.....	86

<i>Figure 44 : Données de Chip-seq et RNA-seq pour LRP au niveau de certains loci génomiques dans un milieu minimum.</i>	88
<i>Figure 45 : Données de Chip-seq et RNA-seq pour LRP au niveau de certains loci génomiques dans un milieu minimum.</i>	89
<i>Figure 46 : Représentation de l'accumulation de surenroulements dans les boucles à 2 gènes.</i>	113
<i>Figure 47 : Représentation de l'activité d'un promoteur en fonction des niveaux de surenroulement négatif accumulés dans et autour de sa séquence.</i>	115
<i>Figure 48 : Représentation schématique de l'organisation chromosomique potentielle permise par l'activité de 3 NAP.</i>	117

TABLE DES TABLEAUX

<i>Tableau 1 : Souches construites.....</i>	<i>94</i>
<i>Tableau 2 : Plasmides construits</i>	<i>94</i>
<i>Tableau 3 : Amorces utilisées.</i>	<i>99</i>

Abréviations

ADN : Acide DésoxyriboNucléique
ADNc : Acide DésoxyriboNucléique complémentaire
ARN : Acide RiboNucléique
ARNm : Acide RiboNucléique messenger
Chip-seq : séquençage Chip
EC : Complexe d'élongation
emGFP : emerald Green Fluorescent Protein
kpb : kilo paires de bases
Lk : linking number ou enlacement
LRP : Leucine-responsive Regulatory Protein
Mpb : Marge paires de bases
NAP : Protéines Associées au Nucléoïde
Ori : origine de réplication
pb : paires de bases
PCR : Réaction en Chaîne par Polymérase
qPCR : Réaction en Chaîne par Polymérase quantitative
rNTP : RiboNucléosides Tri-Phosphate
RNA-seq : séquençage aléatoire du transcriptome entier
RPc : complexe ARN polymérase/Promoteur fermé
RPo : complexe ARN polymérase/promoteur ouvert
RT : Transcription inverse
TBE : Tris, Boric Acid, EDTA
Ter : Terminus de réplication
TPC : Transcrits Par Cellule
TSS : Site d'initiation de la Transcription
Tw : Twist ou torsion
Wr : Writhe ou courbure

TABLE DES MATIERES

1	Introduction générale.....	2
1.1	Bactérie <i>E. coli</i>	3
1.1.1	Description générale.....	3
1.1.2	La souche K12 MG1655	4
1.2	La structure de l'ADN.....	4
1.2.1	ADN type B.....	6
1.2.2	ADN de type A.....	7
1.2.3	ADN de type Z.....	7
1.2.4	ADN dans un nucléoïde bactérien.....	8
2	Expression génique bactérienne.....	10
2.1	Transcription.....	11
2.1.1	ARN polymérase.....	11
2.1.1.1	Initiation de la transcription.....	11
2.1.1.2	Élongation.....	13
2.1.1.3	Terminaison.....	15
2.1.2	Promoteur.....	15
2.1.2.1	Boîtes -35 et -10.....	16
2.1.2.2	Spacer.....	17
2.1.2.3	UP-Element.....	18
2.1.2.4	Région entre la boîte -10 et le TSS.....	19
2.1.3	Termineurs de la transcription.....	20
2.1.3.1	Terminaison intrinsèque de la transcription.....	20
2.1.3.2	Terminaison dépendante de facteurs extérieurs.....	22
2.2	Régulations.....	22
2.2.1	Régulations transcriptionnelles.....	23
2.2.1.1	Régulation par un facteur de transcription.....	24
2.2.1.1.1	Activation en modifiant la conformation de l'ADN.....	24
2.2.1.1.2	Activation par contact direct avec l'ARN polymérase.....	26
2.2.1.1.3	Répression.....	26
2.2.1.1.4	Exemple : la protéine CRP.....	28
2.2.1.2	Promoteurs dépendant du facteur σ^{54}	28
2.2.1.3	Promoteurs dépendants d'autres facteurs σ	29
2.2.1.4	Activation indirecte.....	29
2.2.1.5	Régulation par les petits ligands.....	29
2.2.1.6	Structure du chromosome et protéines associées.....	30
2.2.2	Régulations post-transcriptionnelles des ARNm.....	30
2.3	Organisation chromosomique.....	33
2.3.1	Unités transcriptionnelles : gènes et opérons.....	33
2.3.2	Organisation tridimensionnelle.....	34
2.3.2.1	Macrodomaines.....	35
2.3.2.2	Microdomaines.....	36
2.4	Protéines « architectes ».....	39
2.4.1	Wrapping.....	41
2.4.2	Courbures.....	41
2.4.3	Boucles.....	42
2.4.3.1	Lacl.....	43
2.4.3.2	Lambda-cl.....	44
2.4.4	Implication des NAP dans le maintien de la topologie.....	46
2.5	Surenroulement.....	47
2.5.1	Plectonèmes.....	49
2.5.2	Liens entre surenroulement et transcription.....	50

2.5.2.1	Le surenroulement local affecte la transcription.....	50
2.5.2.2	La transcription affecte le surenroulement local.....	50
2.5.2.2.1	Twin-supercoiled domain model.....	50
2.5.2.2.2	Transcriptional bursting.....	52
2.5.3	Régulation de la topologie	53
2.5.3.1	Topoisomérases de type IA.....	54
2.5.3.2	Topoisomérases de type II	56
2.5.3.2.1	Généralités.....	56
2.5.3.2.2	Gyrase.....	57
2.5.4	Autres facteurs impliqués dans le niveau de surenroulement	58
2.5.5	R-loops	59
2.6	Mon projet de recherche	60
3	Resultats.....	64
3.1	Introduction de la protéine Lambda-cl dans <i>E. coli</i>	64
3.2	L'expression de la cassette BCJ932 dépend uniquement du promoteur p1	66
3.3	L'induction de la boucle d'ADN homogénéise l'expression du gène rapporteur <i>emGFP</i> sous contrôle du promoteur faible p1.....	67
3.4	L'induction de la boucle d'ADN homogénéise l'expression du gène rapporteur <i>emGFP</i> sous contrôle du promoteur fort p3	70
3.5	La suppression de supertours positifs à l'intérieur de la boucle augmente l'expression du gène <i>emGFP</i>	72
3.6	L'expression du gène <i>emGFP</i> est stimulée lorsque les gènes sont co-orientés	76
3.7	La diminution des supertours positifs dans une boucle à deux gènes a un effet positif global sur l'expression.....	80
3.8	La capacité d'oligomérisation de Lambda-cl est essentielle pour la formation de boucles dans l'ADN.....	83
3.9	LRP régule l'expression génique par la formation de boucles dans l'ADN	86
4	Matériel et méthodes.....	91
4.1	Construction des souches bactériennes et plasmides	91
4.1.1	Construction des plasmides	99
4.1.2	Construction des souches	100
4.2	Conditions de culture cellulaire	101
4.3	Méthodes de biologie moléculaire	101
4.4	Extraction ARN.....	103
4.5	Transcription inverse et PCR quantitative (RT-qPCR).....	103
4.6	Analyse de données.....	104
5	Discussion.....	106
5.1	L'expression génique dépend fortement du contexte topologique local.....	106
5.2	Les gènes situés dans une boucle d'ADN artificiellement induite ne sont plus sensibles au contexte topologique local.....	108
5.3	La diminution des supertours positifs au sein de la boucle augmente l'expression de gènes sous contrôle d'un promoteur fort.....	109
5.4	Le changement de surenroulement causé par l'expression d'un gène affecte l'expression du gène voisin et vice versa.....	110
5.5	Le surenroulement négatif affecte la transcription de manière spécifique à chaque promoteur.....	114
5.6	Les boucles d'ADN sont des régulateurs épigénétiques de l'expression de nombreux gènes chez <i>E. coli</i>	115
5.7	Conclusion	118
5.8	Perspectives.....	118

6 Bibliographie122

1 - INTRODUCTION GENERALE

1 INTRODUCTION GENERALE

Le domaine du vivant est composé de 2 types d'organismes : les procaryotes et les eucaryotes. La différence fondamentale entre procaryotes et eucaryotes se trouve au niveau de la présence d'un noyau contenant l'information génétique dans la cellule. Les eucaryotes contiennent des organites, et en particulier un noyau portant l'information génétique. Les procaryotes ne possèdent pas de noyau, et leur information génétique est confinée dans un complexe macromoléculaire appelé nucléoïde. Les procaryotes sont divisés en 2 domaines, les Archées et les Bactéries.

Les bactéries sont des organismes unicellulaires d'une taille avoisinant les quelques micromètres. Elles peuvent avoir des formes différentes dont les plus communes sont sphériques (coques), en bâtonnets (bacilles) ou des formes spiralées. Elles sont considérées comme l'une des premières formes de vie apparue sur Terre, et sont capables de coloniser des types d'habitat très variés : les sols, les eaux, les sources d'eau chaude acides, les déchets radioactifs (1), ou encore la croûte terrestre. Les bactéries peuvent vivre en symbiose ou en parasitisme avec les plantes et les animaux. Chez les humains, les bactéries se trouvent majoritairement dans le tractus digestif, où elles ont un rôle commensal ou bénéfique, mais également sur la peau. Toute vie animale est dépendante des bactéries étant donné qu'elles seules (et quelques archées) sont capables de synthétiser la vitamine B12 (aussi appelée cobalamine), essentielle aux animaux.

Bien que la majorité des bactéries soient inoffensives ou bénéfiques pour l'homme, il existe un nombre conséquent d'espèces bactériennes responsables de maladies infectieuses importantes. Ces maladies peuvent être à transmission respiratoire (Pneumocoques ; Tuberculose ; ...), à transmission orale (Salmonelloses ; Shigellose ; Listériose ; ...), à transmission sexuelle (Chlamydie ; Syphilis ; ...), ou transmises par les animaux (Borréliose ; Brucellose ; ...). L'utilisation d'antibiotiques pour contrer ces maladies, mais aussi dans le secteur de l'agriculture, a fait de la multi-résistance antibiotique un des problèmes de santé publique majeurs de nos jours. C'est pourquoi la microbiologie, et plus précisément la bactériologie, se concentre sur le développement de nouvelles molécules à capacité antibiotique.

Il y a environ 5×10^{30} bactéries sur Terre (2), formant une biomasse excédant celles des plantes et des animaux. Un gramme de terre contient environ 40 millions de bactéries, et 1 millilitre

d'eau peut en contenir plusieurs millions. Au fur et à mesure de son évolution, l'homme a domestiqué l'utilisation de bactéries à son profit : production de fromages et de yaourts, nettoyage des eaux usées, ou encore utilisation en biotechnologies pour la synthèse de molécules d'intérêt.

1.1 Bactérie *E. coli*

1.1.1 Description générale

C'est en 1885 que le pédiatre allemand Théodore Escherich observe pour la première fois le bacille *Bacterium coli commune* (par la suite appelé *Escherichia coli*) isolé à partir de selles de nourrissons. *Escherichia coli*, communément appelée *E. coli*, est une bactérie Gram négatif du genre *Escherichia*, qui appartient à la famille des *Enterobacteriaceae*. Ce bacille anaérobie facultatif en forme de bâtonnet est principalement trouvé dans l'intestin des animaux à sang chaud (3). Ces bactéries sont catalase positives et ne possèdent pas d'oxydase (4). Chez l'homme, ce coliforme fécal colonise le tractus gastro-intestinal dans les premières heures de la vie et constitue 80% de la flore intestinale aérobie (5). La relation entre l'hôte et *E. coli* est souvent commensale, c'est à dire bénéfique pour la bactérie qui va se nourrir de nutriments fournis par l'hôte, et inoffensive pour ce dernier, voire bénéfique par la production de vitamines ou en occupant une niche qui pourrait être la cible d'organismes malveillants. Toutefois, certaines souches d'*E. coli* peuvent être pathogènes et ainsi causer des gastro-entérites, des infections urinaires ou des infections alimentaires. Les comparaisons entre génomes de 20 espèces des genres *Escherichia* et de *Shigella* (très proche d'*E. coli*) ont montré que le core-génome (gènes présents dans toutes les souches d'une espèce) est composé d'environ 2 000 gènes, et que le pangénome (gamme complète de gènes dans une espèce) contient environ 18 000 gènes.

E. coli a été intensivement étudiée en laboratoire à travers le monde depuis plus de 60 années, notamment car sa culture est jugée simple et peu onéreuse. C'est une bactérie chimiotrophe qui a besoin d'une source de carbone et d'énergie dans son milieu de culture, dont le temps de génération en conditions optimales avoisine les 20 minutes. Ces caractéristiques font d'*E. coli* le modèle procaryote le plus utilisé en biotechnologies et microbiologie. C'est pourquoi il faisait partie des premiers modèles utilisés pour le séquençage génomique.

1.1.2 La souche K12 MG1655

E. coli K12 MG1655 est la souche d'*E. coli* la plus utilisée pour les études en laboratoire. Cette souche, a été faiblement modifiée afin de pouvoir survivre dans les conditions de laboratoire et non dans l'intestin humain. Les modifications procédées sont la suppression du bactériophage Lambda et du plasmide F, respectivement grâce à l'utilisation de rayonnements Ultra-Violets (UV) et d'acridine orange (6). Ces traitements ont également entraîné quelques autres modifications au niveau de l'ADN de la bactérie : une mutation dans la séquence du gène *rph* entraînant une auxotrophie de pyrimidine, une mutation dans le gène *ilvG* impactant la voie de synthèse de l'Isoleucine, et une mutation appelée *rfb-50* responsable de l'absence de l'antigène O dans les lipopolysaccharides (7)(8).

Cette souche présente une molécule d'ADN circulaire avoisinant 4,6 millions de paires de bases, contenant environ 4 300 séquences codant des protéines (9). Ce génome contient un grand nombre d'éléments génétiques transposables, de séquences répétées, et des restes de prophages et bactériophages.

1.2 La structure de l'ADN

L'acide désoxyribonucléique, plus communément appelé ADN, est le matériel de l'hérédité chez chaque organisme vivant. Johann Friedrich Miescher fut le premier à isoler des composés riches en phosphates, qu'il appela « nuclein », à partir de globules blancs sanguins en 1869. Ce biologiste et physicien suisse est à l'origine de l'hypothèse que les « nuclein » (l'ADN) seraient impliqués dans l'hérédité, hypothèse confirmée ensuite par Avery *et. al* (10). Puis, en 1953, se basant sur des clichés de l'ADN obtenus par diffraction de rayons X par leur consœur Rosalind Franklin, le biochimiste Watson et le biologiste Crick proposent un modèle pour la structure de l'ADN. Ce modèle, représentant une double hélice antiparallèle (orientées parallèlement mais en sens opposé) est aujourd'hui unanimement reconnu comme une forme majeure de l'ADN, dite forme B (Figure 2A). L'information génétique y est stockée sous forme de code correspondant à la succession de 4 bases chimiques : Adénine (A), Guanine (G), Cytosine (C) et Thymines (T). La cytosine et la thymine sont des bases pyrimidines, c'est à dire qu'elles sont composées d'un seul cycle pyrimidine, contrairement à l'adénine et la guanine qui sont des purines car elles sont composées de deux cycles : un cycle pyrimidine et un cycle

imidazole (Figure 1). Chaque base est liée à un sucre (appelé déoxyribose) et à un groupement phosphate. L'ensemble de ces 3 éléments forme un nucléotide. L'ordre d'enchaînement de ces nucléotides détermine l'information fournie par la séquence ADN, à l'image d'une suite de lettres utilisées pour former des mots et des phrases.

L'ADN est une molécule composée de 2 brins qui s'enroulent l'un autour de l'autre, un brin matrice (ou transcrit) et un brin codant (ou non-transcrit), qui forment la double hélice. Le brin matrice présente une séquence complémentaire au brin codant et sera le brin lu lors de l'expression d'un gène. En face de chaque A se trouvera un T (et réciproquement), et en face de chaque C se trouvera un G (et réciproquement). Dans cette double hélice aux allures d'échelle, les bases forment des liaisons hydrogène entre nucléotides de chaque brin : 2 liaisons hydrogène entre A et T, et 3 liaisons hydrogène entre G et C (Figure 1). Les sucres et les phosphates permettent de lier de façon covalente les bases successives au sein d'un même brin. On peut assimiler cette molécule à un squelette sucre-phosphate sur lequel les bases sont fixées (11).

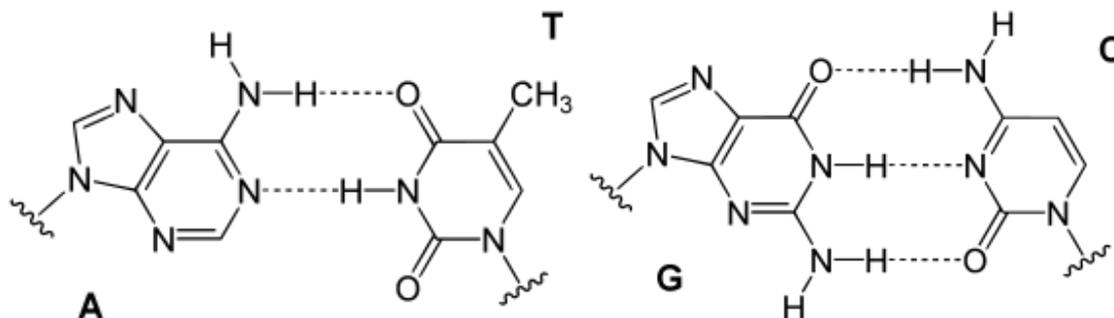


Figure 1 : Représentation de la structure des bases de l'ADN.

A gauche, appariement A : T (deux liaisons hydrogènes) ; A droite, appariement G : C (trois liaisons hydrogènes).

La façon dont les nucléotides sont alignés donne une polarité chimique à la molécule d'ADN. En assimilant le phosphate 5' à une protubérance qui vient se glisser dans l'enfoncement provoqué par le sucre en 3' on observe que tous les nucléotides d'un même brin sont co-orientés. Cette polarité du brin d'ADN est indiquée par les extrémités 5' et 3'. La structure tridimensionnelle de la molécule d'ADN est due aux propriétés chimiques et structurales de ses 2 chaînes polynucléotidiques. L'ADN peut se présenter sous plusieurs types de conformations, qui peuvent être influencés par nombreux facteurs tels que le niveau

d'hydratation, la concentration en sels, la séquence de la molécule d'ADN en question, les niveaux de surenroulements (positifs ou négatifs), la présence de bases modifiées, la présence et la concentration de certains métaux ioniques, et la présence de polyamines dans la solution.

Les types d'ADN les plus communs sont l'ADN de type A, l'ADN de type B et l'ADN Z.

La forme adoptée par l'ADN peut être déterminée par la séquence nucléotidique elle-même : une séquence purement composée de purines (G et A) va entraîner un ADN de type A, une séquence mixte sans réel schéma va entraîner un ADN de type B, et une séquence alternant purine et pyrimidine (ex. CGCGCG) va induire un ADN de type Z (Figure 2B) (12).

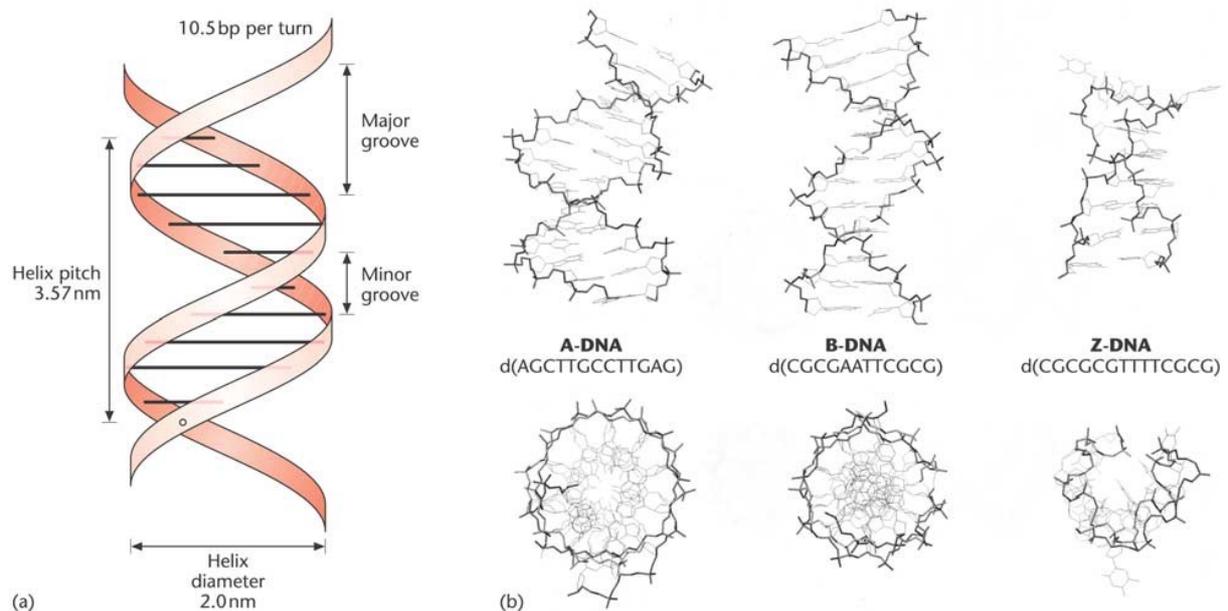


Figure 2 : Différentes représentations de la double hélice ADN (12).

a. Structure de l'ADN de type B proposée par Watson & Crick. b. ADN de type A, B et Z vus de côté (haut) et vus du dessous/dessus (bas).

1.2.1 ADN type B

L'ADN de type B (Figure 2) est la forme la plus commune sous laquelle l'ADN est retrouvé dans une cellule dans des conditions physiologiques naturelles (pH and concentration saline). Les paramètres géométriques de cette conformation ne sont pas stricts mais plutôt définis par un ensemble de conformations apparentées. L'ADN de type B est caractérisé par une double hélice droite. Les deux brins d'ADN étant liés par des liaisons hydrogènes entre les bases, celles-ci se retrouvent dans la partie interne de la double hélice, formant un cœur

hydrophobe, et les groupements phosphates ainsi que les sucres se retrouvent à l'extérieur et portent des charges négatives. Une pyrimidine (un seul cycle) se lie forcément à une purine (deux cycles) sur le brin opposé. Dans cette configuration chaque liaison entre brins est de taille similaire, et l'encastrement entre purine et pyrimidine donne une courbure aux brins d'ADN qui vont s'enrouler l'un autour de l'autre. Toutes les 10 paires de bases (pb) environ, la double hélice aura réalisé un tour complet, appelé « pas », d'une hauteur de 34 Angstrom (3,4 nanomètres). La distance entre 2 paires de bases adjacentes sur un brin est donc de 3,4 Angstrom. Le diamètre du double brin est de 20 Angstrom. La structure de la double hélice montre la présence d'un petit sillon et d'un grand sillon, de 12 et 22 Angstrom respectivement (Figure 2A). La plupart des protéines spécifiques vont interagir avec le grand sillon.

1.2.2 ADN de type A

L'ADN de type A (Figure 2b) est une structure conformationnelle rare sous laquelle on peut retrouver l'ADN dans les conditions de déshydratation. Très similaire à l'ADN de type B, avec une double hélice droite, elle se distingue par une structure plus compacte et courte, et un diamètre de l'hélice de 26 Angstrom. La hauteur de l'ADN de type A est 20 à 25% plus petite que celle de l'ADN B car la distance entre 2 bases est plus petite, environ 2,9 Angstrom. Ainsi, un tour d'hélice correspond à 11,6 pb et à une augmentation de taille de 28,6 Angstrom. Cette structure entraîne l'apparition d'un vide au cœur de la double hélice (Figure 2B).

1.2.3 ADN de type Z

Contrairement aux deux modèles vus précédemment, l'ADN de type Z (Figure 2b) présente une double hélice gauche, en forme de zigzag d'où le nom « ADN Z ». Cette conformation est retrouvée dans une séquence nucléotidique alternant purines et pyrimidines couplée à des fortes concentrations en sels. Le diamètre de la double hélice dans cette conformation est de 18 Angstrom, et la hauteur d'un tour d'hélice est de 44 Angstrom. Chaque tour d'hélice contient 12 nucléotides, et la distance séparant deux nucléotides est de 7,4 Angstrom. La fonction ou les avantages exacts de cette conformation ne sont pas entièrement connus. Ce type d'ADN étant généralement retrouvé en amont des sites de début de

transcription, une des hypothèses est qu'il serait impliqué dans la régulation de l'expression des gènes.

1.2.4 ADN dans un nucléoïde bactérien

La forme dans laquelle se trouve l'ADN au sein d'une cellule bactérienne correspond à un équilibre entre les différents types cités ci-dessus. Une même molécule d'ADN peut présenter des régions nucléotidiques sous différentes formes sous l'impact de stress environnementaux ou topologiques. Additionnement aux types majoritaires B, A et Z, la molécule d'ADN peut adopter d'autres formes. Lorsqu'elle est étirée et surenroulée positivement, elle peut adopter une forme appelée P, dans laquelle les bases se retrouvent à l'extérieur du squelette (13). Inversement, lorsqu'elle est étirée et surenroulée négativement, l'appariement se brise au niveau de certains nucléotides et se met en place une bulle de dénaturation qui fait apparaître une région constituée de deux simples brins d'ADN (13).

L'ADN bactérien est un ADN globalement de type B, contenu dans une structure macromoléculaire appelée nucléoïde. Le nucléoïde contient l'ADN compacté par une combinaison de facteurs, que sont l'encombrement moléculaire (14), la dynamique du polymère d'ADN (15), le surenroulement de l'ADN (16,17), et l'interaction de l'ADN avec d'autres molécules telles que les molécules architectes (18–20).

Cependant, cette compaction doit être réalisée sans entraver le bon déroulement des mécanismes essentiels à la vie de la bactérie. Une importante propriété de l'ADN est qu'il est capable de réplication. Les deux brins d'ADN peuvent se dissocier et chacun va pouvoir être utilisé en tant que matrice afin de synthétiser une molécule d'ADN double brin complète. Ce phénomène, appelé réplication semi-conservative, est très important lors de la division cellulaire car chaque cellule fille doit avoir une copie exacte de l'ADN de la cellule mère. Il faut que la molécule d'ADN soit malléable afin de garantir le bon déroulement de ce mécanisme, dont un ralentissement ou un arrêt peut conduire à la mort cellulaire. De plus, l'expression des gènes doit être facilitée afin d'assurer les fonctions cellulaires de base, et permettre à la bactérie de traiter rapidement les signaux environnementaux. Ainsi, la compaction a pour but de diminuer le volume occupé par l'ADN tout en respectant les mécanismes cellulaires essentiels, tels que la réplication du génome, la ségrégation des chromosomes, et la transcription.

2 – EXPRESSION GENIQUE BACTERIENNE

2 EXPRESSION GENIQUE BACTERIENNE

Tous les organismes, qu'ils soient pluricellulaires ou unicellulaires, macroscopiques ou microscopiques, sont extrêmement riches de complexité. A tel point qu'après des dizaines d'années d'études, à travers des centaines de laboratoires différents, par des milliers de scientifiques sur plusieurs générations, nous ne sommes toujours pas parvenus à décrypter l'ensemble des mécanismes agissant chez une seule des espèces qui nous entoure. L'expression génique fait partie de ces mécanismes très étudiés, et paraît comme simple aux premiers abords : un gène est transcrit en ARN messager (ARNm) par une enzyme appelée ARN polymérase, puis cet ARNm est à son tour traduit en protéine par un complexe macromoléculaire appelé ribosome. Non seulement le mécanisme d'action de l'ARN polymérase est abondant de complexité, mais il est également fortement régulé par des phénomènes de nature très variable, qui eux-mêmes sont régulés par d'autres effecteurs. Il faudrait donc considérer ce système comme un réseau constitué de dizaines d'éléments en interaction, dont la résultante permet de répondre rapidement à des besoins.

Comme stipulé précédemment, l'expression génique se produit en deux étapes : la transcription et la traduction. L'ARN polymérase va se lier à une séquence appelée « promoteur » qui se trouve en amont du gène à transcrire, à partir de laquelle elle va pouvoir initier la transcription. Certaines protéines sont d'usage routinier ou bien requises en grande quantité et donc sont constamment synthétisées, d'autres le sont en réponse à des signaux extérieurs. Les observations ont montré que les mécanismes de la transcription et de la traduction sont couplés au niveau spatial et au niveau temporel chez les bactéries (21). La demi-vie très courte des ARN messagers, de quelques minutes en moyenne (22), a un coût énergétique considérable, mais l'avantage de permettre de très rapides reconversions du métabolisme cellulaire.

Cinq mécanismes moléculaires distincts semblent assurer la distribution fine des molécules d'ARN polymérase entre les différents promoteurs en compétition : les facteurs σ , la séquence ADN, les facteurs de transcription, les ligands, et la conformation du chromosome de la bactérie.

2.1 Transcription

2.1.1 ARN polymérase

Il existe différents types d'ARN, dont les principaux sont les ARN ribosomiaux (ARNr), les ARN messagers (ARNm) et les ARN de transfert (ARNt). Chez les bactéries, tous les différents types d'ARN sont synthétisés par une seule enzyme, l'ARN polymérase. Cette enzyme est composée d'un « core-enzyme », constitué à partir de 5 sous-unités ($\alpha_2\beta\beta'\omega$), d'une masse moléculaire d'environ 400 kDa (23–25). Elle est catalytiquement active mais incapable de reconnaître les séquences promotrices spécifiques et de lancer l'initiation de la transcription. Le core-enzyme va s'associer à une sixième sous-unité appelée facteur sigma (σ). La sous-unité σ va se lier aux deux sous-unités β et β' grâce à des liaisons fortes (26). L'ensemble devient l'holoenzyme, qui va être capable de reconnaître une séquence spécifique et se lier à l'ADN (27). Les sous-unités du core-enzyme ont une séquence très fortement conservée chez les différentes bactéries et ont des homologues chez les eucaryotes et les archées (28,29). Cependant, les sous-unités σ ne sont présentes que chez les bactéries, et très diverses entre les différentes espèces bactériennes, que ce soit au niveau de la séquence, de la taille, ou de la structure (30–32). Les bactéries possèdent plusieurs facteurs σ , parmi eux un est essentiel et assure à lui seul la transcription de milliers de gènes : le facteur σ^{70} chez *E. coli* (analogue du facteur σ^A chez les autres bactéries). Les autres facteurs σ agissent généralement en réponse à des signaux. C'est le cas du facteur σ^{54} , qui diffère structurellement de la famille σ^{70} et qui nécessite la présence d'un activateur protéique mais aussi d'ATP pour engendrer l'initiation de la transcription (33). Il existe au total 7 facteurs σ différents chez *E. coli*. Tous ces facteurs σ sont en compétition pour s'associer avec le core-enzyme de l'ARN polymérase. Chez *E. coli* il y a au total ~2 000 molécules de core-enzyme de l'ARN polymérase, responsables de l'expression d'un peu plus de 4 000 gènes (9).

La transcription se déroule en 3 étapes principales : l'initiation, l'élongation et la terminaison.

2.1.1.1 Initiation de la transcription

Afin d'initier la transcription, l'ARN polymérase doit d'abord se lier à la séquence promotrice (Figure 3a). Cette étape nécessite la présence de la sous-unité σ , qui est responsable

de la reconnaissance de deux hexanucléotides très conservés en amont du gène en position -35 et -10 par rapport au site d'initiation de la transcription (+1) (30,34,35). La liaison de l'holoenzyme à la séquence ADN forme un « complexe fermé » (RPc, RNAPol Promoteur Closed complex), ce qui signifie que les deux brins d'ADN sont toujours appariés et que la transcription ne peut pas encore démarrer (Figure 3b). D'autres éléments présents au niveau de cette région promotrice peuvent faciliter ou limiter la liaison de l'ARN polymérase, ils sont détaillés dans la section 2.1.2 de ce manuscrit.

La séquence consensus de l'hexanucléotide en -10 est TATAAT. C'est le site d'initiation du « complexe ouvert » qui marque la séparation des deux brins d'ADN. En effet, cette région est riche en A/T et il y a seulement deux liaisons hydrogènes entre A et T (alors qu'il y en a 3 entre C et G) facilitant d'un point de vue énergétique l'ouverture du double brin à ce niveau. RPc, le complexe ARN polymérase/promoteur fermé, subit une série de changements conformationnels initiés par la liaison entre un domaine de la sous-unité σ et l'ADN au niveau du nucléotide -11 (36–38). Ces changements induisent la séparation des deux brins et l'apparition d'une « bulle » sur une longueur de 12 à 15 nucléotides au niveau de la boîte -10. C'est le début du « complexe ouvert » (RPo, RNAPol Promoteur Open complex) (Figure 3c). En présence de ribonucléosides (rNTP), bases de l'ARN, le complexe RNAPol/promoteur va pouvoir former les premières liaisons phosphodiester entre les rNTP positionnés en +1 et +2. Le complexe ouvert va alors tirer vers lui l'ADN en aval, jusqu'au nucléotide +15, entraînant une plus grande ouverture de la bulle, qui atteint environ 25 nucléotides (39,40). Simultanément, la synthèse de l'ARN naissant est obstruée par la présence de la liaison entre les sous-unités σ et β (41,42). Ce phénomène va causer des événements abortifs d'initiation, pendant lesquels l'ARNpol va synthétiser de très courts ARN sans se déplacer de la région promotrice (42–44). Une fois que l'ARNpol aura synthétisé un ARN d'environ 11-15 nucléotides, l'ARNpol pourra se détacher de l'ADN en se déliant de sa sous-unité σ .

Se met alors en place le complexe d'élongation (EC) très stable et débute la phase d'élongation de la synthèse d'ARN, dont la séquence est complémentaire au brin d'ADN transcrit (41).

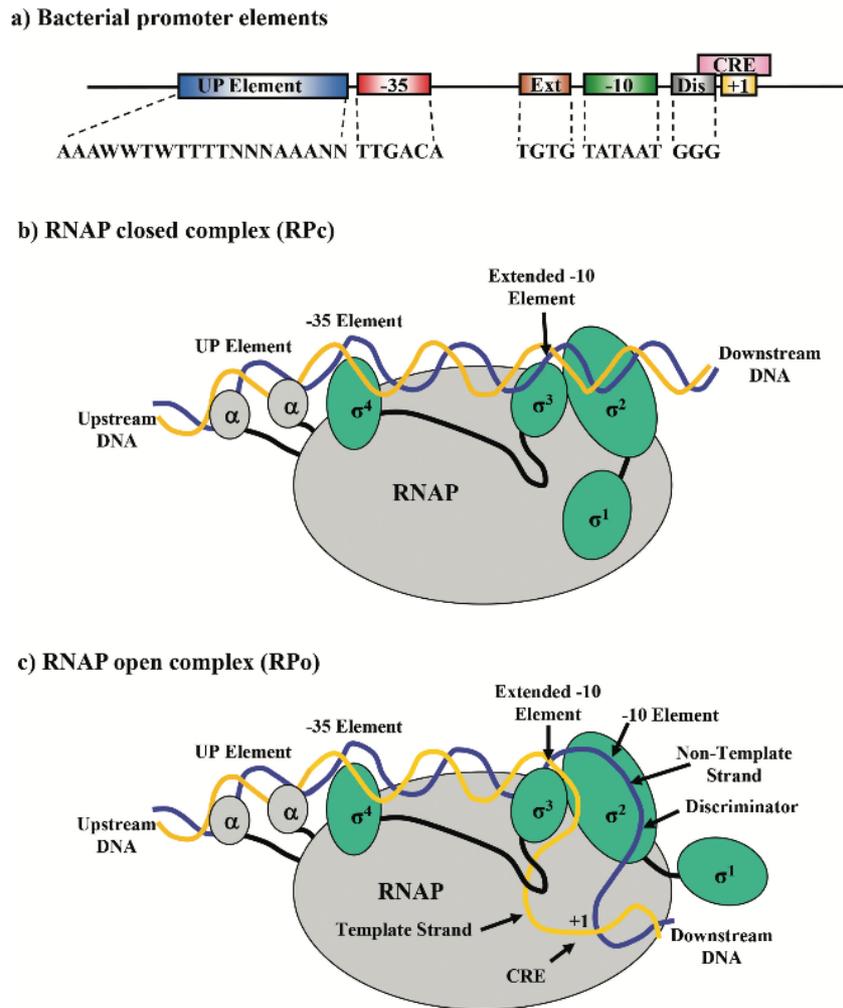


Figure 3 : Reconnaissance du promoteur bactérien durant l'initiation de la transcription (45).

a. Organisation des différents éléments du promoteur bactérien en fonction du site d'initiation de la transcription noté +1. Les différents éléments sont définis par des rectangles colorés et leur séquence consensus est indiquée. La structure du promoteur est détaillée dans la section 2.1.2 ci-après. b. Illustration des interactions entre les différents domaines de l'ARN polymérase et le promoteur pendant le complexe fermé. c. Illustration des interactions entre les différents domaines de l'ARN polymérase et le promoteur pendant le complexe ouvert. Le brin matrice est représenté en jaune.

2.1.1.2 Élongation

La structure globale de l'ARN polymérase est très similaire entre les phases d'initiation et d'élongation. Le détachement de la sous-unité σ va tout de même légèrement changer la conformation de l'ARN polymérase bactérienne, qui va alors avoir une forme de pince plus étroite qui va entourer l'ADN (46,47) et dont les 2 branches sont représentées par les sous-unités β et β' . Ces sous-unités servent de squelette auquel les autres sous-unités vont pouvoir

se fixer. Le vide laissé entre β et β' est divisé en une « chaîne principale » (ou primaire) qui accueille l'ADN double brin en aval et l'hybride ARN naissant/ADN, une « chaîne secondaire » qui sert d'entrée pour les rNTPs, et une « chaîne de sortie de l'ARN » qui est impliquée dans la séparation des hybrides ARN/ADN. L'ARN polymérase possède deux parties malléables qui jouent un rôle crucial dans son site actif : le « F-bridge » et le « G-loop ».

Le G-loop est très flexible et peut adopter plusieurs formes. Pendant la phase de complexe fermé RPC, le G-loop et le F-bridge sont responsables du recrutement et de la liaison du nucléotide complémentaire à l'ADN en +1. Le rôle du G-loop est notamment lié à la fidélité entre la séquence matrice et le rNTP recruté (48). Une fois le rNTP recruté et placé en face du nucléotide +1, les liens entre le rNTP et le G-loop sont rompus, ce dernier quitte le site catalytique et adopte une conformation relâchée (inactive) (47,49).

Durant l'élongation, le complexe RNAPol/ADN progresse comme un cliquet, en alternant les phases actives et les phases inactives. Dès qu'un rNTP correct est lié à la chaîne naissante, l'ARN polymérase avancera d'un cran. Cette avancée graduelle est permise par la flexibilité du F-bridge (46,50,51).

Différentes observations (52,53) réalisés sur les EC de levures suggèrent que l'ouverture du double brin se ferait au niveau des nucléotides entre +3 et +5. Cependant, des modèles proposés plus récemment suggèrent que l'ouverture se ferait entre les nucléotides +1 et +2, ne libérant ainsi que la base qui va être lue et complétée par l'ARN polymérase chez *T. Thermophilus* (47,54). Ce modèle serait appliqué à toutes les ARN polymérases avec sous-unités multiples.

Le taux d'erreur de l'ARN polymérase est compris entre 1×10^{-4} et 1×10^{-5} (55,56). Pour atteindre un niveau de fidélité aussi élevé, l'ARN polymérase est capable de reculer d'un cran afin de couper et remplacer le nucléoside non-désiré pendant l'élongation. Cela suggère que le modèle de cliquet présenté précédemment serait bidirectionnel (48). Globalement, à chaque instant durant l'élongation, le complexe d'élongation (EC) recouvre 13pb de l'ADN en aval, 9pb de l'hybride ARN/ADN, et 7 nucléosides de l'ARN naissant (47).

Le complexe d'élongation est capable de synthétiser des ARN de plusieurs milliers de nucléosides de long sans s'arrêter, de passer les possibles embuches rencontrées, mais devient extrêmement instable dès qu'il atteint la région d'un terminateur de transcription.

2.1.1.3 Terminaison

L'ARN polymérase peut marquer des pauses lorsqu'elle rencontre des embuches, notamment des structures d'ARN en épingle. Ces structures, lorsqu'elles ont une taille inférieure à 5pb, ne semblent pas affecter la liaison de l'ARN polymérase à l'ADN, cependant à partir de 8pb elles peuvent envahir le site actif du complexe d'élongation et entraîner des « pas » infructueux de l'ARN polymérase, qui ne peut continuer la synthèse du brin d'ARN. Cela va entraîner la dissociation du complexe, et l'ARN naissant va être libéré.

Cependant, pour une séparation de l'hybride ARN/ADN efficace, la présence du brin non-matrice d'ADN en amont semble être essentielle afin de directement reformer le double brin et d'empêcher une éventuelle reformation de l'hybride entre ARN et ADN (57,58).

D'autres facteurs, notamment en aval du site actif de l'ARN polymérase, impactent la terminaison de la transcription. Une séquence ADN riche en A/T au niveau de l'ADN entrant dans le site actif de l'ARN polymérase va favoriser la dissociation du complexe, tandis qu'une séquence riche en G/C va inhiber la terminaison (59).

Les effets dus à l'ensemble des facteurs en amont et en aval du site actif semblent entraîner un décalage de la bulle de transcription vers la partie en aval (60). Ce décalage est essentiel et permet la libération complète de l'ARN naissant (61). La taille de l'hybride ARN/ADN dans le complexe d'élongation diminuant, la stabilité de ce dernier est fortement altérée et l'ARN polymérase va se dissocier de l'ADN.

2.1.2 Promoteur

La sous-unité σ de l'holoenzyme ARN polymérase a majoritairement 3 fonctions : assurer la reconnaissance spécifique de la séquence promotrice, positionner l'ARN polymérase sur un promoteur cible, et faciliter l'ouverture du double brin d'ADN au niveau du site d'initiation de la transcription (TSS, Transcript Start Site) (27,62). Le facteur σ est une protéine à plusieurs domaines. Bien que la fonction du domaine I (absent chez certains facteurs σ) est peu comprise, les domaines II, III et IV sont connus pour leur implication dans la reconnaissance du promoteur (27,36,42,46).

Le promoteur est une séquence ADN placée en amont d'un gène et indispensable à la transcription de celui-ci. Chez les bactéries, les promoteurs contrôlent l'expression de tous les

gènes. Les différents promoteurs ont des séquences variables, mais présentent des régions conservées avec des séquences consensus. Ces séquences conservées se trouvent notamment au niveau des Boites -35 et -10 (par rapport au TSS), longtemps considérées comme les seuls éléments à respecter pour la reconnaissance par l'ARN polymérase. Plus récemment, de nouveaux facteurs influents sur le recrutement de l'holoenzyme ont été découverts : la taille et le contenu en A/T de la région entre les 2 boites appelée spacer, la présence d'UP-element (UPstream -35 element) en amont de la boite -35, et la nature des nucléotides entre la boite -10 et le TSS (figure 4).

Au total, l'ARN polymérase recouvre entre 60 et 90pb d'ADN au niveau de la région promotrice.

2.1.2.1 Boites -35 et -10

Les boites -35 et -10 se trouvent respectivement 35 et 10 nucléotides en amont du site d'initiation de la transcription. Ce sont 2 hexanucléotides dans la séquence est fortement conservée, 5'-TTGACA-3' pour la boite -35 et 5'-TATAAT-3' pour la boite -10 (Figure 4), respectivement reconnus par le domaine 4 et le domaine 2 du facteur σ (63,64).

Les nucléotides -35T, -34T, et -33G sont les plus conservés parmi la boite -35 (35). Cet hexanucléotide interagit, notamment par le biais du grand sillon de l'ADN, avec le motif hélice-boucle-hélice du domaine 4.2 de la sous unité σ ($\sigma_{4.2}$).

L'ouverture de la bulle de transcription s'opère au niveau de la boite -10, les observations suggèrent qu'elle interviendrait au niveau de la base -12T par le biais d'une liaison à la glutamine 437 du facteur $\sigma_{2.4}$ (65,66). Les deux brins, transcrit et non-transcrit, interagiraient alors avec l'holoenzyme (67).

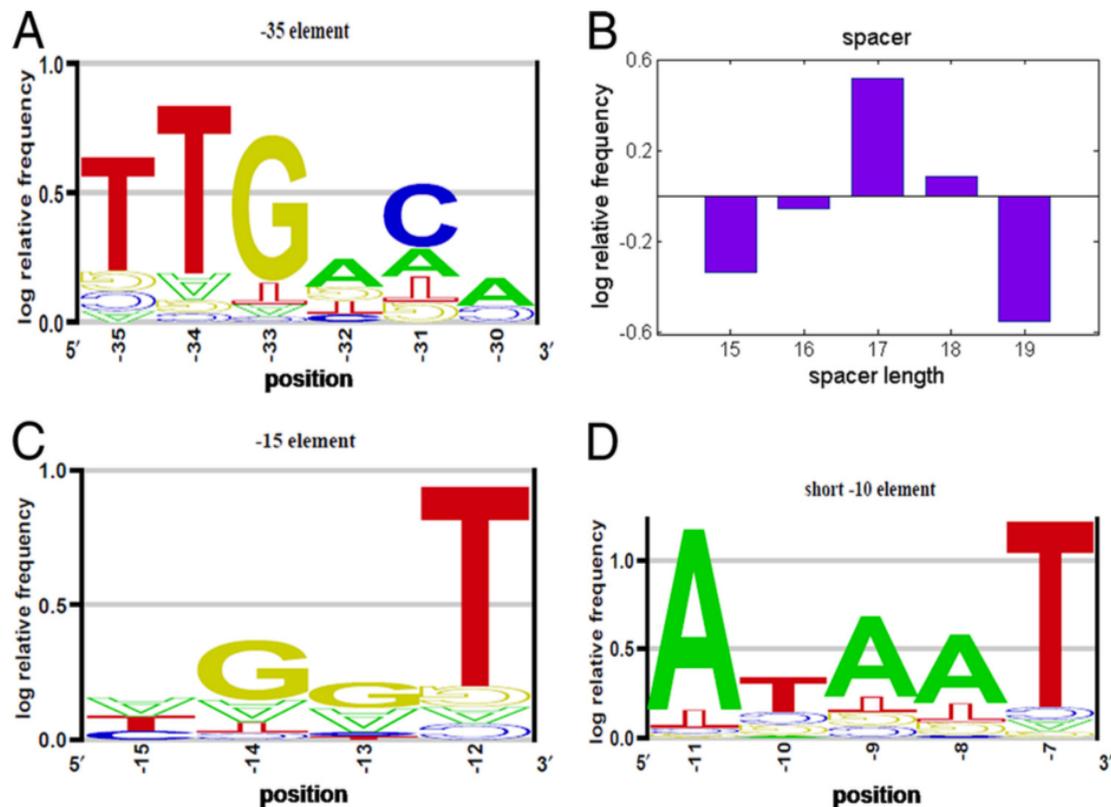


Figure 4 : Specificité des éléments du promoteur (68).

Logos-séquences correspondant aux spécificités des éléments -35 (A) et -10 (C, D). La taille des bases est proportionnelle au logarithme du ratio de la fréquence d'observation des bases dans les alignements de séquence. Pour les bases qui sont sous-représentées, le log-ratio prend une valeur négative, la base est représentée à l'envers. B. Pour le spacer, le logarithme de la fréquence des tailles de spacer observées est représenté.

2.1.2.2 Spacer

Les boîtes -35 et -10 sont séparées par une région ADN non conservée appelée « spacer » (Figure 5). En général, la taille de cet élément est bien plus importante que sa séquence. Chez *E. coli*, la taille optimale du spacer pour la liaison de l'holoenzyme σ^{70} est de 17 paires de bases. Une augmentation ou une diminution de la taille de ce spacer va entraîner une diminution de la force du promoteur (69,70). Modifier la taille du spacer va altérer la disposition rotationnelle des boîtes -35 et -10, et cela pourrait avoir un impact sur l'ouverture du double brin lors du passage au complexe ouvert (RPo) (71,72).

Bien que la séquence du spacer soit souvent considérée comme peu importante, elle peut contenir certains éléments impactant le recrutement de l'ARN polymérase. En effet,

notamment chez *B. subtilis* et dans une moindre mesure chez *E. coli*, une séquence placée en -15 appelée « extended -10 » (en rapport à un allongement de la boîte -10) dont le motif est TnTGn suivie de la séquence consensus de la boîte -10 va augmenter l'affinité de l'ARN polymérase en liant le domaine 3 de la sous-unité σ (36,73–75). Chez *E. coli*, l'extended -10 est souvent associé à une famille de promoteurs chez qui la boîte -35 est absente. Des expériences réalisées avec des mutants de l'extended -10 ont montré l'importance de cette séquence dans la formation du complexe ouvert (RPo) (76,77).

Malgré le fait qu'aucune séquence consensus n'ai été trouvée pour le spacer, les observations récentes montrent qu'une séquence riche en A/T favorise la transcription en permettant à l'ADN d'être plus flexible (78).

Enfin, une étude récente a permis d'identifier un nouvel élément dans le spacer, appelé « Z-element », qui serait impliqué dans l'augmentation des cycles de transcription abortifs. Bien que ce Z-element interagisse avec la sous-unité β de l'ARN polymérase, aucune preuve tangible ne montre que cette interaction soit spécifique et elle pourrait être simplement due à la conformation locale du spacer (79).

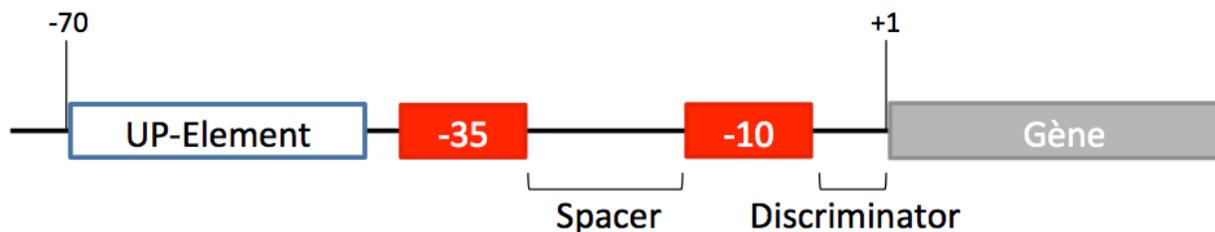


Figure 5 : Composants d'un promoteur bactérien.

