



**HAL**  
open science

# La superhélicité de l'ADN comme cible clef de l'adaptation au cours de 20000 générations d'évolution expérimentale chez *Escherichia coli*

Estelle Crozat

► **To cite this version:**

Estelle Crozat. La superhélicité de l'ADN comme cible clef de l'adaptation au cours de 20000 générations d'évolution expérimentale chez *Escherichia coli*. Autre [q-bio.OT]. Université Joseph-Fourier - Grenoble I, 2007. Français. NNT: . tel-00128144v1

**HAL Id: tel-00128144**

**<https://theses.hal.science/tel-00128144v1>**

Submitted on 30 Jan 2007 (v1), last revised 30 Nov 2008 (v2)

**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.

UNIVERSITE JOSEPH FOURIER – GRENOBLE I

**THESE**

Pour obtenir le grade de  
DOCTEUR DE L'UNIVERSITE JOSEPH FOURIER

Discipline : Biologie

Présentée et soutenue publiquement par

**Estelle CROZAT**

Le 29 janvier 2007

---

**La superhélicité de l'ADN comme cible clef de l'adaptation  
au cours de 20 000 générations  
d'évolution expérimentale chez *Escherichia coli***

---

Composition du jury

Pr. Michel Robert-Nicoud	Président
Pr. Claude Gutierrez	Rapporteur
Dr. Jean-Luc Pernodet	Rapporteur
Dr. Bénédicte Michel	Examinatrice
Pr. Dominique Schneider	Directeur de thèse

Thèse préparée au sein du Laboratoire Adaptation et Pathogénie des Microorganismes

UMR5163 CNRS/UJF

## REMERCIEMENTS

Je remercie tout d'abord les membres de mon jury, Messieurs Gutierrez et Pernodet, ainsi que Madame Michel et Monsieur Robert-Nicoud pour avoir accepté de juger ce travail.

Je voudrais ensuite remercier Dominique, qui a fait le pari de transformer une chimiste en biologiste (j'espère qu'il a réussi !) et qui m'a formée et accompagnée tout au long de cette thèse, en réussissant à m'apprendre presque autant de choses que mon Grand-Père-qui-sait-presque-tout (!). Son enthousiasme et ses qualités pédagogiques en font quelqu'un d'indispensable mais aussi de très agréable à la tête d'une équipe de recherche. Je pars confiante, Nicole s'occupe déjà du café...(merci Nicole !)

Je suis également très reconnaissante à Hans et à Marie-France pour m'avoir accueillie dans leur laboratoire et m'avoir permis d'y rester.

Merci à Richard Lenski pour avoir eu la merveilleuse idée de fonder ces 12 populations, et pour m'avoir de plus accueillie dans son laboratoire de la Michigan State University pendant quelques semaines ; merci à Margareth Riley et Cynthia Winkworth (University of Massachusetts) pour le séquençage de quelques paires de bases ; enfin, merci à Lauriane Kuhn et Jérôme Garin (DRDC, CEA, Grenoble) pour leur aimable collaboration et le séquençage de protéines.

Je souhaite remercier chaleureusement toutes les personnes du labo pour leur aide, leurs conseils, leur présence, enfin tout quoi ! Et entre autres...

Evelyne, Corinne et Nadège pour avoir supporté mes questions sans jamais se moquer, mais aussi pour les longues et nombreuses discussions que nous avons eues. Je leur souhaite beaucoup de courage pour la suite...

Un grand merci à Ludo, le Maître de la Biochimie, pour tout ce qu'il m'a appris...que la Force soit avec Lui.

Ali, Cécile et Stephan pour m'avoir enseigné les secrets de la purification de protéines, les 2D et les footprints.

Christianne et Véro pour les quelques milliers d'érans et de tubes de tampon phosphate stérilisés ces 4 dernières années !

Nicole, Michèle, ainsi que Magali pour avoir géré les commandes hyper-urgentes-qui-doivent-arriver-hier et leur patience dans mes déboires avec l'imprimante...

Sylvain et Alex pour leurs blagues pas drôles, les parties de UT... Et un peu leurs bons conseils, aussi !

Et, juste pour leur amitié et leur soutien, Dorian, Babeth, JeanFi, Polo...je vous souhaite beaucoup de chance, de courage et de réussite pour le futur.

Laetitia, Claire et Mike : vous n'avez jamais été si près de la fin !!! (Mike, t'auras été prévenu...)

Plus personnellement, je souhaite remercier toute ma famille pour leur soutien indéfectible depuis tant d'années...je ne serais pas là sans eux.

Et sans oublier, bien sûr, les copines de toujours, Amé, Sof et Béa !!

# TABLE DES MATIÈRES

TABLE DES ILLUSTRATIONS	3
ABRÉVIATIONS	5
GLOSSAIRE	6
AVANT PROPOS	7
INTRODUCTION PARTIE I	
<b>La Topologie de l'ADN</b>	<b>10</b>
<b>I. Les acteurs de la topologie de l'ADN</b>	<b>13</b>
<b>A. Les topoisomérases</b>	<b>13</b>
1. Les Topos de type I	15
2. Les Topos de type II	19
<b>B. Les protéines de type histone</b>	<b>22</b>
1. Les NAPs ne possédant aucune spécificité de fixation	23
2. La protéine IHF	27
3. La protéine Fis	28
4. Dynamique d'expression des NAPs	37
<b>II. Organisation du chromosome dans la cellule</b>	<b>38</b>
<b>A. Les domaines topologiques</b>	<b>38</b>
1. La transcription	40
2. Les complexes SMC	41
3. Les Topos	42
4. Les protéines liées au nucléoïde	43
5. Les autres protéines	45
<b>B. Les macrodomaines</b>	<b>46</b>
<b>III. Impact de la superhélicité de l'ADN sur l'expression des gènes</b>	<b>49</b>
<b>IV. Régulation et contrôle homéostatique du niveau de superhélicité de l'ADN</b>	<b>50</b>
<b>A. Contrôle homéostatique de la superhélicité de l'ADN en fonction des conditions environnementales</b>	<b>51</b>
<b>B. Autres stress environnementaux et topologie de l'ADN</b>	<b>53</b>
INTRODUCTION PARTIE II	
<b>La Stratégie d'Evolution Expérimentale</b>	<b>56</b>
<b>Manuscrit</b> : Global regulatory networks plasticity during experimental evolution in <i>Escherichia coli</i>	

RÉSULTATS	88
RÉSULTATS PARTIE I	
<b>La Topologie de l'ADN comme Cible de la Sélection Naturelle</b>	<b>90</b>
<b>Article</b> : Long-term experimental evolution in <i>Escherichia coli</i> . XII. DNA topology as a key target of selection	
RÉSULTATS PARTIE II	
<b>Evolution Bactérienne et Superhélicité de l'ADN : Degré Exceptionnel de Parallélisme Génétique et Phénotypique</b>	<b>95</b>
<b>Manuscrit</b> : Bacterial evolution by exceptional levels of parallel genetic changes in DNA superhelicity	
RÉSULTATS PARTIE III	
<b>Une Nouvelle Fonction de Régulation pour FIS</b>	<b>99</b>
<b>I. Matériel et méthodes</b>	<b>100</b>
<b>A. Milieux de cultures, souches et plasmides</b>	<b>100</b>
<b>B. Electrophorèses bidimensionnelles</b>	<b>101</b>
<b>C. Analyse des électrophorèses bidimensionnelles</b>	<b>102</b>
<b>D. Fusions transcriptionnelles</b>	<b>102</b>
<b>E. Retards sur gels</b>	<b>103</b>
<b>F. Empreintes à la DNase I</b>	<b>104</b>
<b>II. Résultats</b>	<b>105</b>
<b>A. Modification des profils protéiques</b>	<b>105</b>
<b>B. Les porines d'<i>E. coli</i></b>	<b>106</b>
<b>C. Effets des mutations <i>topA</i> et <i>fis</i> sur la transcription de <i>ompF</i></b>	<b>108</b>
1. La transcription de <i>ompF</i> dépend de la superhélicité de l'ADN et de Fis	108
2. L'activation de <i>ompF</i> par Fis est directe	109
3. Détermination des sites de fixation de Fis sur <i>ompF</i>	110
<b>D. Régulation des gènes <i>ompF</i>, <i>ompC</i> et <i>ompR</i> chez <i>E. coli</i> B et K12</b>	<b>112</b>
1. Régulation du gène <i>ompF</i> K12	112
2. Régulation des gènes <i>ompC</i> et <i>ompR</i>	114
<b>E. Régulation de <i>ompF</i> au cours de l'évolution expérimentale</b>	<b>116</b>
<b>III. Discussion</b>	<b>117</b>
DISCUSSION GÉNÉRALE	121
<b>I. Parallélisme de l'évolution</b>	<b>123</b>
<b>II. Plasticité des réseaux de régulation globale</b>	<b>125</b>
<b>III. Les mécanismes moléculaires de l'adaptation</b>	<b>127</b>
<b>IV. La superhélicité de l'ADN dans la nature</b>	<b>130</b>
<b>V. Perspectives</b>	<b>131</b>
RÉFÉRENCES BIBLIOGRAPHIQUES	132

## TABLE DES ILLUSTRATIONS

Figure 0 : Les paramètres de définition de la superhélicité de l'ADN	11
Figure 1 : Activité des différentes Topos	13
Figure 2 : Mécanisme d'action d'une Topo de type IIA	14
Figure 3 : Structure d'une Topo de type IA	15
Figure 4 : Mécanisme d'action de la TopoI	15
Figure 5 : Modèle de Liu et Wang et boucles R	17
Figure 6 : Les différents domaines des Topos de type IIA	20
Figure 7 : Mécanisme d'action des Topos de type II, ici de la gyrase	20
Figure 8 : Rôle de la sous-unité GyrA de la gyrase d' <i>E. coli</i> dans la modification de la superhélicité de l'ADN	21
Figure 9 : Effets sur l'ADN de la protéine H-NS	24
Figure 10 : Mécanisme de répression de la transcription par H-NS	24
Figure 11 : Effet sur l'ADN des protéines courbant l'ADN	25
Figure 12 : Effet antagoniste de HU et H-NS sur un plasmide	26
Figure 13 : Protection de l'ADN par Dps	27
Figure 14 : Effet de la protéine Fis sur la structure du nucléoïde et la transcription	29
Figure 15 : Structure d'un dimère de Fis	30
Figure 16 : Régulation du promoteur de <i>fis</i>	31
Figure 17 : Complémentarité entre l'ARNr 16S et l'extrémité 5' non traduite de l'ARNm <i>fis</i>	33
Figure 18 : Région promotrice de <i>rnnB</i>	34
Figure 19 : Modèle d'interaction de Fis avec l'ARN polymérase sur le promoteur <i>rnnB</i> P1	34
Figure 20 : Système de recombinaison par Hin	36
Figure 21 : Types de structures formées par MukB	41
Figure 22 : Système rapporteur sensible à la superhélicité de l'ADN	42
Figure 23 : Modèle de compaction du chromosome par H-NS et Fis	44
Figure 24 : Structure des macrodomaines d' <i>E. coli</i> et leur réponse à la topologie de l'ADN	47
Figure 25 : Réorganisation du chromosome en fonction des conditions de croissance	48
Figure 26 : Boucle de régulation homéostatique de la superhélicité de l'ADN	51
Figure 27 : Détermination du fitness de clones évolués par des expériences de compétitions	59
Figure 28 : Effet sur le fitness des allèles évolués <i>topA</i> <sup>-1</sup> et <i>fis</i> <sup>-1</sup> dans le contexte génétique du clone ancêtre	98
Figure 29 : Les régions promotrices des gènes <i>ompF</i> et <i>ompC</i>	107
Figure 30 : Activités spécifiques $\beta$ -galactosidase correspondant aux fusions <i>ompFB-lacZ</i> dans différents contextes génétiques isogéniques du clone ancêtre	108
Figure 31 : Activités spécifiques $\beta$ -galactosidase des fusions <i>ompFB-lacZ</i> en fonction de la concentration en NaCl	108
Figure 32 : Fixation de Fis sur la région promotrice de <i>ompF</i>	109
Figure 33 : Empreintes à la DNase I de la protéine Fis sur la région promotrice de <i>ompF</i>	110
Figure 34 : Région promotrice de <i>ompF</i> et localisation des sites de fixation de Fis	111
Figure 35 : Activités spécifiques $\beta$ -galactosidase des fusions transcriptionnelles <i>ompFK12-lacZ</i> dans les contextes isogéniques K12 et K12 <sup><math>\Delta</math>fis</sup>	112
Figure 36 : Activités spécifiques $\beta$ -galactosidase des fusions transcriptionnelles <i>ompFB-lacZ</i> ou <i>ompFK12-lacZ</i> dans les contextes génétiques <i>E. coli</i> B ou K12	112

Figure 37 : Activités spécifiques $\beta$ -galactosidase des fusions transcriptionnelles <i>ompR - lacZ</i> ou <i>ompC - lacZ</i> dans les contextes génétiques <i>E. coli</i> B et K12	113
Figure 38 : Fixation de Fis sur les régions promotrices de <i>ompR</i> (A) ou <i>ompC</i> (B)	114
Figure 39 : Détermination des sites de fixation de Fis sur les régions promotrices de <i>ompR</i> ou <i>ompC</i> par des expériences d'empreintes à la DNase I	114
Figure 40 : Régions promotrices de <i>ompR</i> (A) ou <i>ompC</i> (B) et localisation des sites de fixation de Fis	115
Figure 41 : Fixation de OmpR sur <i>ompF</i>	118
Tableau 1 : Nom des souches utilisées dans les 3 parties des résultats	89
Tableau 2 : Modification de l'expression de protéines par les allèles évolués <i>topA</i> et <i>fis</i> isolés dans la population Ara-1	105

## ABRÉVIATIONS

<b>ADN :</b>	Acide DéoxyRibonucléique
<b>ARN (r,m,t) :</b>	Acide Ribonucléique (Ribosomique, Messenger, de Transfert)
<b>ATP :</b>	Adénosine TriPhosphate
<b>CTD :</b>	Domaine C-Terminal
<b>CTP :</b>	Cytosine TriPhosphate
<b>GTP :</b>	Guanine TriPhosphate
<b>IPTG :</b>	Isopropyl- $\beta$ -D-thiogalactopyranoside
<b>LB :</b>	Luria Broth
<b>Lk :</b>	Linking number (nombre de liens)
<b>NAP :</b>	Protéine Associée au Nucléotide
<b>NTD :</b>	Domaine N-terminal
<b><i>ompFB</i> ou K12 :</b>	Gène <i>ompF</i> d' <i>E. coli</i> B ou K12
<b>OmpR-P :</b>	Protéine OmpR phosphorylée
<b>RBS :</b>	Site de Fixation des Ribosomes
<b>TA :</b>	Tétrazolium-Arabinose
<b>TopoI, III, IV :</b>	Topoisomérase I, III ou IV
<b>TopoIA, IIA :</b>	Topoisomérases de type IA, IIA
<b>Topos :</b>	Topoisomérases
<b>Xgal :</b>	5-bromo-4-chloro-3-indolyl- $\beta$ -D-Galactopyranoside

## GLOSSAIRE

**Fitness** : capacité de reproduction d'une souche.

**Fixation d'une mutation** : une mutation est dite « fixée » lorsqu'elle est portée par la majorité des clones composant la population à un temps donné.

**Parallélisme de l'évolution** : Il s'agit de l'évolution indépendante, et dans une direction similaire, d'un même caractère, dans différentes populations soumises à des environnements équivalents. Le parallélisme de l'évolution peut se refléter à différents niveaux : phénotypique, génétique et moléculaire.

**Interférence clonale** : phénomène observé lorsqu'une population est composée de plusieurs sous-populations cellulaires, portant des mutations différentes mais qui apportent un avantage équivalent. Les sous-populations coexistent, empêchant la fixation rapide d'une des mutations bénéfiques. Lorsqu'une mutation bénéfique supplémentaire apparaît dans une des sous-populations, celle-ci devient majoritaire ainsi que les mutations qu'elle porte. Ce phénomène ralentit donc la fixation de mutations bénéfiques.

**Spécialisation écologique** : au cours de l'évolution, des fonctions inutiles dans l'environnement de sélection peuvent être perdues ou réduites. Deux mécanismes peuvent expliquer ce phénomène : l'accumulation de mutations ou la pleïotropie antagoniste. Les bactéries sont alors spécialisées pour leur environnement de sélection, et inadaptées à d'autres environnements nécessitant la présence des fonctions éliminées ou réduites au cours de l'évolution.

**Pleïotropie antagoniste** : il s'agit d'un mécanisme conduisant à la spécialisation écologique. Une mutation, sélectionnée pour le bénéfice qu'elle apporte dans l'environnement de sélection, peut avoir des effets néfastes dans d'autres environnements. Elle a alors des effets pleïotropes (c'est-à-dire multiples) antagonistes.

**Accumulation de mutations** : il s'agit d'un autre mécanisme conduisant à la spécialisation écologique. Au cours de l'évolution, des mutations neutres, voire faiblement délétères dans l'environnement de sélection peuvent s'accumuler dans des gènes non soumis à la sélection naturelle.

# AVANT PROPOS

La superhélicité de l'ADN peut être considérée, chez tous les organismes vivants, comme une fonction biologique cruciale, réalisée en particulier par des topoisomérases et des histones, qui ont été fortement conservées au cours de l'évolution. Elles contrôlent le niveau de superhélicité de façon adéquate et permettent ainsi à la cellule d'effectuer les réactions indispensables à sa survie et à celle de sa descendance. L'absence totale d'une de ces enzymes est par exemple létale chez la souris ; des maladies génétiques rares, liées au déficit d'une de ces protéines, ont été caractérisées chez l'homme, comme par exemple les syndromes de Werner (dysfonctionnement d'une hélicase entraînant un vieillissement prématuré et le développement de pathologies liées au vieillissement) ou de Bloom (échange d'ADN trop fréquent amenant à un retard de croissance et de nombreux cancers). Chez les bactéries, l'altération de ces enzymes peut conduire à des cellules anucléées. Des molécules ciblant les topoisomérases, comme les quinolones, ont d'ailleurs été développées, et sont utilisées contre *Plasmodium falciparum*, parasite causant la malaria, ou comme antibiotiques.

Le maintien d'un niveau précis de superenroulement de l'ADN est primordial. La superhélicité et la structuration de l'ADN au sein de la cellule garantissent l'intégrité du chromosome et sa compaction, mais aussi son accessibilité par différentes enzymes, notamment les polymérases, afin de réaliser toutes les réactions nécessaires à la viabilité de la cellule : la réplication et la réparation de l'ADN, la ségrégation des chromosomes avant la division cellulaire, la transcription globale du génome, et la recombinaison. Les chromosomes sont des structures dynamiques, tant au niveau du degré de superhélicité de l'ADN que de leur organisation spatiale globale. Les conditions de l'environnement (conditions nutritionnelles et stress thermiques, osmotiques, oxydatifs...) sont connues pour affecter cette structuration. Tous les processus vitaux (réplication, transcription...), mais également la régulation de l'expression des génomes, pourraient donc être orchestrés par la structuration même de l'ADN.

Une des principales caractéristiques des bactéries est leur capacité à coloniser toutes les niches écologiques, c'est-à-dire à s'adapter à des conditions de croissance et de stress variées. L'adaptation à de telles conditions requiert une plasticité extrême des phénotypes bactériens. Lors de stress ou de changements d'environnement, ces modifications phénotypiques sont associées à un remaniement total de l'expression du génome, pour s'adapter aux nouvelles conditions. Afin de réaliser cela, il est essentiel pour la bactérie de détecter les conditions de

l'environnement et de transmettre ces signaux afin de modifier l'expression du génome. La superhélicité de l'ADN a été suggérée comme intervenant dans l'adaptation à de nouvelles conditions de croissance, de par sa capacité à être modifiée en fonction des conditions et à remodeler l'expression globale des gènes. Le dogme central des sciences de la vie est que les systèmes biologiques émergent d'une évolution darwinienne, c'est-à-dire de modifications génétiques se produisant au hasard dans un génome, suivies de la sélection des individus mutants ayant un avantage sélectif. Si la superhélicité de l'ADN est impliquée dans l'adaptation des bactéries à leur environnement, elle doit donc constituer une cible évolutive, c'est-à-dire une cible de la sélection naturelle.

L'objectif de ma thèse est de comprendre les mécanismes d'adaptation des bactéries à leur environnement, et en particulier l'impact de la superhélicité de l'ADN sur les processus évolutifs. Peut-elle être la cible de la sélection naturelle ? Et si oui, par quels mécanismes ? La superhélicité peut-elle être impliquée dans la transmission des signaux de l'environnement à l'expression des gènes, pour permettre une réponse bactérienne adaptée à un environnement donné ? Pour répondre à ces questions, j'ai utilisé une stratégie d'évolution expérimentale, où les bactéries sont adaptées pendant plusieurs dizaines de milliers de générations à des transitions nutritionnelles croissance-carence et carence-croissance, et ce à partir d'un ancêtre bien identifié.

Dans l'introduction bibliographique, je présenterai deux parties : la première décrira les différents acteurs protéiques de la superhélicité de l'ADN, l'organisation du chromosome, le rôle de la superhélicité au sein de la cellule, notamment en fonction des conditions environnementales, et enfin son rôle clef dans la régulation de l'expression des gènes. La seconde partie présentera le système d'évolution expérimentale utilisé, ainsi que les résultats obtenus depuis l'initiation de cette expérience.

# INTRODUCTION

INTRODUCTION

PARTIE I

LA TOPOLOGIE DE L'ADN

La relation topologique entre les deux brins d'une même molécule d'ADN, comme par exemple le chromosome circulaire d'*Escherichia coli*, est caractérisée par le « nombre de liens », appelé Linking number, Lk. C'est un entier qui est égal au nombre minimal de passages nécessaires d'un des deux brins par l'autre pour leur séparation. A cette fin, il est bien sûr indispensable de casser un des deux brins. Cela signifie que le nombre d'enlacements entre les deux brins d'une molécule d'ADN est constant tant que les deux brins sont intacts, donc Lk est une propriété topologique. L'enroulement de l'ADN est réalisé à deux niveaux de structuration, géométriques. Le tortillement des brins autour d'eux-mêmes au sein de la double hélice est décrit par le paramètre Tw (pour twist) (**Figure 0Ab**). De ce tortillement va dépendre le pas hélicoïdal de l'ADN, c'est-à-dire le nombre de paires de bases par tour d'hélice. Le vrillage de l'hélice sur elle-même est décrit par le paramètre Wr (pour writh) (**Figure 0Ac**), qui apparaît lorsque la molécule d'ADN circulaire ne se situe pas dans un plan. Lk est composé de ces deux paramètres, qui sont intimement liés. Une double hélice d'ADN adopte une conformation de moindre énergie entre Tw et Wr (**Figure 0Ad**). Un signe est attribué arbitrairement à Lk : positif pour une hélice de pas gauche, et négatif pour une hélice de pas droit (**Figure 0B**). Chez les bactéries, l'ADN est superenroulé négativement dans la cellule, ce qui permet notamment de condenser une molécule d'ADN de 1,5mm de long dans une cellule d'environ 2 $\mu$ m.

Tous les processus cellulaires vitaux, comme la réplication et la réparation de l'ADN, la transcription, et tous les phénomènes de recombinaison et de transposition, dépendent et influent sur le niveau de superhélicité de l'ADN. En effet, ils introduisent naturellement des boucles ou des nœuds dans la double hélice. La progression de la fourche de réplication ainsi que le déplacement des molécules d'ARN polymérase au cours de la transcription des gènes provoquent également des torsions importantes (Liu et Wang, 1987). D'autre part, la réplication et la transcription nécessitent l'ouverture de la double hélice d'ADN.

L'état thermodynamique le plus stable pour une molécule d'ADN circulaire est l'état totalement relaxé. La superhélicité négative introduite dans l'ADN par différentes enzymes est donc une source d'énergie fabuleuse qui va permettre la réalisation des différents processus cellulaires. En effet, une superhélicité négative correspond à un enroulement de l'ADN inférieur à celui adopté par l'ADN sans contrainte ( $Lk_0$ ). Ce phénomène facilite donc l'ouverture de la double hélice. De plus, cette déformation peut également être stabilisée par la fixation de protéines. Tout processus nécessitant ou utilisant ce phénomène, comme la fixation

des protéines de type histone ou de l'ARN polymérase, sera ainsi favorisé sur un ADN superenroulé négativement.

Cependant, le niveau de superhélicité doit rester précis, une déviation du degré optimal étant rapidement défavorable pour la cellule. Par exemple, une superhélicité négative trop élevée génère des structures d'ADN aberrantes (extrusion de structures cruciformes au niveau de séquences répétées, triplex d'ADN et des boucles R d'hybridation ADN-ARNm (Hanvey et al., 1988 ; Drolet et al., 2003). D'autre part, un niveau de superhélicité négative trop faible génère des problèmes de ségrégation du chromosome (Hiraga et al., 1989) et des fonctions chromosomiques déficientes conduisent à la mort cellulaire (Zechiedrich et al., 1997). Le maintien d'un niveau adéquat de superhélicité du chromosome est donc crucial pour le bon fonctionnement de toute cellule. De plus, sans structuration globale du chromosome, toute altération locale de la superhélicité se propagerait sans limite à l'ensemble de la molécule d'ADN. Afin d'éviter ces « accidents topologiques », non seulement le niveau de superhélicité est contrôlé de façon précise dans la cellule, mais le chromosome est organisé de façon globale en domaines topologiques séparés par des barrières qui empêchent la propagation de ces altérations (Espeli et Mariani, 2004). Du niveau de superhélicité va également dépendre l'expression d'un grand nombre de gènes : 7% des gènes d'*E. coli* sont en effet directement affectés par une diminution de la superhélicité (Peter et al., 2004). Parmi ces gènes, certains codent des régulateurs globaux qui, à leur tour, vont avoir des effets variés sur d'autres gènes. De plus, des modifications de l'environnement, comme des stress (carence nutritionnelle, osmotique, oxydatif, thermique), sont connues pour modifier d'une part le niveau de superhélicité, et d'autre part les profils globaux de transcription, permettant une réponse adaptée à ces stress. Superhélicité et transcription sont donc intimement liées.

Le niveau de superhélicité est constitué de deux composantes : la superhélicité dont le niveau est contrôlé par les topoisomérases est dite « non contrainte », car sa valeur est conservée après l'action des topoisomérases. La superhélicité « contrainte » est réalisée par des protéines de type histone se liant au nucléoïde, et y induisant des courbures ou des pontages entre molécules. Cette seconde composante est perdue dès que ces protéines ne sont plus liées à l'ADN.

## I. Les acteurs de la topologie de l'ADN

### A. Les topoisomérases

De par leur rôle cellulaire fondamental, les topoisomérases (Topos) sont présentes chez tous les organismes vivants. Malgré une différence de séquences considérable entre les enzymes bactériennes, archéennes et eucaryotes, des conformations structurales proches sont toutefois retrouvées dans leurs domaines d'activité (Champoux, 2001 ; Corbett et Berger, 2004).

Il existe deux types de Topos, I et II, chacun divisé en deux familles A et B, mais toutes présentent plusieurs points communs : une structure en forme de pince qui va permettre les activités de coupure de l'ADN et de transfert de brins, une tyrosine catalytique responsable de l'activité de clivage et qui crée une liaison covalente avec l'une des deux extrémités de l'ADN clivé, une cavité de stockage temporaire des fragments d'ADN, et un changement conformationnel associé à une rotation ou un mouvement de l'ADN.

Les Topos de type I coupent un seul brin d'ADN (**Figure 1**). Les familles IA et IB diffèrent par la liaison formée entre la tyrosine catalytique et l'ADN. Les Topos IA forment une liaison avec l'extrémité 5' phosphate de l'ADN, transfèrent l'autre brin et referment la coupure. Ceci a pour effet de changer le paramètre Lk par incréments de 1 (Brown et Cozzarelli, 1981). Ces Topos sont capables de relâcher les supertours négatifs seulement, mais pas jusqu'à une relaxation complète (Champoux, 2001). Les Topos IB forment une liaison avec une extrémité 3' phosphate de l'ADN. Contrairement aux Topos IA, elles permettent au brin coupé d'effectuer une ou plusieurs rotations autour du brin intact (**Figure 1**), ce qui permet des modifications de Lk plus importantes (Stivers et al., 1997 ; Stewart et al., 1998). Elles sont par contre capables de relâcher les supertours positifs comme négatifs, la relaxation étant cette fois-ci complète.

Les Topos de type II sont capables de cliver les deux brins d'ADN et de transférer une double hélice d'ADN, provenant de la même molécule, ou d'une autre, à travers la coupure. Elles utilisent pour cela de l'ATP et le changement résultant pour le paramètre Lk est de 2 (**Figure 1** ; Gellert et al., 1976). Elles possèdent des activités diverses de relâchement, de caténation/décation, et d'introduction ou de suppression de nœuds. Parmi toutes les Topos,

la gyrase est la seule capable d'introduire des supertours négatifs dans l'ADN. La classification en Topos de type IIA et IIB est essentiellement basée sur des considérations structurales (Nichols et al., 1999).

Malgré ces différences, des similarités structurales sont tout de même observées au sein des Topos. Les Topos de type IA, IIA et IIB partagent deux domaines essentiels pour la liaison à l'ADN et sa coupure. Le premier est le domaine 5Y-CAP, appelé ainsi car il se retrouve dans un grand nombre de protéines capables de lier l'ADN, dont la protéine CAP (Catabolite Activator Protein) (Schultz et al., 1991). Le résidu catalytique de ce domaine est une tyrosine qui se retrouve liée de façon covalente à l'extrémité 5' de l'ADN après sa coupure (Depew et al., 1978 ; Tse et al., 1980). Le second domaine commun à ces Topos est le domaine « toprim », appelé ainsi car il se retrouve dans certaines Topos mais aussi dans des primases bactériennes, ou encore dans d'autres enzymes qui catalysent des transferts de groupements phosphate ou qui hydrolysent des liaisons phosphodiester (Aravind et al., 1998). Il est constitué de 4 ou 5 feuillets  $\beta$  parallèles entourés d'hélices  $\alpha$ . Un groupement de résidus acides forme un site de liaison pour les ions  $Mg^{2+}$ , nécessaires à l'action des Topos (Osheroff, 1987). Après l'assemblage de la topoisomérase, ce groupement se retrouve proche du domaine 5Y-CAP, avec lequel il permet le mouvement des domaines qui ouvrent ou ferment la structure de la protéine et permettent au brin d'ADN de passer entre eux.

Les Topos de types II partagent également deux domaines dont les activités sont liées à la fixation et à l'hydrolyse de l'ATP (Bergerat et al., 1997 ; Corbett et Berger, 2004). Le site de liaison de l'ATP est le domaine GHKL, appelé ainsi en fonction des différentes protéines dans lesquelles il se trouve (Gyrase, Hsp90, histidine kinase et MutL). Il est composé de 8 feuillets  $\beta$  couverts sur une face par 5 hélices  $\alpha$ , et d'une boucle riche en résidus glycines. Ce site se dimérise une fois l'ATP fixé et les deux boucles entrent alors en contact, formant une porte pour l'hélice d'ADN (**Figure 2**). L'ATP est ensuite hydrolysé par le domaine « transducteur » au niveau d'un résidu lysine. La fixation de l'ATP entraîne une rotation entre les domaines GHKL et transducteur, qui positionne le résidu lysine dans son site catalytique. Cette modification de structure serait un signal pour les régions liées à l'ADN qui dirigerait alors son passage pendant l'ouverture de l'autre brin.

Les Topos de type II sont les cibles de deux types d'antibiotiques : les coumarines et les quinolones (Drlica et Zhao, 1997 ; Levine et al., 1998). Les coumarines se fixent sur les sous-unités ATPases et entrent en compétition avec l'ATP. Les quinolones se fixent sur l'enzyme complexée à l'ADN et l'immobilisent ainsi ; les complexes extraits à la suite de tels traitements montrent la gyrase liée de façon covalente à l'extrémité 5' de l'ADN coupé. Ceci entraîne la fragmentation de l'ADN, aboutissant à la mort cellulaire.

Chez *E. coli*, quatre Topos sont présentes. Deux sont de type IA : la TopoI, codée par *topA* et la TopoIII, codée par *topB*. Deux sont de type IIA : la gyrase, codée par *gyrA* et *gyrB*, et la TopoIV, codée par *parC* et *parE*.

## 1. Les Topos de type I

### a. Mécanisme d'action

Les deux Topos de type IA d'*E. coli* sont monomériques (**Figure 3**). Les Topos IA d'*E. coli* sont composées de 2 domaines essentiels :

- un domaine N-terminal catalytique formant un tore, pouvant cliver l'ADN et s'ouvrir suffisamment pour transférer un brin d'ADN. Ce domaine contient la tyrosine catalytique en position 319 de la TopoI, mais est incapable de l'activité de relâchement. Les domaines N-terminaux des Topos I et III partagent une grande homologie.
- un domaine C-terminal participant à la fixation de l'ADN (Ahumada et Tse-Dinh, 1998), ne montrant par contre aucune homologie entre les deux enzymes.

La TopoI possède de plus un domaine central liant les cations  $Zn^{2+}$ , de 162 acides aminés, indispensable pour l'activité de relaxation. Malgré ces différences, leur mécanisme d'action est très similaire, et nous prendrons ici l'exemple de la TopoI.

Le mécanisme de relâchement de l'ADN par la TopoI (**Figure 4**) a pu être appréhendé grâce à l'obtention de la structure 3D du fragment N-terminal de 67kDa complexé à l'ADN (Perry et Mondragon, 2003). Tout d'abord, la TopoI se fixe sur l'ADN superenroulé et déforme la double hélice (**Figure 4b**). Un changement conformationnel permet alors le positionnement correct du résidu Tyr319 qui engendre la coupure de l'ADN (**Figure 4c**), et entraîne ainsi la

formation d'une liaison covalente entre l'extrémité 5' du brin d'ADN clivé et la tyrosine du site actif. Cette coupure provoque alors un changement de conformation qui permet l'ouverture de l'enzyme et le passage du second brin d'ADN dans la cavité de l'enzyme (**Figure 4d et e**). Après la fermeture de la cavité de l'enzyme, le brin d'ADN clivé est alors religaturé par la TopoI (**Figure 4f**). Un second cycle d'ouverture et de fermeture de l'enzyme libère ensuite l'hélice d'ADN (**Figure 4g et h**) et l'enzyme retrouve alors sa conformation initiale (**Figure 4i**), soit pour se dissocier de l'ADN, soit pour agir de manière processive. Ce mécanisme a pour effet de diminuer la valeur de Lk par incréments de 1.

La principale différence entre la TopoI et la TopoIII réside en deux domaines présents chez la TopoIII et absents de la TopoI. Cette différence pourrait expliquer à elle seule l'activité de décaténation spécifique de la TopoIII. Le premier domaine, de fonction inconnue, correspond aux résidus 241 à 255. Le second est responsable de l'activité de décaténation, car son remplacement par la séquence homologue de la TopoI élimine totalement cette activité (Champoux, 2001). Cette boucle, riche en résidus basiques, est localisée près de l'espace défini par le tore du domaine N-terminal. Elle pourrait interagir directement avec l'ADN et permettre à la TopoIII d'agir de manière processive sur des multi-caténanes, ce dont est incapable la TopoI.

## **b. Fonctions cellulaires**

### **◇ La TopoI**

Son rôle est de limiter un superenroulement négatif trop fort, ce qui est délétère pour la cellule. Cependant, cette enzyme n'est active que lorsqu'un certain niveau de superhélicité négative est dépassé, pour le réajuster à un degré physiologique optimal pour l'expression des gènes (Zechiedrich et al., 2000). Ceci est réalisé en parallèle de l'activité de la gyrase, qui, elle, introduit des supertours négatifs dans l'ADN. Un équilibre entre les activités de ces deux enzymes est ainsi établi dans la cellule en fonction des conditions et stress rencontrés (voir paragraphe IV.B).

Un deuxième rôle majeur de la TopoI dans la cellule a été récemment mis en évidence (Masse et Drolet, 1999), lié à l'activité de la machinerie transcriptionnelle dans la cellule. Au cours de la transcription, le complexe ARN polymérase-ARN progresse le long de l'ADN, qui

est ancré à des structures fixes. Or, la progression du complexe requiert de déplacer la bulle de transcription. Pour ce faire, le complexe pourrait effectuer des rotations afin de suivre l'hélice d'ADN, mais celui-ci, comportant la polymérase, l'ARN et la machinerie traductionnelle, est bien trop imposant pour pivoter. C'est donc l'ADN qui effectue des rotations et, sans l'intervention de Topos pour résoudre ce problème, il s'ensuivrait une accumulation de tours positifs en amont du complexe de transcription et de supertours négatifs en aval (**Figure 5A** ; Liu et Wang, 1987). La TopoI est impliquée dans l'élimination des supertours négatifs qui s'accumulent en aval de l'ARN polymérase au cours de la transcription, la gyrase ayant pour rôle d'éliminer les supertours positifs apparaissant en amont du complexe de transcription (Liu et Wang, 1987).

Une accumulation trop importante de supertours négatifs en aval de l'ARN polymérase entraînerait l'ouverture de la double hélice d'ADN et pourrait favoriser l'hybridation de l'ARNm nouvellement synthétisé avec le brin d'ADN matrice, ces structures ayant été appelées boucles R (Drolet et al., 2003). L'accumulation de ces boucles R entraînerait soit une inhibition de la transcription en bloquant l'avancement de l'ARN polymérase, soit une inhibition de la synthèse protéique en entraînant la dégradation des ARNs synthétisés et hybridés sur le brin matrice (**Figure 5B** ; Drolet, 2006). Un mutant *topA*, dépourvu d'activité TopoI, révèle une accumulation de structures de type boucles R (Masse et Drolet, 1999). La TopoI est donc essentielle dans la cellule, notamment pour l'élimination des supertours négatifs en aval de l'ARN polymérase, ce qui évite la formation de boucles R. Par ailleurs, il a été démontré que le domaine de fixation du zinc de la TopoI était capable d'interagir avec la sous-unité  $\beta'$  de l'ARN polymérase, ce qui pourrait suggérer que la TopoI soit localisée en aval de l'ARN polymérase au cours de la transcription pour éliminer les boucles R (Cheng et al., 2003). La TopoI remplit donc au moins 2 fonctions cruciales dans les cellules bactériennes : le maintien d'un niveau adéquat de superhélicité de l'ADN nécessaire à tous les processus vitaux de la cellule, et l'élimination des boucles R pour assurer le couplage transcription – traduction.

De par ces activités essentielles pour la cellule, des mutants du gène *topA* pourraient s'avérer non viables. Des analyses génétiques ont effectivement démontré que des mutations de *topA* ne pouvaient être observées qu'en association avec des mutations secondaires compensatrices dans les gènes *gyrA* et *gyrB*, codant pour les deux sous-unités de la gyrase (Pruss et al., 1982 ; Raji et al., 1985). Les mutations dans *gyrA* et *gyrB* diminuent l'activité de

la gyrase, ce qui a pour effet de restaurer un niveau de superhélicité physiologique dans un mutant *topA*. Ces travaux suggéraient donc que le gène *topA* était essentiel chez *E. coli*. En revanche, ce phénomène n'a pas été observé chez *Salmonella typhimurium* et *Shigella flexneri*, deux autres entérobactéries (Richardson et al., 1984 ; Bhriain et Dorman, 1993). Afin d'éviter les mutations secondaires, Stupina et Wang (2005) ont alors construit et analysé des souches d'*E. coli* possédant des allèles mutants du gène *topA* codant des TopoI thermosensibles. Ces gènes *topA* sont soit sous le contrôle de leur propre promoteur, soit sous le contrôle du promoteur *Plac*, donc inductibles par l'IPTG dans le second cas. Les souches mutantes sont viables à 37°C et 42°C, sans nécessité de mutation compensatoire. En revanche, ce n'est plus le cas à des températures plus basses. Par ailleurs, ces auteurs ont démontré que la viabilité des mutants du gène *topA* était strictement dépendante de la présence de la TopoIII, l'autre Topo de type IA présente chez *E. coli* et codée par le gène *topB*. En effet, des doubles mutants *topA topB* ne sont pas viables, quelle que soit la température (Zhu et al., 2001 ; Stupina et Wang, 2005). Deux hypothèses ont été émises pour expliquer l'absence de viabilité des mutants *topA* à basse température. Une basse température pourrait soit favoriser la formation de boucles R, ce qui nécessiterait la présence absolue de la TopoI, soit réduire de façon additionnelle l'activité déjà faible de relâchement de l'ADN de la TopoIII (voir paragraphe suivant). La nécessité de mutations compensatoires dans les gènes *gyrAB* lors des premières études était liée à un biais dans la construction des mutations dans le gène *topA* (Stupina et Wang, 2005).

### ◆ La TopoIII

Les fonctions occupées par la TopoIII ne sont pas encore totalement élucidées. En effet, un mutant pour la TopoIII ne montre pas de phénotype clairement différent chez *E. coli*. De plus, en raison de la présence de codons rares dans le gène *topB* d'*E. coli*, la TopoIII est présente dans la cellule en quantités très limitées (1 à 10 molécules/cellule), alors que les autres Topos sont 500 fois plus abondantes (DiGate et Mariani, 1989).

La TopoIII bactérienne est capable d'effectuer la décaténation des chromosomes-fils au cours de la réplication (Hiasa et al., 1994 ; Nurse et al., 2003). Bien que ce rôle soit déjà assuré par la décaténase majeure d'*E. coli*, la TopoIV (voir paragraphe I.A.2.), Nurse *et al.* (2003) ont montré que la TopoIII était capable de séparer les chromosomes-fils après réplication dans un mutant de la TopoIV. La TopoIII étant une Topo de type I, elle ne peut couper qu'un seul brin

d'ADN. Son activité de décaténation nécessite donc des fragments d'ADN simple brin et se réduit, lors de la réplication, à la période où les fragments d'Okazaki sont encore présents (Nurse et al., 2003).

Les mutants du gène *topB* présentent un phénotype d'hyperrecombinaison RecA-indépendante, ce qui suggère que la TopoIII soit impliquée dans le contrôle de la recombinaison illégitime (Lopez et al., 2005). De par son implication dans les phénomènes de décaténation et de recombinaison, la TopoIII pourrait aussi jouer un rôle dans la ségrégation des chromosomes. En effet, des mutants des gènes *topA* et *topB* présentent des phénotypes de filamentation avec des structures de nucléoïdes anormales. Cependant, ces structures ne sont pas dues à un défaut de décaténation des chromosomes-fils, car une surexpression de la TopoIV, autre enzyme possédant une grande activité de décaténation, ne supprime pas ce phénotype (Zhu et al., 2001). En revanche, une délétion de *recA* résout le problème, ce qui confirme que les Topos de type I pourraient être impliquées dans des phénomènes de recombinaison, probablement au niveau des jonctions de Holliday, qui, en reliant deux molécules d'ADN indépendantes, pourraient générer des tensions lors de l'avancement du complexe de réplication. L'activité de décaténation de la TopoIII permettrait de résoudre de tels problèmes.

## 2. Les Topos de type II

### a. Mécanisme d'action

*E. coli* possède deux Topos de type IIA : la TopoIV, codée par les gènes *parC* et *parE*, a une activité majoritaire de décaténation, mais aussi de relâchement des supertours négatifs comme positifs. La gyrase, codée par les gènes *gyrA* et *gyrB*, est la seule Topo à introduire des supertours négatifs dans l'ADN, et peut aussi décaténer l'ADN. Elles sont toutes deux hétérotétramériques de type A<sub>2</sub>B<sub>2</sub>, contrairement aux Topos de type II eucaryotes, qui sont homodimériques (Champoux, 2001). Les sous-unités GyrB et ParE sont similaires entre elles et contiennent le site de fixation de l'ATP et les domaines transducteur et toprim (**Figure 6**). Les sous-unités GyrA et ParC sont elles aussi similaires et comprennent les éléments 5Y-CAP, dans lesquels se situe le résidu tyrosine catalytique. GyrA et ParC forment aussi un élément

appelé « porte C », ou C-gate (**Figure 6**), ainsi qu'un domaine globulaire à l'extrémité C-terminale spécifique aux Topos de type II bactériennes, qui se lie à la double hélice d'ADN fixée qui sera ensuite transférée (Champoux, 2001 ; Corbett et Berger, 2004).