Les différents éléments sont représentés dans la figure. Le +1 correspond au TSS.

2.1.2.3 UP-Element

Le UP-Element correspond à une séquence riche en A/T, d'environ 20pb, qui se trouve en amont du promoteur, entre la boîte -35 et le nucléotide -70 (figure 5). Cette séquence, reconnue par les domaines C-terminaux des sous-unités α de l'ARN polymérase (α CTD I et α CTD 2), va influencer la force de liaison de l'ARN polymérase au promoteur en fonction de la courbure physique qu'elle entraîne dans le polymère. Chacun des α CTD va se lier à l'ADN au niveau des sillons mineurs via un motif hélice-boucle-hélice. Alors que α CTD I se lie dans la partie de l'UP-element la plus proximale du -35 (80,81), α CTD II semble plus mobile et capable

d'occuper plusieurs sites plus en amont sur la même face du double brin (82–85). Ainsi, l'UP-element peut fortement induire la transcription en absence de protéines accessoires liées à l'ADN sur cette séquence. Pour le promoteur P1 du gène *rrnB*, la présence du UP-element augmente les niveaux de transcription de 30 à 70 fois en se liant à la sous unité α de l'ARN polymérase (86–89). Cependant, les mutants de l'ARN polymérase à laquelle ont été enlevé les sites prédictifs de liaison de l'UP-élément (domaine C-terminal de la sous-unité α) ont montré des niveaux d'expression normaux (87).

Estrem et ses collaborateurs (86) ont identifié une séquence consensus pour l'UP-element en sélectionnant *in vitro* les séquences permettant une liaison rapide de l'ARN polymérase au promoteur, suivi par un criblage *in vivo* des promoteurs à forte activité. Cette technique a révélé qu'une séquence alternant successions de A et de T peut induire une augmentation de plus de 300 fois de l'activité transcriptionnelle comparée à un promoteur en absence de UP-element. Une succession de A (entre 4 et 6 nucléotides) dans cette région entraîne une courbure de l'ADN (90), une succession T entraîne une courbure inversement complémentaire. Ce genre de courbures se sont avérées *in vitro* et *in vivo* favorisant la transcription (91), notamment en augmentant les contacts avec les sous-unités α et α_2 .

2.1.2.4 Région entre la boîte -10 et le TSS

La région entre la boîte -10 et le TSS, appelée « discriminator », est une région de 6 à 8pb impliquée dans le maintien du complexe ouvert RPo (92). Le nucléotide en -5 sur le brin non-transcrit est très important pour conserver le complexe ouvert, ainsi un G à cette position va stabiliser l'ouverture du complexe alors qu'un C va entraîner une diminution de la demi-vie du RPo (93). Par exemple, les promoteurs des ARN ribosomiques ont généralement un C deux nucléotides en aval de la boîte -10, entraînant une courte demi-vie du RPo. Cet element est important pour leur régulation *in vitro* et *in vivo* (93).

Le site d'initiation de la transcription correspond au premier nucléoside qui va se trouver sur l'ARN naissant. Il est défini par la valeur +1. Les bases les plus fréquemment retrouvées au niveau du TSS sont A puis G, suivi de T et dans une bien moindre mesure C (94). Dans les promoteurs ribosomiaux, la nature du résidu start est essentielle pour avoir un très fort niveau de transcription (95).

Tous les éléments précédemment cités ont leur importance dans l'efficacité d'un promoteur. Cependant, tous les critères favorisant la transcription ne se retrouvent pas au sein d'un même promoteur, car sinon la liaison chimique de l'ARN polymérase et de la région promotrice serait trop forte, donc peu réversible, et cela serait néfaste pour la dynamique régulatoire de la cellule. Il faudrait plutôt penser le promoteur comme un équilibre entre les différentes tendances. Dans ces conditions, l'absence d'un élément peut alors être compensé par un autre, pour assurer la fonction désirée.

2.1.3 Terminateurs de la transcription

La terminaison de la transcription est un mécanisme essentiel qui (i) empêche l'expression non désirée de gènes en aval de la séquence transcrite, (ii) évite les conflits avec un gène situé en aval et transcrit dans l'autre sens, (iii) façonne la structure en 3' de l'ARN naissant, (iv) permet un recyclage de l'ARN polymérase pour une expression génique efficace, et (v) diminue le risque de collisions entre les machineries transcriptionnelles et répliquatives. Le principal enjeu de la terminaison de la transcription est de suffisamment déstabiliser le complexe d'élongation, d'ordinaire très stable, à des sites génomiques bien précis afin d'entraîner la dissociation du complexe ARN polymérase-ADN. Cependant, la terminaison peut avoir lieu seulement si l'ARN polymérase entre dans le processus de terminaison plus rapidement que le mouvement de la matrice lorsqu'un nucléotide est ajouté. Il faut que l'ARN polymérase marque une pause et ne fasse pas de mouvement sur la matrice que ce soit vers l'avant ou vers l'arrière pendant un laps de temps assez long pour que la terminaison se mette en place. Une terminaison efficace peut alors être induite par des facteurs qui abaissent la vitesse d'élongation. Il y a deux principaux mécanismes de terminaison de la transcription chez les bactéries : la terminaison intrinsèque, impliquant seulement l'ARN polymérase, l'ADN ou l'ARN ; et la terminaison dépendante de facteurs extérieurs, qui implique des protéines qui vont dissocier le complexe d'élongation (EC), telles que Rho ou Mfd (96).

2.1.3.1 Terminaison intrinsèque de la transcription

L'ARN polymérase est capable de marquer des pauses réversibles, allant de quelques secondes à quelques minutes, lorsqu'elle rencontre certaines séquences d'ADN (97,98). Ces pauses pourraient altérer la conformation du complexe d'élongation, mais elles marquent surtout des fenêtres de temps pendant lesquelles la terminaison peut être enclenchée (60,99). La demi-vie des pauses du complexe d'élongation peut être augmentée par la formation de structures secondaires en forme d'épingle dans l'ARN naissant ou par un pas en arrière de l'ARN polymérase (phénomène de « backtracking », pour corriger une erreur). Cependant, l'idée que le pas en arrière de l'ARN polymérase soit un mécanisme de terminaison est peu probable. En effet, ce mécanisme inhiberait une éventuelle formation de structure secondaire en épingle. En réalité, la formation d'une épingle permettrait d'empêcher un éventuel « backtracking » de l'ARN polymérase (100,101).

Les structures secondaires en forme d'épingle sont appelées terminateurs de la transcription (Figure 6). Elles se forment lorsque des séquences complémentaires sont présentes dans le brin d'ARN naissant. Une portion de cet ARN va présenter un appariement de bases complémentaires donnant une région double brin, qui va bloquer la chaîne de sortie de l'ARN polymérase (102,103). Les terminateurs de transcription codent également pour un hybride ARN/ADN de faible stabilité, notamment grâce à une séquence poly-Uracile (99). Cette séquence pourrait causer une pause de l'ARN polymérase tout en déstabilisant le complexe d'élongation, entraînant ainsi la dissociation du complexe de transcription.

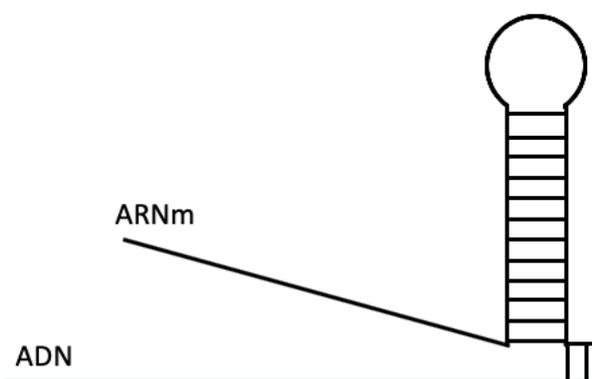


Figure 6 : Représentation d'un ARNm naissant contenant un terminateur de transcription.

L'ARNm est représenté en noir, le terminateur correspond à la tête d'épingle. L'ADN est représenté en bleu. Le complexe de transcription n'est pas représenté volontairement.

2.1.3.2 Terminaison dépendante de facteurs extérieurs

Les facteurs de transcription sont des protéines impliquées dans la régulation de la transcription (voir section 2.2.1.1). Certains peuvent réguler la terminaison de la transcription, notamment en se liant au complexe d'élongation. Ils vont affecter le taux d'élongation, la fréquence et la demi-vie des pauses, la structure de l'ARN naissant, ou la stabilité du complexe d'élongation. Par exemple, chez *Bacillus Subtilis* la protéine NusG est capable de stimuler les pauses liées à la formation de structures secondaires dans l'ARN (104,105), et chez *Mycobacterium bovis* cette même protéine est capable d'induire la terminaison extrinsèque (106). Chez *E. coli*, la protéine NusG induit la transcription en promouvant l'élongation (107). De plus, un paralogue de NusG, la protéine RfaH, est capable de se lier à l'ARN polymérase pour stabiliser le complexe d'élongation, en inhibant l'apparition de pauses et donc la terminaison (108–110). Cependant, des protéines induisant la terminaison de la transcription ont également été caractérisées chez *E. coli*. Ainsi, la protéine NusA est capable de stimuler les pauses et la terminaison du complexe d'élongation en favorisant la formation et la stabilisation de structures secondaires dans l'ARN (110–112).

La terminaison de la transcription est également possible par le facteur Rho. Ce mécanisme repose sur la capacité hélicase (séparation des deux brins de l'ADN) de cette protéine qui va se lier à l'ARN et entraîner des translocations en hydrolysant de l'ATP, jusqu'à induire une pause puis une dissociation du complexe d'élongation (113–115). Rho va d'abord reconnaître et lier l'ARN naissant, ce qui va entraîner une pause du complexe d'élongation, et ensuite agir comme un moteur qui tire l'ARN. La dissociation du complexe a lieu une fois que Rho a réussi à transloquer en aval du complexe tout l'ARN naissant par la chaîne de sortie.

L'étape finale de la terminaison de la transcription est la dissociation de l'ARN polymérase, de l'ADN et de l'ARN. La grande taille du terminateur entraîne la dissociation d'une partie de l'hybride ADN/ARN. Lorsque cet hybride a une taille de 5pb ou moins, il déstabilise entièrement l'EC (101). Cette étape marque la dissolution du complexe d'élongation.

2.2 Régulations

L'étude de la régulation de l'expression génique chez les bactéries a pris un tournant avec les résultats avant-gardistes des recherches sur l'opéron *lac* menées chez *E. coli* (116–119). Depuis, une longue liste de mécanismes impliqués dans la régulation de l'expression des gènes a été dressée. Il est maintenant clair que l'expression d'un gène ou d'un opéron est contrôlée non pas par un seul mais par une combinaison de divers mécanismes, agissant à différents niveaux (transcription, traduction) et à différentes étapes (initiation, élongation, terminaison) de l'expression génique. De plus, l'impact de l'élément régulateur peut varier en fonction de sa cible spécifique, de sa liaison avec de potentiels effecteurs ou protéines, ou encore de son interaction avec d'autres processus régulateurs. Cela rajoute un niveau de complexité supérieur aux systèmes de régulations bactériens. Une grande partie des recherches sur ces mécanismes ont été réalisées chez *E. coli*, cependant, on peut supposer que les principes directeurs de tous ces processus sont valables pour une grande partie des bactéries, qu'elles soient Gram négatifs ou Gram positifs.

Cette étude se focalise sur les régulations qui ont lieu avant et pendant la transcription, ainsi que sur les régulations post-transcriptionnelles des ARN messagers avant l'étape de traduction.

2.2.1 Régulations transcriptionnelles

La transcription est un phénomène fortement sujet à régulation. Les études réalisées dans le domaine ont montré que certains phénomènes permettent d'activer la transcription et d'autres permettent de l'inhiber. Cependant, il existe des mécanismes globaux encore non élucidés qui régissent la régulation des gènes à l'échelle génomique (120). Les observations faites sur les différents mécanismes chez les eucaryotes montrent que la non-expression d'un gène est plus souvent liée à l'absence d'induction de transcription plutôt qu'à une réelle inhibition par une ou plusieurs molécules. Chez les bactéries, la situation est à l'opposée : la non-expression résulte généralement d'une active inhibition.

L'évolution des facteurs de transcription bactériens, familles de protéines capables d'induire et réprimer la transcription, a progressivement permis de lier les signaux environnementaux à la régulation de la transcription. L'activité de chaque facteur de transcription est liée directement ou indirectement à un signal spécifique. Sachant que les bactéries doivent répondre à une grande variété de stimuli, il n'est pas surprenant de voir que les promoteurs géniques sont

régulés par un grand nombre de signaux. Le rôle des promoteurs est donc d'intégrer l'ensemble des signaux activateurs et répresseurs en entrée, pour engendrer un seul niveau de sortie, exprimé comme le nombre de transcrits produits par unité de temps. Les facteurs de transcription peuvent être classifiés en fonction de leur mécanisme d'action, de leur fonction régulatrice ou de leur structure. Cependant, différencier les régulateurs « locaux » (facteurs de transcription qui ciblent spécifiquement des gènes ou opérons) des régulateurs « globaux » (facteurs de transcription qui régulent globalement l'expression génique) est complexe, car le seul nombre de gènes régulés est une caractéristique insuffisante pour différencier les 2 classes. Il faut aussi prendre en compte : (i) le nombre et la nature des facteurs de transcription qui co-régulent, (ii) la capacité de régulation de gènes appartenant à des groupes ciblés par d'autres sous-unité σ (voir ci-dessous), (iii) la capacité de régulation de gènes appartenant à diverses catégories de fonction, (iv) le potentiel de réponse à une grande variété de signaux environnementaux (121).

2.2.1.1 Régulation par un facteur de transcription

Un des scénarii les plus simplistes de l'activation de la transcription est la liaison d'une molécule, appelée facteur de transcription, à une séquence cible sur un promoteur reconnu par l'ARN polymérase contenant la sous-unité σ^{70} . *E. coli* compte plus de 300 gènes qui contiennent des domaines de liaison à l'ADN et qui peuvent induire ou réprimer la transcription (122,123). La plupart des activateurs (de transcription) agissent en stabilisant le complexe ARN polymérase/promoteur ou en accélérant le passage du complexe fermé vers le complexe ouvert (RPc \rightarrow RPo). En principe, il n'y a que deux façons de stimuler le passage vers un complexe ouvert par un facteur de transcription : en modifiant la conformation de la région promotrice qui sera plus facilement reconnaissable par l'holoenzyme, ou en agissant directement sur l'ARN polymérase pour compenser les éventuelles lacunes d'un promoteur.

2.2.1.1.1 Activation en modifiant la conformation de l'ADN

Un des exemples les mieux étudiés concerne la famille des facteurs de transcription MerR. Ces protéines vont lier l'ADN entre les boîtes -35 et -10 des promoteurs ciblés (124).

Cette famille de facteurs de transcription est impliquée dans la résistance au mercure. Les promoteurs cibles de ces facteurs de transcription possèdent un spacer d'une taille non-optimale. Ainsi, lors de la liaison de l'ARN polymérase sur l'UP-element et la boîte -35, la boîte -10 va se retrouver mal placée et l'interaction avec le domaine $\sigma_{2.4}$ ne sera pas optimale, ne permettant pas la formation du complexe ouvert. La liaison de facteurs de transcription de la famille MerR va permettre une courbure au niveau de l'ADN, alignant les boîtes -35 et -10, permettant d'initier la transcription (125,126) (Figure 7).

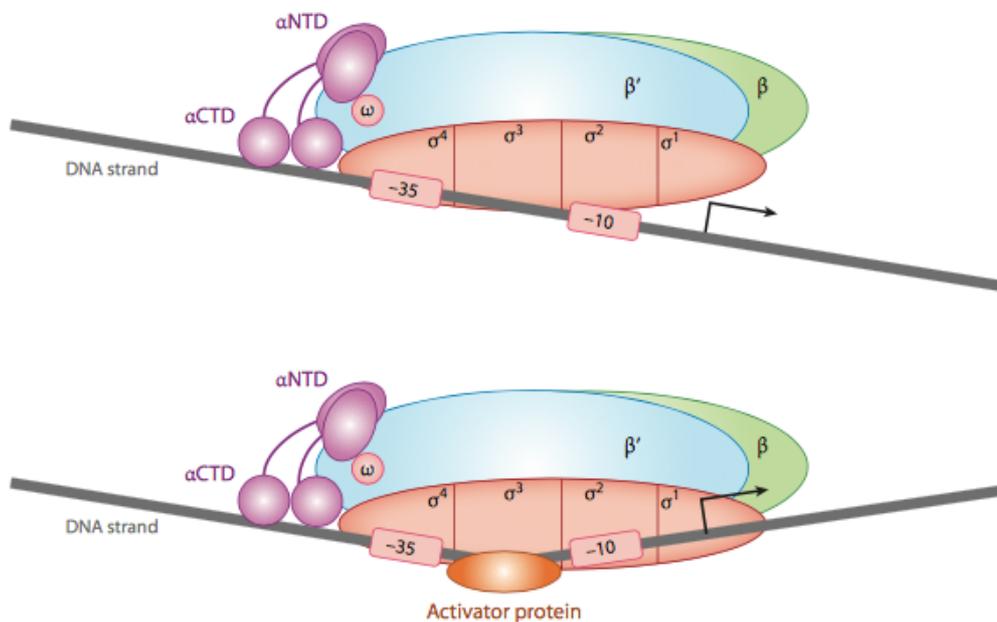


Figure 7 : Activation de la transcription par un changement conformationnel de l'ADN (127).

Le facteur de transcription se lie entre les boîtes -35 et -10 et induit un changement de conformation de la molécule d'ADN, permettant à l'ARN polymérase de mieux interagir avec le promoteur. CTD = domaine C-terminal ; NTD : domaine N-terminal.

Le fait que la modification de la conformation d'un promoteur puisse altérer ou renforcer sa capacité à induire la transcription suggère la possibilité qu'une régulation de l'expression est possible par la superhélicité sans l'intervention directe d'un facteur de transcription. Chez *E. coli*, il semble que le surenroulement (voir section 2.5) joue un rôle global dans l'activité des promoteurs, sans pour autant être directement responsable de l'activation de la transcription au niveau de certains promoteurs en réponse à des signaux (128). Les récentes études ont montré à quel point la liaison spécifique de protéines au niveau des promoteurs pouvait influencer sur la topologie locale de la région promotrice et affecter la transcription des gènes (129–132).

Ainsi, Sheridan et ses collaborateurs ont montré que la liaison de la protéine IHF en amont du promoteur *ilvG* chez *E. coli* influence la conformation des séquences voisines et induit l'activation du promoteur (131). C'est également le cas de la protéine Fis, qui, en se liant au niveau du promoteur *tyrT*, va empêcher la diffusion des contraintes topologiques, ce qui entraîne la formation du complexe de transcription (132).

2.2.1.1.2 Activation par contact direct avec l'ARN polymérase

Les études génétiques et biochimiques sur des promoteurs bactériens ont montré qu'un contact direct entre l'activateur et l'ARN polymérase est essentiel pour activer la transcription au niveau de certaines séquences promotrices. Dans la majorité des promoteurs dépendants de la liaison de l'activateur, ce dernier va se lier en amont ou alors dans le promoteur, dans quel cas il pourra établir un contact direct avec l'ARN polymérase (133–135). Les structures des EC pour les promoteurs dépendants d'activateurs sont les mêmes que celles pour les promoteurs non-dépendants d'activateurs. Ces résultats suggèrent que la liaison ADN/ARN polymérase qui permet le recrutement de l'holoenzyme serait simplement remplacée par la liaison activateur/ARN polymérase (134,136). Cette hypothèse est appuyée par les contrôles positifs mettant en jeu des mutants des activateurs au niveau d'une seule base, stoppant ou réduisant fortement la transcription des gènes ciblés par ces activateurs (133,134). Enfin, il a été démontré que la liaison de ces protéines n'altère pas la structure de l'holoenzyme (137,138), suggérant que n'importe quel phénomène entraînant le recrutement de l'ARN polymérase permettrait d'induire la transcription au niveau de ces promoteurs.

2.2.1.1.3 Répression

Dans un grand nombre de cas, la répression est un phénomène simple qui implique la fixation d'une protéine en aval du site de fixation de l'ARN polymérase (effet de gêne stérique) au niveau du promoteur. Trois mécanismes généraux sont possibles (Figure 8). Parmi ceux-ci, l'occupation stérique est probablement le mécanisme de répression le plus commun. Dans ce phénomène, le répresseur se lie dans ou proche du promoteur, empêchant soit le recrutement de l'ARN polymérase, soit les premières étapes de l'initiation (139). Un deuxième mécanisme

de répression est possible par la mise en place d'une boucle d'ADN, permise par la liaison à des sites distants répresseurs bivalents capables de lier l'ADN à 2 sites. C'est le cas des protéines GalR (140) et LacI (119) (voir section 2.4.3.1). La formation de la boucle va rendre inaccessible le promoteur et l'ARN polymérase ne pourra s'y lier. Enfin, un autre mécanisme par lequel la répression peut avoir lieu met en jeu un répresseur qui va reconnaître et inactiver l'action d'un activateur. Ce mécanisme correspond donc à une anti-activation. C'est le cas du facteur de transcription CytR qui va reconnaître et interagir avec CRP au niveau des promoteurs et empêcher CRP de jouer son rôle d'activateur de transcription (141).

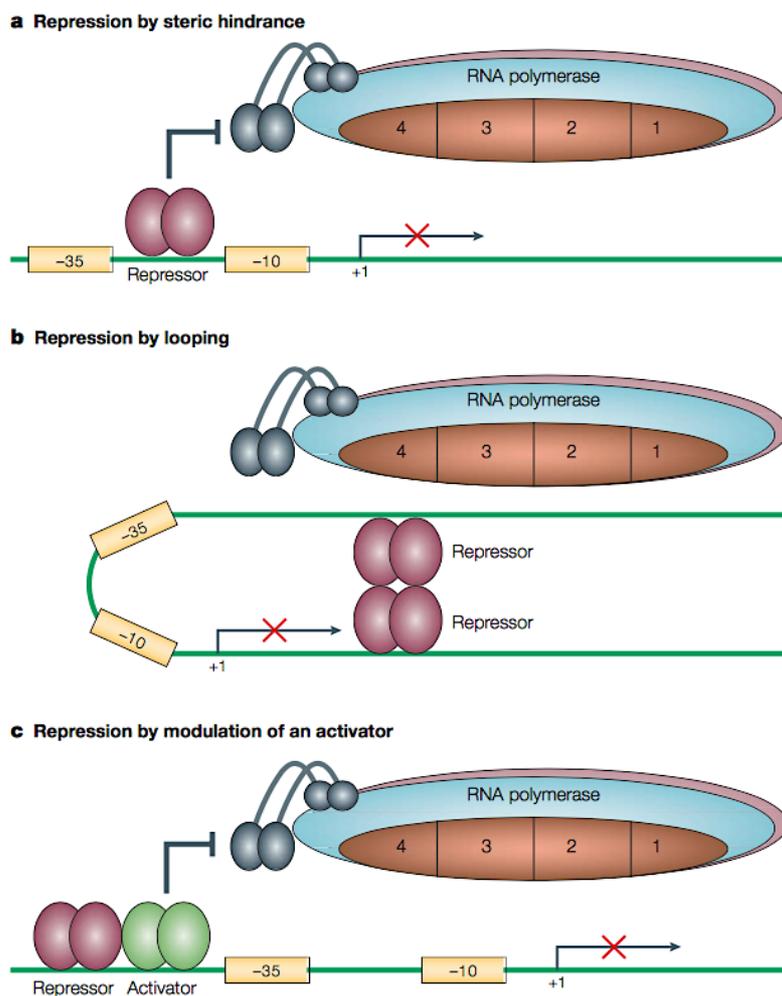


Figure 8 : Mécanismes de répression de l'initiation de la transcription (136).

a. Répression par encombrement stérique. Le site de liaison du répresseur chevauche les éléments du promoteur et empêche la reconnaissance du promoteur par l'ARN polymérase. b. Répression par la formation d'une boucle. Le répresseur lie des sites distants et forme une boucle emprisonnant le promoteur. c. Répression par inactivation d'un activateur. Le répresseur va se lier à l'activateur et l'empêcher de fonctionner, empêchant la reconnaissance par l'ARN polymérase.

2.2.1.1.4 Exemple : la protéine CRP

Le facteur de transcription le mieux étudié chez *E. coli* est sans doute CRP (cAMP receptor protein, également appelée CAP, Catabolite Activator Protein). Cette protéine va interagir avec la sous-unité α de l'ARN polymérase pour généralement stimuler la transcription du gène ciblé (142,143). Cependant, CRP peut aussi agir en tant que répresseur, coactivateur ou corépresseur (144). Cette protéine est continuellement présente au sein de la cellule, mais elle est sous forme inactive lorsque le glucose est la source principale de carbone dans le milieu de culture. La fixation d'une molécule de AMP cyclique (AMPc) avec une liaison forte au niveau du domaine N-terminal de CRP va entraîner un changement dans la conformation de la protéine et exposer le domaine C-terminal de celle-ci (145). Ce domaine contient un motif hélice-boucle-hélice de liaison à l'ADN, qui reconnaît la séquence palindromique TGTGAnnnnnnTCACA appelée CRP-box. L'AMP cyclique est synthétisée par l'adenylate cyclase, une enzyme codée par le gène *cyaA* dont l'expression est activée en absence de glucose. Lorsque le flux de glucose dans la cellule ralentit ou s'arrête, le niveau intracellulaire d'AMPc augmente suffisamment pour déclencher la liaison de CRP à l'ADN. Une fois CRP lié à la région promotrice, CRP va recruter l'ARN polymérase et la transcription va démarrer. Approximativement 200 unités transcriptionnelles ont été identifiées expérimentalement comme étant régulées par CRP, et plus 400 sites de liaison ont été informatiquement prédits (146). A l'heure actuelle, les observations font de CRP le régulateur global de la transcription le plus important chez *E. coli*.

2.2.1.2 Promoteurs dépendant du facteur σ^{54}

Contrairement à la transcription réalisée par l'holoenzyme σ^{70} , la RNA polymérase associée à la sous-unité σ^{54} est incapable d'initier la transcription sans la présence d'un activateur. De façon intrigante, cette sous-unité mal connue est structurellement différente des facteurs σ de la famille σ^{70} (147). Son origine dans l'évolution est cependant ancienne car environ 60% des bactéries connues à ce jour portent des gènes dépendants de σ^{54} . Les éléments clés d'un promoteur reconnus par σ^{54} se trouvent au niveau des positions -12 et -24. L'ARN polymérase reconnaît ce type de promoteur, mais a besoin de la présence d'un activateur de transcription pour former un le complexe ouvert RPo (148). Dans la majorité des

cas, ces activateurs répondent à des signaux métaboliques et se lient à une centaine de paires de bases en amont du promoteur régulé, mais ils peuvent également se lier jusqu'à 1 000pb en amont du gène et agir sur la transcription de ce dernier par le biais de boucles d'ADN (voir section 2.4.3).

2.2.1.3 Promoteurs dépendants d'autres facteurs σ

D'autres facteurs σ , appelés alternatifs, sont utilisés pour reprogrammer l'ARN polymérase afin de modifier ses cibles. Par exemple, sous l'influence de certains signaux spécifiques, un facteur σ alternatif va lier l'ARN polymérase au lieu de σ^{70} et ainsi changer la séquence promotrice recherchée par l'holoenzyme (30). Ces facteurs σ alternatifs sont régulés par le biais de leur disponibilité dans la cellule, c'est pourquoi les promoteurs cibles ont tendance à ne pas avoir besoin d'activateurs pour lancer la transcription.

2.2.1.4 Activation indirecte

Le phénomène le plus répandu d'activation indirecte de la transcription est l'arrêt de la répression. Par exemple, la liaison d'une protéine sur une séquence proche de la région promotrice peut entraîner une augmentation de la transcription non pas en recrutant l'ARN polymérase, mais en permettant la dissociation d'éléments (protéines) qui inhibent le promoteur (18). Bien que ces protéines soient assimilées à des activateurs, elles jouent en réalité le rôle d'anti-répresseurs.

2.2.1.5 Régulation par les petits ligands

La liaison de petits ligands est un mécanisme alternatif permettant une réponse aux changements environnementaux par l'ARN polymérase. L'exemple le plus parlant est la Guanosine 3',5' biphosphate (ppGpp), qui est synthétisée lorsque la disponibilité des acides aminés est trop faible pour la traduction (149). Le ppGpp va déstabiliser le complexe ouvert RPo au niveau des promoteurs responsables de la synthèse de la machinerie de traduction (150,151). Ainsi, ce petit ligand va diminuer drastiquement la synthèse de ribosomes et limiter

la traduction de tous les ARN messagers de la cellule bactérienne. Un tel mécanisme permet de limiter globalement la traduction dans la cellule lorsque la bactérie est en carence d'acides aminés.

D'autres petits ligands peuvent réguler la transcription, non pas en se liant à l'ADN ou à l'ARN polymérase, mais en se liant à un facteur de transcription. Par exemple, le répresseur lactose est un facteur de transcription impliqué dans la régulation de l'opéron lactose. La liaison du petit ligand, l'allolactose, au répresseur lactose va réduire l'affinité de liaison à l'ADN de ce dernier, levant la répression exercée par le répresseur lactose sur l'opéron lactose (152).

2.2.1.6 Structure du chromosome et protéines associées

Les chromosomes bactériens sont fortement compactés par la superhélicité et des interactions avec ARN et des protéines. Ces protéines, appelées NAP (Nucleoid Associated Proteins), sont abondantes dans la cellule mais leur concentration peut varier en fonction de la phase de croissance ou de signaux environnementaux. En se liant à l'ADN elles participent à la compaction du chromosome (voir section 2.4). La liaison de différentes NAP et la variation de leur concentration en fonction des différentes conditions de croissance vont moduler le chromosome et dynamiquement lui conférer une compaction. Les différentes conformations adoptées par l'ADN vont résulter en une distribution différente des ARN polymérases et donc une variation de l'expression de certains gènes. Toutefois, les recherches réalisées se focalisent uniquement sur l'étude de chaque NAP de manière isolée. Certaines revues dressent un modèle plus global de la régulation génique par l'action des NAP, cependant, aucune étude n'a permis d'établir des règles générales qui s'appliquent au niveau chromosomique.

2.2.2 Régulations post-transcriptionnelles des ARNm

Le niveau d'expression d'un gène est principalement défini par 3 facteurs : la fréquence de la transcription, la stabilité de l'ARN messager (ARNm), et la fréquence de la traduction de l'ARNm en protéine. Les recherches menées sur les dernières décennies ont mis en avant la présence de régulateurs d'expression présents après l'étape de transcription, que ce soit chez les eucaryotes ou chez les procaryotes. Différentes régulations peuvent avoir lieu au niveau des

étapes d'épissage ou de transport des ARNm (chez les eucaryotes) mais aussi au niveau des étapes de dégradation et de traduction des ARN.

L'analyse de la dégradation des ARNm chez les procaryotes est particulièrement difficile à cause de la courte durée de demi-vie des ARNm. Chez *E. coli*, la demi-vie moyenne des ARNm à 37°C est d'environ 2,4 minutes. Cela peut être expliqué par le fait que la synthèse puis la dégradation rapide des ARNm permet non seulement une adaptation furtive aux changements environnementaux, mais aussi de réaliser les changements métaboliques nécessaires à la croissance et division cellulaires. Le contrôle de la dégradation des ARNm est permis par une combinaison d'endo- et d'exoribonucléases. Une traduction active, avec un ribosome présent sur les sites ciblés par les nucléases, peut alors protéger l'ARNm. Les endoribonucléases vont couper le brin au niveau des bases internes, alors que les exoribonucléases vont le dégrader à partir de ses extrémités. Les homologues fortes entre endo- et exoribonucléases d'*E. coli* et celles d'organismes procaryotes et eucaryotes suggèrent que les acteurs de la dégradation d'ARN ont été conservés à travers l'évolution (153).

A ce jour, les endoribonucléases connues impliquées dans la dégradation des ARN sont au nombre de 3 : RNase III, RNase P, et RNase E. Le clivage de certains transcrits par la RNase III permet une meilleure stabilité et donc une plus forte expression du gène correspondant, mais la RNase III est notamment connue pour son rôle dans la dégradation de certains ARN messagers (154). La RNase P quant à elle est une enzyme essentielle qui va générer les extrémité 5' des ARN de transfert (ARNt) chez les procaryotes et chez les eucaryotes (155). Elle est impliquée dans la maturation des ARN messagers chez *Salmonella typhimurium* (156), et serait également impliquée dans le traitement ou la dégradation d'autres ARN messagers. Enfin, la RNase E semble être l'élément principal de la dégradation des ARN messagers chez les bactéries. Dans les souches thermosensibles où le gène de la RNase E est inactivée, la plupart des ARN messagers sont très stables (157,158). Cette enzyme clive les ARN simple brin au niveau de régions riches en A/U (159). Les structures secondaires présente au sein des ARN messagers peuvent affecter l'activité de l'enzyme. Celle-ci se lie de manière préférentielle aux extrémités 5' simples brins (160–162).

En addition à l'activité des endoribonucléases, les exoribonucléases participent également à la déstabilisation des ARN messagers. La PNPase et la RNase II sont les 2 exoribonucléases majeures chez *E. coli* qui dégradent les ARN messagers en mononucléotides, dans la direction 3' – 5' (163). L'activité phospholitique de la PNPase et l'activité hydrolytique de la RNase II

génèrent respectivement des oligonucléotides diphosphates et monophosphates. Les mutants simples pour l'une ou l'autre de ces enzymes sont viables tandis que le mutant double ne l'est pas, suggérant que ces exonucléases peuvent se suppléer pour dégrader les mêmes ARNm. Tout comme pour la RNase E, les structures secondaires présentes dans les ARN peuvent ralentir voir stopper l'activité de ces exonucléases (164,165).

Certaines des exo- et des endoribonucléases citées ci-dessus peuvent former un complexe multienzymatique, appelé dégradosome, responsable de la stabilité et de la dégradation des ARN messagers (166). C'est notamment le cas de la RNase E et de la PNPase.

Chez les bactéries, les ARN messagers peuvent être dégradés par un grand nombre de mécanismes qui agissent indépendamment mais simultanément, et qui ciblent différents sites, avec des niveaux d'efficacité différents. L'accessibilité à ces sites de dégradation dépend de la structure de l'ARN, de sa protection par les ribosomes pendant la traduction, et de son degré de polyadénylation (167). En effet, la présence de structures de type tige-boucle dans la région 3' des ARN messagers peut les protéger contre l'action des exonucléases. Cependant, d'autres facteurs peuvent rendre l'ARN messager moins stable. C'est le cas lorsqu'une séquence poly(A) est présente en 3' de l'ARN messager, qui va alors être moins stable (168). Enfin, lors de la traduction, le ribosome va se déplacer le long de l'ARN messager et masquer des sites reconnus par les RNases. Toutefois, le ribosome protège seulement la courte séquence de l'ARN qu'elle recouvre et non la molécule entière (169).

Bien que l'expression des gènes bactériens soit étudiée depuis plusieurs décennies, les exemples pour lesquels tous les aspects de la régulation sont élucidés et compris en détail restent peu nombreux. Heureusement, les connaissances acquises semblent s'appliquer aux nouveaux génomes séquencés. Il est important de noter ici que l'instabilité des ARN messagers peut être un des facteurs limitant dans l'utilisation des microorganismes pour la synthèse de protéines d'intérêts.

Un domaine encore méconnu est l'impact de la conformation des chromosomes bactériens sur l'expression des gènes. Les études récentes montrent que la localisation des unités transcriptionnelles au sein du chromosome a une grande importance sur leur niveau de transcription. Pour élucider les mécanismes impliqués dans cette régulation, un des besoins les plus pressants semble être le développement d'une méthode ou d'un modèle simple d'investigation.

2.3 Organisation chromosomique

2.3.1 Unités transcriptionnelles : gènes et opérons

La définition de ce que représente réellement un gène est variable en fonction de la personne à qui la question est posée. Deux grandes classes de définition existent, la première, issue des théories classiques de l'hérédité, stipule qu'un gène correspond à la plus basique entité physique de l'hérédité. La seconde, centrée sur la génétique moléculaire, énonce qu'un gène est une petite région de l'ADN qui va être transcrite soit en ARN messager lui-même traduit en protéine, soit en une molécule fonctionnelle d'ARN. Un gène représente alors une simple brique dans la géante construction qu'est le génome. Si on considérait l'ensemble de l'ADN d'une bactérie comme une poésie, un nucléotide correspondrait à une lettre, un gène à un mot, un régulon à un vers, et un génome à un poème.

Chez les bactéries, les protéines impliquées dans une même fonction sont souvent transcrites à partir de gènes regroupés les uns à la suite des autres dans un même cluster sur le chromosome et exprimés à partir d'un seul promoteur. Cet ensemble de gènes est appelé opéron. Lors de la transcription, l'ensemble de l'opéron va correspondre à une seule unité transcriptionnelle et va être traduit en un seul ARNm, appelé dès lors ARN messager polycistronique. Cet ARNm va ensuite être traduit en plusieurs protéines correspondant à l'expression des différents gènes de l'opéron (Figure 9). C'est un moyen pour la cellule d'exprimer en même temps et de co-réguler (avec exceptions) des protéines dont la fonction simultanée est nécessaire et la stoechiométrie relative est importante. Par opposition, un ARNm transcrit à partir d'un seul gène et traduit en une seule protéine est appelé ARNm monocistronique.

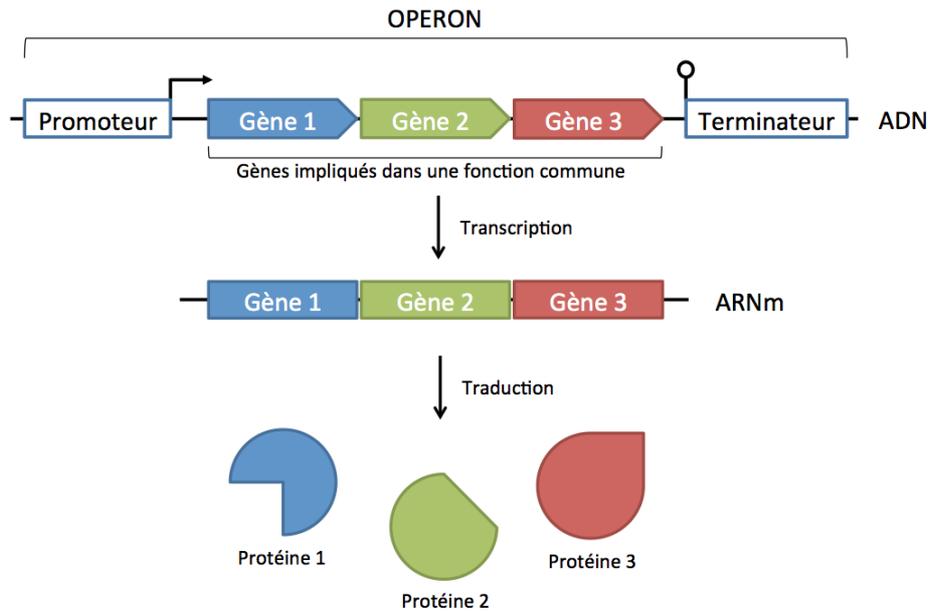


Figure 9 : Organisation d'un opéron contenant 3 gènes.

La transcription de l'opéron donne un ARNm polycistronique qui est ensuite traduit en 3 protéines distinctes.

2.3.2 Organisation tridimensionnelle

Les procaryotes contiennent en leur centre une structure non délimitée par une membrane appelée nucléoïde. Cette structure est un complexe macromoléculaire qui regroupe le matériel génétique et les molécules qui y sont associées. Bien que l'étude du nucléoïde a commencé il y a 60 ans (170), ce n'est que récemment avec le développement des techniques d'imagerie et d'analyse génétique que son organisation spatiale est devenue accessible. On sait désormais que la structure circulaire du chromosome subit une compaction, imposée par plusieurs facteurs : l'encombrement moléculaire, qui correspond à des régions concentrées en ADN et molécules, la dynamique structurale de l'ADN, les niveaux de surenroulement de l'ADN, et l'interaction avec des protéines, appelées NAP. En plus de régir l'architecture de l'ADN, le surenroulement et les NAP vont influencer l'expression des gènes.

Afin d'appréhender la compaction de l'ADN d'*E. coli*, il est essentiel de comparer la taille de l'ADN à celle de la cellule qui l'héberge. Dans le cas de la bactérie *E. coli* K-12, la cellule mesure environ 2µm de long sur 0,5µm de diamètre, alors que son chromosome circulaire contenant 4 288 gènes (9) a une circonférence de 1,5mm. Le chromosome doit donc être compacté environ 1 000 fois pour contenir dans les cellules (171). Cette compaction n'est pas aléatoire,

elle doit être réalisée de façon minutieuse afin de ne pas entraver les processus de réplication, de ségrégation et d'expression génique. En effet, la compaction du chromosome peut représenter une contrainte physique à l'initiation de la transcription (172,173).

2.3.2.1 Macrodomaines

A l'échelle des génomes, la compaction du chromosome de *E. coli* est définie par 6 grandes régions de taille équivalente (Figure 10) : quatre macrodomaines (Ori, Gauche, Droite, Ter) et deux régions non-structurées (NS-gauche et NS-droite) (174). Ces régions correspondent à des régions d'environ 1 Mbp chacune. Elles se distinguent par une faible probabilité d'interactions entre elles, par une forte probabilité d'interactions internes, et par la nature différente des protéines qui y sont liées (175).

Le macrodomaine Ori représente la région centrée autour de l'origine de réplication du chromosome (*oriC*) et le macrodomaine Ter représente la région centrée autour du terminus de réplication. Les quatre autres domaines se répartissent sur chacun des réplichores (les deux « bras ») du chromosome sur lesquels l'ADN polymérase se déplace pendant la réplication (Figure 10).

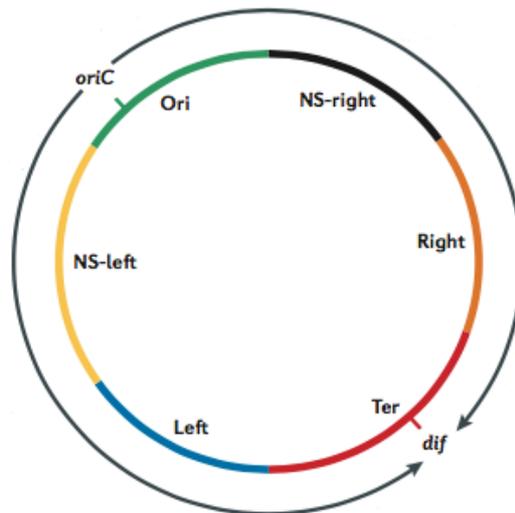


Figure 10 : Organisation du nucléoïde d'*E.coli* (176).

Les macrodomaines sont représentés par des couleurs différentes. oriC correspond à l'origine de réplication et dif correspond au terminus de réplication. Les flèches représentent la direction des réplichores pendant la réplication

Chez *E. coli*, *oriC* est positionné au milieu de la cellule, avec chacun des repli-chores de part et d'autre le long de la cellule. Cet arrangement est permis grâce au complexe de maintenance de la structure du chromosome, MukBEF (177). Le macrodomaine Ter se déplace d'un pôle de la cellule vers son centre au moment de la séparation des cellules filles, et est maintenu en cette position notamment par la protéine MatP. MatP va se lier à l'ADN au niveau d'un motif appelé MatS uniquement présent dans la région Ter (178). De plus, MatP va permettre la formation de boucles d'ADN dans le macrodomaine Ter en se liant à plusieurs séquences MatS et ainsi participer à la compaction de ce macrodomaine.

D'autres protéines, telles que SeqA et SImA, sont également distribuées de manière particulière le long du chromosome. SeqA est un régulateur négatif de l'initiation de la réplication chez *E. coli* qui se lie aux sites 5'-GATC-3' hemi-méthylées (179,180). Ces sites sont abondants dans la région proche d'*oriC*, un peu moins présent dans le reste du chromosome, et complètement absents dans le macrodomaine Ter (181). SImA a un rôle important dans le positionnement des chromosomes pendant la division cellulaire. Tout comme SeqA, la protéine SImA a une distribution qui correspond à l'organisation en macrodomaines du chromosome. Elle se lie à une séquence spécifique qui est absente du domaine Ter. Elle ne se lie quasiment pas dans les macrodomaines gauche et droite. Ainsi, SImA est principalement présente dans la région Ori et dans les régions flanquantes non structurées (182,183).

Ces 3 protéines MatP, SeqA et SImA sont donc étroitement associées aux macrodomaines, ce qui les différencie des NAP, qui ont tendance à être moins macrodomaines-spécifiques. De plus, il n'a pas été démontré, jusqu'à maintenant, que ces 3 protéines sont impliquées dans l'expression globale des gènes, contrairement à une grande partie des NAP. Cela suppose que les NAP pourraient être impliqués dans l'organisation chromosomique à un niveau différent que celui des macrodomaines. Cette hypothèse est appuyée par de nombreuses recherches qui montrent que la liaison des NAP est impliquée dans la formation de structures plus petites, appelées microdomaines.

2.3.2.2 Microdomaines

Un microdomaine correspond à la région relaxée topologiquement par une coupure simple-brin (« nick » en anglais) dans l'ADN. Initialement, le nombre et la taille des microdomaines dans le chromosome étaient soit sur- soit sous-estimés (184,185). Des

estimations bien plus proches de la réalité ont été obtenues par une combinaison d'observations sur les boucles d'ADN, le surenroulement, et la recombinaison au sein de microdomaines (186,187). Il semble maintenant clair que le chromosome d'*E. coli* est divisée en une succession de 400 boucles d'ADN d'une taille avoisinant les 10 kpb. Cette organisation en microdomaines semble être une caractéristique des bactéries en croissance exponentielle (187) (Figure 11A). La structure du chromosome est plus relâchée dans des conditions lentes de croissance (188).

L'observation du nucléoïde isolé d'*E. coli* au microscope montre une structure en rosette contenant une succession de boucles d'ADN (Figure 11). Chaque boucle correspondrait à un seul domaine topologique. Les résultats obtenus par des études physiques et génétiques montrent que les NAP, plus spécifiquement les protéines H-NS (Histone-like Nucleoid Structuring protein) et Fis (Factor for Inversion Stimulation), participent à la formation des microdomaines et agissent en tant que « insulateurs ». En effet, les observations réalisées grâce à la technique de Capture de la Conformation du Chromosome (3C) montrent que H-NS a la faculté de colocaliser les sites connus de liaison de cette protéine à l'ADN (189). Autrement dit, les séquences de fixation de H-NS qui sont espacées sur le chromosome sont retrouvées dans un même espace tridimensionnel ou « point focal ». Cela est permis par la capacité de multivalence d'un monomère ou d'un oligomère de la protéine H-NS, qui va se lier à plusieurs sites ADN, entraînant la formation d'une ou plusieurs boucles avec un seul point d'ancrage.

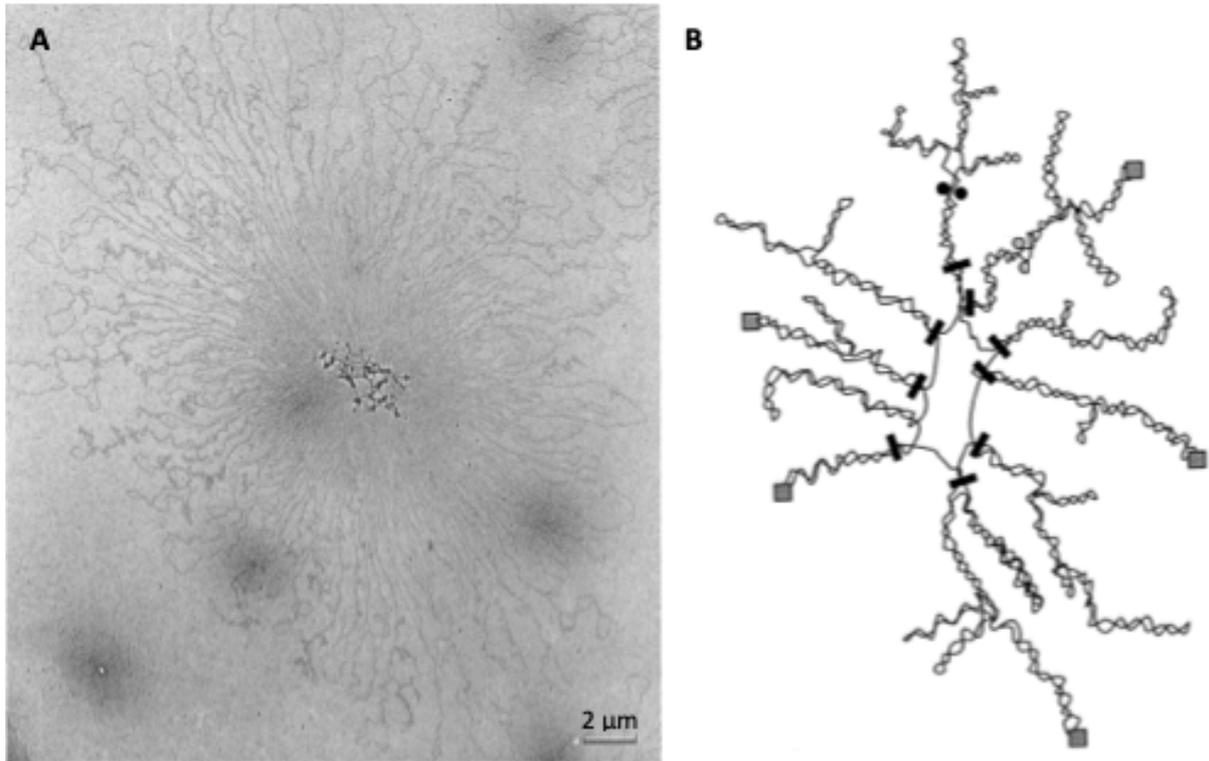


Figure 11 : Chromosome d'*E. coli* organisé en microdomaines.