Le mécanisme d'action de la gyrase est le suivant : l'ADN qui sera clivé (double hélice G pour Gate) vient d'abord se fixer aux éléments catalytiques toprim et 5Y-CAP (**Figure 6 et 7.1**). L'enroulement d'environ 140 pb d'ADN autour des domaines C-terminaux (CTD) des deux sous-unités GyrA (**Figure 7.2** ; Orphanides et Maxwell, 1994 ; Corbett et al., 2004) suivi d'un changement de conformation au niveau des domaines GHKL et transducteurs, formant le domaine ATPase, permet de capturer une seconde double hélice (T pour Transfer) dans une configuration de supertour positif (**Figure 7.2**, vue de dessus et **Figure 7.3**). Ceci nécessite deux molécules d'ATP et entraîne la coupure de l'hélice G à travers laquelle est alors transférée l'hélice T (**Figure 7.4**). Ce transfert a pour effet d'inverser la configuration de supertour positif, et crée ainsi deux supertours négatifs (Brown et Cozzarelli, 1979 ; Kampranis et al., 1999). Cette inversion est permise par un autre changement de configuration, un abaissement d'un des domaines CTD des sous-unités GyrA qui sont reliés aux domaines N-terminaux par une région flexible (**Figure 8** ; Schoeffler et Berger, 2005). L'hélice G est ensuite religaturée puis la double hélice T est probablement libérée par l'ouverture des portes C (**Figure 7.5**). Si les doubles hélices G et T font partie de la même molécule d'ADN, le niveau de superhélicité Lk de celle-ci sera modifié par incréments de 2. En revanche, si les deux doubles hélices correspondent à deux molécules d'ADN différentes, l'activité ainsi générée sera une caténation ou une décaténation.

Le mécanisme d'action de la TopoIV, qui, contrairement à la gyrase, relâche l'ADN, est similaire à celui de la gyrase à deux exceptions près. Tout d'abord, il n'y a pas d'enroulement de l'ADN, qui est dû, pour la gyrase, à la présence au niveau du CTD de GyrA d'une « boîte GyrA » de 7 acides aminés, absente chez la TopoIV (**Figure 7.2** ; Kramlinger et Hiasa, 2006). Deuxièmement, la flexibilité du CTD est propre à GyrA et permet l'ajout de supertours négatifs à l'ADN après abaissement d'un des deux domaines (**Figure 8B**). La TopoIV, ne possédant ni cette activité d'enroulement ni la possibilité d'abaissement des CTD de ParC, ne peut introduire de supertours négatifs à l'ADN.

## b. Fonctions cellulaires

### ◇ La gyrase

La gyrase étant la seule enzyme capable d'introduire des supertours négatifs dans l'ADN, sa fonction est essentielle (Zechiedrich et al., 2000). Elle est nécessaire à l'initiation de la réplication de l'ADN et de la transcription, car l'augmentation de superhélicité négative permet d'abaisser la quantité d'énergie requise pour l'ouverture de la double hélice (Espeli et Mariani, 2004 ; Zechiedrich et al., 2000). De plus, elle convertit en supertours négatifs les supertours positifs qui sont générés en amont des complexes de réplication et de transcription (Liu et Wang, 1987), ce qui est essentiel à la progression de la fourche de réplication et de l'ARN polymérase. La gyrase est donc une enzyme essentielle et les mutations dans les gènes *gyrAB* sont conditionnelles et létales, l'incubation à la température restrictive provoquant un arrêt de la réplication de l'ADN, et donc la mort cellulaire (Kreuzer et Cozzarelli, 1979). Certains de ces mutants entraînent une augmentation de la recombinaison illégitime (Ashizawa et al., 1999), ce qui implique la gyrase dans les phénomènes de recombinaison. Le modèle d'échange de sous-unités de gyrase a été proposé, selon lequel la gyrase se fixe et clive l'ADN, avec chaque sous-unité GyrB liée de façon covalente à l'extrémité 5' de l'ADN (Ikeda et al., 1982). Un échange de sous-unités entre plusieurs molécules de gyrase fixées à des sites différents de l'ADN aboutirait à un échange de brins d'ADN et donc à un événement de recombinaison. Des sites de reconnaissance de la gyrase sont présents sur les chromosomes bactériens, sur les plasmides ou l'ADN des bactériophages. Dans le cas de l'ADN du bactériophage Mu, des sites SGS (Strong Gyrase Site) sont présents, et la fixation de la gyrase sur ces sites permet l'alignement des extrémités de Mu, et donc sa transposition (Sokolsky et Baker, 2003).

### ◇ La topoisomérase IV

La TopoIV est principalement responsable de la décaténation de l'ADN, notamment au cours de la réplication (Hiasa et Mariani, 1996), mais aussi lors des processus de recombinaison site-spécifique (Zechiedrich et al., 1997). Elle est également impliquée dans la ségrégation des chromosomes pendant la réplication.

La TopoIV joue aussi un rôle dans le contrôle du niveau global de superhélicité de l'ADN (Zechiedrich et al., 2000). Ce rôle a longtemps été occulté car les inhibiteurs (coumarines et quinolones) utilisés contre la gyrase inhibent aussi la TopoIV. Zechiedrich *et al.* (2000) ont montré que l'absence simultanée des activités TopoI et TopoIV entraînait un niveau de superhélicité plus élevé que l'absence de la TopoI seulement. De plus, bien que la TopoI ait une activité de relâchement plus importante que la TopoIV, seule cette dernière permet un relâchement complet d'une molécule plasmidique circulaire.

Comme dans le cas de la gyrase, des mutations dans les gènes *parC* et *parE* sont létales conditionnelles. Les cultures de ces mutants à la température restrictive résultent en l'absence de ségrégation des chromosomes qui s'accumulent donc au centre de cellules élonguées (Adams et al., 1992).

## **B. Les protéines de type histone**

Mises à part les topoisomérases, des protéines de type histone, aussi appelées NAP (Protéines Associées au Nucléoïde) (Dorman et Deighan, 2003), interviennent dans la structure et l'état de superenroulement du chromosome (Luijsterburg et al., 2006). Chez *E. coli*, il y en a 5 majeures : Dps, Fis, H-NS, HU et IHF. Elles assurent la compaction et la condensation du chromosome en une structure appelée nucléoïde. Elles peuvent remplir différentes fonctions dans la cellule, mais leur point commun est de lier l'ADN et de modifier sa superhélicité de façon globale ou locale, en introduisant une topologie contrainte. Certaines de ces protéines peuvent courber l'ADN, comme IHF, HU ou Fis, alors que d'autres établissent des ponts entre molécules d'ADN, comme H-NS ou Dps.

Les profils de fixation à l'ADN ont été réalisés pour 3 protéines de type histone (Fis, H-NS et IHF) par immunoprécipitation de la chromatine associée à des puces à ADN (Grainger et al., 2006). Des centaines de sites de fixation ont été mis en évidence sur le génome d'*E. coli*. Environ la moitié de ces sites sont localisés dans des régions intergéniques, qui représentent moins de 10% du génome d'*E. coli*, ce qui pourrait indiquer une préférence pour ces régions, et donc un rôle de ces protéines dans l'organisation générale du chromosome (voir paragraphe II.A.4). Une préférence de fixation pour les régions riches en AT, qui peuvent présenter une courbure intrinsèque, a également été observée (Grainger et al., 2006),

ainsi qu'une association avec les régions promotrices. De plus, des interactions ont été récemment mises en évidence entre certaines NAPs et des sous-unités de l'ARN polymérase et des ribosomes (Butland et al., 2005). Ainsi, en plus de leur rôle dans le contrôle de la superhélicité de l'ADN, certaines NAPs sont également connues pour être des régulateurs de la transcription d'un grand nombre de gènes (Dorman et Deighan, 2003). Elles peuvent en outre affecter le superenroulement de l'ADN par la régulation des gènes de topoisomérases. Comme nous le verrons au paragraphe III, un lien étroit existe donc entre le niveau de superhélicité du chromosome et la transcription globale des gènes.

Enfin, les NAPs sont exprimées différemment au cours du cycle cellulaire. De par leur rôle dans le contrôle de la superhélicité de l'ADN et sur la transcription, les NAPs vont entraîner une dynamique importante du nucléoïde bactérien et de l'expression globale des gènes. Les propriétés et fonctions des 5 NAPs majeures sont décrites dans les paragraphes suivants.

## 1. Les NAPs ne possédant aucune spécificité de fixation

### ◇ **H-NS** (Histone-like Nucleoid Structuring protein)

La protéine de type histone H-NS se fixe de manière non spécifique à l'ADN, bien qu'une préférence pour des séquences riches en AT présentant une courbure ait été suggérée (Dame et al., 2001). Après fixation, elle provoque la condensation de l'ADN chromosomique (Spassky et al., 1984) et contraint des supertours négatifs *in vitro* (Tupper et al., 1994). En effet, des mutants dépourvus de H-NS présentent une diminution de la superhélicité globale (Owen-Hughes et al., 1992).

La protéine H-NS est composée de deux domaines de fonctionnalités différentes, séparés par une partie flexible (**Figure 9**). Le domaine C-terminal est impliqué dans la liaison à l'ADN, alors que le domaine N-terminal permet l'oligomérisation de la protéine (Dame, 2005). H-NS se fixerait à l'ADN sous forme de dimères, mais il a été montré que des oligomères pouvaient se former *in vitro*, qui pourraient servir de squelette pour compacter l'ADN (Dorman, 2004). En effet, les dimères s'oligomériseraient après fixation à l'ADN (Badaut et al., 2002), et une structure du type « fermeture éclair » pourrait se former, aboutissant à la condensation de l'ADN (**Figure 9**). Ces propriétés de H-NS sont à l'origine d'une activité de répression de

l'expression des gènes. En effet, des études génétiques et moléculaires ont montré qu'H-NS était un régulateur global de l'expression des gènes, affectant environ 5% des gènes d'*E. coli* (Hommais et al., 2001), la plupart étant réprimés. Le mécanisme de cette répression a été explicité dans le cas du promoteur P1 des opérons d'ARN ribosomiques (**Figure 10** ; Dame et al., 2002). La caractéristique des promoteurs réprimés par H-NS serait d'être flanqués de régions possédant un fort potentiel de courbure. La fixation de l'ARN polymérase sur ces promoteurs (**Figure 10i**) entraînerait le rapprochement des régions flanquantes intrinsèquement courbées (**Figure 10ii**). H-NS présentant une préférence de fixation sur les séquences d'ADN courbées, plusieurs dimères d'H-NS pourraient alors se fixer sur l'ADN, piégeant ainsi l'ARN polymérase (**Figure 10iii**). Ces complexes ne pourraient se séparer qu'après l'intervention d'activateurs de la transcription ou dans certaines conditions (**Figure 10iv**), comme une élévation de température ou d'osmolarité, qui sont par ailleurs connues pour modifier la superhélicité de l'ADN (voir paragraphe IV.A). Ainsi, le degré de superhélicité de l'ADN va affecter la transcription des gènes réprimés par H-NS, ce qui souligne encore le lien entre superhélicité de l'ADN, protéines de type histone et transcription globale des gènes.

Un certain nombre de gènes peuvent également être activés par H-NS (Johansson et al., 2000). Il s'agit essentiellement des gènes régulés négativement par (p)ppGpp, qui est l'effecteur de la réponse stringente, permettant l'adaptation des bactéries aux périodes de carences nutritionnelles (Cashel et al., 1996). Le mécanisme par lequel H-NS active ces gènes serait lié à sa capacité à augmenter le superenroulement de l'ADN (Johansson et al., 2000).

Au cours de la croissance chez *E. coli*, la quantité de H-NS diminue, passant de 20 000 molécules en phase exponentielle à 8000 en phase stationnaire (Talukder et al., 1999). Son expression est fortement activée en cas de baisse de la température (Free et Dorman, 1995).

Un paralogue de H-NS est également présent dans les bactéries gram-négatives. Cette protéine, appelée StpA chez *E. coli*, présente 58% d'identité de séquence avec H-NS (Free et al., 2001). Elle partage certaines fonctions de régulation et de maintien de la superhélicité avec H-NS (Zhang et al., 1996). Ces deux protéines peuvent former des hétérodimères *in vivo*, ce qui pourrait altérer leurs fonctions respectives (Cusick et Belfort, 1998).

◇ HU (Histone-like protein from *E. coli* strain U93)

HU a été initialement isolée chez *E. coli* comme étant capable de former des structures de type nucléosome (Rouviere-Yaniv et Gros, 1975), sans reconnaître de séquence particulière de l'ADN. En revanche, HU semble se fixer préférentiellement sur l'ADN superenroulé négativement (Shindo et al., 1992). L'étude des effets de HU sur la structure de l'ADN ont conduit à des résultats controversés. HU se lie à l'ADN et y induit une courbure d'environ 160° (**Figure 11B**), ce qui provoque la compaction de molécules d'ADN *in vitro* (Endo et al., 2002). De plus, HU est attirée par des structures spécifiques d'ADN, comme des jonctions en croix ou des coupures simple brin, ce type de liaison à l'ADN lui donnant un rôle dans les processus de recombinaison (Rowland et al., 2006). *In vivo*, HU semble également compacter l'ADN, car des souches dépourvues de cette protéine sélectionnent des mutations secondaires dans *gyrB* qui augmentent la capacité de superenroulement de la gyrase. Ceci a été interprété comme un phénomène de compensation pour pallier une perte de superhélicité engendrée par l'absence de HU (Malik et al., 1996). Cependant, à fortes concentrations, HU donne un aspect circulaire relâché à l'ADN (Dame et Goosen, 2002). HU pourrait alors avoir un effet sur le pas de l'hélice d'ADN, qui serait constitué de 10,9 pb par tour d'hélice au lieu de 10,5 pb. Cet accroissement se ferait par la contrainte de supertours négatifs supplémentaires (Dame et Goosen, 2002) et entraînerait une rigidification de la double hélice. A plus basses concentrations, HU a un effet de courbure sur l'ADN, et l'aspect circulaire est alors moins rigide (**Figure 11B** ; van Noort et al., 2004).

En plus de son rôle structural, HU est impliquée dans l'expression des gènes. De profonds changements de la transcription globale ont été détectés chez les mutants HU (Kar et al., 2005), ainsi que des modifications morphologiques des cellules mutantes. Par son rôle structural sur l'ADN, HU est par ailleurs connue pour assister certains régulateurs au niveau de leur promoteur. Par exemple, HU permet l'interaction entre des dimères du répresseur GalR pour conduire à la répression des gènes *gal* (Kar et Adhya, 2001).

La protéine HU est présente en quantités abondantes dans la cellule, de 30 000 à 55 000 molécules en phase exponentielle (Taluđer et al., 1999). Lors du passage en phase stationnaire, le nombre de molécules de HU diminue pour passer à moins de 20 000 molécules. Chez la plupart des bactéries, HU est un homodimère, alors qu'elle existe sous forme d'hétérodimère

chez *E. coli*. Les deux sous-unités sont codées par les gènes *hupA* et *hupB*. Des formes homodimériques ont également été mises en évidence chez *E. coli* (Ramstein et al., 2003). Leur fonction n'est pas connue, mais pourrait influencer sur les propriétés des dimères de HU.

Enfin, les deux NAPs HU et H-NS exercent des effets antagonistes dans la cellule (**Figure 12**). Alors que H-NS compacte et condense l'ADN, HU lui donne une forme circulaire rigide. De plus, la capacité de H-NS à réprimer la transcription d'un certain nombre de gènes est accrue dans les mutants dépourvus de HU, cette dernière étant capable de lever la répression de ces gènes (Dame et Goosen, 2002).

#### ◇ Dps (DNA-binding Protection from Starved cells)

La protéine Dps se lie de façon non spécifique à l'ADN, et sa principale fonction est de le protéger pendant les périodes de stress, dont la phase stationnaire (Frenkiel-Krispin et al., 2004). Alors que seulement 6000 molécules de Dps sont présentes chez *E. coli* en phase exponentielle, près de 200 000 molécules ont été dénombrées en phase stationnaire, ce qui en fait la protéine la plus abondante de la cellule (Talukder et al., 1999).

Sur des bases structurales, Dps est considérée comme un membre des bactéroferritines, chaque monomère formant une structure entrelacée de 4 hélices. Dps est formée de 12 de ces monomères formant une sphère creuse (**Figure 13**), ces sphères s'assemblant ensuite en nids d'abeilles (Frenkiel-Krispin et al., 2004). La surface ainsi formée ne présente pas de domaine spécifique de fixation à l'ADN, mais des résidus lysines, chargés positivement et présents dans les pores situés entre les sphères, pourraient interagir avec l'ADN (Grant et al., 1998). La liaison à l'ADN pourrait induire des interactions entre dodécamères de Dps, conduisant à la formation de couches superposées de Dps et d'ADN (Wolf et al., 1999). La formation de ces structures très organisées permettrait une forte compaction de l'ADN. Les effets de Dps sur la structure de l'ADN ont été visualisés par microscopie. Après 24h de carence nutritionnelle, des structures toroïdale Dps-ADN sont formées ; après une phase stationnaire prolongée de 48h, ces structures laissent place à des cristaux où les couches de Dps et d'ADN sont alternées (**Figure 13** ; Frenkiel-Krispin et al., 2004).

Les structures formées par Dps permettent de protéger l'ADN des stress oxydatifs, thermiques, et des radiations UV (Minsky et al., 2002 ; Frenkiel-Krispin et al., 2004). Dps fait

d'ailleurs partie du régulon de OxyR, celui de la réponse aux stress oxydatifs (Altuvia et al., 1994). Pendant de tels stress, cette protection stérique serait couplée à une protection chimique, grâce à la liaison des ions  $Fe^{2+}$  à Dps. En effet, les stress oxydatifs sont associés à la production de radicaux libres, toxiques pour les cellules, via les réactions de Fenton impliquant les ions  $Fe^{2+}$ . Ces radicaux peuvent notamment conduire à des cassures simple brin ou double brin de l'ADN par oxydation des bases nucléotidiques. Dps est capable de séquestrer les ions  $Fe^{2+}$  et de les oxyder par l'utilisation, et donc l'élimination, de  $H_2O_2$ , sans relarguer de radicaux libres, préservant ainsi l'intégrité de l'ADN (Zhao et al., 2002).

## 2. La protéine IHF

La protéine IHF (Integration Host Factor) est la NAP présentant la plus forte spécificité de fixation à l'ADN avec une séquence consensus clairement définie, qui ressort de l'analyse de plusieurs sites de fixation (Dorman et Deighan, 2003). IHF est un hétérodimère dont les sous-unités  $\alpha$  et  $\beta$  sont codées par les gènes *himA* et *himB* et dont la séquence et la structure ressemblent à celles de HU. IHF se fixe sur une séquence de 30 pb, dont 13 sont conservées, et y induit une courbure allant jusqu'à  $180^\circ$  (**Figure 11A** ; Dame et al., 2005). Seuls 1000 sites de fixation à haute affinité pour IHF sont présents sur le chromosome (Ussery et al., 2001), alors que 12 000 molécules d'IHF sont présentes chez *E. coli* en phase exponentielle, et 55 000 en début de phase stationnaire (Talukder et al., 1999). Un fort excès d'IHF est donc présent par rapport à ses sites de haute affinité. IHF pourrait donc également se fixer à l'ADN par des interactions non spécifiques, car la majorité d'IHF *in vivo* est liée à l'ADN (Yang et Nash, 1995). La fixation d'IHF à l'ADN *in vitro* provoque une réduction de la longueur de l'ADN de 30%, liée à l'activité de courbure d'IHF et conduisant à la compaction de l'ADN (Ali et al., 2001). L'expression maximale d'IHF lors du passage en phase stationnaire permet d'envisager la compaction du chromosome par IHF à ce moment du cycle cellulaire, avant que Dps ne prenne le relais (Luijsterburg et al., 2006).

En plus de son rôle structural d'organisation de l'ADN, IHF est impliquée dans les phénomènes de régulation de la transcription chez les bactéries (Goosen et van de Putte, 1995), la réplication de l'ADN (Ryan et al., 2002), la recombinaison site-spécifique (Dorman et

Higgins, 1987), et la transposition (Crellin et al., 2004). Son activité de courbure est cruciale pour ces fonctions additionnelles, car elle induit des interactions à grande échelle entre différentes protéines (Goosen et van de Putte, 1995). D'autre part, l'influence d'IHF sur la structure de l'ADN contribue notamment au contrôle de la transcription. Ainsi, la liaison d'IHF sur un de ses sites de fixation en amont du promoteur *ilvP<sub>G</sub>* améliore l'ouverture de la bulle de transcription en transférant la tension induite par la courbure de l'ADN de sites riches en AT en amont du promoteur vers celui-ci (Parekh et al., 1996 ; Sheridan et al., 1998).

Des profils globaux de transcription ont été réalisés chez *S. typhimurium* afin de comparer la souche de référence et des mutants dépourvus d'IHF (Mangan et al., 2006). De profonds changements de la transcription ont été observés suite à l'absence d'IHF, les gènes affectés étant notamment impliqués dans la transition entre les phases exponentielle et stationnaire, mais aussi dans la virulence et l'invasion des cellules épithéliales, ce qui montre le rôle d'IHF dans la régulation coordonnée des traits pathogéniques et de l'adaptation aux carences (Mangan et al., 2006).

### **3. La protéine Fis**

La protéine Fis (Factor for Inversion Stimulation), un homodimère de 98 acides aminés, fut découverte dans les années 1980 chez *E. coli* et *S. typhimurium* pour son implication dans les phénomènes de recombinaison site-spécifique (Johnson et al., 1986 ; Koch et Kahmann, 1986 ; Thompson et al., 1987 ; Ross et al., 1990). Depuis, un grand nombre de fonctions lui ont été attribuées dans la cellule, et elles ont pu être corrélées à sa structure, connue depuis 1991 (Kostrewa et al., 1991).

#### **a. Rôle de Fis dans la superhélicité de l'ADN**

Des sites à haute affinité pour Fis ont été mis en évidence, ce qui a permis de déduire une séquence consensus de fixation de Fis sur l'ADN, de 15pb (Hengen et al., 1997). Cette séquence est cependant très dégénérée. De plus, le nombre de sites à haute affinité répertoriés sur le chromosome d'*E. coli*, estimé à 6000 (Ussery et al., 2001), est bien inférieur au nombre de molécules de Fis dans la cellule en phase exponentielle (plus de 50 000 molécules). Ainsi,

Fis est également capable de se fixer de manière non spécifique à l'ADN (Pan et al., 1996), avec une préférence pour des régions courbées. La fixation de Fis sur l'ADN provoque une courbure de 50 à 90°, ainsi qu'un phénomène d'extrusion de boucles de l'ADN (**Figure 11C** ; Schneider et al., 2001). La capacité de Fis à se fixer de manière à la fois non spécifique et sur des sites à haute affinité lui permet d'influer sur la topologie de l'ADN aussi bien au niveau global que local. En effet, de par sa capacité à se fixer de façon non spécifique et à former des « branches » et des boucles d'ADN, Fis est impliquée dans la structuration globale du chromosome. Ainsi, Fis forme des barrières qui contribuent au maintien de domaines topologiques, et son absence engendre un relâchement général du chromosome (voir paragraphe II.A.4 ; Hardy, Cozzarelli, 2005).

Au niveau local, Fis se fixe sur des sites de reconnaissance à haute spécificité, qui sont souvent localisés dans les régions promotrices de certains gènes, ce qui fait de Fis un régulateur global de la transcription (Dorman et Deighan, 2003). Ces sites de fixation sont appelés Upstream Activating Sequence (UAS) et ont été particulièrement bien étudiés dans le cas des opérons d'ARN ribosomiques, qui sont activés par Fis. Leur région régulatrice possède en effet plusieurs sites de fixation de Fis qui sont séparés par une distance correspondant au pas de l'hélice, le site majeur étant localisé à la position -71 par rapport au site d'initiation de la transcription (Newlands et al., 1992 ; Hirvonen et al., 2001). Cette structuration des UAS leur permet d'être tous localisés du même côté de la double hélice d'ADN. La fixation de Fis sur ces UAS entraîne des courbures de l'ADN et la formation de boucles d'ADN (Travers et Muskhelishvili, 1998 ; Schneider et al., 2001), ces boucles favorisant la fixation de l'ARN polymérase (Heggeler-Bordier et al., 1992). Elles pourraient donc favoriser la transcription des opérons d'ARNr par l'intermédiaire de Fis. Ainsi, les deux effets de Fis sur l'ADN, aux niveaux global et local, impliquent le même mécanisme structural : une forte courbure de l'ADN avec pour conséquence la formation de boucles (**Figure 14** ; Muskhelishvili et Travers, 2003). Un autre exemple est constitué par la région promotrice de *tyrT*, qui contient trois sites où Fis se fixe de manière coopérative. L'effet de cette fixation est de stimuler l'ouverture du promoteur, probablement à l'aide de la torsion introduite par les trois dimères de Fis (Auner et al., 2003). L'effet local de Fis sur les promoteurs possédant des sites de fixation à haute affinité serait de moduler, en l'augmentant, le superenroulement local de l'ADN de ces promoteurs. De cette façon, Fis permettrait à ces promoteurs de rester actifs même si la superhélicité globale de

l'ADN devenait défavorable à la transcription de ces promoteurs (Muskhelishvili et Travers, 2003). Par exemple, en augmentant localement la superhélicité de l'ADN en amont des opérons d'ARNr, Fis favorise leur transcription qui nécessite une superhélicité forte (Zacharias et al., 1992), et ce même si la superhélicité globale du chromosome venait à diminuer. Ce n'est que lorsque la concentration de Fis diminue, lors du passage en phase stationnaire, que la transcription de ces opérons va diminuer.

Enfin, Fis agit également sur le niveau de superhélicité de l'ADN de façon indirecte, en régulant les gènes codant les topoisomérases. En effet, Fis réprime l'expression des gènes codant la gyrase (Schneider et al., 1999) et active *topA*, qui code la TopoI, lors de stress oxydatifs (Weinstein-Fischer et al., 2000). Fis est ainsi impliquée dans le contrôle homéostatique du degré de superhélicité de l'ADN (voir paragraphe IV.B.). En effet, la transcription de *fis* est sensible au niveau de superhélicité.

Tous ces mécanismes font de Fis une protéine centrale du contrôle de la superhélicité, agissant comme senseur des conditions de l'environnement, organisateur de la structure du chromosome et enfin régulateur des topoisomérases (Schneider et al., 2000b).

## b. Structure de Fis

La structure d'un monomère de Fis révèle 4 hélices  $\alpha$  entremêlées A à D (**Figure 15A**). Ces hélices s'étendent des résidus 27 à 42 (A), 50 à 70 (B), 74 à 81 (C) et 85 à 94 (D). La structure formée par les 25 premiers acides aminés a longtemps été inconnue, et était donc supposée non structurée. Un mutant a été isolé, dont la structure a révélé une organisation en épingle à cheveux formée par deux feuillets  $\beta$  et rattachés au corps de la protéine par des interactions hydrophobes et des liaisons hydrogène (Safo et al., 1997).

Les hélices A et B sont localisées dans la partie N-terminale de la protéine et sont impliquées dans des interactions entre monomères pour construire une structure compacte dont la taille est d'environ 25×35×50 Å (Yuan et al., 1991). L'hélice A d'un monomère n'a aucun contact avec les hélices C' et D' de l'autre monomère, mais elle établit de nombreuses liaisons hydrogène et de Van der Waals avec l'hélice B' de l'autre monomère. L'hélice B est courbée en son milieu en direction de l'hélice C' avec laquelle elle forme une hélice virtuelle plus longue par l'intermédiaire de liaisons hydrogène (**Figure 15A**).

Les hélices C et D sont situées dans la région C-terminale de la protéine et forment un motif hélice-tour-hélice de liaison à l'ADN (Yuan et al., 1991). Les protéines contenant un tel motif s'insèrent généralement dans le grand sillon de l'ADN. L'orientation des groupements phosphate de l'ADN responsables des interactions avec Fis est telle que la distance entre deux grands sillons varie de 29 à 34Å. Bien que les deux hélices D d'un dimère de Fis soient orientées correctement pour s'insérer dans les grands sillons, elles ne sont espacées que de 25Å. La fixation de Fis résulte ainsi en une courbure de l'ADN (**Figure 11C** et **15B**) d'environ 50 à 90° (Pan et al., 1996).

### c. La régulation de Fis

Le gène *fis* est le deuxième gène d'un opéron comportant également *dusB*, codant une dihydrouridine synthase (Bishop et al., 2002). L'opéron *dusB-fis* est sous le contrôle d'un promoteur unique, *fisP* (Ball et al., 1992 ; Walker et al., 1999 ; Mallik et al., 2004).

Le nombre de molécules de Fis varie de façon importante au cours du cycle cellulaire. Il est maximal lors de l'entrée en phase exponentielle, environ 1h après dilution dans du milieu frais (plus de 50 000 molécules), puis diminue jusqu'à une quasi-disparition de Fis en phase stationnaire (moins de 100 molécules).

#### ◇ Régulation transcriptionnelle

Le promoteur *fisP* est reconnu par le facteur sigma végétatif  $\sigma^{70}$  (**Figure 16**), même si la boîte -35 est plus éloignée de la séquence consensus. Quatre caractéristiques de ce promoteur sont à l'origine d'une forte transcription en phase exponentielle, suivie d'un arrêt en phase stationnaire.

Premièrement, ce promoteur se caractérise par la présence d'une séquence riche en GC entre la boîte -10 et le site d'initiation de la transcription, appelée discriminateur (**Figure 16** ; Walker et al., 1999). Cette séquence crée une barrière énergétique pour l'ouverture de la double hélice lors de l'initiation de la transcription. Elle constitue la signature de gènes dont la transcription nécessite une superhélicité négative de l'ADN élevée. En effet, *fis* est activée par une superhélicité forte, caractéristique de la phase exponentielle de croissance (voir paragraphe IV.A.). Ceci est à la base du contrôle homéostatique de la superhélicité de l'ADN par Fis.

Deuxièmement, deux sites d'initiation de la transcription sont observés, dont le majeur est situé 8 pb après la boîte -10 (**Figure 16**) et fait intervenir un CTP comme nucléotide initiateur. Le second est situé 2 pb en amont et implique un GTP, mais ne participe que très faiblement à l'initiation de la transcription. Le fait d'initier la transcription par un CTP est un phénomène relativement rare chez *E. coli*, du fait de sa faible abondance (Walker et al., 2004), et de la faible interaction qu'il établit avec l'ARN polymérase (Wu et Goldthwait, 1969). Ceci en fait ainsi un remarquable mécanisme de régulation de la transcription pour des gènes qui sont transcrits rapidement puis réprimés tout aussi rapidement, ce qui est le cas de *fis*. En effet, à l'entrée en phase exponentielle de croissance, la concentration en CTP augmente, ce qui favorise une activation de la transcription de *fis* par la stabilisation du complexe ouvert (Walker et al., 2004). En milieu de phase exponentielle, la quantité de CTP disponible diminue fortement, et la transcription de *fis* aussi et, bien que cette diminution soit due à différents facteurs, elle suit la concentration en CTP de manière remarquable (Walker et al., 2004). Le remplacement de ce C par un A entraîne une prolongation de l'expression de *fis* jusqu'en phase stationnaire (Walker et al., 1999).

Troisièmement, le promoteur de *fis* est soumis à une autorépression. En effet, Fis se fixe sur son propre promoteur, sur 6 sites répartis des résidus -231 à +30 (Finkel et Johnson, 1992 ; Ball et al., 1992). Fis présente une grande affinité pour les sites I et II, mais c'est surtout le site II qui participe à la répression (Pratt et al., 1997). En effet, il chevauche la région -35 du promoteur et, avec le site I, se superpose au site de fixation de l'ARN polymérase. L'hypothèse la plus probable est que la fixation de Fis empêche l'ARN polymérase d'initier la transcription.

Quatrièmement, le promoteur de *fis* est soumis au contrôle stringente, c'est-à-dire à la répression par (p)ppGpp, l'effecteur de la réponse stringente permettant la réponse aux carences nutritionnelles (Cashel et al., 1996). Les nucléotides (p)ppGpp sont synthétisés lors de carences nutritionnelles. Ils répriment alors les gènes nécessaires à la croissance et activent certains gènes nécessaires à la survie. Une des caractéristiques des promoteurs régulés négativement par la réponse stringente est la présence du discriminateur riche en GC entre la région -10 et le +1 de transcription. Le promoteur de *fis* est ainsi soumis à la répression par (p)ppGpp (Ninnemann et al., 1992). Le contrôle stringente ainsi exercé sur *fisP* participe donc également à la chute de transcrits observée avant le passage en phase stationnaire.

A côté des différents niveaux de régulation transcriptionnelle précédents, un autre régulateur est impliqué dans l'activation de l'opéron *dusB-fis*: IHF. Un site de fixation d'IHF est centré 114 pb en amont du +1 de transcription du gène *fis* (Pratt et al., 1997). La fixation d'IHF sur ce site entraîne une activation de la transcription de 3 ou 4 fois.

#### ◇ Régulation traductionnelle

La régulation de l'expression de *fis* s'exerce également au niveau de la traduction par une protéine à activité GTPase, appelée BipA, et qui présente des homologies avec des facteurs d'élongation de type EF-G (Owens et al., 2004). Son profil d'expression est similaire à celui de Fis : de très abondante en début de phase exponentielle, BipA devient quasiment indétectable après 3h de culture. BipA interagit avec les ribosomes et son rôle est de favoriser l'initiation de la traduction de l'ARNm de *fis*. En effet, des interactions particulièrement fortes existent entre la région 5' non traduite de l'ARNm de *fis* et les ribosomes, et BipA permettrait de déstabiliser ces interactions pour débiter l'initiation de la traduction par les ribosomes (**Figure 17** ; Owens et al., 2004). Les auteurs émettent deux hypothèses principales quant au rôle physiologique de BipA : la première est que l'interaction BipA-ribosomes permettrait de recruter ceux-ci pour traduire *fis*, ce qui est indispensable en phase exponentielle, c'est-à-dire au moment où la traduction de nombreux ARNs est en cours. Ceci permettrait ainsi une synthèse préférentielle et rapide de Fis en début de phase exponentielle et une synthèse continue pendant la phase exponentielle. La seconde hypothèse est que l'activité GTPase de BipA aurait un rôle de senseur de la concentration en GTP, car elle possède une activité GTPase, et ne permettrait la traduction de *fis* que lorsque la concentration est optimale, la traduction étant un phénomène particulièrement consommateur de GTP.

Un second niveau de régulation de la traduction de Fis est réalisé par l'intermédiaire des polyamines (putrescine, spermidine et spermine), la quantité de Fis augmentant en leur présence (Yoshida et al., 2004). Ces petites molécules interagissent avec les ARNs messagers et affectent ainsi la traduction des protéines (Miyamoto et al., 1993). En déstabilisant les ARNm doubles brins, les polyamines rendent libre la séquence de Shine-Dalgarno, sur laquelle la sous-unité 30S du ribosome peut ensuite venir se fixer (Yoshida et al., 2004).

#### d. Autres rôles de Fis dans la cellule

En plus de son rôle dans le maintien d'un niveau optimal de superhélicité de l'ADN, Fis intervient dans de nombreux processus cellulaires fondamentaux. Fis est en effet un régulateur de la transcription de nombreux gènes, mais est également impliquée dans la réplication du chromosome et dans des phénomènes de recombinaison et d'inversion site-spécifique.

##### ◇ Régulateur global de l'expression des gènes

Fis est connue pour réguler la transcription d'un grand nombre de gènes. Parmi ceux-ci, Fis active la transcription des promoteurs d'ARN ribosomiques et d'ARN de transfert (Bosch et al., 1990 ; Ross et al., 1990 ; Lazarus et Travers, 1993). Fis interagit avec 3 sites de fixation situés en amont du promoteur P1 de *rrnB*, mais c'est la fixation sur le site I qui est la plus importante pour l'activation du promoteur (**Figure 18**). En plus de modifications structurales de l'ADN (voir plus haut), la fixation d'un dimère de Fis sur ce site permet son interaction avec la sous-unité  $\alpha$  de l'ARN polymérase (**Figure 19**), par l'intermédiaire des résidus situés à des positions comprises entre 68 et 74 d'un des monomères de Fis, c'est-à-dire situés à la connexion entre les hélices B et C (**Figure 15** ; Bokal et al., 1997 ; Aiyar et al., 2002). La présence de 3 sites de fixation de Fis consécutifs pourrait induire une plus grande torsion de l'ADN et ainsi un plus grand contact avec l'ARN polymérase (Muskhelishvili et al., 1995).

Fis est également un régulateur des gènes de topoisomérases, en inhibant la transcription des gènes *gyrA* et *gyrB* codant les sous-unités de la gyrase (Schneider et al., 1999) et en activant le gène *topA* codant la TopoI (Weinstein-Fischer et al., 2000). Fis se lie au promoteur de *gyrA*, et la transcription est inhibée directement, Fis empêchant la fixation de l'ARN polymérase sur le promoteur. Pour *gyrB*, le mécanisme est moins clair : l'ARN polymérase se fixe sur le promoteur même en présence de Fis, mais une probable stabilisation de la structure empêcherait le début de la transcription (Schneider et al., 1999).

Fis intervient également dans le contrôle de différentes voies métaboliques en régulant, de façon indirecte cette fois, l'expression de gènes impliqués dans le catabolisme de sources de carbone ou d'acides nucléiques, comme *glpK*, *nt1D*, *rbsB*, *cdd* et *udp* (Gonzalez-Gil et al., 1996). La recherche de gènes régulés par Fis a permis à Xu et Johnson (Xu et Johnson, 1995b) d'en identifier 13 autres, dont 5 seulement ont une fonction connue. Trois d'entre eux codent

pour des protéines nécessaires à la croissance en milieu pauvre : des transporteurs du D-xylose (XylF), de la glutamine (GlnQ) et des méthylgalactosides (MglA). Les deux autres gènes identifiés codent SdhA, une sous-unité de la succinate déshydrogénase, et AldB, une aldéhyde déshydrogénase qui participe à la détoxification des alcools et aldéhydes (Xu et Johnson, 1995a). D'autres gènes impliqués dans différentes voies métaboliques et réprimés par Fis ont été identifiés : *acs*, qui code l'acétyl-coenzyme A synthétase, qui permet à la cellule d'utiliser l'acétate lors de carences en sources de carbone et à l'entrée en phase stationnaire (Browning et al., 2004) ; l'opéron *nir*, codant une nitrite-réductase cytoplasmique intervenant en conditions d'anaérobiose (Browning et al., 2000) ; ou encore l'opéron *nrf*, actif lui aussi en anaérobiose et responsable de la réduction des nitrites par le formate (Browning et al., 2005). Sur un certain nombre de ces promoteurs, Fis agit de concert avec d'autres régulateurs de la transcription et assure ainsi une expression très fine et précise de ces gènes (Browning et al., 2005).

De par les gènes qu'elle contrôle, notamment les opérons d'ARNr, la protéine de type histone Fis est donc un point de contrôle clef de la croissance bactérienne. Il a d'ailleurs été récemment démontré que Fis participait à la régulation du facteur  $\sigma^S$  spécifique des conditions de stress nutritionnel, en réprimant la transcription de *spoS* en phase exponentielle chez *S. typhimurium* (Hirsch et Elliott, 2005). Fis agirait ici par le même mécanisme que sur son propre promoteur, c'est-à-dire en empêchant la fixation de l'ARN polymérase à l'ADN.

Enfin, Fis est également impliquée dans le contrôle de l'expression des gènes de virulence dans des souches pathogènes d'*E. coli* (Goldberg et al., 2001), chez *S. flexneri* (Falconi et al., 2001) et chez *S. typhimurium* (Schechter et al., 2003). En effet, chez *S. typhimurium*, Fis active des gènes faisant partie des îlots de pathogénicité. La délétion de *fis* entraîne la diminution de l'expression de *hilA* et *invF*, associée à un déficit sérieux de virulence, chez la souris (Wilson et al., 2001). Ces deux gènes sont responsables de l'activation de plusieurs opérons de SPI-1, un îlot de pathogénicité de *S. typhimurium* qui contient en partie le système de sécrétion de type III. De plus, la comparaison des profils globaux de transcription entre une souche de référence de *S. typhimurium* et un mutant dépourvu de Fis a révélé que Fis activait les gènes du système de sécrétion de type III, d'invasion des cellules épithéliales, de la survie dans les macrophages et de fabrication des flagelles, ces gènes étant répartis dans les îlots de pathogénicité SPI-1 à 5 (Kelly et al., 2004). Des gènes impliqués dans

des voies métaboliques nécessaires à la survie dans l'intestin des mammifères sont régulés négativement par Fis (utilisation du propanediol, biosynthèse de la biotine, transport de la vitamine B12, métabolisme de l'acétate,...). Ainsi Fis joue un rôle central en coordonnant l'expression de facteurs nécessaires à la croissance mais également à la virulence.

#### ◇ La réplication

Un autre rôle fondamental de Fis est celui joué dans l'initiation de la réplication. Fis possède en effet un ou plusieurs sites de fixation au niveau de l'origine de réplication *oriC* (Gille et al., 1991 ; Filutowicz et al., 1992). Bien que la raison de cette fixation demeure inconnue, des mutants *fis* produisent une asynchronie de l'initiation de la réplication dans des conditions de croissance rapide (Boye et al., 1993). Une courbure induite par la fixation de Fis près d'*oriC* pourrait faciliter l'initiation de la réplication. Des interactions existent également au niveau d'*oriC* entre Fis, IHF et DnaA (Ryan et al., 2004).

#### ◇ Les phénomènes d'inversion et de recombinaison site-spécifique

Fis a été découverte par sa capacité à favoriser les réactions d'inversion catalysées par les recombinases du type Hin, Gin et Cin (Finkel et Johnson, 1992). Durant ces réactions, Fis se lie à deux sites entourant une séquence de recombinaison (Recombinational Enhancer) d'environ 60 pb (**Figure 20**). D'autre part, la recombinase Hin se fixe au niveau des deux sites de recombinaison *hixR* et *hixL* et y est responsable du clivage et de la ligation des brins d'ADN (Osuna et al., 1991). La fixation de Fis permet le rapprochement des deux sites *hixR* et *hixL* et du site stimulateur de la recombinaison. Ceci génère alors un intermédiaire de recombinaison au sein duquel l'interaction entre Fis et la recombinase Hin induit l'événement de recombinaison, par extrusion sous forme de boucles de l'ADN situé entre *hixR* et *hixL* (**Figure 20** ; Osuna et al., 1991). On retrouve ainsi l'effet de la fixation de Fis sur l'ADN décrit précédemment.

Enfin, Fis stimule également l'intégration et l'excision de l'ADN du bactériophage  $\lambda$  en se liant au site *attR* du génome phagique (Ball et Johnson, 1991a, b). En introduisant une courbure dans l'ADN, Fis favorise l'intégration et l'excision de l'ADN du phage (Esposito et Gerard, 2003).

#### 4. Dynamique d'expression des NAPs

Comme nous venons de le voir, Fis est la protéine de type histone majoritaire pendant la phase exponentielle. Son niveau augmente très rapidement au moment de la reprise de la croissance, pour atteindre plus de 50 000 molécules par cellule (Talukder et al., 1999). Deux autres NAPs, HU (30 000 à 50 000 molécules) et H-NS (20 000 molécules) sont abondantes en phase exponentielle. En fin de phase exponentielle, leur niveau décroît, en même temps que le degré de superhélicité de l'ADN : Fis disparaît presque totalement alors que HU reste à un niveau basal correspondant à moins d'un tiers de sa concentration en phase exponentielle (Talukder et al., 1999). H-NS subit une diminution de concentration d'environ deux fois, et reste ainsi présente en phase stationnaire. Lors de la transition entre les phases exponentielle et stationnaire, deux autres NAP deviennent majoritaires : Dps (près de 200 000 molécules) et IHF. Le rôle principal d'IHF se situe pendant cette transition, puisque son niveau atteint un pic (55 000 molécules), puis diminue par la suite, tout en restant à la moitié de son niveau maximal (Talukder et al., 1999). Enfin, en phase stationnaire prolongée, la protéine Dps est de loin majoritaire.

Une certaine hiérarchie entre ces protéines a donc été mise en évidence au cours du cycle cellulaire d'*E. coli*. L'expression de ces protéines de type histone est très dynamique tout au long de la croissance bactérienne. Ces protéines sont impliquées dans la régulation de la transcription de nombreux gènes, mais aussi dans les processus de réplication du chromosome et de recombinaison. Leurs différentes fonctions sont souvent réalisées de par leur effet sur la topologie de l'ADN. Ainsi, la dynamique d'expression de ces histones va conduire à des variations de superhélicité de l'ADN, locales ou globales, et ceci pourrait avoir un impact important sur les capacités de réponse des cellules aux fluctuations de leur environnement. De plus, comme nous allons le voir dans la partie suivante, les NAP sont également impliquées dans la structuration du chromosome en domaines plus ou moins larges et leur expression différentielle entraîne une plasticité importante de la structure du chromosome.