A. Microscopie électronique d'un chromosome isolé d'E. coli en phase exponentielle de croissance présentant plus de 140 boucles d'ADN émanant du noyau central (190). B. Représentation schématique de l'organisation du chromosome d'E. coli en microdomaines surenroulés (191).

Une autre caractéristique fondamentale du repliement et de la compaction de l'ADN chromosomique est son niveau de surenroulement négatif (171) (voir section 2.5), qui est également essentiel pour les nombreux processus qui nécessitent le déroulement et la séparation des deux brins d'ADN (192,193). Cependant, le surenroulement est très fragile, en absence de barrières topologiques une seule cassure simple brin de l'ADN suffirait à relâcher les enroulements dans le chromosome entier. Dans la réalité, les effets des cassures d'ADN sont limités par la liaison des NAP (et peut-être d'ARN) tout au long du chromosome. Ces liaisons divisent le chromosome en microdomaines successifs isolés topologiquement les uns des autres.

2.4 Protéines « architectes »

Chez tous les êtres vivants le matériel génétique doit être organisé et stocké de manière compatible avec les grands mécanismes cellulaires : la réplication de l'ADN, la ségrégation des chromosomes et la transcription des gènes. Des protéines architectes qui compactent l'ADN ont été identifiées dans les 3 domaines de la vie (Eucaryotes, Bactéries, et Archées). Chez les eucaryotes, ces protéines sont connues sous le nom d'histones. Leur capacité à influencer la structure de la chromatine et la transcription a été étudiée et approfondie en détail depuis des années. Ces connaissances ont orienté les travaux réalisés chez les procaryotes dans la recherche de protéines qui auraient une fonction analogue aux histones trouvées chez les eucaryotes. Ces protéines procaryotes, d'abord qualifiées comme « histone-like proteins » (194,195), sont désormais regroupées sous le nom de « protéines associées au nucléoïde » (NAP) (Figure 12).

Il est difficile de différencier les NAP des facteurs de transcription car ces derniers sont souvent impliqués dans l'architecture du génome (121,196). Les définitions basées sur l'abondance protéique, les mécanismes de régulation, le nombre de gènes ciblés, ou les sites de liaison à l'ADN ne permettent pas une réelle distinction entre les NAP et les facteurs de transcription. Le point clé marquant la différence entre ces deux types de protéines concerne la capacité qu'ont les NAP à compacter le chromosome. Une définition commune serait alors que les NAP sont un petit groupe de protéines très abondantes qui jouent un rôle dans le repliement et la compaction des chromosomes bactériens (197) mais aussi dans la régulation de la transcription.

Une caractéristique importante des NAP est leur capacité à se lier à l'ADN de manière à la fois spécifique (séquence ou structure spécifique) et non-spécifique. Les NAP sont donc définies comme des protéines à fonction double (« dual fonction »). La liaison de NAP à des séquences spécifiques est impliquée dans le contrôle de la transcription, de la réplication, de la recombinaison et de la réparation de l'ADN. Leur fixation sur des séquences non-spécifiques dépendrait du très grand excès (de plusieurs ordres de grandeurs) de ces protéines par rapport aux sites de liaison spécifique dans le génome. Il faut cependant noter que les séquences non-spécifiques ne sont probablement pas aléatoires. Il est possible qu'elles contiennent faible spécificité de séquence, de structure, ou de reconnaissance d'une structure induite par la liaison d'autres NAP.

Bien que les mécanismes moléculaires utilisés par les NAPs pour condenser l'ADN sont peu connus, les résultats obtenus à ce jour permettent de dire que ces protéines induisent ou stabilisent des structures dans l'ADN, et ainsi diminuent l'amplitude de la surface occupée par le double brin. En se liant à l'ADN, les NAP peuvent le courber (« bend »), l'embobiner (« wrap »), ou le ponter (« bridge ») (Figure 12). Elles permettent également d'emprisonner les surenroulements au sein des boucles d'ADN et donc de contribuer à la compaction du chromosome et au contrôle de l'expression génique par la topologie.

Il y a au moins 12 NAPs identifiées chez *E. coli* (198). Parmi elles, les plus étudiées sont HU, IHF, H-NS et Fis. Cependant, des observations récentes suggèrent qu'un grand nombre de protéines, certes moins abondantes, seraient capables de se fixer transitoirement à l'ADN et pourraient participer à l'organisation du nucléoïde (199).

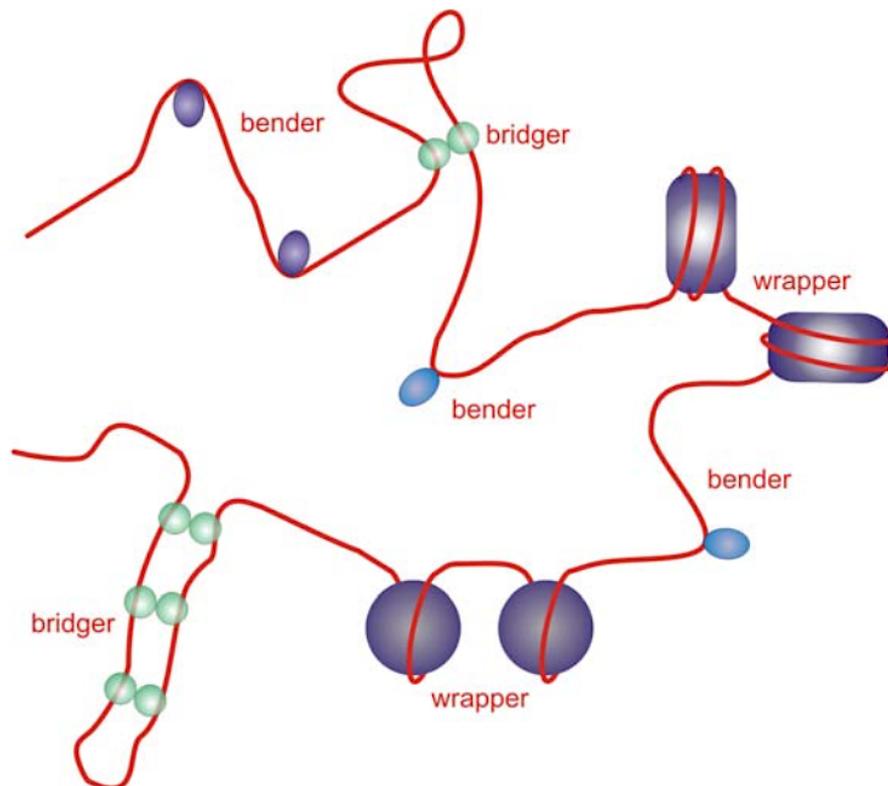


Figure 12 : Courbures, embobinages, et pontages de l'ADN par des protéines architecturales (19).

Le trait rouge correspond à l'ADN. Les protéines entraînant un pontage sont représentées en vert et nommées « bridger ». Les protéines qui embobinent l'ADN sont représentées par des ronds ou rectangles violets et nommées « wrapper ». Les protéines qui courbent l'ADN sont représentées par des ovales violets ou bleus et nommés « bender ».

2.4.1 Wrapping

Certaines protéines, dites de « wrapping », peuvent enrouler l'ADN autour d'elles réduisant considérablement le volume occupé par l'ADN (Figure 12). Les protéines de « wrapping » les plus connues sont les histones, présentes dans les cellules eucaryotes, chez qui elles permettent la formation de nucléosomes grâce à une octamérisation. Les nucléosomes sont présents environ toutes les 200 paires de bases, et permettent d'enrouler environ 146 paires de bases de séquence ADN, qui forme 1,6 tour autour de l'octamère d'histone.

Bien que moins fréquentes que les histones, des protéines de « wrapping » sont également présentes chez les bactéries. C'est le cas des protéines de la famille Lrp/AsnC, qui forment des octamères d'apparence semblables à des disques qui vont enrouler l'ADN autour d'elle dans une hélice droite (200,201). Ce mode d'enroulement va permettre d'emprisonner les surenroulements ainsi qu'une compaction de l'ADN (201,202).

2.4.2 Courbures

Bien que le « wrapping » semble être le mode de compaction dominant chez les eucaryotes, il semblerait que ce ne soit pas le cas pour les procaryotes puisque ces derniers ne produisent pas ou peu ces protéines. Dans ces organismes, les protéines induisant des courbures (« bending ») sont plus répandues et jouent un rôle important dans la compaction du chromosome (Figure 12). L'exemple le plus étudié concerne la famille des protéines HU/IHF. La protéine HU correspond à un dimère qui va se lier à l'ADN et entraîner une courbure de 160° de la molécule d'ADN. Cette liaison n'est pas spécifique d'une séquence mais HU semble se lier principalement à des régions flexibles de l'ADN (203,204). La protéine IHF, quant à elle, va se lier à une séquence spécifique (205). Les courbures engendrées par la liaison de ces protéines permettent de modifier la forme locale de l'ADN pour la rendre plus compacte tout en empêchant l'agrégation.

2.4.3 Boucles

Certaines protéines sont capables de former des liens dynamiques entre différents duplex ADN, ou entre un duplex ADN et une autre protéine architecturale, entraînant la formation de boucles (206,207). Ces protéines sont dites di- ou multi-valentes car capables de se fixer simultanément à plusieurs sites ADN. Cette classe de protéines permet donc de mettre en place un niveau supérieur d'organisation du chromosome. Une protéine architecte formant des boucles dans l'ADN très bien caractérisée est H-NS (Histone-like Nucleoid Structuring protein). H-NS est un dimère qui possède deux sites de liaison à l'ADN et qui va donc pouvoir interagir simultanément avec deux duplex ADN (208,209). Chez *E. coli* et *Salmonella enterica*, la protéine va se lier sur des séquences riches en AT avec une forte affinité (210–214). La distribution de ces sites riches en AT correspond à l'emplacement probable des ancrages des microdomaines dans le chromosome (207). Une autre protéine associée au nucléoïde et capable de former des boucles d'ADN est appelée Fis (Factor for Inversion Stimulation) (215). Elle est aussi capable de se lier tout au long du génome, et comme H-NS, préférentiellement dans les régions intergéniques (216). Des protéines capables de former des ponts ont été identifiées dans nombreuses autres bactéries plus ou moins proches phylogénétiquement d'*E. coli* (217–220). Elles présentent ou non des similarités de séquence avec H-NS. Cette observation suggère que le pontage de l'ADN est un processus de compaction des génomes très fréquent chez les bactéries.

Cependant, les bactéries privées de NAP telles que H-NS et Fis maintiennent la structure de leur nucléoïde (221), indiquant que ces NAP ne sont pas essentielles pour l'organisation du nucléoïde. Ces observations montrent que nos connaissances sur la compaction des génomes bactériens sont incomplètes et que d'autres facteurs de compaction importants restent à découvrir.

Le positionnement des sites de liaison des protéines de pontage et l'effet de la liaison sur la molécule d'ADN déterminent l'impact éventuel de la formation d'un pont sur l'expression génique. La formation du pont peut inhiber l'expression d'un gène, par exemple en piégeant l'ARN polymérase, ou en l'empêchant d'accéder au promoteur. C'est le cas de la protéine LacI dans l'opéron lactose, et de la protéine Lambda-cl chez le phage Lambda.

2.4.3.1 Lacl

La régulation génique a été découverte grâce aux travaux de François Jacob et Jacques Monod en 1961 sur l'opéron lactose (222). L'opéron lactose est composé de 3 gènes responsables de l'utilisation du lactose par la cellule qui sont transcrits en un ARN polycistronique (Figure 13a). En amont, se trouve le gène *lacI*, codant pour la protéine lacl également appelée Lac repressor, qui est responsable de la régulation de l'expression de l'opéron. Les premières études ont montré qu'en absence de lactose dans le milieu de culture, la protéine Lacl se lie à une séquence opératrice appelée O₁ présente au niveau du promoteur des gènes *lacZYA* et réprime la transcription de cet opéron (118). Additionnellement à O₁, les alignements de séquence avec faible stringence ont permis la découverte de deux séquences opératrices (O₂ et O₃) dans cette région, appelées opérateurs auxiliaires. Les opérateurs auxiliaires ont une séquence similaire à O₁ et peuvent également lier Lacl. En premier lieu, ces sites auxiliaires étaient considérés comme des vestiges de l'évolution n'ayant aucune fonction (152). Cette idée était due au fait que ces sites sont loin du promoteur, suggérant qu'une interaction avec eux n'affecterait pas l'ARN polymérase. De plus, l'affinité de O₂ pour Lacl est 10 fois plus faible que celle de O₁, et celle de O₃ est plus 300 fois plus faible. En outre, la suppression de l'un ou l'autre des opérateurs auxiliaires laisse les niveaux de répression quasi inchangés (152). Cependant, le rôle des opérateurs O₂ et O₃ est essentiel dans la régulation de l'opéron lactose : la suppression simultanée des deux opérateurs auxiliaires entraîne la diminution d'un facteur 100 de la répression par Lacl.

En réalité, la répression des gènes impliqués dans le catabolisme du lactose est permise par la formation d'une boucle dans l'ADN. La protéine régulatrice va se lier au site opérateur O₁ puis à un opérateur auxiliaire et induire la formation d'une boucle d'ADN (223) (Figure 13b,c). De plus, la présence de Lacl sur O₂ ou O₃ augmente la probabilité d'une liaison du répresseur à O₁.

Malgré une complexité apparente, la régulation génique par la formation de boucles d'ADN semble être un mécanisme largement utilisé chez les bactéries. D'abord découvert au niveau de l'opéron *ara*, ce mécanisme est également présent sur d'autres systèmes tels que *lac*, *deo*, *gal* ou encore *gln* (224). La formation de boucle joue également un rôle essentiel dans la régulation chez le phage lambda (λ).

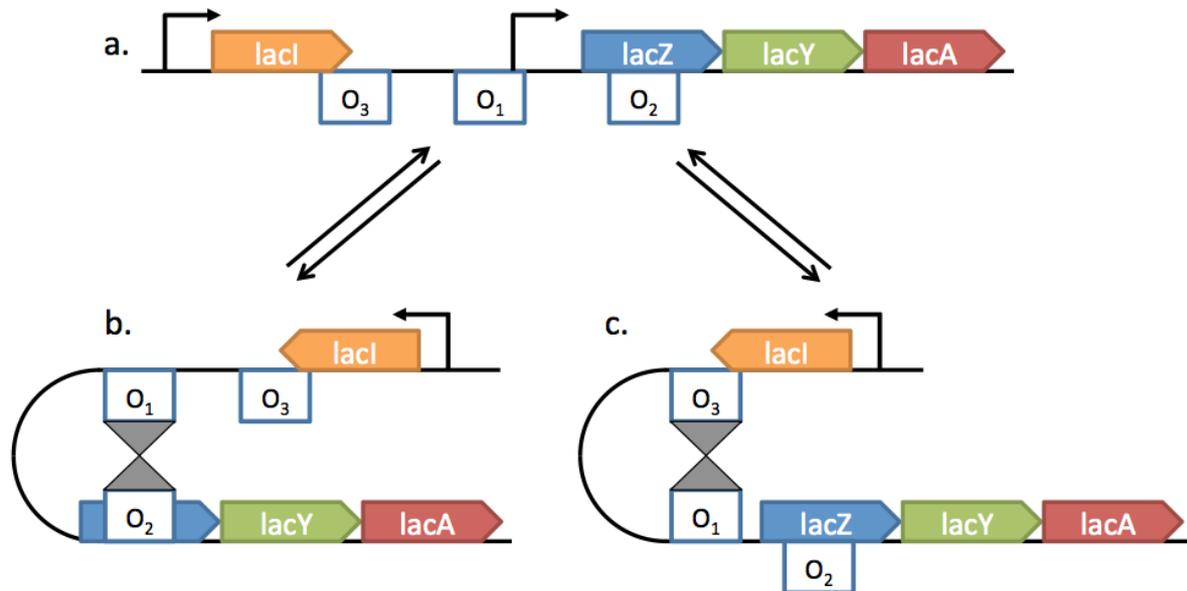


Figure 13 : Représentation de l'opéron lactose.

La représentation n'est pas à l'échelle. Les gènes lacZ, lacY et lacA sont co-transcrits en un ARNm polycistronique. Le gène lacI est immédiatement en amont de l'opéron. a. Opéron en absence de protéine lac repressor. b. Répression de la transcription de l'opéron lac par la formation d'une boucle par la liaison du lac repressor entre les opérateurs O_1 et O_2 . c. Répression de la transcription de l'opéron lac par la formation d'une boucle par la liaison du lac repressor entre les opérateurs O_1 et O_3 .

2.4.3.2 Lambda-cl

Le phage lambda fait partie des premiers systèmes modèles chez qui la régulation génique a été étudiée, en partie en raison de ses deux différents modes de croissance chez son hôte *E. coli*. Pendant le cycle lytique, le bactériophage utilise la machinerie cellulaire pour produire massivement des nouvelles particules phagiques, qui sont ensuite relâchés par la lyse de la cellule hôte. Au contraire, pendant le cycle lysogénique, l'ADN du phage est intégré dans le génome de l'hôte et transféré aux cellules filles après la division cellulaire. Le phage est maintenu dans ce mode inactif par une protéine, appelée cl ou lambda-cl, qui empêche la transcription des gènes impliqués dans le passage en phase lytique. Cette répression est maintenue jusqu'à ce que la cellule subisse des dommages au niveau de son ADN et active RecA pour les réparer. RecA va catalyser la dégradation de la protéine Lambda-cl, ce qui va faire basculer la cellule dans un cycle lytique (225).

La protéine Lambda-cl maintient l'état lysogénique en se liant aux régions opératrices O_L et O_R . Cette liaison a pour conséquence de réprimer les promoteurs P_L et P_R qui permettent le passage dans le cycle lytique (Figure 14A). Les régions opératrices O_L et O_R sont espacées d'environ 2,3kpb et contiennent chacune trois sites opérateurs. Lambda-cl va se lier aux opérateurs sous forme de dimères. La coopérativité entre dimères de Lambda-cl facilite la liaison de la protéine aux sites adjacents, suggérant la formation de tétramères préférentiellement sur les sites O_{R1} . O_{R2} . La liaison de Lambda-cl à O_{R1} . O_{R2} augmente l'expression au niveau de son propre promoteur P_{RM} d'un facteur 10 (226). Le troisième site de liaison dans O_R présente une affinité beaucoup plus faible pour Lambda-cl, mais lorsqu'il est occupé, la transcription est bloquée au niveau du promoteur P_{RM} . Ce rétrocontrôle négatif est permis par la formation d'une boucle d'ADN induite par un pontage entre les régions opératrices O_L et O_R (227). Dans ce complexe, les régions opératrices sont alignées de façon spécifique (O_{L1} avec O_{R1} , O_{L2} avec O_{R2} et O_{L3} avec O_{R3}) (228) (Figure 14B). Les dimères de Lambda-cl liés aux opérateurs 1 et 2 de chaque région opératrice vont se lier sous forme d'un octamère, induisant la formation d'une boucle dans l'ADN (227). Les dimères présents au niveau de O_{R3} et O_{L3} peuvent stabiliser la boucle en formant un tétramère. Les données expérimentales obtenues par Ding *et. al* (229) ont montré la protéine lambda-cl peut séquestrer de très forts niveaux de surenroulement au sein des boucles d'ADN qu'elle forme et empêcher leur diffusion dans le reste du génome. Dans ce contexte, la protéine lambda-cl agit comme une NAP.

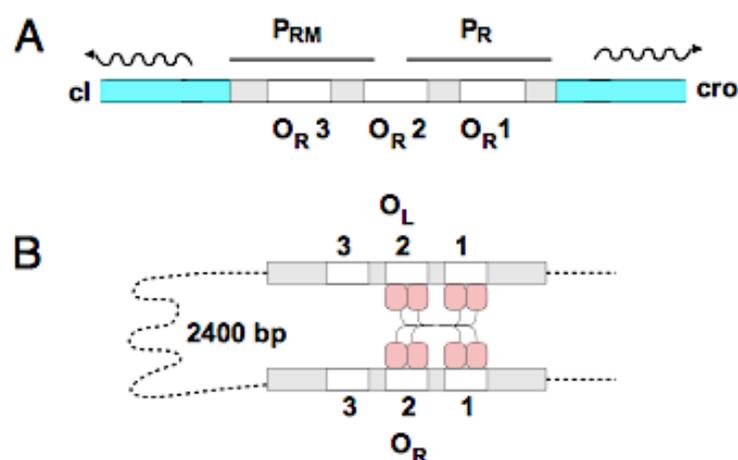


Figure 14 : Représentation schématique des régions opératrices du phage lambda (230).

A. Région du phage lambda contenant les promoteurs P_{RM} et P_R responsables respectivement de l'expression de lambda-cl et de cro et la région opératrice O_R . B. Induction d'une boucle d'ADN par octamérisation de lambda-cl et liaison aux régions opératrices O_R et O_L .

2.4.4 Implication des NAP dans le maintien de la topologie

La transcription varie en fonction de l'état physiologique de la cellule. Les données expérimentales montrent qu'il y a une forte corrélation entre l'activité transcriptionnelle et le nombre et la stabilité des boucles d'ADN dans une région (186,231,232) (Figure 15). L'impact de la liaison des NAP sur l'ADN ne se limite pas à la compaction du nucléoïde. Cette liaison permet également aux NAP d'agir comme des barrières topologiques, qui définissent les microdomaines. Sans présence de barrière topologique dans l'ADN, une seule coupure dans l'ADN serait potentiellement suffisante pour relâcher les surenroulements dans toute la molécule. Chez *E. coli*, environ la moitié des surenroulements de l'ADN sont maintenus par les interactions avec des protéines (233,234). En maintenant des niveaux de surenroulements variables tout au long de la molécule d'ADN, les NAP participent indirectement à la régulation de l'expression génique par le surenroulement.

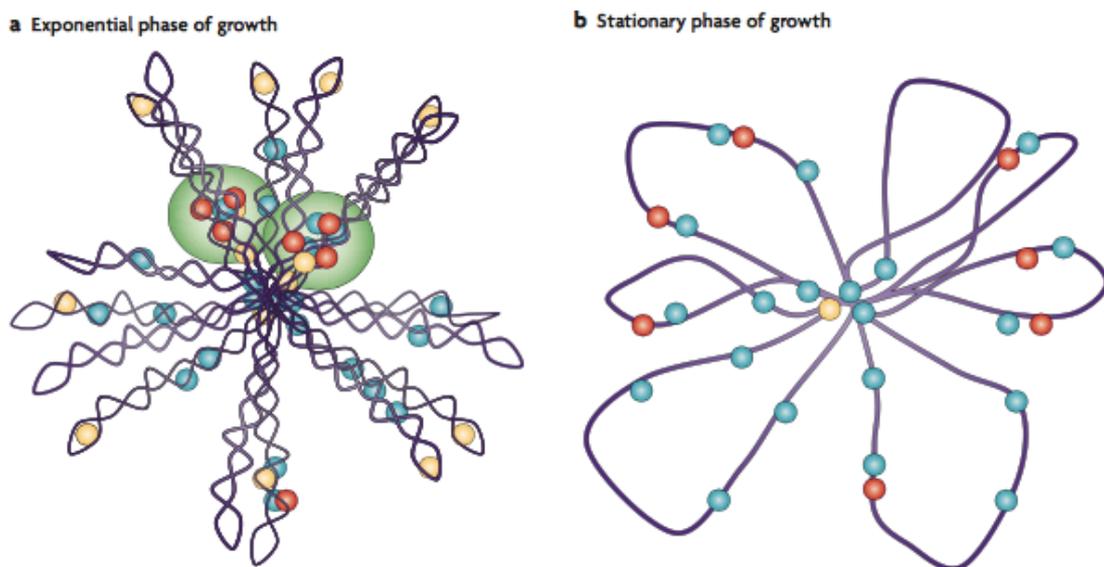


Figure 15 : La liaison de NAP et le surenroulement influencent la structure du nucléoïde (18).

Représentation de la structuration du nucléoïde par les NAP (H-NS et Fis) en fonction de la phase de croissance. En conditions de croissance rapide les mécanismes bactériens présentent une activité plus forte, notamment au niveau des gènes impliqués dans la machinerie traductionnelle qui vont se retrouver dans des points focaux, et l'ADN de la bactérie contient plus de domaines. Lorsque la croissance bactérienne est ralentie, le taux global de transcription diminue et une baisse du nombre de boucles d'ADN est observée. a. Organisation du nucléoïde pendant la phase exponentielle de croissance. b. Organisation du nucléoïde pendant la phase stationnaire de croissance. En vert sont représentés les points focaux, en rouge sont représentés les ARN polymérases, en bleu sont représentés les protéines H-NS, et en jaune sont représentés les protéines Fis.

2.5 Surenroulement

Depuis plusieurs décennies, les recherches menées sur la régulation de l'expression génique se sont concentrées sur les protéines (et plus récemment les petits ARN) qui affectent l'initiation de la transcription. Les modèles de régulation génique présentés mettent de côté l'impact de la molécule d'ADN elle-même sur l'expression des gènes. Or, la molécule d'ADN et avant tout la conformation dans laquelle elle se trouve, joue un rôle essentiel dans la transcription des gènes grâce à sa topologie dynamique.

La découverte du surenroulement de l'ADN a dévoilé la structure tertiaire adoptée par l'ADN (235), c'est-à-dire le repliement sur elle-même de la double hélice dans l'espace. Ce phénomène est présent dans les molécules circulaires qui sont sujettes à une forte contrainte topologique (Figure 16). Le surenroulement de l'ADN est dû à la présence de supertours négatifs (dans le même sens que la double hélice) ou de supertours positifs (dans le sens opposé) sur la molécule d'ADN (Figure 17). Dans la plupart des cellules bactériennes, l'ADN présente un défaut de tours d'hélices et se trouve dans un état de stress de torsion (« torsional stress » en anglais) appelé surenroulement négatif ou sous-enroulement. Ce stress de torsion peut être quantifié sous forme d'enlacement ou « linking number » (Lk) de la molécule d'ADN. Le linking number correspond à l'addition des niveaux de courbure ou « writhe » (Wr) et de torsion ou « twist » (Tw) de l'ADN comme défini par l'équation suivante :

$$Lk = Tw + Wr$$

L'empilement des bases, leur appariement et la conformation des sucres contribuent à l'obtention d'une torsion spécifique par la double hélice. Ces paramètres définissent Tw, le nombre de tours d'hélices de l'ADN ou nombre de révolutions qu'un brin fait autour de l'autre. Par convention, Lk est un chiffre entier qui sera positif dans une hélice droite et négatif dans une hélice gauche. Dans un ADN bicaténaire plan, Wr est égal à zéro, alors Lk sera égal au nombre de torsion Tw, la molécule d'ADN est alors dans un état dit « relâché » (Figure 18). Dans une structure ne montrant pas de variation de Lk, la moindre variation de Tw crée une tension mécanique qui va déplacer l'équilibre. La variation de Tw va être compensée par un changement de Wr, c'est-à-dire par le nombre de fois que le duplex d'ADN va s'enrouler autour de lui-même, ou nombre de supertours. C'est une formation de structure tertiaire, une superhélice, dont le sens de rotation est inverse à celui de la double hélice. Le nombre d'enlacements Lk peut alors s'exprimer sous différentes combinaisons de courbure et de

torsion, et ne variera que si la molécule d'ADN subit une rupture des brins. Dans ce cas, Lk ne change que d'un nombre entier d'unité.

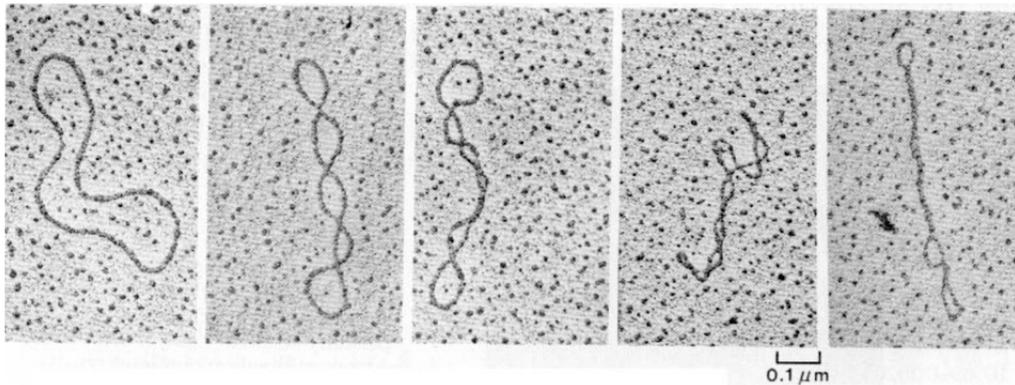


Figure 16 : Structure tertiaire de l'ADN d'un plasmide de 5 kpb observée par microscopie électronique (236).

Le degré de surenroulement augmente progressivement de gauche à droite. Ces structures, composées d'une série de boucles d'ADN successives, sont appelées plectonèmes.

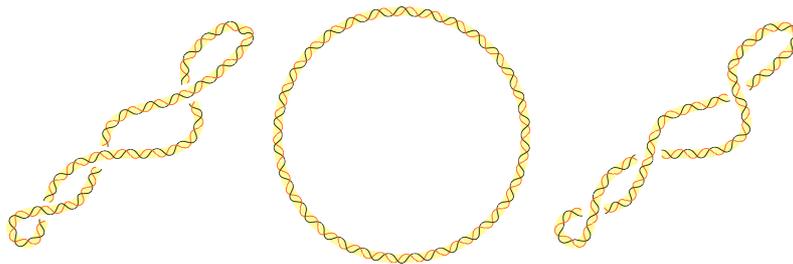


Figure 17 : Représentation du surenroulement dans un ADN circulaire (237).

Trois topoisomères d'un même ADN circulaire : à gauche, ADN surenroulé négativement (superhélice droite) ; au centre, ADN relâché ; à droite, ADN surenroulé positivement (superhélice gauche).

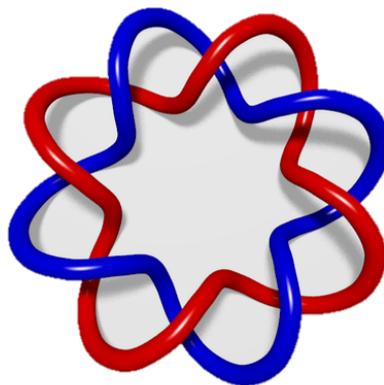


Figure 18 : Liens entre 2 courbes fermées.

Le linking number vaut $Lk = \pm 4$ (le signe dépend de l'orientation relative des deux courbes).

2.5.1 Plectonèmes

Les courbures dans la molécule d'ADN vont se traduire par la formation de structures secondaires. En effet, les défauts d'appariement de nucléotides introduits par le surenroulement sont coûteux en énergie libre. Afin de rétablir l'appariement, la molécule doit rétablir sa torsion en diminuant son tour d'hélice. En conservant l'équilibre en place, la diminution de torsion (T_w) est permise par l'augmentation des niveaux de courbure de la molécule d'ADN.

Bauer et Vinograd ont postulé que le toroïde (ou solénoïde) et le plectonème sont les deux principales formes de structuration du surenroulement négatif chez tous les organismes (238). Le toroïde correspond à une succession de supertours négatifs, alors que le plectonème linéaire est une accumulation en un seul segment de la molécule d'ADN de boucles de supertours négatifs tressés (Figure 19). Des formes alternatives telles que le plectonème branché et le solénoïde (ou toroïde linéaire) ont également été décrites (239–242).

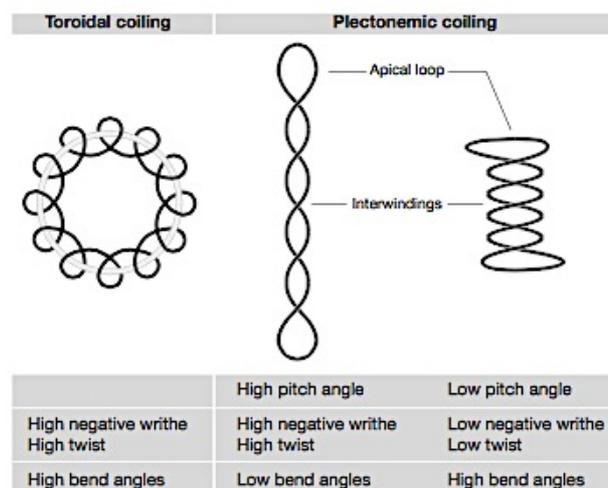


Figure 19 : Représentation des géométries adoptées par un ADN surenroulé (243).

A gauche se trouve le toroïde. Au milieu et droite se trouvent deux géométries différentes de plectonèmes présentant le même nombre de supertours. Le passage de l'une à l'autre se fait par un changement de courbure et de torsion de l'ADN.

Lorsque l'ADN circulaire est observé au microscope, on voit qu'il est essentiellement sous forme plectonémique (244,245) (Figure 16). Cette structure permet une meilleure stabilité de l'ADN surenroulé négativement, grâce aux contacts entre les différents segments du double brin d'ADN. De plus, la formation de plectonèmes permet la juxtaposition de sites distants de plusieurs kpb à plusieurs dizaines de kpb (246,247).

2.5.2 Liens entre surenroulement et transcription

Le chromosome bactérien est surenroulé négativement (184). Cet état topologique est important pour la transcription mais aussi pour la réplication et la recombinaison (248–250). En effet, la perte de surenroulement négatif peut entraîner des déficiences au niveau de ces processus et conduire à la mort cellulaire. La réplication et la transcription nécessitent l'ouverture du double brin d'ADN, ce qui cause des changements de la topologie locale.

2.5.2.1 Le surenroulement local affecte la transcription

Le surenroulement de la molécule d'ADN peut affecter la transcription à plusieurs niveaux. Le niveau le plus évident serait l'étape d'isomérisation pendant l'initiation de la transcription. En effet, l'énergie nécessaire à la cassure des liaisons hydrogènes lors du passage du complexe fermé R_{Pc} au complexe ouvert R_{Po} est fournie par le surenroulement négatif de l'ADN. Un autre niveau auquel le surenroulement pourrait affecter la transcription est le recrutement de l'ARN polymérase sur les boîtes -35 et -10 du promoteur. En effet, dans les promoteurs qui contiennent un spacer dont la taille est non conventionnelle (autre que 17pb), la torsion de l'ADN est un élément essentiel pour la reconnaissance par la sous-unité σ (17). Ainsi, le recrutement de l'ARN polymérase sur un promoteur peut dépendre de la topologie de la région (251). De la même façon, le surenroulement peut également affecter l'interaction d'un facteur de transcription avec l'ARN polymérase. En effet, l'activation du complexe fermé formé sur des promoteurs dépendants de σ^{54} nécessite un contact sensible au surenroulement local avec un facteur de transcription (251).

2.5.2.2 La transcription affecte le surenroulement local

2.5.2.2.1 Twin-supercoiled domain model

Il y a une trentaine d'années, Liu et Wang ont proposé un modèle afin de décrire les liens entre la transcription et les niveaux locaux de surenroulement (252). Les études menées depuis, qu'elles soient à l'échelle de molécules uniques ou à l'échelle génomique, ont renforcé ce modèle, qui semble désormais universellement adopté (253–256). Ce modèle, appelé

« twin-supercoiled domain model » en anglais, montre l'induction de surenroulements dans l'ADN au niveau local par la transcription : l'avancée de l'ARN polymérase provoque du surenroulement positif devant la fourche de transcription et du surenroulement négatif derrière cette fourche (252) (Figure 20). Le complexe ouvert va rapidement atteindre un point où sa progression sera difficile, à moins de diminuer les contraintes topologiques : soit en entraînant une rotation du complexe autour de la molécule d'ADN, soit en entraînant une rotation de la molécule d'ADN, soit par l'intervention d'agents extérieurs permettant de restaurer les niveaux habituels de surenroulement (voir section 2.5.3). La rotation de l'ARN polymérase autour de la molécule d'ADN semble peu probable de par la taille du complexe et la difficulté des mouvements dans les milieux concentrés en molécules (257–259). De plus, l'ARN naissant grandit et la liaison de ribosome sur celui-ci rend encore moins probable cette hypothèse. Dans ces conditions, le complexe est considéré comme ancré (252). La rotation libre de la molécule d'ADN peut quant à elle être rendue difficile par les points d'ancrage au niveau des différents domaines. La transcription d'un gène va donc générer des surenroulements locaux (254,255) qui pourraient être emprisonnés dans des boucles délimitées par différentes molécules liées à l'ADN, notamment les NAP (Figure 20).

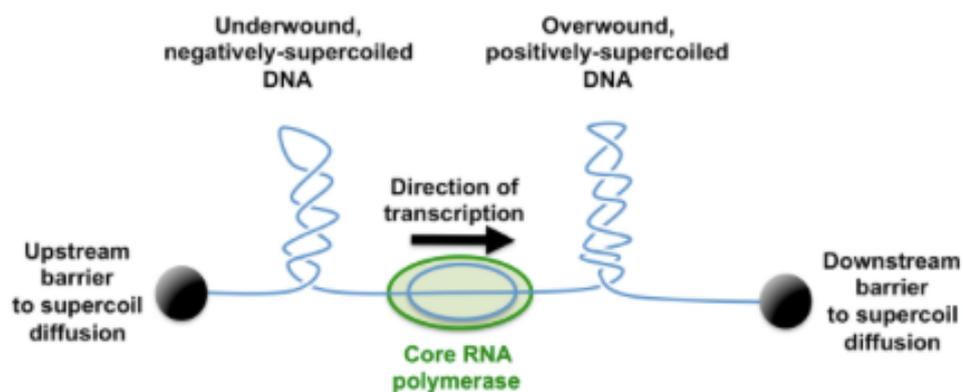


Figure 20 : Modèle « Twin-supercoiled domain » (260).

L'ARN polymérase est engagée dans l'élongation de la transcription (l'ARNm, les ribosomes, et le polypeptide naissant sont omis volontairement pour plus de clarté). Avec l'avancée du complexe de transcription/traduction de la gauche vers la droite, l'ADN en aval de l'ARN polymérase devient de plus en plus surenroulé positivement alors que l'ADN en amont de l'ARN polymérase devient de plus en plus surenroulé négativement. Les barrières topologiques (ronds noirs) présentes de part et d'autre du complexe de transcription vont empêcher la diffusion de ces surenroulements. L'ADN est représenté en bleu, l'ARN polymérase est représentée en vert.

2.5.2.2.2 Transcriptional bursting

Le modèle twin-supercoiled domain fournit une explication à un phénomène inattendu et universel : la stochasticité de la transcription des gènes fortement exprimés. En effet, différentes données montrent que ces gènes sont transcrits par rafales (« transcriptional bursting » en anglais), c'est-à-dire par cycles de transcription successifs suivi de périodes d'inactivités (261–266). Dans une étude récente, Chong *et al.* ont suivi la synthèse d'ARN *in vitro* par l'ARN polymérase T7 lorsque la matrice ADN était sujette à de fortes contraintes topologiques (267). Ils ont observé qu'en présence de la topoisomérase I seule, dont le rôle est d'enlever les surenroulements négatifs, l'initiation et l'élongation de la transcription sont ralenties puis éventuellement arrêtées à cause de l'accumulation de surenroulements positifs causés par la transcription. Conformément à cette hypothèse, la transcription est restaurée quand on ajoute au milieu réactionnel une enzyme, la gyrase, dont le rôle est d'enlever les surenroulements positifs. Les auteurs ont ensuite montré que l'inhibition de la transcription par le surenroulement positif a aussi lieu *in vivo* (267). Ils ont alors proposé une explication au phénomène de transcription en rafales qui tient compte de l'organisation en microdomaines du chromosome de *E. coli*. Dans un microdomaine contenant un gène fortement transcrit, des surenroulements s'accumulent. Les surenroulements négatifs en amont de l'ARN polymérase sont rapidement enlevés par l'action de la Topoisomérase I, alors que l'action de la gyrase est insuffisante pour empêcher l'accumulation progressive de surenroulements positifs en aval de la fourche de transcription. Cette accumulation va inhiber la transcription et « désactiver » le gène. La liaison d'une molécule de gyrase va ensuite libérer les surenroulements positifs accumulés (voir section 2.5.3.2.2) et ainsi « activer » le gène. Les gènes vont donc sans cesse alterner entre un état actif et un état inactif, donnant une explication au phénomène de transcription en rafales.

Contrôler le surenroulement pendant la transcription ou entre différents cycles de transcription est donc primordial. A la fin d'un cycle de transcription, le changement topologique induit par le passage de l'ARN polymérase pendant le cycle précédent peut modifier l'accessibilité de la région promotrice et affecter négativement l'expression du gène pour un autre cycle de transcription. En effet, les études *in vitro* et *in vivo* primitives dans ce domaine ont montré qu'une légère variation du niveau de surenroulement dans la région

promotrice peut entraîner des différences significatives du niveau d'expression du gène correspondant (268–273). De plus, les surenroulements positifs et négatifs ont un effet différent sur la transcription. L'accumulation de surenroulement négatifs peut faciliter l'ouverture du double brin d'ADN pendant la l'initiation de la transcription et donc entraîner le passage de complexe fermé à complexe ouvert par l'ARN polymérase (274). Cependant, comme pour le surenroulement positif, une trop forte accumulation de surenroulements négatifs a un effet inhibiteur sur la transcription génique.

On peut dès lors penser que le surenroulement local causé par la transcription au niveau d'un promoteur peut impacter l'expression au niveau d'un promoteur proche (275). Ainsi, deux promoteurs qui seraient placés dos-à-dos (divergents) peuvent accumuler des surenroulements négatifs entre eux, ce qui va mutuellement influencer leurs niveaux respectifs de transcription. C'est également le cas pour les unités transcriptionnelles placées dans des directions convergentes, qui vont accumuler des surenroulements positifs dans la région intergénique, toujours avec un impact probablement significatif sur la transcription. Enfin, les promoteurs placés l'un après l'autre dans la même direction (en tandem) vont influencer les gènes en amont et en aval. D'un point de vue uniquement topologique, l'organisation des gènes en tandem semble celle permettant le moins d'accumulation de surenroulements, car le surenroulement induit par un gène sera contré par un surenroulement opposé induit par l'expression du gène voisin. Ce phénomène peut expliquer certains arrangements géniques au sein des génomes, et devrait être pris en compte lors de la conception de nouvelles constructions plasmidiques ou chromosomiques (276,277). Cependant, ces aspects n'ont pas été assez profondément étudiés pour comprendre les phénomènes en jeu et en déduire des schémas généraux.

2.5.3 Régulation de la topologie

Les problèmes de topologie semblent être traités principalement par l'intervention d'enzymes appelées topoisomérases. Cette idée est appuyée par nombreux modèles dressés à partir de données expérimentales (278). Les topoisomérases agissent également dans d'autres processus impliquant une ouverture du double brin d'ADN, tel que la réplication.

Les topoisomérases sont des enzymes qui vont engendrer des coupures simple ou double brins dans l'ADN. Le type de coupure réalisée dans l'ADN par ces enzymes est à l'origine de leur classification : les topoisomérases de type I coupent un seul brin (279–282), tandis que les topoisomérases de type II coupent les deux brins (283–286). Les coupures générées vont permettre de supprimer l'accumulation de surenroulement. Il y a plusieurs sous familles pour chaque type de topoisomérases, classées en fonction de leur séquence protéique ou de leur mécanisme d'action. Les topoisomérases de type I sont divisées en trois sous-familles : IA, IB, et IC. Les topoisomérases de type II sont divisées en deux sous-familles : IIA et IIB.

Chez *E. coli*, il y a quatre topoisomérases : deux topoisomérases de type IA (topoisomérase I et topoisomérase III) et deux topoisomérases de type IIA (topoisomérase IV et gyrase). Malgré quelques fonctions communes, chacune de ces topoisomérases semble être optimisée pour accomplir ses propres actions.

2.5.3.1 Topoisomérases de type IA

Les topoisomérases de type IA sont des monomères qui vont cliver un seul des deux brins ADN. Suite au clivage, elles vont lier de manière covalente une extrémité du brin coupé grâce à une liaison 5' phosphodiester sur un résidu tyrosine du site actif. Afin de pouvoir relâcher les contraintes topologiques accumulées dans l'ADN, il est essentiel de pouvoir faire pivoter un brin autour de l'autre. La structure de la topoisomérase I montre plusieurs domaines, dont un permet la liaison à l'ADN. Cette enzyme va cliver un brin de l'ADN et rester liée aux deux extrémités issues du clivage en formant un pont, par lequel le brin intact va être passé. Ce mécanisme est appelé « enzyme-bridging model » (287,288) (Figure 21). Ces réactions nécessitent la présence d'ions magnésium (Mg^{2+}) et d'une région simple brin exposée dans l'ADN cible. Elles sont actives sur les plasmides seulement si ceux-ci contiennent des surenroulements négatifs. Leur action entraîne une relaxation de l'ADN d'un tour par réaction. Les membres de la sous-famille de topoisomérase IA se distinguent par deux caractéristiques qui reflètent le rôle qu'elles occupent dans la cellule. D'abord, ces enzymes se distinguent par le degré de surenroulement négatif de l'ADN sur lequel elles agissent. Par exemple, la topoisomérase I peut relaxer topologiquement des plasmides isolés mais pas la topoisomérase III qui nécessite une matrice avec un niveau de surenroulement très important (289–291). En fait, la dépendance vis à vis du niveau de surenroulement de l'ADN matrice dépend de la facilité

qu'aura l'enzyme à séparer les brins d'ADN (292). *In vivo*, le niveau de surenroulement final, c'est-à-dire celui qui doit être atteint suite à l'action de la topoisomérase, semble également très important. Ainsi, dans le cas de la topoisomérase I, l'enzyme n'agit qu'à partir du moment où le surenroulement dépasse un certain seuil, et son action s'arrête quand le surenroulement revient au seuil de base. L'action des topoisomérases semble dépendante de la présence d'hélicases, dont le rôle est de catalyser l'ouverture du double brin d'ADN.

Le second point qui permet de distinguer les membres de la sous-famille IA est leur efficacité à relâcher les supertours dans l'ADN en fonction de leur capacité de décaténation (ouverture des deux brins ADN). La topoisomérase III est bien plus efficace pour la décaténation que la topoisomérase I et serait donc préférentiellement impliquée dans la décaténation des génomes en fin de réplication (289,293). Contrairement à la topoisomérase I, la topoisomérase III possède la capacité de cliver et d'entraîner la décaténation de molécules d'ARN (294). Le mécanisme de décaténation de l'ADN réalisé par la topoisomérase III est différent de celui de la topoisomérase I. La présence d'une « boucle de décaténation » dans la structure de la topoisomérase III permet l'entrée et la liaison d'une molécule d'ADN double brin, ensuite, par une seconde liaison, l'enzyme va se lier à un segment d'ADN simple brin (295). Après clivage simple brin, l'ADN double brin initialement lié dans la boucle de décaténation est passé à travers la cassure. La boucle de décaténation va ensuite permettre la ligation de l'ADN et donc la reformation du double brin. Cela correspond à un cycle de décaténation.

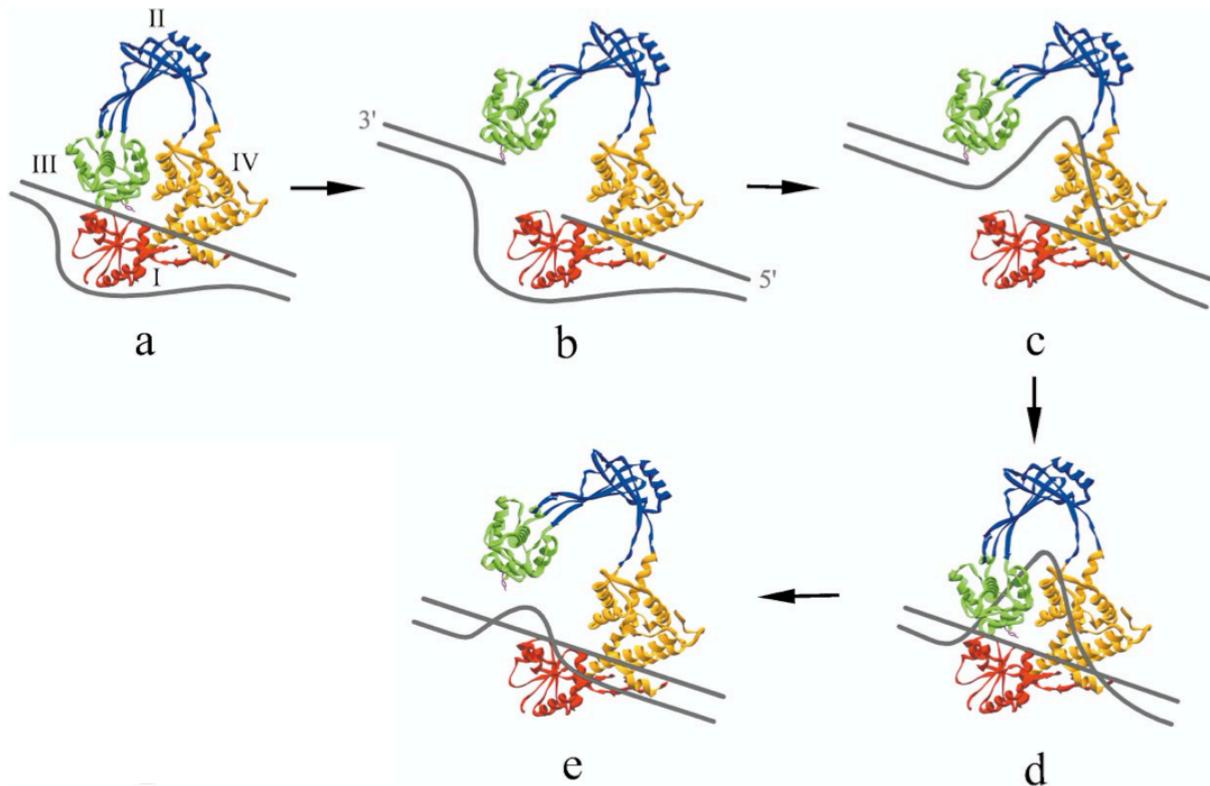


Figure 21 : Un cycle de décaténation par une Topoisomérase I (296)

a. Liaison de la Topoisomérase I à l'ADN. b. Clivage de l'ADN par la Topoisomérase I. c. Passage du brin intact dans le site de clivage. d. Ligation du brin clivé. e. La Topoisomérase I relâche l'ADN. L'enzyme est représentée par ses 4 domaines, l'ADN est représenté par les 2 traits gris.

2.5.3.2 Topoisomérases de type II

2.5.3.2.1 Généralités

Plusieurs caractéristiques sont partagées par toutes les topoisomérases de type II (292). D'abord, ce sont des enzymes dimériques qui vont se lier à l'ADN et couper les deux brins en laissant des extrémités cohésives longues de 4 bases (297). Ensuite, le clivage qu'elles réalisent est permis par une liaison covalente de chaque sous-unité de l'enzyme à l'extrémité 5' de l'ADN. Puis, un changement conformationnel va séparer les 2 extrémités d'ADN au niveau du site de clivage pour créer une ouverture, appelée segment-G. Une seconde région de l'ADN double brin, appelée segment-T est alors transportée à travers le segment-G (298,299). Une fois ce transport effectué, l'ADN est religaturé et le cycle peut reprendre. Ces réactions nécessitent du Magnésium (Mg^{2+}) et l'hydrolyse d'ATP (300,301). D'autres caractéristiques partagées par les topoisomérases de type II sont que les tyrosines du site actif sont localisées

dans les motifs hélice-boucle-hélice (297,302) et la présence d'un résidu arginine très conservé qui est impliqué dans la catalyse grâce à sa proximité avec le site actif tyrosine (297,303).

Parmi toutes les topoisomérases de type II présentes chez les différents organismes, la gyrase est la seule enzyme connue, à ce jour, capable d'hydrolyser l'ATP et d'introduire des supertours négatifs dans l'ADN.

2.5.3.2.2 Gyrase

La gyrase est une enzyme ubiquitaire qui a été découverte en 1976 à la suite de travaux sur des molécules antibactériennes et sur le bactériophage λ (304). L'enzyme de *E. coli* la plus étudiée à ce jour, est codée par les genes *gyrA* et *gyrB*. Elle est composée de deux sous-unités GyrA (97 kDa chacune) et de deux sous-unités *gyrB* (90 kDa chacune) donnant un hétérotétramère A_2B_2 (305,306). Dans bon nombre d'organismes comme *Bacillus subtilis*, les gènes *gyrA* et *gyrB* sont contigus (307). *E. coli* fait exception à cette règle, *gyrA* se trouvant dans la région proche du terminus et *gyrB* dans la région proche de l'origine de réplication, à quelques 2 Mpb. Bien que toutes les topoisomérases de type II nécessitent l'énergie obtenue par l'hydrolyse de l'ATP pour libérer les surenroulements, la gyrase est capable d'utiliser cette énergie pour générer des supertours négatifs. Cette propriété est possible grâce à sa sous-unité A, puisque la délétion de celle-ci va entraîner une enzyme capable de relâcher les surenroulements accumulés dans l'ADN mais incapable d'introduire des supertours négatifs (308). En absence d'ATP, la gyrase peut relâcher les surenroulements négatifs également (309,310).

L'introduction de supertours négatifs par la gyrase se fait grâce à un double mécanisme (298). D'abord, une section de l'ADN, appelée Segment G, se lie au domaine N-terminal des sous-unités GyrA (311,312). Cette liaison va induire un mouvement du domaine C-terminal des sous-unités GyrA. Une seconde partie de l'ADN va alors pouvoir s'enrouler autour du domaine C-terminal des sous-unités GyrA (313,314). Cet enroulement va positionner une troisième section ADN, appelée segment T, au niveau du segment G avec un angle d'environ 60° (313,315). Les sous-unités GyrA orientent le segment T de manière à pouvoir induire des surenroulements négatifs. Au total, la gyrase va lier et recouvrir entre 120 et 150 paires de bases de l'ADN (316–318). Bien que la gyrase ne nécessite pas d'ATP pour les étapes d'enroulement de l'ADN et de recrutement du segment T, l'ATP est essentielle pour permettre

le passage du segment T à travers le segment G clivé. L'ATP va se lier au domaine N-terminal de chaque sous-unité GyrB, entraînant un changement conformationnel qui va emprisonner le segment T (319,320). Les tyrosines du site actif de l'enzyme vont former des liaisons phosphates avec le segment G, entraînant un clivage double brin de l'ADN et laissant des extrémités cohésives longues de 4 bases (284,317). Un changement conformationnel va entraîner un mouvement du dimère de GyrA, entraînant avec lui le segment G. Le segment T va passer à travers la cavité causée par le clivage double brin. Le segment G est ensuite ressoudé, et le segment T libéré (321). L'hydrolyse de l'ATP permet de réinitialiser la gyrase, afin qu'elle puisse retrouver sa conformation. Ce cycle permet l'introduction d'un supertour négatif dans l'ADN. Il y a environ 500 molécules de gyrase par cellule chez *E. coli* (252,322,323), ce qui correspond environ au nombre de microdomaines qui structurent l'ADN. Il y a donc environ une gyrase par boucle d'ADN. La liaison de la gyrase à la boucle d'ADN va permettre le relâchement des supertours positifs, ainsi l'ARN polymérase pourra continuer à transcrire les gènes. La dissociation de la gyrase de la boucle d'ADN va entraîner une accumulation de supertours positifs en aval du gène transcrit, entraînant un ralentissement potentiel de l'élongation de la transcription et un arrêt de l'initiation de la transcription.

2.5.4 Autres facteurs impliqués dans le niveau de surenroulement

Le surenroulement a un impact global sur l'expression des gènes comme l'ont démontré études réalisées sur des gènes seuls et sur des données de transcriptomique, dans lesquels l'action des topoisomérases était sauvage ou inhibée. Ces enzymes jouent donc un rôle essentiel dans le maintien du niveau de surenroulement et dans l'expression des génomes. Potentiellement, tous les promoteurs de la bactérie peuvent être impactés par un changement dans le niveau de surenroulement. Cependant, pour que ce surenroulement puisse avoir un rôle de régulateur global de la transcription, il faut que le niveau de surenroulement varie de façon dynamique à des changements dans l'environnement. Ceci semble effectivement être le cas chez les bactéries puisque des changements d'environnements (osmolarité, milieu de culture, phase de croissance, température, pH, ...) sont accompagnés par des changements dans le surenroulement et l'expression de nombreux gènes (324–328).

2.5.5 R-loops

Les R-loops sont des structures d'acides nucléiques à 3 brins constituées d'un hybride ARN/ADN et d'un ADN simple brin (329). Elles peuvent se former en réponse à l'invasion d'un ADN double brin par un ARN simple brin. Ce processus dépend de l'enzyme de la recombinaison homologue RecA et peut aboutir à la formation de nouvelles origines de réplication (330,331). Des R-loops peuvent également être formés pendant la transcription lorsque l'ARN naissant reste hybridé avec le brin d'ADN matrice en amont de l'ARN polymérase qui avance (332). Le brin d'ADN non matrice étant libre, il va pouvoir s'enrouler autour de l'hybride ARN/ADN.

La formation de R-loops a été observée dans tous les génomes où ils ont été recherchés, mais également dans les plasmides bactériens (333,334). Ces structures peuvent se former favorablement si leur énergie à l'équilibre est plus faible que celle de l'ADN de type B. La formation de R-loops nécessite d'abord la séparation du double brin ADN. Si l'énergie demandée pour l'appariement ARN/ADN est plus faible que l'énergie nécessaire pour la formation du duplex ADN alors l'énergie de la structure globale sera réduite. L'analyse des dinucléotides a montré que pour sept des seize combinaisons possibles la balance penche en la faveur des hybrides ARN/ADN (335). Cette analyse suggère que les R-loops se forment favorablement dans les régions du transcrit riches en G ou en G/A, ce qui a été confirmé par des données expérimentales (336,337).

Les souches d'*E. coli* présentant une déficience de Topoisomérase I (permettant de réduire les surenroulements négatifs) montrent une accumulation de R-loops, mettant en avant le rôle des surenroulements négatifs dans la formation de R-loops pendant la transcription (334). Dans ces souches, la surexpression de RNase H, une enzyme qui dégrade spécifiquement l'ARN dans les hybrides ARN/ADN, a permis de restaurer en partie les conséquences de la déficience de Topoisomérase I (338). Ces résultats suggèrent que la formation de R-loops est un phénomène commun chez les procaryotes et qui est régulés par des facteurs pro-R-loops (surenroulement négatif) et des facteurs anti-R-loops (Topoisomérase I, RNase H).

Une possible interprétation de ces observations serait que l'ADN surenroulé négativement en amont de l'unité transcriptionnelle (252) se déforme et permet une introduction de l'ARN naissant ce qui initie la formation d'un R-loop. Cependant, il a été montré que la séparation des deux brins d'ADN se fait préférentiellement au niveau de régions riches en A/T (339) alors que

la formation de R-loops est facilitée par la présence de séquence riches en G, rendant cette hypothèse peu plausible. Une explication plus probable serait que le surenroulement négatif accumulé en amont de l'ARN polymérase pendant la transcription correspondrait à un état fortement stressé de l'ADN, demandant une forte énergie qui serait réduite par l'introduction de R-loops (340). Dans un R-loop, les deux brins d'ADN étant séparés, ils ne s'enlacent plus, ce qui peut permettre aux surenroulements voisins de migrer dans le R-loop et relâcher les tensions dans cette région de l'ADN. La capacité des R-loops à baisser l'énergie demandée pour la stabilisation donne une explication au rôle des surenroulements négatifs dans leur formation et stabilité.

Ces informations dressent un aspect hiérarchique de la régulation génique chez les bactéries, dans lequel le surenroulement négatif global serait au sommet de la pyramide, d'où il peut affecter un très grand nombre de gènes. A l'étage juste en dessous se trouveraient les NAP, qui vont également contrôler l'expression d'un grand nombre de gènes. Puis les facteurs de transcription généralistes, tels que Crp ou Lrp, qui vont induire ou réprimer l'expression de groupes de gènes. Puis les facteurs de transcription dédiés (Lac, Gal, ...). Et enfin, les interactions entre promoteurs proches occupent l'étage inférieur de cette pyramide.

2.6 Mon projet de recherche

Pour étudier l'effet de la position sur l'expression génique et comprendre les variations dues au contexte génomique, notre équipe a introduit dans un travail antérieur une cassette rapportrice dans 209 positions aléatoires du génome de la bactérie *E. coli* par mutagenèse par transposition. Cette cassette nommée BCJ932 (Figure 22) contient, entre deux terminateurs de transcription forts, le gène rapporteur *emGFP* codant une protéine fluorescente sous contrôle d'un promoteur « faible » appelé p1 (induisant une très faible expression génique comparativement à la moyenne d'expression des gènes chromosomiques).



Figure 22 : Représentation de la construction introduite dans le génome par mutagénèse par transposition.

Cassette BCJ932 insérée dans 209 sites dans le chromosome d'E. coli par mutagénèse par transposition. La flèche représente le promoteur, les ronds représentent les terminateurs de transcription bidirectionnels, les rectangles bleus représentent des sites de liaison pour la protéine Lambda-cl, et le gène rapporteur est représenté en vert.

Les mesures de la transcription du gène rapporteur par PCR quantitative (qPCR) dans les 209 positions du génome ont révélé que cette expression génique pouvait varier d'un facteur 1300 en fonction du site d'insertion de la cassette (Figure 23). Cette variation est surprenante car le gène est isolé par des terminateurs de transcription caractérisés expérimentalement comme très efficaces (341).

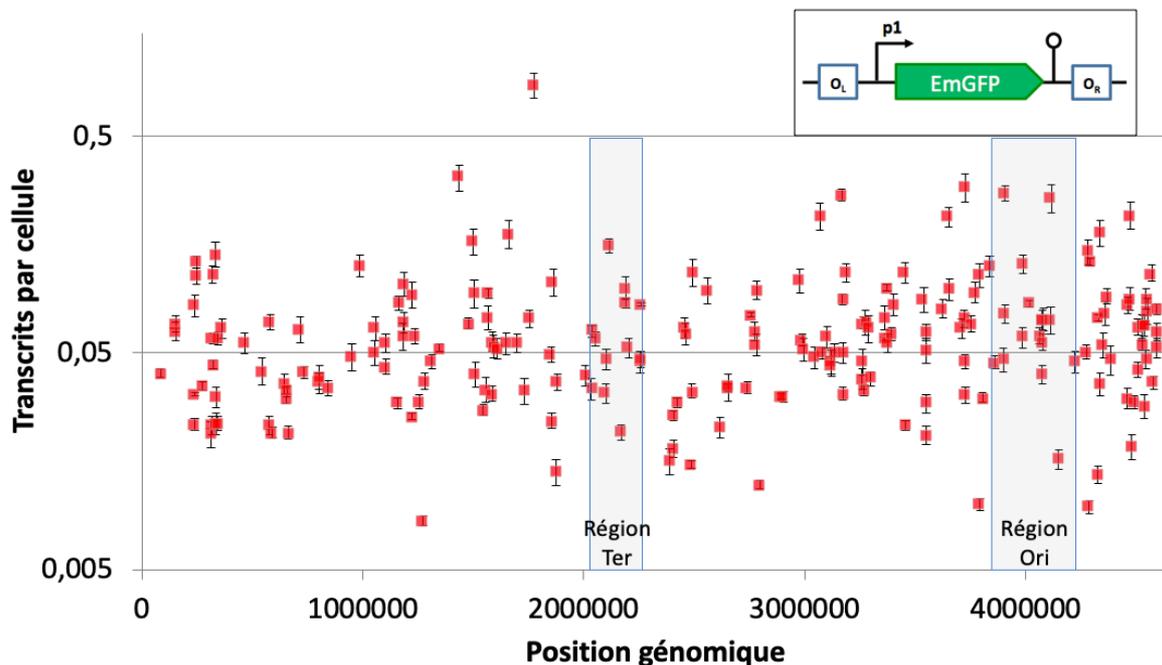


Figure 23 : Transcription du gène *emGFP* dans la cassette BCJ932 dans 209 positions le long du chromosome d'*E. coli*.

*Transcription du gène *emGFP* mesurée par RT-qPCR dans 209 souches portant la construction dans une position unique le long du chromosome. L'axe des abscisses représente le site d'insertion de la construction portant le gène rapporteur *emGFP*. L'axe des ordonnées représente le nombre de transcrits par cellule. Les régions Ori et Ter correspondent aux régions dans lesquelles la suite des travaux a été conduite.*

La même construction, contenant cette fois le gène sous contrôle de promoteurs « forts » (promoteur p3 ou p10) a été introduite dans le génome dans les mêmes loci. Ces promoteurs ont été caractérisés comme induisant une expression élevée du gène devant lequel elles sont placées (342), significativement supérieure à la moyenne d'expression des gènes chromosomiques. Les mesures d'expression ont montré une variation allant jusqu'à un facteur 100 en fonction des positions testées. Ces mesures confirment la variabilité observée avec le promoteur faible, avec toutefois une amplitude moindre.

Nous supposons que ces fortes différences d'expression sont dues au contexte génomique dans lequel la cassette a été introduite. En effet, le génome d'*E. coli* présentant plus de 400 microdomaines avec des topologies et des organisations géniques différentes, l'introduction de la cassette dans l'ADN met celle-ci dans un environnement distinct dans chaque microdomaine et peut-être à chaque loci. Autrement dit, nous postulons que les niveaux d'expressions seraient quasi identiques si le même contexte topologique était uniformément présent.

En réponse à ce postulat, mon projet consiste à étudier l'impact de contraintes topologiques sur la transcription au sein des microdomaines chez *E. coli*. Afin de rendre cette étude possible, j'ai utilisé la protéine Lambda-cl, issue du phage lambda, pour induire artificiellement des boucles dans l'ADN d'*E. coli* et ainsi isoler la cassette rapportrice des contraintes topologiques locales.

3 – RESULTATS

3 RESULTATS

Notre premier objectif a été de tester l'effet d'une boucle locale d'ADN sur l'expression d'un gène placé dans cette boucle. Un outil classique et bien caractérisé pour créer à volonté une boucle locale d'ADN est la protéine Lambda-cl issue du phage lambda. Sa capacité à fixer l'ADN simultanément par chacun de ses 8 monomères lui permet de créer des boucles dans l'ADN de manière spécifique en se liant à 2 régions opératrices O_L et O_R , chacune composée de 3 sites opérateurs. La forte spécificité de la protéine Lambda-cl et l'absence des séquences O_L et O_R dans le chromosome naturel de la souche utilisée permet la formation d'une boucle seulement dans les régions ciblées : la protéine ne se lie qu'aux sites introduits artificiellement dans l'ADN de la bactérie.

3.1 Introduction de la protéine Lambda-cl dans *E. coli*

Afin d'introduire le système de pontage lambda-cl dans la bactérie, une cassette contenant le gène codant Lambda-cl et un gène marqueur de sélection conférant une résistance à la phléomycine (Figure 24) a été introduite dans le génome de la souche sauvage de référence BCJ952 (Tableau 1). Cette cassette produite par PCR, flanquée de part et d'autre par des séquences homologues aux régions cibles dans l'opéron rhamnose, a été intégrée dans le génome de la bactérie à la place des gènes *rhaB*, *rhaA* et *rhaD* comme indiqué dans la Figure 24. Une fois la cassette introduite dans le génome, le gène responsable de la résistance à la phléomycine a été excisé et la souche obtenue a été appelée $\alpha 1$ (Figure 24c) (voir section 4.1.2 pour la construction détaillée de la souche).

L'opéron RhaBAD est un ensemble de cistrons impliqués dans l'utilisation du sucre rhamnose. Le gène *lambda-cl* a été placé sous contrôle du promoteur pRhaB de l'opéron rhamnose. Ainsi, l'addition de rhamnose dans le milieu de culture induit l'expression de la protéine Lambda-cl. La production de Lambda-cl a été vérifiée grâce à une expérience de Western Blot. La souche $\alpha 1$ a été mise en croissance dans un milieu minimum avec des concentrations croissantes de rhamnose (Figure 25). Les résultats en Western Blot montrent des bandes pour les concentrations en rhamnose supérieures à 0,6 mM à la taille attendue (26 kDa) pour la protéine

Lambda-cl. Dans la suite des travaux, 10 mM de rhamnose ont été ajoutés dans le milieu de cultures des souches pour induire l'expression de *lambda-cl*.

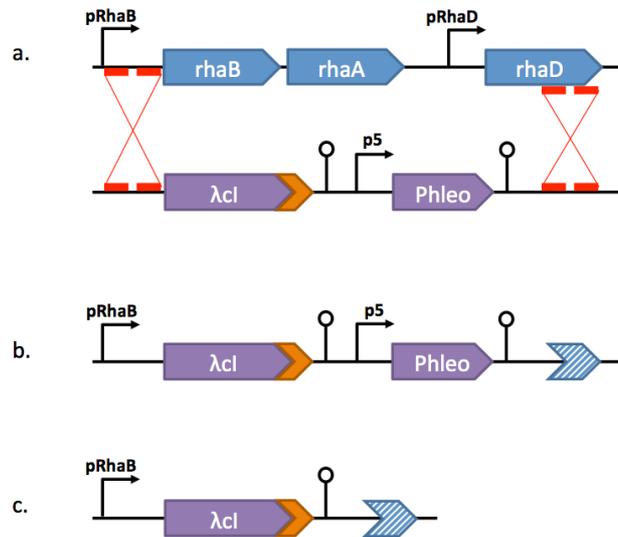


Figure 24 : Construction de la souche exprimant *lambda-cl* sous contrôle du promoteur P_{rha} .

a. Recombinaison homologue au niveau de l'opéron RhaBAD remplacé par la construction comprenant la protéine hétérologue Lambda-cl et un gène de résistance à la phléomycine. La recombinaison homologue permet le remplacement de la région comprise entre le promoteur $pRhaB$ et le gène *rhaD*, laissant ce dernier tronqué. b. Représentation du site d'insertion après recombinaison. c. Représentation du site d'insertion final après excision du gène de résistance pour la phléomycine. La souche obtenue est appelée $\alpha 1$. Les flèches angulaires représentent les promoteurs, les épingles représentent les terminateurs de transcription, les flèches pleines représentent les gènes, la flèche pleine orange représente le tag FLAG fusionné en C-terminal à la protéine Lambda-cl, la flèche pleine hachurée représente le gène tronqué *rhaD*, les lignes noires représentent l'ADN, et les lignes rouges représentent les sites de recombinaison.

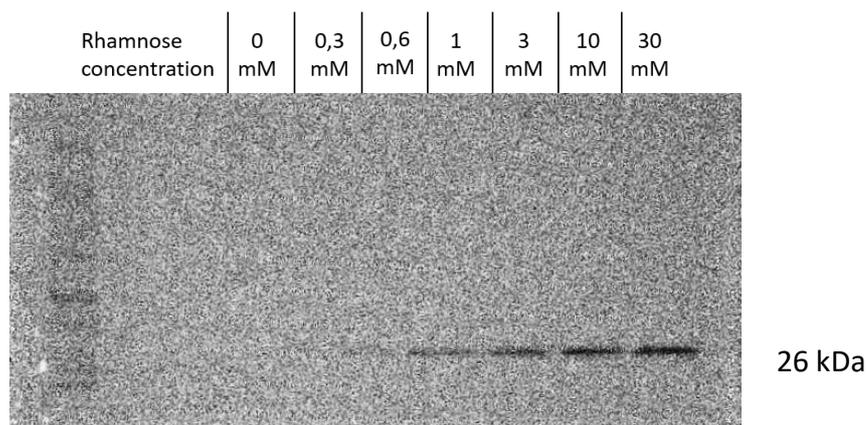


Figure 25 : Gel de migration de Western Blot montrant l'expression de *lambda-cl*.

Expression de la protéine Lambda-cl dans la souche $\alpha 1$ en croissance dans un milieu M9 + glycérol en présence de différentes concentrations de rhamnose (de 0 à 30mM). L'anticorps primaire cible le tag FLAG, l'anticorps secondaire détecté par chimiluminescence cible l'anticorps primaire.

3.2 L'expression de la cassette BCJ932 dépend uniquement du promoteur p1

Afin de vérifier qu'aucune transcription provenant des séquences de part et d'autre de la cassette n'interfère pas avec les mesures de la transcription du gène *emGFP*, des souches contrôles ont été construites. Ces souches arborent la cassette BCJ932 avec une séquence aléatoire de 44bp en remplacement de la séquence promotrice. Deux souches arborant cette cassette à une position chromosomique différentes ont été construites. La présence de transcription résiduelle a été mesurée par RT-qPCR pour les deux directions dans les loci 4 046 827 et 4 158 229 en absence et en présence de la protéine Lambda-cl (et donc d'une boucle d'ADN contenant le gène rapporteur, voir résultats présentés dans la Figure 43 ci-après). Les résultats ont montré des niveaux d'expression compris quasi-nuls, entre 0,003 transcrits par cellules (TPC) et 0,0003 TPC. Ces niveaux d'expression extrêmement faibles montrent que la transcription du gène *emGFP* dans la cassette BCJ932 dépend uniquement du promoteur p1.

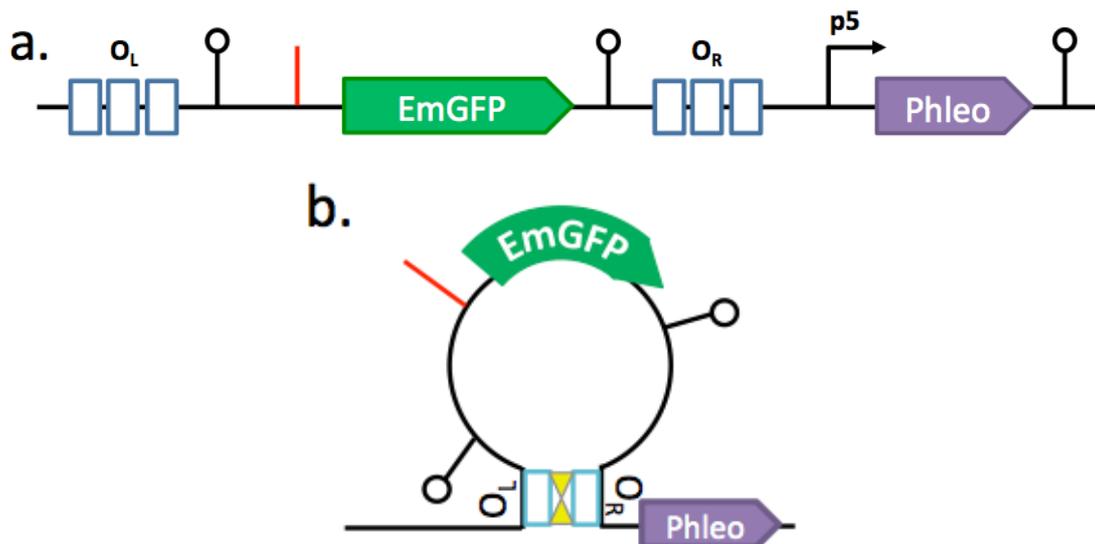


Figure 26 : Représentation du gène *emGFP* sous contrôle du promoteur muté.

a. Représentation de la cassette arborant emGFP en aval de la séquence aléatoire remplaçant le promoteur p1 en absence de la protéine Lambda-cl. b. Représentation de la cassette arborant emGFP en aval de la séquence aléatoire remplaçant le promoteur p1 en présence de la protéine Lambda-cl. Le trait rouge représente le promoteur muté, les épingles représentent les terminateurs de transcription, les flèches pleines représentent les gènes, les lignes et cercles noirs représentent l'ADN, les rectangles bleus représentent les sites de liaisons pour la protéine Lambda-cl, et les triangles jaunes représentent la protéine Lambda-cl.

3.3 L'induction de la boucle d'ADN homogénéise l'expression du gène rapporteur *emGFP* sous contrôle du promoteur faible p1

Dans le but d'étudier l'effet de la formation d'une boucle d'ADN sur l'expression de la cassette BCJ932, cette cassette a été insérée à différentes positions dans le chromosome d'*E. coli* (Figure 27). Deux régions chromosomiques ont été ciblées, une proche de l'origine de réplication, appelée région Ori longue de 300 kpb, et une proche du terminus, appelée région Ter longue de 100 kpb. Ces deux régions ont été sélectionnées en se basant sur les résultats obtenus avec la banque de 209 mutants obtenus par mutagenèse par transposition (Figure 23). L'intérêt porté à la région Ter résulte de la présence de 24 sites d'insertion intergéniques par transposition du gène rapporteur dans un segment de moins de 100 kpb, ce qui représente un taux d'insertion ~5 fois supérieur par rapport au reste du génome. L'intérêt porté à la région Ori est dû aux niveaux d'expressions globaux supérieurs par rapport au reste du génome. Ces niveaux élevés ne peuvent pas être expliqués par la seule proximité avec l'origine de réplication (effet de dosage de gènes dû à la réplication (343)) puisque les conditions de cultures sont lentes et que l'effet de dosage de gènes provoqué par une faible activité de réplication est négligeable. De plus, d'autres régions proches de l'origine de réplication présentent des niveaux d'expression significativement plus faibles.

Dans chacune de ces deux régions, 8 sites d'insertion ont été sélectionnés pour la suite de l'étude. Dans la région Ter, le nombre d'insertions de la cassette étant élevé, les sites d'intérêt ont été sélectionnés en se basant sur leur proximité relative et/ou leur niveau d'expression (Figure 23). Les sites présentant les niveaux d'expression les plus extrêmes ou situés à moins de 10kbps (moyenne de la taille d'un microdomaine) ont été privilégiés.

Le gène *lambda-cl* a été introduit dans ces 16 souches, et l'expression du gène *emGFP* a été mesurée en présence de la protéine Lambda-cl. Les niveaux d'expression mesurés par RT-qPCR en présence de Lambda-cl (et donc d'une boucle dans l'ADN) ont été comparés aux niveaux d'expression mesurés en absence de boucle dans l'ADN (déjà présentés en Figure 23). Ces résultats sont présentés dans la Figure 28.

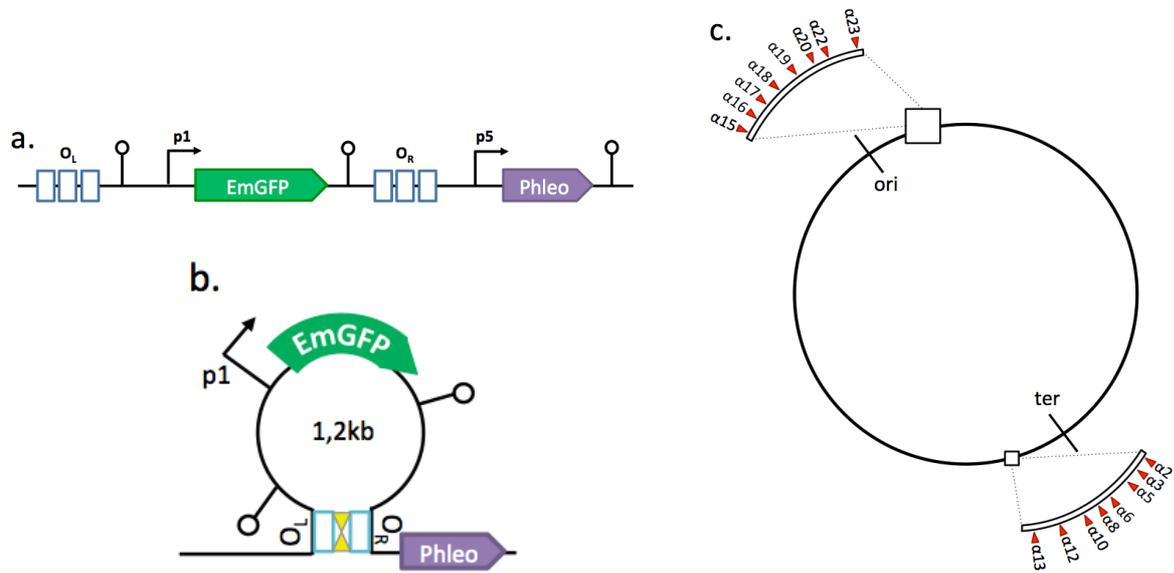


Figure 27 : Représentation de la construction 932 et ses sites génomiques d'insertion

a. Construction 932 contenant le gène rapporteur emGFP bordé de part et d'autre par des terminateurs de transcription et des sites de liaison pour la protéine Lambda-cl, et portant dans sa région 3' le marqueur de sélection conférant une résistance à la phléomycine. b. La fixation de la protéine Lambda-cl sur ses 2 sites de liaison (O_L et O_R) entraîne la formation d'une boucle dans l'ADN. c. Sites d'insertion de la construction dans les régions Ori et Ter. Les flèches angulaires représentent les promoteurs, les épingles représentent les terminateurs de transcription, les flèches pleines représentent les gènes, les lignes noires représentent l'ADN, les rectangles bleus représentent les sites de liaisons pour la protéine Lambda-cl, les triangles jaunes représentent la protéine Lambda-cl, et les triangles rouges représentent les différents loci d'insertion de la construction 932 et le nom des souches correspondantes est indiqué.

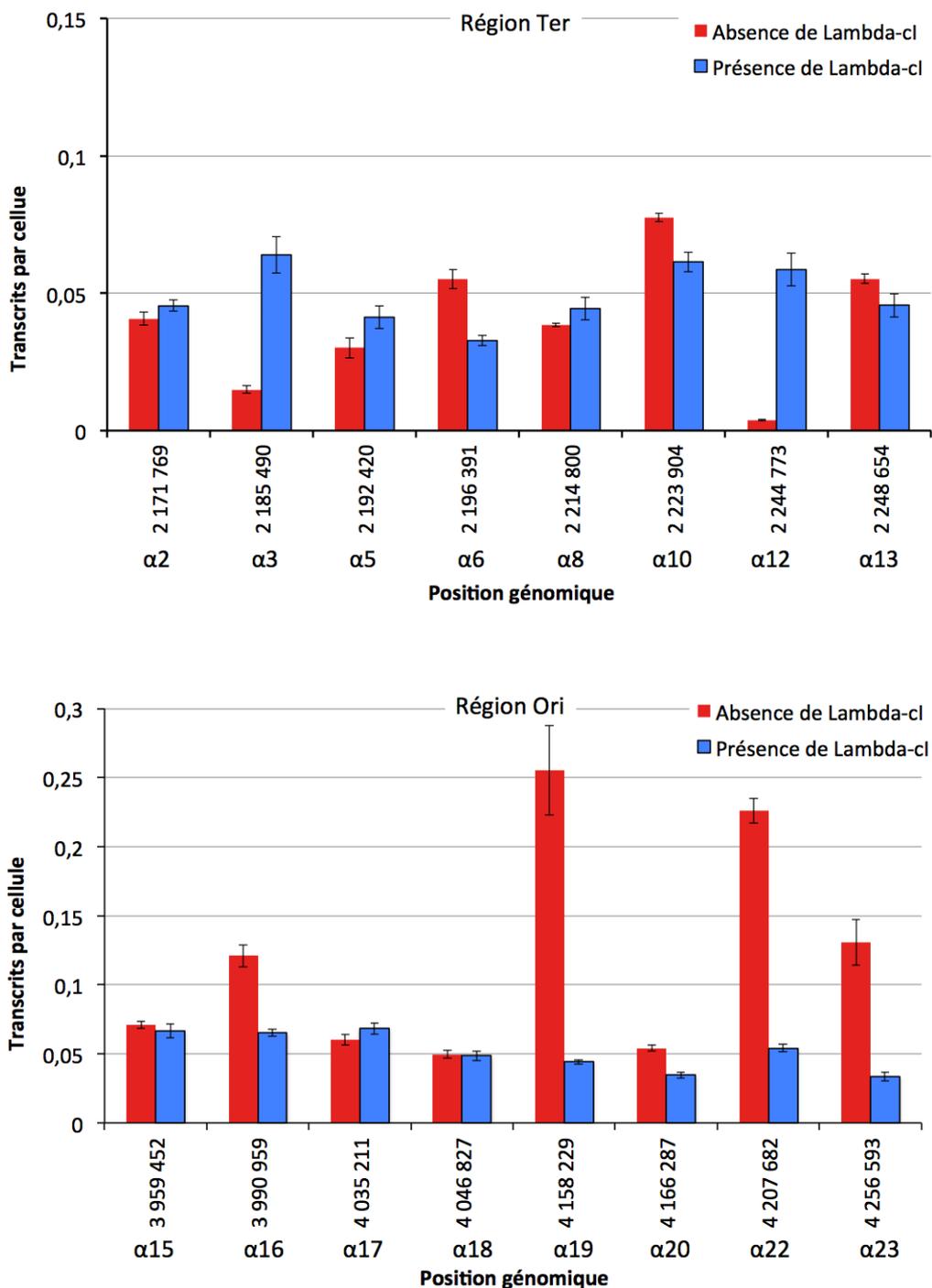


Figure 28 : Impact de la formation d’une boucle d’ADN sur l’expression du gène *emGFP* placé sous contrôle du promoteur p1 dans les régions Ori et Ter.

Transcription du gène emGFP sous contrôle du promoteur p1 mesurée par RT-qPCR dans les régions Ori et Ter en fonction du site d’insertion génomique et de la présence ou non de Lambda-cl. L’axe des abscisses représente le site d’insertion de la construction portant le gène rapporteur emGFP ainsi que le nom de la souche. L’axe des ordonnées représente le nombre de transcrits par cellule. Les niveaux d’expression en absence de boucle sont représentés par les barres rouges. Les niveaux d’expression en présence de boucle sont représentés par les barres bleues.

Les niveaux d'expression du gène rapporteur sous contrôle du promoteur faible p1 sont relativement faibles. En absence de Lambda-cl, le nombre de TPC varie entre ~ 0 et 0,08 dans la région Ter et entre 0,05 et 0,25 dans la région Ori. Il y a une forte variabilité de l'expression, jusqu'à un facteur 5 dans chaque région (sans compter la position où l'expression est quasi-nulle). Avec la présence de Lambda-cl, le nombre de TPC varie entre 0,03 et 0,06 dans les 2 régions testées, avec une moyenne autour de 0,05 TPC. L'induction de la boucle entraîne donc (i) une homogénéisation de l'expression de *emGFP* autour de la valeur de 0,05 TPC et (ii) une diminution de la variabilité de l'expression génique intra- et inter-région.

3.4 L'induction de la boucle d'ADN homogénéise l'expression du gène rapporteur *emGFP* sous contrôle du promoteur fort p3

La construction BCJ927 (Figure 29) diffère de la construction BCJ932 par la présence du promoteur fort p3 en lieu et place du promoteur faible p1. Elle a été introduite dans 4 positions chromosomiques, 2 dans la région Ori et 2 autres dans la région Ter. Ces sites d'insertion ont été sélectionnés en se basant sur les résultats obtenus avec la construction sous contrôle du promoteur faible (Figure 28). Ainsi, les positions 2 185 490 et 2 244 773 ont été choisies car l'induction de la boucle d'ADN a entraîné une forte (>5 fois) augmentation des niveaux d'expression. La position 4 158 229 a été choisie car l'induction de la boucle d'ADN a entraîné une forte diminution (5 fois) du nombre de transcrits. Enfin, une position témoin a également été sélectionnée, l'induction de la boucle d'ADN au niveau de la position 4 046 827 n'ayant pas de réel impact sur l'expression. Les niveaux d'expression ont été mesurés par RT-qPCR en absence et en présence de Lambda-cl pour les 4 sites d'insertion (Figure 30).

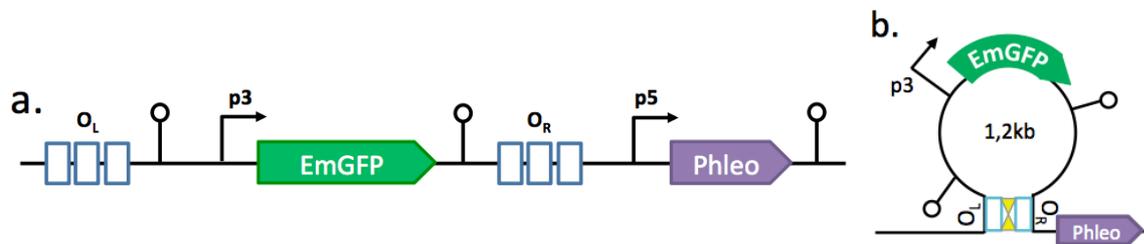


Figure 29 : Représentation de la construction BCJ927.

Construction BCJ927 contenant le gène rapporteur emGFP, sous contrôle du promoteur fort p3, bordé de part et d'autre par des terminateurs transcriptionnels et des sites de liaison pour la protéine Lambda-cl, ainsi que le marqueur de sélection conférant une résistance à la phléomycine. a. Configuration en absence d'expression de lambda-cl. b. Configuration en présence d'expression de lambda-cl. Les flèches angulaires représentent les promoteurs, les épingles représentent les terminateurs de transcription, les flèches pleines représentent les gènes, les lignes et cercles noirs représentent l'ADN, les 3 rectangles bleus représentent les sites de liaisons pour la protéine Lambda-cl, les triangles jaunes représentent la protéine Lambda-cl.

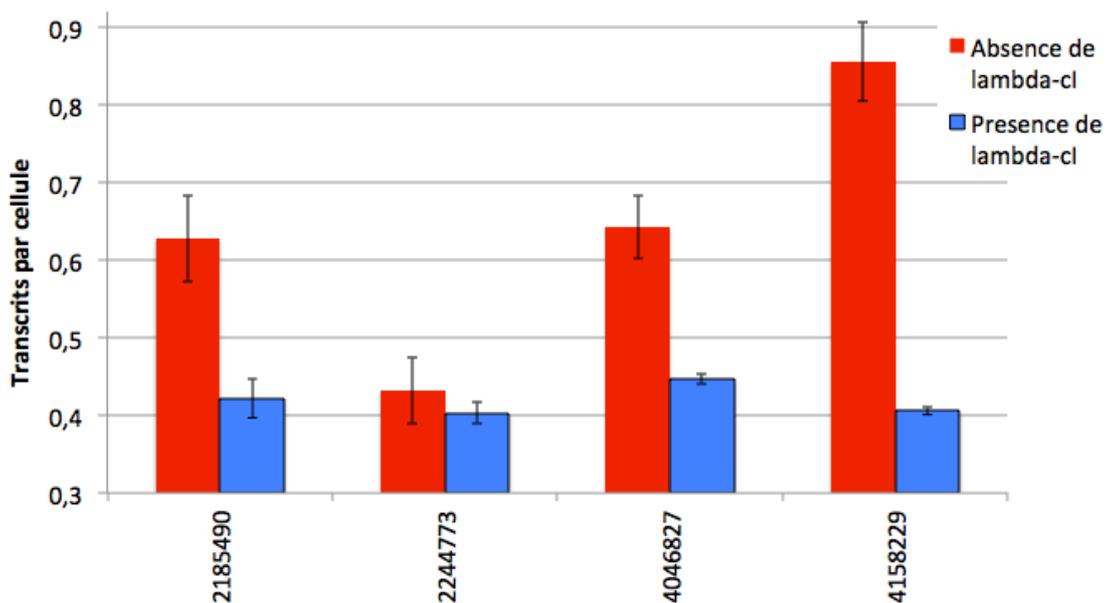


Figure 30 : Impact de la formation d'une boucle d'ADN sur l'expression du gène *emGFP* sous contrôle du promoteur fort p3 dans les régions Ori et Ter.

Expression du gène emGFP sous contrôle du promoteur p3 dans les régions Ori et Ter en fonction du site d'insertion génomique. L'axe des abscisses représente le site d'insertion de la construction portant le gène rapporteur emGFP. L'axe des ordonnées représente le nombre de transcrits par cellule. La présence de Lambda-cl permet l'induction de la boucle d'ADN. Les niveaux d'expression en absence de boucle sont représentés par les barres rouges. Les niveaux d'expression en présence de boucle sont représentés par les barres bleues. Les niveaux d'expression ont été mesurés par RT-qPCR dans les différentes souches testées.

Les niveaux d'expression du gène *emGFP* sous contrôle du promoteur fort p3 sont plus élevés (environ 10 fois) que ceux observés pour le promoteur faible p1. En absence de boucle d'ADN, l'expression du gène rapporteur varie d'un facteur environ 2 en fonction du site d'insertion, les niveaux de transcrits allant de $\sim 0,43$ à $\sim 0,85$ par cellule. L'amplitude de la variation d'expression semble plus faible pour la construction avec le promoteur fort comparé à celle avec le promoteur faible. Comme avec le promoteur p1, c'est le locus 2 244 773 qui présente le niveau d'expression le plus faible avec $\sim 0,43$ TPC, et le locus 4 158 229 qui présente le niveau d'expression le plus fort avec $\sim 0,85$ TPC. L'induction de la boucle d'ADN entraîne généralement une réduction des niveaux d'expression et les homogénéise autour de la valeur moyenne de $\sim 0,42$ TPC. Ces résultats montrent que les promoteurs forts sont également sensibles au contexte. Contrairement aux observations réalisées sur le promoteur p1, ici l'induction de la boucle n'entraîne une augmentation d'expression du gène rapporteur dans aucun des différents loci testés. L'expression semble bridée au niveau de la valeur moyenne $\sim 0,4$ TPC. Nous émettons l'hypothèse que cette expression muselée est causée par l'effet inhibiteur des surenroulements positifs induits par la transcription du gène rapporteur à l'intérieur de la boucle. La suite des travaux a visé à étudier les facteurs internes à la boucle d'ADN capables d'influer la transcription, notamment le rôle du surenroulement positif à l'intérieur de cette boucle.

3.5 La suppression de supertours positifs à l'intérieur de la boucle augmente l'expression du gène *emGFP*

Dans le but d'étudier de tester l'hypothèse selon laquelle l'accumulation de supertours positifs inhibe l'expression de gènes situés au sein d'une boucle, nous avons décidé d'étudier l'effet de la diminution de cette accumulation sur l'expression de *emGFP*. Des différentes topoisomérases étudiées, seule la gyrase est capable d'éliminer les supertours positifs et d'introduire des surenroulements négatifs dans l'ADN. Les travaux de Fisher et ses associés ont permis la découverte de séquences spécifiques impliquées dans la reconnaissance par la gyrase (316). Ces sites de reconnaissance par la gyrase (GRS) ont été observés chez différents organismes. L'étude de ces séquences a permis de déterminer laquelle était la plus efficace pour le recrutement et l'introduction de surenroulement négatifs par la gyrase chez *E. coli* (344,345). Ainsi, nous avons sélectionné le GRS du phage Mu et l'avons introduit dans la boucle

d'ADN. Selon l'effet de la transcription sur la topologie locale (252), la position à l'intérieur de la boucle dans laquelle il serait le plus judicieux d'introduire le GRS pour réduire une accumulation de surenroulement positif est en aval du gène rapporteur, comme indiqué dans la Figure 31a. La taille du GRS est d'environ 150 pb, à laquelle on ajoute environ 100 pb comprenant les éléments techniques permettant l'intégration dans la séquence ADN. L'incorporation de cet élément dans la construction 927 va entraîner une augmentation significative de la taille de la boucle d'ADN (~0,2 kpb). Afin de vérifier qu'un éventuel effet du GRS ne serait pas dû à la seule augmentation de la taille de la boucle, nous avons réalisé un contrôle dans lequel une séquence « neutre » de taille identique à l'élément comprenant le GRS a été introduite dans la même position (Figure 31b). Cet ADN neutre correspond à une séquence aléatoire débarrassée de toute séquence connue pouvant interagir avec les mécanismes cellulaires (sites de reconnaissance pour enzyme de restriction, motifs conservés de promoteurs, séquence Shine-Dalgarno, ...) (346). Enfin, le GRS a également été introduit dans une seconde position à l'intérieur de la boucle, en amont du gène transcrit (Figure 31c). La transcription entraînant des surenroulements négatifs en amont du gène transcrit, l'action de la gyrase dans cette région pourrait augmenter l'accumulation de surenroulement négatif dans ce segment ADN. Dans cette configuration, il est attendu que le GRS ai un impact négatif sur la transcription.

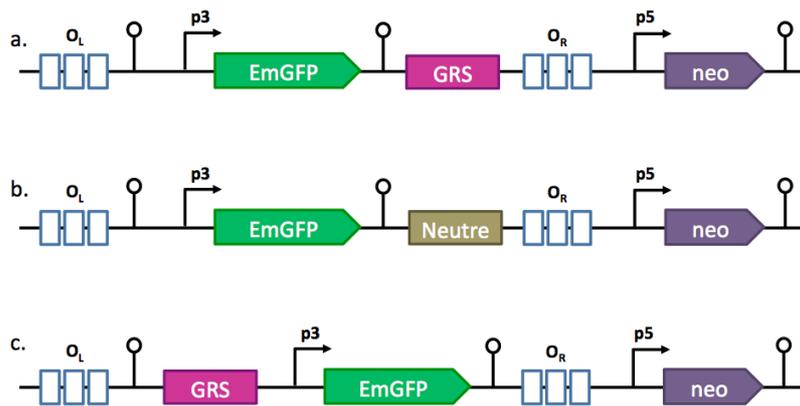


Figure 31 : Representation des constructions contenant le gène *emGFP* et le GRS.

*Constructions contenant le gène rapporteur *emGFP*, sous contrôle du promoteur fort p3 et un site de reconnaissance pour gyrase ou de l'ADN neutre de même taille, le tout bordé de part et d'autre par des terminateurs transcriptionnels et des sites de liaison pour la protéine Lambda-cl. En aval des constructions se trouve un marqueur de sélection neo conférant une résistance à la kanamycine. Les flèches angulaires représentent les promoteurs, les épingles représentent les terminateurs de transcription, les flèches pleines représentent les gènes, les lignes et cercles noirs représentent l'ADN, les 3 rectangles bleus représentent les sites de liaisons pour la protéine Lambda-cl, les rectangles rose et olive représentent respectivement le site de reconnaissance pour gyrase et l'ADN neutre.*

Les différentes constructions ont été insérées dans les 4 mêmes loci étudiés pour l'expression du gène *emGFP* sous contrôle du promoteur fort p3 (Section 3.4). Les mesures de l'expression du gène rapporteur en présence de la boucle d'ADN et d'un GRS ou ADN neutre sont répertoriées dans la Figure 32 et comparées aux résultats obtenus pour le gène rapporteur sous contrôle du promoteur fort p3 sans et avec boucle (Figure 30). Dans les 4 loci testés, l'expression de la construction comprenant l'ADN neutre est homogène autour de la valeur de 0,4 TPC. La présence de l'ADN neutre à l'intérieur de la boucle d'ADN, et donc l'augmentation de la taille de la boucle de 0,2 kpb (soit 20%), n'entraîne pas de différence des niveaux d'expression du gène rapporteur (Figure 32b-c). Lorsque le GRS est présent dans la boucle dans la région en aval du gène rapporteur (Figure 32d), l'expression du gène *emGFP* est comprise entre 0,5 et 0,7 TPC en fonction du site d'insertion de la cassette. La présence du GRS en aval du gène rapporteur permet d'augmenter les niveaux de transcription peu importe le locus génomique. Cependant, cette stimulation de transcription n'est pas corrélée aux variations d'expression observées d'un site à l'autre en absence de boucle (Figure 32a). Lorsque le GRS est inséré en amont du gène rapporteur, le nombre de transcrit est homogène à environ 0,55 TPC dans les 4 sites d'insertion de la cassette. Ces résultats suggèrent que la diminution

de l'accumulation de surenroulement positifs par la présence du GRS dans la boucle (quelque soit sa position par rapport au gène rapporteur) entraîne une augmentation des niveaux d'expression.