## II. Organisation du chromosome dans la cellule

Chez les eucaryotes, les chromosomes sont organisés en plusieurs niveaux de compaction, l'unité de base étant la structure nucléosomique réalisée par les histones (Kornberg et Lorch, 1999). Le nucléosome est le point central du contrôle de la transcription. Chez les bactéries, une telle structuration est absente, mais l'ADN est aussi organisé au sein du nucléoïde par son association avec différentes protéines (Pettijohn, 1996). Une telle organisation est nécessaire pour compacter le chromosome de manière à le contenir dans la cellule (compaction de 1000 fois pour le chromosome d'*E. coli*), mais aussi pour permettre la flexibilité nécessaire à la réplication du chromosome et à la transcription des gènes, processus qui vont par ailleurs être modifiés en fonction des conditions de l'environnement. Une partie de cette organisation repose sur le fait que le chromosome bactérien présente une superhélicité négative, qui facilite les transitions entre les conformations double et simple brin, nécessaires à tous les processus vitaux (réplication, transcription, recombinaison). En revanche, le niveau global d'organisation du nucléoïde reste peu connu, bien que des travaux récents permettent de mettre en évidence la structuration spatiale du chromosome bactérien sous forme de domaines topologiques et de macrodomaines.

### A. Les domaines topologiques

Le premier niveau de compaction du chromosome est sa division en domaines topologiques, un domaine étant défini comme la région relaxée par une coupure simple ou double brin de l'ADN (Pettijohn, 1996). En l'absence de domaines topologiques, une coupure aurait pour effet la rotation des brins d'ADN autour de cette coupure, conduisant à la perte rapide et catastrophique de la superhélicité négative, qui pourrait alors se propager à l'ensemble du chromosome et conduire ainsi à la mort cellulaire. De telles coupures peuvent en effet être générées *in vivo* par des altérations chimiques (radicaux oxygénés, alkylation, dépurination), des endonucléases cellulaires, ou par le processus même de la réplication. Ainsi, la répartition de domaines topologiques tout au long du chromosome permet de préserver celui-ci : suite à une coupure, seul un domaine sera relâché, et non le chromosome entier. Les

changements de superhélicité seront ainsi restreints à un niveau local, grâce à la présence de barrières empêchant la diffusion des supertours entre différents domaines. Les premiers travaux effectués ont montré la nécessité de multiples coupures pour relâcher totalement le chromosome *in vitro*, ce qui suggérait la présence de différents domaines topologiques (Worcel et Burgi, 1972). Ces études, ainsi que d'autres basées sur l'utilisation de triméthylpsoralène tritié (Sinden et Pettijohn, 1981), ont conclu à l'existence d'une quarantaine de domaines, d'environ 100kb chacun. Plus récemment, deux méthodes indépendantes ont été utilisées pour étudier plus finement l'organisation chromosomique en domaines topologiques.

La première est basée sur l'identification de 306 gènes dont la transcription est modifiée suite à un relâchement de l'ADN, ces gènes étant dispersés sur le chromosome d'*E. coli* (Peter et al., 2004). Postow et al. (2004) se sont servis de cette étude pour déterminer l'impact d'une coupure du chromosome sur la transcription de ces 306 gènes. Des coupures doubles brins ont été introduites à des positions définies du chromosome par l'induction *in vivo* de l'expression de l'enzyme de restriction *SwaI*, qui présente 117 sites de coupures sur le chromosome d'*E. coli* (Postow et al., 2004). L'effet de ces coupures étant un relâchement de l'ADN, la transcription des gènes situés à proximité, notamment ceux sensibles à la superhélicité, va être affectée (Peter et al., 2004). Le profil global de transcription a donc été déterminé avant et après coupure par *SwaI*. Connaissant la distance entre les sites de restriction et les gènes sensibles au relâchement de l'ADN, les auteurs ont pu conclure à un nombre de domaines d'environ 450, chacun ayant en moyenne une taille de 10kb. Ces résultats ont été appuyés par une mesure directe de la taille des boucles chromosomiques formées par microscopie électronique (Postow et al., 2004).

La deuxième étude porte sur des souches de *S. typhimurium* dans lesquelles des paires de sites *res* de recombinaison site-spécifique ont été introduites à différentes distances chromosomiques. Ces souches contiennent également les gènes, inductibles par l'IPTG, codant les résolvases des transposons Tn3 ou  $\gamma\delta$  (Higgins et al., 1996). Un système de sélection permet de détecter et de quantifier les événements de recombinaison entre les sites *res*. Ce système de recombinaison site-spécifique est basé sur l'alignement précis de 2 sites *res* dont la taille est de 140pb, et dépend d'un niveau adéquat de superhélicité négative (Stark et Boocock, 1995). Ainsi, 2 sites *res* ne peuvent recombiner qu'à l'intérieur d'un même domaine topologique, afin

de permettre la diffusion de superhélicité négative entre les deux sites. L'utilisation de ce système ingénieux a révélé que la taille moyenne d'un domaine était d'environ 25kb. Cette valeur, basée sur la stabilité des barrières entre domaines et sur celle de la résolvasse Tn3 ou  $\gamma\delta$  utilisée, est probablement sous-estimée. En effet, la stabilité élevée de ces résolvases rend non détectables des domaines dont la durée de vie serait trop courte. Des mutants de la résolvasse présentant des demi-vies variées (de 5min à plusieurs heures) ont alors été utilisés pour étudier la dynamique du chromosome (Stein et al., 2005). Ces mutants ont permis de montrer que *S. typhimurium*, comme *E. coli*, possédait plus de 400 domaines topologiques par chromosome, de taille moyenne comprise entre 11 et 13kb. D'autre part, ces domaines présentent une structure extrêmement dynamique, ce qui avait été montré précédemment (Postow et al., 2004). La dynamique de ces domaines est indispensable pour les processus de réplication et de transcription. En effet, ceux-ci provoquent une forte altération du superenroulement et les domaines topologiques doivent se refermer rapidement après le passage de la fourche de réplication ou de l'ARN polymérase (Deng et al., 2005).

Tous ces résultats impliquent non seulement la présence de barrières délimitant les différents domaines topologiques, mais également un grand dynamisme quant à leur localisation. Plusieurs hypothèses sont formulées quant à leur nature.

## 1. La transcription

Le processus de transcription de gènes codant des protéines membranaires pourrait générer des barrières à la diffusion des supertours. En effet, il a été démontré que la transcription de tels gènes provoquait une augmentation drastique de la superhélicité de l'ADN (Lynch et Wang, 1993). Le phénomène d'ancrage à la membrane cytoplasmique issu de la transcription/traduction de protéines membranaires pourrait ainsi contribuer à la formation de barrières topologiques. Ceci a été récemment confirmé en utilisant le système de recombinaison site-spécifique entre deux sites *res* (Deng et al., 2004). Ces deux sites, ainsi que le gène *tetA* codant une pompe à efflux membranaire pour la tétracycline, ont été introduits à l'intérieur d'un domaine topologique chez *S. typhimurium*. L'induction de la transcription de *tetA* engendre une réduction d'un facteur 20 de la recombinaison entre les deux sites *res*, ce qui suggère que la transcription d'un gène codant une protéine membranaire crée des

barrières topologiques (Deng et al., 2004). Cependant, la transcription de gènes codant des protéines cytoplasmiques induit la même diminution de recombinaison entre sites *res* (Deng et al., 2004). Ainsi, le phénomène d'ancrage à la membrane n'est pas indispensable, et ce serait l'activité transcriptionnelle qui serait en elle-même responsable de la formation de domaines topologiques. Deng *et al.* (2004) suggèrent cependant qu'un niveau élevé de transcription serait nécessaire, et qu'une telle activité transcriptionnelle est rare dans la cellule. Ainsi, bien que la transcription contribue à la formation de barrières topologiques, elle ne peut expliquer à elle seule la présence de plus de 400 domaines topologiques (Deng et al., 2005).

## 2. Les complexes SMC

Dans les cellules eucaryotes, les protéines SMC (Maintenance Structurale des Chromosomes) assurent la structuration et l'organisation des chromosomes (Heck, 1997). Cette famille de protéines inclut les condensines (ségrégation des chromosomes), et les cohésines (alignement des chromatides sœur). Dans la cellule bactérienne, un seul complexe a été répertorié comme membre de la famille des protéines SMC : il a été appelé MukBEF chez *E. coli* (Yamazoe et al., 1999 ; Wang et al., 2006), et BsSMC chez *Bacillus subtilis* (Hirano et Hirano, 2006).

Les protéines SMC jouent un rôle dans le maintien du niveau de superhélicité de l'ADN. En effet, des mutants de ces gènes chez *E. coli* et *B. subtilis* présentent un changement de superhélicité de l'ADN plasmidique et chromosomique, ainsi qu'une décondensation du nucléoïde (Weitao et al., 2000 ; Lindow et al., 2002). Ces mutants montrent également une baisse de viabilité à des températures supérieures à 30°C. De plus, certaines de ces mutations interfèrent avec les phénotypes conférés par d'autres mutations connues pour affecter le niveau de superenroulement dans les cellules (Weitao et al., 2000 ; Sawitzke et Austin, 2000). Par exemple, la sensibilité thermique de ces mutants est supprimée par des mutations du gène *topA*, codant la TopoI.

Les protéines de type SMC sont actives sous forme de dimères dont la forme est celle d'un V, chaque sous-unité étant repliée et enroulée sur elle-même de manière anti-parallèle (Hirano et Hirano, 2006). Les extrémités de cette structure en V se regroupent et forment un site de liaison capable d'hydrolyser l'ATP après liaison à l'ADN. Dans le cas de MukB, le

dimère est stabilisé par les co-facteurs MukE et MukF (Yamazoe et al., 1999 ; Fennell-Fezzie et al., 2005), qui permettraient la formation de polymères de complexes qui, en se liant à l'ADN, le compacteraient en structures de type « rosettes » (**Figure 21** ; Matoba et al., 2005 ; Hirano et Hirano, 2006). Cependant, le nombre total de molécules de MukB dans la cellule est de l'ordre de 150 (Kido et al., 1996), ce qui est bien trop faible pour être impliqué dans la formation de l'ensemble des 400 domaines topologiques. Comme dans le cas de la transcription, les complexes MukBEF ne peuvent contribuer qu'à la formation d'une partie des domaines topologiques.

### 3. Les Topos

L'utilisation du système de recombinaison site-spécifique entre deux sites *res*, décrit ci-dessus, a montré que le nombre de domaines topologiques augmentait dans certains mutants des gènes codant la gyrase (Staczek et Higgins, 1998 ; Stein et al., 2005). Les Topos pourraient donc intervenir dans la mise en place de domaines topologiques. Plus récemment, Moulin *et al.* (2005) ont démontré la capacité de la gyrase à former des barrières en utilisant un autre système expérimental. Il s'agit d'une construction comportant d'une part un promoteur inductible fusionné au gène rapporteur *uidA*, codant la  $\beta$ -glucuronidase (GUS), suivi d'un terminateur fort de la transcription, et d'autre part le promoteur du gène *gyrA*, sensible à la superhélicité, et qui a été fusionné à *lacZ*, codant la  $\beta$ -galactosidase (**Figure 22**). Selon le modèle de Liu et Wang (1987), l'avancée de l'ARN polymérase sur l'ADN lors de la transcription provoque des supertours positifs en amont de la machinerie transcriptionnelle et des supertours négatifs en aval (**Figure 5A**). L'induction de la transcription de *uidA* va ainsi provoquer des supertours positifs qui pourraient se transmettre jusqu'au promoteur *gyrA*. Suite à l'induction de la transcription du gène *uidA*, Moulin *et al.* (2005) ont démontré un accroissement de l'activité  $\beta$ -galactosidase, et donc de l'activité du promoteur *gyrA* sensible au niveau de superhélicité du chromosome. La diffusion de supertours positifs créés par la transcription de *uidA*, et donc l'induction du promoteur de *gyrA*, peut se faire à des distances d'environ 4kb. Des séquences d'ADN connues pour contenir des sites de fixation de la gyrase ont alors été insérées entre le promoteur inductible et le promoteur de *gyrA* (Moulin et al., 2005). Ces insertions entraînent une diminution de la transcription à partir du promoteur

*gyrA* et empêchent donc la diffusion des supertours positifs générés par l'induction transcriptionnelle du gène *uidA*. Des sites de fixation de la gyrase peuvent donc faire office de barrières pour les domaines topologiques, et cette capacité est fortement corrélée à l'activité de coupure de l'ADN par l'enzyme (Moulin et al., 2005).

La gyrase est connue pour se fixer sur des sites spécifiques appelés SGS (Strong Gyrase Site) dans le cas du génome du bactériophage Mu (Pato et al., 1990), mais également présents sur des plasmides, à différentes positions sur le chromosome, et sur le génome d'autres bactériophages (Morrison et Cozzarelli, 1979 ; Lockshon et Morris, 1985 ; Franco et Drlica, 1988). D'autre part, la gyrase se fixe sur des séquences répétées présentes en plusieurs centaines de copies chez *E. coli* et *S. typhimurium*, nommées séquences REP (Repetitive Extragenic Palindromic) (Yang et Ames, 1988) et BIME (Bacterial Interspersed Mosaic Elements) (Moulin et al., 2005). Le fait que la gyrase puisse constituer un type de barrières en se fixant sur ces sites rappelle la situation observée chez les eucaryotes, où la topoisomérase II est localisée à la base des boucles d'ADN (Earnshaw et Heck, 1985).

#### **4. Les protéines liées au nucléoïde**

Les protéines de type histone jouent un rôle majeur dans l'organisation du nucléoïde grâce à deux propriétés fondamentales : elles sont présentes en quantités abondantes dans les cellules bactériennes et se fixent à l'ADN de manière non spécifique. Même pour celles présentant une certaine spécificité de fixation, comme IHF ou Fis, les quantités présentes sont bien supérieures au nombre de sites de fixation à forte affinité, et elles vont donc également se fixer de manière aspécifique (Johnson et al., 2005 ; Deng et al., 2005).

Hardy et Cozzarelli (2005) ont tenté d'identifier les constituants des barrières topologiques en utilisant un crible génétique afin d'identifier des mutants de ces barrières chez *E. coli*. L'hypothèse de départ consiste à imaginer qu'un mutant de barrières topologiques va présenter un ADN globalement plus relâché que la souche sauvage. Une souche a été construite, contenant les gènes rapporteurs *bla* et *lacZ*, codant respectivement la  $\beta$ -lactamase et la  $\beta$ -galactosidase, chacun sous le contrôle du promoteur de *gyrB*, dont la transcription dépend du degré de superhélicité et augmente suite à un relâchement de l'ADN (Menzel et Gellert, 1987 ; Peter et al., 2004). Chacun des gènes rapporteurs a été intégré dans le

chromosome de la souche rapporteur, les deux gènes étant insérés à deux loci différents (Hardy et Cozzarelli, 2005). Ainsi, la sélection sur les deux phénotypes simultanés, c'est-à-dire la résistance à l'ampicilline et une forte activité  $\beta$ -galactosidase, permet de visualiser l'activation de la transcription du promoteur *gyrB* probablement due à un relâchement de l'ADN. Celui-ci a lieu au niveau global du chromosome, plutôt qu'à un niveau local, car les deux gènes rapporteurs sont insérés à deux loci différents. Une mutagenèse par insertion utilisant le transposon mini-Tn10 a été effectuée, et les clones portant une insertion aléatoire du mini-Tn10 ont alors été sélectionnés et criblés pour une fréquence de résistance à l'ampicilline accrue, ainsi que pour une production accrue de  $\beta$ -galactosidase, par étalement sur un milieu contenant de l'ampicilline et du Xgal.

Parmi les gènes identifiés par cette approche génétique, Hardy et Cozzarelli (2005) ont identifié *hns*, codant la protéine de type histone H-NS, qui constitue donc une barrière de domaines topologiques. Afin d'analyser de façon plus exhaustive le rôle potentiel des protéines de type histone, Hardy et Cozzarelli (2005) ont également utilisé une approche plus ciblée en délétant chacun des gènes d'histone dans leur souche rapporteur. Ceci leur a permis de montrer que Fis constituait également une barrière, ce qui n'était pas le cas d'IHF, de StpA ou de HU. Les protéines de type histone Fis et H-NS ont été particulièrement étudiées au niveau structural pour leur interaction avec l'ADN, sur lequel elles forment des boucles (Dame et al., 2000 ; Schneider et al., 2001). Elles peuvent donc constituer des barrières de domaines topologiques. H-NS forme des oligomères et peut se fixer de façon coopérative sur plusieurs centaines de paires de bases (Dame et al., 2000). Ces auteurs ont montré par microscopie à force atomique que de telles structures ADN/H-NS pouvaient s'associer entre elles, provoquant ainsi non seulement la condensation de l'ADN, mais aussi une extrusion sous forme de boucles d'ADN libre (**Figure 23A**). Par ailleurs, des mutants H-NS présentent une fréquence élevée de cellules anucléées (Kaidow et al., 1995). Toutes ces données font d'H-NS une frontière idéale de domaines topologiques.

Fis est capable de se lier à l'ADN sur une séquence spécifique, bien que dégénérée, mais également de façon non spécifique (voir paragraphe I.B.3). *In vitro*, Fis se fixe préférentiellement à des points de croisements et de branchements sur l'ADN superenroulé, suggérant que Fis puisse, comme H-NS, contraindre des boucles d'ADN (**Figure 23B**). Des mutants *fis* présentent des défauts de ségrégation des chromosomes et de division cellulaire à

haute température (Filutowicz et al., 1992). De plus, Fis est présente en quantités extrêmement importantes dans la cellule en phase exponentielle de croissance (plus de 50 000 molécules par cellule), moment où le nombre de domaines topologiques est élevé, puis diminue fortement en phase stationnaire, avec le nombre de domaines (Deng et al., 2005).

## 5. Les autres protéines

La stratégie utilisée par Hardy et Cozzarelli (2005) leur a également permis de mettre en évidence 3 autres protéines pouvant contribuer à la formation des barrières topologiques.

La première, codée par le gène *pgm*, est une phosphoglucomutase impliquée dans la glycolyse. Ce gène forme un opéron avec *seqA*, codant la protéine SeqA, qui se lie à l'ADN hémiméthylé et séquestre l'origine de réplication chez *E. coli*, assurant ainsi la coordination entre l'initiation de la réplication et le cycle cellulaire (Lu et al., 1994). SeqA est également impliquée dans la condensation du chromosome (Weitao et al., 1999). Cependant, le rôle précis de Pgm et/ou de SeqA en tant que barrières topologiques n'est pas connu (Hardy et Cozzarelli, 2005).

La deuxième protéine identifiée est DksA. Des mutants du gène *dksA* chez *E. coli* et *S. flexneri* présentent des cellules filamenteuses. Ces mutants révèlent également des interactions génétiques entre *dksA* et d'autres gènes, tels que *dnaK*, qui code un facteur de l'hôte nécessaire à la réplication de l'ADN du bactériophage  $\lambda$  (Georgopoulos, 1977), *mukB*, codant une protéine de type SMC (voir paragraphe II.A.2), *ruv*, impliqué dans les phénomènes de recombinaison (West, 1997), et *recN*, codant une protéine de type SMC induite par le système SOS (Meddows et al., 2005). Les phénotypes des mutants *dksA* suggèrent ainsi un rôle dans la maintenance structurale du chromosome (Hardy et Cozzarelli, 2005). Ce rôle pourrait également être lié à l'activité de l'ARN polymérase, car DksA peut se fixer au niveau du canal secondaire de l'ARN polymérase où elle stabiliserait la fixation de (p)ppGpp, l'effecteur de la réponse stringente (Perederina et al., 2004 ; Paul et al., 2004).

La troisième protéine identifiée est TktA, la transcétolase majeure du métabolisme des sucres chez *E. coli* (Josephson et Fraenkel, 1969). Il s'agit de l'enzyme majeure du cycle des pentoses phosphate. Là aussi, son rôle dans les barrières topologiques reste mystérieux, mais *tktA* a été récemment identifié lors d'un criblage de mutants sensibles à la norfloxacine, un

antibiotique ciblant les Topos de type II chez *Bacteroides thetaiotaomicron* (Ueda et Yoshimura, 2003). De plus, TktA est un homologue de RecP, une protéine impliquée dans les processus de recombinaison chez *Streptococcus pneumoniae* (Reizer et al., 1993). TktA pourrait donc avoir une double fonctionnalité, dans le métabolisme des sucres et l'organisation du chromosome.

En plus des protéines identifiées par l'étude de Hardy et Cozzarelli (2005), les protéines intervenant dans la ségrégation du chromosome au cours de la division cellulaire pourraient également jouer un rôle de barrières de domaines topologiques. D'une part, FtsK coordonne la division cellulaire avec la ségrégation du chromosome. Elle est ancrée par son domaine N-terminal dans la membrane plasmique au niveau du septum par une interaction avec la protéine FtsZ, pendant que le domaine C-terminal aide à la translocation de l'ADN en dehors de l'espace du septum et facilite la décaténation (Sherratt, 2003 ; Aussel et al., 2002). D'autre part, MreB est une protéine dont la structure est homologue à l'actine des eucaryotes et participe aussi au déplacement de l'ADN après la réplication (Kruse et Gerdes, 2005).

## **B. Les macrodomaines**

L'organisation du chromosome en domaines topologiques ne reflète qu'un premier niveau de structuration, sous-jacent à un degré d'organisation impliquant de grandes distances chromosomiques bien supérieures à 10kb. En particulier, des analyses d'hybridation *in situ* par fluorescence (FISH) chez *E. coli* ont révélé l'existence de 2 grandes régions d'environ 1Mb chacune, appelées macrodomaines, l'une située autour de l'origine de réplication *oriC* et l'autre autour du terminus de réplication *ter* (Niki et al., 2000). Comme dans le cas des domaines topologiques, les macrodomaines ont une structure dynamique au cours du cycle cellulaire. Cette organisation en macrodomaines a été confirmée et précisée par une approche génétique (Valens et al., 2004), basée sur le système de recombinaison site-spécifique du bactériophage  $\lambda$ , afin de détecter la proximité spatiale de séquences d'ADN éloignées sur la structure primaire de la double hélice. Différentes souches ont été construites, comportant 2 sites d'attachement *att* du génome de  $\lambda$ , l'un à une position définie du chromosome et l'autre à des positions variées suivant les souches. Celles pour lesquelles un événement de recombinaison a eu lieu entre 2 sites *att* ont été sélectionnées. Les fréquences de

recombinaison reflètent ainsi la capacité d'interactions et donc la proximité spatiale de séquences d'ADN. Valens *et al.* (2004) ont montré que les événements de recombinaison étaient fortement limités à certaines régions chromosomiques, ce qui a permis de révéler la présence de 4 macrodomaines sur le chromosome d'*E. coli*, séparés par deux régions moins structurées (**Figure 24**). Deux de ces macrodomaines entourant *oriC* et *ter* et d'environ 1Mb chacun, correspondent à ceux déterminés précédemment par FISH (Niki *et al.*, 2000). Cette organisation en macrodomaines permettrait d'une part la condensation de l'ADN conduisant à l'interaction entre séquences d'ADN à l'intérieur d'un même macrodomaine, et d'autre part la séquestration des macrodomaines, inhibant ainsi la recombinaison de séquences d'ADN localisées dans des macrodomaines différents (Valens *et al.*, 2004).

Cette organisation physique du chromosome en macrodomaines a des répercussions sur l'expression globale du génome. En effet, l'analyse du profil global de transcription des gènes chez *E. coli* en fonction de leur position sur le chromosome a permis de détecter une organisation spatiale de la transcription des gènes (Jeong *et al.*, 2004). La transcription globale a été analysée en présence et en absence de norfloxacine, un inhibiteur de la gyrase. Les profils de transcription obtenus ont été analysés sur la base de l'activité transcriptionnelle de chaque gène individuellement, mais également en fonction de similitudes régionales. Trois profils spatiaux de transcription ont été mis en évidence (Jeong *et al.*, 2004). Tout d'abord, dans un premier groupe de gènes, séparés par une distance allant jusqu'à 16kb, la transcription est dépendante du niveau de superhélicité de l'ADN, c'est-à-dire du traitement à la norfloxacine. Les deux autres profils spatiaux de transcription correspondent à des régions plus grandes, d'une part de l'ordre de 100kb, et d'autre part de 600 à 800 kb. La formation de ces deux derniers profils de transcription est quant à elle indépendante du degré de superhélicité du chromosome, mais semble être corrélée à la distribution des sites de fixation de la gyrase. De plus, Jeong *et al.* (2004) ont également montré que les profils de transcription sont plus importants au niveau du bras gauche de la fourche de réplication à partir d'*oriC* (**Figure 24**). La transcription de certains gènes proches de *oriC* et autour de *ter* est également sensible au niveau de superhélicité (Peter *et al.*, 2004 ; Willenbrock *et al.*, 2004 ; Travers *et al.*, 2005).

Ainsi, l'organisation physique du chromosome en domaines topologiques et en macrodomaines a de fortes répercussions sur les profils spatiaux de transcription des gènes. De

plus, les conditions de l'environnement vont influencer sur ces deux fonctions cellulaires, structure du chromosome et transcription globale des gènes, qui sont donc non seulement intimement connectés, mais également dynamiques et modulables. Des expériences de microscopie en fluorescence, utilisant une fusion du gène *rpoC*, codant la sous-unité  $\beta'$  de l'ARN polymérase, avec le gène *gfp* codant la protéine de fluorescence verte de méduse, ont permis d'étudier la distribution des molécules d'ARN polymérase dans la cellule et en particulier au niveau du nucléoïde et de sa périphérie (Cabrera et Jin, 2003). La distribution des molécules d'ARN polymérase est fortement dépendante des conditions de culture. En conditions de croissance rapide, l'ARN polymérase est concentrée au niveau de spots de fluorescence qui correspondent aux opérons d'ARNr. Cette forte activité transcriptionnelle semble d'ailleurs être corrélée à une condensation du chromosome (Cabrera et Jin, 2003). Tout traitement inhibant la transcription des ARNr, par exemple l'induction de la réponse stringente par une carence en acides aminés, fait d'une part disparaître les spots de fluorescence et d'autre part diminuer la condensation du chromosome.

Ainsi, non seulement l'activité transcriptionnelle a une influence sur la structuration du chromosome (voir paragraphe II.A.1), mais la structure du chromosome, ainsi que les quantités des différentes NAPs, influent sur la transcription et sont modifiées au cours du cycle cellulaire chez *E. coli* (**Figure 25**) et dans différentes conditions de l'environnement, par exemple les carences nutritionnelles (voir paragraphe IV.A.). Le passage de la phase exponentielle à la phase stationnaire est caractérisé par une restructuration complète du chromosome bactérien, aboutissant tout d'abord à des structures toroïdales puis à des structures cristallines au sein desquelles l'ADN est séquestré et protégé (Frenkiel-Krispin et al., 2004). La structure du nucléoïde est donc dynamique et varie en fonction des conditions de l'environnement, qui en retour conduisent à des fluctuations des profils d'expression du génome. Ces deux processus pourraient être associés afin de permettre une réponse adéquate des bactéries aux fluctuations de leur environnement.

### III. Impact de la superhélicité de l'ADN sur l'expression des gènes

Les fluctuations du niveau de superhélicité de l'ADN ont des conséquences importantes sur l'expression de nombreux gènes (Steck et al., 1993 ; Peter et al., 2004), dont la transcription des opérons d'ARN ribosomiques (Muskhelishvili et Travers, 2003). Une superhélicité négative est en effet requise pour leur transcription. Ainsi, ne serait-ce que par son effet sur la synthèse des ARNr, le niveau de superhélicité de l'ADN doit être régulé de façon précise pour l'adaptation des bactéries à des conditions de croissance ou de carences. D'autre part, une augmentation de la superhélicité négative du chromosome provoquée par une mutation du gène *topA* entraîne une modification de l'expression des gènes, observée par électrophorèse bidimensionnelle de protéines (Steck et al., 1993). Des effets similaires sont obtenus suite à une mutation du gène *hupA*, codant la sous-unité  $\alpha$  de HU, qui provoque une compaction du chromosome (Kar et al., 2005). De façon générale, des mutations affectant les gènes codant les NAPs (H-NS, Fis, HU, IHF) ont des effets drastiques sur la transcription globale des gènes (Kar et al., 2005 ; Peter et al., 2004). Des modifications de superhélicité de l'ADN provoquent également de profonds changements d'expression chez d'autres bactéries, comme par exemple chez *Haemophilus influenzae* (Gmuender et al., 2001), ou même chez la levure *Saccharomyces cerevisiae* (Eriksson et al., 2005).

Plus récemment, la comparaison des profils de transcription globale d'une souche d'*E. coli* dont le chromosome présente un défaut de superhélicité (inactivation des gènes codant la gyrase) avec une souche de référence, ou de cette souche de référence cultivée en présence ou en absence d'inhibiteurs de la gyrase, révèle que l'expression de 7% des gènes du chromosome est affectée par le niveau de superhélicité (Peter et al., 2004). La relaxation du chromosome provoque l'activation de la transcription de 106 gènes et l'inhibition de celle de 200 gènes. Les gènes dont la transcription est modifiée remplissent des fonctions diverses et sont dispersés sur le chromosome. Les gènes induits présentent des séquences riches en AT à la fois aux niveaux des promoteurs et des régions codantes, ce qui n'est pas le cas des gènes dont l'expression est réprimée dans ces conditions. Une superhélicité négative élevée réprimerait l'échappement de l'ARN polymérase au niveau de promoteurs riches en AT ainsi que l'élongation de la

transcription (Peter et al., 2004). A l'opposé, une superhélicité négative élevée est connue pour favoriser l'initiation de la transcription au niveau de promoteurs possédant une séquence discriminatrice riche en GC entre la boîte -10 et le site +1 d'initiation de la transcription. Un discriminateur est présent au niveau des promoteurs des opérons d'ARNs ribosomiques, ce qui explique la nécessité d'un superenroulement négatif élevé pour leur transcription. Celui-ci permettrait de lever la barrière énergétique posée par le discriminateur riche en GC. Un discriminateur est présent pour tous les promoteurs connus pour être réprimés par (p)ppGpp, l'effecteur de la réponse stringente (Cashel et al., 1996). Ainsi, la plupart des gènes exprimés dans les conditions optimales de croissance nécessitent une superhélicité négative élevée.

La superhélicité de l'ADN est donc intimement liée à l'expression des gènes et pourrait représenter un élément régulateur majeur de l'adaptation des bactéries à leur environnement, en ajustant la transcription globale des gènes aux variations de l'environnement (**Figure 25**). Cependant, malgré une organisation dynamique de la structure de l'ADN, sa superhélicité doit être contrôlée de façon précise pour éviter des modifications trop importantes qui auraient des conséquences létales pour la cellule. Ceci est réalisé par l'intermédiaire d'un contrôle homéostatique du niveau de superhélicité de l'ADN.

#### **IV. Régulation et contrôle homéostatique du niveau de superhélicité de l'ADN**

La régulation de la superhélicité de l'ADN est finement contrôlée à plusieurs niveaux. Le premier est une régulation croisée des gènes de Topos en fonction du degré de superhélicité de l'ADN. L'expression et l'activité de la gyrase sont en effet inhibées par un superenroulement de l'ADN trop élevé (Menzel et Gellert, 1983). A l'opposé, les promoteurs des gènes *gyrA* et *gyrB* sont activés en cas de relâchement de l'ADN (Franco et Drlica, 1989). Le promoteur *gyrA* est d'ailleurs utilisé comme senseur de la topologie de l'ADN (Moulin et al., 2005). Le mécanisme par lequel ce promoteur est activé par un relâchement de l'ADN reste inconnu, mais la boîte -10 du promoteur serait impliquée (Straney et al., 1994). La formation du complexe ouvert lors de l'initiation de la transcription de *gyrA* et *gyrB* pourrait également

constituer une étape sur laquelle le relâchement de l'ADN agirait comme activateur (Menzel et Gellert, 1987). Contrairement à la gyrase qui est activée par une diminution de la superhélicité de l'ADN, la TopoI est elle activée par une augmentation du niveau de superhélicité (Tse-Dinh, 1985 ; Zechiedrich et al., 2000). L'état de superenroulement du chromosome établit ainsi une boucle d'autorégulation des Topos dans la cellule, permettant le maintien d'un niveau adéquat de superhélicité.

Les fluctuations de l'environnement influencent fortement le niveau de superhélicité du chromosome bactérien, comme par exemple les carences nutritionnelles (Balke et Gralla, 1987), les stress osmotiques (Higgins et al., 1988), oxydatifs (Weinstein-Fischer et al., 2000), thermiques et de pH (Lesley et al., 1990). La seule autorégulation des Topos sur elles-mêmes ne peut pas expliquer les variations de superhélicité observées en fonction des conditions environnementales. Des niveaux supplémentaires de régulation du degré de superhélicité, et donc des gènes codant les Topos, doivent donc exister pour détecter et répondre aux signaux de l'environnement.

### **A. Contrôle homéostatique de la superhélicité de l'ADN en fonction des conditions environnementales**

Le mécanisme moléculaire par lequel le niveau de superhélicité de l'ADN varie en fonction des conditions de l'environnement, tout en étant maintenu à des niveaux physiologiques adéquats, est le mieux décrit dans le cas des transitions entre apports de nutriments et carences nutritionnelles, et vice versa. En coordination avec certains facteurs de transcription, c'est la protéine de type histone Fis (voir paragraphe I.B.3) qui joue le rôle majeur dans cette régulation, qui consiste en un contrôle homéostatique de la superhélicité de l'ADN (**Figure 26**). En effet, Fis exerce un contrôle transcriptionnel sur les gènes de Topos (Schneider et al., 1999) et en retour, le niveau de superhélicité de l'ADN régule la transcription du gène *fis* (Schneider et al., 2000b).

Lors d'un apport de nutriments, c'est-à-dire en début de phase exponentielle de croissance, l'activité de la gyrase augmente suite à l'accroissement de la charge énergétique cellulaire, en l'occurrence du rapport [ATP]/[ADP] (van Workum et al., 1996 ; Wijker et al., 1995). Il s'ensuit une augmentation du niveau de superhélicité négative de l'ADN (Balke et

Gralla, 1987), qui active alors la transcription de *fis* (Schneider et al., 2000b). La protéine Fis atteint sa concentration maximale lors de la phase de croissance exponentielle. Cette augmentation de la superhélicité de l'ADN et de Fis est alors favorable à la transcription des gènes nécessaires à la croissance bactérienne, dont les opérons d'ARN ribosomiques (Muskhelishvili et al., 1997). En parallèle, Fis va également réprimer la transcription des gènes *gyrA* et *gyrB* (Schneider et al., 1999), ce qui va entraîner en fin de phase exponentielle une diminution de la superhélicité de l'ADN. La transcription de Fis nécessitant une superhélicité élevée, ce relâchement de l'ADN entraîne la diminution de la transcription de Fis (Schneider et al., 2000b). A ce moment, lors du passage en phase stationnaire, il y a également induction de la réponse stringente et donc de la synthèse de (p)ppGpp (Cashel et al., 1996), qui va inhiber la transcription de *fis* (Mallik et al., 2004). Tous ces mécanismes conduisent à une forte diminution de la transcription des opérons d'ARNr. La forte diminution de concentration de Fis va alors lever l'inhibition sur la transcription des gènes *gyrA* et *gyrB*. Malgré cela, un relâchement de l'ADN est observé lors du passage en phase stationnaire (Balke et Gralla, 1987 ; Reyes-Dominguez et al., 2003), et cela s'explique en partie par l'accumulation de  $\sigma^S$  (Hengge-Aronis, 2002). En effet, parmi les nombreux gènes-cibles de  $\sigma^S$ , se trouvent *gyrI*, codant un inhibiteur de la gyrase (Baquero et al., 1995 ; Oh et al., 2001), et *topA*, dont un des promoteurs est  $\sigma^S$ -dépendant (Qi et al., 1997). Une interaction directe de la gyrase avec GyrI a été montrée (Nakanishi et al., 1998), et ceci a été confirmé par l'étude du mécanisme d'action de cet inhibiteur, qui séquestrerait la gyrase probablement en empêchant son interaction avec l'ADN (Chatterji et Nagaraja, 2002). D'autre part, la diminution de la quantité de nutriments dans le milieu va empêcher l'activité de la gyrase, sensible au rapport [ATP]/[ADP] (van Workum et al., 1996). Ainsi, malgré la transcription des gènes codant la gyrase suite à la levée de la répression par Fis, le degré de superhélicité de l'ADN diminue en phase stationnaire.

Le relâchement de l'ADN en phase stationnaire est alors favorable à la transcription des gènes  $\sigma^S$ -dépendants. En effet, même en phase stationnaire,  $\sigma^S$  ne représente que 30% de la quantité du facteur sigma végétatif  $\sigma^{70}$ , et un des principaux facteurs discriminants pour la transcription des gènes entre  $\sigma^{70}$  et  $\sigma^S$  est le degré de superhélicité du promoteur. La diminution de superhélicité permet ainsi d'activer les gènes essentiels à la survie en phase stationnaire (Kusano et al., 1996 ; Bordes et al., 2003), et permet donc l'adaptation des bactéries aux conditions de carences.

Bien que l'activité de la gyrase soit diminuée en phase stationnaire, sa quantité reste constante. Lors d'un apport en nutriments, ceci permet probablement une augmentation immédiate du niveau de superenroulement de l'ADN suite à l'augmentation de la charge énergétique, et donc une reprise de la croissance cellulaire (Reyes-Dominguez et al., 2003). La protéine Fis va être à nouveau synthétisée et jouer son rôle majeur sur le contrôle homéostatique de la superhélicité de l'ADN et sur la croissance cellulaire. Fis est donc un régulateur essentiel de la croissance bactérienne en fonction des conditions nutritionnelles de l'environnement.

## **B. Autres stress environnementaux et topologie de l'ADN**

Le niveau de superhélicité de l'ADN est aussi modifié suite à différents stress, tels que le stress osmotique ou oxydatif. Cette modification est transitoire et associée à des changements, eux aussi temporaires, de l'expression des gènes. Ceci permet ainsi la synthèse de protéines nécessaires à la survie de la cellule. Quand les conditions de stress disparaissent, le niveau de superenroulement de l'ADN revient à celui d'origine, comme les profils de transcription des gènes.

Lors de stress oxydatifs, c'est à nouveau la protéine Fis qui intervient dans le contrôle de la superhélicité et de l'expression des gènes nécessaires à la survie dans ces conditions. Un stress oxydatif a pour effet de diminuer la superhélicité de l'ADN. Ceci est réalisé par l'induction du gène *topA*, codant la TopoI. La protéine de type histone Fis a été démontrée comme étant indispensable à la survie au stress oxydatif, car Fis active *topA* dans ces conditions (Weinstein-Fischer et al., 2000). Un mutant dépourvu de Fis est en effet sensible au stress oxydatif. De plus, un tel stress induit la synthèse de polyamines (Tkachenko et Nesterova, 2003), connues pour activer la traduction de Fis (Yoshida et al., 2004). La diminution de superhélicité, liée à l'activation de *topA* par Fis, entraîne alors l'induction des gènes nécessaires à la réparation des dommages de l'ADN suite au stress oxydatif et à la détoxification des espèces réactives de l'oxygène, permettant ainsi d'éliminer le stress (Weinstein-Fischer et al., 2000). Ensuite, la superhélicité augmente à nouveau suite à la diminution de transcription de *topA* et les gènes nécessaires à la survie ne sont plus transcrits.

Fis est donc un régulateur global de l'expression des gènes, qui permet l'adaptation des bactéries à leur environnement via un contrôle du degré de superhélicité de l'ADN.

En plus du stress oxydatif, la transcription du gène *topA* est capable de répondre à d'autres stress grâce à la présence de 4 promoteurs (Qi et al., 1997), qui sont activés dans des conditions différentes : P1 est activé par un choc à haute température par l'intermédiaire de  $\sigma^{32}$  (Lesley et al., 1990), le facteur sigma induit au cours de chocs thermiques (Grossman et al., 1984) ; P2 et P4 sont les promoteurs majoritaires au cours de la phase exponentielle à 37°C et 30°C ; P5 (ou PX1) est induit en phase stationnaire par l'intermédiaire de  $\sigma^S$  (Qi et al., 1997), le facteur sigma spécifique de la phase stationnaire et d'autres stress chez *E. coli* (McCann et al., 1991). Ces multiples promoteurs permettraient la synthèse *de novo* de TopoI en cas de changement de conditions de croissance, et donc un changement rapide du niveau de superhélicité.

Lors d'un choc osmotique, le superenroulement de l'ADN augmente, à la suite de l'augmentation du rapport [ATP]/[ADP] dans la cellule, qui stimule l'activité de la gyrase (Higgins et al., 1988 ; Cheung et al., 2003). Ceci active la transcription des gènes impliqués dans la réponse à ce stress, notamment des gènes codant des osmorégulateurs ou impliqués dans la synthèse du peptidoglycane. Le gène *proU* par exemple, codant un système de transport d'osmoprotecteurs tels que la glycine et la bétaine, est fortement induit (Higgins et al., 1988). Une mutation dans le gène *topA*, par l'augmentation de superhélicité qu'elle induit, peut mimer l'effet d'un stress osmotique et induire l'expression de *proU*. Après un stress osmotique, la superhélicité revient à son niveau d'origine, ainsi que l'expression de *proU* et des autres gènes.

D'autres stress (pH, thermique) influent également de façon transitoire sur le niveau de superhélicité de l'ADN (Lesley et al., 1990 ; Jones et al., 1992). Comme les deux cas décrits ci-dessus, ces changements de superhélicité sont transitoires et vont également influencer de façon temporaire sur la transcription des gènes nécessaires à la survie à ces stress, ce qui permet aux bactéries de s'adapter à ces conditions de l'environnement.

La superhélicité de l'ADN représente donc un point de contrôle clef de l'adaptation des bactéries aux transitions nutritionnelles et aux stress environnementaux. Les fluctuations de l'environnement entraînent des modifications du niveau de superenroulement de l'ADN, qui en retour provoquent des changements dans les profils globaux de transcription des gènes.

Ceci permet ainsi l'adaptation des bactéries à leur environnement. Un contrôle homéostatique du niveau de superenroulement permet également son maintien à un niveau physiologique, nécessaire au bon fonctionnement de tous les processus vitaux pour la cellule. Toutes ces propriétés font de la superhélicité de l'ADN un réseau de régulation global de l'expression des gènes, indispensable à la détection de signaux de l'environnement et à la transmission de ces signaux à la machinerie cellulaire. Ce réseau pourrait donc constituer une cible de la sélection naturelle et donc des processus évolutifs.

INTRODUCTION

PARTIE II

LA STRATÉGIE  
D'ÉVOLUTION  
EXPÉRIMENTALE

L'évolution fait depuis longtemps l'objet de recherches dont le but est d'en comprendre les mécanismes. Cependant, il est difficile d'appréhender les différentes étapes et les modifications qui amènent un organisme à s'adapter à son environnement. Les nombreuses études comparatives menées (Losos et al., 1998 ; Rundle et al., 2000 ; Derome et al., 2006) présentent en effet deux inconvénients majeurs : les fluctuations de l'environnement et l'absence de dénominateur commun, c'est-à-dire la non-connaissance de l'organisme ancêtre. Ainsi, l'évolution darwinienne, c'est-à-dire la diversification des organismes suivie de la sélection des individus les plus adaptés à leur environnement, est soumise à deux contraintes qui rendent son étude difficile : l'environnement et l'histoire évolutive passée. C'est pourquoi des stratégies d'évolution expérimentale ont vu le jour au cours des dernières décennies. Ces stratégies utilisent pour la plupart des microorganismes (virus, bactéries ou levures), qui présentent différents avantages : temps de génération court, culture aisée (par rapport à des organismes plus complexes), et présence de plusieurs réplicats pour chaque expérience. Un certain nombre de stratégies d'évolution expérimentale ont utilisé des bactériophages (Bull et al., 1997). Cependant, ce sont des organismes beaucoup moins complexes que les bactéries, ne présentant par exemple pas de réseaux complexes de régulation des gènes.