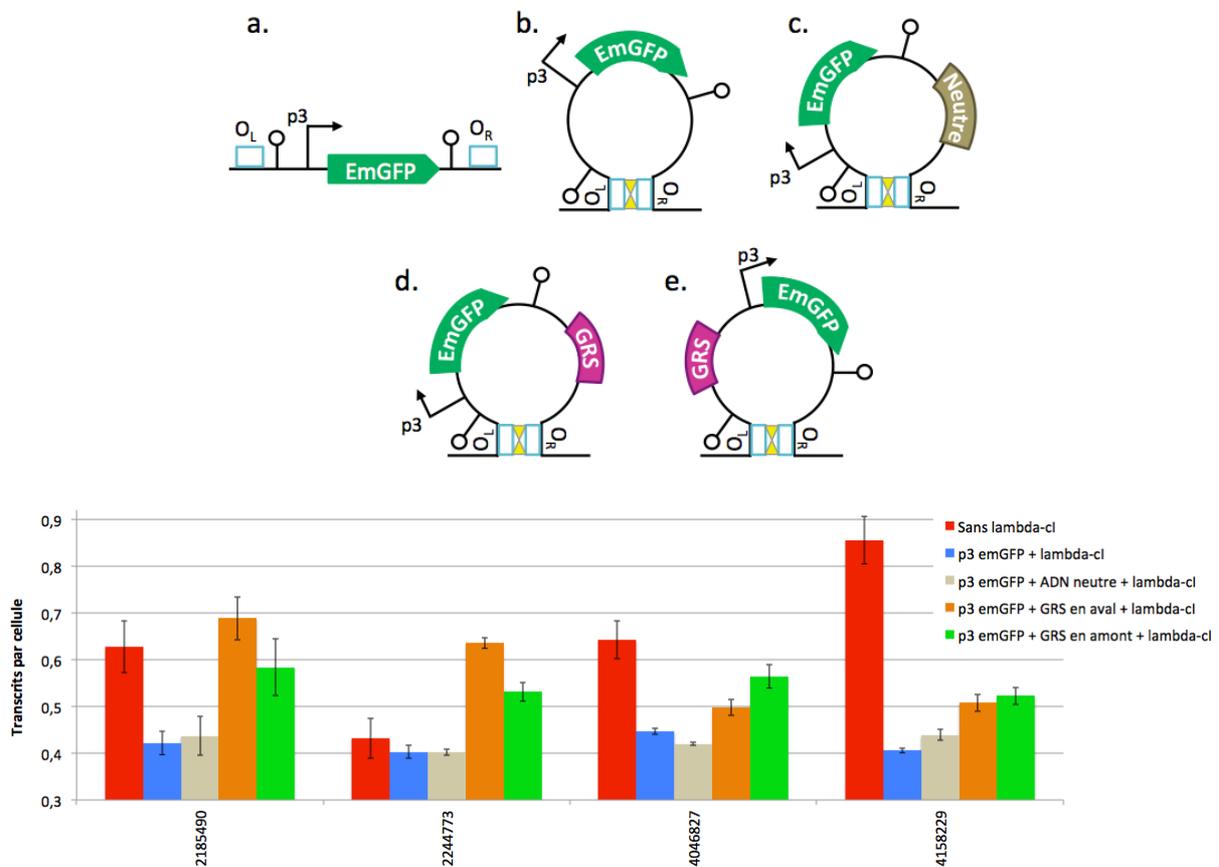


Figure 32 : Impact de la formation d'une boucle d'ADN sur l'expression du gène *emGFP* sous contrôle du promoteur p3 dans les régions Ori et Ter en présence du GRS ou d'un fragment d'ADN neutre.

Les niveaux d'expression du gène *emGFP* ont été mesurés par RT-qPCR en présence d'un site de reconnaissance pour gyrase (GRS) à différentes positions dans la boucle. L'axe des abscisses représente le site d'insertion de la construction portant le gène rapporteur *emGFP*. L'axe des ordonnées représente le nombre de transcrits par cellule. a. Construction 927 en absence de *Lambda-cl*. b. Construction 927 en présence de *Lambda-cl*. c. Construction 927 en présence de *Lambda-cl* dans laquelle un segment d'ADN neutre a été introduit. d. Construction 927 en présence de *Lambda-cl* dans laquelle un GRS a été introduit en aval de *emGFP* dans la boucle. e. Construction 927 en présence de *Lambda-cl* dans laquelle un GRS a été introduit en amont de *emGFP* dans la boucle. Les flèches angulaires représentent les promoteurs, les épingles représentent les terminateurs de transcription, la flèche verte représente *emGFP*, les lignes et cercles noirs représentent l'ADN, les rectangles représentent les sites de liaisons pour la protéine *Lambda-cl*, les triangles jaunes représentent la protéine *Lambda-cl*, les rectangles rose et jaune représentent respectivement le site de reconnaissance pour gyrase et l'ADN neutre.

3.6 L'expression du gène *emGFP* est stimulée lorsque les gènes sont co-orientés

Afin d'étendre nos études à des boucles d'ADN plus complexes, des constructions contenant un second gène dans la boucle d'ADN ont été conçues. L'introduction d'un second gène dans la boucle peut être réalisée sous 3 configurations : (i) les gènes sont en tandem (co-orientés), (ii) les gènes sont convergents (orientés l'un vers l'autre), (iii) les gènes sont divergents (orientés en directions opposées). L'étude de ces 3 configurations géniques va permettre d'éclaircir l'importance mutuelle de l'orientation relative des gènes sur leur expression et d'aborder l'impact des contextes topologiques locaux générés en réponse au « twin-supercoiled domain model ». Nous avons pensé qu'il était judicieux de placer le second gène sous contrôle de différents promoteurs afin de comparer l'impact d'un promoteur faible et d'un promoteur fort sur l'expression du gène rapporteur *emGFP* exprimé à partir du promoteur fort p3. Le second gène introduit dans la boucle est *mCherry*, il code lui aussi pour une protéine fluorescente. Il a été placé sous contrôle du promoteur faible p1 ou d'un promoteur fort p10 (Figure 33). Une construction dans laquelle le promoteur de *mCherry* a été muté (remplacé par une séquence neutre) a également été conçue (Figure 33g). Ces différentes constructions ont été insérées dans 2 positions dans le génome, une dans la région Ori et une dans la région Ter, dans des souches exprimant la protéine Lambda-cl.

Lorsque le gène *mCherry* est placé sous contrôle du promoteur faible (Figure 33a,c,e), l'expression du gène rapporteur *emGFP* est : de ~0,95 TPC dans la configuration en tandem, de ~0,85 TPC dans la configuration convergente, et de ~0,5 TPC dans la configuration divergente (Figure 34). Ces niveaux d'expression sont similaires pour les 2 sites d'insertion. Il est à noter que les niveaux d'expression de *emGFP* dans ces boucles à deux gènes sont de 1 à 2 fois supérieurs à ce qui a été observé avec une boucle portant uniquement *emGFP* sous p3. Dans les conditions où le second gène est peu exprimé, la configuration en tandem est celle permettant la plus forte stimulation du gène *emGFP*.

Lorsque *mCherry* est placé sous contrôle du promoteur fort p10 (Figure 33b,d,f), l'expression du gène rapporteur *emGFP* est de ~1,7 TPC dans la configuration tandem, de ~0,6 TPC dans la configuration convergente, et de ~1 TPC dans la configuration divergente (Figure 35). Ces niveaux d'expression sont similaires pour les 2 sites d'insertion. Lorsque le second gène à

l'intérieur de la boucle est très exprimé, la configuration en tandem est à nouveau la plus favorable pour l'expression de *emGFP*.

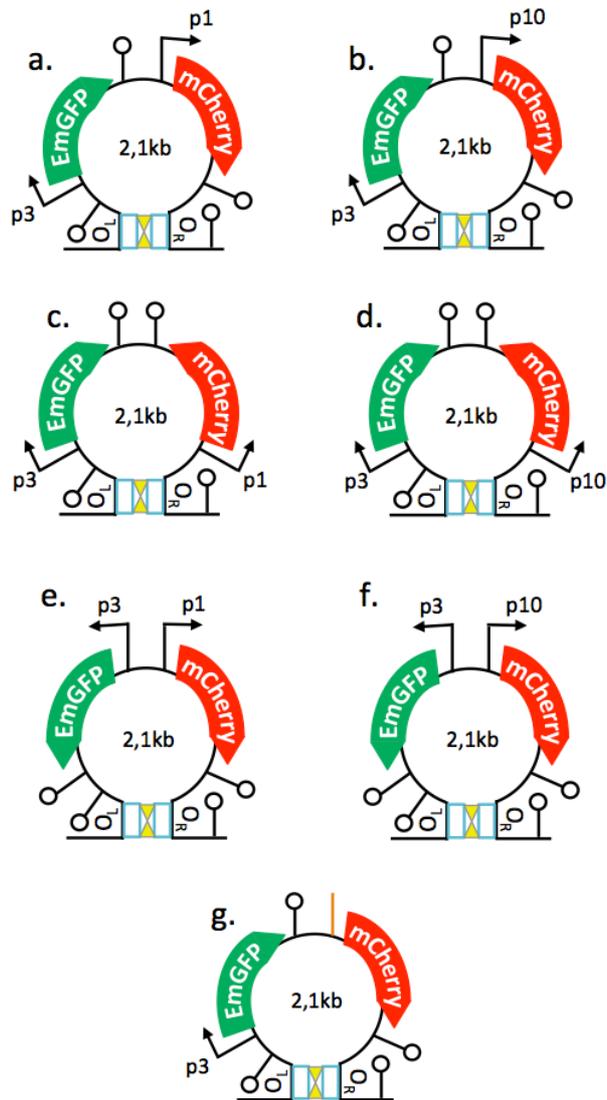


Figure 33 : Représentation des constructions arborant 2 gènes à l'intérieur de la boucle d'ADN.

a. Construction comprenant les gènes en tandem avec *mCherry* sous contrôle du promoteur faible p1. *b.* Construction comprenant les gènes en tandem avec *mCherry* sous contrôle du promoteur fort p10. *c.* Construction comprenant les gènes convergents avec *mCherry* sous contrôle du promoteur faible p1. *d.* Construction comprenant les gènes convergents avec *mCherry* sous contrôle du promoteur fort p10. *e.* Construction comprenant les gènes divergents avec *mCherry* sous contrôle du promoteur faible p1. *f.* Construction comprenant les gènes divergents avec *mCherry* sous contrôle du promoteur fort p10. *g.* Construction comprenant les gènes en tandem avec une séquence aléatoire remplaçant le promoteur du gène *mCherry*. Les flèches angulaires représentent les promoteurs, les épingles représentent les terminateurs de transcription, la flèche verte représente *emGFP*, la flèche rouge représente *mCherry*, les lignes et cercles noirs représentent l'ADN, les rectangles représentent les sites de liaisons pour la protéine λ -cl, et les triangles jaunes représentent la protéine λ -cl.

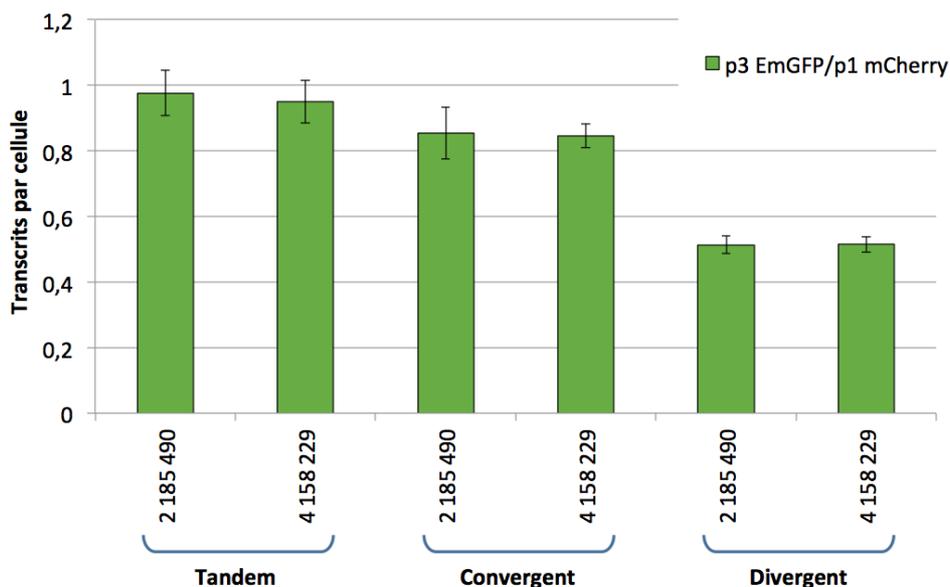


Figure 34 : Niveaux d'expression du gène *emGFP* dans les 3 configurations et lorsque le gène *mCherry* est faiblement transcrit.

Les niveaux d'expression du gène emGFP ont été mesurés par RT-qPCR en présence d'un second gène sous contrôle d'un promoteur faible dans la boucle d'ADN. L'axe des abscisses représente le site d'insertion de la construction et l'orientation relative des gènes. L'axe des ordonnées représente le nombre de transcrits par cellule.

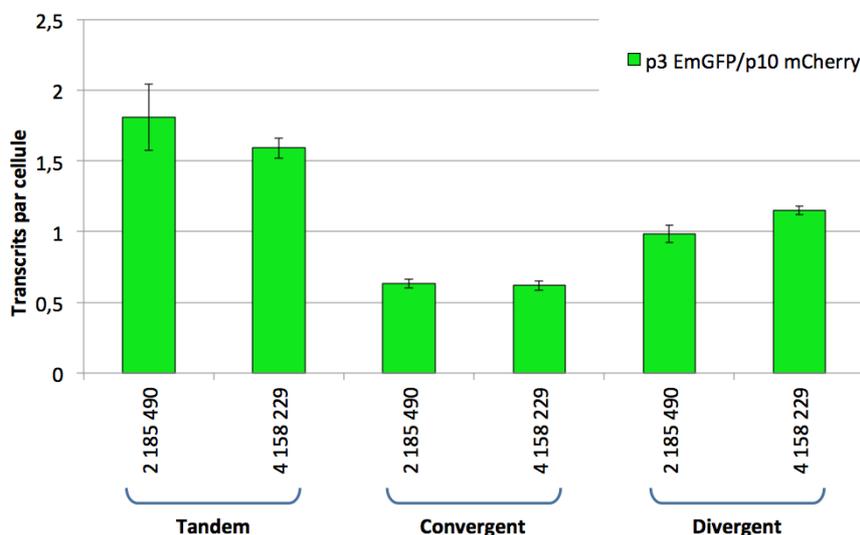


Figure 35 : Niveaux d'expression du gène *emGFP* dans les 3 configurations et lorsque le gène *mCherry* est fortement transcrit.

Les niveaux d'expression du gène emGFP ont été mesurés par RT-qPCR en présence d'un second gène sous contrôle d'un promoteur fort dans la boucle d'ADN. L'axe des abscisses représente le site d'insertion de la construction et l'orientation relative des gènes. L'axe des ordonnées représente le nombre de transcrits par cellule.

La comparaison configuration par configuration, et pour les deux loci, des données d'expression de *emGFP* en présence de *mCherry* sous contrôle d'un promoteur faible, fort ou d'une séquence mutée permet d'observer l'effet de la force d'un promoteur avoisinant sur l'expression de *emGFP* à l'intérieur d'une boucle d'ADN (Figure 36). Ainsi, on constate que dans la configuration en tandem, l'expression de *emGFP* augmente avec la force du promoteur de *mCherry*. Cet effet est plus fort dans la configuration en tandem que dans la configuration divergente. Par contre, dans la configuration convergente, un résultat inverse est observé : l'expression de *emGFP* diminue d'environ 25% quand l'expression de *mCherry* augmente.

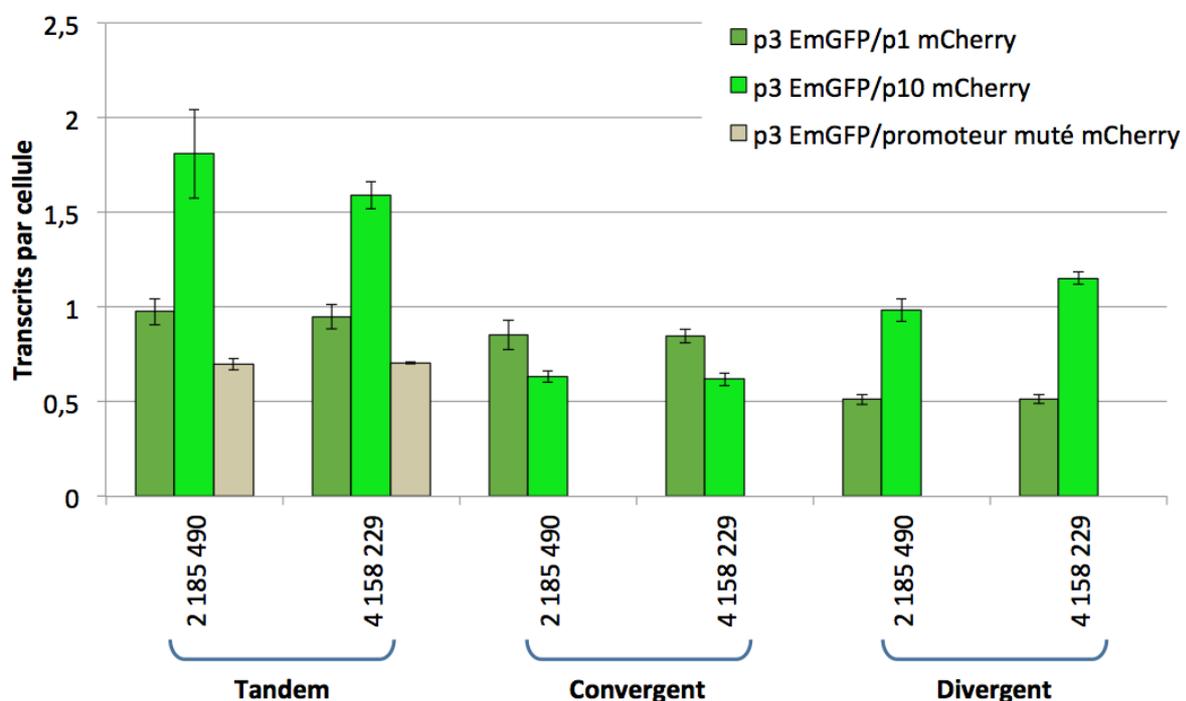


Figure 36 : Comparaison des niveaux d'expression du gène *emGFP* dans les 3 configurations, en fonction du niveau de transcription du gène *mCherry*.

Comparaison des niveaux d'expression du gène emGFP en présence d'un second gène sous contrôle d'un promoteur faible, fort ou muté dans la boucle d'ADN. L'axe des abscisses représente le site d'insertion de la construction et l'orientation relative des gènes. L'axe des ordonnées représente le nombre de transcrits par cellule.

La Figure 37 rassemble les niveaux d'expression de la protéine *mCherry* sous contrôle de p1 et de p10 dans les 3 configurations et aux deux loci. Dans leur ensemble, ces données confirment que le promoteur p10 est plus fort que p1. Cependant, cette différence dépend fortement de l'orientation relative de *mCherry* et *emGFP* : alors que l'activité du promoteur p10

varie faiblement (1,5 fois) d'une configuration à l'autre, celle du promoteur p1 est fortement réduite (10 fois) dans la configuration convergente par rapport à la configuration en tandem et quasi nulle dans la configuration divergente (l'activité est au moins 100 plus faible en configuration divergente par rapport à tandem). Lorsque le promoteur est muté, l'expression de *mCherry* est de $\sim 0,05$ TPC dans la configuration en tandem. Comme observé pour le gène *emGFP*, la configuration en tandem semble être celle permettant la plus forte expression du gène *mCherry*, et la configuration convergente semble celle causant la plus mauvaise expression pour des gènes fortement exprimés.

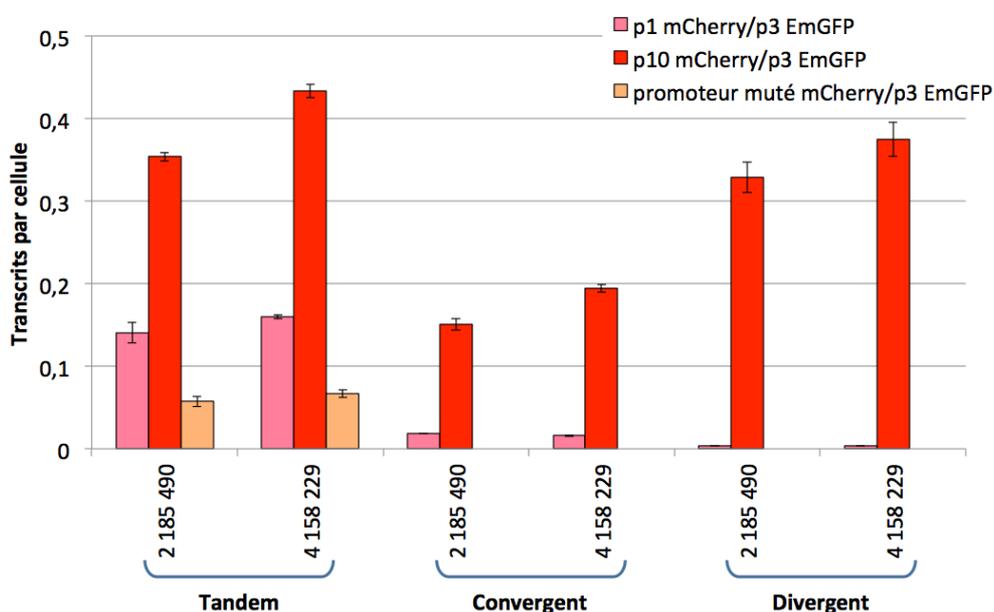


Figure 37 : Niveaux d'expression du gène *mCherry* en fonction du promoteur et de l'orientation relative par rapport au gène *emGFP*.

Les niveaux d'expression du gène mCherry sous contrôle d'un promoteur faible, fort, ou muté ont été mesurés par RT-qPCR lorsque le gène emGFP est fortement exprimé dans la boucle d'ADN. L'axe des abscisses représente le site d'insertion de la construction et l'orientation relative des gènes. L'axe des ordonnées représente le nombre de transcrits par cellule.

3.7 La diminution des supertours positifs dans une boucle à deux gènes a un effet positif global sur l'expression

La réduction des supertours positifs dans les constructions comprenant le gène *emGFP* seul à l'intérieur de la boucle grâce à l'utilisation d'un GRS augmente les niveaux de transcription de *emGFP* (Figure 32). Afin d'étendre cette étude à des boucles plus complexes,

un GRS a été intégré dans la région intergénique de boucles contenant 2 gènes. L'un de ces gènes (*emGFP*) était sous contrôle du promoteur fort p3, l'autre (*mCherry*) sous contrôle du promoteur fort p10 (Figure 38). Plus un gène est transcrit, plus il va générer des surenroulements (négatifs en amont de l'ARN polymérase, et positifs en aval). En ciblant les boucles d'ADN dans lesquelles les 2 gènes sont sous contrôle de promoteurs forts et dans les différentes organisations (tandem, convergent, divergent) nous avons pu étudier les constructions présentant les plus fortes accumulations locales de surenroulement positif ou négatif.

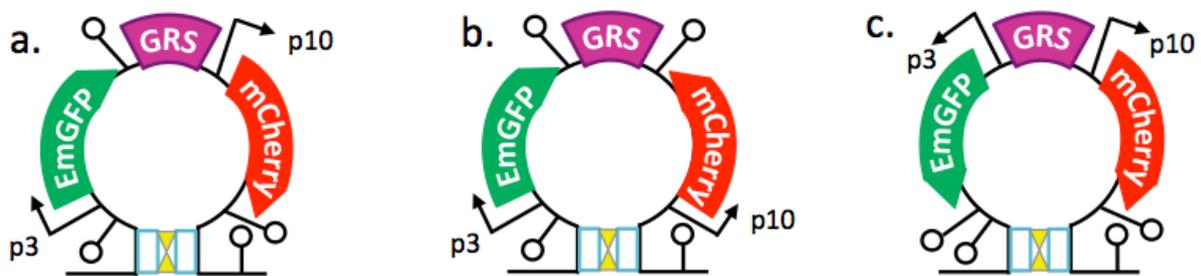


Figure 38 : Représentation des constructions à 2 gènes arborant le GRS.

a. Construction comprenant les gènes emGFP et mCherry en tandem sous contrôle de promoteurs forts (p3 et p10 respectivement) et un site GRS dans la région intergénique. b. Construction comprenant les gènes emGFP et mCherry convergents sous contrôle de promoteurs forts (p3 et p10 respectivement) et un site GRS dans la région intergénique. c. Construction comprenant les gènes emGFP et mCherry divergents sous contrôle de promoteurs forts (p3 et p10 respectivement) et un site GRS dans la région intergénique. Les flèches angulaires représentent les promoteurs, les épines représentent les terminateurs de transcription, la flèche verte représente emGFP, la flèche rouge représente mCherry, les lignes et cercles noirs représentent l'ADN, les rectangles représentent les sites de liaisons pour la protéine Lambda-cl, et les triangles jaunes représentent la protéine Lambda-cl.

La présence du GRS à l'intérieur de la boucle dans les constructions à 2 gènes permet d'augmenter l'expression de *emGFP* dans toutes les configurations (Figure 39). Les effets sont cependant variés : on observe une forte augmentation (~6 fois) dans la configuration convergente et des augmentations plus faibles dans les configurations en tandem (~4 fois) et divergent (2 à 3 fois). Encore une fois, la configuration en tandem est celle permettant la plus forte expression du gène *emGFP*.

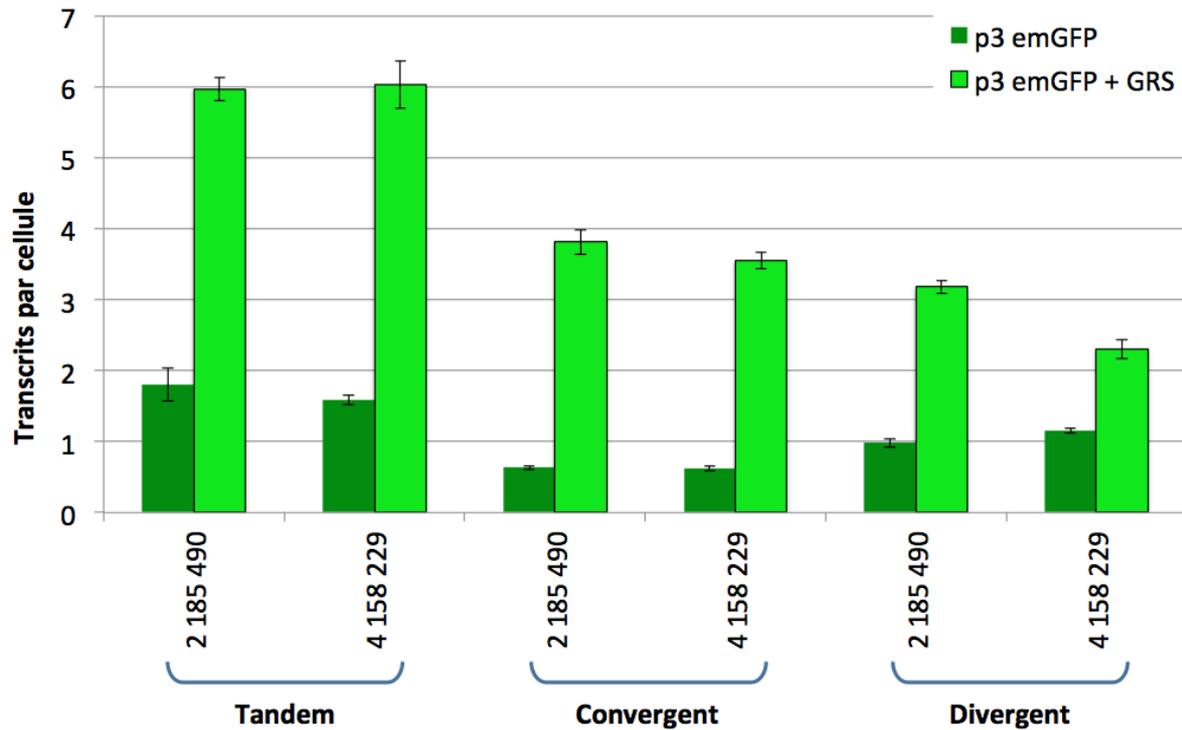


Figure 39 : Niveaux d'expression du gène *emGFP* dans les boucles à 2 gènes avec GRS.

Les niveaux d'expression du gène emGFP ont été mesurés par RT-qPCR en présence d'un second gène sous contrôle d'un promoteur fort dans la boucle d'ADN en présence ou absence d'un GRS dans la région intergénique. L'axe des abscisses représente le site d'insertion de la construction et l'orientation relative des gènes. L'axe des ordonnées représente le nombre de transcrits par cellule.

Une analyse comparable réalisée au niveau de *mCherry* montre que la présence du GRS stimule l'activité de p10 d'un facteur ~3 dans la configuration en tandem et d'un facteur ~2 dans la configuration convergente. La présence du GRS à l'intérieur de la boucle ne semble pas avoir d'effet significatif dans la configuration divergente. Comme pour *emGFP*, la configuration en tandem est celle permettant la plus forte expression du gène *mCherry*.

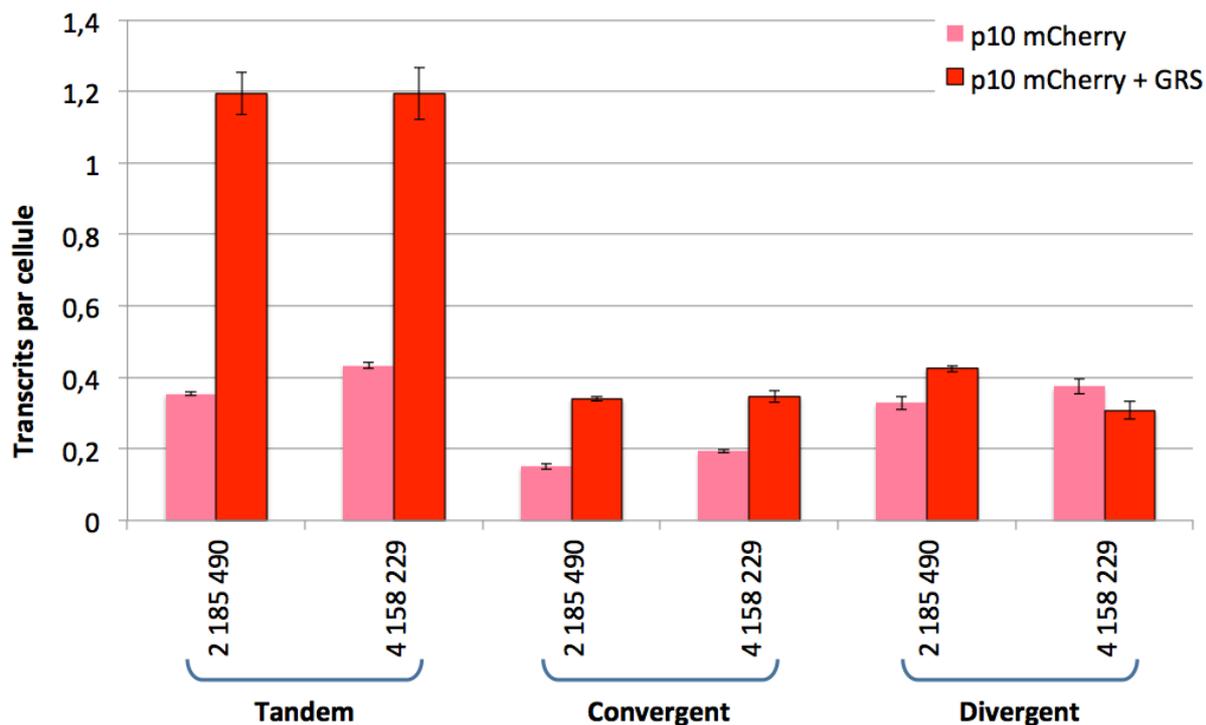


Figure 40 : Niveaux d'expression du gène *mCherry* dans les boucles à 2 gènes avec GRS.

Les niveaux d'expression du gène mCherry sous contrôle d'un promoteur fort ont été mesurés par RT-qPCR lorsque le gène emGFP est fortement exprimé dans la boucle d'ADN en présence ou absence d'un GRS dans la région intergénique. L'axe des abscisses représente le site d'insertion de la construction et l'orientation relative des gènes. L'axe des ordonnées représente le nombre de transcrits par cellule.

3.8 La capacité d'oligomérisation de Lambda-cl est essentielle pour la formation de boucles dans l'ADN

L'aptitude de Lambda-cl à former des boucles d'ADN dépend strictement de sa capacité d'oligomérisation. Suivant ses interactions avec l'ADN, cette protéine forme des dimères, des tétramères et des octamères (227,228). Les dimères se lient sur chacun des sites opérateurs 1 et 2 des régions O_L et O_R et forment des tétramères (Figure 41a). Le tétramère lié à O_{L1-2} et celui lié à O_{R1-2} interagissent pour former ensuite un octamère. Cette liaison permet la formation d'une boucle d'ADN (Figure 41b). Cette boucle est stabilisée par la formation d'un tétramère à partir d'un dimère lié à O_{L3} et d'un autre lié à O_{R3} (Figure 41c).

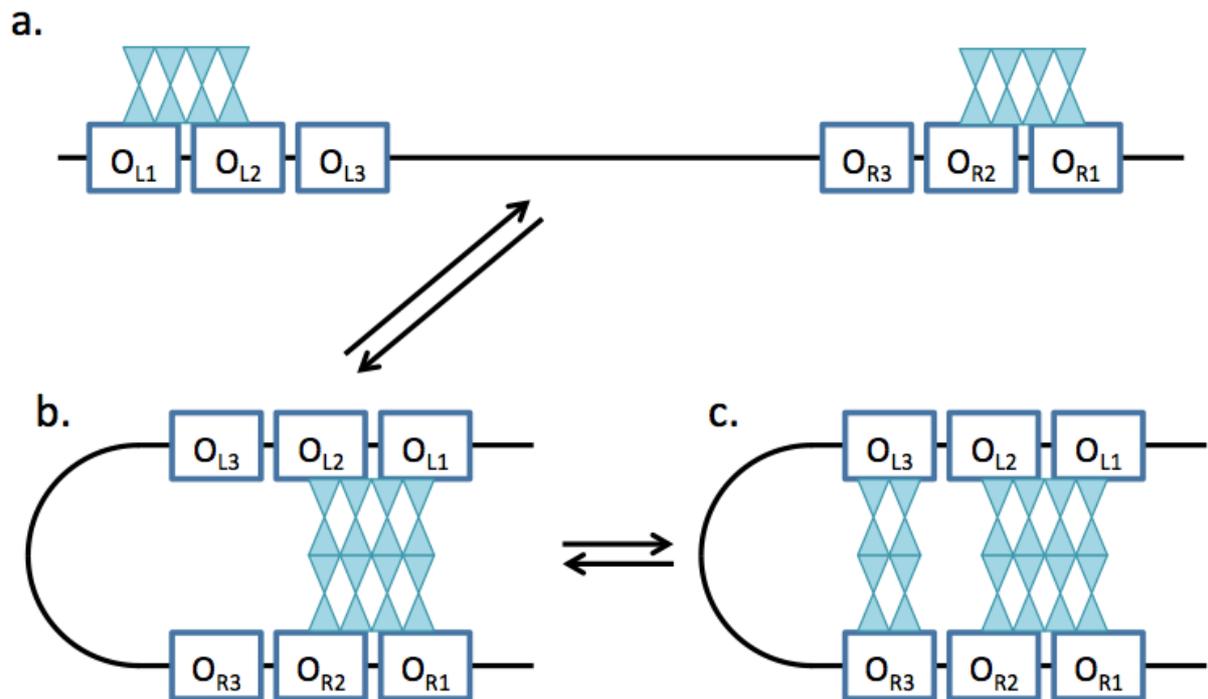


Figure 41 : Représentation des différentes conformation adoptées par une région ADN en fonction de la liaison de Lambda-cl.

a. Liaison des tétramères sur les sites O_{L1-2} et O_{R1-2} . b. Octamérisation entre les tétramères liés aux sites O_{L1-2} et O_{R1-2} . c. Tétramérisation entre deux dimères liés à O_{L3} et O_{R3} . Les rectangles bleus représentent les sites opérateurs, les sabliers représentent Lambda-cl, et les lignes noires représentent l'ADN.

Dans le but de confirmer que c'est la formation de la boucle d'ADN qui permet l'homogénéisation de l'expression des gènes *emGFP* et *mCherry* et non la seule liaison de Lambda-cl sur ses sites opérateurs, des contrôles dans lesquels la protéine Lambda-cl a été mutée ont été réalisés. Parmi les différentes mutations caractérisées pour Lambda-cl par Burz et Ackers (347), 3 ont été sélectionnées : P158 → T ; Y210 → H ; S228 → R (Figure 42). Le mutant P158T présente une altération de sa capacité de dimérisation, le mutant S228R peut former des dimères mais ne peut former ni des tétramères ni des octamères, et le mutant Y210H n'est pas capable de liaison coopérative à 2 sites opérateurs adjacents. Ces mutations empêchent la formation d'une boucle dans l'ADN sans inhiber la fixation aux opérateurs.

La construction p3:emGFP/GRS/p10:mCherry tandem (Figure 38a) a été insérée dans des souches exprimant chacune un mutant de Lambda-cl. Le niveau d'expression des gènes a ensuite été mesuré par RT-qPCR et comparé aux niveaux d'expression en présence et en absence de Lambda-cl sauvage (Figure 43). En absence (Figure 43) et en présence (Figure 39)

de la protéine Lambda-cl sauvage, l'expression de *emGFP* est de $\sim 2,1$ et ~ 6 TPC respectivement. Dans les cellules produisant Lambda-cl P158T ou S228R, cette expression est de $\sim 1,4$ TPC et de $\sim 2,1$ TPC en présence de Lambda-cl Y210H. Des résultats relatifs similaires, mais à un niveau d'expression plus faible ont été observés pour *mCherry* (Figure 43). Ces résultats montrent que les mutants de Lambda-cl permettent des niveaux d'expression des gènes *emGFP* et *mCherry* similaires aux niveaux d'expression en absence de Lambda-cl. Il est conclu que l'effet de Lambda-cl sur l'expression des cassettes rapportrices dépend de la formation de boucles dans l'ADN et non de la seule fixation de la protéine à l'ADN.

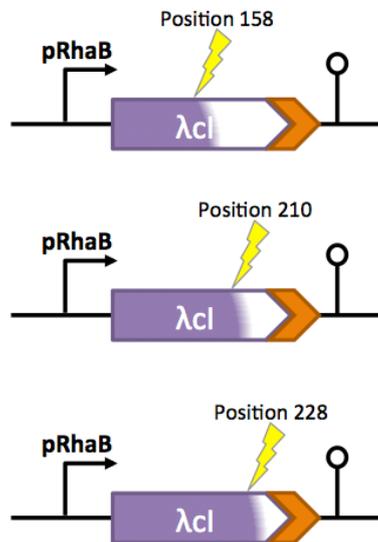


Figure 42 : Représentation des 3 mutants de *lambda-cl*.

Les flèches angulaires représentent les promoteurs, les épingles représentent les terminateurs de transcription, les éclairs représentent les sites de mutation dans le gène, les flèches pleines violettes et blanches représentent le gène lambda-cl, la flèche pleine orange représente le tag FLAG combiné au gène lambda-cl, et les lignes noires représentent l'ADN.

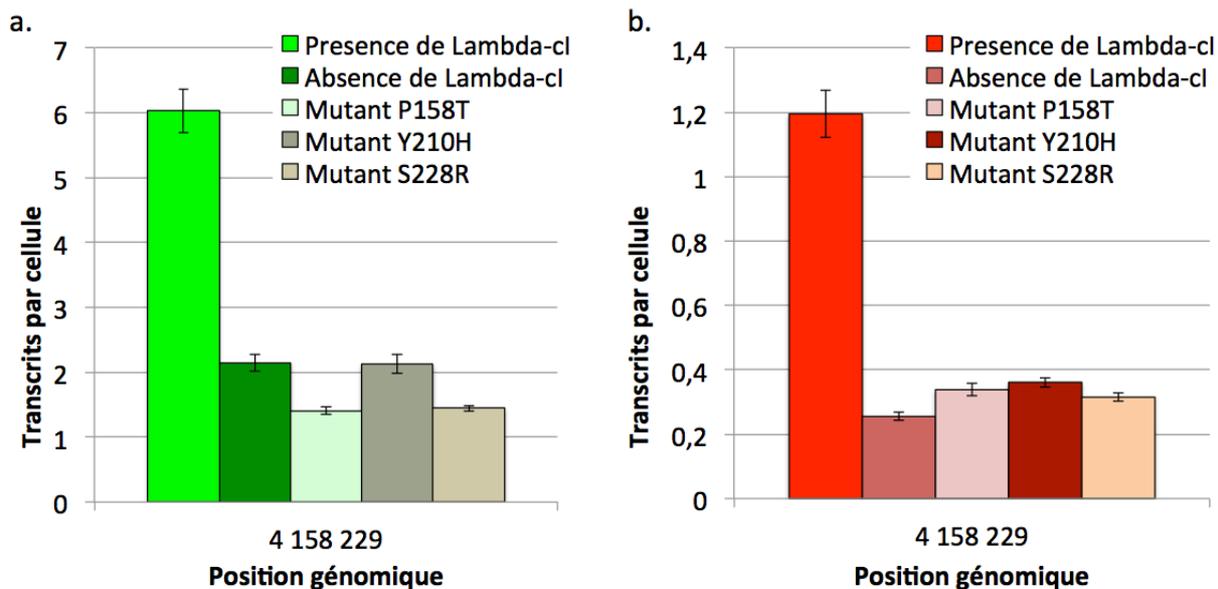


Figure 43 : Expression de *emGFP* (en vert) et *mCherry* (en rouge) organisés en tandem en fonction des mutants de Lambda-cl

Les niveaux d'expression des gènes emGFP et mCherry sous contrôle respectifs des promoteurs forts p3 et p10 ont été mesurés par RT-qPCR lorsqu'un GRS est présent dans la région intergénique. L'axe des abscisses représente le site d'insertion de la construction et l'orientation relative des gènes. L'axe des ordonnées représente le nombre de transcrits par cellule.

3.9 LRP régule l'expression génique par la formation de boucles dans l'ADN

La protéine LRP est impliquée directement ou indirectement dans la régulation de près d'un tiers du génome d'*E. coli*, et de ~70% des gènes qui montrent une expression modifiée lors de l'entrée en phase stationnaire de croissance (348,349). Dans une étude récente, Kroner et collaborateurs (348) ont exploré le rôle de la protéine LRP dans la régulation globale des gènes chez *E. coli*. Leur objectif était de mettre en lumière le spectre complet des cibles de cette NAP et les mécanismes utilisés par cette protéine pour réguler les gènes. Ils ont réalisé des expériences de Chip-seq et de RNA-seq dans 3 milieux de cultures différents (milieu minimum, milieu minimum supplémenté, et milieu riche) avec à chaque fois 3 points de mesure (milieu de phase exponentielle, phase de transition, phase stationnaire). Si leur approche a permis d'agrandir la liste des gènes régulés par LRP, elle n'a cependant pas permis de déterminer clairement par quel mécanisme LRP régulait indirectement de nombreux gènes.

Dans certains cas, les auteurs proposent que cette régulation impliquerait des interactions de LRP avec d'autres acteurs cellulaires. Mais dans la majorité des autres cas, aucune explication claire n'est avancée. De plus, bien que constamment liée à l'ADN au niveau de certains loci, LRP n'impacte la transcription génique que dans certaines conditions. Malgré leurs efforts, les chercheurs n'ont pas réussi à expliquer ce phénomène.

Dans cette étude, les chercheurs ont omis de prendre en compte le rôle de la protéine LRP dans la structuration du nucléoïde. En effet, LRP est capable de s'oligomériser et de former des boucles dans l'ADN (350–353). Nous avons exploité les données de Chip-seq et de RNA-seq dans une approche plus globale, tridimensionnelle et topologique. Nous avons découvert que la régulation d'un bon nombre de gènes par LRP donne des réponses similaires à celles obtenues avec les boucles induites par Lambda-cl.

Nous avons observé que LRP se lie de part et d'autre d'un bon nombre de gènes décrits comme régulés directement ou indirectement par cette protéine. Dans ces cas, plus le signal de liaison de LRP est fort, plus ces gènes sont réprimés. Cette réponse est observée pour les gènes *frmA*, *frmR*, *ybhK*, *ddlA*, *ybaB*, *nfsB*, *ybdF*, *ybdH*, *dpiB*, *ybiT*, *ybjN*, *ompF*, *asnS*, *pncB*, *ghrA*, *rluB*, *btuR*, *sohB*, *topA*, *feaB*, *yddW*, *ydcl*, *yneG*, *yneH*, *manA*, *ydgH*, *ves*, *gdhA*, *yeeN*, *sdcb*, *xseA*, *radD*, *rply*, *alaS*, *recJ*, *rpoN*, *sstT*, *mtr*, *yicR*, *menA*, *iclR*, *aphA*, *yjiT*, *yjgH*, et *nhaA*, et certains résultats sont présentés en détail dans la Figure 44. Nous postulons que la régulation de ces gènes résulte de la formation de boucles d'ADN induites par la liaison de LRP en amont et en aval de ces gènes. Ainsi, comme pour les cassettes p3:emGFP, l'expression de certains gènes régulés indirectement par LRP pourrait résulter de la formation de boucles de petites tailles où l'expression des gènes est inhibée par l'accumulation de supertours positifs au sein de la boucle.

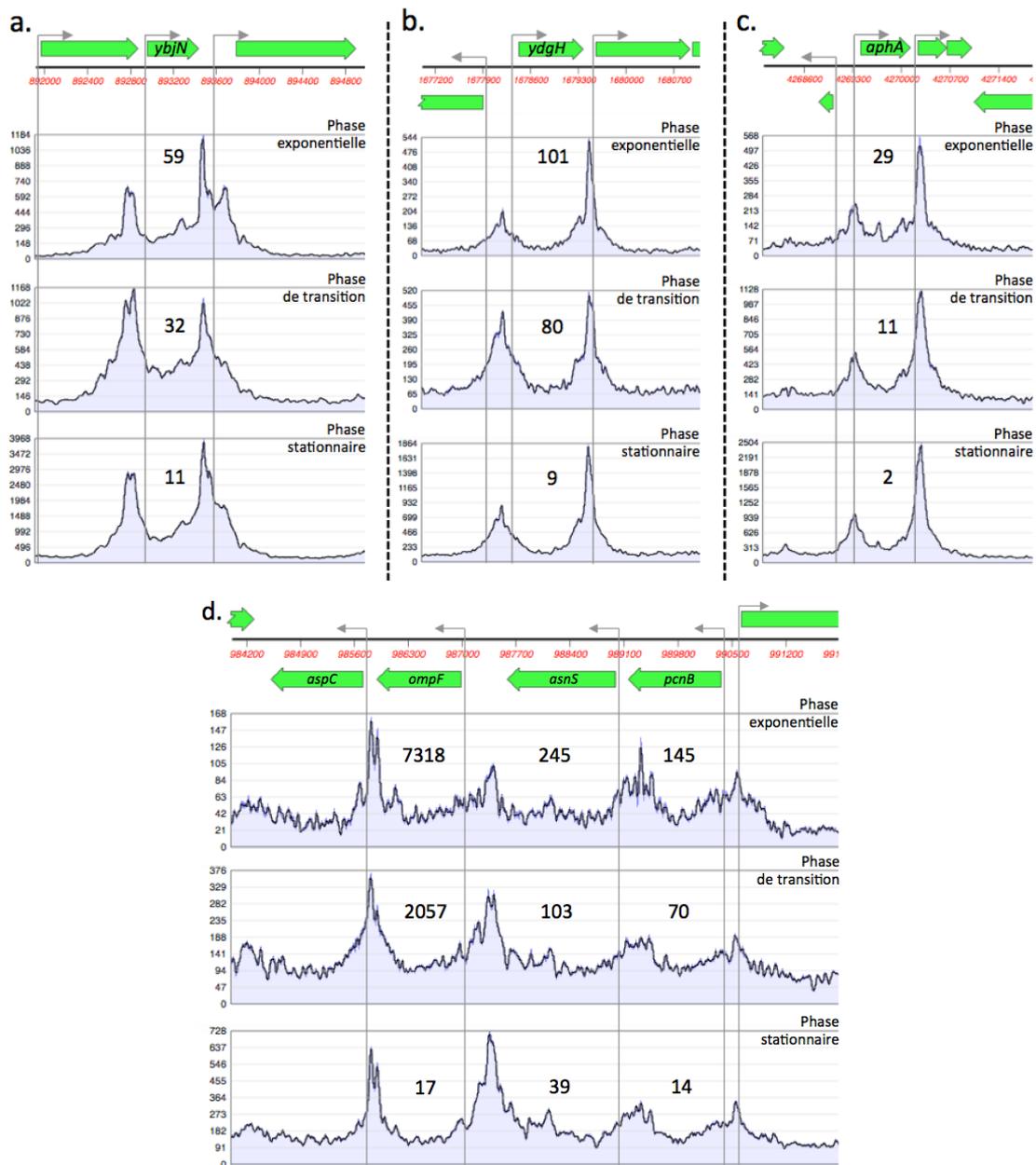


Figure 44 : Données de Chip-seq et RNA-seq pour LRP au niveau de certains loci génomiques dans un milieu minimum.

Les données de Chip-seq et RNA-seq ont été extraites de l'étude de Kroner et collaborateurs (348). Les mesures ont été réalisées dans un milieu de culture minimum avec du glucose comme source de carbone. L'intensité du signal de Chip-seq est indiquée par la courbe en fonction de l'échelle à gauche de chaque courbe. Les valeurs de RNA-seq sont directement indiquées sous les gènes correspondants, dans les 3 conditions de culture : phase exponentielle, phase de transition, et phase stationnaire. a. Expression du gène *ybjN* en fonction de la phase de croissance. b. Expression du gène *ydgH* en fonction de la phase de croissance. c. Expression du gène *aphA* en fonction de la phase de croissance. d. Expression des gènes *ompF*, *asnS*, et *pcnB* en fonction de la phase de croissance. Les gènes sont représentés en vert et les flèches courbées correspondent aux promoteurs. Les illustrations ont été réalisées sur le logiciel MacVector.