La bactérie *E. coli* a été très souvent utilisée dans de telles expériences d'évolution. Plusieurs conditions de croissance ont été étudiées, et notamment une adaptation à de longues périodes de carence. Les bactéries s'adaptent par des modifications du gène *rpoS*, codant le facteur sigma de la phase stationnaire (Zinser et Kolter, 2000), mais également en augmentant leurs capacités à utiliser les acides aminés comme sources de carbone (Zinser et Kolter, 1999). Le second type d'évolution expérimentale est une adaptation à une croissance continue en chimostat, dans laquelle les bactéries sont maintenues dans un état physiologique constant. Des modifications parallèles conduisant à un transport amélioré du glucose à travers la membrane ont été détectées dans des populations indépendantes (Ferencsik, 2001). Enfin, le dernier type d'évolution expérimentale est celui initié par Richard Lenski, et qui consiste à adapter les bactéries à des transitions quotidiennes entre phase de croissance et carences nutritionnelles. Cette expérience possède plusieurs particularités : c'est tout d'abord la plus longue qui ait été effectuée à ce jour, puisqu'elle n'a quasiment pas été interrompue pendant 18 ans. Ensuite, tous les intermédiaires d'évolution ont été conservés et peuvent donc être

« ressuscités » à tout moment. Enfin, les nombreux travaux effectués à partir de cette stratégie en font une des mieux caractérisées, jusqu'à présent.

Cette stratégie d'évolution expérimentale, utilisée dans ce travail, fait l'objet d'une revue intitulée « Global regulatory networks plasticity during experimental evolution in *E. coli* », qui sera soumise à la revue *BioEssays* (Philippe N., Crozat E., Lenski RE., and Schneider D.). Une présentation d'ensemble de cette stratégie d'évolution expérimentale est donnée ici.

Le système d'évolution expérimentale a été initié dans le laboratoire du Pr. Richard Lenski (Michigan State University, USA) en 1988. Un clone ancêtre d'*E. coli* B, de phénotype Ara-, c'est-à-dire incapable d'utiliser l'arabinose comme source de carbone, ainsi qu'un variant isogénique de phénotype Ara+, ont été utilisés pour initier 12 populations indépendantes, 6 étant issues de l'ancêtre Ara- (appelées Ara-1 à Ara-6) et 6 de l'ancêtre Ara+ (appelées Ara+1 à Ara+6). Le phénotype d'utilisation de l'arabinose est utilisé comme marqueur dans les expériences de compétition et a été démontré comme étant neutre dans ces conditions (Lenski et al., 1991). Toutes ces populations sont propagées pendant 24h à 37°C, sous agitation, dans 10mL de milieu minimum Davis (DM) contenant une quantité limitée de glucose (25µg/mL, DM25). Chaque jour, elles sont diluées au 1/100<sup>ème</sup> dans le même milieu. Cette dilution permet environ 6,64 générations par 24h ( $\log_2 100$ ), et une concentration d'environ  $5.10^7$  cellules/mL est obtenue après 24h de culture. Depuis 1988, les 12 populations ont donc évolué pendant plus de 40 000 générations dans ce milieu constant, de façon indépendante à partir d'un ancêtre commun, ce qui, à l'échelle humaine, correspondrait environ à un million d'années.

Chaque jour, ces populations rencontrent d'abord une phase de latence, une phase exponentielle, puis elles entrent en phase stationnaire jusqu'au jour suivant, où les cellules recommencent à croître dans le milieu frais. Ces populations sont donc propagées dans des conditions où elles alternent tous les jours entre des conditions de croissance et de carences nutritionnelles, et vice versa. Outre les nombreux avantages liés à *E. coli* (multiplication rapide et clonale, grande taille de populations, données disponibles très nombreuses), ce système d'évolution expérimentale présente un certain nombre d'atouts :

- des échantillons des 12 populations ont été prélevés et congelés toutes les 500 générations, permettant ainsi d'isoler des individus évolués à chacun des temps d'évolution et dans chacune des 12 populations. N'importe quel clone évolué peut donc

être comparé à son clone ancêtre, par exemple par des expériences de compétitions, mais aussi pour tout paramètre phénotypique ou moléculaire.

- les mutations identifiées au cours de cette évolution expérimentale peuvent être introduites par recombinaison homologue dans le chromosome de l'ancêtre et de tout clone évolué, au moyen de plasmides suicides, et leurs effets peuvent être analysés dans différents contextes génétiques et environnementaux. Le stade d'apparition et de fixation de chaque mutation peut aussi être déterminé.

Les expériences de compétition consistent, après 24h d'adaptation de chaque compétiteur au milieu étudié, à mélanger un clone évolué et le clone ancêtre de marqueur Ara opposé (**Figure 27**). Au temps zéro, c'est-à-dire à l'instant du mélange, un premier prélèvement est effectué et une dilution adéquate est étalée sur un milieu solide contenant du tétrazolium et de l'arabinose. Sur ce milieu, les colonies Ara+ apparaissent roses alors que les Ara- apparaissent rouges, ce qui permet de les distinguer et de les dénombrer. Après 24h de co-culture, un nouvel échantillon est prélevé, dilué puis étalé, et les deux compétiteurs sont à nouveau dénombrés. Les proportions des 2 compétiteurs déterminées au début et à la fin de la co-culture permettent de calculer le fitness : c'est le ratio des taux de croissance de ces deux compétiteurs (**Figure 27**).

La trajectoire d'évolution du fitness de chacune des 12 populations a été déterminée au cours du temps, par rapport au clone ancêtre. Le fitness augmente de 70% en moyenne après 20 000 générations d'évolution (Lenski et Travisano, 1994 ; Cooper et Lenski, 2000), avec une augmentation très forte pendant les 2000 premières générations, puis plus lente par la suite (voir figure 1 de l'article suivant). Cette augmentation de fitness est similaire au sein des 12 populations, ce qui révèle une évolution phénotypique parallèle, caractéristique de l'adaptation des bactéries à ces conditions expérimentales.

D'autres phénotypes ont été identifiés sur la base de leur évolution parallèle dans la majorité des 12 populations. Le phénotype le plus visible est une modification de la morphologie cellulaire, associée à une augmentation de la taille et du volume cellulaire, de 50 à 100%, qui suit de manière similaire l'augmentation du fitness (Mongold et Lenski, 1996). Les paramètres de la croissance et les capacités cataboliques des cellules sont également affectés de manière parallèle au cours de l'évolution (Vasi et al., 1994 ; Cooper et Lenski, 2000). En effet,

chacune des 12 populations évoluées révèle une forte augmentation du taux de croissance, ainsi qu'une diminution du temps de latence. D'autre part, la croissance prolongée de ces populations avec du glucose comme seule source de carbone a conduit à une diminution de 42% en moyenne de la capacité de croissance sur une grande diversité de substrats (Cooper et Lenski, 2000). Sur 64 substrats testés, les clones évolués de la majorité des 12 populations présentent une capacité réduite à utiliser 16 sources de carbone après 10 000 générations (Cooper et Lenski, 2000). Cette diminution touche vraisemblablement des fonctions non indispensables dans l'environnement de l'évolution où le glucose est la seule source de carbone. L'adaptation des bactéries se caractérise ainsi par un phénomène de spécialisation écologique. Dans un cas extrême, une perte de fonction a été observée pour les 12 populations : la capacité à utiliser le ribose comme source de carbone. Ceci est dû à des délétions de l'opéron *rbs* d'utilisation du ribose, associées à l'activité d'une séquence d'insertion IS150 (Cooper et al., 2001). Des modifications parallèles de phénotypes au cours de l'évolution représentent une signature de la sélection naturelle, ces changements phénotypiques pouvant alors représenter des modifications bénéfiques dans ces conditions (Cunningham et al., 1997). Il a ainsi été démontré que les délétions de l'opéron ribose étaient effectivement bénéfiques dans ces conditions d'évolution expérimentale, procurant un effet positif d'environ 2% sur le fitness. D'autre part, le taux de mutation est un autre caractère présentant des changements parallèles au cours de cette évolution expérimentale. En effet, 4 des 12 populations présentent un taux de mutation environ 100 fois plus élevé que celui de l'ancêtre (Sniegowski et al., 1997 ; Shaver et al., 2002), suite à des mutations dans les gènes codant les protéines impliquées dans le système de réparation des mésappariements de l'ADN (Friedberg et al., 1995).

Les mutations responsables de l'adaptation des bactéries à leur environnement, et pouvant conférer un fitness accru dans ces conditions, ont été recherchées par 2 types d'approches : l'utilisation de marqueurs génomiques et l'analyse moléculaire de phénotypes modifiés de façon parallèle dans la majorité des 12 populations au cours de l'évolution.

Une analyse de l'évolution des génomes a été effectuée par polymorphisme de longueur de fragments de restriction (RFLP), en utilisant les séquences d'insertion IS connues chez *E. coli* comme marqueurs génomiques au sein de deux populations appelées Ara-1 et Ara+1 (Papadopoulos et al., 1999). Une vingtaine de clones évolués ont été isolés, dans chacune des deux populations, à différents temps d'évolution, et leur ADN génomique a été hybridé avec

les 7 séquences IS présentes chez *E. coli* B utilisées comme sondes moléculaires. Leurs profils d'hybridation ont été comparés à celui du clone ancêtre et des arbres phylogénétiques ont été construits (Papadopoulos et al., 1999). Ceux-ci ont permis la mise en évidence de mutations, détectées de façon précoce pendant les 2000 premières générations, période où l'augmentation de fitness est la plus forte. Ces mutations sont conservées chez tous les descendants dans chacune des deux populations. Elles mettent donc en valeur des gènes qui constituent de bons candidats pour la recherche des mutations bénéfiques et qui ont été caractérisés par la détermination des sites d'insertion des IS correspondantes (Papadopoulos et al., 1999 ; Schneider et al., 2000a).

En plus des délétions de l'opéron ribose (voir plus haut), 4 autres mutations correspondant à des transpositions d'IS ont été mises en évidence. Elles affectent les gènes *pykF* (codant la pyruvate kinase I) (Ponce et al., 1995), *nadR* (codant un régulateur du métabolisme du Nicotinamide Adénine Di-nucléotide) (Tritz et Chandler, 1973), *hokB-sokB* (un locus homologue des systèmes de maintien de plasmides) (Pedersen et Gerdes, 1999), et *pbpA-rodA* (codant les protéines PBP2 et RodA impliquées dans la biosynthèse de la paroi) (Asoh et al., 1986). Les loci *nadR* (Ara+1), *pykF* (Ara-1) et *hokB-sokB* (Ara+1) sont mutés par des transpositions d'IS150 à l'intérieur des séquences codantes, aboutissant probablement à l'inactivation de ces gènes (Schneider et al., 2000a). L'opéron *pbpA-rodA* est muté par l'insertion d'une copie d'IS150 11pb en amont du promoteur, provoquant une dérégulation de l'expression de ces gènes (Philippe, 2006).

Le séquençage de ces 4 gènes dans les 12 populations à 20 000 générations a révélé un fort degré de parallélisme génétique (Woods et al., 2006) : en effet, des mutations ont été détectées dans les gènes *pykF* et *nadR* dans toutes les populations, et dans les gènes *pbpA-rodA* et *hokB-sokB* dans la moitié des populations. Ces résultats ont été comparés aux séquences obtenues pour 36 gènes choisis au hasard chez des clones évolués isolés à 20 000 générations de chacune des 12 populations d'évolution expérimentale (Lenski et al., 2003). Seulement 6 mutations ont alors été détectées dans ces 36 gènes, et uniquement dans les populations mutatrices, présentant un taux de mutation 100 fois supérieur à celui de l'ancêtre. De plus, chacune des 6 mutations n'a été observée que dans une seule population. Le fort parallélisme génétique observé pour les 4 loci *nadR*, *pykF*, *pbpA-rodA* et *hokB-sokB* contraste donc de façon importante avec la quasi-absence de mutations mise en évidence pour le groupe

de 36 gènes choisis aléatoirement. Ceci suggère que les 4 loci sont soumis à la sélection naturelle et que les mutations correspondantes pourraient conférer un avantage sélectif au cours de l'évolution expérimentale (Woods et al., 2006).

La seconde stratégie d'identification de mutations bénéfiques a consisté à mesurer et comparer, par rapport à celui du clone ancêtre, les profils globaux d'expression des gènes de deux clones évolués indépendants, isolés à 20 000 générations des populations Ara-1 et Ara+1. Ces analyses globales d'expression ont été réalisées par détermination des profils globaux de transcription (Cooper et al., 2003) et protéiques (Pelosi et al., 2006). Les deux types d'analyses ont révélé un fort degré de parallélisme, car un nombre important de gènes et de protéines subissent des changements d'expression similaires, allant dans la même direction par rapport au clone ancêtre, et ce dans les deux clones évolués indépendants. En effet, dans ces deux clones évolués, les profils transcriptionnels révèlent la modification parallèle de la transcription de 59 gènes (Cooper et al., 2003), tandis que les profils protéiques montrent une modification parallèle de 38 protéines, sur 300 à 400 visualisées (Pelosi et al., 2006). Conformément à l'hypothèse nulle selon laquelle ces différences reflèteraient des changements aléatoires pendant l'évolution ou la préparation des échantillons, il n'est pas attendu d'observer une telle correspondance au niveau de la direction des changements d'expression des gènes chez les clones évolués. Ainsi, les différences parallèles observées par ces deux approches ont sans doute une signification physiologique et sont vraisemblablement le reflet de mutations bénéfiques qui se sont produites au cours de l'évolution expérimentale. Une analyse plus détaillée des gènes dont l'expression est modifiée de façon parallèle a révélé que la moitié sont connus pour être régulés par (p)ppGpp, l'effecteur de la réponse stringente (Cashel et al., 1996), un réseau de régulation permettant l'adaptation des bactéries à des conditions de carences nutritionnelles. Le séquençage des gènes impliqués directement dans le métabolisme de (p)ppGpp, *relA* et *spoT*, a révélé des mutations dans le gène *spoT* dans 8 des 12 populations, démontrant une fois de plus une évolution génétique parallèle. L'effet sur le fitness apporté par la mutation *spoT* a été mesuré par des expériences de compétitions entre la souche ancêtre dans laquelle une des mutations a été introduite et la souche ancêtre de départ. Cette mutation apporte un avantage considérable, d'environ 10% (Cooper et al., 2003).

Un autre point commun révélé par la comparaison des profils transcriptionnels et protéiques consiste en une forte diminution de l'expression des gènes impliqués dans

l'utilisation du maltose dans les deux populations analysées (Pelosi et al., 2006). Le séquençage du gène *malT*, codant l'activateur transcriptionnel de ces gènes, a révélé la présence de mutations dans 8 des 12 populations, induisant ainsi la diminution parallèle de la capacité à utiliser le maltose dans la majorité des populations. Des expériences de compétitions entre souches ancêtres isogéniques, excepté pour les mutations du gène *malT*, ont démontré leur effet positif sur le fitness d'environ 1,5% (Pelosi et al., 2006).

Seuls 3 types de mutations ont été démontrés comme apportant un bénéfice dans ces conditions d'évolution expérimentale chez *E. coli*. Elles affectent les gènes suivant : *rhs* (avantage de 2%), *spoT* (avantage de 10%), et *malT* (avantage de 1,5%). Si ces mutations ont un effet additif sur le fitness, le bénéfice total apporté n'est que de 13 à 14%, ce qui est très éloigné de la valeur de fitness de 70% en moyenne mesurée dans les clones évolués à 20 000 générations dans les 12 populations (Cooper et Lenski, 2000). De même, les clones évolués isolés à 2000 générations, après la période de forte augmentation du fitness (Lenski et Travisano, 1994), présentent en moyenne un fitness accru de 20 à 25%. Ainsi, de nombreuses mutations bénéfiques sont encore inconnues, et certaines d'entre elles pourraient avoir un effet important sur le fitness.

L'objectif de ma thèse entre dans le cadre de la thématique générale de l'équipe « Dynamique et Evolution des Génomes Microbiens », qui est de comprendre les mécanismes écologiques et moléculaires de l'adaptation des bactéries à leur environnement. En particulier, la recherche de mutations bénéfiques, responsables de l'augmentation du fitness au cours de l'évolution expérimentale chez *E. coli*, en est le point d'orgue.

Je me suis particulièrement intéressée à l'importance de la superhélicité de l'ADN dans les processus adaptatifs, les 12 populations ayant été adaptées à des transitions nutritionnelles pendant plus de 40 000 générations. Ces conditions sont en effet connues pour affecter le degré de superhélicité de l'ADN. Cette étude permettra notamment de répondre aux questions suivantes : la superhélicité peut-elle représenter une cible de la sélection naturelle, c'est-à-dire être impliquée dans les processus évolutifs et adaptatifs ? Par quels types de modifications génétiques ? Un parallélisme phénotypique et génétique est-il sous-jacent à des changements éventuels de superhélicité de l'ADN ? Quels sont les mécanismes impliqués ?

**Global regulatory networks plasticity during experimental evolution in *Escherichia coli*.**

Nadège Philippe,<sup>1</sup> Estelle Crozat,<sup>1</sup> Richard E. Lenski,<sup>2</sup> and Dominique Schneider<sup>1\*</sup>

<sup>1</sup> Laboratoire Adaptation et Pathogénie des Micro-organismes, CNRS UMR5163, Université Joseph Fourier, Grenoble, France.

<sup>2</sup> Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA.

\* Correspondence to: Dominique Schneider, Laboratoire Adaptation et Pathogénie des Micro-organismes, CNRS UMR5163, Université Joseph Fourier, Institut Jean Roget, Domaine de la Merci, BP170, 38042 Grenoble Cedex 9, France. E-mail : dominique.schneider@ujf-grenoble.fr

## Summary

Natural selection and evolution shaped all living organisms on Earth. Strategies of experimental evolution in the laboratory, where a common ancestral cell is propagated for hundreds or thousands of generations in defined environments, provide the opportunity to analyze evolutionary processes “by eye”. The longest running evolution experiment has developed since 1988 and involves propagation of twelve evolving populations from a common ancestor of *Escherichia coli* for 40,000 generations in a constant environment. The hallmark of this evolution is adaptation through parallel phenotypic changes. Gene targets of natural selection were identified, revealing a high level of genetic parallelism. Beneficial mutations affect all levels of gene regulation in the cells: from the lower level of the gene or operon, to the higher level of global regulatory networks. In particular, two well interconnected networks have been involved in these adaptation processes: the DNA superhelicity and stringent response networks.

## Abbreviations

Ara<sup>-</sup>: deficiency to grow on arabinose as a carbon source.

Ara<sup>+</sup>: ability to grow on arabinose as a carbon source.

IS: Insertion Sequence.

NAD: nicotinamide adenine di-nucleotide.

PEP: phosphoenolpyruvate.

Rbs: deficiency to grow on ribose as a carbon source.

RFLP: restriction fragment length polymorphism.

rRNA: ribosomal RNA.

TLC: thin-layer chromatography.

tRNA: transfer RNA.

## Introduction

Natural selection shaped all the beautiful and extraordinary diverse phenotypes of living organisms and mutations are the raw material of these evolutionary processes.<sup>(Darwin 1859)</sup> Bacteria have evolved on Earth for billions of years, leading to both the ability to colonize any known ecological niches, and an extraordinary degree of physiological and metabolic diversity. In their natural environments, bacteria have to cope with and survive to multiple stresses and to repeated switches between conditions of nutritional feast and famine. These adaptation capabilities are usually manifest as phenotypic changes occurring in different environments. At one extreme, some bacteria develop differentiation processes, yielding cell-type diversity. Specialized bacterial cell types, occurring after asymmetric cell division, enable adaptation to starvation and harsh environments. Gram-positive bacteria can form spores, either by engulfment inside the mother cells, like *Bacillus* or *Clostridium* species<sup>(Errington 2003)</sup>, or as exospores at the end of a complex developmental program reminiscent to the biology of filamentous fungi, like *Streptomyces* species.<sup>(Flårdh 2003)</sup> After nitrogen deprivation, some cyanobacteria differentiate into heterocysts, specialized in the fixation of nitrogen.<sup>(Zhang et al. 2006)</sup> Other differentiating bacteria even evolved a socially multicellular life style, like myxobacteria constructing fruiting bodies within which they sporulate.<sup>(Kaiser 2004)</sup> At the other extreme, non-differentiating enteric bacteria, even if they divide symmetrically, undergo morphological and physiological changes when nutrients become exhausted, including cell volume decrease, cell shape modifications, nucleoid compaction, alteration in cell wall composition and in cytoplasm constituents.<sup>(Huisman et al. 1996)</sup> These phenotypic changes are associated with increased resistance to multiple stresses. Some bacterial pathogens also alternate between two phenotypically different states, an infectious non-replicating form promoting transmission to a new host and a non-infectious but metabolically-active replicative form for multiplication inside the host.<sup>(Moulder 1991; Swanson and Hammer 2000)</sup> Although it may take different forms, transitions between favourable growth conditions and stressful environments are characterized by adaptation processes associated with various phenotypic features. These rapid and drastic environmental fluctuations are sensed by complex interconnected regulatory networks, which in turn fine-tune the global gene expression of the genome allowing an adapted response of bacteria to specific or combined stress conditions.<sup>(McAdams et al 2004)</sup> After reversal of the harsh conditions, the structural topology of the regulatory networks, and as a consequence the global gene expression pattern, eventually returns to the initial “pre-stress” state. Genetic regulatory networks have been most intensively studied in *Escherichia coli* and are characterized by a multi-layer structure, from a local control of genes or operons, through regulation of several operons either closely localized on chromosomes (regulons) or not (stimulons), to a global control of genomic expression.<sup>(Hatfield and Benham 2002)</sup> Individual gene or operon regulation is mediated by transcription factors binding to a restricted number of DNA target sites, while regulons or stimulons are controlled by more global regulators able to bind to multiple targets. Regulatory networks are therefore organized as a power-law distribution with a small number of well-connected hub genes, encoding so-called global regulators of gene expression, and many genes with few connections.<sup>(Barabasi and Oltvai 2004)</sup> The highest level of regulation must integrate all the environmental signals and transmit them to express a required set of genes while repressing others in order to precisely adjust the global expression pattern and optimize the bacterial response to the current environmental conditions. The dynamics of the interconnected

network between the global regulators is therefore the key player in bacterial adaptation and evolutionary processes. In this review, we propose that two regulatory networks, the stringent response<sup>(Cashel et al 1996)</sup> and the DNA superhelicity,<sup>(Travers and Muskhelishvili 2005)</sup> constitute these key players able to detect environmental changes and subsequently adjust the overall gene expression pattern leading to the optimal bacterial response to the external stimuli.

While the different hierarchical levels of gene expression are widely studied, the evolutionary dynamics of genetic regulatory networks is largely unknown. To understand this dynamics in bacterial adaptation processes, it is necessary to relate the topology of regulatory networks to the “fitness” feature, which is a measure of ecological success, i.e. the reproductive ability of genotypes. Characterization of the relationship between regulatory networks and improvement of organismal fitness will help to understand how natural selection is able, and at which regulatory level, to shape and improve these networks, leading to well-adapted individuals. At one extreme, one could imagine that bacteria will evolve through modifications of structural genes, leaving intact regulatory genes, while at the other extreme one could envision modifications of regulatory networks from the basal to the highest levels.

The last decade has witnessed the development of experimental evolution strategies using bacteria which were propagated from a common ancestor in defined environments for several hundreds to thousands generations.<sup>(Elena and Lenski 2003)</sup> In this review, we will focus on the longest evolution experiment during which bacteria evolve under conditions of batch cultures, therefore witnessing each day, since 1988, transitions of growth - starvation and starvation-growth.<sup>(Lenski 2004)</sup> This evolution experiment was repeated twelve times, leading to twelve evolving populations derived from a common ancestor for more than 40,000 cell generations. This experiment provides the raw material to relate phenotypic and genomic evolution over this period of time, to study how bacteria adapted to these transitions between growth and starvation, in particular if they improved structural or regulatory elements. The dynamics of regulatory networks in relation to changes in fitness can also be investigated. In this review, we will emphasize on the type of beneficial mutations that allowed bacteria to adapt to their environment and on the evolution of regulatory networks.

## The long-term evolution experiment

A clone of *E. coli* B was used as the common ancestor to found the longest-running evolution experiment.<sup>(LENSKI 2004)</sup> Twelve independent populations were initiated from two variants of this ancestor and propagated by daily serial transfer (a 1/100 dilution) in the same defined glucose-limited environment for more than 40,000 generations.<sup>(Lenski et al. 1991, Lenski and Travisano 1994, Cooper and Lenski 2000)</sup> The growth medium is Davis Minimal medium containing 25mM glucose (DM25). Six populations (called Ara-1 to Ara-6) were initiated from the original ancestor clone, which is unable to use arabinose as a sole carbon source (Ara-), while the six others (Ara+1 to Ara+6) were founded from an Ara+, but otherwise isogenic, revertant of the ancestor. The arabinose utilization phenotype serves as a marker in competition experiments and was shown to be neutral under these conditions.<sup>(Lenski et al 1991)</sup> At each cycle of 24 hours, the evolving cells experienced growth-starvation and starvation-growth transitions. Hence, each day cells undergo lag phase, exponential growth, stationary phase, before being transferred back into fresh medium. Under those conditions, the total size of each of the twelve populations varied between about  $5 \times 10^6$  to about  $5 \times 10^8$  cells. Mixtures of populations were sampled and frozen at regular time intervals during evolution, providing the possibility of resurrecting the frozen

fossils and isolating individuals from any population and time sample. Direct comparisons can therefore be performed between any evolved individual and its progenitor in any type of environments including the one used to propagate the twelve populations. The use of *E. coli* cells for experimental evolution offers multiple advantages, as they reproduce quickly, are easy to grow, count, control and manipulate. Large populations can be obtained supplying detectable mutation diversity. Moreover, *E. coli* has also been used as a model organism for studying its genetics, physiology, metabolic pathways, biochemistry, genomics. A huge amount of information is therefore available together with computer analyses of regulation factors and regulatory networks.

### *Phenotypic evolution*

Fitness values were measured by directly competing evolved populations or individuals with the ancestor carrying the opposite arabinose marker in the same glucose-limited environment.<sup>(Lenski et al 1991)</sup> Fitness is calculated as the ratio of the realized (net) population growth rates obtained for the two strains while they compete with one another. All the replicate populations achieved parallel gains in competitive fitness, of about 70% after 20,000 generations, indicative of substantial adaptation.<sup>(Lenski and Travisano 1994, Cooper and Lenski 2000)</sup> The rate of fitness increase was very fast during the first 2,000 generations, and much slower afterwards, although still significant (Fig. 1A).

Several other phenotypic traits also evolved in parallel in most of the twelve populations. The most visual change corresponds to a substantial increase in cell size in all twelve populations,<sup>(Lenski and Mongold 2000)</sup> with a trajectory very similar to the one observed for evolution of fitness (Fig. 1B). The conditions to which these bacteria adapt include a succession of growth-starvation cycles, and evolved populations are characterized by shorter lag phases and higher maximum growth rates.<sup>(Vasi et al 1994)</sup>

The selection environment consists in a glucose-limiting medium and adaptation to these conditions may have generated trade-offs with decreased ability to grow in other environments, a phenomenon known as ecological specialization. Growth of the evolved populations was quantified on a wide range of substrates (64 informative substrates), revealing a significant and parallel decrease of the catabolic breadth by 42% on average after 20,000 generations.<sup>(Cooper and Lenski 2000)</sup> In one case, a complete loss of the ability to grow was observed, when ribose was used as the sole carbon source. It was suggested that this ecological specialization was mostly accounted for by the mechanism of antagonistic pleiotropy, according to which the same mutations are beneficial in the selection environment and have detrimental effects in other conditions, as opposed to the mechanism of mutation accumulation, where upon the mutations causing decreased growth abilities in other environments are neutral or even slightly deleterious in the glucose selection environment.<sup>(Cooper and Lenski 2000)</sup> This conclusion was confirmed by the discovery of beneficial mutations conferring higher fitness in the glucose environment, but lower fitness when assayed with other sugars (see also section “Parallel changes in global gene expression: evolution of the stringent response regulatory network and of carbon utilization genes”). In particular, as stated above, we found that all twelve populations lost the ability to grow on ribose as a sole carbon source. Moreover, this parallel evolved phenotype (Rbs<sup>-</sup>) occurred very early during the experimental evolution, i.e. during the first period of fast fitness increase,<sup>(Cooper et al 2001)</sup> suggesting that loss of ribose utilization may be linked to mutations conferring increased fitness. We subsequently characterized that this evolved trait was due to

deletions of the ribose-utilization genes, which are clustered within the *rbs* operon and which occurred in all twelve populations. These deletions occurred after transposition events of *IS150* elements within the *rbs* operon followed by a recombination event between each of these newly arising copies and an *IS150* copy present upstream of *rbs* in the ancestral genetic background. (Cooper et al 2001) These IS element activities were also responsible for a high genetic instability of the *rbs* locus, leading to a 10 times higher than expected rate of loss of the ability to use ribose as a carbon source for the ancestor strain. To check if these deletion mutations were advantageous in the selection environment, we constructed an ancestral isogenic strain except for carrying a deletion allele of the *rbs* operon, and performed competition experiments between two ancestor strains carrying either the ancestral or the deleted *rbs* allele. The mutated allele was shown to confer a 2% benefit to the cells. The combination of both the high genetic instability and the selective advantage of the mutations was able to explain the fast and parallel evolution of the Rbs<sup>-</sup> phenotype. (Cooper et al 2001)

### *Chasing mutations*

A simple calculation allows one to estimate the overall number of mutational events that may theoretically have happened in a typical population during 40,000 generations, (Lenski 2004) as the product of the effective population size (adjusted for serial dilution,  $3 \times 10^7$ ), the number of generations (40,000), the base-pair mutation rate ( $5 \times 10^{-10}$  (Drake et al 1998)) and the genome size ( $4.7 \times 10^6$  bp). The total number of mutations during 40,000 generations in each population is expected to be about  $3 \times 10^9$ . Therefore, all possible mutations at each base pair position was available to these evolving populations, without meaning that all mutations really occurred. However, based on estimations on the combined effects of natural selection and random drifts, probably fewer than 100 mutations were substituted in a typical population. (Lenski 2004) This estimate was based on the probability of substitution of a beneficial mutation and on the overall rate of neutral substitution (see (Lenski 2004) for a complete description of this estimation). In four populations, the number of expected mutations will be higher since they developed a mutator phenotype during the course of the evolution, two around generations 2,500 to 3,000 and one around 8,500 generations, (Sniegowski et al 1997) and the last one after 15,000 generations. (Cooper and Lenski 2000) These populations revealed a 100-fold increase in mutation rates and all retained their mutator state at least through generation 20,000. Three of the populations were shown to be defective in the methyl-directed mismatch repair pathways (Sniegowski et al 1997) and mutations were subsequently detected in the *mutL* gene, (Shaver and Sniegowski 2002) whereas the underlying mutation is yet unknown in the fourth population. There is no evidence that these “mutator” mutations confer any advantage in the experimental evolution populations. (Shaver and Sniegowski 2002)

A random sequencing strategy was employed to detect point mutations in these twelve evolving populations. (Lenski et al 2003) The sequence of 36 randomly chosen genes was performed and compared between both ancestors (Ara<sup>-</sup> and Ara<sup>+</sup>) and two clones from each of two time points (10,000 and 20,000 generations) isolated from each of the 12 populations. Only exactly 10 mutations were discovered this way, and only strictly in the populations that were mutators. (Lenski et al 2003) This data set was used to estimate the ancestral mutation rate, about  $1.4 \times 10^{-10}$ . This highlights the difficulty of finding point mutations in this long-term evolution experiment even after 12 times 20,000 generations of independent evolution. Beneficial mutations, which are known to represent only a tiny fraction of all mutations, (Fisher 1930, Kimura 1983) will be even more difficult to identify and investigate. Therefore, to characterize beneficial

mutations, we designed two different strategies: first, we used Insertion Sequence (IS) elements (Chandler and Mahillon 2002) as genotypic markers to follow the genetic diversity over evolutionary time. The IS elements are mobile bacterial elements carrying all genetic information required for their movement. They promote high-frequency mutational events, either through their transposition activity or as substrates for recombination events generating chromosomal rearrangements. (Schneider and Lenski 2004) They are often used as markers for species typing and for epidemiological purposes. Second, we analyzed parallel molecular phenotypes evolving in most of the 12 populations, since parallel and convergent changes in phenotypic traits are widely regarded as hallmarks of adaptive evolution. (Simpson 1953, Cunningham et al 1997) For example, parallel deletions of the *rbs* operon, which were demonstrated to be beneficial under these evolutionary conditions, were the bases of the parallel loss of the ribose utilization phenotype (see section “*Phenotypic evolution*” above).

## Genetic diversity

We analyzed the genetic diversity and genomic evolution of two of the twelve populations over the first 10,000 generations by restriction fragment length polymorphism (RFLP) using the seven IS elements known in *E. coli* B as molecular probes. (Papadopoulos et al 1999) Many evolved clones were isolated at several time points from the two focal populations, called Ara-1 and Ara+1, and their IS genomic fingerprints were compared to the ancestor of the experimental evolution. Construction of phylogenetic trees revealed a huge amount of diversity, almost all individuals tested in each population at 10,000 generations being different from each other (Fig. 2), and contrasting with the failure to detect any point mutation in 36 genes at 20,000 generations in all the non-mutator populations. (Lenski et al 2003) In addition to the *rbs* deletions that were detected (see section “*Phenotypic evolution*” above), five interesting IS-associated mutations were detected, which were substituted early in the populations i.e. during the period of fast fitness increase and were shared by all descendent clones. These “pivotal” mutations therefore constituted potential candidates for conferring higher fitness in these conditions. The underlying mutations were characterized by localizing the corresponding IS elements. (Schneider et al 2000) Two kinds of mutations were discovered: insertions of IS elements after transposition events and chromosomal rearrangements after recombination events between homologous IS copies (Table 1). In population Ara-1, we detected a transposition event of *IS150* within the *pykF* gene, encoding pyruvate kinase I one of the two glycolytic *E. coli* pyruvate kinases catalyzing the transformation of phosphoenolpyruvate (PEP) into pyruvate, (Bledig et al 1996) potentially leading to its inactivation. In population Ara+1, four mutations were analyzed. One consists in an inversion of about one third of the chromosome through recombination between two *IS150* copies, while the three others corresponded each to a transposition event of *IS150*. One *IS150* element jumped 11 bp upstream of the *pbpA-rodA* operon, encoding penicillin-protein 2 and the morphogenic protein RodA both involved in cell wall biosynthesis and cell shape. (Begg and Donachie 1998) The two others corresponded to *IS150* insertions within the coding sequence of two loci, *nadR* and *hokB-sokB*. The *nadR* gene encodes a regulatory protein involved in NAD metabolism by both repressing the NAD biosynthetic genes and influencing the transport of NAD precursors. (Penfound and Foster 1999, Grose et al 2005) The *hokB-sokB* locus is homologous to addiction modules involved in maintenance of plasmids inside bacterial host cells. (Pedersen and Gerdes 1999)

The IS transposition in each of these four genes (*pykF*, *pbpA-rodA*, *nadR*, and *hokB-sokB*) was detected in one of the two analyzed populations.<sup>(Papadopoulos et al 1999, Schneider et al 2000)</sup> We sequenced the same four genes in all the other 11 populations in evolved clones isolated after 20,000 generations.<sup>(Woods et al 2006)</sup> Mutations were substituted in both *pykF* and *nadR* in all 12 populations, and in half of them in both *pbpA-rodA* and *hokB-sokB*. This is in marked contrast with the sequencing of the 36 random genes, where no mutation was found in the 8 non-mutator populations.<sup>(Lenski et al 2003)</sup> Random genes had substitutions only in the four mutator populations, and no single gene was affected in more than one population. Parallel evolution is a hallmark of adaptation through natural selection,<sup>(Cunningham et al. 1997)</sup>, suggesting that the genetic changes in the four candidate genes may be adaptive in these conditions. Four different statistical tests were performed using both sets of sequence data (the 36 random genes and the four candidate loci:<sup>(Woods et al 2006)</sup> i) overall substitution rates were higher for the candidate genes, ii) an excess of non synonymous over synonymous mutations was measured for the candidate genes, iii) a higher trend of mutations in random genes was observed in the mutator populations compared to the candidate genes, iv) the number of mutations in candidate genes follow a uniform distribution different from a Poisson distribution. These four statistical tests all revealed that natural selection drove parallel genetic changes in the four genes, and that these mutations are therefore beneficial in this experimental evolution context.

Several hypotheses can be suggested for the functional bases of the benefit that may be conferred by the different mutations.<sup>(Schneider et al. 2000, Woods et al. 2006)</sup> The *pykF* mutations may be linked to a decreased conversion of PEP to pyruvate, leading to an increase availability of PEP, which constitutes the energy source of the PTS-transport system of the only available nutrient, glucose. The *nadR* mutations may decrease NadR repressor activity and therefore increase the intracellular levels of NAD, which may be required in the evolved clones since they show increased growth rates. Inactivation of the *hokB-sokB* locus, encoding a toxin-antitoxin-like addiction module may be beneficial to the cells. The last mutations, in the regulatory region of *pbpA-rodA*, may change the levels of these two proteins which are involved in the coupling between cell wall biosynthesis and cell division. Improvement of this coupling process may be required since the 12 populations evolved larger cell size and volume (see section “*Phenotypic evolution*”).

In summary, natural selection targeted genes involved in central metabolism and cell wall biosynthesis during long-term experimental evolution in *E. coli*. Parallel genetic changes in these genes may explain some of the parallel phenotypic modifications that occurred in these populations.

## Parallel genomic evolution

### *Parallel changes in DNA supercoiling*

The 12 populations experienced each day since 1988 cycles of growth, starvation and re-growth, conditions known to affect the DNA supercoiling level in bacterial cells.<sup>(Hatfield and Benham 2002, Reyes-Dominguez et al. 2003)</sup> We measured the DNA supercoiling state of a reporter plasmid which was introduced into three clones from each of the 12 populations at each of three time points: 2,000, 10,000, and 20,000 generations.<sup>(Croizat et al. 2005)</sup> Substantial parallel changes were detected since 10 out of the 12 populations revealed increased DNA supercoiling levels, with

modifications starting before 2,000 generations in most populations and still increasing over evolutionary time. These parallel increases in the level of DNA supercoiling may therefore reflect adaptive changes, with the underlying mutations conferring enhanced fitness to the cells. The DNA supercoiling level is a precisely controlled essential process, and two types of enzymes are involved in maintaining an adequate superhelicity of the chromosome: topoisomerases,<sup>(Champoux 2001)</sup> and histone-like proteins.<sup>(Dorman and Deighan 2003)</sup> Sequencing relevant genes in the model population Ara-1, which presents two successive DNA supercoiling increases (one before 2,000 and one after 10,000 generations), revealed two mutations in the *topA* and *fis* genes.<sup>(Croizat et al. 2005)</sup> The *topA* gene encodes topoisomerase I involved in the relaxation of DNA by removing negative supercoils<sup>(Wang 1971)</sup> and also involved in the coupling process between transcription and translation by removing hypernegative supercoils behind the transcribing RNA polymerase molecules.<sup>(Drolet 2006)</sup> The *fis* gene encodes a histone-like protein, which is also a global regulator of gene expression, including ribosomal and tRNA genes, coupling the genetic and physiological states of cells to the nutritional conditions.<sup>(Mallik et al. 2004)</sup> The Fis protein adjusts DNA supercoiling by directly binding to DNA and by repressing the *gyrAB* genes encoding DNA gyrase which is the only enzyme able to introduce negative supercoils into DNA molecules.<sup>(Schneider et al. 1999)</sup> Analysis of the temporal dynamics of the appearance and substitution of both mutations revealed the fixation of the *topA* evolved allele well before 2,000 generations, whereas the *fis* evolved allele was substituted after 10,000 generations, well correlating with the two successive steps of DNA supercoiling increase observed in this focal population. A set of isogenic strains was constructed by moving each evolved allele of *topA* and *fis*, alone and in combination, in the ancestral genetic background.<sup>(Croizat et al. 2005)</sup> Competition experiments between these ancestral strains differing only in the *topA* and/or *fis* alleles revealed that both mutations were beneficial and contributed additively to the increased DNA supercoiling, reproducing precisely the net change that was observed in the population. The *topA* evolved allele therefore was responsible for a decreased activity of the topoisomerase I, and we showed that the evolved *fis* mutation led to a decrease of the level of this global regulator. We therefore demonstrated that DNA supercoiling was an important target of selection during this evolution experiment, which suggests that global regulatory networks may have been improved to allow increased fitness. The extent of genetic parallelism in these genes is currently under investigation.

### *Parallel changes in global gene expression: evolution of the stringent response regulatory network and of carbon utilization genes*

Global gene expression profiles were compared between the ancestor and two independently evolved clones sampled at 20,000 generations from populations Ara-1 and Ara+1. Several hundreds to thousands of phenotypic expression traits can be simultaneously assessed by these global expression analyses. Two different methodologies and sets of conditions were used: DNA microarrays, with total RNAs extracted from cells sampled during exponential phase in the same medium as the one used during experimental evolution;<sup>(Cooper et al. 2003)</sup> and two-dimensional protein electrophoresis with proteins extracted from cells isolated after stationary phase in the same medium but containing a ten-time higher glucose concentration.<sup>(Pelosi et al. 2006)</sup> Whereas DNA microarrays allows one to detect the transcription of almost all *E. coli* genes under a particular condition, a typical proteomic profile separates about 300 to 400 proteins.

Despite two different methodologies and two completely different environmental conditions (exponential versus stationary phase, two different glucose concentrations), an exceptional level of parallelism was detected in the expression profiles of both evolved clones compared to the ones obtained from the ancestor. The transcription profiles revealed parallel changes in the expression of 59 genes after 20,000 generations of experimental evolution,<sup>(Cooper et al. 2003)</sup> while 38 proteins showed quantitative parallel changes on the global protein profiles.<sup>(Pelosi et al. 2006)</sup> Remarkably, all the transcriptional and protein parallel changes evolved in the same direction compared to the ancestral expression level, reflecting their physiological and adaptive significance. Each of these parallel changes per se is unlikely due to a particular adaptive mutation. Most of these parallel changes may rather represent a coordinated response to regulatory modifications, while some of them may be linked to more specific mutations. Since the transcriptional and protein global analyses were performed under two strikingly different environments, most genes affected by parallel changes for the two methodologies were different.<sup>(Pelosi et al. 2006)</sup> However, a close examination revealed that half of them belong to the same regulatory network, i.e. the stringent response which is involved in adaptation to nutritional stresses.<sup>(Cashel et al. 1996)</sup> The effector of the stringent response is the nucleotide guanosine tetraphosphate, called ppGpp, whose concentration increases in the cells during transition from feast to starvation conditions. Following starvation, this signal molecule reprograms the global transcription profile in the cell by inhibiting transcription of genes involved in growth (for example synthesis of stable RNAs) while activating transcription of genes necessary for survival like the amino-acid biosynthetic operons and the *ipoS* gene, encoding the stationary phase-specific sigma factor RpoS.<sup>(Hengge-Aronis 2002)</sup> Sequencing genes known to be involved in the metabolism of ppGpp revealed a mutation in the *spoT* gene, encoding a protein with both synthesis and degradation activities of ppGpp,<sup>(Gentry and Cashel 1996)</sup> in the population Ara-1. Introducing this particular mutation in the genomic context of the ancestor, and competing both ancestor clones carrying the ancestral and evolved *spoT* alleles, respectively, revealed a strong beneficial effect conferred by the *spoT* evolved allele.<sup>(Cooper et al. 2003)</sup> Therefore, bacterial adaptation occurred through improved regulatory networks and this evolutionary change caused parallel gene expression modifications, with different subsets of genes belonging to the stringent response network affected by the two growth conditions used in the transcriptomic and proteomic studies.

Despite very little strict identity in the gene targets detected by both methodologies, two parallel gene expression changes were common to the macroarrays and proteomics. Hence, decrease in the expression of the ribose operon and the maltose regulon was detected in both cases.<sup>(Pelosi et al. 2006)</sup> We already demonstrated previously the adaptive loss of the ribose consumption capacities under these experimental evolution conditions.<sup>(Cooper et al. 2001)</sup> We therefore further investigated the molecular bases of the evolved decreased ability to use maltose as a carbon source. Mutations in *malT*, the transcriptional activator of the maltose operons,<sup>(Danot 2001)</sup> were discovered in both focal Ara-1 and Ara+1 populations.<sup>(Pelosi et al. 2006)</sup> Isogenic ancestral strains were constructed by introducing each of these *malT* mutations within the chromosome of the ancestor. Competition experiments between ancestor clones carrying the ancestral and each evolved *malT* alleles revealed that both mutations enhanced fitness in these conditions.

The *spoT* and *malT* genes were subsequently sequenced in all twelve evolving populations.<sup>(Cooper et al. 2003, Pelosi et al. 2006)</sup> Mutations were found in *spoT* in eight populations (Fig. 3), and the same situation occurred for *malT* although not in the same precise eight populations.

As for the DNA superhelicity, parallel genetic changes affecting another global regulatory network, i.e. the stringent response, is one of the key evolutionary pathways in this experimental evolution. Parallel genetic changes in the stringent response regulatory network underlie the parallel phenotypic changes observed in the global gene expression patterns. Likewise, at least some of the parallel catabolic decreases detected during evolution of these bacterial populations <sup>(Cooper et al. 2000)</sup> can be explained by parallel mutations in the corresponding sugar utilization operons, which are beneficial in the glucose environment. These mutations confirm that these populations evolved by ecological specialization through, at least partially, the mechanism of antagonistic pleiotropy (see section “*Phenotypic evolution*”).

Comparison of global expression profiles in independently evolving individuals highlighted adaptation through parallel genetic modifications at two levels of gene regulation: the lower level of gene regulation, i.e. the gene or operon level, and the higher level of global gene regulatory network. While mutations affecting the first level generated clear-cut phenotypes (decrease of catabolic functions), it is more difficult to predict the phenotypic consequences of substitutions affecting the second level. We therefore decided to investigate into more details the *spoT* mutations.

## Evolutionary modifications of the stringent response regulatory network

### *The stringent response*

The stringent response is a regulatory network allowing bacteria to cope with changing environments, in particular with different nutritional states. <sup>(Cashel et al. 1996)</sup> The effectors of this response are two nucleotide molecules, ppGpp which is derived from pppGpp by hydrolysis. Both molecules are collectively called (p)ppGpp. In response to various starvation conditions, the levels of (p)ppGpp increase intracellularly through the action of two enzymes RelA and SpoT, encoded by the *relA* and *spoT* genes. The (p)ppGpp effector is then able to reprogram the entire transcription machinery of the cell, through direct interactions with RNA polymerase and the help of protein partners like DksA. <sup>(Magnusson et al. 2005)</sup> These overall transcription changes result in inhibition of the synthesis of stable RNAs, together with all processes involved in cell growth. At the same time, (p)ppGpp leads to a stimulation of transcription of genes necessary for survival like the amino-acid biosynthetic operons or the stress survival gene encoding RpoS, an alternative sigma factor. This coordinated response allows cells to survive under starvation conditions. After starvation, when nutrients become available, (p)ppGpp is degraded by the SpoT protein, <sup>(Gentry and Cashel 1996)</sup> allowing transcription to promote growth again.

The metabolism of (p)ppGpp is directed essentially by two proteins, RelA and SpoT. The RelA protein is a ribosome-associated (p)ppGpp synthetase involved in synthesis as a response to amino-acid starvation, the induction signal being accumulation of uncharged tRNAs. <sup>(Wendrich et al. 2002)</sup> The SpoT enzyme is a bifunctional (p)ppGpp synthetase and hydrolase, both activities being localized at the N terminal part of the protein, <sup>(Gentry and Cashel 1996)</sup> followed by a C terminal domain (CTD) of unknown function (Fig. 3), suggested to be involved in the regulation of both activities. <sup>(Mechold et al. 2002)</sup> By contrast to RelA which is specific to amino-acid starvation, SpoT is involved in (p)ppGpp regulation in response to most starvation conditions, but the SpoT-dependent signals leading to (p)ppGpp synthesis or degradation are mostly unknown.

### *Evolution of the stringent response during 20,000 generations*

Six out of the eight *spoT* mutations affected the CTD of the protein. We further investigated three of these mutations: the originally identified mutation in population Ara-1, and the two mutations substituted in populations Ara+2 and Ara-4 (Fig. 3). These three mutations affect amino-acid residues localized at the extremities and at the middle of the CTD domain of SpoT. Investigating the substitution dynamics of these three *spoT* mutations in their respective populations revealed that all three were fixed well before 2,000 generations during the period of strong fitness increase. Three isogenic ancestral strains carrying each of the *spoT* evolved alleles were constructed and compared to the ancestor bearing the ancestral *spoT* allele. Competition experiments, conducted between each of these strains and the ancestor in the same environment as the evolution experiment, confirmed the high benefit conferred by the *spoT* mutation of population Ara-1,<sup>(Cooper et al. 2003)</sup> and demonstrated that the two other mutations were also highly beneficial in these conditions with a fitness effect of about 10% (Fig. 4A). The same competition experiments were performed but samples were removed and plated at various time points during growth and not only after 24 hours of competitions, to dissect whether the *spoT*-linked benefit was associated with the lag, exponential or stationary phases. The entire benefit conferred by all three evolved alleles was shown to be specific to growth, i.e. to exponential phase. These results highlighted the strong level of genetic and phenotypic parallelism in these evolving populations. The same increase in fitness was conferred by the three *spoT* evolved alleles in MOPS minimal medium containing 0.2% glucose, which is the medium used for measurements of (p)ppGpp levels (see next).

Using the same four strains, we carried out three types of quantitative assays of (p)ppGpp levels. After cultures of all strains in minimal MOPS medium containing 0.2% glucose as the sole carbon source and labeling with [<sup>32</sup>P]H<sub>3</sub>PO<sub>4</sub>, nucleotides were extracted by formic acid and samples were resolved by KH<sub>2</sub>PO<sub>4</sub> one-dimensional thin-layer chromatography (TLC).<sup>(Cashel 1994)</sup> The first assay consists in measuring accumulation of (p)ppGpp by artificially provoking the stringent response after addition of serine hydroxamate which inhibits tRNA<sup>Ser</sup> aminoacylation therefore leading to a serine starvation.<sup>(Tosa and Pizer 1971)</sup> Serine hydroxamate (0.2 mg/mL) was added to [<sup>32</sup>P]H<sub>3</sub>PO<sub>4</sub>-labelled cultures of the four strains, and we removed samples every 5 minutes for 30 minutes. Accumulation of (p)ppGpp was followed by extracting nucleotides and resolving them on TLC plates, with mutants devoid of pppGpp and/or ppGpp as controls. Three independent experiments were performed and (p)ppGpp accumulation over starvation time was calculated as the ratio of the relative intensity of the ppGpp spots to the GTP + pppGpp + ppGpp spot intensities. The second assay consists in following (p)ppGpp degradation. Induction of (p)ppGpp accumulation was performed exactly as described with serine hydroxamate addition, followed by a reversal of the starvation by adding serine (2 mg/mL), chloramphenicol (0.1 mg/mL) and a mixture of the 20 amino-acids (0.1 mg/mL each). The reversal of starvation therefore leads to rapid degradation of (p)ppGpp through the activity of SpoT. Immediately after starvation, we sampled cells every 10 seconds during 2 minutes for each of the four strains, and the samples were treated as described previously. Three independent experiments were performed and accurate half-lives of (p)ppGpp can be calculated from straight-line slopes of decay on semi-logarithmic graphs, because the rates of decay are first-order reactions.<sup>(Cashel 1994)</sup> The third measurements consisted in sampling cells from the four strains bearing the four *spoT* alleles during growth in [<sup>32</sup>P]H<sub>3</sub>PO<sub>4</sub>-labelled MOPS minimal medium. Samples were removed every 30 minutes during

12 hours, and subjected to the extraction procedures and to TLC. All three biochemical measurements of (p)ppGpp accumulation, degradation and levels revealed no differences whatsoever between the ancestral strains bearing the ancestral and evolved *spoT* alleles. The high fitness increase conferred by the three mutations are therefore independent of the intracellular (p)ppGpp levels. To confirm this, we introduced the four *spoT* alleles (ancestral and three evolved) in a 606 ancestral genetic background where we previously deleted the *relA* gene. These four strains have therefore very low levels of (p)ppGpp and we competed them against the Ara+ variant of the ancestor. These competition experiments confirmed a high benefit of the *spoT* evolved alleles even in this *relA*-deleted background. Therefore, the *spoT* evolved alleles conferred higher fitness independently of the (p)ppGpp levels.

Since the stringent response negatively regulates the synthesis of ribosomal RNAs upon transition into stationary phase, we determined the intracellular content of DNA, RNAs and proteins in the four previous strains (Fig. 4B). Although a stringent control is clearly observed in the four strains upon entry into starvation, the ratios of DNA, RNA and protein per cell in stationary phase relative to growth phase is higher when each of the three *spoT* evolved alleles is present, compared to the *spoT* ancestral allele (Fig. 4B). In other words, it seems that the *spoT* evolved alleles conferred a very mild relaxed phenotype in those strains. This fine-tuning leaves the strain in a “growth state” even upon entry into stationary phase. This is reminiscent to what happens in a strain devoid of (p)ppGpp,<sup>(Magnusson et al. 2003)</sup> although not to the same extent since we observed no changes in the (p)ppGpp levels. Cells containing higher levels of macromolecules in stationary phase should be at a disadvantage upon prolonged starvation. This is indeed the case, since isogenic strains bearing each of the *spoT* evolved alleles have lower fitness during competitions with the ancestral strain under prolonged starvation conditions.

## Conclusions

The hallmark of evolution of the bacterium *E. coli* during 40,000 generations in a constant environment is a very high level of parallelism, that can be characterized at various levels. At the phenotypic level, all populations revealed increased fitness associated with changes in the morphology and size of the cells. Most populations also revealed consistent changes in catabolic abilities, global gene expression patterns, and DNA supercoiling levels. Investigation of some of these parallel phenotypic changes uncovered several beneficial mutations and revealed a high level of genetic parallelism, the same genes being the targets of natural selection in most evolving populations. Despite this high genetic parallelism at the gene level, the mutations discovered were different in the various populations. Analysis of different mutations in some target genes, like *malt* or *spoT*, however uncovered a third level of parallelism: the molecular level, the different mutations leading to similar effects. Both genetic and molecular parallelisms underlie the strong parallelism observed at the phenotypic level. This is in marked contrast with sequences of randomly-chosen genes in evolved clones isolated at 20,000 generations from all twelve populations, where mutations were found only in the four mutator populations; and no single mutation was observed in more than one population.

Beneficial mutations discovered can be classified into two main categories, targeting either the lower regulation level in the cell, i.e. the gene or operon level, or the larger regulation level, i.e. the global regulator level. This reflects the overall architecture of regulatory interactions

in many organisms, with many genes with one or a few connections and a small number of well-connected “hub” genes.<sup>(Barabasi and Oltvai 2004)</sup> All our data suggest that the evolutionary ability of bacterial organisms acts by plasticity of all levels of gene regulation, from a single low-connected gene whose function may be improved (or cancelled) to give a better fit to the current environment to a highly connected “hub” gene whose modification will have profound consequences onto the overall cell physiology. The first class of beneficial mutation-containing genes affects cellular functions involved in carbon source catabolism (*rbs*, *malT*), cell wall biosynthesis (*pbpA*), and metabolism (*pykF*, *nadR*). Some of them explain parallel phenotypic changes that were observed during long-term evolution. For example, repeated deletions of the *rbs* operon and mutations within *malT* are the molecular bases of the reduced ability to use ribose and maltose as carbon sources of the evolved clones. The *pbpA* mutations may be the bases of the size and morphology modifications, and mutations in *pykF* and *nadR* may explain better growth in the medium used in the evolution experiment with glucose as the sole carbon source.<sup>(Schneider et al. 2000, Woods et al. 2006)</sup>

The second class of mutations affect two global regulatory networks which are well conserved among bacteria and even in eukaryotic cells: the DNA supercoiling and the stringent response. Both networks are able to sense environmental fluctuations and induce modifications of the global gene expression pattern to respond to the external conditions.<sup>(Cashel et al. 1996, Magnusson et al. 2005, Hatfield and Benham 2002, Travers and Muskhelishvili 2005)</sup>

The long-term evolution experiment provides an exquisite training ground for bacteria to adapt to transitions growth-starvation and starvation-regrowth. Adaptation occurred through changes in genes known to be global regulators under these conditions: *topA*, *fis*, and *spoT*. During growth, synthesis of stable RNAs (rRNA and tRNA) is maximal and directly related to the growth rates, whereas it strongly decreases upon transition to stationary phase.<sup>(Bremer and Dennis 1996)</sup> Regulation of this tight control is performed mainly through (p)ppGpp which represses transcription from the rRNA promoters,<sup>(Cashel and Gallant 1969, Travers 1976)</sup> the level of DNA superhelicity which is required to be highly negative,<sup>(Lamond 1985)</sup> and the Fis histone-like protein which binds to them for activation.<sup>(Muskhelishvili et al. 1995)</sup>

There is a high regulatory connection between these different levels of regulation of stable RNA synthesis, and it is therefore striking that all three revealed modifications during 40,000 generations of bacterial adaptation. The (p)ppGpp molecules were shown to destabilize the stability of the transcription initiation complexes, which is particularly drastic on stringent promoters, i.e. promoters regulated negatively by (p)ppGpp, which are already intrinsically unstable.<sup>(Zhou and Jin 1998)</sup> A crucial determinant of this negative regulation is a specific sequence, called the discriminator, which is localized between the -10 promoter box and the transcription initiation site and which is characterized by a high GC content for stringent promoters.<sup>(Travers 1980)</sup> At the same time, these stringent promoters require high negative DNA superhelicity to be optimally transcribed, since this high GC discriminator sequence provides a high energy barrier, that can be overcome by high DNA supercoiling.<sup>(Pemberton et al. 2000)</sup> During growth, DNA superhelicity is negative and high and there is no (p)ppGpp, conditions favoring transcription of rRNA. During transition into stationary phase, (p)ppGpp levels strongly and rapidly increase and the DNA is more relaxed,<sup>(Reyes-Dominguez et al. 2003)</sup> conditions inhibiting transcription of rRNA. Under starvation conditions, (p)ppGpp also have a positive effect on transcription of other genes, like the amino-acid biosynthetic operons or genes under the control of alternative sigma factors.<sup>(Magnusson et al. 2005)</sup> These positive effects have been proposed to be related to mainly two mechanisms, a passive<sup>(Barker et al. 2001)</sup> and an active<sup>(Ishage et al. 2002)</sup> model of positive regulation by (p)ppGpp. During growth, most RNA polymerase molecules are

engaged in transcription of the stable rRNA operons, whereas transition into stationary phase leads to destabilization of the transcription open complexes on the stringent promoters through the action of (p)ppGpp. The passive model for activation through (p)ppGpp is therefore linked to the subsequent availability of RNA polymerase molecules to transcribe promoters for which they have much less affinity, for example the amino-acid biosynthetic operons.<sup>(Barker et al. 2001)</sup> Associated with this process, (p)ppGpp is also involved in sigma factor competition for the RNA polymerase core enzyme, by increasing the binding ability of alternative sigma factors ( $\sigma^S$ ,  $\sigma^{32}$ , and  $\sigma^{54}$ ) compared to the vegetative  $\sigma^{70}$  factor.<sup>(Jishage et al. 2002, Kvint et al. 2000, Laurie et al. 2003)</sup> Therefore, (p)ppGpp is actively involved in gene expression alterations by inducing a shift in the ability of sigma factors to recruit RNA polymerase enzyme. Collectively, the negative and positive effects of (p)ppGpp reprograms the entire transcription machinery in the cell during the transitions between nutritional states, leading in particular to a decrease in the quantity of stable RNAs, and consequently of the total RNAs, and of ribosomes upon entry into stationary phase.

During 40,000 generations of experimental evolution of *E. coli* cells, the beneficial *spoT* mutations, although having no effect on the overall synthesis and degradation of (p)ppGpp, are responsible for a reduced decrease of the quantity of RNAs and proteins upon transition into stationary phase, compared to the ancestral strain. This is reminiscent to a relaxed phenotype, with a stringent response having a milder effect. Although the mechanism is still unknown, these *spoT* evolved alleles seem to counter-balance the effects of the (p)ppGpp effectors. This phenotypic effect may “lock” the cells into a growth state which may confer a selective advantage under these conditions where nutrients are supplied each day. Also, the beneficial increase of the DNA superhelicity level may amplify this phenomenon by facilitating transcription through the discriminator sequence of stringent promoters. Evolution of *E. coli* under these growth-starvation transition conditions therefore involved the improvement of the interplay between these two highly connected regulatory networks: the DNA superhelicity and the stringent response. Dissecting the mechanisms that allow improvement of bacterial regulatory circuits is of crucial importance to understand how natural selection shapes fitter individuals as a function of environmental conditions. The experimental evolution populations offer unique opportunities to quantify dynamics in a complex adaptive system and elucidate the interplay among these regulatory mutations. Hence, we are in the process of reconstructing and analyzing the overall evolutionary history of one model population by successively introducing the various beneficial mutations into the ancestral background.

## Acknowledgements

We thank Université Joseph Fourier and Centre National de la Recherche Scientifique for financial support. NP acknowledges the french ministry of « Education Nationale, de l'Enseignement Supérieur et de la Recherche » for a research fellowship and the « Université Joseph Fourier » for an ATER funding. EC thanks the Centre National de la Recherche Scientifique for a BDI fellowship.

## References

- Barabasi A-L, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. *Nature Rev Genet* 5:101-113.
- Barker MM, Gaal T, Gourse RL. 2001. 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* 305:689-702.
- Begg KJ, Donachie WD. 1998. Division planes alternate in spherical cells of *Escherichia coli*. *J Bacteriol* 180:2564-2567.
- Bledig SA, Ramseier TM, Saier Jr MH. 1996. FruR mediates catabolite activation of pyruvate kinase (*pykF*) gene expression in *Escherichia coli*. *J Bacteriol* 178:280-283.
- Bremer H, Dennis PP. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umberger HE. ed; *Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press: Washington DC. p 1553-1569.
- Cashel M, Gallant J. 1969. Two compounds implicated in the function of the RC gene of *Escherichia coli*. *Nature* 221:838-841.
- Cashel M, Gentry VJ, Hernandez VJ, Vinella D. 1996. The stringent response. In Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umberger HE. ed; *Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press: Washington DC. p 1458-1496.
- Cashel M. 1994. Detection of (p)ppGpp accumulation patterns in *Escherichia coli* mutants. In Adolph KW. ed; *Methods in molecular genetics* vol 3 part A. New York NY: Academic Press. p 341-356.
- Champoux JJ. 2001. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* 70:369-413.
- Chandler M, Mahillon J. 2002. Insertion sequences revisited. In Craig NL, Craigie R, Gellert M, Lambowitz AM. ed; *Mobile DNA II*. Washington DC: American Society of Biology. p 305-366.
- Cooper TF, Rozen DE, Lenski RE. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc Natl Acad Sci USA* 100:1072-1077.
- Cooper VS, Lenski RE. 2000. The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407:736-739.
- Cooper VS, Schneider D, Blot M, Lenski RE. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *J Bacteriol* 183:2834-2841.
- Crozat E, Philippe N, Lenski RE, Geiselman J, Schneider D. 2005. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169:523-532.
- Cunningham CW, Jeng K, Husti J, Badgett M, Molineux IJ, Hillis DM, Bull JJ. 1997. Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. *Mol Biol Evol* 14:113-116.
- Danot O. 2001. A complex signalling module governs the activity of MalT, the prototype of an emerging transactivator family. *Proc Natl Acad Sci USA* 98:435-440.
- Darwin C. 1859. *On the origin of species by means of natural selection*. Murray, London.
- Dorman CJ, Deighan P. 2003. Regulation of gene expression by histone-like proteins in bacteria. *Curr Opin Genet Dev* 13:179-184.
- Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. rates of spontaneous mutation. *Genetics* 148:1667-1686.
- Drolet M. 2006. Growth inhibition mediated by excessive negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol Microbiol* 59:723-730.
- Elena SF, Lenski RE. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Rev Genet* 4:457-469.
- Errington J. 2003. Regulation of endospore formation in *Bacillus subtilis*. *Nature Rev Microbiol* 1:117-126.
- Fisher RA. 1930. *The genetical theory of natural selection*. Oxford:Oxford University Press.

- Flårdh K. 2003. Growth polarity and cell division in *Streptomyces*. *Curr Opin Microbiol* 6:564-571.
- Gentry DR, Cashel M. 1996. Mutational analysis of the *Escherichia coli spoT* gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation. *Mol Microbiol* 19:1373-1384.
- Grose JH, Bergthorsson U, Roth JR. 2005. Regulation of NAD synthesis by the trifunctional NadR protein of *Salmonella enterica*. *J Bacteriol* 187:2774-2782.
- Hatfield GW, Benham CJ. 2002. DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu Rev Genet* 36:175-203.
- Hengge-Aronis R. 2002. Signal transduction and regulatory mechanisms involved in control of the  $\sigma^S$  (RpoS) subunit of RNA polymerase. *Microbiol Mol Biol Rev* 66:373-395.
- Huisman GW, Siegele DA, Zambrano MM, Kolter R. 1996. Morphological and physiological changes during stationary phase. In Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umberger HE. ed; *Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press: Washington DC. p 1672-1682.
- Jishage M, Kvint K, Shingler V, Nyström T. 2002. Regulation of sigma factor competition by the alarmone ppGpp. *Genes Dev* 16:1260-1270.
- Kaiser D. 2004. Signaling in myxobacteria. *Annu Rev Microbiol* 58:75-98.
- Kimura M. 1983. The neutral theory of molecular evolution. Cambridge:Cambridge University Press.
- Kvint K, Farewell A, Nyström T. 2000. RpoS-dependent promoters require guanosine tetraphosphate for induction even in the presence of high levels of sigma(S). *J Biol Chem* 275:14795-14798.
- Lamond AI. 1985. Supercoiling response of a bacterial tRNA gene. *EMBO J* 4:501-507.
- Laurie AD, Bernardo LM, Sze CC, Skarfstad E, Szalewska-Palasz A, Nystrom T, Shingler V. 2003. The role of the alarmone (p)ppGpp in sigma N competition for core RNA polymerase. *J Biol Chem* 278:1494-1503.
- Lenski RE, Mongold J. 2000. cell size, shape, and fitness in evolving populations of bacteria. In Brown JH, West GB. ed; *Scaling in biology*. Oxford: Oxford University Press. p 221-235.
- Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat* 138:1315-1341.
- Lenski RE, Travisano M. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc Natl Acad Sci USA* 91:6808-6814.
- Lenski RE, Winkworth CL, Riley MA. 2003. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *J Mol Evol* 56:498-508.
- Lenski RE. 2004. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. *Plant Breed Rev* 24:225-265.
- Magnusson LU, Farewell A, Nyström T. 2005. ppGpp: a global regulator in *Escherichia coli*. *Trends Microbiol* 13:236-242.
- Magnusson LU, Nyström T, Farewell A. 2003. Underproduction of  $\sigma^{70}$  mimics a stringent response. *J Biol Chem* 278:968-973.
- Mallik P, Pratt TS, Beach MB, Bradley MD, Undamatla J, Osuna R. 2004. Growth-phase dependent regulation and stringent control of *fis* are conserved processes in enteric bacteria and involve a single promoter (*fis P*) in *Escherichia coli*. *J Bacteriol* 186:122-135.
- McAdams HH, Srinivasan B, Arkin AP. 2004. The evolution of genetic regulatory systems in bacteria. *Nature Rev Genet* 5:1-9.
- Mechold U, Murphy H, Brown L, Cashel M. 2002. Intramolecular regulation of the opposing (p)ppGpp catalytic activities of Rel<sub>Seq</sub>, the Rel/Spo enzyme from *Streptococcus equisimilis*. *J Bacteriol* 184:2878-2888.
- Moulder JW. 1991. Interaction of *Chlamydiae* and host cells *in vitro*. *Microbiol Rev* 55:143-190.
- Muskhelishvili G, Travers AA, Heumann H, Kahmann R. 1995. FIS and RNA polymerase holoenzyme form a specific nucleoprotein complex at a stable RNA promoter. *EMBO J* 14:1446-1452.
- Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE, Blot M. 1999. Genomic evolution during a 10,000-generation experiment with bacteria. *Proc. Natl. Acad. Sci. USA* 96:3807-3812.

- Pedersen K, Gerdes K. 1999. Multiple *hok* genes on the chromosome of *Escherichia coli*. *Mol Microbiol* 32:1090-1102.
- Pelosi L, Kühn L, Guetta D, Garin J, Geiselman J, Lenski RE, Schneider D. 2006. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics* 173:1851-1869.
- Pemberton IK, Muskhelishvili G, Travers AA, Buckle M. 2000. The GC-rich discriminator region of the *tyrT* promoter antagonises the formation of stable pre-initiation complexes. *J Mol Biol* 299:859-864.
- Penfound T, Foster JW. 1999. NAD-dependent DNA-binding activity of the bifunctional NadR regulator of *Salmonella typhimurium*. *J Bacteriol* 181:648-655.
- Reyes-Dominguez Y, Contreras-ferrat G, Ramirez-Santos J, Membrillo-Hernandez J, Gomez-Eichelmann MC. 2003. Plasmid DNA supercoiling and gyrase activity in *Escherichia coli* wild-type and *spoS* stationary-phase cells. *J Bacteriol* 185:1097-1100.
- Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M. 2000. Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics* 156:477-488.
- Schneider D, Lenski RE. 2004. Dynamics of insertion sequence elements during experimental evolution of bacteria. *Res Microbiol* 155:319-327.
- Schneider R, Travers A, Kutateladze T, Muskhelishvili G. 1999. A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol Microbiol* 34:953-964.
- Shaver AC, Sniegowski PD. 2002. Spontaneously arising *mutL* mutators in evolving *Escherichia coli* populations are the result of changes in repeat length. *J Bacteriol* 185:6076-6082.
- Simpson GG. 1953. The major features of evolution. New York: Columbia University Press.
- Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature* 387:703-705.
- Stephens JC, Artz SW, Ames BN. 1975. Guanosine 5'-diphosphate 3'-diphosphate (ppGpp) positive effector for histidine operon transcription and general signal for amino-acid deficiency. *Proc Natl Acad Sci USA* 72:4389-4393.
- Swanson MS, Hammer BK. 2000. *Legionella pneumophila* pathogenesis: a fateful journey from amoebae to macrophages. *Annu Rev Microbiol* 54:567-613.
- Tosa T, Pizer LI. 1971. Biochemical basis for the antimetabolite action of L-serine hydroxamate. *J Bacteriol* 106:972-982.
- Travers A, Muskhelishvili G. 2005. DNA supercoiling – a global transcriptional regulator for enterobacterial growth? *Nature Rev Microbiol* 3:157-169.
- Travers A. 1976. Modulation of RNA polymerase specificity by ppGpp. *Mol Gen Genet* 147:225-232.
- Travers AA. 1980. Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. *J Bacteriol* 141:973-976.
- Vasi F, Travisano M, Lenski RE. 1994. Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. *Am Nat* 144:432-456.
- Wang JC. 1971. Interaction between DNA and an *Escherichia coli* protein omega. *J Mol Biol* 55:523-533.
- Wendrich TW, Blaha G, Wilson DN, Marahiel MA, Nierhaus KH. 2002. Dissection of the mechanism for the stringent factor RelA. *Mol Cell* 10:779-788.
- Woods R, Schneider D, Winkworth C, Riley MA, Lenski RE. 2006. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc Natl Acad Sci USA* 103:9107-9112.
- Zhang C-C, Laurent S, Sakr S, Peng L, Bédou S. 2006. Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. *Mol Microbiol* 59:367-375.
- Zhou YN, Jin DJ. 1998. The *spoB* mutants destabilizing initiation complexes at stringently controlled promoters behave like “stringent” RNA polymerases in *Escherichia coli*. *Proc Natl Acad Sci USA* 95:2908-2913.

**Table 1.** IS-associated mutations that occurred during the first 2,000 Generations of experimental evolution (adapted from <sup>(Schneider et al 2000)</sup>).

<b>Mutational event</b>	<b>Gene function</b>	<b>Population *</b>
Insertion of <i>IS150</i> into <i>pykF</i>	Pyruvate kinase	Ara-1 (12)
Insertion of <i>IS150</i> into <i>nadR</i>	NAD** metabolism	Ara+1 (12)
Insertion of <i>IS150</i> into <i>hokB-sokB</i>	Addiction module	Ara+1 (6)
Inversion between <i>IS150</i> elements		Ara+1 (1)
Insertion of <i>IS150</i> upstream of <i>pbpA-rodA</i> operon	Cell wall biosynthesis	Ara+1 (6)

\* The population in which the initial mutation was found by RFLP-IS is indicated, together with the number of populations (within parentheses) in which a mutation in the corresponding genes was subsequently found (see text).

\*\* Nicotinamide adenine di-nucleotide.

## Figure legend.

**Figure 1. A.** Fitness trajectory during 20,000 generations of experimental evolution with *E. coli*.<sup>(Cooper and Lenski 2000)</sup> Each dotted line represents the relative fitness values of each evolved population compared to the ancestor. The black line is the mean of the twelve populations. **B.** Cell size trajectories during 10,000 generations in each of the twelve populations.

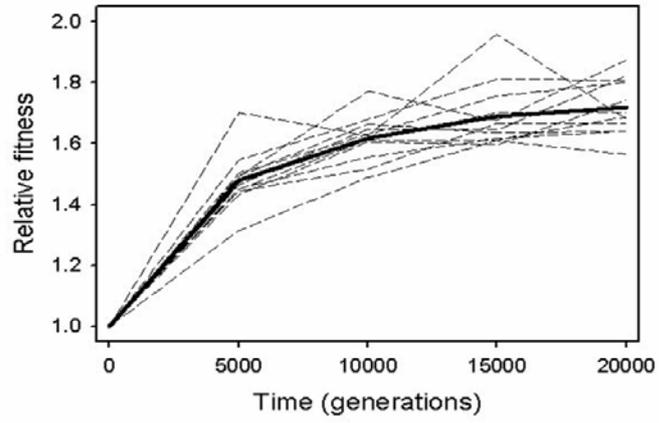
**Figure 2.** Phylogenetic trees constructed from RFLP analyses of the ancestor and evolved clones.<sup>(after Papadopoulos et al 1999)</sup> Each individual is depicted by the generation at which it was sampled followed by an arbitrary number. Inside the box are located individuals presenting the same fingerprints as the ancestor. Pivotal mutations, representing potential beneficial mutations, are shown with arrows. **A.** Population Ara+1. **B.** Population Ara-1. (Copyright 1999 National Academy of Sciences, USA).

**Figure 3.** Map of the *spoT* mutations found in eight evolving populations.<sup>(Cooper et al. 2003)</sup> The amino-acid changes are shown together with their locations within the SpoT protein. The SpoT protein is divided into three functional domains: the ppGpp hydrolase domain and the ppGpp synthetase domain,<sup>(Gentry and Cashel 1996)</sup> and the carboxy terminal domain (CTD) suggested to be involved in the regulation of the two previous domains.<sup>(Mechold et al. 2002)</sup>

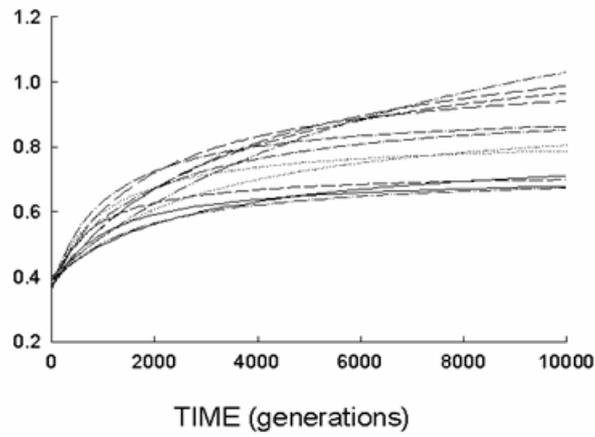
**Figure 4. A.** Fitness effects of three evolved *spoT* mutations, all moved into the ancestral genetic background. Competitions were performed in the same medium used in the long-term evolution experiment. Error bars are 95% confidence intervals based on 6 replicated competition assays for each genotype. From left to right, the four genotypes are as follows: 606 is the ancestral strain with the ancestral *spoT* allele; *spoT*<sup>-1</sup>, *spoT*<sup>+2</sup>, and *spoT*<sup>-4</sup> are alleles that evolved in populations Ara-1,<sup>(Cooper et al. 2003)</sup> Ara+2, and Ara-4, respectively, which were then moved into the ancestral chromosome. In each case, the genotype of interest competed against a variant of the 606 ancestor bearing a neutral marker, the arabinose utilization ability, that allows the distinction between both competitors. As shown previously,<sup>(Lenski et al. 1991)</sup> this marker had no discernible effect on fitness ( $H_0 = 1$ ,  $n = 12$ ,  $t_s = 0.48$ ,  $p = 0.6379$ ). All three *spoT* mutations were beneficial, with fitness relative to the ancestral *spoT* allele of 1.0935 for the Ara-1 mutation ( $H_0 = 1$ ,  $n = 6$ ,  $t_s = 10.53$ ,  $p = 0.0001$ , confirming previous data<sup>(Cooper et al. 2003)</sup>), 1.1195 for the Ara+2 mutation ( $H_0 = 1$ ,  $n = 6$ ,  $t_s = 10.52$ ,  $p = 0.0001$ ), and 1.1196 for the Ara-4 mutation ( $H_0 = 1$ ,  $n = 6$ ,  $t_s = 11.79$ ,  $p = 0.0001$ ). All these gains were eliminated when each evolved allele was replaced again by the ancestral allele, confirming that they were caused by the *spoT* mutations rather than by some hypothetical secondary mutations that might have accidentally occurred during the strain constructions. **B.** Intracellular content of cells from ancestral strains bearing the ancestral and evolved alleles of *spoT* from populations Ara-1, Ara+2, and Ara-4 (from left to right). Each strain was grown in the same medium as the evolution experiment (Davis Minimal) but containing ten times more glucose, and samples were removed in mid-exponential phase and in stationary phase after 24 hours of growth. At each time point, viable cell numbers were determined as well as the concentration of DNA, RNA and proteins (Quant-it™ high-sensitivity kits, Molecular Probes). The ratio of the concentration of each macromolecule per cell at stationary phase relative to exponential phase is shown. Error bars obtained from three independent experiments are indicated.

**Figure 1**

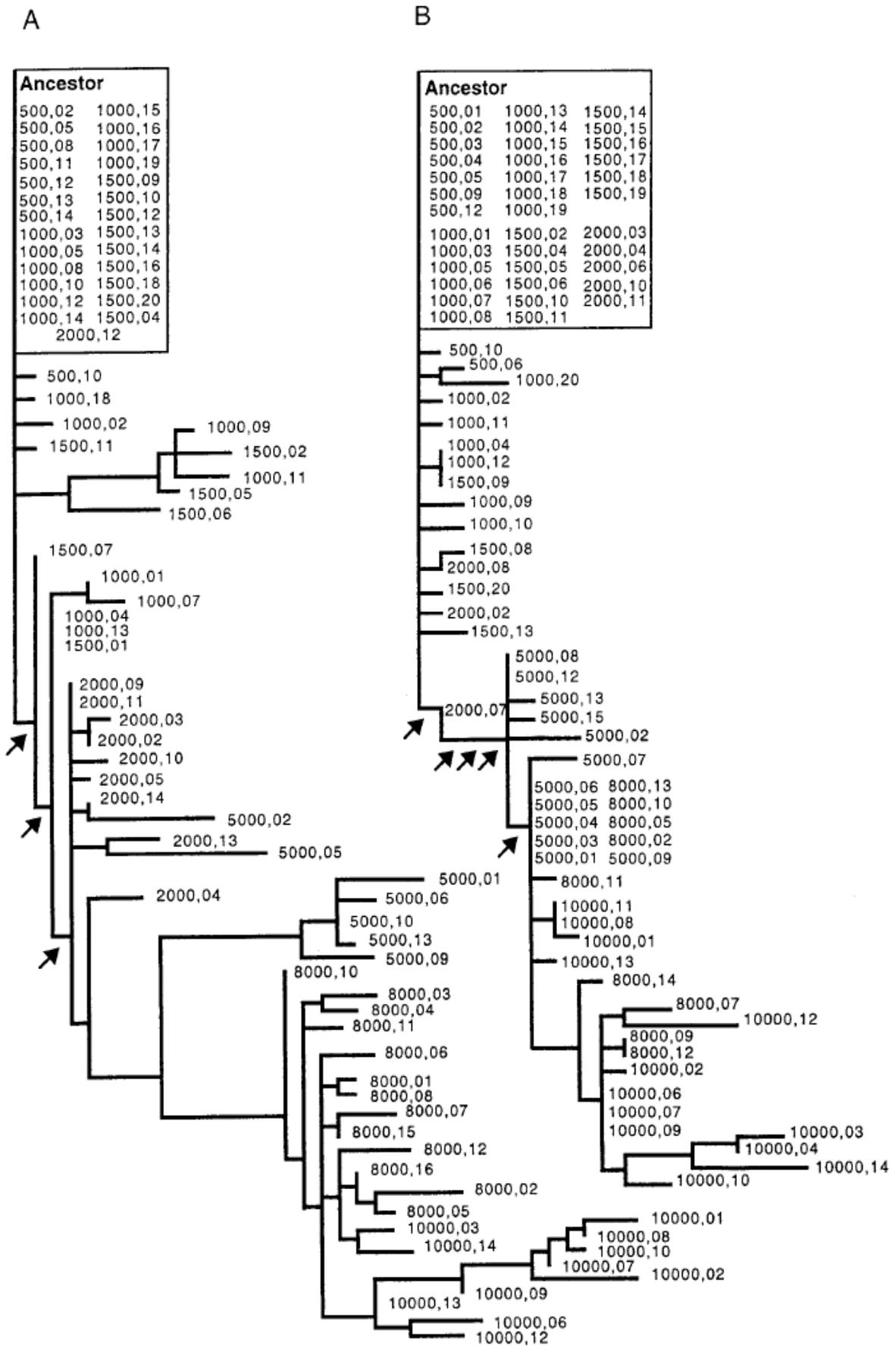
**A**



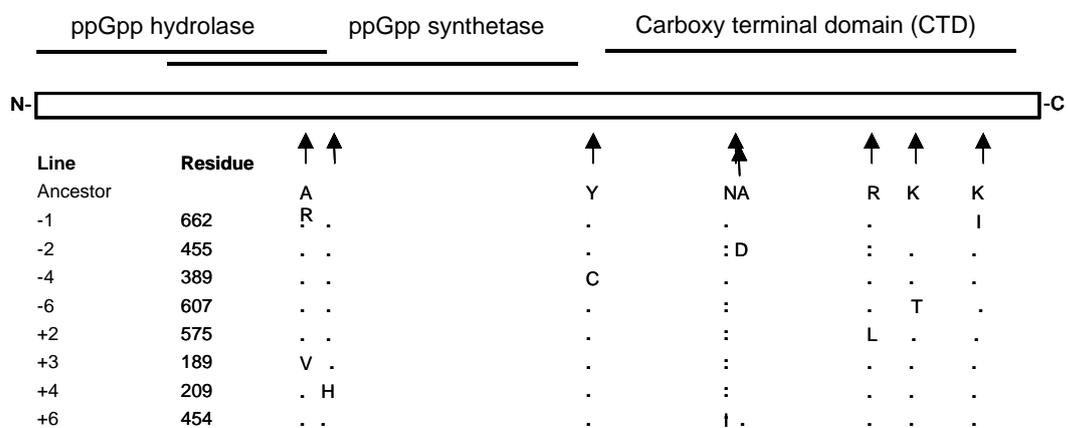
**B**



**Figure 2**



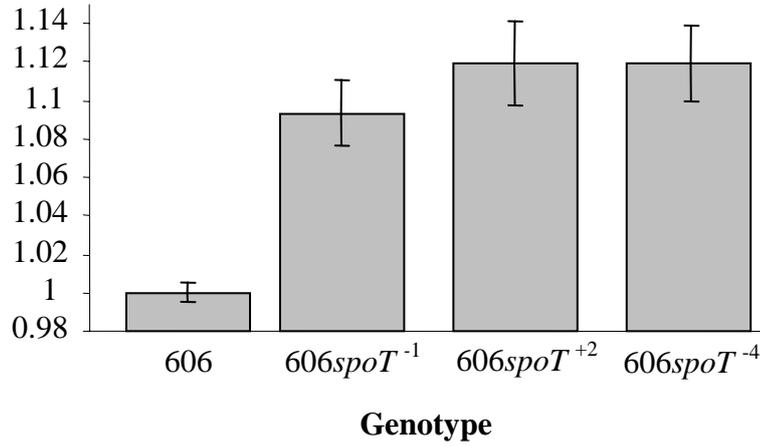
**Figure 3**



**Figure 4**

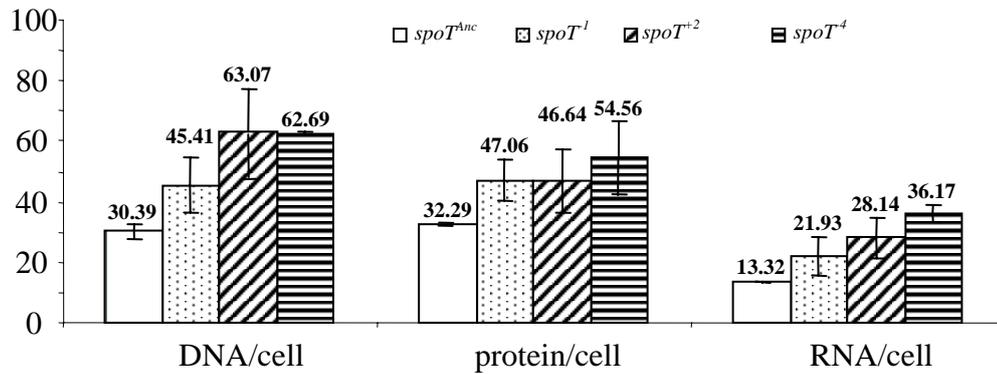
**A**

**Relative fitness**



**B**

**Ratio Stat/Expo**



# RÉSULTATS

Les résultats de ce travail sont exposés ici en trois parties. Dans la première partie, nous avons montré que la superhélicité de l'ADN était la cible de la sélection naturelle au cours de l'évolution expérimentale. Des modifications parallèles de superhélicité ont ainsi été mises en évidence dans la majorité des 12 populations, et une population modèle, Ara-1, a été étudiée plus en détails. La seconde partie porte sur le prolongement de cette étude aux 11 autres populations et montre que le parallélisme phénotypique observé au niveau de la superhélicité de l'ADN a une base génétique. Enfin, la dernière partie détaille et caractérise au niveau moléculaire une des cibles de la sélection naturelle, le gène *fts*.

Toutes les souches utilisées au cours de ce travail sont résumées dans le **Tableau 1**. La nomenclature utilisée pour certaines souches a été modifiée entre la partie I des résultats et les parties II et III. Les correspondances sont données dans le **Tableau 1**. Ces différences sont dues au fait que seule la population Ara-1 ait été étudiée dans un premier temps et ces résultats ont été publiés (Crozat et al., 2005). Ils ont ensuite été étendus aux autres populations, ce qui a nécessité une modification de nomenclature.

RÉSULTATS

PARTIE I

LA TOPOLOGIE DE L'ADN  
COMME CIBLE DE LA  
SÉLECTION NATURELLE

Cette partie a fait l'objet de la publication ci-jointe intitulée « Long-term Experimental Evolution in *Escherichia coli*. XII. DNA Topology as a Key Target of Selection » (E. Crozat, N. Philippe, J. Geiselman and D. Schneider, *Genetics*, 2005, 169 : 523-532).

Tous les jours depuis 1988, les cultures des douze populations sont diluées dans du milieu frais, après 24h d'incubation à 37°C. Elles sont alors soumises à une phase de latence, suivie d'une période de croissance active qui est stoppée lorsque les nutriments s'épuisent. Les cellules entrent alors en phase stationnaire jusqu'au lendemain, où la dilution dans du milieu frais les place à nouveau dans des conditions favorables à leur croissance. Ces transitions croissance-carence et carence-croissance sont connues pour affecter la superhélicité de l'ADN (Balke et Gralla, 1987 ; Hatfield et Benham, 2002 ; Reyes-Dominguez et al., 2003). Des modifications de topologie de l'ADN sont par ailleurs connues pour modifier la transcription globale des gènes (Pruss et Drlica, 1989 ; Steck et al., 1993 ; Bordes et al., 2003 ; Peter et al., 2004). Ceci pourrait ainsi permettre aux bactéries d'adapter rapidement l'expression de leurs gènes aux conditions de croissance, notamment lors des transitions nutritionnelles.