De manière intéressante, une région de 5kb concentre 4 sites de liaison pour la protéine LRP espacés d'environ 1,3 kpb chacun. Potentiellement, 3 boucles successives, portant chacune un gène, peuvent se former dans cette région (Figure 44d). Il est observé que ces gènes sont également réprimés quand le signal LRP est fort. Le fait que l'évolution ait façonné cette région de façon à former 3 boucles successives au lieu d'une seule plus grande indique que la régulation de gènes par la formation de boucles a un maximum d'efficacité quand les boucles sont petites.

Notre travail a également permis d'observer le comportement de gènes convergents placés entre deux sites de liaison de LRP. A l'image de ce qu'il se passe quand 2 gènes convergents sont placés dans les boucles induites par Lambda-cl (Figure 33c,d), une augmentation de l'expression d'un gène résulte toujours en une diminution de l'expression du second gène et vice versa. C'est le cas pour les gènes *smrA* & *dgcM*, *yhjE* & *yhjG*, *eco* & *mgo*, et *ytfK* & *ytfL*, et certains résultats sont présentés dans la Figure 45.

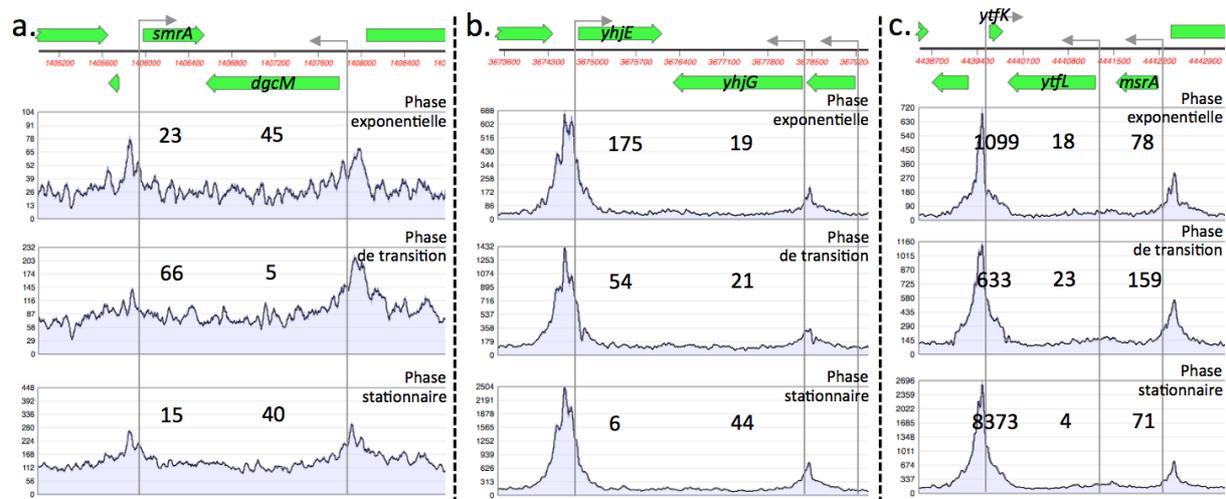


Figure 45 : Données de Chip-seq et RNA-seq pour LRP au niveau de certains loci génomiques dans un milieu minimum.

*Les données de Chip-seq et RNA-seq ont été extraites de l'étude de Kroner et collaborateurs (348). Les mesures ont été réalisées dans un milieu de culture minium avec du glucose comme source de carbone. L'intensité du signal de Chip-seq est indiquée par la courbe en fonction de l'échelle à gauche de chaque courbe. Les valeurs de RNA-seq sont directement indiquées sous les gènes correspondants, dans les 3 points de mesure : phase exponentielle, phase de transition, et phase stationnaire. a. Expression des gènes *smrA* et *dgcM* en fonction de la phase de croissance. b. Expression des gènes *yhjE* et *yhjG* en fonction de la phase de croissance. c. Expression des gènes *ytfK*, *ytfL* et *msrA* en fonction de la phase de croissance. Les gènes sont représentés en vert, et les flèches courbées correspondent aux promoteurs.*

4 – MATERIEL ET METHODES

4 MATERIEL ET METHODES

4.1 Construction des souches bactériennes et plasmides

Les souches, plasmides et amorces utilisés dans ce travail sont listés dans les tableaux 1 à 3.

Espèce	Souche parentale	Nom de souche	Caractéristiques	Resistance
<i>E. coli</i>	MG1665	BCJ952	pTKRED + λ cl (Δ RhaB Δ RhaA Δ RhaD) + Phleomycine	Spectinomycine + Phleomycine
<i>E. coli</i>	BCJ952.4	α 1	pTKRED + λ cl (Δ RhaB Δ RhaA Δ RhaD)	Spectinomycine
<i>E. coli</i>	MG1665	2	pTKRED + cassette rapportrice pBCJ932 dans locus 2,171,769	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	5	pTKRED + cassette rapportrice pBCJ932 dans locus 2,185,490	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	7	pTKRED + cassette rapportrice pBCJ932 dans locus 2,192,420	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	8	pTKRED + cassette rapportrice pBCJ932 dans locus 2,196,391	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	12	pTKRED + cassette rapportrice pBCJ932 dans locus 2,214,800	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	14	pTKRED + cassette rapportrice pBCJ932 dans locus 2,223,904	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	19	pTKRED + cassette rapportrice pBCJ932 dans locus 2,244,773	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	20	pTKRED + cassette rapportrice pBCJ932 dans locus 2,248,654	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	26	pTKRED + cassette rapportrice pBCJ932 dans locus 3,959,452	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	28	pTKRED + cassette rapportrice pBCJ932 dans locus 3,990,959	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	29	pTKRED + cassette rapportrice pBCJ932 dans locus 4,035,211	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	30	pTKRED + cassette rapportrice pBCJ932 dans locus 4,046,827	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	31	pTKRED + cassette rapportrice pBCJ932 dans locus 4,158,229	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	32	pTKRED + cassette rapportrice pBCJ932 dans locus 4,166,287	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	35	pTKRED + cassette rapportrice pBCJ932 dans locus 4,207,682	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	36	pTKRED + cassette rapportrice pBCJ932 dans locus 4,256,593	Spectinomycine + Phleomycine
<i>E. coli</i>	α 1	α 2	pTKRED + λ cl + cassette rapportrice pBCJ932 dans locus 2,171,769	Spectinomycine + Phleomycine
<i>E. coli</i>	α 1	α 3	pTKRED + λ cl + cassette rapportrice pBCJ932 dans locus 2,185,490	Spectinomycine + Phleomycine
<i>E. coli</i>	α 1	α 5	pTKRED + λ cl + cassette rapportrice pBCJ932 dans locus 2,192,420	Spectinomycine + Phleomycine
<i>E. coli</i>	α 1	α 6	pTKRED + λ cl + cassette rapportrice pBCJ932 dans locus 2,196,391	Spectinomycine + Phleomycine

Matériel et méthodes

<i>E. coli</i>	α1	α8	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 2,214,800	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α10	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 2,223,904	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α12	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 2,244,773	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α13	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 2,248,654	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α15	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 3,959,452	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α16	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 3,990,959	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α17	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 4,035,211	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α18	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 4,046,827	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α19	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 4,158,229	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α20	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 4,166,287	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α22	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 4,207,682	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α23	pTKRED + λcl + cassette rapportrice pBCJ932 dans locus 4,256,593	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α26	pTKRED + λcl + cassette rapportrice pBCJ927 dans locus 2,185,490	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α28	pTKRED + λcl + cassette rapportrice pBCJ927 dans locus 2,244,773	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α35	pTKRED + λcl + cassette rapportrice pBCJ927 dans locus 4,046,827	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α36	pTKRED + λcl + cassette rapportrice pBCJ927 dans locus 4,158,229	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α61	pTKRED + λcl + cassette rapportrice β5 dans locus 2,185,490	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α62	pTKRED + λcl + cassette rapportrice β5 dans locus 4,046,827	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α63	pTKRED + λcl + cassette rapportrice β5 dans locus 4,158,229	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α64	pTKRED + λcl + cassette rapportrice β5 dans locus 2,244,773	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α66	pTKRED + λcl + cassette rapportrice β7 dans locus 2,185,490	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α67	pTKRED + λcl + cassette rapportrice β7 dans locus 4,046,827	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α68	pTKRED + λcl + cassette rapportrice β7 dans locus 4,158,229	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α69	pTKRED + λcl + cassette rapportrice β7 dans locus 2,244,773	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	α73	pTKRED + cassette rapportrice pBCJ927 dans locus 2,185,490	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	α74	pTKRED + cassette rapportrice pBCJ927 dans locus 4,046,827	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	α75	pTKRED + cassette rapportrice pBCJ927 dans locus 4,158,229	Spectinomycine + Phleomycine

Matériel et méthodes

<i>E. coli</i>	MG1665	α76	pTKRED + cassette rapportrice pBCJ927 dans locus 2,244,773	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	α92	pTKRED + cassette rapportrice β5 dans locus 2,185,490	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	α93	pTKRED + cassette rapportrice β5 dans locus 4,046,827	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	α94	pTKRED + cassette rapportrice β5 dans locus 4,158,229	Spectinomycine + Phleomycine
<i>E. coli</i>	MG1665	α95	pTKRED + cassette rapportrice β5 dans locus 2,244,773	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α120	pTKRED + λcl + cassette rapportrice β16 dans locus 2,185,490	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α121	pTKRED + λcl + cassette rapportrice β16 dans locus 4,046,827	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α122	pTKRED + λcl + cassette rapportrice β16 dans locus 4,158,229	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α123	pTKRED + λcl + cassette rapportrice β16 dans locus 2,244,773	Spectinomycine + Phleomycine
<i>E. coli</i>	α26	α125	pTKRED + λcl + p3 emGFP et p1 mCherry (Tandem) dans locus 2,185,490	Spectinomycine + Kanamycin
<i>E. coli</i>	α36	α127	pTKRED + λcl + p3 emGFP et p1 mCherry (Tandem) dans locus 4,158,229	Spectinomycine + Kanamycin
<i>E. coli</i>	α26	α129	pTKRED + λcl + p3 emGFP et p10 mCherry (Tandem) dans locus 2,185,490	Spectinomycine + Kanamycin
<i>E. coli</i>	α36	α131	pTKRED + λcl + p3 emGFP et p10 mCherry (Tandem) dans locus 4,158,229	Spectinomycine + Kanamycin
<i>E. coli</i>	α26	α147	pTKRED + λcl + p3 emGFP et p1 mCherry (Convergent) dans locus 2,185,490	Spectinomycine + Kanamycin
<i>E. coli</i>	α36	α148	pTKRED + λcl + p3 emGFP et p1 mCherry (Convergent) dans locus 4,158,229	Spectinomycine + Kanamycin
<i>E. coli</i>	α26	α149	pTKRED + λcl + p3 emGFP et p10 mCherry (Convergent) dans locus 2,185,490	Spectinomycine + Kanamycin
<i>E. coli</i>	α36	α150	pTKRED + λcl + p3 emGFP et p10 mCherry (Convergent) dans locus 4,158,229	Spectinomycine + Kanamycin
<i>E. coli</i>	α1	α153	pTKRED + λcl + cassette rapportrice β22 dans locus 2,185,490	Spectinomycine + Phleomycine
<i>E. coli</i>	α1	α154	pTKRED + λcl + cassette rapportrice β16 dans locus 4,158,229	Spectinomycine + Phleomycine
<i>E. coli</i>	α161	α165	pTKRED + λcl + p3 emGFP et p1 mCherry (Divergent) dans locus 2,185,490	Spectinomycine + Kanamycin
<i>E. coli</i>	α162	α166	pTKRED + λcl + p3 emGFP et p1 mCherry (Divergent) dans locus 4,158,229	Spectinomycine + Kanamycin
<i>E. coli</i>	α163	α167	pTKRED + λcl + p3 emGFP et p10 mCherry (Divergent) dans locus 2,185,490	Spectinomycine + Kanamycin
<i>E. coli</i>	α164	α168	pTKRED + λcl + p3 emGFP et p10 mCherry (Divergent) dans locus 4,158,229	Spectinomycine + Kanamycin
<i>E. coli</i>	α26	α169	pTKRED + λcl + p3 emGFP et promoteur muté mCherry (Tandem) dans locus 2,185,490	Spectinomycine + Kanamycin
<i>E. coli</i>	α36	α170	pTKRED + λcl + p3 emGFP et promoteur muté mCherry (Tandem) dans locus 4,158,229	Spectinomycine + Kanamycin
<i>E. coli</i>	α171	α179	pTKRED + λcl + p3 emGFP, GRS et p10 mCherry (Tandem) dans locus 2,185,490	Spectinomycine + Kanamycin

Matériel et méthodes

<i>E. coli</i>	α 172	α 180	pTKRED + λ cl + p3 emGFP, GRS et p10 mCherry (Tandem) dans locus 4,158,229	Spectinomycine + Kanamycin
<i>E. coli</i>	α 173	α 181	pTKRED + λ cl + p3 emGFP, GRS et p10 mCherry (Convergent) dans locus 2,185,490	Spectinomycine + Kanamycin
<i>E. coli</i>	α 174	α 182	pTKRED + λ cl + p3 emGFP, GRS et p10 mCherry (Convergent) dans locus 4,158,229	Spectinomycine + Kanamycin
<i>E. coli</i>	α 175	α 183	pTKRED + λ cl + p3 emGFP, GRS et p10 mCherry (Divergent) dans locus 2,185,490	Spectinomycine + Kanamycin
<i>E. coli</i>	α 176	α 184	pTKRED + λ cl + p3 emGFP, GRS et p10 mCherry (Divergent) dans locus 4,158,229	Spectinomycine + Kanamycin

Tableau 1 : Souches construites.

Nom	Vecteur	Contenu	Resistance
pBCJ932	pTKIP	Opérateurs lambda-cl, p1-EmGFP	Ampiciline + Phleomycine
pBCJ927	pTKIP	Opérateurs lambda-cl, p3-EmGFP	Ampiciline + Phleomycine
pMC48	pTKIP	mCherry sous contrôle du promoteur p1	Ampiciline
β 5	pTKIP	Opérateurs lambda-cl, p3-EmGFP, GRS (en amont de EmGFP)	Ampiciline + Phleomycine
β 7	pTKIP	Opérateurs lambda-cl, p3-EmGFP, Neutral DNA (en amont de EmGFP)	Ampiciline + Phleomycine
β 9	pTKIP	Opérateurs lambda-cl, GRS (en aval EmGFP), p3-EmGFP	Ampiciline + Phleomycine
β 22	pTKIP	Plasmide pBCJ927 avec promoteur p3 muté	Ampiciline + Phleomycine
β 23	pTKIP	Promoteur muté + mCherry + opérateurs lambda-cl+	Ampiciline + Kanamycin

Tableau 2 : Plasmides construits

Nom	Sequence (5' - 3')	Tm	Site de liaison
147F	TGAATTCATCGTCATTTACCCATATTCAATTGTGGCTAG TGTAACGAAGTACGGCCCCAAGGTCCAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 2 171 769 dans génome</u>
147R	GTGATGCTGGCCCGTATTGTGCAAAACAGATCATT ACCAATGGTCCCCTTGGCTTCAGGGATGAGGCG	57° C	<u>Liaison à la construction 927/932 extrémité 3', queue pour cibler le site 2 171 769 dans génome</u>
150F	GATAAATCGCAGAGGAGGATGGTAATGTCCAGCGCAC GCGTTGTAACGATACGGCCCCAAGGTCCAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 2 185 490 dans génome</u>
150R	TTCATATATCAAATAATTTATTAACGCGATTGTAAACT GCCGTTTTCTTGGCTTCAGGGATGAGGCG	57° C	<u>Liaison à la construction 927/932 extrémité 3', queue pour cibler le site 2 185 490 dans génome</u>
152F	AATATTTAAGAGTATTAATACTATTTATCGCATCTCAAT TAATGTAGATTACGGCCCCAAGGTCCAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 2 192 420 dans génome</u>

Matériel et méthodes

152R	ATATAAATGATTTTCGGCTTTTTTATTGATATCAACAATA CCATTTACATATTGGCTTCAGGGATGAGGCG	57° C	<u>Liaison à la construction 927/932 extrémité 3', queue pour cibler le site 2 192 420 dans génome</u>
153F	CATCGACAGCGCCTTTTCTTTATAAATTCCTAAAGTTGT TTTCTTGCGATTACGGCCCCAAGGTCCAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 2 196 391 dans génome</u>
153R	CAGTTGAATGCAGATGCTACCAGTATTTATGCGGGTTA GAGAGAGACAAATTGGCTTCAGGGATGAGGCG	57° C	<u>Liaison à la construction 927/932 extrémité 3', queue pour cibler le site 2 196 391 dans génome</u>
157F	AAATCGCCTGGCAAAAATAAAATCACCCCTATAGATGC ACAAAAACGGGCTACGGCCCCAAGGTCCAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 2 214 800 dans génome</u>
157R	GCGAGGTCCCGGTTTAACTTTAGACGCAGTTTTGCGAA CCAGGTAGTTTTTGGCTTCAGGGATGAGGCG	57° C	<u>Liaison à la construction 927/932 extrémité 3', queue pour cibler le site 2 214 800 dans génome</u>
159F	GGATCGGTAAAACAGTAAACGGAAAACTGGCAGG AAGTGGAGTAAAAATACGGCCCCAAGGTCCAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 2 223 904 dans génome</u>
159R	GGCACAGAACGATTAAGTGAATTCGGATGCGGATACT CTGCCATCCGTAATTGGCTTCAGGGATGAGGCG	57° C	<u>Liaison à la construction 927/932 extrémité 3', queue pour cibler le site 2 223 904 dans génome</u>
164F	CATGCGCAGTATTTACTGAAGTGAAAGTCCGCCCGGT TCGCCGGCATCTTACGGCCCCAAGGTCCAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 2 244 773 dans génome</u>
164R	GAAGACGGACGTCGTTACTTTATGGCAGTGGATTATC GCTTCTGATGAGATTGGCTTCAGGGATGAGGCG	57° C	<u>Liaison à la construction 927/932 extrémité 3', queue pour cibler le site 2 244 773 dans génome</u>
165F	ACACGGTTATAAGACACCTTCATGATCGCCAGGGAT TATAAGTAAAGCATACGGCCCCAAGGTCCAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 2 248 654 dans génome</u>
165R	ATTCGTGACGCAGCGCCTTCAGCATGCATTGCCAG AAAAGAGATTGGCTTGGCTTCAGGGATGAGGCG	57° C	<u>Liaison à la construction 927/932 extrémité 3', queue pour cibler le site 2 248 654 dans génome</u>
245F	GAAACTGCGCGGCTATATGACAGATATGAAACGTATT GCTGTTGCGGGTAAAGTGC GCGCTACGGCCCCAAGGT CCAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 3 959 452 dans génome</u>
245R	AGGAGAAGGCCTTGAGTGTTTTCTCCCTCTCCCTGTGG GAGAGGGTCCGGGTGAGGGCATCATTGGCTTCAGGG ATGAGGCG	57° C	<u>Liaison à la construction 927/932 extrémité 3', queue pour cibler le site 3 959 452 dans génome</u>
247F	AATAAGGAATTTACAGAGAATAAACGGTGCTACACTT GTATGTAGCGCATCTTACGGCCCCAAGGTCCAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 3 990 959 dans génome</u>
247R	ATGGTCTAAAACGTGATCAATTTAACACCTTGCTGATT GACCGTAAAGAATTGGCTTCAGGGATGAGGCG	57° C	<u>Liaison à la construction 927/932 extrémité 3', queue pour cibler le site 3 990 959 dans génome</u>
248F	AAGCATAAAGAATAAAAAATGCGCGGTCAGAAAATTA TTTTAAATTTCTCTTGTGAGGCTACGGCCCCAAGGTC CAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 4 035 211 dans génome</u>
248R	CCGGCGCGGTGTTTGCCGTTGTTCCGTGTCAGTGGTG GCGCATTATAGGGAGTTATCCGTTGGCTTCAGGGATG AGGCG	57° C	<u>Liaison à la construction 927/932 extrémité 3', queue pour cibler le site 4 035 211 dans génome</u>
249F	ACAAACATCATGCTGTAAAAAGCATGATAATAAATTA AAGCGATGTAATAATTTATGCTACGGCCCCAAGGTC CAAAC	57° C	<u>Liaison à la construction 927/932 extrémité 5', queue pour cibler le site 4 046 827 dans génome</u>

Matériel et méthodes

249R	TCGTGAAAATCTTTTGTAGATCTTCTGGATCGCTCGCA AATCGTCATGTGGATAACTTTGTTGGCTCAGGGATG <u>AGGCG</u>	57° C	<u>Liaison à la construction 927/932 extrémité 3',</u> queue pour cibler le site 4 046 827 dans génome
250F	TCACCGCAGCAGGTGGCGCAGGCGATTGCTTTGCGC AGGCTCGGTTAGGGTAAGAACATTACGGCCCCAAGGT <u>CCAAAC</u>	57° C	<u>Liaison à la construction 927/932 extrémité 5',</u> queue pour cibler le site 4 158 229 dans génome
250R	CCGCTTTTTTTTGCATAAAAAAGCCGGCGATAAGCC AGGCTCAAATTTATACATATAATTGGCTCAGGGATGA <u>GGCG</u>	57° C	<u>Liaison à la construction 927/932 extrémité 3',</u> queue pour cibler le site 4 158 229 dans génome
251F	TATTGCCCGTTTTACAGCGTTACGGCTCGAAACGCTC GAAAACTGGCAGTTTTAGGCTGATTACGGCCCCAA <u>GGTCCAAAC</u>	57° C	<u>Liaison à la construction 927/932 extrémité 5',</u> queue pour cibler le site 4 166 287 dans génome
251R	AAATAATTTTCTGACCGCGCAACATTCACCAAATCAG CCTAAAACGTCAGTTTTTCGAGCTTGGCTTCAGGGAT <u>GAGGCG</u>	57° C	<u>Liaison à la construction 927/932 extrémité 3',</u> queue pour cibler le site 4 166 287 dans génome
254F	TTTTCGCCCGCATTGTAACGAAAACGTTTGCACAACGC TCGCGAATTTTTCTTTCAATTACGGCCCCAAGGTCC <u>AAAC</u>	57° C	<u>Liaison à la construction 927/932 extrémité 5',</u> queue pour cibler le site 4 207 682 dans génome
254R	AATCATCAATGTAATTTCTGTATTTGCCACGTAACCC ACAGTCAAAATTTGTGATCACCTTGGCTTCAGGGATGA <u>GGCG</u>	57° C	<u>Liaison à la construction 927/932 extrémité 3',</u> queue pour cibler le site 4 207 682 dans génome
255F	TCTGTGGTATCCGCTCATGTTTCGCGCGGCGCTACGCA AACCCGAATCATCGGATTTAACTACGGCCCCAAGGTC <u>CAAAC</u>	57° C	<u>Liaison à la construction 927/932 extrémité 5',</u> queue pour cibler le site 4 256 593 dans génome
255R	GTGGTATTATTGGCCATTGAAAGAACCTTTTACATTAT GAGCGTCAATATCAGTGTACCTTGGCTTCAGGGATGA <u>GGCG</u>	57° C	<u>Liaison à la construction 927/932 extrémité 3',</u> queue pour cibler le site 4 256 593 dans génome
147s 1	TACTCCC GCCAGATCCTGAA	54° C	Contrôle de l'insertion génomique de la construction 927/932
147s 2	AGCCATTAATCGCTCACCGT	54° C	
150s 1	CGAGTCGTTTGTTGCGATG	54° C	Contrôle de l'insertion génomique de la construction 927/932
150s 2	AACACATCGTTCACCAGGGG	54° C	
152s 1	GCGGCAGCAATAATTGAACG	54° C	Contrôle de l'insertion génomique de la construction 927/932
152s 2	GAAGTGGTTGTTGCAGGTAGC	54° C	
153s 1	ATAATCCCCCTTCAAGGCGC	54° C	Contrôle de l'insertion génomique de la construction 927/932
153s 2	GCAACTCGTTCCCTGTTGC	54° C	
157s 1	TACCCGCAAACCACACTTCC	54° C	Contrôle de l'insertion génomique de la construction 927/932
157s 2	CAAAGCAACGCCACTTCACC	54° C	
159s 1	CGACCAACAGCATGAATCCG	54° C	Contrôle de l'insertion génomique de la construction 927/932
159s 2	ACAAAAAGCCGCCAAATGG	54° C	

Matériel et méthodes

164s 1	GATGTATGCCAGTAACGCGC	54° C	Controle de l'insertion genomique de la construction 927/932
164s 2	ACTATACCGGGCAGAAACGC	54° C	
165s 1	AAATGAGGCGCGATTGTAGC	54° C	Controle de l'insertion genomique de la construction 927/932
165s 2	GGCGCTTTCTGGACTATTGC	54° C	
245s 1	CTTGTGTCCGTTGCTGAAA	54° C	Controle de l'insertion genomique de la construction 927/932
245s 2	TGGTCCGCGTTCGATAAA	54° C	
247s 1	TGGATCCTGACAGGCGTTTC	54° C	Controle de l'insertion genomique de la construction 927/932
247s 2	TTGGAATGCAGGCCCCATAG	54° C	
248s 1	GTCAGGCGGTGAAACGGATA	54° C	Controle de l'insertion genomique de la construction 927/932
248s 2	TTTGCTTTCTCTGCCGGAGT	54° C	
249s 1	GCAAGGCGCAGATTTTAGCA	54° C	Controle de l'insertion genomique de la construction 927/932
249s 2	GATCTTCTGGATCGCTCGCA	54° C	
250s 1	CGATTCTGTCGCTGCAATCG	54° C	Controle de l'insertion genomique de la construction 927/932
250s 2	ACTGCCCGTTTCGAGAGTTT	54° C	
251s 1	TTACAAGTGCTGCCAGAGGG	54° C	Controle de l'insertion genomique de la construction 927/932
251s 2	CCGGCCTGACAAGAGGAAAT	54° C	
254s 1	TTCGACGATACCGGCTTTGT	54° C	Controle de l'insertion genomique de la construction 927/932
254s 2	TGACTGACTGCTGCATTCCC	54° C	
255s 1	AGAGGTATCCAGCCCCAGTT	54° C	Controle de l'insertion genomique de la construction 927/932
255s 2	CTTCGTTGATCCATGCAGCG	54° C	
90a	GAGAGATGGGTAAGCACAAAC	54° C	Contrôle de l'insertion génomique de Lambda-cl
263F	TCAGGTGGCTCATCACGCTA	54° C	
92c	TCGTGACCACCTTGACCTAC	54° C	Sondes qPCR pour emGFP

Matériel et méthodes

92d	GTCGTCCTTGAAGAAGATGG	54° C	
A115 F	TCACCCGGGTCGTATAATGTGTGGATACTAGTGAAAG AGGAGAAATACT	57° C	Queue contenant promoter p3 pour pBCJ927
A115 R	CGACCCGGGTGAATTAATAAGCAAATAAATTGTCGCT ATTGAGTGAGCTAACTCACA	57° C	
A143 R	ATTCAGGCGCTTTTTAGACTGGTCGTAATGAAATTCAG CAGGATCACCATATGAGCACAAAAAGAAACCATTAA C	57° C	Insertion de Lambda-Ci dans l'opéron Rha
A144 R	GCTCCTTTGTCTGTCGTGTAGAGTGAATCTGCGCCACA TTTGGCTTCAGGGATGAGGCG	57° C	
N1F	GCTTGATATCGAATCCGAAG	53° C	Insertion du GRS en aval de emGFP
N1R	CTTGACAGCTCGTCCATGC	53° C	
N2F	GGCCGATTCATTAATGCAGCTGGCACGAC	54° C	Insertion du GRS en amont de emGFP
N2R	AACGCGCGGGGAGAGGCGTTTGCGTATT	54° C	
N3F	TAGCTTGGCTGCAGGTCGTCGAAATTC	54° C	Insertion du GRS en dehors de la boucle
N3R	GCTTGGCTGGACGTAAACTC	54° C	
N6F	CATGGTCATAGCTGTTTCCAGCGCCGGAAGTTGTGTA CAGTCATGCCCGGCTTCTGGTCCATTC	54° C	Insertion de l'ADN neutre en aval de emGFP
N6R	ACTGGCCGTCGTTTTACACACCAGTTTACCTATCGTT CTGATTGCATATCCGGTAACTGCGG	54° C	
N16F	TACGGCCCCAAGGTCCAAACG	54° C	Se lie à Landing pad 1
N16R	TTGGCTTCAGGGATGAGGCG	54° C	Se lie à Landing pad 2
N53F	GCATCCGGGCGTGCCTCTATACTAGTGAAAGAGGAG AAATACTAGATGGTAAGCAAGGCGGAG	54° C	Pour mutation du promoteur p3
N53R	TAGAGCGCACGCCCGGATGCGATGAATGTGCTATGGA CCATAAGGTCGCTACGGGACCTGTAACAACGTTAGA GCCTGTAAGTGCCTG	54° C	
N64F	TGTAACCTCACTCAATAGCGACGCGAGTTGTAACGCTCC AGCTGCTAGTAGATAGACATAGCTCTAGGCGTACTAG TGAAAGAGGAGAAATACTAGATGGTGAGCAAGGGCG AGG	54° C	Amplification de mCherry depuis pMC48 et introduction du promoteur muté
N19R	GCATCTTCGGCATTGTTGCCCCATGCAAACGGGCGGTG GGAATGGACCAGAACGCCGGGCAATAAAAAAGCCCC CGGAATGATCTTCCGGGGGCTTACTTGTACAGCTCGTC CATGCC	54° C	
N64R	CTCGCTCGCTATTGAGTGAGTTACA	54° C	Amplification du plasmide pBCJ927 pour integration de promoteur muté mCherry
N22R	CGCCTCATCCCTGAAGCCAA	54° C	
N80F	GAACGGCCACGAGTTCGAGA	54° C	Sondes qPCR pour mCherry

N80R	CTTGGAGCCGTACATGAACTGAGG	54° C	Amplification du GRS
pUCF	CCCAGTCACGACGTTGTAAAACG	53° C	
pUCR	AGCGGATAACAATTCACACAGG	53° C	

Tableau 3 : Amorces utilisées.

4.1.1 Construction des plasmides

Un fragment d'ADN synthétique contenant le gène *emGFP* flanqué par 2 terminateurs bidirectionnels et les sites opérateurs O_L et O_R pour Lambda-cl a été intégré dans un vecteur pTKIP contenant un gène de résistance à la phléomycine par ligation, donnant le plasmide pBCJ932. Pour exprimer *emGFP* à partir du promoteur fort p3, pBCJ932 a été amplifié par PCR inverse en utilisant les amorces A115F & A115R, résultant en la création du plasmide BCJ927 après circularisation par ligation (HiFi, NEB®).

Pour construire le plasmide β_5 contenant le GRS en aval du gène *emGFP*, le plasmide BCJ927 a été amplifié par PCR inverse en utilisant les amorces N1F & N1R. Un fragment d'ADN de synthèse (acheté à IDT-DNA) contenant le GRS flanqué par des sites homologues au vecteur a été inséré dans le plasmide BCJ927 par ligation (HiFi, NEB®). La même approche a été utilisée pour construire les plasmides β_7 et β_9 , en utilisant respectivement les amorces N6F & N6R et les amorces N2F & N2R.

Le plasmide pMC48 a été construit par intégration d'un fragment d'ADN synthétique contenant le gène *mCherry* (acheté à IDT-DNA) dans un vecteur pJet2.1.

Pour la construction du plasmide β_{23} , les amorces N64F & N19R ont été utilisées afin d'amplifier le gène *mCherry* depuis le plasmide pMC48. L'amorce N64F contient une queue portant une séquence aléatoire de 44 pb et une séquence de 20 pb homologue au vecteur. L'amorce N19R se lie à *mCherry* et possède une queue portant un terminateur de transcription et une séquence de 20 pb homologue au vecteur. Les amorces N64R et N22R ont été utilisées pour amplifier le vecteur à partir du plasmide pBCJ927. Les deux fragments d'ADN ont été assemblés par ligation (HiFi, NEB®).

Pour la construction du plasmide β_{22} contenant le gène *emGFP* dont le promoteur est muté, le plasmide BCJ927 a été amplifié par PCR inverse en utilisant les amorces N53F & N53R. Ces

amorces contiennent une séquence de remplacement pour p3 et 20 pb d'homologie entre eux. Une étape de ligation a permis de circulariser le plasmide.

4.1.2 Construction des souches

Pour la construction d'une souche exprimant *lambda-cl* depuis le promoteur inductible P_{rha} , le gène *lambda-cl* a été amplifié à partir d'un plasmide en utilisant les amorces A143R & A144R. Ce fragment contient également un gène de résistance à la phléomycine et a été directement inséré dans le génome de la souche *E. coli* MG1655 grâce au système de recombinaison LambdaRED (354,355), en remplaçant les gènes *rhaB*, *rhaA* et *rhaD*. La phléomycine a été supprimée du génome de la bactérie par recombinaison homologue d'un fragment d'ADN contenant un gène de résistance à la kanamycine, flanqué de part et d'autre par des sites de liaison pour l'enzyme I-SceI. L'induction du gène I-SceI permise par le système LambdaRED a permis de couper le chromosome de part et d'autre du gène de résistance à la kanamycine. Après religation du chromosome, la souche présentant *lambda-cl* sans gène de résistance antibiotique artificiellement introduit a été nommée $\alpha 1$.

Le système LambdaRED a été utilisé pour l'intégration de la cassette BCJ932 dans les différents loci. Cette cassette a été générée par PCR sur le plasmide pBCJ932 en utilisant les amorces répertoriées dans le Tableau 3 pour cibler chaque locus. Les fragments PCR ont été insérés dans la souche *E. coli* MG1655 pour les souches n'exprimant pas Lambda-cl (2 à 36) et dans la souche $\alpha 1$ pour les souches exprimant Lambda-cl ($\alpha 2$ à $\alpha 23$).

Le système LambdaRED a également été utilisé pour l'intégration dans les différents loci de la cassette BCJ927 amplifiée à partir du plasmide pBCJ927 en utilisant les amorces listées dans le Tableau 3. Pour la construction des souches $\alpha 61$ à $\alpha 64$, des réactions de PCR ont été réalisées sur le plasmide $\beta 5$ en utilisant les amorces listées dans le Tableau 3 puis insérées dans les différents loci par LambdaRED. Pour la construction des souches $\alpha 66$ à $\alpha 69$, des réactions de PCR ont été réalisées sur le plasmide $\beta 7$ en utilisant les amorces listées dans le Tableau 3 puis insérées dans les différents loci par LambdaRED. Pour la construction des souches $\alpha 120$ à $\alpha 123$, des réactions de PCR ont été réalisées sur le plasmide $\beta 16$ en utilisant les amorces listées dans le Tableau 3 puis insérées dans les différents loci. Pour la construction des souches $\alpha 125$ à $\alpha 168$, des réactions de PCR ont été réalisées sur le plasmide 48 pour amplifier *mCherry* en

utilisant les amorces listées dans le Tableau 3. Lorsque c'est nécessaire, ces amorces possèdent une queue contenant une séquence mutée en remplacement du promoteur. Ces produits PCR ont été insérés dans les loci cibles par LambdaRED dans les souches $\alpha 26$ ou $\alpha 36$. Pour la construction des souches $\alpha 153$ et $\alpha 154$, des réactions de PCR ont été réalisées sur le plasmide $\beta 22$ en utilisant les amorces listées dans le Tableau 3 puis insérées dans les différents loci par LambdaRED. Pour la construction des souches $\alpha 169$ et $\alpha 170$, des réactions de PCR ont été réalisées sur le plasmide $\beta 23$ en utilisant les amorces listées dans le Tableau 3 puis insérées dans les différents loci par LambdaRED. Pour la construction des souches $\alpha 179$ et $\alpha 184$, des réactions de PCR ont été réalisées sur le plasmide $\beta 5$ en utilisant les amorces listées dans le Tableau 3 pour amplifier le GRS seul. Le produit PCR a ensuite été intégré dans les souches $\alpha 129$, $\alpha 131$, $\alpha 149$, $\alpha 150$, $\alpha 167$, ou $\alpha 168$ par le système LambdaRED.

Les amorces et construction génétiques ont été conçues en utilisant le logiciel MacVector et la base de données EcoCyc (356)

4.2 Conditions de culture cellulaire

Le milieu Luria Broth (LB) a été utilisé pour les cultures de routine des souches d'*E. coli*. Le milieu M9 avec glycérol a été utilisé pour la culture des souches avant extraction ARN et mesures de qPCR. Le milieu est composé de 6 g/L Na₂HPO₄ x 2H₂O, 3 g/L KH₂PO₄, 0.5 g/L NaCl, 0.002% Casamino acids, 2 mM MgSO₄, 100 μ M CaCl₂ et 0.8% de glycerol comme source de carbone. Des antibiotiques ont été ajoutés aux milieux en fonction des besoins selon ces concentrations : spectinomycine (60 μ g/mL), kanamycine (25 μ g/mL pour les insertions génomiques, 50 μ g/mL pour les plasmides), phleomycine (10 μ g/mL), gentamicine (10 μ g/mL), and ampicilline (100ug/ml). Les cultures ont été réalisées à 30 ou 37°C avec 200 rpm d'agitation.

4.3 Méthodes de biologie moléculaire

Pour les amplifications PCR de routine, les enzymes ADN polymérases Q5 et OneTaq ont été utilisées en respectant le protocole du fournisseur (NEB®). Les produits PCR ont été purifiés par utilisation du kit Monarch DNA CleanUp (NEB®) et les plasmides ont été purifiés

par utilisation du kit Monarch Plasmid Miniprep (NEB®) en respectant le protocole du fournisseur. Les extractions d'ADN génomique ont été réalisées par utilisation du kit PureLink™ Genomic DNA Extraction (ThermoFisher®).

Les souches électrocompétentes ont été obtenues par des techniques courantes décrites dans la littérature (357). Brièvement, les souches bactériennes ont été mises en précultures dans un milieu LB avec l'antibiotique correspondant. Ces cultures ont ensuite été diluées au 400^{ème} pour inoculer 200mL de LB contenant l'antibiotique correspondant et si nécessaire de l'IPTG (pour le système LambdaRED) puis cultivées jusqu'à $DO_{600} = 0,5$. Les culots bactériens ont été obtenus par centrifugation à 3 900xg à 4°C pendant 6 minutes. Après 2 étapes de lavage avec du glycérol 10% froid, les culots ont été resuspendus dans 2mL de glycérol 10%.

Les électroporations ont été réalisées dans un électroporateur Eppendorf 2510 en respectant le protocole du fournisseur. Avant les étapes d'électroporation dans les constructions utilisant le système LambdaRED, les produits PCR ont été digérés avec l'enzyme DpnI pendant 1 heure à 37°C pour éliminant tout reste de plasmide matrice. Ensuite, 200ng de fragment ADN purifié contenant la construction d'intérêt ont été mélangés avec 50µL de cellules électrocompétentes dans des cuves de 0,1cm. Les cellules ont subi une électroporation à 1,8kV et ont été immédiatement placées dans 1mL de LB. Ces cellules ont été cultivées à 30°C avec 200rpm pendant 3h puis 100µL de culture ont été utilisés pour ensemercer des boîtes de pétri contenant du LB + agar avec l'antibiotique correspondant. Les colonies présentes sur les boîtes de pétri après une nuit de croissance ont été utilisées pour ensemercer 2mL de LB. Après 24h de croissance, l'ADN génomique a été extrait et utilisé comme matrice pour une réaction de PCR permettant d'amplifier la région d'intégration cible de la construction. Le séquençage des amplicons a permis de confirmer les souches.

Les migrations sur gels d'électrophorèses ont été réalisées dans une machine RunOne Electrophoresis System (Embitec) en utilisant un colorant Thermo Scientific Loading Dye, un tampon TBE (Tris, Boric Acid, EDTA) et un gel d'agarose 1 ou 2% en fonction de la taille du fragment (2% si <500bp, 1% pour >500bp) pendant 20 minutes à 100V. Les gels ont été colorés au Bromure d'Ethidium (Sigma) pour les ARN ou au Midori Green (Nippon Genetics) pour les ADN, avant d'être visualisés sous lampe UV ou dans un lecteur G-box iChemi (Syngene).

4.4 Extraction ARN

300µL de milieu M9 + Glycérol ont étéensemencés avec les souches d'intérêt et cultivés pendant une nuit à 37°C. Le jour suivant, 200µL de ces cultures ont été utilisés pour ensemencer 10mL de M9 + Glycérol dans lesquels 10mM de rhamnose ont été ajoutés pour induire l'expression de *lambda-cl* lorsque nécessaire. Les cultures ont été stoppées lorsque la $DO_{600} = 0,55$ et centrifugées pendant 10 minutes à 3 900xg à 4°C. Les culots ont été rapidement gelés par bain dans un mélange éthanol/glacé carbonique et stockés à -80°C. Pour l'extraction de l'ARN total, les culots ont été transférés sur de la glace et resuspendus dans 1mL de Ribozol RNA Extraction Reagent (VWR). L'extraction a été réalisée en suivant le protocole du fournisseur. Les culots d'ARN ont été resuspendus dans 225µL d'eau ultrapure et traités à la DNase I (NEB) en suivant le protocole du fournisseur. L'ARN a ensuite été précipité par addition de 20µL d'acétate de sodium et 500µL d'isopropanol. Après une étape de centrifugation à 21 130xg pendant 30 minutes, le culot d'ARN a été lavé avec 500µL d'éthanol 75% et resuspendu dans 80µL d'eau. L'intégrité, la qualité, et la quantité d'ARN purifié a été déterminée par gel d'électrophorèse et mesures de Nanodrop.

4.5 Transcription inverse et PCR quantitative (RT-qPCR)

500ng d'ARN total ont été utilisés pour la transcription inverse en utilisant le kit Protoscript II RT en suivant le protocole du fournisseur (NEB®). Les ADNc obtenus ont été purifiés par utilisation du kit GeneJet PCR Purification (Thermofisher®) et élués dans un volume final de 50µL. Les échantillons ont été dilués 10 fois et les qPCR ont été réalisées en utilisant le kit SYBR Premix Ex Taq (Takara) en suivant le protocole du fournisseur. Les sondes utilisées sont listées dans le Tableau 3. La qPCR a été réalisée sur une machine Realplex2 Mastercycler (Eppendorf®) en suivant le protocole du fournisseur et les paramètres optimisés suivants : 40 cycles avec dénaturation pendant 5 secondes à 95°C, hybridation des sondes pendant 30 secondes à 95°C et extension à 72°C pendant 20 secondes. Un gamme étalon a été utilisée dans chaque plaque de mesure de qPCR pour quantifier les signaux. Tous les échantillons ont été quantifiés en duplicat technique sur la plaque de mesure. La mesure pour chaque construction a été répétée au minimum 3 fois avec des cultures indépendantes. Au total, au

minimum 6 répétitions ont été réalisées : 2 techniques et 3 biologiques. L'erreur standard est inférieure à 25% pour toutes les valeurs moyennes.

4.6 Analyse de données

Les valeurs absolues d'expression des gènes cibles ont été obtenues par quantification avec une courbe étalon. L'analyse des données a été réalisée sur le logiciel Microsoft Excel. Afin d'obtenir le nombre de transcrits par cellule, la formule suivante a été utilisée :

Nombre de copies =

$$(X \text{ nanogrammes} * \text{Nombre d'Avogadro}) / (\text{Masse Moléculaire} * 1 \times 10^9)$$

X correspond à la quantité d'amplicons obtenue par qPCR, le nombre d'Avogadro ($6,0221 \times 10^{23}$) correspond au nombre de molécules par mole, la masse moléculaire de emGFP est 233703,9, la masse moléculaire de mCherry est 232269,38, la multiplication par 1×10^9 permet d'obtenir le nombre de molécules par nanogrammes d'ARN total. Ce nombre est ensuite divisé par 10 000 pour obtenir le nombre de molécules par cellule.

5 – DISCUSSION

5 DISCUSSION

Notre travail a permis de mettre en lumière un nouveau mécanisme épigénétique de régulation de l'expression des gènes bactériens. Cependant, il est essentiel d'augmenter la robustesse de certains résultats expérimentaux que nous avons obtenus, d'abord par l'intermédiaire de tests statistiques pour les constructions présentant des expressions géniques peu différentes, mais aussi par de nouvelles mesures d'expressions dans certaines constructions pour éclaircir les résultats inattendus. Par exemple, la construction à 2 gènes dont le promoteur de *mCherry* est muté (p3:emGFP/promoteur muté:mCherry) a présenté un niveau anormalement élevé d'expression du gène *mCherry*. Pourtant, le terminateur utilisé pour arrêter la transcription émanant de *emGFP* a été caractérisé comme étant très efficace. Nous pensons que l'expression de *mCherry* dans cette construction est due à la présence au niveau de la position -40 dans la séquence mutée d'un motif 5'-TGTAAC-3' semblable à la séquence consensus de la boîte -10 des promoteurs bactériens. Il est essentiel de supprimer cette séquence et de procéder à de nouvelles mesures dans cette construction. Enfin, des mesures d'expression supplémentaires sont nécessaires pour les souches dans lesquelles des mutants inactivant tel ou tel aspect de la fonction de lambda-cl sont exprimés. En effet, une construction p3:emGFP insérée dans des souches exprimant chacun des 3 mutants (P158T ; Y210H ; S228R) a donné des résultats inattendus (données non montrées). Dans ces souches, l'expression de *emGFP* est semblable à l'expression en présence d'une boucle d'ADN. Des mesures complémentaires devraient permettre d'appuyer les observations faites sur le rôle de la formation de la boucle dans l'homogénéisation de l'expression génique (Figure 43).

5.1 L'expression génique dépend fortement du contexte topologique local

La transcription d'un gène est souvent présentée comme directement contrôlée par son promoteur et l'action de facteurs de transcription. Cela laisse penser à un mécanisme simpliste dans lequel l'expression serait soit induite soit réprimée. Dans la réalité, le niveau d'expression d'un gène est la résultante de différentes tendances. Malgré le nombre important d'études conduites depuis plusieurs décennies, l'ensemble des facteurs impliqués dans la transcription n'est pas connu et loin d'être maîtrisé. Ceci est d'autant plus frustrant que des études récentes

montrent que ces facteurs jouent un rôle important dans l'expression génique. En effet, l'expression d'un même gène, sous contrôle d'un même promoteur, pouvait varier fortement en fonction du site dans lequel il a été inséré dans le génome (358–361). Une explication proposée à cette variabilité concerne l'accessibilité de la région promotrice au facteur de transcription, qui serait changeante en fonction du locus d'insertion, à cause des différentes topologies chromosomiques et concentrations locales de facteur de transcription. Cependant, cette variabilité est également observée lorsque le gène étudié n'est pas connu pour être régulé par un facteur de transcription. Cela suppose la présence de mécanismes locaux encore non élucidés impliqués dans l'expression génique. Dans leurs travaux, Bryant et associés ont observé, sur un nombre de sites trop limité pour tirer des conclusions fermes, que les loci génomiques dans lesquels le gène rapporteur *gfp* était le moins exprimé correspondaient à des régions présentant une forte concentration de NAP liées à l'ADN (359). Lorsque le gène rapporteur remplace ces régions, au lieu d'y être intégré, son expression augmente. L'amuisement du gène intégré dans ces régions est probablement permis par l'architecture de l'ADN au niveau du site d'insertion définie elle-même par la liaison de NAP (362).

Nous avons également observé que l'expression d'un gène rapporteur, *emGFP*, dépend fortement de son site d'insertion. Cette variabilité est particulièrement forte (facteur 1 300) quand *emGFP* est sous le contrôle d'un promoteur faible et encore très importante (facteur 100) quand ce gène est sous contrôle d'un promoteur fort. L'activité d'un promoteur faible est donc plus sensible au contexte local que celle d'un promoteur fort et nos données suggèrent que cette sensibilité réduite des promoteurs forts résiderait dans leur résistance aux signaux locaux inhibiteurs. Cependant, un effet de saturation de l'expression au niveau des promoteurs fort est également plausible.

Le chromosome d'*E. coli* peut présenter jusqu'à environ 400 microdomaines (363). Les surenroulements y sont emprisonnés par la liaison dynamique de différentes protéines bivalentes qui forment des boucles de différentes taille et robustesse dans l'ADN. Les quelques 10 gènes contenus dans ces microdomaines sont hétérogènes en termes de séquence, de longueur, d'expression, de régulation et d'orientation relative. Chaque microdomaine présente donc une topologie qui peut être plus ou moins variable inter- et intra-microdomaine. Ainsi, l'insertion aléatoire d'un gène d'intérêt dans différents loci va positionner ce gène dans une multitude de topologies locales différentes de l'ADN. Nous proposons que ces variabilités locales de topologie jouent un rôle important dans la sensibilité de l'expression génique à la

position dans le génome, en affectant le recrutement de l'ARN polymérase au niveau des promoteurs et/ou en affectant les étapes d'initiation et d'élongation de la transcription

5.2 Les gènes situés dans une boucle d'ADN artificiellement induite ne sont plus sensibles au contexte topologique local

Dans le but de comprendre l'origine de la variabilité due au contexte génomique, nous avons utilisé la protéine Lambda-cl pour générer un microdomaine artificiel emprisonnant le gène rapporteur *emGFP*. Il est attendu que ce système induit un micro-environnement topologique similaire quel que soit sa localisation dans le chromosome d'*E. coli*. Alors que les niveaux d'expression de *emGFP* sous contrôle du promoteur p1 sont très variables en absence de boucle d'ADN, nous avons observé que l'emprisonnement de cette unité de transcription dans le microdomaine artificiel Lambda-cl réduit la variabilité et homogénéise son expression à 0,05 TPC dans les 16 loci testés. Des expériences contrôles avec des mutants d'oligomérisation de Lambda-cl montrent que l'effet de cette protéine sur l'expression du gène rapporteur dépend de sa capacité à ponter les séquences situées de part et d'autre du gène rapporteur et donc, dépend de sa capacité à former des boucles dans l'ADN. De manière intéressante, l'homogénéisation de l'expression autour de la valeur de 0,05 TPC a lieu que l'expression du gène en absence de boucle soit inférieure ou supérieure à cette valeur moyenne. Cette valeur moyenne de 0,05 TPC, également relevée dans l'étude de transposition par mutagenèse, semble être le niveau basal d'expression à partir du promoteur p1. Toute variation à la hausse ou à la baisse par rapport à cette valeur moyenne semble être le résultat de l'impact du contexte génomique sur l'expression du gène rapporteur. Dans leur ensemble, ces résultats indiquent que la variabilité d'expression au sein du génome est due à une variabilité des contraintes topologiques locales et qu'une boucle d'ADN isole le gène qu'elle porte de ces contraintes locales. Cette idée est appuyée par les mesures d'expression lorsque le gène rapporteur est sous contrôle d'un promoteur fort. Toutefois, dans ces nouvelles conditions, l'expression du gène rapporteur dans la boucle est toujours inférieure à l'expression en absence de boucle (Figure 30) : l'expression du gène rapporteur semble donc bridée à l'intérieur de la boucle d'ADN. Nous formulons l'hypothèse que ce plafonnement d'expression est dû à une accumulation de surenroulements positifs causés par la forte transcription du

gène rapporteur qui, en retour, inhibe la transcription du gène. Si tel est le cas, alors une relaxation des surenroulements devrait lever ce bridage.

5.3 La diminution des supertours positifs au sein de la boucle augmente l'expression de gènes sous contrôle d'un promoteur fort

Lorsque le pontage est induit, le contexte génomique ne semble pas influencer l'expression du gène dans la boucle d'ADN. Les éléments régissant l'expression génique sont alors les éléments internes à la boucle tels que la force du promoteur ou la topologie de la séquence. Plus le gène est transcrit, plus il y aura induction de surenroulement (252) et plus la transcription se fera en rafales (267). L'ADN bactérien étant naturellement surenroulé négativement, les surenroulements négatifs ont tendance à diffuser le long de la molécule plus facilement que les surenroulements positifs, qui vont s'accumuler en aval de la fourche de transcription. Pour augmenter l'expression d'un gène exprimé à partir d'un promoteur fort et situé au sein d'une boucle, il semble donc essentiel de diminuer les surenroulements positifs à l'intérieur de la boucle, notamment en aval du gène *emGFP*. Dans une molécule d'ADN présentant une forte accumulation de surenroulements, un simple clivage simple brin devrait potentiellement permettre de relaxer le duplex ADN. Malgré nos investigations, nous n'avons pas trouvé dans la littérature scientifique d'autres sites de reconnaissance pour topoisomérase que celui de la gyrase. Nous nous sommes donc limités à l'utilisation d'un site de reconnaissance pour l'enzyme gyrase dans nos constructions, afin que celle-ci empêche l'accumulation de surenroulements positifs dans la boucle. La présence du site de reconnaissance pour gyrase (GRS) à l'intérieur de la boucle a permis dans n'importe quelle configuration d'augmenter l'expression du gène rapporteur *emGFP* de 20 à 700%, tout en gardant une expression fortement homogène entre les différents sites chromosomiques testés.

Les observations réalisées avec les constructions génétiques contenant le GRS suggèrent que le surenroulement positif est un acteur majeur des mécanismes de régulation des niveaux d'expression des gènes. Ainsi, tous les éléments entraînant une modification dans l'équilibre local entre les surenroulements positifs et négatifs peuvent influencer sur l'expression des gènes. Dans ce contexte, deux unités transcriptionnelles proches pourraient affecter

mutuellement leur expression génique l'une de l'autre sans interaction moléculaire directe, en modifiant la topologie locale par leur transcription.

5.4 Le changement de surenroulement causé par l'expression d'un gène affecte l'expression du gène voisin et vice versa

Nos résultats suggèrent que la variabilité d'expression au sein du génome dépend de contraintes topologiques locales. Les données de la littérature montrent que ces contraintes sont influencées par la présence, l'orientation, et la transcription d'unités transcriptionnelles singulières proches. Afin d'étudier plus systématiquement les interférences transcriptionnelles entre deux gènes proches dans un contexte affranchi des contraintes topologiques locales, nous avons choisi d'introduire un second gène dans des boucles d'ADN. Deux gènes contigus peuvent se présenter sous 3 configurations relatives différentes : en tandem, convergents, ou divergents. Le second gène, *mCherry*, a donc été introduit sous contrôle d'un promoteur faible (p1) ou d'un promoteur fort (p10) dans les différentes constructions de façon à respecter ces 3 configurations. La comparaison des données d'expression de *emGFP* à deux loci différents et en présence de *mCherry* sous contrôle de p1, p10 ou d'une séquence mutée a permis d'observer l'effet de la force d'un promoteur sur l'expression d'un gène voisin à l'intérieur de d'une boucle d'ADN.

Lorsque les gènes sont placés en tandem, l'expression de *emGFP* augmente avec le niveau d'expression de *mCherry*. Selon le « twin-supercoiled domain model », il est attendu que le surenroulement négatif induit par l'expression de *mCherry* s'oppose aux surenroulements positifs dus à la forte expression de *emGFP* dans la région intergénique et favorise l'expression de ces 2 gènes (Figure 46a,b). La diminution des contraintes topologiques intergéniques est donc bénéfique pour l'expression des deux gènes. Pour *emGFP*, l'ARN polymérase subit moins de contraintes dans sa progression, pour *mCherry* le recrutement et/ou le passage en complexe ouvert sont facilités.

Lorsque les gènes sont transcrits de manière convergente, une plus forte transcription de *mCherry* se traduit par une diminution de l'expression de *emGFP* de ~25%. Cette observation peut être expliquée par l'accumulation dans la région intergénique des surenroulements positifs générés par la transcription des 2 gènes qui inhiberait la progression de l'ARN polymérase le long de *emGFP* et peut-être également l'initiation de la transcription de ce gène

(Figure 46c, d). De manière intéressante, on peut noter que cette baisse d'environ 0,2 TPC observée pour *emGFP* est permise par une augmentation de $\sim 0,16$ TPC du gène *mCherry*. Cette observation suggère que la transcription est directement proportionnelle au niveau de surenroulements accumulés.

Dans leur étude, Bryant et collaborateurs (359) ont également observé expérimentalement que l'expression d'un gène pouvait réprimer l'expression d'unités transcriptionnelles convergentes. Leurs travaux montrent que l'induction du gène *gfp* entraîne une répression d'un facteur 2 de l'opéron *melAB* convergent. C'est également le cas lorsque le gène *mntH* est réprimé, le gène *gfp* convergent est exprimé 3 fois plus. Nos observations sont compatibles avec leurs résultats. La configuration convergente entrave la transcription optimale des gènes.

Une autre étude, réalisée par Yeung et collaborateurs (364), a défini la configuration convergente comme étant celle permettant la plus forte expression. Cependant, dans cette étude le vecteur utilisé pour étudier les gènes convergents était un plasmide et non le chromosome. Malgré les similarités avec un chromosome, les plasmides présentent des caractéristiques très différentes, notamment au niveau biophysique. La petite taille du plasmide n'est pas représentative des conditions naturellement retrouvées dans le chromosome d'*E. coli*. Ces plasmides ne présentent probablement pas de points d'ancrage qui emprisonnent les surenroulements à l'image des microdomaines, ainsi les surenroulements peuvent diffuser tout au long de la molécule, ayant un impact différent sur la transcription par rapport aux observations réalisées sur le chromosome naturel. De plus, leurs observations ne sont pas en adéquation avec les constats réalisés sur l'organisation naturelle des gènes au sein des chromosomes bactériens. Les unités transcriptionnelles sont majoritairement organisés en tandem (276,277). Cela est expliqué par une optimisation de l'expression génique façonnée par l'évolution, évitant la collision entre différentes machineries cellulaires et différentes ARN polymérases en action. Nous postulons que le « twin-supercoiled domain model » de la transcription a également joué un rôle dans cette optimisation par l'évolution. Chez les bactéries les cistrons impliqués dans une fonction commune sont souvent placés en opéron, donc en tandem. L'expression des gènes impliqués dans une même fonction doit être simultanée. Placer ces gènes dans une configuration convergente serait néfaste pour la fonction car l'expression d'un gène altérerait celle de l'autre. En outre, l'évolution a pu placer des gènes dans une configuration convergente afin de réguler de manière parcimonieuse et

peu coûteuse pour la cellule l'expression d'un gène lorsqu'un autre est exprimé. Les observations que nous avons réalisées sur les données de Chip-seq et de RNA-seq de la protéine LRP (348) ont confirmé les observations réalisées lors de nos expérimentations. Lorsque deux gènes convergents sont compris entre 2 sites de liaison de LRP, on voit que l'augmentation de l'expression d'un gène entraîne directement une diminution de l'expression du second gène, conformément à nos observations. Plus précisément, lorsque 2 gènes sont convergents, la diminution d'expression de l'un d'entre eux peut être attribuée soit à sa répression par un acteur extérieur, soit à l'induction de l'expression du second gène par un acteur extérieur, soit à l'action combinée de ces 2 paramètres (moins parcimonieux). Grâce à ce mécanisme, 2 gènes peuvent être régulés par l'action d'une seule molécule sur une seule cible, ce qui est énergétiquement moins coûteux pour la cellule.

L'ensemble de ces résultats donne une image de proportionnalité entre l'accumulation de surenroulements positifs et l'expression génique permise. Toutefois, le lien qui unit les niveaux de surenroulement négatifs à l'expression génique semble bien plus complexe. Les constructions dans lesquels les gènes sont placés dans une configuration divergente ont montré des effets variables du surenroulement négatif en fonction du promoteur utilisé.

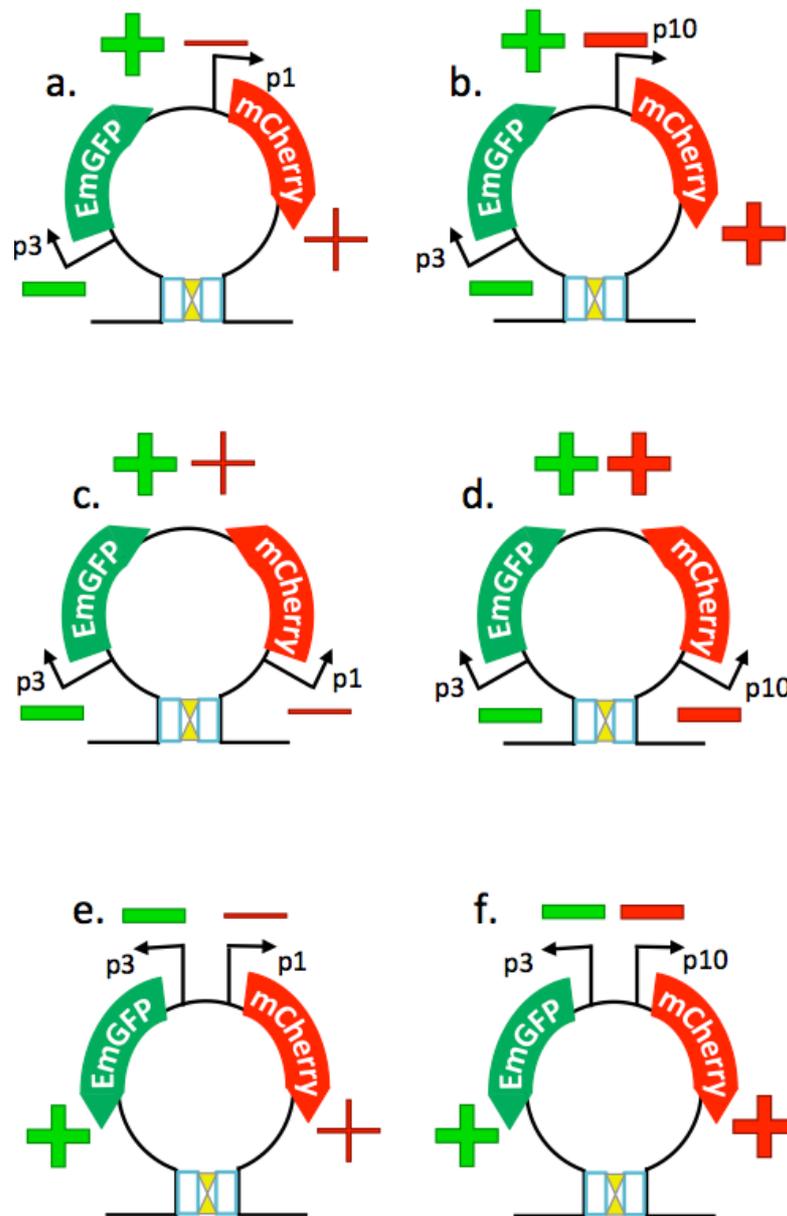


Figure 46 : Représentation de l'accumulation de surenroulements dans les boucles à 2 gènes

a. Construction comprenant les gènes en tandem avec mCherry sous contrôle du promoteur faible p1. *b.* Construction comprenant les gènes en tandem avec mCherry sous contrôle du promoteur fort p10. *c.* Construction comprenant les gènes convergents avec mCherry sous contrôle du promoteur faible p1. *d.* Construction comprenant les gènes convergents avec mCherry sous contrôle du promoteur fort p10. *e.* Construction comprenant les gènes divergents avec mCherry sous contrôle du promoteur faible p1. *f.* Construction comprenant les gènes divergents avec mCherry sous contrôle du promoteur fort p10. *g.* Construction comprenant les gènes en tandem avec une séquence aléatoire remplaçant le promoteur du gène mCherry. Les flèches angulaires représentent les promoteurs, la flèche verte représente emGFP, la flèche rouge représente mCherry, les lignes et cercles noirs représentent l'ADN, les rectangles représentent les sites de liaisons pour la protéine Lambda-cl, les triangles jaunes représentent la protéine Lambda-cl, les - et + verts ou rouges représentent le surenroulement de l'ADN.

5.5 Le surenroulement négatif affecte la transcription de manière spécifique à chaque promoteur

La structure d'un promoteur a été initialement définie par la séquence ADN elle-même. En définissant arbitrairement la taille d'un promoteur à 40 nucléotides, dont une dizaine seulement seraient conservés, il y aurait plus d'un milliard de milliard de séquences possibles. En conséquence, chaque promoteur est unique et possède une structure particulière qui peut lui conférer des sensibilités variées aux surenroulements locaux. Par exemple, ces surenroulements peuvent modifier l'alignement des boîtes -35 et -10, la taille physique du spacer, la courbure de la séquence, et/ou l'accessibilité à certaines séquences.

Dans nos expérimentations, lorsque 2 gènes sont présents au sein de la boucle d'ADN et qu'ils sont placés de manière divergente, une plus forte expression du gène *mCherry* résulte en une augmentation de l'expression de *emGFP*. Ceci semble dû au fait que la transcription plus forte du gène *mCherry* entraîne une accumulation plus forte de surenroulements négatifs dans la région promotrice (Figure 46e,f), ayant pour conséquence un effet positif sur la transcription du gène *emGFP* à partir du promoteur p3. Ce résultat est confirmé par les expériences incluant le GRS dans les constructions (GRS/p3:emGFP et p3:emGFP/GRS/p10:mCherry) : la présence du GRS, et donc la probable l'introduction de surenroulements négatifs dans la région promotrice, induit une augmentation de l'expression du gène *emGFP* à partir du promoteur p3. Cependant, l'augmentation des niveaux de surenroulements négatifs n'a pas le même effet sur le promoteur p10. En effet, dans la construction à 2 gènes divergents, l'introduction du GRS entraîne une modification mineure des niveaux d'expression du gène *mCherry* sous contrôle du promoteur p10 (Figure 40).

La Figure 47 est une représentation qualitative des probables niveaux d'expression permis par un promoteur en fonction des niveaux de surenroulement négatif présents en son sein. Par exemple, dans les constructions divergentes p3:emGFP/p1:mCherry (~0,5 TPC pour *emGFP*), p3:emGFP/p10:mCherry (~1 TPC pour *emGFP*), et p3:emGFP/GRS/p10:mCherry (~2,7 TPC pour *emGFP*), le promoteur p3 se trouverait respectivement au niveau du point A, puis B, puis C (ou D).

En revanche, les constructions arborant le promoteur p10 ne présentent pas d'augmentation d'expression lorsque les niveaux de surenroulement négatif sont supérieurs. Dans les constructions p3:emGFP/p10:mCherry (~0,36 TPC pour *mCherry*), et

p3:emGFP/GRS/p10:mCherry (~0,36 TPC pour *mCherry*), le promoteur p10 serait au niveau du point C en absence de GRS, et au niveau du point D en présence de GRS, expliquant les niveaux d'expression équivalents malgré des niveaux de surenroulements négatifs différents.

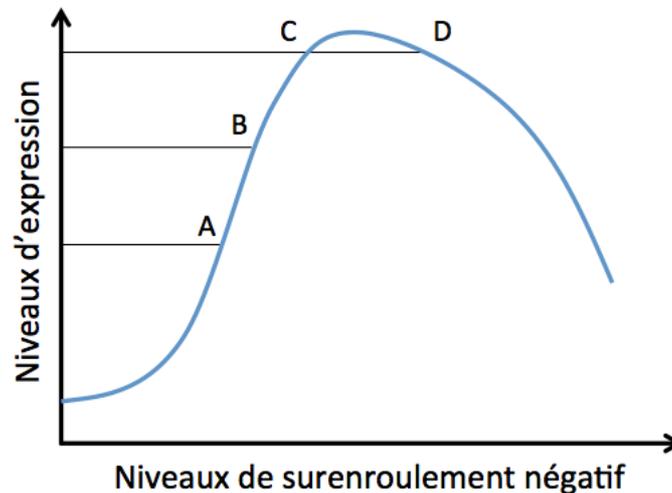


Figure 47 : Représentation de l'activité d'un promoteur en fonction des niveaux de surenroulement négatif accumulés dans et autour de sa séquence.

L'axe des abscisses représente les niveaux de surenroulement négatif dans la région promotrice. L'axe des ordonnées représente les niveaux d'expression du gène induits par le promoteur.

5.6 Les boucles d'ADN sont des régulateurs épigénétiques de l'expression de nombreux gènes chez *E. coli*

Nous avons observé que l'expression génique au sein d'une boucle d'ADN dépend des surenroulements positifs qui s'y accumulent et est par contre insensible aux facteurs affectant la transcription dans la région du chromosome portant la boucle. Nous avons voulu déterminer si *E. coli* utilise de telles boucles pour réguler l'expression de gènes de façon épigénétique. Pour cela, nous avons utilisé une approche bio-informatique sur des données Chip-seq et RNA-seq du régulon LRP (348).

LRP est une protéine de *E. coli* importante pour la régulation transcriptionnelle des gènes lorsque les acides aminés se raréfient dans le milieu de culture et pour l'organisation du chromosome. Cette protéine est principalement présente sous forme d'hexadecamère puis d'octamère. La quantité de LRP libre dans la cellule diminue d'un facteur 16 lorsque le milieu

de culture est enrichi (351), et la présence de leucine dans le milieu de culture induit la dissociation des hexadecamères en octamères liés à la leucine. Les résultats de Chip-seq ont montré que LRP est constamment présente sur certains sites de liaison dans le génome même si la régulation transcriptionnelle n'a lieu qu'à l'approche et durant la phase stationnaire (348). Nous postulons qu'en phase exponentielle LRP est sous forme d'octamères liés à la leucine. Sous cette forme, LRP se lierait à l'ADN sans pouvoir former de boucles dans l'ADN. Par contre, à l'approche et durant la phase stationnaire, la chute en leucine provoquerait la dissociation de la leucine avec les octamères de LRP qui pourraient alors former des hexadecamères et, concomitamment, des boucles dans l'ADN (200,350,352). Ainsi, à l'image de Lambda-cl, l'altération de l'oligomérisation de LRP modulerait sa capacité à former des boucles d'ADN.

E. coli utilise-t-elle ces propriétés d'oligomérisation et de pontage de LRP pour réguler l'expression d'un grand nombre de gènes à travers un processus épigénétique impliquant la formation de boucles ?

Les résultats que nous avons obtenus par l'étude bio-informatique montrent que lorsqu'un gène est flanqué de part et d'autre par des sites de liaison pour LRP, l'expression du gène est réprimée dans ~90% des cas. Au niveau de ces gènes, l'augmentation du signal de liaison à l'ADN de LRP en fonction de la phase de croissance entraîne une plus forte répression. Ces résultats suggèrent donc fortement que LRP utilise les boucles d'ADN pour réguler l'expression de nombreux gènes en les piégeant dans de petites boucles d'ADN dans certaines conditions de croissance.

LRP n'est qu'une parmi plusieurs NAP ayant été décrites comme impliquées dans l'organisation tridimensionnelle du chromosome en réponse à différents stimuli. De nombreuses études ont permis de mettre en lumière le rôle et les mécanismes d'actions des NAP dans la compaction du chromosome. L'activité de ces protéines change en fonction des conditions de culture et de la phase de croissance. De plus, des changements conformationnels induits par des modifications post-traductionnelles (phosphorylations, acétylations, ...) et probablement par des interactions avec des métabolites (comme proposé ci-dessus pour LRP) sont capables d'affecter les propriétés de liaison à l'ADN de certaines NAP. Ces modifications se produisent très rapidement dans la cellule et pourraient expliquer la réponse rapide des bactéries aux changements environnementaux. Ainsi, en passant d'un état conformationnel à un autre suite à une simple modification post-traductionnelle, les NAP pourraient changer

l'organisation du génome et l'expression de centaines de gènes en quelques minutes. Formellement, ce mécanisme permet une réponse beaucoup plus rapide que n'importe quel mécanisme dépendant de la transcription puis la traduction de nouvelles protéines. Ce modèle est appuyé par les recherches récentes de Yus et collaborateurs (365), qui ont découvert que la régulation transcriptionnelle chez *Mycoplasma pneumoniae* dépend principalement de facteurs tels que le surenroulement de l'ADN et l'organisation génomique plutôt que de facteurs de transcription.

Ainsi, la combinaison des différentes NAP, leur production et leur conformation vont permettre de changer l'organisation 3D du chromosome et ainsi de réguler de manière vive et dynamique l'expression génique bactérienne en fonction des conditions de croissance et des signaux environnementaux. La Figure 48 est une représentation simpliste de ce modèle, dans laquelle la conformation d'une région de l'ADN est façonnée en fonction de l'action distincte de 3 protéines architectes. Un même gène peut donc se retrouver dans plusieurs contextes génomiques différents en réponse à un signal environnemental et être induit ou réprimé rapidement par la conformation et la topologie locale. En se basant sur les observations de nos travaux, nous estimons que le gène B aura un niveau d'expression significativement différent dans chacune des configurations présentées en Figure 48.

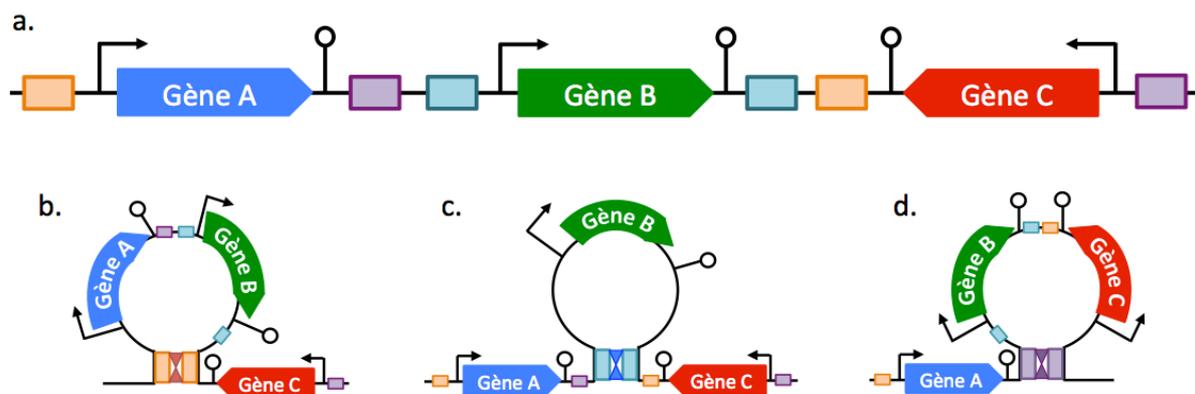


Figure 48 : Représentation schématique de l'organisation chromosomique potentielle permise par l'activité de 3 NAP.

Les flèches angulaires représentent les promoteurs, les épingles représentent les terminateurs de transcription, les flèches pleines représentent les gènes, les lignes et cercles noirs représentent l'ADN, les rectangles oranges, violets et bleus représentent les sites de liaisons pour chaque NAP, les triangles représentent chaque NAP.

5.7 Conclusion

Ce travail a permis une meilleure compréhension des mécanismes moléculaires impliqués dans les variations de l'expression génique dues au contexte génomique. Dans un premier temps, nous avons dressé à l'échelle génomique un aperçu de ce phénomène en utilisant la mutagenèse par transposition. Cette approche a permis d'avoir une vue globale de la variabilité de l'expression en fonction du site d'insertion du gène rapporteur et en fonction du promoteur. Les résultats que nous avons obtenus grâce à la formation de boucles artificielles ont permis de déterminer des causes importantes de la fluctuation de l'expression génique. La topologie locale est le moteur de cette variabilité. En se focalisant sur les niveaux de surenroulement dans la région promotrice ou dans la région en aval des gènes, nous avons développé des outils pour optimiser les niveaux et l'homogénéité de l'expression génique. L'utilisation de boucles d'ADN induites par Lambda-cl, couplée à la présence d'un GRS et d'un second gène dans la boucle, a permis de créer un microenvironnement optimal pour l'expression du gène *emGFP*, dont l'expression a augmenté jusqu'à 950% au niveau d'un même locus.

5.8 Perspectives

Nous avons envisagé d'introduire des boucles d'ADN artificielles emprisonnant des gènes naturels dans le chromosome de la bactérie. Notre objectif était de trouver une région présentant une organisation génique similaire à la représentation en Figure 48. Dans cette région, nous avons pour objectif d'insérer des sites de liaison pour Lambda-cl afin de produire 3 souches présentant chacune une conformation différente. Cette approche a soulevé beaucoup de difficultés, les gènes candidats devant répondre à de nombreuses caractéristiques : (i) présenter l'organisation souhaitée, (ii) ne pas être trop éloignés les uns des autres, (iii) ne pas être en opéron, (iv) être exprimés fortement, (v) posséder des terminateurs forts de transcription, (vi) ne pas posséder d'action de régulation entre eux, (vii) ne pas s'autoréguler. A ces difficultés se rajoute 2 inconnues : des phénomènes de régulation pas encore décrits, et la liaison de cette région par des NAP. Ainsi, une boucle d'ADN que nous aurions induit aurait potentiellement pu être physiquement altérée par la formation d'une

boucle naturelle. L'ensemble de ces difficultés nous a poussé à privilégier les analyses bio-informatiques réalisées sur LRP.

Les résultats bio-informatiques obtenus par l'analyse des données récoltées par Kroner et collaborateurs (348) prouvent que le mécanisme que nous suggérons, sur la base de l'induction contrôlée de boucles artificielles, serait utilisé naturellement chez les bactéries pour réguler épigénétiquement de nombreux gènes. Cependant, notre travail s'est limité à l'analyse du régulon de LRP dans une condition de culture seulement. Nous avons choisi le milieu le plus pauvre en nutriment car c'est probablement celui dans lequel LRP est le plus actif. Ce genre d'analyses pourrait être réalisé dans les milieux supplémenté et riche, mais aussi pour toutes les NAP capables de former des boucles dans le chromosome de la bactérie. Ces analyses pourraient ainsi donner une carte détaillée de la conformation et de la topologie du chromosome en entier en fonction des conditions de culture. Elles permettraient également de découvrir les NAP qui ont une fonction complémentaire et ainsi expliquer pourquoi la délétion de certaines NAP ne change pas ou peu la conformation du chromosome (366).

Une autre approche pour l'étude de l'impact de la topologie sur l'expression génique concerne les promoteurs. Certains promoteurs, tel que le promoteur $P_{\text{leu-500}}$ présent chez la bactérie *Salmonella typhimurium* (367), présentent une activité dépendante des niveaux de surenroulement de la molécule ADN. En utilisant un promoteur de ce type dans des constructions avec des gènes présentant différentes configurations (tandem, divergente) il serait théoriquement possible de mesurer le surenroulement positif et négatif provoqué par la transcription d'un gène en amont en tandem ou divergent. A titre d'exemple, des constructions en tandem du type $p1:\text{emGFP}/P_{\text{leu-500}}:\text{mCherry}$, $p3:\text{emGFP}/P_{\text{leu-500}}:\text{mCherry}$ et $p10:\text{emGFP}/P_{\text{leu-500}}:\text{mCherry}$ permettraient une mesure du surenroulement positif généré en aval par la transcription du gène *emGFP*. De la même façon, ces mêmes unités transcriptionnelles cette fois placées de manière divergentes permettraient une mesure du surenroulement négatif généré par la transcription du gène *emGFP*.

La durée de demi-vie des ARNm bactériens est relativement faible, en moyenne 2,4 minutes chez *E. coli*, mais elle connaît une grande variabilité. La stabilité d'un même ARNm peut varier pour plusieurs raisons: (i) la concentration d'un ARNm peut influencer sa stabilité (368), (ii) des protéines se liant aux ARN peuvent fixer la molécule et la rendre plus stable ou la déstabiliser (369,370), (iii) les ribosomes peuvent affecter la dégradation des ARNm (371). En

fonction de la région chromosomique, la concentration en protéines liant les ARNm ou en ribosomes peut varier. Quoique peu probable, cette éventuelle variation peut entraîner une stabilité différente des ARNm messenger en fonction du site d'insertion du gène. Pour vérifier cette possibilité, il serait utile de mesurer la stabilité des ARNm des gènes *emGFP* et *mCherry* dans les régions testées dans ce travail.

Un autre aspect des ARNm intéressant à étudier concerne leur taille. Lorsque l'ARN polymérase rencontre une forte accumulation de surenroulement durant l'élongation, la transcription peut être ralentie et stoppée. Ce phénomène donnerait lieu à des ARN tronqués. Une mesure de la taille de ces ARNm pourrait alors donner des indications sur les niveaux de surenroulement positifs présents dans la partie 3' des gènes.

6 – BIBLIOGRAPHIE

6 BIBLIOGRAPHIE

1. Fredrickson JK, Zachara JM, Balkwill DL, Kennedy D, Li SW, Kostandarithes HM, et al. Geomicrobiology of High-Level Nuclear Waste-Contaminated Vadose Sediments at the Hanford Site, Washington State. *Appl Environ Microbiol.* 2004 Jul;70(7):4230–41.
2. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: The unseen majority. *Proc Natl Acad Sci U S A.* 1998 Jun 9;95(12):6578–83.
3. J.P. Euzéby : Dictionnaire de Bactériologie Vétérinaire - PDF Téléchargement Gratuit [Internet]. [cited 2020 Feb 13]. Available from: <https://docplayer.fr/42172333-J-p-euzeby-dictionnaire-de-bacteriologie-veterinaire.html>
4. Le Minor L, Popoff MY, Bockemühl J. Supplement 1989 (n° 33) to the Kauffmann-White scheme. *Res Microbiol.* 1990 Nov;141(9):1173–7.
5. Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev.* 1998 Jan;11(1):142–201.
6. Bachmann BJ. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K12. In 1987.
7. Lawther RP, Calhoun DH, Gray J, Adams CW, Hauser CA, Hatfield GW. DNA sequence fine-structure analysis of *ilvG* (*IlvG*⁺) mutations of *Escherichia coli* K-12. *J Bacteriol.* 1982 Jan;149(1):294–8.
8. Liu D, Reeves PR. *Escherichia coli* K12 regains its O antigen. *Microbiol Read Engl.* 1994 Jan;140 (Pt 1):49–57.
9. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, et al. The Complete Genome Sequence of *Escherichia coli* K-12. *Science.* 1997 Sep 5;277(5331):1453–62.
10. Avery OT, Macleod CM, McCarty M. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *J Exp Med.* 1944 Feb 1;79(2):137–58.
11. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. The Structure and Function of DNA. *Mol Biol Cell* 4th Ed [Internet]. 2002 [cited 2020 Feb 19]; Available from: <https://www.ncbi.nlm.nih.gov/books/NBK26821/>
12. Ussery DW. DNA Structure: A-, B- and Z-DNA Helix Families. In: eLS [Internet]. American Cancer Society; 2002 [cited 2020 Mar 5]. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1038/npg.els.0003122>
13. Allemand JF, Bensimon D, Lavery R, Croquette V. Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proc Natl Acad Sci U S A.* 1998 Nov 24;95(24):14152–7.
14. de Vries R. DNA condensation in bacteria: Interplay between macromolecular crowding and nucleoid proteins. *Biochimie.* 2010 Dec;92(12):1715–21.
15. Pelletier DA, Hurst GB, Foote LJ, Lankford PK, McKeown CK, Lu T-Y, et al. A general system for studying protein-protein interactions in Gram-negative bacteria. *J*

Proteome Res. 2008 Aug;7(8):3319–28.

16. Hatfield GW, Benham CJ. DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu Rev Genet*. 2002;36:175–203.
17. Travers A, Muskhelishvili G. DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nat Rev Microbiol*. 2005 Feb;3(2):157–69.
18. Dillon SC, Dorman CJ. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol*. 2010 Mar;8(3):185–95.
19. Luijsterburg MS, White MF, van Driel R, Dame RT. The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol*. 2008 Dec;43(6):393–418.
20. Touzain F, Petit M-A, Schbath S, El Karoui M. DNA motifs that sculpt the bacterial chromosome. *Nat Rev Microbiol*. 2011 Jan;9(1):15–26.
21. McGary K, Nudler E. RNA polymerase and the ribosome: the close relationship. *Curr Opin Microbiol*. 2013 Apr;16(2):112–7.
22. Rauhut R, Klug G. mRNA degradation in bacteria. *FEMS Microbiol Rev*. 1999 Jun;23(3):353–70.
23. Sweetser D, Nonet M, Young RA. Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc Natl Acad Sci U S A*. 1987 Mar;84(5):1192–6.
24. Ebright RH. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol*. 2000 Dec 15;304(5):687–98.
25. Darst SA. Bacterial RNA polymerase. *Curr Opin Struct Biol*. 2001 Apr;11(2):155–62.
26. Borukhov S, Nudler E. RNA polymerase holoenzyme: structure, function and biological implications. *Curr Opin Microbiol*. 2003 Apr;6(2):93–100.
27. Gross CA, Chan C, Dombroski A, Gruber T, Sharp M, Tupy J, et al. The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol*. 1998;63:141–55.
28. Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, et al. Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLoS Biol* [Internet]. 2009 May [cited 2020 Mar 5];7(5). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2675907/>
29. Werner F, Grohmann D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol*. 2011 Feb;9(2):85–98.
30. Gruber TM, Gross CA. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol*. 2003;57:441–66.
31. Haugen SP, Ross W, Gourse RL. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Microbiol*. 2008 Jul;6(7):507–19.
32. Feklistov A, Sharon BD, Darst SA, Gross CA. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol*. 2014;68:357–76.
33. Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD. The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. *J Bacteriol*. 2000 Aug;182(15):4129–36.

34. Paget MSB, Helmann JD. The sigma70 family of sigma factors. *Genome Biol.* 2003;4(1):203.
35. Lisser S, Margalit H. Compilation of E. coli mRNA promoter sequences. *Nucleic Acids Res.* 1993 Apr 11;21(7):1507–16.
36. Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science.* 2002 May 17;296(5571):1285–90.
37. deHaseth PL, Helmann JD. Open complex formation by Escherichia coli RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. *Mol Microbiol.* 1995 Jun;16(5):817–24.
38. Aiyar SE, Juang YL, Helmann JD, deHaseth PL. Mutations in sigma factor that affect the temperature dependence of transcription from a promoter, but not from a mismatch bubble in double-stranded DNA. *Biochemistry.* 1994 Sep 27;33(38):11501–6.
39. Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebright RH. Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism. *Science.* 2006 Nov 17;314(5802):1144–7.
40. Revyakin A, Liu C, Ebright RH, Strick TR. Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. *Science.* 2006 Nov 17;314(5802):1139–43.
41. Murakami KS, Darst SA. Bacterial RNA polymerases: the whole story. *Curr Opin Struct Biol.* 2003 Feb 1;13(1):31–9.
42. Murakami KS, Masuda S, Darst SA. Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 Å Resolution. *Science.* 2002 May 17;296(5571):1280–4.
43. Sen R, Nagai H, Hernandez VJ, Shimamoto N. Reduction in Abortive Transcription from the λ P R Promoter by Mutations in Region 3 of the σ 70 Subunit of Escherichia coli RNA Polymerase. *J Biol Chem.* 1998 Apr 17;273(16):9872–7.
44. Kulbachinskiy A, Mustaev A. Region 3.2 of the σ Subunit Contributes to the Binding of the 3'-Initiating Nucleotide in the RNA Polymerase Active Center and Facilitates Promoter Clearance during Initiation. *J Biol Chem.* 2006 Jul 7;281(27):18273–6.
45. Busby SJW. Transcription activation in bacteria: ancient and modern. *Microbiology.* 2019 Apr 1;165(4):386–95.
46. Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva MN, Borukhov S, et al. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature.* 2002 Jun;417(6890):712–9.
47. Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature.* 2007 Jul 12;448(7150):157–62.
48. Bar-Nahum G, Epshtein V, Ruckenstein AE, Rafikov R, Mustaev A, Nudler E. A ratchet mechanism of transcription elongation and its control. *Cell.* 2005 Jan 28;120(2):183–93.

49. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell*. 2006 Dec 1;127(5):941–54.
50. Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell*. 1999 Sep 17;98(6):811–24.
51. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science*. 2001 Jun 8;292(5523):1876–82.
52. Westover KD, Bushnell DA, Kornberg RD. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*. 2004 Nov 12;119(4):481–9.
53. Kettenberger H, Armache K-J, Cramer P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell*. 2004 Dec 22;16(6):955–65.
54. Korzheva N, Mustaev A, Kozlov M, Malhotra A, Nikiforov V, Goldfarb A, et al. A structural model of transcription elongation. *Science*. 2000 Jul 28;289(5479):619–25.
55. Blank A, Gallant JA, Burgess RR, Loeb LA. An RNA polymerase mutant with reduced accuracy of chain elongation. *Biochemistry*. 1986 Oct 7;25(20):5920–8.
56. Zaher HS, Green R. Fidelity at the molecular level: lessons from protein synthesis. *Cell*. 2009 Feb 20;136(4):746–62.
57. Naryshkina T, Kuznedelov K, Severinov K. The role of the largest RNA polymerase subunit lid element in preventing the formation of extended RNA-DNA hybrid. *J Mol Biol*. 2006 Aug 25;361(4):634–43.
58. Touloukhonov I, Landick R. The role of the lid element in transcription by *E. coli* RNA polymerase. *J Mol Biol*. 2006 Aug 25;361(4):644–58.
59. Telesnitsky A, Chamberlin MJ. Terminator-distal sequences determine the in vitro efficiency of the early terminators of bacteriophages T3 and T7. *Biochemistry*. 1989 Jun 13;28(12):5210–8.
60. Yarnell WS, Roberts JW. Mechanism of intrinsic transcription termination and antitermination. *Science*. 1999 Apr 23;284(5414):611–5.
61. Santangelo TJ, Roberts JW. Forward Translocation Is the Natural Pathway of RNA Release at an Intrinsic Terminator. *Mol Cell*. 2004 Apr;14(1):117–26.
62. Wösten MM. Eubacterial sigma-factors. *FEMS Microbiol Rev*. 1998 Sep;22(3):127–50.
63. Gross CA, Lonetto M, Losick R. In *Transcription Regulation, Volume 1*. 1992.
64. Dombroski AJ, Walter WA, Record MT, Siegele DA, Gross CA. Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. *Cell*. 1992 Aug 7;70(3):501–12.
65. Guo Y, Lew CM, Gralla JD. Promoter opening by ζ 54 and ζ 70 RNA polymerases: ζ factor-directed alterations in the mechanism and tightness of control. *Genes Dev*. 2000 Jan 9;14(17):2242–55.
66. Zuo Y, Steitz TA. Crystal structures of the *E. coli* transcription initiation

- complexes with a complete bubble. *Mol Cell*. 2015 May 7;58(3):534–40.
67. Feklistov A, Darst SA. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. *Cell*. 2011 Dec 9;147(6):1257–69.
68. Djordjevic M. Redefining Escherichia coli σ 70 Promoter Elements: –15 Motif as a Complement of the –10 Motif. *J Bacteriol*. 2011 Nov 15;193(22):6305–14.
69. Ayers DG, Auble DT, deHaseh PL. Promoter recognition by Escherichia coli RNA polymerase. Role of the spacer DNA in functional complex formation. *J Mol Biol*. 1989 Jun 20;207(4):749–56.
70. Stefano JE, Gralla JD. Spacer mutations in the lac ps promoter. *Proc Natl Acad Sci U S A*. 1982 Feb;79(4):1069–72.
71. Amouyal M, Buc H. Topological unwinding of strong and weak promoters by RNA polymerase. A comparison between the lac wild-type and the UV5 sites of Escherichia coli. *J Mol Biol*. 1987 Jun 20;195(4):795–808.
72. Travers AA. Structure and function of E. coli promoter DNA. *CRC Crit Rev Biochem*. 1987;22(3):181–219.
73. Bown JA, Barne KA, Minchin SD, Busby SJW. Extended –10 Promoters. In: Eckstein F, Lilley DMJ, editors. *Mechanisms of Transcription* [Internet]. Berlin, Heidelberg: Springer; 1997 [cited 2020 Mar 9]. p. 41–52. (Nucleic Acids and Molecular Biology). Available from: https://doi.org/10.1007/978-3-642-60691-5_4
74. Keilty S, Rosenberg M. Constitutive function of a positively regulated promoter reveals new sequences essential for activity. *J Biol Chem*. 1987 May 5;262(13):6389–95.
75. Sanderson A, Mitchell JE, Minchin SD, Busby SJW. Substitutions in the Escherichia coli RNA polymerase sigma70 factor that affect recognition of extended -10 elements at promoters. *FEBS Lett*. 2003 Jun 5;544(1–3):199–205.
76. Grimes E, Busby S, Minchin S. Different thermal energy requirement for open complex formation by Escherichia coli RNA polymerase at two related promoters. *Nucleic Acids Res*. 1991 Nov 25;19(22):6113–8.
77. Burns H, Minchin S. Thermal energy requirement for strand separation during transcription initiation: the effect of supercoiling and extended protein DNA contacts. *Nucleic Acids Res*. 1994 Sep 25;22(19):3840–5.
78. Hook-Barnard IG, Hinton DM. The promoter spacer influences transcription initiation via σ 70 region 1.1 of Escherichia coli RNA polymerase. *Proc Natl Acad Sci*. 2009 Jan 20;106(3):737–42.
79. Yuzenkova Y, Tadigotla VR, Severinov K, Zenkin N. A new basal promoter element recognized by RNA polymerase core enzyme. *EMBO J*. 2011 Jul 26;30(18):3766–75.
80. Ross W, Schneider DA, Paul BJ, Mertens A, Gourse RL. An intersubunit contact stimulating transcription initiation by E coli RNA polymerase: interaction of the alpha C-terminal domain and sigma region 4. *Genes Dev*. 2003 May 15;17(10):1293–307.
81. Chen H, Tang H, Ebright RH. Functional interaction between RNA polymerase alpha subunit C-terminal domain and sigma70 in UP-element- and activator-dependent transcription. *Mol Cell*. 2003 Jun;11(6):1621–33.

82. Estrem ST, Ross W, Gaal T, Chen ZW, Niu W, Ebright RH, et al. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev.* 1999 Aug 15;13(16):2134–47.
83. Meng W, Belyaeva T, Savery NJ, Busby SJW, Ross WE, Gaal T, et al. UP element-dependent transcription at the *Escherichia coli* rrnB P1 promoter: positional requirements and role of the RNA polymerase α subunit linker. *Nucleic Acids Res.* 2001 Oct 15;29(20):4166–78.
84. Naryshkin N, Revyakin A, Kim Y, Mekler V, Ebright RH. Structural organization of the RNA polymerase-promoter open complex. *Cell.* 2000 Jun 9;101(6):601–11.
85. Newlands JT, Josaitis CA, Ross W, Gourse RL. Both fis-dependent and factor-independent upstream activation of the rrnB P1 promoter are face of the helix dependent. *Nucleic Acids Res.* 1992 Feb 25;20(4):719–26.
86. Estrem ST, Gaal T, Ross W, Gourse RL. Identification of an UP element consensus sequence for bacterial promoters. *Proc Natl Acad Sci U S A.* 1998 Aug 18;95(17):9761–6.
87. Rao L, Ross W, Appleman JA, Gaal T, Leirimo S, Schlax PJ, et al. Factor independent activation of rrnB P1. An “extended” promoter with an upstream element that dramatically increases promoter strength. *J Mol Biol.* 1994 Feb 4;235(5):1421–35.
88. Ross W, Gosink KK, Salomon J, Igarashi K, Zou C, Ishihama A, et al. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science.* 1993 Nov 26;262(5138):1407–13.
89. Ross W, Aiyar SE, Salomon J, Gourse RL. *Escherichia coli* Promoters with UP Elements of Different Strengths: Modular Structure of Bacterial Promoters. *J Bacteriol.* 1998 Oct;180(20):5375–83.
90. Crothers DM, Haran TE, Nadeau JG. Intrinsically bent DNA. *J Biol Chem.* 1990 May 5;265(13):7093–6.
91. Pérez-Martín J, Rojo F, de Lorenzo V. Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. *Microbiol Rev.* 1994 Jun;58(2):268–90.
92. Shimada T, Yamazaki Y, Tanaka K, Ishihama A. The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of *Escherichia coli*. *PLoS One.* 2014;9(3):e90447.
93. Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, Gourse RL. rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell.* 2006 Jun 16;125(6):1069–82.
94. Liu J, Turnbough CL. Effects of transcriptional start site sequence and position on nucleotide-sensitive selection of alternative start sites at the pyrC promoter in *Escherichia coli*. *J Bacteriol.* 1994 May;176(10):2938–45.
95. Krásný L, Tiserová H, Jonák J, Rejman D, Sanderová H. The identity of the transcription +1 position is crucial for changes in gene expression in response to amino acid starvation in *Bacillus subtilis*. *Mol Microbiol.* 2008 Jul;69(1):42–54.