La superhélicité de l'ADN a été mesurée au sein des 12 populations au cours de l'évolution. Ces mesures ont été réalisées en utilisant un plasmide rapporteur, ici pUC18, dont le niveau de superhélicité reflète qualitativement celui du chromosome (Weitao et al., 2000). Ce plasmide rapporteur a été introduit dans 3 clones évolués isolés à 2000, 10 000 et 20 000 générations de chacune des 12 populations, ainsi que dans le clone ancêtre. La distribution des topoïsomères plasmidiques de tous les clones évolués a été comparée à celle du clone ancêtre par électrophorèse sur gels d'agarose contenant de la chloroquine (Higgins et al., 1988). Cette molécule s'intercale dans l'ADN et permet de visualiser les différents topoïsomères d'un plasmide. Les clones évolués de 10 populations sur 12 présentent une augmentation de superhélicité du chromosome à 20 000 générations par rapport à l'ancêtre. Parmi ces populations, 9 présentent des modifications importantes au cours des 2000 premières générations d'évolution, période d'amélioration rapide du fitness (Lenski et Travisano, 1994 ; Cooper et Lenski, 2000). L'augmentation de superhélicité de l'ADN est donc une modification phénotypique supplémentaire observée de façon parallèle dans la majorité des populations au cours de l'évolution (voir Introduction Partie II). Les changements phénotypiques apparaissant de façon parallèle au sein de lignées indépendantes suggèrent une

évolution adaptative (Bull et al., 1997 ; Cooper et al., 2001 ; Nosil et al., 2002). L'accroissement du niveau de superhélicité de l'ADN observé dans la majorité des populations dès 2000 générations, pendant la période de forte augmentation de fitness, pourrait donc être impliqué dans l'adaptation des bactéries aux conditions de l'évolution expérimentale.

Les mutations susceptibles d'être à l'origine de cet accroissement de superhélicité ont été recherchées dans la population modèle Ara-1, utilisée dans la plupart des autres études génétiques réalisées (Papadopoulos et al., 1999 ; Schneider et al., 2000a ; Cooper et al., 2003 ; Pelosi et al., 2006). Cette population présente deux augmentations successives de superhélicité, l'une intervenant avant 2000 générations, l'autre après 10 000 générations. Ceci suggère l'implication d'au moins deux mutations, qui, après séquençage des gènes connus pour être impliqués dans la topologie de l'ADN, ont effectivement été mises en évidence dans deux de ces gènes. Une mutation cible le gène *topA* (H33Y), codant la topoisomérase I, qui relâche les supertours négatifs, tandis que l'autre modifie le site de fixation des ribosomes (RBS) du gène *fis*, codant une protéine de type histone (Schneider et al., 1999 ; Dorman et Deighan, 2003). La présence de ces mutations a alors été recherchée chez une cinquantaine de clones isolés à 500, 1000, 1500, 2000 et 10 000 générations, ce qui a permis de déterminer leur dynamique de fixation au sein de la population Ara-1. La mutation du gène *topA* est apparue avant 500 générations et a été fixée entre 1000 et 1500 générations, bien avant la mutation du gène *fis*, fixée après 10 000 générations. Ceci correspond aux deux périodes où des changements de superhélicité ont été mesurés dans cette population Ara-1.

Afin d'analyser les effets phénotypiques de ces mutations, elles ont été introduites seules ou en combinaison dans le contexte génétique ancestral, au moyen du plasmide suicide pKO3 (Link et al., 1997). Ceci a permis d'obtenir plusieurs souches ancestrales isogéniques à l'exception des allèles *topA* et/ou *fis*. De la même façon, un clone isolé à 2000 générations dans la population Ara-1 a été modifié par remplacement de l'allèle évolué de *topA* par son allèle ancestral. Ces différentes constructions ont permis d'étudier l'effet des deux mutations sur la topologie de l'ADN, le fitness des bactéries, mais aussi sur l'activité de la topoisomérase I et l'expression de *Fis*.

Des mesures de topologie de l'ADN ont été effectuées, de la même façon que précédemment, sur les souches portant les allèles évolués ou ancestraux de *topA* et *fis*.

Chacune des deux mutations augmente la superhélicité négative de l'ADN plasmidique dans le contexte génétique de l'ancêtre. La mutation *topA* reproduit le changement de topologie observé chez les clones évolués isolés à 2000 générations dans la population Ara-1, tandis que les deux mutations combinées reproduisent les changements de topologie observés chez les clones évolués isolés à 20 000 générations. Ce résultat est étroitement corrélé à la dynamique de fixation des mutations : ainsi, la première augmentation de superhélicité correspondrait à la fixation de la mutation dans le gène *topA*, tandis que la deuxième correspondrait à la fixation de la mutation dans le gène *fis*. Ces deux mutations semblent donc être responsables de toutes les modifications de superhélicité observées dans la population Ara-1.

Des expériences de compétition ont été réalisées entre les clones ancêtres portant les allèles évolués et ancestraux des gènes *topA* et *fis*. Ces expériences ont permis de montrer que l'allèle évolué du gène *topA* apporte un bénéfice important de 13,3%, alors que l'allèle évolué du gène *fis* est bénéfique d'environ 3%. Lorsque les deux mutations sont combinées, le gain de fitness est d'environ 16%, ce qui n'est pas statistiquement différent d'un effet additif. Ces deux mutations permettent donc l'adaptation des bactéries aux conditions de l'évolution expérimentale. Elles ont un effet additif aussi bien en terme d'augmentation de superhélicité de l'ADN que de fitness.

La mutation dans le gène *topA* touche un résidu très conservé (Chen et Wang, 1998) du domaine I de la topoisomérase I, situé dans le domaine N-terminal, et probablement impliqué dans l'interaction avec l'ADN (Lima et al., 1994). La topoisomérase I a pour rôle de diminuer la superhélicité négative de l'ADN. L'allèle *topA* évolué entraînant une augmentation de superhélicité, il provoque probablement une baisse de l'activité de la topoisomérase I. L'effet de la mutation découverte dans le RBS du gène *fis* sur l'expression de la protéine a été analysé par western-blot en utilisant des anticorps dirigés contre Fis. Cette mutation provoque une diminution de la quantité de Fis, qui est un répresseur transcriptionnel de l'expression de *gyrA* et *gyrB* codant la gyrase (Schneider et al., 1999). Ceci pourrait donc expliquer l'augmentation de superhélicité de l'ADN induite par cette mutation.

La mutation *topA* est détectée de façon très précoce (avant 500 générations), et ce dans 1/3 des individus testés à ce temps évolutif. Une mutation apportant un tel avantage (13%) aurait dû être fixée en moins de 200 générations (Lenski et al., 1991), c'est-à-dire bien avant

1000 générations. Or il faut attendre 1500 générations pour que tous les individus analysés présentent l'allèle évolué du gène *topA*. Ceci pourrait être dû à un phénomène appelé interférence clonale. A une même étape de l'évolution, la population peut en effet être composée de plusieurs sous-populations constituées de cellules portant des mutations bénéfiques différentes mais apportant un avantage équivalent (Gerrish et Lenski, 1998). Cette compétition entraîne alors un retard de fixation des mutations. Il faut en effet attendre qu'une seconde mutation bénéfique apparaisse dans une des sous-populations, et que cette mutation confère une augmentation suffisante du fitness, pour permettre la fixation de cette sous-population aux dépens des autres. A 1000 générations, la population Ara-1 serait donc constituée d'au moins 2 sous-populations, l'une composée de cellules portant l'allèle évolué de *topA*, l'autre de bactéries portant l'allèle ancestral de *topA* mais comportant une ou plusieurs autres mutations bénéfiques. Le fitness et le niveau de superhélicité de 4 clones évolués isolés à 1000 générations au sein de la population Ara-1 ont été mesurés, 2 des clones ne portant pas l'allèle évolué de *topA*, les 2 autres le possédant. Or ces 4 clones présentent un fitness et un niveau de superhélicité de l'ADN similaires. Ce résultat signifie donc que les clones ne possédant pas l'allèle évolué de *topA* portent une ou plusieurs autres mutations ayant un effet équivalent sur le fitness. La topologie de l'ADN étant la même que pour la sous-population portant l'allèle évolué de *topA*, il est probable qu'une de ces mutations bénéfiques affecte également un gène impliqué dans le contrôle de la superhélicité de l'ADN.

Ces expériences ont permis de mettre en évidence une nouvelle classe de mutations bénéfiques, permettant l'adaptation des bactéries aux conditions de l'évolution expérimentale. Elles affectent le degré de superhélicité de l'ADN, qui est un point de contrôle des transitions nutritionnelles et de la réponse aux stress chez les bactéries. La superhélicité de l'ADN est donc une cible majeure de la sélection naturelle. En effet, l'accroissement de superhélicité a été mis en évidence de façon parallèle dans la majorité des populations mais également dans différentes sous-populations éléments d'une même population.

## Long-Term Experimental Evolution in *Escherichia coli*. XII. DNA Topology as a Key Target of Selection

Estelle Crozat,\* Nadège Philippe,\* Richard E. Lenski,<sup>†</sup> Johannes Geiselmann\*  
and Dominique Schneider\*<sup>1</sup>

\*Laboratoire Adaptation et Pathogénie des Microorganismes, Université Joseph Fourier, Institut Jean Roget, CNRS UMR 5163, F-38041 Grenoble, France and <sup>†</sup>Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824

Manuscript received September 3, 2004  
Accepted for publication October 29, 2004

### ABSTRACT

The genetic bases of adaptation are being investigated in 12 populations of *Escherichia coli*, founded from a common ancestor and serially propagated for 20,000 generations, during which time they achieved substantial fitness gains. Each day, populations alternated between active growth and nutrient exhaustion. DNA supercoiling in bacteria is influenced by nutritional state, and DNA topology helps coordinate the overall pattern of gene expression in response to environmental changes. We therefore examined whether the genetic controls over supercoiling might have changed during the evolution experiment. Parallel changes in topology occurred in most populations, with the level of DNA supercoiling increasing, usually in the first 2000 generations. Two mutations in the *topA* and *fis* genes that control supercoiling were discovered in a population that served as the focus for further investigation. Moving the mutations, alone and in combination, into the ancestral background had an additive effect on supercoiling, and together they reproduced the net change in DNA topology observed in this population. Moreover, both mutations were beneficial in competition experiments. Clonal interference involving other beneficial DNA topology mutations was also detected. These findings define a new class of fitness-enhancing mutations and indicate that the control of DNA supercoiling can be a key target of selection in evolving bacterial populations.

**E**VOOLUTION experiments allow one to study the dynamics and genetic bases of microbial evolution, and this approach has become increasingly widespread in recent years (ELENA and LENSKI 2003). The longest-running evolution experiment involves 12 populations of *Escherichia coli* founded from the same ancestral strain, which have been serially propagated in the same glucose-limited medium for >20,000 generations (LENSKI *et al.* 1991; LENSKI and TRAVISANO 1994; COOPER and LENSKI 2000; LENSKI 2004). The populations improved their performance as demonstrated by the large fitness gains seen when the evolved bacteria compete against their ancestor in the same environment (COOPER and LENSKI 2000). Many genetic changes have been found in these evolved lines including point mutations, deletions, inversions, and insertions of insertion sequence elements (SNIEGOWSKI *et al.* 1997; PAPADOPOULOS *et al.* 1999; SCHNEIDER *et al.* 2000; COOPER *et al.* 2001a, 2003; LENSKI *et al.* 2003). To date, mutations at two loci (*rbs* and *spoT*) have been conclusively demonstrated to be beneficial in the evolutionary environment (COOPER *et al.* 2001a, 2003), but their effects explain only a small

fraction of the total fitness improvement. Both loci are mutated in most or all of the independently derived lineages, indicating parallel adaptive evolution.

In the experiment cited above, each day the evolving populations experienced a lag phase followed by exponential growth, then depletion of the limiting glucose, and finally stationary phase (until transfer into fresh medium the next day). Such transitions are known to influence DNA topology (HATFIELD and BENHAM 2002; REYES-DOMINGUEZ *et al.* 2003). DNA supercoiling is dynamically modified during many environmental challenges, including the transition from growth to stationary phase (BALKE and GRALLA 1987), nutrient upshift (REYES-DOMINGUEZ *et al.* 2003), anaerobiosis (BHRIAN *et al.* 1989), thermal stress (GOLDSTEIN and DRLICA 1984), oxidative stress (WEINSTEIN-FISCHER *et al.* 2000), osmotic stress (HIGGINS *et al.* 1988), and acid stress (KAREM and FOSTER 1993). Phenotypic acclimation by bacteria to these challenges requires rapid changes in expression of many genes, and one important acclimatory response is the transient modification of supercoiling, which can produce genome-wide changes in rates of transcription (JOVANOVICH and LEBOWITZ 1987; PRUSS and DRLICA 1989; STECK *et al.* 1993; GMUENDER *et al.* 2001). The topology of DNA therefore helps to coordinate the gene regulatory networks of bacteria in response to varying environments. The maintenance of

<sup>1</sup>Corresponding author: Laboratoire Adaptation et Pathogénie des Microorganismes, Facultés Médecine/Pharmacie, Université Joseph Fourier, Institut Jean Roget, 38700 La Tronche, France.  
E-mail: dominique.schneider@ujf-grenoble.fr

proper DNA topology is also required for cell viability, since it affects many processes in bacteria including replication, repair, transcription, recombination, and transposition. The level of DNA supercoiling is tightly regulated in the cell by the combined activities of topoisomerases (CHAMPOUX 2001) and histone-like proteins (DORMAN and DEIGHAN 2003). In *E. coli*, three main topoisomerases are involved in maintaining DNA supercoiling: topoisomerases I (WANG 1971) and IV (ZECHIEDRICH *et al.* 2000) relax DNA, while DNA gyrase introduces negative supercoils (GELLERT *et al.* 1976). Histone-like proteins constrain the supercoiling level by binding to DNA and regulating the expression of the topoisomerase-encoding genes (DORMAN and DEIGHAN 2003).

In this study, we sought to determine whether the level of DNA supercoiling might have changed during the evolution experiment, because the populations experienced daily challenges of nutrient upshift and exhaustion. If so, such changes would suggest candidate loci for further study by sequencing, genetic manipulation, and analyses of phenotypic effects.

## MATERIALS AND METHODS

**Strains, plasmids, and culture conditions:** Twelve populations of *E. coli* B were started from two genotypes that differed by only a single neutral marker (arabinose utilization), and they were propagated for 20,000 generations at 37° in a glucose-limited defined medium (LENSKI *et al.* 1991; LENSKI and TRAVISANO 1994; COOPER and LENSKI 2000; LENSKI 2004). Six populations, designated Ara-1–Ara-6, were founded from the ancestor that is Ara<sup>-</sup> (unable to use arabinose as a carbon source) and six others, Ara+1–Ara+6, from an Ara<sup>+</sup> revertant of the ancestor. The arabinose utilization phenotype serves as a marker in competition experiments and was shown to be neutral under these conditions (LENSKI *et al.* 1991). We used three clones sampled at random from each of the 12 populations at generations 2000, 10,000, and 20,000 (COOPER and LENSKI 2000). Population Ara-1 served as the focal population in this study and was the source of the evolved *topA* and *fis* alleles. Strain 606 *fis*<sup>Δ</sup> has an in-frame *fis* deletion and was constructed (D. SCHNEIDER, unpublished data) using the suicide-plasmid pKO3 (LINK *et al.* 1997). This deletion construct is presumed to be otherwise isogenic to the ancestor (REL606) and was used as a control in immunoblot analyses.

We used plasmids pUC18 (YANISCH-PERRON *et al.* 1985) for DNA supercoiling measurements, pKO3 (LINK *et al.* 1997) for allele replacements, and pCRII-Topo (Invitrogen, San Diego) for cloning experiments.

All experiments were performed by growing strains either in the same Davis glucose (25 μg/ml) minimal medium (DM25) that was used in the evolution experiment (LENSKI *et al.* 1991) or in rich Luria broth (LB) medium (SAMBROOK *et al.* 1989). Ampicillin (100 μg/ml), chloramphenicol (30 μg/ml), or kanamycin (50 μg/ml) were added as needed.

**DNA supercoiling measurements:** Topological changes during evolution were measured by analyzing the relative abundance of different topoisomers of a reporter plasmid. The pUC18 plasmid was introduced to various strains by electrotransformation (Bio-Rad, Richmond, CA, 200 ohm, 25 μF, 2.5 kV). Electrocompetent cells were prepared as described elsewhere (SAMBROOK *et al.* 1989). It should be noted that the way we measured DNA topology allowed us to assess changes

only in the activity of the nicking-closing enzymes (*i.e.*, topoisomerases). The superhelix density *in vivo* is also affected by other DNA-binding proteins (LILLEY 1986; BLISKA and COZZARELLI 1987).

Bacteria containing pUC18 were grown in LB to an OD<sub>600nm</sub> = 2 (or in DM25 to the end of exponential growth). Cells were harvested and pUC18 was extracted using the QIAprep Spin Miniprep Kit (QIAGEN, Chatsworth, CA). The same amount of plasmid DNA (200 ng) from each strain was loaded onto 1% agarose gels containing 1.5 μg/ml chloroquine. The gels were run overnight at 4° at 3 V/cm. Under these conditions, more highly supercoiled topoisomers migrate faster (HIGGINS *et al.* 1988). Chloroquine was washed from the gels by soaking for 4 hr in water, and photographs were taken under UV light after ethidium bromide staining. Because such gels are sensitive to buffer conditions, we checked that negative supercoils occur under the chloroquine concentration that we used. It is known that the relative mobility of the different topoisomers varies with chloroquine concentrations (HIGGINS *et al.* 1988); at a concentration of 25 μg/ml, more supercoiled topoisomers migrate more slowly, while more relaxed ones migrate faster. As expected, at that higher concentration, we obtained the symmetrical outcomes compared to those obtained at a chloroquine concentration of 1.5 μg/ml (data not shown).

Quantitative comparisons of DNA topology between each evolved clone and the ancestor were performed by scanning the relevant gels. Densitometric analyses were performed to obtain the relative abundance of all of the plasmid topoisomers for each clone, and the average topoisomer value was computed from these data for each clone, including the ancestor. To estimate the level of DNA supercoiling, mean  $\sigma$ -values were calculated according to the formula  $\sigma = \Delta Lk/Lk_0$ , where  $\Delta Lk$  is the difference in the linking number (*i.e.*, difference in average topoisomer value) between two clones and  $Lk_0$  for pUC18 is 2686 bp/10.45 bp/turn = 257. The  $\sigma$ -value for the ancestor was expressed relative to the midpoint topoisomer of plasmid that was completely relaxed by using calf-thymus topoisomerase I and assigned  $\sigma = 0$ . Comparing the topoisomer distribution of plasmid DNA from the ancestor with the completely relaxed plasmid control yielded an estimate of  $\sigma = -0.066$  for the ancestor. We emphasize that the midpoint topoisomer of the completely relaxed plasmid was used only to establish a baseline value. All comparisons between evolved clones and the ancestor and between the various isogenic constructs used the complete distribution of topoisomers obtained by densitometry. The  $\sigma$ -value for each evolved clone or isogenic construct was obtained by adding or subtracting its average topoisomer value to that of the ancestor for more supercoiled or more relaxed molecules, respectively. For conversion, one negative supercoil difference corresponds to a difference in  $\sigma$  of  $-0.0039$ . The resulting  $\sigma$ -value for each evolved clone and isogenic construct was then compared to that of the ancestor. These analyses thus allowed a precise quantitative comparison of the distribution of plasmid topoisomers between each clone and the ancestor.

Changes in DNA topology were measured after growing the various clones in LB medium. However, the populations had evolved in a glucose-limited minimal medium. Therefore, we also measured the supercoiling of the reporter plasmid after growth in the same minimal medium as used in the evolution experiment. Cell densities and molecular yields are lower in the minimal medium, but similar changes in DNA topology were observed in minimal and LB media (data not shown). All subsequent experiments on supercoiling were performed using cells grown in LB medium.

**Electrophoresis and immunoblot analyses of proteins:** Triplicate cultures of the ancestor carrying the ancestral *fis* allele,

an evolved *fis* allele, or a complete deletion of *fis* were prepared in LB, and aliquots were sampled after 1, 1.5, 2.5, and 5 hr of incubation at 37°. (Under these conditions, the duration of the lag phase was ~0.75 hr, and after ~5 hr the cultures enter into stationary phase.) Cell pellets were resuspended in lysis buffer (20 mM HEPES pH 7.4, 1 mM EDTA, 10% glycerol, 1 mM DTT) before sonication. After centrifugation at 10,000 × *g* for 30 min at 4°, total protein concentrations were determined using the Bradford protein assay kit (Bio-Rad) and with bovine serum albumin as a standard. Equal amounts of protein samples were subjected to 16% SDS-polyacrylamide gel electrophoresis, and the separated proteins were then electrotransferred onto polyvinylidene difluoride membranes (Amersham Pharmacia). Prestained protein standards (Amersham, Buckinghamshire, UK) were used for molecular weight estimation. Immunoblot analyses were performed with antibodies against Fis and RpoA (courtesy of A. Ishihama and M. Cashel, respectively). The blots were developed using ECL (Amersham Pharmacia) or nitroblue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate (Sigma) systems, and the staining intensity was quantified using the ImageGauge (Fujifilm) software.

**Strain construction:** The *topA* and *fis* mutations, alone and in combination, were introduced into the ancestral genetic background by using the suicide plasmid pKO3 (LINK *et al.* 1997). Briefly, a PCR product containing ~500 bp of adjacent sequences on each side of the mutation first was cloned into pCRII-Topo and then subcloned into pKO3. After electrotransformation of the ancestor with the appropriate plasmid, chloramphenicol-resistant cells with a copy of this nonreplicative plasmid that had integrated into the chromosome were selected at high temperature. Subsequent selection on 5% sucrose-LB plates without NaCl (BLOMFIELD *et al.* 1991) allows detection of plasmid loss, because the plasmid carries *sacB*, which renders cells sensitive to killing by sucrose. These plasmid-free cells were then screened for the presence of the evolved *topA* allele by PCR/restriction fragment length polymorphism (PCR/RFLP) using *Nmu*CI (Fermentas) to distinguish between ancestral and evolved alleles or for the presence of the *fis* allele by direct sequencing. As a further control for the precision of these manipulations, constructed strains were “deconstructed” (the introduced evolved alleles were replaced again by their ancestral counterparts, using pKO3) to check for unintended secondary mutations (COOPER *et al.* 2003); in all cases, the deconstructed strains were confirmed to have reversed the phenotypic effects of the introduced *topA* and *fis* evolved alleles. The same approach was used to replace the evolved alleles by the ancestral ones in some evolved clones.

**Fitness assays:** The relative fitness of two strains was estimated by directly competing them, as described elsewhere (LENSKI *et al.* 1991). Briefly, isogenic ancestral strains with the introduced evolved *topA* allele, the evolved *fis* allele, or both were allowed to compete against the ancestor carrying the opposite neutral Ara marker (LENSKI *et al.* 1991). Each pairwise competition was replicated at least fourfold, with competitions running for 1 day or, with 1:100 daily serial transfer, for 6 days (longer duration allows detection of smaller fitness effects) in the same medium as the evolution experiment. Both 1-day and 6-day competitions include all the same phases of population growth that occur in the serial-transfer regime, including a lag-phase prior growth, exponential growth, depletion of the limiting glucose, and stationary phase (VASI *et al.* 1994). In both cases, competitors were separately acclimated to the same regime, then transferred together from stationary-phase cultures into fresh medium (with 1:200 dilution each, thus 1:100 combined). Samples of the competition assays were taken to measure the abundance of the competitors; these samples were always taken in stationary phase (24 hr after the previous transfer), such that only the duration of the competi-

tions varied and not the physiological state of the cells or the time they spent in various states. Fitness was calculated as the ratio of the realized (net) population growth rates obtained for the two strains while they competed with one another; *t*-tests were performed to evaluate whether measured fitness values differed significantly from the null hypothetical value of one. The same basic protocol was also used to assess fitness variation among clones sampled from the evolution experiment; a one-way ANOVA was then performed to test whether there was significant variation in fitness among the clones.

**Evolutionary dynamics of mutation substitutions:** PCR/RFLP strategies were used to track the origin and spread of the mutations discovered in the focal population, using clones isolated from samples frozen during the evolution experiment. For the *topA* locus, numerous clones from generations 500, 1000, 1500, and 2000 were subjected to PCR using primers ODS177 (5'-GACTACGTGGTCAAATCCAG-3') and ODS178 (5'-ACCAGGCAACACTTCATAGTG-3'). The presence or absence of the evolved *topA* mutation was determined by *Nmu*CI digestion of the PCR products. The evolved *fis* allele does not change any restriction sites. Therefore, PCR products obtained with primers ODS65 (5'-GTACCGAATTGCACGTAAAG-3') and ODS141 (5'-CACCGTGAATAATGGTCTTAG-3') for many clones isolated from 2000, 10,000, and 20,000 generations were directly sequenced using primer ODS65.

## RESULTS

**DNA topology changed during experimental evolution:** To investigate DNA supercoiling, we examined reporter plasmid DNA that was introduced into, and then isolated from, 108 evolved clones (3 clones from each of the 12 long-term populations for each of three time points: 2000, 10,000, and 20,000 generations; COOPER and LENSKI 2000). All measures were compared to the ancestor clone harboring the same reporter plasmid. The ancestor and evolved clones were grown in rich medium and, upon reaching stationary phase, plasmid DNA was extracted, quantified, and loaded onto chloroquine-agarose gels to separate topoisomers and visualize their distribution. In all cases, the 3 evolved clones from the same population and same generation gave identical results.

Three main trends emerge from the levels of DNA supercoiling observed in the various evolved clones relative to that measured in the ancestor (Figure 1). First, 10 of the 12 populations at generation 20,000 show increased supercoiling relative to their ancestor. Second, alterations of the DNA topology began early in the evolution experiment, with 9 populations showing changes by 2000 generations. These initial 2000 generations were also the period of most rapid fitness improvement (LENSKI and TRAVISANO 1994; COOPER and LENSKI 2000). Third, three populations showed continued increases in supercoiling over time, in one case reaching a final  $\sigma$ -value difference of ~12% and in two others of ~17% relative to the ancestral state.

Only a few exceptions to these general trends were observed (Figure 1). In population Ara+1, the change in supercoiling was seen only transiently at 2000 generations. In Ara+3, supercoiling initially increased but be-

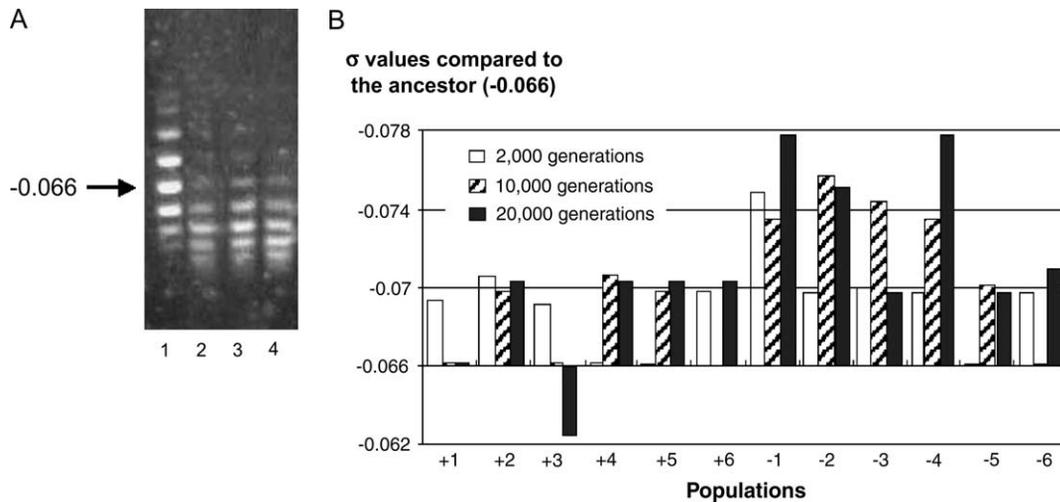


FIGURE 1.—Parallel changes in DNA topology during experimental evolution. Plasmid DNA was isolated from ancestral and evolved cells carrying reporter plasmid pUC18 and analyzed by electrophoresis on gels that allow visualization of the topoisomer distribution, with more tightly supercoiled topoisomers migrating faster (HIGGINS *et al.* 1988). (A) Gel for population Ara-1 with pUC18 extracted from the ancestor clone (lane 1) and evolved clones from 2000, 10,000, and 20,000 generations (lanes 2–4, respectively). Quantitative comparisons between each evolved clone and the ancestor are based on densitometric analyses of the gels (see MATERIALS AND METHODS). The indicated  $\sigma$ -value for the ancestor clone ( $-0.066$ ) was obtained relative to the midpoint topoisomer of plasmid completely relaxed with calf-thymus topoisomerase I (see MATERIALS AND METHODS). (B) The change in supercoiling level is shown by bars for each evolved population, labeled as +1–+6 (Ara+1–Ara+6) and –1 to –6 (Ara-1–Ara-6). Open, hatched, and solid bars represent clones isolated at 2000, 10,000, and 20,000 generations, respectively. The  $\sigma$ -values calculated in evolved clones are compared to the ancestral level ( $-0.066$ ), such that higher absolute values show increased supercoiling and lower absolute values indicate relaxation of DNA. The absence of a bar for +6 at 10,000 generations indicates that no measurement could be obtained (see text). Otherwise, short bars indicate the absence of any change in supercoiling.

came more relaxed than the ancestral state at 20,000 generations. In population Ara+6, the supercoiling level could not be measured for the 10,000-generation clones, because they grew slowly under the ampicillin selection required for maintenance of the reporter plasmid, and no plasmid DNA could be recovered. However, increased supercoiling was seen at 2000 and 20,000 generations in this population. Finally, Ara-6 showed no change in supercoiling at 10,000 generations, whereas it was increased at 2000 and 20,000 generations. During the course of the evolution experiment, four populations became mutators (SNIĘGOWSKI *et al.* 1997; COOPER and LENSKI 2000), including Ara+3, Ara+6, Ara-2, and Ara-4. There is no apparent association of the evolved changes in DNA topology with mutator status.

Summarizing to this point, our results show consistent and parallel changes in the DNA supercoiling level in most evolved populations, with many changes already seen by 2000 generations. Rapid and parallel changes across independent lineages are indicative of adaptive evolution (BULL *et al.* 1997; COOPER *et al.* 2001a; NOSIL *et al.* 2002). These data therefore strongly suggest that the parallel changes in DNA topology are adaptive. To test this hypothesis further, we sought to identify the mutations responsible for increased supercoiling among the various genes involved in the control of DNA topology. We focused on changes in population Ara-1 for two

reasons. First, this population has served as a focal lineage in several other studies of the long-term lines (PAPADOPOULOS *et al.* 1999; COOPER and LENSKI 2000; SCHNEIDER *et al.* 2000; COOPER *et al.* 2003). Second, the supercoiling level in this population increased in at least two successive steps, implicating the involvement of at least two mutations (Figure 1).

**Identification of mutations altering DNA topology:** By sequencing candidate loci in a clone isolated at 20,000 generations from population Ara-1, we found two mutations. One mutation is in *topA*, which encodes topoisomerase I, which in turn relaxes DNA (WANG 1971). The other mutation occurs in *fis*, which encodes the Fis protein that controls the level of DNA supercoiling (SCHNEIDER *et al.* 1999). The mutation in *topA* is a C-to-T substitution, leading to the change of the histidine 33 residue of topoisomerase I into a tyrosine. The mutation in *fis* is an A-to-C substitution, which lies four nucleotides upstream of the ATG translational start codon.

To analyze precisely the phenotypic effects of these two mutations, we constructed a set of isogenic strains by allelic replacements. We moved each mutation alone and in combination into the ancestral chromosome, replacing the corresponding ancestral allele. We also replaced the evolved alleles with their ancestral counterparts in some evolved clones. These isogenic strains, with

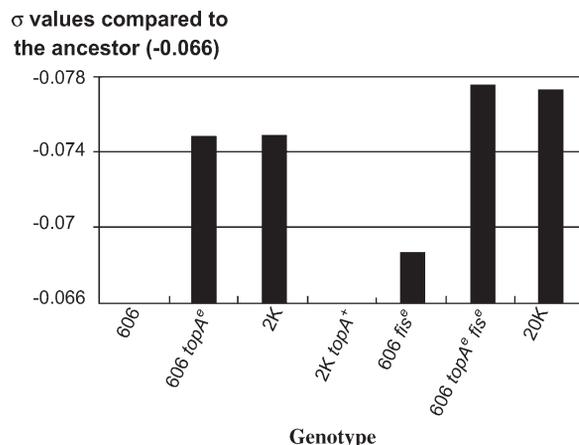


FIGURE 2.—Changes in DNA supercoiling generated by evolved *topA* and *fis* mutations. The changes in  $\sigma$ -values relative to the ancestral level ( $-0.066$ ) are shown for various strains. The strain 606 is the ancestral strain; 2K and 20K are clones isolated from population Ara-1 at 2000 and 20,000 generations, respectively. All others are isogenic constructs made by replacing ancestral alleles with evolved alleles (denoted by superscript *e*) or by replacing an evolved allele with its ancestral counterpart (superscript +).

and without the *topA* and *fis* mutations, were then used to examine the effects of the evolved alleles on DNA supercoiling and on competitive fitness.

#### Both mutations contribute to altered DNA topology:

The reporter plasmid was introduced into the isogenic strains, which were grown in LB, their plasmids extracted, and DNA topology measured as before. As shown in Figure 2, both the *topA* and *fis* mutations, when introduced separately into the ancestral chromosome, caused increases in DNA supercoiling. The *topA* mutation increased the  $\sigma$ -value by  $\sim 12\%$ , whereas the *fis* mutation increased it by  $\sim 5\%$ . When combined in the ancestral background, the two mutations had an additive effect, leading to a net increase of 17% relative to the ancestor. The effect of the *topA* mutation reproduces the DNA supercoiling level measured in clones from generation 2000 in population Ara-1, while the combined effects of the *topA* and *fis* mutations match the supercoiling observed in the 20,000-generation clones (Figure 2). Thus, the two mutations are sufficient to explain the observed changes in DNA topology in this population.

These changes suggest the substitution of the *topA* mutation before 2000 generations and the substitution of the *fis* mutation sometime later. Sequencing *topA* in a 2000-generation clone confirmed the presence of the same evolved allele as found at 20,000 generations. Replacing the evolved *topA* allele in the same 2000-generation clone with the ancestral allele restored supercoiling to the ancestral level (Figure 2). This result also precludes the presence at 2000 generations of compensatory mutations in the *gyrAB* genes, which have been

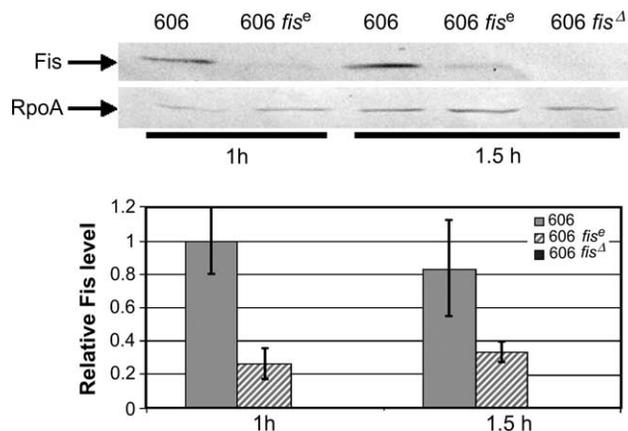


FIGURE 3.—Relative levels of Fis protein. Extracts were prepared from the ancestral strain (606), an isogenic strain with the evolved *fis* allele (606 *fis*<sup>+</sup>), and a *fis*-deleted derivative of the ancestor (606 *fis*<sup>Δ</sup>) after 1 and 1.5 hr of exponential growth in LB medium. Western blots and immunodetection assays are shown for an anti-Fis antibody (top) and, as a control, an anti-RpoA antibody (middle). (Bottom) The ratio of Fis to RpoA levels in the same strains, based on three replicate experiments with error bars indicating standard deviations. The ratio is arbitrarily set to 1 for the ancestral strain after 1 hr of exponential growth. No Fis protein was detected in the control strain with the *fis* gene deleted.

found elsewhere to compensate for certain mutations in *topA* (DiNARDO *et al.* 1982; PRUSS *et al.* 1982). The absence of mutations in *gyrAB* was confirmed by sequencing the entirety of these genes in the 2000-generation clone. These data also imply a reduction of the DNA-relaxing activity of topoisomerase I in the evolved clones, because this protein otherwise reduces supercoiling.

To analyze the effect of the evolved *fis* mutation, we performed Western blot experiments and immunodetection using an antibody raised against Fis, with an antibody raised against RpoA serving as a control (Figure 3). We used three variants of the ancestral strain that carried the ancestral *fis* allele, the evolved *fis* allele, or a deletion of the *fis* gene. As expected (MALLIK *et al.* 2004), the Fis level peaked during exponential growth in the ancestor (data not shown). The evolved *fis* allele had somewhat less than half the ancestral level of Fis protein, while the *fis*-deletion strain produced no detectable Fis (Figure 3). The level of the control protein did not differ between these strains.

#### Beneficial fitness effects of *topA* and *fis* mutations and their dynamics:

To evaluate the fitness effects of the mutations responsible for the altered DNA topology in the focal population, we performed competitions between the ancestral strain with the evolved *topA* allele only, the evolved *fis* allele only, both evolved alleles, or both ancestral alleles, under the same culture conditions that prevailed during the long-term evolution experiment (Figure 4). In each case, the genotype of interest competed against a variant of the ancestor bearing a neutral marker (arabinose utilization) that allowed

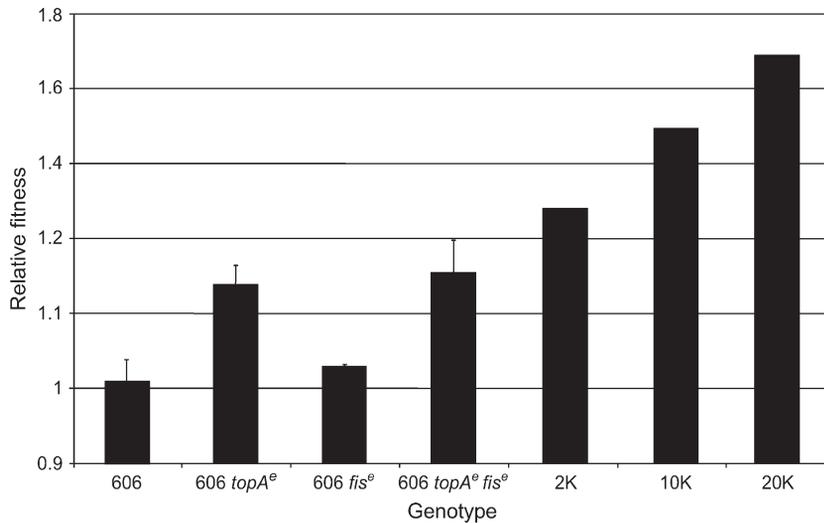


FIGURE 4.—Fitness effects of the evolved DNA topology-altering mutations in the ancestral genetic background. Competition experiments were performed in the same medium used in the long-term evolution. Error bars are 95% confidence intervals based on 16, 11, 4, and 6 replicate competition assays for each genotype (left to right). The strain 606 is the ancestor; *topA*<sup>e</sup> and *fis*<sup>e</sup> are alleles that evolved in focal population Ara-1 and then were moved, alone or in combination, into the ancestral chromosome. For comparison, we also show the mean fitness for population Ara-1 at three time points: 2000 (2K), 10,000 (10K), and 20,000 (20K) generations. The value shown for 2K is the average of values obtained by LENSKI *et al.* (1991) and by LENSKI and TRAVISANO (1994); the value for 10K is the average of values obtained by LENSKI and TRAVISANO (1994) and by COOPER and LENSKI (2000); and the value for 20K is from COOPER and LENSKI (2000).

the strains to be distinguished. As shown in prior studies (LENSKI *et al.* 1991), this marker had no discernible effect on fitness ( $H_0 = 1$ ,  $t$ -test,  $t_s = 0.4293$ ,  $P = 0.6738$ ). Each evolved allele was beneficial by itself, with fitness relative to the ancestral allele of 1.133 for the *topA* mutation ( $H_0 = 1$ ,  $t$ -test,  $t_s = 10.58$ ,  $P < 0.0001$ ) and 1.029 for the *fis* mutation ( $H_0 = 1$ ,  $t$ -test,  $t_s = 41.14$ ,  $P < 0.0001$ ). Combining the two mutations gave a relative fitness of 1.155, which does not differ significantly from the fitness predicted by a simple additive model (13.3% + 2.9% = 16.2%). Also, these gains were eliminated when each evolved allele was replaced again by the ancestral allele (data not shown), confirming that the gains were caused by the *topA* and *fis* mutations rather than by some hypothetical secondary mutations that might have accidentally occurred during the strain constructions. In addition, replacing each evolved allele with the corresponding ancestral allele in evolved clones reduced their fitness relative to the otherwise isogenic evolved clones, as would be expected if these mutations were beneficial (data not shown). Thus, increasing DNA supercoiling had a beneficial effect during evolution in the experimental culture regime.

To examine in detail the temporal dynamics of the appearance and substitution of the evolved *topA* allele in population Ara-1, we performed PCR/RFLP experiments (see MATERIALS AND METHODS) using a total of 180 evolved clones isolated at generations 500, 1000, 1500, and 2000. The *topA* mutation was present in 22/45 clones at 500 generations and in 19/45 clones at 1000 generations, and it was present in 45/45 clones at both 1500 and 2000 generations. Thus, it contributed to the early large gain in fitness, existed as a polymorphism for some time (probably indicating clonal interference from other beneficial mutations: see below), and then reached fixation. Another beneficial mutation,

in the *spoT* gene, was recently shown to have arisen somewhat later in the same population (COOPER *et al.* 2003).

We performed similar analyses for the evolved *fis* allele. This allele was not found in any of 50 clones isolated at 2000 generations, but was present in 15/37 clones at 10,000 generations. Therefore, the evolved *fis* allele contributed to the later, slower increase in fitness. The dynamics of the evolved *topA* and *fis* alleles match the observed changes in supercoiling (Figures 1 and 2). The *topA* allele was substituted by 2000 generations, and it induced the ~12% change seen at that time in the Ara-1 population. The *fis* mutation, absent at 2000 generations and substituted later, induces an additional ~5% change, leading to the ~17% net change in supercoiling seen at 20,000 generations.

**Clonal interference and DNA topology as a key target of selection:** Given the large fitness benefit of the *topA* mutation and its presence in about half the clones at 500 generations, one might have expected its complete fixation by 1000 generations. On the basis of its fitness advantage of >10% (Figure 4), the *topA* mutation, if unimpeded, would require <500 generations to reach 100% frequency (LENSKI *et al.* 1991). However, competition between subpopulations carrying different beneficial mutations—known as “clonal interference”—can prolong the time required for substitution (GERRISH and LENSKI 1998). To examine this possibility, we sought to determine whether other clones present at the same time had similar fitness but lacked the evolved *topA* allele. Recall that the focal population at generation 1000 had two subpopulations: one had the evolved *topA* allele, and the other retained the ancestral allele. We isolated two clones of each class from this sample; none of them had the previously characterized *spoT* mutation (COOPER *et al.* 2003), another beneficial mutation that

arose somewhat later in the same population. We then measured fitness and DNA supercoiling for these four clones. Competitions between each clone and the ancestor were performed with sixfold replication and gave similar fitness values for all four clones with no significant variation among them (one-way ANOVA,  $F_{3,20} = 1.956$ ,  $P = 0.1531$ ). The finding that co-occurring clones that retained the ancestral *topA* allele had similar fitness to those bearing the evolved allele supports the hypothesis that clonal interference impeded the substitution of the evolved allele. Moreover, and surprisingly, all four clones had approximately the same 12% increase in supercoiling compared with the ancestor. In other words, even clones from generation 1000 without the evolved *topA* allele had an altered DNA-topology phenotype similar to those carrying that allele. Therefore, the subpopulation that retained the ancestral *topA* allele acquired some other topology-altering and fitness-enhancing mutation, although the identity of this mutation is not known at present. By showing that similar beneficial mutations arose in parallel even within the same population (as well as across many replicate populations), these data further demonstrate that DNA topology was a key target of selection in the environment of the evolution experiment.