96. Selby CP, Sancar A. Molecular mechanism of transcription-repair coupling. *Science*. 1993 Apr 2;260(5104):53–8.
97. Lee DN, Phung L, Stewart J, Landick R. Transcription pausing by *Escherichia coli* RNA polymerase is modulated by downstream DNA sequences. *J Biol Chem*. 1990 Sep 5;265(25):15145–53.
98. Touloukhonov I, Zhang J, Palangat M, Landick R. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol Cell*. 2007 Aug 3;27(3):406–19.
99. Gusarov I, Nudler E. The mechanism of intrinsic transcription termination. *Mol Cell*. 1999 Apr;3(4):495–504.
100. Komissarova N, Kashlev M. Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc Natl Acad Sci U S A*. 1997 Mar 4;94(5):1755–60.
101. Komissarova N, Becker J, Solter S, Kireeva M, Kashlev M. Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Mol Cell*. 2002 Nov;10(5):1151–62.
102. Hein PP, Kolb KE, Windgassen T, Bellecourt MJ, Darst SA, Mooney RA, et al. RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat Struct Mol Biol*. 2014 Sep;21(9):794–802.
103. Sekine S, Murayama Y, Svetlov V, Nudler E, Yokoyama S. The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Mol Cell*. 2015 Feb 5;57(3):408–21.
104. Yakhnin AV, Yakhnin H, Babitzke P. Function of the *Bacillus subtilis* transcription elongation factor NusG in hairpin-dependent RNA polymerase pausing in the *trp* leader. *Proc Natl Acad Sci U S A*. 2008 Oct 21;105(42):16131–6.
105. Yakhnin AV, Babitzke P. Mechanism of NusG-stimulated pausing, hairpin-dependent pause site selection and intrinsic termination at overlapping pause and termination sites in the *Bacillus subtilis* *trp* leader. *Mol Microbiol*. 2010 May;76(3):690–705.
106. Czyz A, Mooney RA, Iaconi A, Landick R. Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *mBio*. 2014 Apr 8;5(2):e00931.
107. Herbert KM, Zhou J, Mooney RA, Porta AL, Landick R, Block SM. *E. coli* NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J Mol Biol*. 2010 May 28;399(1):17–30.
108. Svetlov V, Belogurov GA, Shabrova E, Vassylyev DG, Artsimovitch I. Allosteric control of the RNA polymerase by the elongation factor RfaH. *Nucleic Acids Res*. 2007;35(17):5694–705.
109. Sevostyanova A, Belogurov GA, Mooney RA, Landick R, Artsimovitch I. The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol Cell*. 2011 Jul 22;43(2):253–62.
110. Kolb KE, Hein PP, Landick R. Antisense oligonucleotide-stimulated transcriptional pausing reveals RNA exit channel specificity of RNA polymerase and

- mechanistic contributions of NusA and RfaH. *J Biol Chem.* 2014 Jan 10;289(2):1151–63.
111. Schmidt MC, Chamberlin MJ. nusA protein of *Escherichia coli* is an efficient transcription termination factor for certain terminator sites. - PubMed - NCBI [Internet]. 1987 [cited 2020 Mar 12]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/2821282>
112. Ha KS, Touloukianov I, Vassylyev DG, Landick R. The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. *J Mol Biol.* 2010 Sep 3;401(5):708–25.
113. Lowery-Goldhammer C, Richardson JP. An RNA-dependent nucleoside triphosphate phosphohydrolase (ATPase) associated with rho termination factor. *Proc Natl Acad Sci U S A.* 1974 May;71(5):2003–7.
114. Zalatan P, Platt T. Effects of decreased cytosine content on rho interaction with the rho-dependent terminator trp t' in *Escherichia coli*. - PubMed - NCBI [Internet]. 1992 [cited 2020 Mar 12]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/1388163>
115. Rutherford ST, Villers CL, Lee JH, Ross W, Gourse RL. Allosteric control of *Escherichia coli* rRNA promoter complexes by DksA. - PubMed - NCBI [Internet]. 2009 [cited 2020 Mar 12]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/19171784>
116. Pardee AB, Jacob F, Monod J. The genetic control and cytoplasmic expression of "Inducibility" in the synthesis of β -galactosidase by *E. coli*. *J Mol Biol.* 1959 Jun 1;1(2):165–78.
117. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol.* 1961 Jun 1;3(3):318–56.
118. Gilbert W, Müller-Hill B. Isolation of the lac repressor. *Proc Natl Acad Sci U S A.* 1966 Dec;56(6):1891–8.
119. Gilbert W, Müller-Hill B. The lac operator is DNA. *Proc Natl Acad Sci U S A.* 1967 Dec;58(6):2415–21.
120. Keren L, Zackay O, Lotan-Pompan M, Barenholz U, Dekel E, Sasson V, et al. Promoters maintain their relative activity levels under different growth conditions. *Mol Syst Biol.* 2013 Oct 29;9:701.
121. Martínez-Antonio A, Collado-Vides J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol.* 2003 Oct;6(5):482–9.
122. Pérez-Rueda E, Collado-Vides J. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* 2000 Apr 15;28(8):1838–47.
123. Madan Babu M, Teichmann SA. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.* 2003 Feb 15;31(4):1234–44.
124. Brown NL, Stoyanov JV, Kidd SP, Hobman JL. The MerR family of transcriptional regulators. *FEMS Microbiol Rev.* 2003 Jun;27(2–3):145–63.
125. Heldwein EE, Brennan RG. Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature.* 2001 Jan 18;409(6818):378–82.
126. Newberry KJ, Brennan RG. The structural mechanism for transcription

- activation by MerR family member multidrug transporter activation, N terminus. *J Biol Chem.* 2004 May 7;279(19):20356–62.
127. Lee DJ, Minchin SD, Busby SJW. Activating Transcription in Bacteria. *Annu Rev Microbiol.* 2012;66(1):125–52.
128. Dorman CJ. DNA supercoiling and bacterial gene expression. *Sci Prog.* 2006;89(Pt 3-4):151–66.
129. Auner H, Buckle M, Deufel A, Kutateladze T, Lazarus L, Mavathur R, et al. Mechanism of transcriptional activation by FIS: role of core promoter structure and DNA topology. *J Mol Biol.* 2003 Aug 8;331(2):331–44.
130. Opel ML, Aeling KA, Holmes WM, Johnson RC, Benham CJ, Hatfield GW. Activation of transcription initiation from a stable RNA promoter by a Fis protein-mediated DNA structural transmission mechanism. *Mol Microbiol.* 2004 Jul;53(2):665–74.
131. Sheridan SD, Opel ML, Hatfield GW. Activation and repression of transcription initiation by a distant DNA structural transition. - PubMed - NCBI [Internet]. 2001 [cited 2020 Mar 12]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/11359573>
132. Travers A, Muskhelishvili G. DNA microloops and microdomains: a general mechanism for transcription activation by torsional transmission. *J Mol Biol.* 1998 Jun 26;279(5):1027–43.
133. Busby S, Ebright RH. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell.* 1994 Dec 2;79(5):743–6.
134. Hochschild A, Dove SL. Protein–Protein Contacts that Activate and Repress Prokaryotic Transcription. *Cell.* 1998 Mar 6;92(5):597–600.
135. Pérez-Rueda E, Gralla JD, Collado-Vides J. Genomic position analyses and the transcription machinery. *J Mol Biol.* 1998 Jan 16;275(2):165–70.
136. Browning DF, Busby SJW. The regulation of bacterial transcription initiation. *Nat Rev Microbiol.* 2004 Jan;2(1):57–65.
137. Benoff B, Yang H, Lawson CL, Parkinson G, Liu J, Blatter E, et al. Structural basis of transcription activation: the CAP- α CTD-DNA complex. *Science.* 2002 Aug 30;297(5586):1562–6.
138. Jain D, Nickels BE, Sun L, Hochschild A, Darst SA. Structure of a ternary transcription activation complex. *Mol Cell.* 2004 Jan 16;13(1):45–53.
139. Müller-Hill B. Some repressors of bacterial transcription. *Curr Opin Microbiol.* 1998 Apr;1(2):145–51.
140. Neidhardt FC. *Escherichia coli* and *Salmonella* : cellular and molecular biology [Internet]. 1996 [cited 2020 Mar 13]. Available from: /paper/Escherichia-coli-and-Salmonella-%3A-cellular-and-Neidhardt/5720475eea38e9a3bf493a7cfdbeddadb248ac7
141. Shin M, Kang S, Hyun S-J, Fujita N, Ishihama A, Valentin-Hansen P, et al. Repression of deoP2 in *Escherichia coli* by CytR: conversion of a transcription activator into a repressor. *EMBO J.* 2001 Oct 1;20(19):5392–9.
142. Busby S, Ebright RH. Transcription activation by catabolite activator protein (CAP). *J Mol Biol.* 1999 Oct 22;293(2):199–213.

143. Harman JG. Allosteric regulation of the cAMP receptor protein. *Biochim Biophys Acta*. 2001 May 5;1547(1):1–17.
144. Kolb A, Busby S, Buc H, Garges S, Adhya S. Transcriptional Regulation by cAMP and Its Receptor Protein. *Annu Rev Biochem*. 1993 Jun;62(1):749–97.
145. Passner JM, Steitz TA. The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer. *Proc Natl Acad Sci U S A*. 1997 Apr 1;94(7):2843–7.
146. Cameron ADS, Redfield RJ. Non-canonical CRP sites control competence regulons in *Escherichia coli* and many other gamma-proteobacteria. *Nucleic Acids Res*. 2006;34(20):6001–14.
147. Merrick MJ. In a class of its own--the RNA polymerase sigma factor sigma 54 (sigma N). *Mol Microbiol*. 1993 Dec;10(5):903–9.
148. Ghosh T, Bose D, Zhang X. Mechanisms for activating bacterial RNA polymerase. *FEMS Microbiol Rev*. 2010 Sep;34(5):611–27.
149. Chatterji D, Kumar Ojha A. Revisiting the stringent response, ppGpp and starvation signaling. *Curr Opin Microbiol*. 2001 Apr 1;4(2):160–5.
150. Barker MM, Gaal T, Josaitis CA, Gourse RL. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. *J Mol Biol*. 2001 Jan 26;305(4):673–88.
151. Barker MM, Gaal T, Gourse RL. Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. *J Mol Biol*. 2001 Jan 26;305(4):689–702.
152. Müller-Hill B. *The lac Operon: a short history of a genetic paradigm* /. Berlin ; Walter de Gruyter,; 1996.
153. Mian I. Comparative sequence analysis of ribonucleases HII, III, II PH and D. *Nucleic Acids Res*. 1997 Aug 15;25(16):3187–95.
154. Court DL, Gan J, Liang Y-H, Shaw GX, Tropea JE, Costantino N, et al. RNase III: Genetics and Function; Structure and Mechanism. *Annu Rev Genet*. 2013 Nov 23;47(1):405–31.
155. Pace NR, Smith D. Ribonuclease P: function and variation. *J Biol Chem*. 1990 Mar 5;265(7):3587–90.
156. Alifano P, Rivellini F, Piscitelli C, Arraiano CM, Bruni CB, Carlomagno MS. Ribonuclease E provides substrates for ribonuclease P-dependent processing of a polycistronic mRNA. *Genes Dev*. 1994 Dec 15;8(24):3021–31.
157. Misra TK, Apirion D. RNase E, an RNA processing enzyme from *Escherichia coli*. *J Biol Chem*. 1979 Nov 10;254(21):11154–9.
158. Cohen SN, McDowall KJ. RNase E: still a wonderfully mysterious enzyme. *Mol Microbiol*. 1997 Mar;23(6):1099–106.
159. McDowall KJ, Lin-Chao S, Cohen SN. A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J Biol Chem*. 1994 Apr 8;269(14):10790–6.
160. Bouvet P, Belasco JG. Control of RNase E-mediated RNA degradation by 5'-terminal base pairing in *E. coli*. *Nature*. 1992 Dec 3;360(6403):488–91.

161. Mackie GA, Genereaux JL. The role of RNA structure in determining RNase E-dependent cleavage sites in the mRNA for ribosomal protein S20 in vitro. *J Mol Biol.* 1993 Dec 20;234(4):998–1012.
162. Mackie GA. Ribonuclease E is a 5'-end-dependent endonuclease. *Nature.* 1998 Oct 15;395(6703):720–3.
163. Nicholson AW. Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiol Rev.* 1999 Jun;23(3):371–90.
164. Hajnsdorf E, Steier O, Coscoy L, Teyssset L, Régnier P. Roles of RNase E, RNase II and PNPase in the degradation of the rpsO transcripts of *Escherichia coli*: stabilizing function of RNase II and evidence for efficient degradation in an *ams pnp rnb* mutant. *EMBO J.* 1994 Jul 15;13(14):3368–77.
165. Coburn GA, Mackie GA. Overexpression, purification, and properties of *Escherichia coli* ribonuclease II. *J Biol Chem.* 1996 Jan 12;271(2):1048–53.
166. Py B, Higgins CF, Krisch HM, Carpousis AJ. A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature.* 1996 May 9;381(6578):169–72.
167. Régnier P, Arraiano CM. Degradation of mRNA in bacteria: emergence of ubiquitous features. *BioEssays News Rev Mol Cell Dev Biol.* 2000 Mar;22(3):235–44.
168. Mohanty BK, Kushner SR. Analysis of the function of *Escherichia coli* poly(A) polymerase I in RNA metabolism. *Mol Microbiol.* 1999 Dec;34(5):1094–108.
169. Joyce SA, Dreyfus M. In the absence of translation, RNase E can bypass 5' mRNA stabilizers in *Escherichia coli*. *J Mol Biol.* 1998 Sep 18;282(2):241–54.
170. Kellenberger E, Ryter A, Sechaud J. Electron microscope study of DNA-containing plasmids. II. Vegetative and mature phage DNA as compared with normal bacterial nucleoids in different physiological states. *J Biophys Biochem Cytol.* 1958 Nov 25;4(6):671–8.
171. Holmes VF, Cozzarelli NR. Closing the ring: links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proc Natl Acad Sci U S A.* 2000 Feb 15;97(4):1322–4.
172. Willenbrock H, Ussery DW. Chromatin architecture and gene expression in *Escherichia coli*. *Genome Biol.* 2004 Dec 1;5(12):252.
173. Marr C, Geertz M, Hütt M-T, Muskhelishvili G. Dissecting the logical types of network control in gene expression profiles. *BMC Syst Biol.* 2008 Feb 19;2:18.
174. Valens M, Penaud S, Rossignol M, Cornet F, Boccard F. Macrodome organization of the *Escherichia coli* chromosome. *EMBO J.* 2004 Oct 27;23(21):4330–41.
175. Lioy VS, Cournac A, Marbouty M, Duigou S, Mozziconacci J, Espéli O, et al. Multiscale Structuring of the *E. coli* Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell.* 2018 Feb;172(4):771–783.e18.
176. Dorman CJ. Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat Rev Microbiol.* 2013;11(5):349–55.
177. Badrinarayanan A, Lesterlin C, Reyes-Lamothe R, Sherratt D. The *Escherichia coli* SMC complex, MukBEF, shapes nucleoid organization independently of DNA replication. *J Bacteriol.* 2012 Sep;194(17):4669–76.

178. Espéli O, Borne R, Dupaigne P, Thiel A, Gigant E, Mercier R, et al. A MatP-divisome interaction coordinates chromosome segregation with cell division in *E. coli*. *EMBO J*. 2012 May 11;31(14):3198–211.
179. Lu M, Campbell JL, Boye E, Kleckner N. SeqA: a negative modulator of replication initiation in *E. coli*. *Cell*. 1994 May 6;77(3):413–26.
180. von Freiesleben U, Rasmussen KV, Schaechter M. SeqA limits DnaA activity in replication from *oriC* in *Escherichia coli*. *Mol Microbiol*. 1994 Nov;14(4):763–72.
181. Sánchez-Romero MA, Busby SJW, Dyer NP, Ott S, Millard AD, Grainger DC. Dynamic distribution of *seqA* protein across the chromosome of *Escherichia coli* K-12. *mBio*. 2010 May 18;1(1).
182. Dame RT, Kalmykova OJ, Grainger DC. Chromosomal macrodomains and associated proteins: implications for DNA organization and replication in gram negative bacteria. *PLoS Genet*. 2011 Jun;7(6):e1002123.
183. Tonthat NK, Arold ST, Pickering BF, Van Dyke MW, Liang S, Lu Y, et al. Molecular mechanism by which the nucleoid occlusion factor, *SImA*, keeps cytokinesis in check. *EMBO J*. 2011 Jan 5;30(1):154–64.
184. Worcel A, Burgi E. On the structure of the folded chromosome of *Escherichia coli*. *J Mol Biol*. 1972 Nov 14;71(2):127–47.
185. Sinden RR, Pettijohn DE. Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling. *Proc Natl Acad Sci U S A*. 1981 Jan;78(1):224–8.
186. Postow L, Hardy CD, Arsuaga J, Cozzarelli NR. Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev*. 2004 Jul 15;18(14):1766–79.
187. Deng S, Stein RA, Higgins NP. Organization of supercoil domains and their reorganization by transcription. *Mol Microbiol*. 2005 Sep;57(6):1511–21.
188. Hadizadeh Yazdi N, Guet CC, Johnson RC, Marko JF. Variation of the folding and dynamics of the *Escherichia coli* chromosome with growth conditions. *Mol Microbiol*. 2012 Dec;86(6):1318–33.
189. Wang W, Li G-W, Chen C, Xie XS, Zhuang X. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*. 2011 Sep 9;333(6048):1445–9.
190. Kavenoff R, Bowen BC. Electron microscopy of membrane-free folded chromosomes from *Escherichia coli*. *Chromosoma*. 1976 Dec 16;59(2):89–101.
191. Higgins NP, Yang X, Fu Q, Roth JR. Surveying a supercoil domain by using the gamma delta resolution system in *Salmonella typhimurium*. *J Bacteriol*. 1996 May;178(10):2825–35.
192. Funnell BE, Baker TA, Kornberg A. Complete enzymatic replication of plasmids containing the origin of the *Escherichia coli* chromosome. *J Biol Chem*. 1986 Apr 25;261(12):5616–24.
193. Pruss GJ, Drlica K. DNA supercoiling and prokaryotic transcription. *Cell*. 1989 Feb 24;56(4):521–3.
194. Drlica K, Rouviere-Yaniv J. Histone-like proteins of bacteria. *Microbiol Rev*. 1987 Sep;51(3):301–19.
195. Dorman CJ, Deighan P. Regulation of gene expression by histone-like proteins

- in bacteria. *Curr Opin Genet Dev.* 2003 Apr;13(2):179–84.
196. Azam TA, Hiraga S, Ishihama A. Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells Devoted Mol Cell Mech.* 2000 Aug;5(8):613–26.
197. Browning DF, Grainger DC, Busby SJ. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr Opin Microbiol.* 2010 Dec 1;13(6):773–80.
198. Azam TA, Ishihama A. Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J Biol Chem.* 1999 Nov 12;274(46):33105–13.
199. Robison K, McGuire AM, Church GM. A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J Mol Biol.* 1998 Nov 27;284(2):241–54.
200. de los Rios S, Perona JJ. Structure of the *Escherichia coli* leucine-responsive regulatory protein Lrp reveals a novel octameric assembly. *J Mol Biol.* 2007 Mar 9;366(5):1589–602.
201. Thaw P, Sedelnikova SE, Muranova T, Wiese S, Ayora S, Alonso JC, et al. Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. *Nucleic Acids Res.* 2006;34(5):1439–49.
202. Pul U, Wurm R, Wagner R. The role of LRP and H-NS in transcription regulation: involvement of synergism, allostery and macromolecular crowding. *J Mol Biol.* 2007 Feb 23;366(3):900–15.
203. Tanaka H, Goshima N, Kohno K, Kano Y, Imamoto F. Properties of DNA-binding of HU heterotypic and homotypic dimers from *Escherichia coli*. *J Biochem (Tokyo).* 1993 May;113(5):568–72.
204. Goshima N, Kano Y, Tanaka H, Kohno K, Iwaki T, Imamoto F. IHF suppresses the inhibitory effect of H-NS on HU function in the *hin* inversion system. *Gene.* 1994 Apr 8;141(1):17–23.
205. Goodrich JA, Schwartz ML, McClure WR. Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). *Nucleic Acids Res.* 1990 Sep 11;18(17):4993–5000.
206. Hardy CD, Cozzarelli NR. A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure. *Mol Microbiol.* 2005 Sep;57(6):1636–52.
207. Noom MC, Navarre WW, Oshima T, Wuite GJL, Dame RT. H-NS promotes looped domain formation in the bacterial chromosome. *Curr Biol CB.* 2007 Nov 6;17(21):R913-914.
208. Dame RT, Wyman C, Goosen N. H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res.* 2000 Sep 15;28(18):3504–10.
209. Dame RT, Noom MC, Wuite GJL. Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature.* 2006 Nov 16;444(7117):387–90.

210. Dorman CJ. H-NS, the genome sentinel. *Nat Rev Microbiol.* 2007;5(2):157–61.
211. Bouffartigues E, Buckle M, Badaut C, Travers A, Rimsky S. H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nat Struct Mol Biol.* 2007 May;14(5):441–8.
212. Lang B, Blot N, Bouffartigues E, Buckle M, Geertz M, Gualerzi CO, et al. High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic Acids Res.* 2007;35(18):6330–7.
213. Lucchini S, Rowley G, Goldberg MD, Hurd D, Harrison M, Hinton JCD. H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog.* 2006 Aug;2(8):e81.
214. Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, Libby SJ, et al. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science.* 2006 Jul 14;313(5784):236–8.
215. Schneider R, Lurz R, Lüder G, Tolksdorf C, Travers A, Muskhelishvili G. An architectural role of the *Escherichia coli* chromatin protein FIS in organising DNA. *Nucleic Acids Res.* 2001 Dec 15;29(24):5107–14.
216. Grainger DC, Hurd D, Goldberg MD, Busby SJW. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.* 2006;34(16):4642–52.
217. Tendeng C, Soutourina OA, Danchin A, Bertin PN. MvaT proteins in *Pseudomonas* spp.: a novel class of H-NS-like proteins. *Microbiol Read Engl.* 2003 Nov;149(Pt 11):3047–50.
218. Dame RT, Luijsterburg MS, Krin E, Bertin PN, Wagner R, Wuite GJL. DNA bridging: a property shared among H-NS-like proteins. *J Bacteriol.* 2005 Mar;187(5):1845–8.
219. Chen JM, Ren H, Shaw JE, Wang YJ, Li M, Leung AS, et al. Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. *Nucleic Acids Res.* 2008 Apr;36(7):2123–35.
220. Gordon BRG, Imperial R, Wang L, Navarre WW, Liu J. Lsr2 of *Mycobacterium* represents a novel class of H-NS-like proteins. *J Bacteriol.* 2008 Nov;190(21):7052–9.
221. Zimmerman SB. Cooperative transitions of isolated *Escherichia coli* nucleoids: implications for the nucleoid as a cellular phase. *J Struct Biol.* 2006 Feb;153(2):160–75.
222. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol.* 1961 Jun;3(3):318–56.
223. Oehler S, Eismann ER, Krämer H, Müller-Hill B. The three operators of the lac operon cooperate in repression. *EMBO J.* 1990 Apr;9(4):973–9.
224. Dunn TM, Hahn S, Ogden S, Schleif RF. An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc Natl Acad Sci U S A.* 1984 Aug;81(16):5017–20.
225. Ptashne M. A genetic switch: phage lambda revisited. 3rd ed. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press; 2004. 154 p.
226. Meyer BJ, Ptashne M. Gene regulation at the right operator (OR) of

- bacteriophage lambda. III. lambda repressor directly activates gene transcription. *J Mol Biol.* 1980 May 15;139(2):195–205.
227. Révet B, von Wilcken-Bergmann B, Bessert H, Barker A, Müller-Hill B. Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. *Curr Biol CB.* 1999 Feb 11;9(3):151–4.
228. Dodd IB, Shearwin KE, Perkins AJ, Burr T, Hochschild A, Egan JB. Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev.* 2004 Feb 1;18(3):344–54.
229. Ding Y, Manzo C, Fulcrand G, Leng F, Dunlap D, Finzi L. DNA supercoiling: a regulatory signal for the λ repressor. *Proc Natl Acad Sci U S A.* 2014 Oct 28;111(43):15402–7.
230. Morelli MJ, ten Wolde PR, Allen RJ. DNA looping provides stability and robustness to the bacteriophage switch. *Proc Natl Acad Sci.* 2009 May 19;106(20):8101–6.
231. Stein RA, Deng S, Higgins NP. Measuring chromosome dynamics on different time scales using resolvases with varying half-lives. *Mol Microbiol.* 2005 May;56(4):1049–61.
232. Hillebrand A, Wurm R, Menzel A, Wagner R. The seven *E. coli* ribosomal RNA operon upstream regulatory regions differ in structure and transcription factor binding efficiencies. *Biol Chem.* 2005 Jun;386(6):523–34.
233. Bliska JB, Cozzarelli NR. Use of site-specific recombination as a probe of DNA structure and metabolism in vivo. *J Mol Biol.* 1987 Mar 20;194(2):205–18.
234. Pettijohn DE, Pfenninger O. Supercoils in prokaryotic DNA restrained in vivo. *Proc Natl Acad Sci U S A.* 1980 Mar;77(3):1331–5.
235. Vinograd J, Lebowitz J, Radloff R, Watson R, Laipis P. The twisted circular form of polyoma viral DNA. *Proc Natl Acad Sci.* 1965 May 1;53(5):1104–11.
236. Kornberg A, Baker TA. *DNA Replication.* 2d ed. W. H. Freeman and Company; 1992.
237. Surenroulement de l'ADN. In: Wikipédia [Internet]. 2016 [cited 2020 Jul 13]. Available from: https://fr.wikipedia.org/w/index.php?title=Surenroulement_de_l%27ADN&oldid=129981458
238. Bauer W, Vinograd J. The interaction of closed circular DNA with intercalative dyes. *J Mol Biol.* 1968 Apr;33(1):141–71.
239. Upholt WB, Gray HB, Vinograd J. Sedimentation velocity behavior of closed circular SV40 DNA as a function of superhelix density, ionic strength, counterion and temperature. *J Mol Biol.* 1971 Nov 28;62(1):21–38.
240. Képès F, Vaillant C. Transcription-Based Solenoidal Model of Chromosomes. *Complexus.* 2003;1(4):171–80.
241. Képès F. Periodic transcriptional organization of the *E. coli* genome. *J Mol Biol.* 2004 Jul 23;340(5):957–64.
242. Carpentier A-S, Torrèsani B, Grossmann A, Hénaut A. [No title found]. *BMC*

- Genomics. 2005;6(1):84.
243. Travers A, Muskhelishvili G. A common topology for bacterial and eukaryotic transcription initiation? *EMBO Rep.* 2007 Feb;8(2):147–51.
244. Boles TC, White JH, Cozzarelli NR. Structure of plectonemically supercoiled DNA. *J Mol Biol.* 1990 Jun 20;213(4):931–51.
245. Adrian M, ten Heggeler-Bordier B, Wahli W, Stasiak AZ, Stasiak A, Dubochet J. Direct visualization of supercoiled DNA molecules in solution. *EMBO J.* 1990 Dec;9(13):4551–4.
246. Huang J, Schlick T, Vologodskii A. Dynamics of site juxtaposition in supercoiled DNA. *Proc Natl Acad Sci.* 2001 Jan 30;98(3):968–73.
247. Le TBK, Imakaev MV, Mirny LA, Laub MT. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science.* 2013 Nov 8;342(6159):731–4.
248. Drlica K. Biology of bacterial deoxyribonucleic acid topoisomerases. *Microbiol Rev.* 1984 Dec;48(4):273–89.
249. Drlica K. Control of bacterial DNA supercoiling. *Mol Microbiol.* 1992 Feb;6(4):425–33.
250. Wang JC. DNA topoisomerases. *Annu Rev Biochem.* 1985;54:665–97.
251. Wagner R. Transcription regulation in prokaryotes. Oxford ; New York: Oxford University Press; 2000. 365 p.
252. Liu LF, Wang JC. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A.* 1987 Oct;84(20):7024–7.
253. Lilley DM, Higgins CF. Local DNA topology and gene expression: the case of the leu-500 promoter. *Mol Microbiol.* 1991 Apr;5(4):779–83.
254. Ma J, Bai L, Wang MD. Transcription under torsion. *Science.* 2013 Jun 28;340(6140):1580–3.
255. Ma J, Wang M. Interplay between DNA supercoiling and transcription elongation. *Transcription.* 2014;5(3):e28636.
256. Zhi X, Dages S, Dages K, Liu Y, Hua Z-C, Makemson J, et al. Transient and dynamic DNA supercoiling potentially stimulates the leu-500 promoter in *Escherichia coli*. *J Biol Chem.* 2017 01;292(35):14566–75.
257. Cook PR. How mobile are active RNA polymerases? *J Cell Sci.* 1990 Jun;96 (Pt 2):189–92.
258. Cook PR, Gove F. Transcription by an immobilized RNA polymerase from bacteriophage T7 and the topology of transcription. *Nucleic Acids Res.* 1992 Jul 25;20(14):3591–8.
259. Papantonis A, Cook PR. Fixing the model for transcription. *Transcription.* 2011;2(1):41–4.
260. Dorman CJ. DNA supercoiling and transcription in bacteria: a two-way street. *BMC Mol Cell Biol.* 2019 18;20(1):26.
261. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. *Cell.* 2005 Dec 16;123(6):1025–36.
262. Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J, et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science.*

- 2010 Jul 30;329(5991):533–8.
263. Zong C, So L, Sepúlveda LA, Skinner SO, Golding I. Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene. *Mol Syst Biol*. 2010 Nov 30;6:440.
264. So L-H, Ghosh A, Zong C, Sepúlveda LA, Segev R, Golding I. General properties of transcriptional time series in *Escherichia coli*. *Nat Genet*. 2011 Jun;43(6):554–60.
265. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 2011 Apr 22;332(6028):472–4.
266. Levens D, Larson DR. A new twist on transcriptional bursting. *Cell*. 2014 Jul 17;158(2):241–2.
267. Chong S, Chen C, Ge H, Xie XS. Mechanism of transcriptional bursting in bacteria. *Cell*. 2014 Jul 17;158(2):314–26.
268. Sanzey B. Modulation of gene expression by drugs affecting deoxyribonucleic acid gyrase. *J Bacteriol*. 1979 Apr;138(1):40–7.
269. Menzel R, Gellert M. Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. *Cell*. 1983 Aug;34(1):105–13.
270. Brahms JG, Dargouge O, Brahms S, Ohara Y, Vagner V. Activation and inhibition of transcription by supercoiling. *J Mol Biol*. 1985 Feb;181(4):455–65.
271. Borowiec JA, Gralla JD. Supercoiling response of the *lac ps* promoter in vitro. *J Mol Biol*. 1985 Aug 20;184(4):587–98.
272. Borowiec JA, Gralla JD. All three elements of the *lac ps* promoter mediate its transcriptional response to DNA supercoiling. *J Mol Biol*. 1987 May 5;195(1):89–97.
273. Rudd KE, Menzel R. *his* operons of *Escherichia coli* and *Salmonella typhimurium* are regulated by DNA supercoiling. *Proc Natl Acad Sci U S A*. 1987 Jan;84(2):517–21.
274. Tabuchi H, Handa H, Hirose S. Underwinding of DNA on binding of yeast TFIID to the TATA element. *Biochem Biophys Res Commun*. 1993 May 14;192(3):1432–8.
275. Wu HY, Tan J, Fang M. Long-range interaction between two promoters: activation of the *leu-500* promoter by a distant upstream promoter. *Cell*. 1995 Aug 11;82(3):445–51.
276. Képès F, Jester BC, Lepage T, Rafiei N, Rosu B, Junier I. The layout of a bacterial genome. *FEBS Lett*. 2012 Jul 16;586(15):2043–8.
277. Junier I, Rivoire O. Conserved Units of Co-Expression in Bacterial Genomes: An Evolutionary Insight into Transcriptional Regulation. *PLoS ONE* [Internet]. 2016 May 19 [cited 2020 Jul 24];11(5). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4873041/>
278. Dorman CJ. Genetics of bacterial virulence. Oxford ; Boston: Blackwell Scientific Publications; 1994. 369 p.
279. Depew RE, Liu LF, Wang JC. Interaction between DNA and *Escherichia coli* protein omega. Formation of a complex between single-stranded DNA and omega protein. *J Biol Chem*. 1978 Jan 25;253(2):511–8.
280. Liu LF, Wang JC. Interaction between DNA and *Escherichia coli* DNA

- topoisomerase I. Formation of complexes between the protein and superhelical and nonsuperhelical duplex DNAs. *J Biol Chem.* 1979 Nov 10;254(21):11082–8.
281. Tse YC, Kirkegaard K, Wang JC. Covalent bonds between protein and DNA. Formation of phosphotyrosine linkage between certain DNA topoisomerases and DNA. *J Biol Chem.* 1980 Jun 25;255(12):5560–5.
282. Champoux JJ. DNA is linked to the rat liver DNA nicking-closing enzyme by a phosphodiester bond to tyrosine. *J Biol Chem.* 1981 May 25;256(10):4805–9.
283. Brown PO, Cozzarelli NR. A sign inversion mechanism for enzymatic supercoiling of DNA. *Science.* 1979 Nov 30;206(4422):1081–3.
284. Morrison A, Cozzarelli NR. Site-specific cleavage of DNA by *E. coli* DNA gyrase. *Cell.* 1979 May;17(1):175–84.
285. Liu LF, Rowe TC, Yang L, Tewey KM, Chen GL. Cleavage of DNA by mammalian DNA topoisomerase II. *J Biol Chem.* 1983 Dec 25;258(24):15365–70.
286. Sander M, Hsieh T. Double strand DNA cleavage by type II DNA topoisomerase from *Drosophila melanogaster*. *J Biol Chem.* 1983 Jul 10;258(13):8421–8.
287. Brown PO, Cozzarelli NR. Catenation and knotting of duplex DNA by type 1 topoisomerases: a mechanistic parallel with type 2 topoisomerases. *Proc Natl Acad Sci U S A.* 1981 Feb;78(2):843–7.
288. Tse Y, Wang JC. *E. coli* and *M. luteus* DNA topoisomerase I can catalyze catenation of decatenation of double-stranded DNA rings. *Cell.* 1980 Nov;22(1 Pt 1):269–76.
289. DiGate RJ, Marians KJ. Identification of a potent decatenating enzyme from *Escherichia coli*. *J Biol Chem.* 1988 Sep 15;263(26):13366–73.
290. Hanai R, Caron PR, Wang JC. Human TOP3: a single-copy gene encoding DNA topoisomerase III. *Proc Natl Acad Sci U S A.* 1996 Apr 16;93(8):3653–7.
291. Wilson TM, Chen AD, Hsieh T. Cloning and Characterization of *Drosophila* Topoisomerase III β : RELAXATION OF HYPERNEGATIVELY SUPERCOILED DNA. *J Biol Chem.* 2000 Jan 21;275(3):1533–40.
292. Wang JC. DNA topoisomerases. *Annu Rev Biochem.* 1996;65:635–92.
293. Hiasa H, DiGate RJ, Marians KJ. Decatenating activity of *Escherichia coli* DNA gyrase and topoisomerases I and III during *oriC* and pBR322 DNA replication in vitro. *J Biol Chem.* 1994 Jan 21;269(3):2093–9.
294. DiGate RJ, Marians KJ. *Escherichia coli* topoisomerase III-catalyzed cleavage of RNA. *J Biol Chem.* 1992 Oct 15;267(29):20532–5.
295. Li Z, Mondragón A, Hiasa H, Marians KJ, DiGate RJ. Identification of a unique domain essential for *Escherichia coli* DNA topoisomerase III-catalysed decatenation of replication intermediates. *Mol Microbiol.* 2000 Feb;35(4):888–95.
296. Champoux JJ. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem.* 2001;70:369–413.
297. Nichols MD, DeAngelis K, Keck JL, Berger JM. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. *EMBO J.* 1999 Nov 1;18(21):6177–88.
298. Mizuuchi K, Fisher LM, O’Dea MH, Gellert M. DNA gyrase action involves the

- introduction of transient double-strand breaks into DNA. *Proc Natl Acad Sci U S A*. 1980 Apr;77(4):1847–51.
299. Liu LF, Liu CC, Alberts BM. Type II DNA topoisomerases: enzymes that can unknot a topologically knotted DNA molecule via a reversible double-strand break. *Cell*. 1980 Mar;19(3):697–707.
300. Sugino A, Higgins NP, Brown PO, Peebles CL, Cozzarelli NR. Energy coupling in DNA gyrase and the mechanism of action of novobiocin. *Proc Natl Acad Sci U S A*. 1978 Oct;75(10):4838–42.
301. Roca J, Wang JC. DNA transport by a type II DNA topoisomerase: evidence in favor of a two-gate mechanism. *Cell*. 1994 May 20;77(4):609–16.
302. Berger JM, Fass D, Wang JC, Harrison SC. Structural similarities between topoisomerases that cleave one or both DNA strands. *Proc Natl Acad Sci U S A*. 1998 Jul 7;95(14):7876–81.
303. Berger JM, Gamblin SJ, Harrison SC, Wang JC. Structure and mechanism of DNA topoisomerase II. *Nature*. 1996 Jan 18;379(6562):225–32.
304. Gellert M, Mizuuchi K, O’Dea MH, Nash HA. DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc Natl Acad Sci U S A*. 1976 Nov;73(11):3872–6.
305. Klevan L, Wang JC. Deoxyribonucleic acid gyrase-deoxyribonucleic acid complex containing 140 base pairs of deoxyribonucleic acid and an $\alpha_2\beta_2$ protein core. *Biochemistry*. 1980 Nov 11;19(23):5229–34.
306. Adachi T, Mizuuchi M, Robinson EA, Appella E, O’Dea MH, Gellert M, et al. DNA sequence of the *E. coli* *gyrB* gene: application of a new sequencing strategy. *Nucleic Acids Res*. 1987 Jan 26;15(2):771–84.
307. Moriya S, Ogasawara N, Yoshikawa H. Structure and function of the region of the replication origin of the *Bacillus subtilis* chromosome. III. Nucleotide sequence of some 10,000 base pairs in the origin region. *Nucleic Acids Res*. 1985 Apr 11;13(7):2251–65.
308. Kampranis SC, Maxwell A. Conversion of DNA gyrase into a conventional type II topoisomerase. *Proc Natl Acad Sci*. 1996 Dec 10;93(25):14416–21.
309. Gellert M, Mizuuchi K, O’Dea MH, Itoh T, Tomizawa JI. Nalidixic acid resistance: a second genetic character involved in DNA gyrase activity. *Proc Natl Acad Sci U S A*. 1977 Nov;74(11):4772–6.
310. Sugino A, Peebles CL, Kreuzer KN, Cozzarelli NR. Mechanism of action of nalidixic acid: purification of *Escherichia coli* *nalA* gene product and its relationship to DNA gyrase and a novel nicking-closing enzyme. *Proc Natl Acad Sci U S A*. 1977 Nov;74(11):4767–71.
311. Morais Cabral JH, Jackson AP, Smith CV, Shikotra N, Maxwell A, Liddington RC. Crystal structure of the breakage-reunion domain of DNA gyrase. *Nature*. 1997 Aug 28;388(6645):903–6.
312. Sugino A, Higgins NP, Cozzarelli NR. DNA gyrase subunit stoichiometry and the covalent attachment of subunit A to DNA during DNA cleavage. *Nucleic Acids Res*. 1980 Sep 11;8(17):3865–74.

313. Heddle JG, Mittelheiser S, Maxwell A, Thomson NH. Nucleotide binding to DNA gyrase causes loss of DNA wrap. *J Mol Biol.* 2004 Mar 26;337(3):597–610.
314. Lanz MA, Klostermeier D. Guiding strand passage: DNA-induced movement of the gyrase C-terminal domains defines an early step in the supercoiling cycle. *Nucleic Acids Res.* 2011 Dec;39(22):9681–94.
315. Papillon J, Ménétret J-F, Batisse C, Hélye R, Schultz P, Potier N, et al. Structural insight into negative DNA supercoiling by DNA gyrase, a bacterial type 2A DNA topoisomerase. *Nucleic Acids Res.* 2013 Sep;41(16):7815–27.
316. Fisher LM, Mizuuchi K, O’Dea MH, Ohmori H, Gellert M. Site-specific interaction of DNA gyrase with DNA. *Proc Natl Acad Sci U S A.* 1981 Jul;78(7):4165–9.
317. Morrison A, Cozzarelli NR. Contacts between DNA gyrase and its binding site on DNA: features of symmetry and asymmetry revealed by protection from nucleases. *Proc Natl Acad Sci U S A.* 1981 Mar;78(3):1416–20.
318. Kirkegaard K, Wang JC. Mapping the topography of DNA wrapped around gyrase by nucleolytic and chemical probing of complexes of unique DNA sequences. *Cell.* 1981 Mar;23(3):721–9.
319. Kampranis SC, Bates AD, Maxwell A. A model for the mechanism of strand passage by DNA gyrase. *Proc Natl Acad Sci U S A.* 1999 Jul 20;96(15):8414–9.
320. Basu A, Schoeffler AJ, Berger JM, Bryant Z. ATP binding controls distinct structural transitions of *Escherichia coli* DNA gyrase in complex with DNA. *Nat Struct Mol Biol.* 2012 Apr 8;19(5):538–46, S1.
321. Wang JC. Moving one DNA double helix through another by a type II DNA topoisomerase: the story of a simple molecular machine. *Q Rev Biophys.* 1998 May;31(2):107–44.
322. Higgins NP, Peebles CL, Sugino A, Cozzarelli NR. Purification of subunits of *Escherichia coli* DNA gyrase and reconstitution of enzymatic activity. *Proc Natl Acad Sci U S A.* 1978 Apr;75(4):1773–7.
323. Baker TA, Funnell BE, Kornberg A. Helicase action of dnaB protein during replication from the *Escherichia coli* chromosomal origin in vitro. *J Biol Chem.* 1987 May 15;262(14):6877–85.
324. Higgins CF, Dorman CJ, Stirling DA, Waddell L, Booth IR, May G, et al. A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. *Cell.* 1988 Feb 26;52(4):569–84.
325. Dorman CJ, Barr GC, Ni Bhriain N, Higgins CF. DNA supercoiling and the anaerobic and growth phase regulation of tonB gene expression. *J Bacteriol.* 1988 Jun;170(6):2816–26.
326. Dorman CJ, Ni Bhriain N, Higgins CF. DNA supercoiling and environmental regulation of virulence gene expression in *Shigella flexneri*. *Nature.* 1990 Apr 19;344(6268):789–92.
327. Karem K, Foster JW. The influence of DNA topology on the environmental regulation of a pH-regulated locus in *Salmonella typhimurium*. *Mol Microbiol.* 1993 Oct;10(1):75–86.
328. Cheung KJ, Badarinarayana V, Selinger DW, Janse D, Church GM. A microarray-

- based antibiotic screen identifies a regulatory role for supercoiling in the osmotic stress response of *Escherichia coli*. *Genome Res.* 2003 Feb;13(2):206–15.
329. Santos-Pereira JM, Aguilera A. R loops: new modulators of genome dynamics and function. *Nat Rev Genet.* 2015 Oct;16(10):583–97.
330. Zaitsev EN, Kowalczykowski SC. A novel pairing process promoted by *Escherichia coli* RecA protein: inverse DNA and RNA strand exchange. *Genes Dev.* 2000 Mar 15;14(6):740–9.
331. Kasahara M, Clikeman JA, Bates DB, Kogoma T. RecA protein-dependent R-loop formation in vitro. *Genes Dev.* 2000 Feb 1;14(3):360–5.
332. Chédin F. Nascent Connections: R-Loops and Chromatin Patterning. *Trends Genet TIG.* 2016;32(12):828–38.
333. Itoh T, Tomizawa J. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proc Natl Acad Sci U S A.* 1980 May;77(5):2450–4.
334. Drolet M, Bi X, Liu LF. Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *J Biol Chem.* 1994 Jan 21;269(3):2068–74.
335. Huppert JL. Thermodynamic prediction of RNA-DNA duplex-forming regions in the human genome. *Mol Biosyst.* 2008 Jun;4(6):686–91.
336. Reaban ME, Griffin JA. Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature.* 1990 Nov 22;348(6299):342–4.
337. Daniels GA, Lieber MR. RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Res.* 1995 Dec 25;23(24):5006–11.
338. Drolet M, Phoenix P, Menzel R, Massé E, Liu LF, Crouch RJ. Overexpression of RNase H partially complements the growth defect of an *Escherichia coli* delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc Natl Acad Sci U S A.* 1995 Apr 11;92(8):3526–30.
339. Benham CJ. Duplex destabilization in superhelical DNA is predicted to occur at specific transcriptional regulatory regions. *J Mol Biol.* 1996 Jan 26;255(3):425–34.
340. Stolz R, Sulthana S, Hartono SR, Malig M, Benham CJ, Chedin F. Interplay between DNA sequence and negative superhelicity drives R-loop structures. *Proc Natl Acad Sci U S A.* 2019 26;116(13):6260–9.
341. Cambray G, Guimaraes JC, Mutalik VK, Lam C, Mai Q-A, Thimmaiah T, et al. Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.* 2013 May;41(9):5139–48.
342. Mutalik V, Guimaraes J, Cambray G, Lam C, Christoffersen M, Mai Q-A, et al. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods.* 2013 Mar 10;10.
343. Chandler MG, Pritchard RH. The effect of gene concentration and relative gene dosage on gene output in *Escherichia coli*. *Mol Gen Genet MGG.* 1975 Jun 1;138(2):127–41.
344. Pato ML, Howe MM, Higgins NP. A DNA gyrase-binding site at the center of the bacteriophage Mu genome is required for efficient replicative transposition. *Proc*

- Natl Acad Sci U S A. 1990 Nov;87(22):8716–20.
345. Oram M, Pato ML. Mu-like prophage strong gyrase site sequences: analysis of properties required for promoting efficient mu DNA replication. *J Bacteriol.* 2004 Jul;186(14):4575–84.
346. Casini A, Christodoulou G, Freemont PS, Baldwin GS, Ellis T, MacDonald JT. R2oDNA Designer: Computational Design of Biologically Neutral Synthetic DNA Sequences. *ACS Synth Biol.* 2014 Aug 15;3(8):525–8.
347. Burz DS, Ackers GK. Single-site mutations in the C-terminal domain of bacteriophage lambda cl repressor alter cooperative interactions between dimers adjacently bound to OR. *Biochemistry.* 1994 Jul 19;33(28):8406–16.
348. Kroner GM, Wolfe MB, Freddolino PL. *Escherichia coli* Lrp Regulates One-Third of the Genome via Direct, Cooperative, and Indirect Routes. *J Bacteriol.* 2019 01;201(3).
349. Tani TH, Khodursky A, Blumenthal RM, Brown PO, Matthews RG. Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis. *Proc Natl Acad Sci U S A.* 2002 Oct 15;99(21):13471–6.
350. Tapias A, López G, Ayora S. *Bacillus subtilis* LrpC is a sequence-independent DNA-binding and DNA-bending protein which bridges DNA. *Nucleic Acids Res.* 2000 Jan 15;28(2):552–9.
351. Chen S, Hao Z, Bieniek E, Calvo JM. Modulation of Lrp action in *Escherichia coli* by leucine: effects on non-specific binding of Lrp to DNA. *J Mol Biol.* 2001 Dec 14;314(5):1067–75.
352. Chen S, Rosner MH, Calvo JM. Leucine-regulated self-association of leucine-responsive regulatory protein (Lrp) from *Escherichia coli*. *J Mol Biol.* 2001 Sep 28;312(4):625–35.
353. Chen S, Iannolo M, Calvo JM. Cooperative Binding of the Leucine-Responsive Regulatory Protein (Lrp) to DNA. *J Mol Biol.* 2005 Jan 14;345(2):251–64.
354. Kuhlman TE, Cox EC. Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res.* 2010 Apr;38(6):e92.
355. Kuhlman TE, Cox EC. A place for everything: chromosomal integration of large constructs. *Bioeng Bugs.* 2010 Aug;1(4):296–9.
356. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, et al. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* 2017 04;45(D1):D543–50.
357. Green MR, Sambrook J. *Molecular Cloning: A Laboratory Manual* (Fourth Edition). Cold Spring Harbor Laboratory Press; 2012.
358. Sousa C, de Lorenzo V, Cebolla A. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiol Read Engl.* 1997 Jun;143 (Pt 6):2071–8.
359. Bryant JA, Sellars LE, Busby SJW, Lee DJ. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res.* 2014 Oct 13;42(18):11383–92.
360. Scholz SA, Diao R, Wolfe MB, Fivenson EM, Lin XN, Freddolino PL. High-Resolution Mapping of the *Escherichia coli* Chromosome Reveals Positions of High

- and Low Transcription. *Cell Syst.* 2019 27;8(3):212-225.e9.
361. Brambilla E, Sclavi B. Gene regulation by H-NS as a function of growth conditions depends on chromosomal position in *Escherichia coli*. *G3 Bethesda Md.* 2015 Feb 19;5(4):605–14.
362. Yousuf M, Iuliani I, Veetil RT, Seshasayee ASN, Sclavi B, Cosentino Lagomarsino M. Early fate of exogenous promoters in *E. coli*. *Nucleic Acids Res.* 2020 Mar 18;48(5):2348–56.
363. Postow L, Hardy CD, Arsuaga J, Cozzarelli NR. Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev.* 2004 Jul 15;18(14):1766–79.
364. Yeung E, Dy AJ, Martin KB, Ng AH, Del Vecchio D, Beck JL, et al. Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks. *Cell Syst.* 2017 26;5(1):11-24.e12.
365. Yus E, Lloréns-Rico V, Martínez S, Gallo C, Eilers H, Blötz C, et al. Determination of the Gene Regulatory Network of a Genome-Reduced Bacterium Highlights Alternative Regulation Independent of Transcription Factors. *Cell Syst.* 2019 Aug 28;9(2):143-158.e13.
366. Kano Y, Imamoto F. Requirement of integration host factor (IHF) for growth of *Escherichia coli* deficient in HU protein. *Gene.* 1990 Apr 30;89(1):133–7.
367. Chen D, Bowater R, Dorman CJ, Lilley DM. 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.* 1992 Sep 15;89(18):8784–8.
368. Nouaille S, Mondeil S, Finoux A-L, Moulis C, Girbal L, Cocaign-Bousquet M. The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. *Nucleic Acids Res.* 2017 Nov 16;45(20):11711–24.
369. Nogueira T, Springer M. Post-transcriptional control by global regulators of gene expression in bacteria. *Curr Opin Microbiol.* 2000 Apr;3(2):154–8.
370. Moll I, Afonyushkin T, Vytvytska O, Kaberdin VR, Bläsi U. Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. *RNA N Y N.* 2003 Nov;9(11):1308–14.
371. Sharp JS, Bechhofer DH. Effect of Translational Signals on mRNA Decay in *Bacillus subtilis*. *J Bacteriol.* 2003 Sep;185(18):5372–9.

Titre : Identification de contraintes locales impactant l'expression des gènes chez *Escherichia coli*

Mots clés : surenroulement, topologie, boucles d'ADN, régulation génique, transcription, épigénétique

Résumé : Les études menées chez *Escherichia coli* ont montré que l'organisation 3D de leur chromosome circulaire s'échelonne à plusieurs niveaux. Au niveau cellulaire, le chromosome est organisé en 6 macrodomaines, d'une taille proche de 1Mb. Au niveau moléculaire, les NAP (Protéines Associées au Nucléotide) organisent l'ADN en microdomaines d'environ 10 kb qui présentent des surenroulements de l'hélice d'ADN indépendants. Pour mesurer l'impact de l'organisation du génome sur l'expression génique d'une façon systématique, une étude à grande échelle a été menée par notre équipe. Ce travail a montré que l'expression d'un gène rapporteur isolé de part et d'autre par des terminateurs de transcription peut varier d'un facteur 1300 en fonction de sa position sur le chromosome.

Nous avons démontré que ces variations d'expression sont au moins en partie dues aux microdomaines qui présentent des niveaux locaux de surenroulements différents. Par l'utilisation d'une protéine de pontage hétérologue, nous avons induit des boucles d'ADN dans le chromosome d'*E. coli* permettant d'isoler un gène d'intérêt de son contexte génomique. Nous avons montré que la topologie locale est le moteur de la variabilité de l'expression génique. Enfin, les analyses bioinformatiques que nous avons réalisées prouvent que les phénomènes que nous avons observés grâce aux boucles artificiellement induites sont naturellement présents chez la bactérie, et permettent une régulation globale de l'expression des gènes à l'échelle génomique.

Title : Identification of local constraints impacting gene expression in *Escherichia coli*.

Keywords : supercoiling, DNA-loops, epigenetic, NAP, genome architecture, context sensitivity

Abstract : Recent studies conducted in *Escherichia coli* showed the macrodomain and microdomain organization of the chromosome. Macrodomains correspond to long DNA region of ~1 Mbp. Microdomains are smaller DNA segments, allowed by the binding of NAPs (Nucleoid Associated Proteins), that shape the bacterial chromosome. These microdomains are successive DNA-loops of ~10kbs and present heterogeneous local supercoil levels. Our team conducted a genome-wide study showing the impact of the genomic context on the expression of a transcriptionally-isolated reporter gene. The expression of the reporter gene varies 1,300-fold depending on its chromosomal position.

Focusing on two chromosomal regions, we showed that this variability is at least partially due to microdomains, that present heterogeneous superhelicity. Using heterologous Lambda-cl protein, we developed a tool to artificially induce loops in the DNA. We used this loop to isolate the gene from its genomic context and showed that it was indeed the local topology that was responsible of the expression variability. Using bio-informatic approach, we discovered that *E. coli* Lrp NAP naturally uses the DNA-looping mechanism to regulate expression of its target genes. This mechanism allows transcription regulation of genes at the genomic scale.