#### DISCUSSION

We describe the discovery of a new class of fitness-enhancing mutations that were substituted in evolving populations of *E. coli* during 20,000 generations in a culture environment in which cells experienced daily fluctuations between glucose abundance and depletion (LENSKI *et al.* 1991; LENSKI and TRAVISANO 1994; COOPER and LENSKI 2000; LENSKI 2004). Of 12 independent populations, 10 underwent changes in DNA topology such that the level of supercoiling was higher (more negative) than that in the ancestor. We closely examined one focal population that showed a two-step increase in the level of DNA supercoiling. Two mutations were found in candidate genes known to affect DNA topology, *topA* and *fis*. The two mutations were moved, alone and in combination, into the chromosome of the ancestor. We measured supercoiling levels as well as relative fitness in these isogenic strains and found that both mutations were beneficial and that they contributed additively to the increased DNA supercoiling (Figures 2 and 4). By using numerous clones sampled from the focal population at various time points, we showed that the *topA* mutation was substituted prior to the *fis* mutation, as was likely given its larger fitness effect (Figure 4). We also found that the substitution of the evolved *topA* allele was impeded by competition with other clones that evidently had acquired other beneficial mutations that similarly affected DNA topology. The parallel phenotypic changes across the replicate populations

as well as the spread of multiple mutations affecting DNA topology within at least the focal population demonstrate that the level of DNA supercoiling was an important target of selection during the evolution experiment.

The changes in DNA topology were assessed by measuring the distribution of topoisomers of a reporter plasmid. This method is widely used to compare superhelical density between different strains or under different environmental conditions. ZECHIEDRICH *et al.* (2000) compared the effect of inhibiting topoisomerase I and topoisomerase IV on the superhelical level of a reporter plasmid and on the “effective” supercoiling level in the cell. The latter was estimated by the level of site-specific recombination by the bacteriophage  $\lambda$  integrase (BOLES *et al.* 1990). Loss of the activity of both topoisomerases led to DNA topology changes of the reporter plasmid that reflected the supercoiling changes in the cell (ZECHIEDRICH *et al.* 2000). In another study, changes in negative supercoiling of the *E. coli* chromosome in two cell-cycle mutants, *mukB* and *seqA*, were investigated by direct measurement of the superhelicity of membrane-free nucleoids using titration of the intercalating agent ethidium bromide (WEITAO *et al.* 2000). The two mutants have opposing effects on the chromosome topology. Measurements of superhelicity for a reporter plasmid and for the chromosome using ethidium bromide titration showed the same qualitative effects of the mutations, with the magnitude of the effects measured for the chromosome somewhat greater than that measured with the plasmid (WEITAO *et al.* 2000). These results therefore strongly support the interpretation that the DNA topology changes detected using a reporter plasmid qualitatively reflect what is happening *in vivo* for the chromosome.

Beneficial mutations play a central role in evolution, but they are only a tiny fraction of all mutations (FISHER 1930; KIMURA 1983) and are therefore difficult to identify and investigate. Evolution experiments with bacteria and other microorganisms offer the opportunity to study the adaptive process in the laboratory (ELENA and LENSKI 2003). Even so, sequencing dozens of 500-bp regions chosen at random after 20,000 generations in the long-term evolution experiment with *E. coli* revealed almost no mutations (LENSKI *et al.* 2003). By contrast, parallel phenotypic changes in the replicate evolving population clones led to the discovery of two loci (*rbs* and *spoT*) in which beneficial mutations occurred in most or all of the populations (COOPER *et al.* 2001a, 2003). Here, we discovered parallel changes in DNA topology, and we used this information to identify two loci, *topA* and *fis*, that acquired beneficial mutations in a focal population.

The *topA* mutation replaced the histidine at position 33, which is a conserved polar residue (CHEN and WANG 1998), with a tyrosine. Although it was previously shown that an H33A mutant topoisomerase I is unaffected in its relaxation catalytic activity (CHEN and WANG 1998),

our data indicate that the H33Y mutant protein causes an increase in DNA supercoiling, probably reflecting a conformational change. Moreover, the H33 residue lies in domain I in the structure of the 67K N-terminal fragment of topoisomerase I (LIMA *et al.* 1994) and, together with other conserved residues, it surrounds the core of residues that cluster around the active-site tyrosine. Domain I resembles a nucleotide-binding domain and may interact with one end of the cut DNA (LIMA *et al.* 1994). Our H33Y mutant probably retains some activity because complete inactivation of *topA* is lethal unless compensatory mutations in the *gyrAB* genes decrease supercoiling (DINARDO *et al.* 1982; PRUSS *et al.* 1982), and we found no such compensatory mutations. The *fis* mutation lies 4 bases upstream of the ATG translational start codon, and it reduced the level of Fis protein (Figure 3). Fis represses transcription of *gyrAB* and reduces the activity of DNA gyrase (SCHNEIDER *et al.* 1999), which may explain the increased level of supercoiling caused by the evolved *fis* allele. We have also found mutations in *topA*, *fis*, or both in some other evolved populations in addition to the focal population (D. SCHNEIDER, C. L. WINKWORTH, E. CROZAT, H. GEISELMANN, M. A. RILEY and R. E. LENSKE, unpublished data). A full survey and characterization of the mutations substituted at these loci in all 12 populations is now in progress, and the identity of these other mutations may provide further insight into the molecular basis of the adaptive changes at these loci.

Why does an increase of DNA supercoiling enhance fitness of the bacteria? The exact nature of the physiological link may be difficult to decipher, because changes in DNA topology simultaneously influence the expression of many genes (JOVANOVIĆ and LEBOWITZ 1987; PRUSS and DRLICA 1989; STECK *et al.* 1993; GMUENDER *et al.* 2001). One interesting possibility is that transcription initiation from the promoters synthesizing stable RNAs is highly dependent on the level of DNA supercoiling (OOSTRA *et al.* 1981; FREE and DORMAN 1994; FIGUEROA-BOSSI *et al.* 1998). An increase in supercoiling would facilitate transcription of the rRNA operons. A higher rate of rRNA synthesis could be advantageous because the evolved lines have substantially higher exponential growth rates than does the ancestor (VASI *et al.* 1994; COOPER *et al.* 2001b). Cells growing faster may need more rRNA than cells growing slowly, and the ratio of RNA to DNA appears to have increased in the evolved lines (LENSKE *et al.* 1998). A seemingly contradictory result to that hypothesis is the observed decline in the level of Fis, which activates rRNA transcription (ROSS *et al.* 1990). However, the loss of *fis* *in vivo* has no obvious effect on transcriptional activity of rRNA operons, owing to a compensatory mechanism (ROSS *et al.* 1990), which has been suggested to involve an increase in the superhelical density of the DNA after *fis* mutation (AUNER *et al.* 2003).

In addition to *topA* and *fis* mutations that were substi-

tuted in the focal population, another mutation that increased DNA supercoiling also reached high frequency but was eventually lost. This mutation was discovered by examining the time course of the spread of the *topA* mutation, which became “stuck” at a frequency of ~50% from 500 to 1000 generations, despite its large selective advantage measured in direct competition between isogenic strains. This dynamic suggested clonal interference (GERRISH and LENSKE 1998), which was confirmed by finding that co-occurring clones from generation 1000 that lacked the evolved *topA* allele nonetheless had similarly high fitness. Moreover, these co-occurring clones had increased DNA supercoiling, indicating that they, too, carried a mutation affecting some topology-related gene. The identity of that other mutation is not known, and it was presumably eventually out-competed as additional beneficial mutations arose in the subpopulation carrying the beneficial evolved *topA* allele. All these results show that DNA supercoiling was a very important aspect of fitness in the focal population, providing one of the earliest and largest adaptive changes.

The identification of *topA* and *fis* mutations gives a total of four confirmed beneficial mutations that were substituted in the focal population Ara-1, with the others being a deletion of the *rbs* operon (COOPER *et al.* 2001a) and a point mutation in *spoT* (COOPER *et al.* 2003). Assuming that the measured fitness effects of these four mutations were additive, they would collectively increase fitness by ~28%, which is still well below the overall improvement of ~67% measured after 20,000 generations (Figure 4; COOPER and LENSKE 2000). Evidently, some beneficial mutations remain to be discovered, which is consistent with the estimate that each evolved population would have substituted between 10 and 20 beneficial mutations (LENSKE 2004). The eventual identification of all the beneficial mutations will allow the reconstruction of the entire adaptive evolutionary history of this focal population by successively introducing the various beneficial mutations into the ancestral background.

We thank Akira Ishihama (Nippon Institute for Biological Science) and Michael Cashel (National Institutes of Health) for the generous gifts of anti-Fis and anti-RpoA antibodies, respectively. We thank M. Cashel for critical reading of the manuscript. This work was supported by grants from the French Centre National de Recherche Scientifique, Rhône-Alpes Region, and Université Joseph Fourier and from the National Science Foundation (to R.E.L.).

#### LITERATURE CITED

- AUNER, H., M. BUCKLE, A. DEUFEL, T. KUTATELADZE, L. LAZARUS *et al.*, 2003 Mechanism of transcriptional activation by FIS: role of core promoter structure and DNA topology. *J. Mol. Biol.* **331**: 331–344.
- BALKE, V. L., and J. D. GRALLA, 1987 Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *J. Bacteriol.* **169**: 4499–4506.
- BHRIAIN, N. N., C. J. DORMAN and C. F. HIGGINS, 1989 An overlap

- between osmotic and anaerobic stress responses: a potential role for DNA supercoiling in the coordinate regulation of gene expression. *Mol. Microbiol.* **3**: 933–942.
- BLISKA, J. B., and N. R. COZZARELLI, 1987 Use of site-specific recombination as a probe of DNA structure and metabolism *in vivo*. *J. Mol. Biol.* **194**: 205–218.
- BLOMFIELD, I. C., V. VAUGHN, R. F. REST and B. I. EISENSTEIN, 1991 Allelic exchange in *Escherichia coli* using the *Bacillus subtilis* *sacB* gene and a temperature-sensitive pSC101 replicon. *Mol. Microbiol.* **5**: 1447–1457.
- BOLES, T. C., J. H. WHITE and N. R. COZZARELLI, 1990 Structure of plectonemically supercoiled DNA. *J. Mol. Biol.* **213**: 931–951.
- BULL, J. J., M. R. BADGETT, H. A. WICHMAN, J. P. HUELSENBECK, D. M. HILLIS *et al.*, 1997 Exceptional convergent evolution in a virus. *Genetics* **147**: 1497–1507.
- CHAMPOUX, J. J., 2001 DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* **70**: 369–413.
- CHEN, S.-J., and J. C. WANG, 1998 Identification of active site residues in *Escherichia coli* DNA topoisomerase I. *J. Biol. Chem.* **273**: 6050–6056.
- COOPER, T. F., D. E. ROZEN and R. E. LENSKI, 2003 Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **100**: 1072–1077.
- COOPER, V. S., and R. E. LENSKI, 2000 The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* **407**: 736–739.
- COOPER, V. S., D. SCHNEIDER, M. BLOT and R. E. LENSKI, 2001a Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli*. *J. Bacteriol.* **183**: 2834–2841.
- COOPER, V. S., A. F. BENNETT and R. E. LENSKI, 2001b Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. *Evolution* **55**: 889–896.
- DI NARDO, S., K. A. VOELKEL, R. STERNGLANZ, A. E. REYNOLDS and A. WRIGHT, 1982 *Escherichia coli* DNA topoisomerase I mutants have compensatory mutations in DNA gyrase genes. *Cell* **31**: 43–51.
- DORMAN, C. J., and P. DEIGHAN, 2003 Regulation of gene expression by histone-like proteins in bacteria. *Curr. Opin. Genet. Dev.* **13**: 179–184.
- ELENA, S. F., and R. E. LENSKI, 2003 Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* **4**: 457–469.
- FIGUEROA-BOSSI, N., M. GUÉRIN, R. RAHMOUNI, M. LENG and L. BOSSI, 1998 The supercoiling sensitivity of a bacterial tRNA promoter parallels its responsiveness to stringent control. *EMBO J.* **17**: 2359–2367.
- FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.
- FREE, A., and C. J. DORMAN, 1994 *Escherichia coli* *tyrT* gene transcription is sensitive to DNA supercoiling in its native chromosomal context: effect of DNA topoisomerase IV overexpression on *tyrT* promoter function. *Mol. Microbiol.* **14**: 151–161.
- GELLERT, M., K. MIZUUCHI, M. H. O'DEA and H. A. NASH, 1976 DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc. Natl. Acad. Sci. USA* **73**: 3872–3876.
- GERRISH, P. J., and R. E. LENSKI, 1998 The fate of competing beneficial mutations in an asexual population. *Genetica* **102/103**: 127–144.
- GMUENDER, H., K. KURATLI, K. DI PADOVA, C. P. GRAY, W. KECK *et al.*, 2001 Gene expression changes triggered by exposure of *Haemophilus influenzae* to novobiocin or ciprofloxacin: combined transcription and translation analysis. *Genome Res.* **11**: 28–42.
- GOLDSTEIN, E., and K. DRLICA, 1984 Regulation of bacterial DNA supercoiling: plasmid linking numbers vary with growth temperature. *Proc. Natl. Acad. Sci. USA* **81**: 4046–4050.
- HATFIELD, G. W., and C. J. BENHAM, 2002 DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu. Rev. Genet.* **36**: 175–203.
- HIGGINS, C. F., C. J. DORMAN, D. A. STIRLING, L. WADDELL, I. R. BOOTH *et al.*, 1988 A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *Salmonella typhimurium* and *Escherichia coli*. *Cell* **52**: 569–584.
- JOVANOVICH, S. B., and J. LEBOWITZ, 1987 Estimation of the effect of coumermycin A1 on *Salmonella typhimurium* promoters by using random operon fusions. *J. Bacteriol.* **169**: 4431–4435.
- KAREM, K., and J. W. FOSTER, 1993 The influence of DNA topology on the environmental regulation of a pH-regulated locus in *Salmonella typhimurium*. *Mol. Microbiol.* **10**: 75–86.
- KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.
- LENSKI, R. E., 2004 Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. *Plant Breed. Rev.* **24**: 225–265.
- LENSKI, R. E., and M. TRAVISANO, 1994 Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl. Acad. Sci. USA* **91**: 6808–6814.
- LENSKI, R. E., M. R. ROSE, S. C. SIMPSON and S. C. TADLER, 1991 Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138**: 1315–1341.
- LENSKI, R. E., J. A. MONGOLD, P. D. SNIEGOWSKI, M. TRAVISANO, F. VASI *et al.*, 1998 Evolution of competitive fitness in experimental populations of *E. coli*: What makes one genotype a better competitor than another? *Antonie van Leeuwenhoek* **73**: 35–47.
- LENSKI, R. E., C. L. WINKWORTH and M. A. RILEY, 2003 Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *J. Mol. Evol.* **56**: 498–508.
- LILLEY, D. M. J., 1986 Bacterial chromatin. A new twist to an old story. *Nature* **320**: 14–15.
- LIMA, C. D., J. C. WANG and A. MONDRAGON, 1994 Three-dimensional structure of the 67K N-terminal fragment of *E. coli* DNA topoisomerase I. *Nature* **367**: 138–146.
- LINK, A. J., D. PHILLIPS and G. M. CHURCH, 1997 Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *J. Bacteriol.* **179**: 6228–6237.
- MALLIK, P., T. S. PRATT, M. B. BEACH, M. D. BRADLEY, J. UNDRAMATLA *et al.*, 2004 Growth phase-dependent regulation and stringent control of *fis* are conserved processes in enteric bacteria and involve a single promoter (*fis P*) in *Escherichia coli*. *J. Bacteriol.* **186**: 122–135.
- NOSIL, P., B. J. CRESPI and C. P. SANDOVAL, 2002 Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* **417**: 440–443.
- OOSTRA, B. A., A. J. VAN VLIET, G. AB and M. GRUBER, 1981 Enhancement of ribosomal ribonucleic acid synthesis by deoxyribonucleic acid gyrase activity in *Escherichia coli*. *J. Bacteriol.* **148**: 782–787.
- PAPADOPOULOS, D., D. SCHNEIDER, J. MEIER-EISS, W. ARBER, R. E. LENSKI *et al.*, 1999 Genomic evolution during a 10,000-generation experiment with bacteria. *Proc. Natl. Acad. Sci. USA* **96**: 3807–3812.
- PRUSS, G. J., and K. DRLICA, 1989 DNA supercoiling and prokaryotic transcription. *Cell* **56**: 521–523.
- PRUSS, G. J., S. H. MANES and K. DRLICA, 1982 *Escherichia coli* DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. *Cell* **31**: 35–42.
- REYES-DOMINGUEZ, Y., G. CONTRERAS-FERRAT, J. RAMIREZ-SANTOS, J. MEMBRILLO-HERNANDEZ and M. C. GOMEZ-EICHELMANN, 2003 Plasmid DNA supercoiling and gyrase activity in *Escherichia coli* wild-type and *rpoS* stationary-phase cells. *J. Bacteriol.* **185**: 1097–1100.
- ROSS, W., J. F. THOMPSON, J. T. NEWLANDS and R. L. GOURSE, 1990 *Escherichia coli* Fis protein activates ribosomal RNA transcription *in vitro* and *in vivo*. *EMBO J.* **9**: 3733–3742.
- SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Plainview, NY.
- SCHNEIDER, D., E. DUPERCHY, E. COURSENGE, R. E. LENSKI and M. BLOT, 2000 Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics* **156**: 477–488.
- SCHNEIDER, R., A. TRAVERS, T. KUTATELADZE and G. MUSKHELISHVILI, 1999 A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol. Microbiol.* **34**: 953–964.
- SNIEGOWSKI, P. D., P. J. GERRISH and R. E. LENSKI, 1997 Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature* **387**: 703–705.
- STECK, T. R., R. J. FRANCO, J. Y. WANG and K. DRLICA, 1993 Topo-

- isomerase mutations affect the relative abundance of many *Escherichia coli* proteins. *Mol. Microbiol.* **10**: 473–481.
- VASI, F., M. TRAVISANO and R. E. LENSKI, 1994 Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. *Am. Nat.* **144**: 432–456.
- WANG, J. C., 1971 Interaction between DNA and an *Escherichia coli* protein omega. *J. Mol. Biol.* **55**: 523–533.
- WEINSTEIN-FISCHER, D., M. ELGRABLY-WEISS and S. ALTUVIA, 2000 *Escherichia coli* response to hydrogen peroxide: a role for DNA supercoiling, topoisomerase I and Fis. *Mol. Microbiol.* **35**: 1413–1420.
- WEITAO, T., K. NORDSTRÖM and S. DASGUPTA, 2000 *Escherichia coli* cell cycle control genes affect chromosome superhelicity. *EMBO Rep.* **1**: 494–499.
- YANISCH-PERRON, C., J. VIEIRA and J. MESSING, 1985 Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103–119.
- ZECHIEDRICH, E. L., A. B. KHODURSKY, S. BACHELLIER, R. SCHNEIDER, D. CHEN *et al.*, 2000 Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. *J. Biol. Chem.* **275**: 8103–8113.

Communicating editor: G. R. SMITH

# RÉSULTATS

## PARTIE II

### EVOLUTION BACTÉRIENNE ET SUPERHÉLICITÉ DE L'ADN : DEGRÉ EXCEPTIONNEL DE PARALLÉLISME GÉNÉTIQUE ET PHÉNOTYPIQUE

Cette partie s'inscrit dans la continuité des travaux décrits dans l'article précédent et fait l'objet de la publication ci-jointe intitulée « Bacterial Evolution by Exceptional Levels of Parallel Genetic Changes in DNA Superhelicity » (E. Crozat, C. Winkworth, P. Fisher Hallin, M.A. Riley, R.E. Lenski and D. Schneider, à soumettre à la revue PLoS Biology).

Dans la première partie des résultats (Crozat et al., 2005), nous avons démontré une augmentation parallèle de la superhélicité négative de l'ADN dans 10 des 12 populations, cette augmentation intervenant dès 2000 générations pour la majorité des populations. Au sein de la population modèle Ara-1, la découverte et l'étude de mutations responsables de ces changements de superhélicité nous ont permis de suggérer que ces modifications étaient liées à une sélection positive au cours de l'évolution. Dans cette seconde partie des résultats, nous nous sommes intéressés aux changements génétiques liés aux modifications de superhélicité de l'ADN dans les 11 autres populations, afin de répondre aux questions suivantes : le parallélisme phénotypique observé au niveau de la superhélicité de l'ADN s'explique-t-il par un parallélisme génétique ? Où sont localisées les mutations responsables de l'augmentation de superhélicité ? Quels sont leurs effets ? Les mêmes gènes sont-ils touchés par des mutations dans la majorité des populations ? Si oui, trouve-t-on les mêmes mutations plusieurs fois de façon indépendante ?

Au moins deux hypothèses peuvent être envisagées a priori. A un extrême, il se peut que seul un accroissement de superhélicité de l'ADN soit important pour l'adaptation des bactéries à leur environnement. Dans ce cas, les différentes populations peuvent présenter des mutations dans tous types de gènes impliqués dans la superhélicité, pourvu qu'elles conduisent à une augmentation de superhélicité. A l'autre extrême, l'augmentation de fitness pourrait être liée à des fonctions particulières d'un petit nombre de gènes impliqués dans le contrôle de la superhélicité. Dans ce cas, seuls certains gènes de contrôle de la topologie seront les cibles de la sélection naturelle. Ainsi, la mise en évidence du degré de parallélisme génétique à la base des changements de superhélicité de l'ADN pourrait nous éclairer sur les mécanismes mis en jeu.

Pour répondre à ces questions, nous avons séquencé 9 loci (17 gènes) connus pour être impliqués dans la régulation de la topologie de l'ADN. Ces loci comprennent des gènes codant les topoisomérases (*gyrA*, *gyrB*, *parC*, *parE*, *topA*, *topB*) et des protéines de type histone (*dps*,

*fis*, *hns*). Dans les cas où les gènes se situent dans des opérons (*fis*, *parC*, *parE* et *topB*), la totalité de ceux-ci ont été séquencés, ce qui donne un total de 17 gènes et 23 102pb. Chacun de ces loci a été séquencé dans un clone évolué isolé à 20 000 générations de chacune des 12 populations. Un total de 27 mutations a été mis en évidence (Tableau 1 du manuscrit joint).

Ces différentes séquences ont été comparées à celles réalisées pour 36 gènes choisis au hasard (pour un total de 18 374pb) dans des clones évolués isolés à 20 000 générations dans chacune des 12 populations (Woods et al., 2006). Cette dernière étude n'a permis d'isoler que 6 mutations au sein de ces 36 gènes, toutes apparues dans des gènes différents et dans les populations devenues mutatrices au cours de l'évolution. (Il est important de rappeler ici que 4 populations sur 12 présentent, bien avant 20 000 générations, un taux de mutation 100 fois supérieur à celui de l'ancêtre, suite à des mutations dans les gènes de réparation des mésappariements). Quatre tests statistiques ont été utilisés pour comparer ces deux catégories de gènes (liés à la topologie de l'ADN ou choisis au hasard).

Un degré exceptionnel de parallélisme génétique a été mis en évidence, avec seulement 3 gènes (*topA*, *fis* et *dusB*) soumis à sélection positive au cours de l'évolution. Malgré ce parallélisme génétique, les mutations identifiées dans chacun des 3 gènes dans les différentes populations sont différentes les unes des autres. L'analyse moléculaire des différents allèles évolués du gène *fis* (modifié aux niveaux transcriptionnel, traductionnel et de la séquence codante) nous a permis de montrer que ces mutations divergentes conduisaient cependant à des effets moléculaires similaires : une diminution de la quantité ou de l'activité de la protéine Fis. Cette seconde partie des résultats nous a permis également d'identifier de nouvelles mutations bénéfiques, encore inconnues dans ces populations et affectant le gène *dusB*. De plus, nos résultats suggèrent des fonctions communes entre *fis* et *dusB*. Enfin, l'effet du contexte génétique sur l'accroissement de superhélicité de l'ADN a été étudié. Tous ces résultats sont présentés en anglais sous la forme de l'article suivant.

Cette étude a été complétée par l'analyse de l'effet de l'environnement sur le fitness conféré par les mutations isolées au sein de la population Ara-1 affectant le niveau de superhélicité de l'ADN. En effet, l'adaptation prolongée au milieu de l'évolution est associée à une augmentation de la superhélicité de l'ADN, mais celle-ci pourrait avoir des effets pleiotropes dans d'autres milieux. Nous avons donc réalisé des compétitions entre chacune des souches décrites précédemment (Croizat et al., 2005), 606<sup>*topA-1*</sup>, souche isogénique à l'ancêtre

sauf pour la mutation *topA* de la population Ara-1, 606<sup>*fis-1*</sup>, isogénique à l'ancêtre sauf pour la mutation *fis* de la population Ara-1, et la souche ancêtre de départ. Les résultats des expériences de compétitions sont résumés dans la **Figure 28**. Les compétitions sont réalisées dans les mêmes conditions que celles décrites dans la Partie I des résultats. Seul l'environnement, c'est-à-dire le milieu de culture ou la température d'incubation, est différent. Les effets initialement mesurés par des expériences de compétitions dans le milieu DM25 de l'évolution expérimentale sont une augmentation du fitness de 13% pour la souche 606<sup>*topA-1*</sup> par rapport à l'ancêtre et de 3% pour la souche 606<sup>*fis-1*</sup>. Dans tous les cas et dans tous les milieux testés, le marqueur arabinose a été vérifié comme étant neutre par compétition de l'ancêtre Ara+ (REL607) contre l'ancêtre Ara- (REL606).

En milieu riche (LB) et à 37°C, les deux mutations *fis* et *topA* apportent chacune un bénéfice considérable de 65 à 70% par rapport à l'ancêtre. Cette augmentation de fitness est visible après seulement 24h de compétition. Pour donner un ordre de grandeur, rappelons que les clones évolués des 12 populations ont en moyenne un fitness accru de 70% après 20 000 générations d'évolution et qu'ils possèdent plusieurs dizaines de mutations bénéfiques. Quand le milieu LB est dilué 2 fois, les valeurs de fitness sont toujours élevées : +22% pour la mutation *topA* et +6,5% pour la mutation *fis*. Ces valeurs sont intermédiaires entre celles mesurées en DM25 et en LB. Un lien potentiel entre taux de croissance et bénéfice apporté par ces deux mutations pourrait exister.

Nous avons également testé l'effet des mutations dans un milieu contenant une autre source de carbone. Des compétitions ont été réalisées dans les mêmes conditions que l'évolution (milieu minimum DM25, 37°C) mais le glucose a été remplacé par du maltose. Encore une fois, les deux mutations confèrent une augmentation de fitness de 6% pour la mutation *topA* et 2% pour la mutation *fis*.

Enfin, une sensibilité au froid ayant été observée dans des mutants  $\Delta topA$  (Stupina et Wang, 2005), nous avons testé l'effet d'une baisse de température sur le fitness de la souche 606<sup>*topA-1*</sup>. A 30°C, cette mutation s'est également révélée bénéfique en augmentant le fitness de 12%, ce qui est similaire à l'effet observé sur le fitness à 37°C. Ceci confirme l'effet modéré de la mutation sur l'activité de la topoisomérase I, celle-ci étant diminuée, mais pas supprimée.

# Bacterial Evolution by Exceptional Levels of Parallel Genetic Changes in DNA Superhelicity

Running head: Evolutionary changes in DNA superhelicity

**Estelle Crozat<sup>1</sup>, Cynthia Winkworth<sup>2</sup>✉, Peter Fisher Hallin<sup>3</sup>, Margareth A. Riley<sup>2</sup>,  
Richard E. Lenski<sup>4</sup> and Dominique Schneider<sup>1\*</sup>**

**1** Laboratoire Adaptation et Pathogénie des Microorganismes, Université Joseph Fourier, CNRS UMR 5163, F-38041 Grenoble, France, **2** Department of Biology, University of Massachusetts, Amherst, MA 01003, **3** Center for Biological Sequence Analysis, Technical University of Denmark, 2800, Lyngby, Denmark, **4** Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824

✉ Current address: Department of Zoology, University of Otago, Dunedin, New Zealand

\* To whom correspondence should be addressed. Email: dominique.schneider@ujf-grenoble.fr

## Summary

DNA supercoiling, the master bacterial function interconnecting chromosome structuration and global gene transcription, was recently shown to be under positive selection during evolution of 12 initially identical populations of *Escherichia coli* that have been propagated in a defined environment (alternating between growth phase and starvation conditions) for 20,000 generations. Parallel increases in DNA supercoiling were observed in 10 of the 12 populations and fitness-enhancing mutations have been found in one studied population in the *topA* and *fis* genes, encoding respectively topoisomerase I and one of the major histone-like proteins. As a first step in deciphering the molecular bases of DNA supercoiling as the target of natural selection, we quantified the level of phenotypic, molecular and genetic parallelism of DNA supercoiling in all 12 evolving populations. Such quantitation of evolution repeatability is inherently difficult due to two major hints, which can be circumvented by experimental evolution strategies, i.e. subtle differences in both environmental conditions and founding genotypes. By sequencing 9 DNA topology-related loci (17 genes in total) in all 12 populations, we found an exceptional level of genetic parallelism, with mutations in only 3 genes (*topA*, *fis* or *dusB*) found to be under positive selection. The other topology-related genes behave no differently from a set of 36 randomly sequenced genes. Evolved alleles discovered in a single gene were highly divergent in different populations, and this allelic divergence was further analyzed in the case of *fis*. Molecular analyses of four different mutations affecting the transcription, translation and protein activity of Fis revealed an exceptional level of parallelism, leading to highly similar outcomes : a decrease either in the amount or in the activity of Fis. Moreover, artificially increasing DNA supercoiling in one of the two populations in which no DNA topology changes was initially observed led to a strong fitness increase, indicating the absence of historical contingency. Exceptional genetic parallelism therefore indicates that changes in DNA superhelicity drove the evolution of these bacterial populations. Divergent mutations at similar topology-related genes led to exceptional parallel molecular outcomes, finally explaining phenotypic adaptation. Exceptional level of molecular and genetic parallelism, targeting a small subset of genes even inside a cellular function like DNA supercoiling, can give insights into the molecular mechanisms underlying evolutionary processes. As a molecular bonus, we identified a novel positively-selected gene, *dusB*. Our statistical analyses revealed that *fis* and *dusB* might have related functions, and this may have some significance in protein function evolution.

## Introduction

DNA superhelicity is one of the most important regulatory networks in bacteria [1], involved in many vital processes, from chromosome replication and segregation, transcription, translation, recombination, to responses to environmental conditions. In particular, it has been suggested that DNA supercoiling can act as both a sensor of external stimuli and a transmitter to the global expression profile of the cells. Hence, DNA superhelicity has a dynamic structural organization, transiently modified during all environmental changes tested so far, like nutritional conditions [2,3], oxidative stress [4], osmotic stress [5], thermal stress [6], pH stress [7], presence or absence of oxygen [8], or intracellular growth of pathogens [9]. Upon different stresses, transient alterations of the DNA superhelicity level generate transient modifications of global transcriptional patterns, including activation of genes involved in stress survival, leading to phenotypic acclimation of bacteria to their environment [4,5]. Changes in DNA superhelicity are known to affect transcription of at least 10% of the *Escherichia coli* genes [10]. The DNA superhelicity control of gene transcription involves two levels: one local whereupon it directly and dynamically alters the physical properties of promoter sequences, for example by influencing the opening of DNA duplex [11], and one global. Hence, the bacterial chromosome is organized first into fluid topological domains of 10kb on average [12,13], and second into large macrodomains of several hundreds of kilobase pairs [14] which may lead to expression coupling of spatially separated genes involved in phenotypic responses to particular stresses. These different levels of chromosomal organization correlate with the existence of spatial patterns of transcription in the *E. coli* genome [15], which can be altered by modulation of the DNA superhelicity level. Structural organization of the chromosome, DNA superhelicity and global transcription patterns are therefore highly interconnected processes allowing cells to cope with fluctuating environmental conditions.

The DNA superhelicity level is precisely adjusted by the combined activities of topoisomerases [16] and histone-like proteins [17]. Topoisomerases are enzymes that introduce or remove supercoils by cutting, transferring and religating DNA strands. In *E. coli*, three major topoisomerases maintain the appropriate level of DNA supercoiling: topoisomerases I, encoded by *topA* [18] and IV, encoded by *parC* and *parE* [19] relax DNA, whereas DNA gyrase, encoded by *gyrA* and *gyrB* is the only known topoisomerase able to introduce negative supercoils into the bacterial chromosome [20]. Binding of histone-like

proteins leads to both local effects on DNA superhelicity like bending, bridging, looping out of DNA or constraining negative supercoils [21], and global effects constituting diffusion barriers for the formation of topological domains [22]. Most histone-like proteins have also been shown to be global regulators of gene transcription in bacteria, like Fis [23], H-NS [24,25], HU [26] or IHF [27]. They are expressed at various levels during the bacterial cell cycle, generating yet another level of highly dynamic interplay between environmental signals, DNA superhelicity and global transcription profiles.

Of primary importance among these histone-like proteins connecting DNA superhelicity and global transcription to environmental signals is the Fis protein [28]. After binding DNA, Fis leads to the extrusion of DNA loops [29]. This constitutes the common mechanism by which Fis can regulate gene expression by local binding to high-affinity sites [28], organize the chromosome at the global level into topological domains by binding non-specifically to DNA [22], or facilitate site-specific recombination events catalysed by serine invertase and tyrosine integrase recombinases [30]. Several lines of evidence suggest that Fis plays a crucial role in the adaptation of bacteria to the nutritional state of the environment by sensing nutrient upshifts and deficiencies, modifying DNA superhelicity at both local and global levels, thereby adjusting global gene expression profiles to provide bacteria with the appropriate phenotypic response: i) Fis exerts a homeostatic control on DNA superhelicity, allowing a physiologically relevant supercoiling level *in vivo*. Fis represses transcription of *gyrA* and *gyrB* [31] and in turn, *fis* transcription is strongly dependent on the level of DNA supercoiling [32]; ii) Nutrient upshifts lead to high negative superhelicity, which triggers high Fis levels (50,000 molecules per cell) [33], which in turn activates transcription of ribosomal RNA operons and other genes necessary to cell growth. Transition to stationary phase is accompanied by DNA relaxation, leading to a dramatic decrease in Fis levels (100 molecules per cell), as well as in transcription of rRNA operons. This favours transcription of genes by the  $\sigma^S$  sigma factor of RNA polymerase, specific to the stationary phase [34–36]. Besides being a sensor of the nutritional state of the environment, Fis also helps bacterial cells to cope with certain stresses, in particular oxidative stress, again through its ability of adjusting the DNA superhelicity level [4]; iii) After binding to numerous sites on the chromosome, Fis acts as a global regulator of gene transcription [23,37], regulating ribosomal RNA operons [38], genes involved in cellular metabolism [39], and virulence genes necessary to intracellular bacterial growth [9].

The last two decades have witnessed the development of experimental evolution strategies, which consist in propagating living organisms for hundreds or thousands of

generations in a defined environment [40], representing a powerful tool to study the dynamics and genetic bases of adaptation. Using the longest-running evolution experiment, we recently further extended the crucial role of DNA supercoiling in cellular life by showing that it can be involved in evolutionary processes [41]. Hence, DNA supercoiling was shown to be one of the major targets of natural selection. In particular, we demonstrated that both a topoisomerase and a histone-like protein could be directly involved in adaptive evolution. In this long-term evolution experiment, Lenski and colleagues have propagated 12 populations of *E. coli* by serial daily transfers for more than 40,000 generations at 37°C in a minimal medium supplemented with limiting amounts of glucose [42–44]. Six of the 12 populations, called Ara-1 to Ara-6, were founded from an *E. coli* B ancestor clone unable to use arabinose as a carbon source (Ara-) [45], while the 6 others (Ara+1 to Ara+6) were founded from an Ara+ revertant of the ancestor clone. The arabinose utilization phenotype was used as an internal neutral marker in competition experiments to assess the fitness of evolved clones compared to the ancestor [46]. All replicate populations achieved parallel gains in competitive fitness, indicative of substantial adaptation. The fitness increase reaches about 70% in all 12 populations after 20,000 generations of evolution [43]. Interestingly, four populations evolved defects in their DNA repair pathways, leading to a hypermutable phenotype [47,48]. During 40,000 generations, evolution is characterized by a strong level of phenotypic parallelism. Hence, several phenotypic traits in addition to fitness evolved in parallel in most populations, including cell size [49], growth parameters [50], catabolic functions [43,51], global gene expression [51,52] and DNA topology [41]. Various genetic studies, specifically aimed to understand these parallel phenotypic changes, revealed that the same genes were reproducibly altered by mutations in most populations : *rbs*, the ribose utilization operon [53]; *spoT*, one of the genes involved in (p)ppGpp metabolism during the stringent response [51,52] and *malT*, encoding the transcriptional activator of the maltose utilization operon [51]. These genetic changes were demonstrated to be beneficial under the environmental conditions of the evolution experiment. We investigated the observed parallel changes in DNA supercoiling levels only in one model population, where two successive increases in DNA superhelicity were detected [41]. We found two mutations, one in *topA*, which encodes topoisomerase I, and one in *fis*, encoding the histone-like protein. By genetically manipulating these mutations, we could demonstrate that both were beneficial in these conditions. However, and this constitutes one of the objectives of this manuscript, this study was not extended in the other 11 populations to analyse the level of genetic parallelism linked to these parallel increases of DNA supercoiling.

An independent strategy, unlinked to parallel phenotypic changes, but using IS elements as genomic markers [54], revealed 4 insertion mutations in one of the 12 populations [55]. These mutations affect *nadR* involved in NAD metabolism, *pykF* encoding pyruvate kinase I, *pbpA-rodA* involved in cell wall biosynthesis and *hokB-sokB* encoding a plasmid maintenance module [55,56]. Sequencing these 4 loci in all 12 populations revealed mutations in most populations [56]. Statistical tests were used to compare these results to a set of 36 randomly chosen genes that were sequenced in evolved clones isolated at 20,000 generations from all 12 populations. Almost no mutations were detected [57]. These statistical tests of the pattern of molecular evolution all indicate that the parallel genetic changes in the 4 loci were driven through adaptation by natural selection [56], although no experimental proof confirmed these conclusions.

Phenotypic and genetic parallelism occurring in several lineages is a hallmark of this evolution experiment. In the conditions prevailing in this experiment, all possible genetic changes are potentially possible, given the predicted mutation rate (about  $2 \cdot 10^{-10}$  mutations/bp/generation), the genome and population sizes and the number of generations [44,57]. Therefore, parallel evolution provides strong evidence of adaptation by natural selection for the parallel evolved traits. Parallel evolution, both at the phenotypic and genetic levels, is widespread in nature. Similar phenotypes evolved repeatedly when independent populations colonize similar environments, including the morphology of lizards and fishes [58,59], the wing length in *Drosophila* [60]. Parallel genomic changes have been described for viral and bacterial pathogens [61,62]. Linkage between environmental parameters and evolution of organismal phenotypes unlikely arises by random genetic drift and strongly involves natural selection. Despite numerous cases of parallel evolution, the only way to precisely quantify parallelism and the level it can reach during evolutionary processes is to measure phenotypic repeatability in independent populations during experimental evolution strategies where the common ancestor is available to experimental and statistical tests. This would however be crucial to understand the underlying molecular bases of adaptation and to use evolutionary biology as a predicting tool. The DNA supercoiling target of natural selection constitutes an ideal candidate to analyse in deep details the level of parallelism of evolutionary processes, since more than 20 genes (encoding topoisomerases, histone-like proteins and gene expression regulators) are involved in its regulation in *E. coli*.

Here, we sought to gain further insights into both the level of parallelism characterizing evolutionary processes and the underlying molecular and ecological mechanisms by analyzing

the relationships between DNA supercoiling and adaptation by natural selection. We raised three different types of questions. We first asked whether the observed parallel changes in DNA supercoiling were due to an underlying genetic parallelism, by sequencing 9 topology-related loci (including *topA* and *fis*) in both the ancestral and evolved clones isolated from all 12 populations. On one extreme, mutations in various genes involved in DNA superhelicity control could explain the observed parallel phenotypic increases in DNA supercoiling. On the other extreme, only a small number of genes may be repeatedly targets of natural selection in the 12 independently evolving populations. Second, as we detected genetic parallelism at the level of the gene, but genetic divergence at the level of mutation alleles, we asked how this is related to the level of molecular gene regulation. Third, only few of the 12 populations did not show any increase in the DNA superhelicity level [41]. We asked whether the absence of changes was linked to any molecular explanation or was simply due to the random processes of mutation fixation within the evolving populations. Our results revealed an exceptional high level of genetic and molecular parallelism. During the course of our work, we also discovered a new beneficial mutation.

## Results

### Evolutionary Trajectories of Topology-Related Genes

We sequenced 9 loci known to be involved in DNA supercoiling regulation in *E. coli*, 6 encoding all topoisomerases and 3 encoding major histone-like proteins, in an evolved clone isolated at 20,000 generations from each of the 12 populations and in both REL606 and REL607 ancestor clones. In 4 cases, the genes are located within operon structures that were entirely sequenced, leading to a total of 17 sequenced genes (Table 1). Both ancestor clones revealed identical sequences for all 17 genes. Compared to both ancestor clones, we detected a total of 27 mutations distributed in the 12 populations.

To evaluate whether any of the candidate genes involved in DNA topology, or any other genes in the same operons, were targets of selection during the evolution experiment, we applied four statistical tests [56]. The first one compares the overall rate of substitution in these genes versus 36 gene regions, chosen at random, that were previously sequenced [57]. For the topology-related genes, we sequenced 23,102 bp, while 18,374 bp were sequenced for the random genes. Table 2 shows that 11 of the 12 populations had higher rates of change in

the topology-related than in the random genes. This difference is highly significant (sign test,  $p = 0.0032$ ) in the direction predicted if one or more of the topology-related genes are undergoing accelerated selection, consistent with being genetic targets of selection. The difference is significant even if we exclude population Ara-1 ( $p = 0.0059$ ), where we knew from our previous research that it substituted beneficial mutations in both *topA* and *fis* [41].

Of the 27 mutations we found in the topology-related genes, 11 (40.7%) were in *topA*, *fis*, or their regulatory regions, although together they constituted only 3,634 bp (15.7%). Another gene, *dusB*, had 5 mutations (18.5%) in only 966 bp of coding sequence (4.2%). The *dusB* gene encodes one of three dihydrouridine synthases in *E. coli* [63]. Dihydrouridine is the most abundant modified base found at conserved positions in the D-loop of tRNA in prokaryotes and eukaryotes, but its physiological role is unknown [63]. The other 11 mutations (40.7%) were dispersed across 10 different genes that comprised 80.1% of the sequencing effort, including two in *ydjA* and one each in *topB*, *parC*, *dps*, *selD*, *yqiB*, *yqiA*, *hns*, *gyrB*, and *gyrA*. Based on these data, we conclude that mutations in *dusB*, as well as those in *topA* and *fis*, give evidence of positive selection based on this first test.

Moreover, the important fact that *dusB* mutations were in five populations (Ara-6, Ara+3, Ara+4, Ara+5, and Ara+6) that evolved changes in DNA supercoiling, while lacking any mutations in *topA* or *fis*, suggests that these mutations were involved with changes in DNA topology. We therefore focus the further statistical tests on *topA*, *fis*, and *dusB*.

The second test compares the pattern of synonymous and non-synonymous substitutions in the randomly chosen genes and this refined set of three topology-related genes. Table 3 shows that the random genes had equal number of synonymous and non-synonymous mutations, while non-synonymous mutations (including insertions and point mutations in coding regions) largely predominated in the topology-related genes. This difference is again in the direction expected if these topology-related genes are under selection for changes in their encoded proteins, although the difference is only marginally significant (Fisher's exact test,  $p = 0.0833$ ).

The third test compares the distribution of point mutations between the four evolved populations that became mutators, owing to evolved defects in DNA repair, and the eight other populations that retained the much lower ancestral mutation rate. Table 4 shows that mutations in the topology genes were more common in the non-mutator populations than in the mutators (about equally common per population). By contrast, all the mutations in the randomly chosen genes were found in the four mutator populations. This difference is in the expected direction under the hypothesis that both mutator and non-mutator populations

accumulate mutations in target genes, whereas mutators differentially accumulate mutations in other loci not under positive selection to change (Fisher's exact test,  $p = 0.0054$ ). Interestingly, and consistent with our earlier separation of *topA*, *fis*, and *dusB* from the other genes in topology-related loci, all 10 point mutations discovered in those other genes were in the 4 mutator lines (1 mutation in these other genes was in a non-mutator population but corresponded to an *ISI* insertion within *dps*). This distributional difference between *topA*, *fis*, and *dusB* and the other mutations in the topology-related genes is also highly significant (Fisher's exact test,  $p = 0.0005$ ). The other topology-related genes therefore behave as the set of random genes.

The fourth, and final, sequence-based test evaluates the distribution of mutations across the 12 independently evolved populations. If the mutations accumulated randomly by genetic drift, and if populations had the same mutation rate, then we would expect a Poisson distribution. Given the variation in mutation rates associated with the losses of DNA repair in four populations, one would expect a highly clustered distribution under the drift hypothesis. By contrast, if selection drives the sequence evolution, and if the different mutations at a given locus, or set of loci, produce the same phenotypic benefit (such that multiple mutations produce diminishing returns), then one would expect a more uniform distribution of mutations across the replicate lines. In the next section, we explicitly test and confirm the supposition that different mutations in the *fis* gene produce very similar phenotypic effects. Two of the candidate genes, *fis* and *dusB*, show maximally uniform distributions, each harboring single mutations in five different populations, although the number of mutations for either gene alone is too small for statistical significance. It is very interesting, however, that if we combine these two genes, which together comprise the full extent of an operon [64], we see that 10 populations each have exactly one mutation in either *fis* or *dusB*. That distribution is maximally uniform and very significantly different from a random Poisson distribution ( $p = 0.0039$ ). This outcome is driven not only by the uniform distribution of mutations at each separate locus but also by a significant negative association between the two loci (Fisher's exact test,  $p = 0.0278$ ). This pattern suggests that future research should examine the molecular and phenotypic interactions of *fis* and *dusB* and their associated gene products.

## Phenotypic Effects of *fis* Mutations

Despite an exceptional level of genetic parallelism, with only three topology-related genes accumulating positively-selected mutations, the mutated alleles within a single gene in

different populations were divergent. Both experimental tests by competition assays [41] and statistical analyses of DNA sequences (this work) demonstrated the fitness-enhancing effects of the mutations in the topology-related genes. In an attempt to further characterize the extent of parallelism associated with evolutionary processes, we analyzed the phenotypic effects of divergent mutations in the same gene across populations. We focused on 4 different mutations targeting the *fis* gene (Table 1). These mutations were substituted in populations Ara+1, Ara-1, Ara-3 and Ara-4 and affected different levels of the *fis* gene regulation: the transcriptional level since the Ara+1 evolved allele lies in the -35 promoter box, translation initiation since the Ara-1 evolved allele lies in the RBS, and activity of the protein since the Ara-3 and Ara-4 evolved alleles both affect the coding sequence (Figure 1A). The *fis* mutation substituted in population Ara-1 is an A-to-C nucleotide change located in the RBS of *fis* and was previously described in [41]. A 3-fold decrease in the Fis amount was detected as a consequence of this mutated allele. The *fis* mutation substituted in population Ara+1 affects the 6<sup>th</sup> position of the *fis* promoter -35 element [65], replacing the T residue at position -31 relative to the +1 transcription start site by a G residue. This 6<sup>th</sup> position was not included in the molecular analysis of the *fis* promoter [65], and the evolved allele did not bring the -35 element closer to the consensus 5'TTGACA3' sequence. The *fis* mutation substituted in population Ara-3 changes the Y51 residue into a serine. The Y51 residue is very-well conserved and is located at the beginning of the helix B of a Fis monomer, which is involved in the formation of the active Fis dimers [66,67]. The *fis* mutation substituted in population Ara-4 leads to a modification of the C-terminal region of Fis. This allele is a frameshift mutation, corresponding to the insertion of an A residue at position +275 relative to the translation start codon. This leads to a change of the C-terminal domain of the protein product. Hence, the last four YGMN aminoacids are replaced by IRHELIQVS. This part of the protein is involved in the DNA-binding activity [66,67].

In order to analyse the phenotypic effects of the mutations, we constructed a set of isogenic strains by allelic replacements in the ancestral genetic background. We had already constructed the two  $606^{fis-1}$  and  $606^{\Delta fis}$  strains [41], corresponding to the ancestor strain with the evolved *fis* allele of Ara-1 population and an in-frame *fis* deletion, respectively. Here, we completed this set of isogenic strains by constructing  $606^{fis+1}$  and  $606^{fis-3}$ , i.e. the ancestral strains bearing respectively the Ara+1 and Ara-3 *fis* evolved alleles. Despite numerous attempts, we were unable to introduce the Ara-4 *fis* allele within the chromosome of the ancestor, for unknown reasons.

As a first phenotypic test, we measured the effect of the *fis* mutations on the amount of the Fis protein by extraction of soluble proteins from early exponential phase cultures, where the level of Fis is maximal, and immunodetection with an antibody raised against Fis (Figure 1B). Each evolved allele (except Ara-4) was analyzed in two genetic backgrounds: the ancestral one using the set of isogenic strains, and the evolved one using an evolved clone isolated at 20,000 generations from each of the Ara-1, Ara+1, Ara-3 and Ara-4 populations (the evolved Ara-4 allele was analyzed only in the evolved background). Both genetic contexts gave similar results for each *fis* allele. The in-frame *fis*-deleted strain and an anti-RpoA ( $\alpha$  subunit of RNA polymerase) antibody were used as controls. Consistent with our previous results [41], the Ara-1 evolved allele resulted in a 3-fold decrease in the Fis level. The same quantitative decrease of Fis amount was detected for the Ara+1 evolved allele (Figure 1B). By contrast, no such decrease was detected for the two other Ara-3 and Ara-4 mutations, consistent with their localization in the coding sequence of Fis.

As a second phenotypic test, we determined whether the effect of the Ara+1 allele occurred at the level of transcription. We constructed single-copy chromosomal *lacZ* transcriptional fusions with both ancestral and Ara+1 evolved *fis* promoters, using bacteriophage  $\lambda$  derivatives [68]. Both fusions were introduced into the chromosomal  $\lambda$  *att* site of the REL606 ancestor clone and we measured the  $\beta$ -galactosidase activities during exponential phase, transition into stationary phase and stationary phase (Figure 1C). The ancestral promoter revealed the typical growth rate control of *fis* transcription, with maximal transcription during exponential phase followed by a decreased activity when cells shift into stationary phase. This growth phase-related regulation of *fis* transcription was maintained for the evolved promoter, although its activity was reduced by about 3-fold compared to the ancestral promoter at all growth phases, thereby confirming the protein data (Figure 1B and C).

As a third phenotypic test, we analyzed the Ara-3 and Ara-4 *fis* mutations by purifying both evolved proteins together with the ancestral one and characterizing their DNA binding activities. As a preliminary experiment, we used an *in vivo* genetic approach previously developed [69]. This *in vivo* screen addresses the DNA binding activities of various Fis proteins. It is based upon two compatible plasmids, one expression plasmid containing the gene encoding *fis* or a variant under the control of the *tac* promoter, and one reporter plasmid carrying the *cat* reporter gene whose expression is repressed by Fis binding [69]. The *cat* gene, encoding chloramphenicol acetyl-transferase and conferring resistance to chloramphenicol, is under the dependence of modified *lac* -35 and -10 promoter elements.

The *lac* promoter has been modified by introducing just upstream of the -35 box, the UP element sequences of the *rrnBP1* promoter to strongly activate transcription, and just downstream of the -10 box a strong Fis binding site. Therefore, binding of a Fis protein will repress *cat* transcription and decrease the resistance to high levels of chloramphenicol of a strain carrying both plasmids. Moreover, the level of resistance decrease will be proportional to the strength of Fis binding. We cloned each of the ancestral, Ara-3 and Ara-4 *fis* genes in the expression plasmid under the control of *Ptac*, introduced each of them together with the reporter plasmid into the ancestral strain carrying the in-frame *fis* deletion, and plated all constructed strains onto two types of LB plates, one containing ampicillin (to select the presence of the expression plasmid) and 100 µg/mL of chloramphenicol and one containing ampicillin and 600 µg/mL of chloramphenicol. Percent survival (number of colonies on plates with high chloramphenicol concentration per number of colonies on plates with low chloramphenicol concentration) was used as a measure of Fis binding, as described [69]. Survival was about 10-fold higher when the expression plasmid carried either the Ara-3 or the Ara-4 evolved *fis* gene (3.5% and 3.4% respectively) compared to the one carrying the ancestral gene (0.4%), indicating a DNA binding defect for the evolved proteins (data not shown).

To validate these data, ancestral and evolved Fis proteins were purified as previously described [70] and we assessed their DNA binding abilities by performing gel mobility shift assays. Three representative promoter sequences, all known to contain Fis binding sites, were selected for use as binding targets : the *fis* promoter [64], the *rrnBP1* promoter [71], and the *proP* promoters [72] (Figure 2). In the case of the Fis ancestral protein, a shift in electrophoretic mobility was seen at the lowest concentration (0.2 nM) of Fis used (Figure 2A). Multiple retarded bands were detected typical of the presence of several Fis binding sites [73]. Maximal retardation was observed at a Fis concentration of 5 nM. Both Fis evolved proteins revealed much less DNA binding ability, with higher concentrations of protein (2 to 5 nM) needed to observe a shift (Figure 2A). Also, maximal retardation was observed for concentrations of 30 nM or higher. Specificity of binding was checked by using first a 5 to 1000-fold excess of a cold 30 bp-oligonucleotide containing a consensus Fis binding site, second the same DNA fragment but with the Fis site mutated, and third an antibody against Fis in the DNA binding reactions. The first control reversed the binding observed in Figure 2A, while the second had no detectable effect on binding, and the third one generated a strong supershift (data not shown). To precisely compare the behaviour of the Fis ancestral and evolved proteins, we quantified the amount of DNA bound by Fis based on the gel mobility

shift assays. Figure 2B shows a typical quantitation curve and Figure 2C the dissociation constants of each of the three proteins for each of the three promoters regions. The  $Kd$  values are about 15 times higher for both evolved proteins compared to the ancestral protein, thus confirming their decreased DNA binding ability. We therefore demonstrated that both mutations in the *fis* coding region led to a decreased activity of the protein.

To summarize this section, a strong genetic parallelism was observed during the long-term evolution, with *fis* being repeatedly a target of natural selection. This was however achieved through different mutations, affecting all regulation levels: transcription, translation, and activity of the protein. Despite this mutational divergence, they all lead to a strong molecular parallelism yielding a decrease either in the amount or in the activity of Fis in the evolved cells.

## Effect of the Genetic Context on the Evolutionary Trajectory of DNA Supercoiling

In our previous study [41], we demonstrated parallel phenotypic increases in DNA supercoiling in most of the 12 independently evolving populations. Here, we showed that an exceptional level of genetic parallelism underlies these phenotypic changes and both studies reveal that DNA topology-related genes, in particular *topA*, *fis* and *dusB* are the targets of natural selection. However, two populations (Ara+1 and Ara+3) did not reveal this phenotypic pattern after 20,000 generations of evolution, but still shows identical fitness increases. We therefore asked for the non-mutator population Ara+1 if the genetic context, i.e. the past history of that population, may explain the absence of DNA supercoiling changes. To do so, we introduced the *topA* evolved allele that was substituted in population Ara-1 [41], and which strongly increases both fitness (+13%) and DNA superhelicity (from -0.066 to -0.075), in the genetic context of an evolved clone isolated at 2,000 generations from population Ara+1, called 1158C, which carries the *topA* ancestral allele. Hence, no *topA* mutation was substituted in that Ara+1 population. The 1158C clone shows only a slight increase in DNA supercoiling at this evolutionary time ( $\sigma=-0.069$  instead of -0.066 for the ancestor) [41], and has already undergone a 22%-fitness increase compared to the ancestor (Student test,  $p=1.4977 \cdot 10^{-6}$ ). The 1158C and its 1158C<sup>*topA*-1</sup> counterpart, carrying the Ara-1 *topA* allele, were compared for their levels of DNA supercoiling and fitness. Replacing the *topA* ancestral allele by the *topA*<sup>-1</sup> evolved allele increased the DNA supercoiling in the 1158C clone, at a

level equivalent to the one conferred by the *topA*<sup>-1</sup> allele in the ancestor clone (Figure 3A). We also measured the fitness conferred by the *topA*<sup>-1</sup> evolved allele in the genetic context of the Ara+1 population by competing each of the 1158C and 1158C<sup>*topA*<sup>-1</sup></sup> against the ancestor REL606 (Figure 3B). The Ara-1 evolved *topA* allele increased fitness by 13% in the ancestral background [41] and by 9% in the Ara+1 evolved clone background, and this increase is significant (Student test, p=0.0191).

In at least one population showing no phenotypic increase in DNA supercoiling during the evolution experiment, introduction of an evolved allele known to associate a fitness increase to the modification of the DNA topology generated similar phenotypic changes. The absence of substitution of supercoiling-related changes in that particular population was therefore not the result of historical contingency, i.e. the substitution of a previous mutation either having the same effect as *topA* or antagonizing its effect.

## Discussion

In a previous work, we described parallel phenotypic increases in DNA supercoiling in 10 out of the 12 evolved populations [41]. Our model population showing 2 successive steps of DNA supercoiling increase was further analyzed and we identified two mutations, one in *topA* and one in *fis*, responsible for all the DNA topology changes in this population. Both mutations were also found to enhance fitness, suggesting DNA supercoiling as a target of natural selection [41].

Here, we further confirmed the crucial role of DNA supercoiling in evolutionary processes by extending our genetic and molecular analyses to all 12 evolving populations. Four lines of evidence demonstrate that an exceptional level of genetic parallelism drives the phenotypic parallel changes observed in these evolving populations: i) DNA supercoiling-related genes were the targets of natural selection; ii) by sequencing 9 DNA topology-related loci (17 genes in total), encoding both topoisomerases and histone-like proteins, we found evidence that among those genes, only three were under positive selection: *topA*, *fis*, and *dusB*. Statistical analyses revealed also a putative negative functional link between *fis* and *dusB*. In the course of this work, we therefore discovered new fitness-enhancing mutations, affecting the *dusB* gene. Together with other studies [41,51–53], this is a crucial step for a comprehensive and detailed description of all beneficial mutations. Hence, beneficial

mutations play central roles in evolution, but they are difficult to identify and investigate [74,75]; iii) despite this exceptionally high level of genetic parallelism, a high allelic divergence was observed in the different populations for the same three genes. By analysing the *fis* alleles, we however demonstrated that the different mutations, while affecting almost all regulation levels of *fis* (transcription, translation, protein activity), generated again an exceptional level of molecular parallelism by reducing either the amount or the activity of Fis. This molecular parallelism may therefore explain the parallel increase of DNA supercoiling associated to enhanced fitness in these evolution conditions; iv) only 2 of the 12 populations did not show an increase of DNA supercoiling. By introducing the well-described beneficial *topA* mutation from our model population [41] into an evolved clone from one of these 2 populations, we could show that the same phenotypic changes were linked to this *topA* allele in this different genetic context : increase in DNA supercoiling associated to enhanced fitness. This further reveals the absence of historical contingency. Therefore, the exceptional level of phenotypic, genetic and molecular parallelism, through divergent mutational pathways, is a hallmark of this evolution experiment.

Parallel evolution is usually considered as a strong evidence of adaptation by natural selection. However, despite many examples of parallel evolution in nature, reproducibility of evolutionary processes, i.e. the percentage of parallel outcomes out of the total number of mutational events, is difficult to assess mainly for two reasons: differences in environmental conditions and in ancestral genotypes. Furthermore, complex organisms like bacteria are characterized by interconnected cellular functions, each one involving a large number of genes. It is also not known quantitatively what level of genetic parallelism underlies phenotypic parallel changes in a single cellular function. On one extreme, every single gene involved in that cell function could be the target of natural selection, whereas on the other extreme, only a few genes regulating that function could be modified through mutational events. This kind of data may provide invaluable insights into both the molecular mechanisms underlying evolutionary processes, and protein function evolution. Targeting genes involved in the control of DNA superhelicity, we found an exceptional level of genetic parallelism. Hence, 16 out of the 27 identified mutations (about 60%) affected the three genes *topA*, *fis* and *dusB*, which represent only 20% of the sequencing effort. The other 11 mutations are located in the remaining 80% of the sequences, including 10 point mutations found in the 4 populations that have evolved mutator phenotypes. Both statistical and experimental data [41] allowed us to show that mutated alleles identified in these loci represent *bona fide* fitness-enhancing mutations. This high level of genetic parallelism was not detected at the sequence

level, since different mutations in *topA*, *fis* or *dusB* were found in the different populations (Table 1). To the best of our knowledge, only two previous studies quantified the level of genetic parallelism using experimental evolution strategies. Our data are in marked contrast with the parallelism level detected in the first milestone study where independent populations of the  $\Phi$ X174 bacteriophage were propagated on a new host [76]. In that particular case, parallelism was detected even at the level of nucleotide changes in the different evolving populations. Hence, about half of the substitutions were identical. The second study used the same experimental evolution strategy as in this work [56]. However, instead of measuring the genetic parallelism of a single cellular function known to be involved in increased fitness like the DNA supercoiling in our work, they sequenced 4 loci on the basis that Insertion Sequence-associated mutations had previously been detected in those loci in a single evolving population [54–56]. They also found a high genetic parallelism, at the gene level but not at the sequence level, supporting our data. However, no molecular analyses of the different mutations were performed and it was suggested that the discrepancies between the viral and bacterial organisms may reflect differences in genome size and complexity [56]. Another way to reconcile both set of data, still related to differences in genome organization, would be to hypothesize that divergent mutations in genes under positive selection may lead to parallel molecular outcomes. We focused on different mutations affecting *fis* in four different populations. All regulatory levels of the *fis* genes were affected by these four mutations: transcription in the Ara+1 population, translation in Ara-1 and protein activity in both Ara-3 and Ara-4. Both Ara+1 and Ara-1 evolved alleles led to a decreased amount of Fis protein in the cell (Figure 1B), whereas both Ara-3 and Ara-4 alleles decreased the DNA binding activity of the Fis protein (Figure 2). Therefore, even if all these evolved alleles are divergent in nature, they all lead to parallel molecular outcomes, as was observed with the  $\Phi$ X174 bacteriophage. This may also be the case for the four loci studied by Woods *et al.* [56]. The crystal structure of the Fis protein revealed that each monomer of the Fis dimer is constituted by four  $\alpha$ -helices called A to D [66]. The Ara-4 evolved allele is an A residue insertion, leading to a change in the C-terminal part of the protein. Each monomer is therefore a 103 amino-acid protein, instead of 98. The evolved protein has therefore a modified helix D, which, together with helix C, carries the DNA binding helix-turn-helix motif. Also, the helix D inserts into the DNA major groove [66] and the R85, T87, R89, K90 and K93 are involved in DNA interaction. This may explain the reduced DNA binding activity of the Ara-4 evolved protein. In particular the Y95 residue, which is lost in the population Ara-4 participates in van der Waals interactions with residue L63 and they are known to be crucial in configuring the

helix-turn-helix motif in proper orientation [66]. The Ara-3 allele leads to the substitution of the Y51 residue by a S residue. The Y51 residue is located at the beginning of helix B, which comprises amino-acid residues 50 to 70. The N-terminus of each helix B is hydrogen-bounded to the C-terminus of helix C' from the opposite subunit, the four helices B, C', B' and C therefore defining a platform right at the middle of the Fis molecule. In particular, among four main chain hydrogen bonds connecting helices B' and C, the Y51 of one monomer makes contact with the M81 of the other monomer. In addition, the Y51 residue of one helix B also established side chain hydrogen bonds with E59 and K91 residues of helices B' and D' of the other subunit [66], thus being involved in the dimerization of Fis. The Y51 residue is therefore an important part of a hydrogen-bounded network, involved both in maintaining the HTH DNA-binding motif in proper conformation and in allowing Fis dimerization. The substitution of the Y residue for a S probably alters this interaction network, leading to the decrease of DNA binding activity of the evolved Y51S protein. However, we cannot rule out the possibility that it alters Fis dimerization, thereby indirectly reducing DNA binding. The crystal structure of Fis clearly demonstrated that mutations affecting helices C or D, or their interactions with helix B, decrease or eliminate binding to DNA [66]. By contrast, the two other *fis* evolved alleles did not alter the activity of the protein, but reduced its amount, either by affecting the promoter [65] thereby reducing transcription in population Ara+1, or by affecting the RBS element in population Ara-1. In this later case, the A to C mutation reduced the complementarity between the translational initiation region of *fis* mRNA and the 3' end of 16S rRNA [77].

The benefit linked to an increase of DNA supercoiling may be related to optimized transcription from the promoters synthesizing stable RNAs [41], which is known to be strictly dependent upon a high negative superhelicity [78,79]. In the evolution experiment, the cells have substantially improved their growth-rate [50,80], and increased rRNA synthesis may therefore be beneficial. A decrease of Fis levels or activity may therefore at first glance be contradictory, since Fis is an activator of the transcription of rRNA operons [71]. However, removing Fis *in vivo* do not lead to any effect on rRNA synthesis because of a compensatory mechanism involving an increase in DNA superhelicity [81]. Moreover, the exceptional levels of molecular and genetic parallelism, characterized by discovering positively-selected mutations in only a subset of the DNA topology-related genes, may suggest that their benefit could be linked to specific functions of these genes, and not just by any type of DNA supercoiling increase. In that sense, the *fis* evolved alleles may give crucial information about the molecular mechanisms underlying organismal performance.

The histone-like protein Fis is known to bind DNA both specifically to high-affinity sites and non-specifically. Moreover, Fis binding sites differ by their relative affinities for Fis [28,70]. Fis binding results in the formation and stabilization of DNA loops [21,29,82,83]. At the global level, these loop-forming activities are suggested to be involved in Fis being a topological barrier helping the structuration of the entire chromosome [22]. At a more local DNA level, Fis binding to high-affinity sites, often arranged in helical phase, results in strongly bent DNA structures and again formation of loops, thereby locally increasing DNA superhelicity which is required for transcription of some genes like rRNA operons [28]. Fis was therefore suggested to act as a local topological homeostat to preserve intermediately supercoiled forms of DNA, protecting them from global DNA supercoiling fluctuations [84]. Therefore, Fis has marked effects on global gene transcription either by directly regulating genes through upstream local binding or by an indirect global effect by binding non-specifically to DNA and organizing the chromosome into topological domains known to influence transcription [1]. All positively-selected mutations in *fis* in the evolution experiment resulted in decreased amount or activity of Fis. This may have pronounced effects on both global gene transcription and chromosomal organization and our future work will address these questions. On one hand, the decrease of Fis concentration or activity may lead to the loss of regulation of some genes with weak Fis-affinity binding sites. On the other hand, it was shown to lead to a decrease of the number of topological domains in the chromosome [22], which will alter transcription of at least the genes known to be susceptible to the level of DNA supercoiling [10]. The *fis* evolved alleles may therefore adjust and enhance bacterial growth in the evolution experiments by combined global and local effects on the overall expression and structural pattern of the *E. coli* chromosome.

During the course of our work, we discovered a new gene under positive selection, repeatedly affected by mutations in many of the 12 populations. This gene is *dusB*, encoding one of the three dihydrouridine synthases in *E. coli* [63]. It is particularly intriguing that *dusB* is forming an operon structure with *fis*. Although the precise biochemical and physiological roles of *dusB* is currently unknown, our statistical sequence analyses of the topology-related genes suggested to us that both *dusB* and *fis* functions may be related. Hence, both genes show maximally uniform distribution, each harbouring single mutations in 5 independent populations, thereby leading to 10 populations with exactly one mutation in either *fis* or *dusB*. This distribution was shown to be statistically different from a random Poisson distribution ( $p=0.0039$ ), and a significant negative association is suggested between both genes. This confirmed a previous study which suggested a functional link between *dusB* and *fis* by

performing a phylogenetic analysis of Fis and DusB proteins [85]. Multiple sequence alignments between different  $\gamma$ -proteobacterial Fis proteins and  $\alpha$  and  $\gamma$ -proteobacterial NtrC proteins showed that Fis derives from the C-terminal DNA-binding domain of the NtrC protein from  $\alpha$ -proteobacteria. Moreover, DusB has strong homology with NifR3, encoded by a gene lying in the same operon as *ntrC* in several  $\alpha$ -proteobacteria. The sequence analyses suggest that the *dusB-fis* operon resulted from a rearrangement event that was acquired by horizontal transfer by a lineage ancestral to the  $\gamma$ -proteobacteria [85]. The two genes later diverged to their present functions inside  $\gamma$ -proteobacteria. These data were suggested to provide evidence for a functional interaction between both proteins to “carry out an as yet unknown function”. Hence, co-regulation within an operon structure is often considered as important to insure related functions. This previous study was performed in 1998 with only 20 to 30 proteins used for alignments. Based on our data, we decided to revisit and refine these sequence alignments by analysing the *fis* and *dusB* sequences in 323 bacteria with available genome sequences (Table 5). Four major trends can be observed. First, high occurrence of the *dusB-fis* operon was detected in both  $\gamma$ -proteobacteria, where it was found in 67% (53/79) of the available genomes, and  $\beta$ -proteobacteria where 95.6% (22/23) of the available genomes carry the operon. Second, only few occurrences were detected in both  $\alpha$ -proteobacteria (1/46) and  $\delta$ -proteobacteria (1/11). Third, inside  $\gamma$ - and  $\beta$ -proteobacterial groups, absence of the *dusB-fis* operon structure can be observed for 26 and 1 bacteria respectively, with only 7 cases of absence of *fis* (6 in  $\gamma$ -proteobacteria and 1 in  $\beta$ -proteobacteria). Fourth, the *fis* gene was never detected outside the proteobacteria. All these observations allow three major conclusions. First, while the phylogenetic analyses performed previously suggested that the *dusB-fis* operon was ancestral to the *Enterobacteriaceae* and *Pasteurellaceae* [85], our data suggest that it was actually ancestral to the split between  $\gamma$ - and  $\beta$ -proteobacteria. The *E. coli* O157:H7 was recently shown to have diverged from the B and K12 ancestor about 1.8 million years ago [86], while *E. coli* and *Salmonella enterica* species diverged for about 160 million years [87]. The split between  $\beta$ - and  $\gamma$ -proteobacteria is therefore much more ancient, giving enough evolutionary time for evolution of the DusB and Fis functions. Second, we proposed that the absence of the operon in  $\gamma$ -proteobacteria results from further protein function evolution. In some cases, the selective pressure for the related function between DusB and Fis may have been reduced, resulting to the loss of the operon structure. For example, endosymbiotic bacteria of the *Enterobacteriaceae* do not show the *dusB-fis* operon (Table 5), and this may be linked to their specialized life conditions which is well known to lead to gene loss events [88]. Hence, their narrow ecological niches have resulted in reduced selective

constraint on gene function. In some cases, like in the *Xanthomonadaceae*, a new gene, encoding a protein with unknown function, is present upstream *fis* in all representative bacteria of the group. This new operon may have resulted from an event similar to the one that generated the *dusB-fis* operon. We suggest that the protein encoded by this gene may have a function related to *fis* in the *Xanthomonadaceae*. Third, the occurrence of the *dusB-fis* operon in only rare cases in  $\alpha$  and  $\delta$ -proteobacteria may indicate either acquisition through horizontal transfer or loss of the operon structure in other members of that group.

The high level of parallel phenotypic DNA supercoiling increase was observed in all populations, except two. Artificially increasing DNA supercoiling in one evolved clone isolated from one of these populations, by introducing a fitness-enhancing and DNA supercoiling-increasing mutation from our model population, resulted in the exact similar phenotypes. This first suggests that we did not miss any mutation in DNA topology-related genes not sequenced in our analysis. Furthermore, it suggests that the absence of DNA supercoiling alterations in those populations do not result from historical contingency, i.e. from the substitution of previous mutations that may have either antagonize or neutralize later DNA topology-related mutations. More likely, it reflects the absence of substitution of mutations affecting DNA supercoiling. Hence, only 1% of the cells are transferred each day to fresh medium in the evolution experiment, and most beneficial mutations are therefore lost during this process.

Finally, a total of 10 different genes or operons have been shown to be targets of selection in this evolution experiment, based either on competition assays using isogenic strains: *rbs* [53], *spoT* [52], *topA* and *fis* [41], and *malT* [51], or on sequence analyses: *pykF*, *nadR*, *pbpA-rodA*, *hokB-sokB* [56], and *dusB* (this work). Among these loci, *spoT*, *fis*, *topA* and *dusB* affect two of the major regulatory networks in the cell: DNA supercoiling and the stringent response [1,89]. Both regulatory networks coordinate global gene expression to allow bacterial cells to cope with various stresses. Any alteration, positive or negative for bacterial fitness, is therefore expected to yield many pleiotropic effects. It is therefore highly unlikely that all the changes resulting from the *spoT*, *fis*, *topA* and *dusB* evolved alleles will confer increased fitness. Most probably, the average effects of all pleiotropic changes are beneficial, but these mutations may also yield potentially neutral and harmful side effects. One of the possibilities for further evolution of the cells would then be to reverse these side effects. It would therefore be extremely informative to analyze at the molecular levels all evolutionary steps along the adaptive pathway of a population. In our model population Ara-

1, 9 of these previously cited loci were shown to be altered and the precise order of mutation substitution has been documented (unpublished data). To relate the beneficial changes and the changes in gene expression to the successively substituted mutational alterations, the entire evolutionary history of the population can now be constructed by introducing successively, and in their precise substitution order, the different beneficial mutations into the ancestral background. Such an analysis would also give invaluable insights into the interconnected dynamics between regulatory networks.

## **Acknowledgments**

The authors wish to thank Mohamed-Ali Hakimi (Laboratoire Adaptation et Pathogénie des Micro-organismes) for help in Fis purification, Akira Ishihama (Nippon Institute for Biological Sciences, Japan), Michael Cashel (National Institute of Health, USA) and Tamas Gaal and Richard Gourse (University of Wisconsin Madison, USA) for the anti-Fis, anti-RpoA antibodies and plasmid pKG13. This work was supported by grants from the french Centre National de la Recherche Scientifique, Université Joseph Fourier and the National Science Foundation. EC acknowledges the CNRS for a BDI fellowship.

## Materials and Methods

**The long-term evolution experiment.** The long-term evolution experiment is described in details elsewhere [44]. In brief, 12 populations were founded using an *E. coli* B clone as the ancestor [45,90] and then propagated by serial daily transfers at 37°C for 20,000 generations (3,000 days) in the same Davis minimal medium supplemented with glucose at 25 mg/mL (DM25) [42]. Each day, after 24 h incubation, cells were diluted 100-fold into fresh medium, the population size fluctuating daily between about  $5 \cdot 10^6$  and  $5 \cdot 10^8$  cells. The source ancestral strain, REL606, cannot grow on arabinose (Ara-) and was used to initiate 6 populations, called Ara-1 to Ara-6, whereas the other 6 populations called Ara+1 to Ara+6, were started with a spontaneous Ara+ revertant of the source strain, called REL607. The Ara phenotype is used as a selectively neutral marker in competition experiments [42].

**Strains, plasmids and culture conditions.** All single evolved clones used in this study have been isolated either at generation 20,000 (for sequence analyses) or at generation 2,000 (for gene allele exchange experiments) from each of the 12 populations [43], and have been stored as glycerol suspensions at -80°C. Isogenic ancestral strains, except for *topA* and *fis*, were previously constructed [41] using the suicide-plasmid pKO3 [91] and were used in this study : 606<sup>*topA-1*</sup> is the ancestral REL606 strain bearing the beneficial evolved *topA* allele identified in population Ara-1 [41]; 606<sup>*fis-1*</sup> and 606 <sup>$\Delta$ *fis*</sup> are ancestral REL606 strains bearing respectively the beneficial evolved *fis* allele previously identified in population Ara-1 and an in-frame *fis* deletion allele. Strains TOP10 (Invitrogen, San Diego) and JM109 [92] were used for DNA cloning experiments.

We used plasmids pUC18 [92] for DNA supercoiling measurements, pKO3 [91] for allele replacements, and pCRII-Topo (Invitrogen, San Diego) for cloning experiments. Two plasmids, pKG13 and pKK223-3 [69] were used to address the DNA binding activity of ancestral and evolved Fis proteins.

All experiments were performed by growing strains either in the same DM25 medium that was used in the evolution experiment [42], or in rich Luria Broth (LB) medium [93]. Ampicillin (100  $\mu$ g/mL), chloramphenicol (30  $\mu$ g/mL), or kanamycin (50  $\mu$ g/mL) were added as needed.

**Sequencing.** Nine loci, encoding topoisomerases or histone-like proteins, were sequenced in one evolved clone isolated at 20,000 generations from each of the 12 populations. In four cases, the genes are located within operons and the whole operons were then sequenced. A total of 17 genes were sequenced in each evolved clone. Table 6 lists the genes sequenced in this study, as well as function of the gene products and number of base pairs sequenced. All 17 genes (23,102 bp) were sequenced in all 14 clones (12 evolved clones and both ancestors, totaling 323,428 bp). Genomic DNA was prepared from each isolate using a boiling technique. Cells from a single colony were transferred into 100  $\mu$ L of sterile water, boiled for 10 min and stored at 4°C for subsequent experiments. Typically 30-60 ng of nucleic acid were employed for PCR amplifications. Primers were designed using the Primer3 program (<http://www.genome.wi.mit.edu>). Several primers pairs were used to amplify and sequence the entire genes, each pair producing a PCR product between 500 and 750 bp in length. All primers pairs designed for the 17 genes listed in Table 6 are available upon request to the authors. All reactions were performed using Qiagen Taq polymerase in a Gene Amp 9700 thermocycler (PE biosystems, Foster City, USA), using the following program: 95°C for 5 min, 25 cycles at 95, 57, and 72°C for 45s each, and finally 72°C for 7min. The resulting PCR products were visualized on 1%-agarose gels containing 5 ng/ $\mu$ L ethidium bromide to determine the correct size and concentration. Amplified products were purified using a QIAquick 8 PCR purification kit (Qiagen, Valencia, USA) following the manufacturer's instructions and sequencing reactions prepared for analysis on an ABI 377 automated sequencer using the Big Dye Terminator Cycle Sequencing Kit (ABI Applied Biosystems, Foster City, USA). All PCR products were sequenced in both forward and reverse directions. Sequencing reactions were purified using a 96-well filtration block (Edge Biosystems, Gaithersburg, USA) following the manufacturer's instructions. Sequencing results were analyzed using Sequencher 3.1.1 (Gene Codes Corporation, Ann Arbor, USA). Polymorphic sites were always confirmed by resequencing relevant regions twice with independent DNA extractions and reanalyzing the sequence chromatogram using Sequencher 3.1.1. Polymorphism was analyzed using the Molecular Evolutionary Analysis Package MEA (Etsuko Moriyama, Yale University version 6/22/00), based on the algorithm of Nei and Gojobori [94].

**Construction of isogenic strains.** Evolved *fis* alleles were introduced into the chromosome of the ancestral strain REL606 by homologous recombination using suicide plasmid pKO3 [91], as described previously [41]. Briefly, the pKO3-derived plasmids containing the cloned

*fis* evolved alleles were electrotransformed into the ancestor clone. Incubation at high temperature on chloramphenicol-containing LB plates selects for integration of each evolved allele into the ancestral *fis* gene. Chloramphenicol-resistant cells were then plated on sucrose-containing LB plates selecting for plasmid loss, as the plasmid carries *sacB*, which renders cells sensitive to killing by sucrose. Sucrose-resistant and chloramphenicol-sensitive plasmid-free clones were screened for the presence of the evolved *fis* allele by a PCR-restriction fragment length polymorphism (PCR-RFLP) approach using *AluI* and *DdeI* (Euromedex, France) to distinguish between ancestral and evolved alleles in the case of the Ara+1 and Ara-3 alleles, respectively. All constructed strains were also deconstructed such that the evolved alleles were replaced by their ancestral counterparts, to confirm the absence of secondary mutations during the construction process. These experiments allowed us to construct isogenic ancestral strains, except for the *fis* alleles: 606<sup>*fis*+1</sup> and 606<sup>*fis*-3</sup> are the ancestor REL606-derived strains bearing the evolved Ara+1 and Ara-3 *fis* alleles, respectively. We could however never obtain the ancestor bearing the evolved Ara-4 *fis* allele, despite numerous attempts.

The same suicide plasmid-based strategy was used to recombine the evolved *topA* allele that was identified in population Ara-1 [41] into the chromosome of an evolved clone, called 1158C, isolated from Ara+1 population at 2,000 generations. The 1158C clone contains the ancestral *topA* allele, which was therefore replaced by the evolved Ara-1 *topA* allele, leading to strain 1158C<sup>*topA*-1</sup>. Handling of the Ara-1 *topA* allele for construction experiments is described in [41].

**Fitness assays.** The relative fitness of two strains was estimated by directly competing them, as described elsewhere [51]. Briefly, isogenic strains were competed against the ancestor carrying the opposite arabinose marker. Each pairwise competition was replicated 6-fold. Competitions were performed in DM25, the same medium used for the evolution experiment, for 1 day. Fitness assays thus encompass all phases of population growth that occurred in the evolution experiment (lag, exponential and stationary phases). Immediately before each assay, both competitors were separately acclimated to the same conditions and then transferred together from stationary phase cultures into fresh medium (with 1:200 dilution each and thus 1:100 combined). Samples of the competition assays were taken immediately after mixing at day 0 and again after 1 day of competition to measure the abundance of both competitors. At time of sampling, all cells are in stationary phase (24 h after the previous transfer). Knowing the initial and final cell counts and the dilution factors allows one to calculate the realized

(net) population growth of each competitor. Fitness was then calculated as the ratio of the realized growth rates of the two strains during their direct competition, and *t*-tests were performed to evaluate whether the measured fitness values differed significantly from the null hypothetical value of one.

**DNA superhelicity measurements.** Topological changes were measured by the relative abundance of different topoisomers of a reporter plasmid, as described previously [41]. Briefly, the pUC18 plasmid was introduced by electrotransformation in the strains of interest. After growing cells in LB-ampicillin media to an  $OD_{600nm} = 2$ , plasmids were extracted (Qiagen Qiaprep Spin Miniprep kit), quantified and 200 ng were loaded onto 1%-agarose gels containing 1.5  $\mu$ g/mL chloroquine. The gels were run overnight at 4°C at 3 V/cm. The gels were washed in water to remove chloroquine and visualized under UV light after ethidium bromide staining. Under these conditions, more highly supercoiled topoisomers migrate faster [5]. Densitometric analyses were performed using the Image Gauge software (Fujifilm) to obtain the average topoisomer value for each clone from the overall relative abundance of all plasmid topoisomers. The level of DNA supercoiling was estimated by mean sigma values  $\sigma = \Delta Lk / Lk_0$ , where  $\Delta Lk$  is the difference in the linking number (i.e. difference in average topoisomer value) between two clones and  $Lk_0$  for pUC18 is 2,686 bp / 10.45 bp / turn = 257. Values for  $\sigma$  are usually expressed relative to the mid-point topoisomer of plasmid that was completely relaxed by using calf-thymus topoisomerase I and assigned  $\sigma = 0$ . Comparing the topoisomer distribution of plasmid DNA from the ancestor with the completely relaxed plasmid control yielded an estimate of  $\sigma = -0.066$ . The  $\sigma$  value for each evolved clone and isogenic construct were compared to that of the ancestor, allowing a precise quantitative comparison of the distribution of plasmid topoisomers between each clone and the ancestor [41].

**Electrophoresis and immunoblot analysis of proteins.** Triplicate cultures of the ancestor strains bearing the different *fis* alleles and evolved clones sampled at 20,000 generations of evolution from the Ara-1, Ara-3, Ara-4 and Ara+1 populations were grown in LB medium at 37°C, and aliquots were sampled during exponential phase after 1.5 h of incubation. After centrifugation, cells were resuspended in lysis buffer (70 mM Tris pH7.4, 1 mM EDTA, 1 mM DTT, 10% glycerol) before sonication. Lysates were centrifuged at 10,000 g for 30 min at 4°C, and total protein concentrations were determined using the Bradford protein assay kit

(Bio-Rad) and bovine serum albumin as a standard. Equal amounts of proteins were loaded on 16% SDS-polyacrylamide gels, separated by electrophoresis and electrotransferred (Bio-rad) on nitrocellulose membranes (Amersham Pharmacia). Prestained protein standards (Euromedex, France) were used for molecular weight estimation. Immunoblot analyses were performed with antibodies raised against Fis (courtesy of A. Ishihama, Nippon Institute for Biological Sciences, Japan) and RpoA (courtesy of M. Cashel, National Institute of Health, USA) and developed using the ECL plus Western Blotting Detection System kit (Amersham Pharmacia).

**Transcriptional fusions and  $\beta$ -galactosidase assays.** Ancestral and evolved Ara+1 *fis* promoter alleles were PCR amplified from positions -241 to +84 relative to the +1 *dusB* transcription start site [64] using primers (*fisEcoRI* 5'-cacgtaattgcccgaattctacgcaaatt-3' and *fisBamHI* 5'-ccatgggcgctggggatccaggcgattctg-3'). The primers were designed to contain *EcoRI* and *BamHI* restriction sites (underlined in the sequence). Both 325 bp-products were cloned into the pRS551 plasmid, upstream of the *lacZ* reporter gene encoding  $\beta$ -galactosidase [68]. This generates plasmid transcriptional fusions with each of the ancestral and evolved promoters directing transcription of *lacZ*. These transcriptional fusions were transferred onto  $\lambda$  bacteriophage DNA by recombination, using  $\lambda$ RS45 as described [68]. Both  $\lambda$ RS45 derivatives were introduced into the  $\lambda$  *att* site of the chromosome of REL606. For both fusions, the sequences of the *fis* promoters were verified and single lysogeny of fusion-carrying  $\lambda$  phages was checked by a previously described PCR method [95].

For  $\beta$ -galactosidase assays, cultures were performed in kanamycin-containing LB media, and samples were taken at different times during growth. The  $\beta$ -galactosidase specific activities were assayed by use of *o*-nitrophenyl  $\beta$ -D-galactopyranoside (ONPG) as a substrate and were calculated in nanokatals per milligram (nanomoles of ONPG transformed per second per milligram of cellular proteins, nkat/mg). The amount of total cellular proteins was quantified using the Bradford Protein Assay kit (Bio-Rad). All reported activity values are the average of three independent experiments.

**Fis purifications.** Two of the mutations identified in *fis* in the evolving populations affect its coding sequence. We purified both evolved Fis proteins together with the ancestral protein, according to a previously described protocol [70]. The three *fis* alleles were cloned in the pET41 expression vector and overexpressed in BL21(DE3) cells (Novagen). Plasmid-carrying cells were grown in 250 mL LB containing kanamycin at 37°C up to an OD<sub>600nm</sub> of 0.6 to 0.8,

and Fis expression was induced with 0.8mM of isopropyl-beta-D-thiogalactopyranoside (IPTG) for 4h. After centrifugation, pellets were washed and resuspended in extraction buffer (70 mM Tris pH7.4, 1 M NaCl, 10% glycerol). After lysis with a French Press (1000 psi, Thermo Spectroni, Rochester, USA), the lysates were centrifugated 15 min at 5,000 g and sonicated for 1 min on ice. DNA was removed by the addition of 0.35% of polyethylenimine (Sigma) and the precipitate was eliminated by centrifugation at 100,000 g for 1 h. Remaining PEI was eliminated by incubation in 0.2 volume activated phosphocellulose (P11, Whatman). The lysates were then dialyzed into FB (20 mM Tris pH7.5, 1 mM EDTA, 300 mM NaCl, 10% glycerol) and passed through an SP sepharose High Trap column (Amersham) equilibrated with the same buffer. Fis was eluted with a linear gradient from 0.3 to 0.7 M NaCl in FB. Fis concentrations were determined with the Bradford Protein Assay reagent using bovine serum albumin as a standard.

**DNA labelling.** Three promoters known to contain Fis-binding sites were analyzed : i) a fragment with the *fis* promoter extending from positions -215 to +58 was PCR-amplified using primers ODS234 (5'-cgcaaattgcgtgccaaaatt-3') and ODS435 (5'-tgagctgatattgctcgtatg-3'); ii) a fragment with the *rrnB* P1 promoter, extending from position -210 to +56 [96] was PCR-amplified using primers ODS436 (5'-aacatgaagccccggatgca-3') and ODS437 (5'-tctcaggagaaccccgct-3'); iii) a fragment with the *proP* promoters, extending from positions -114 to +197 [97] was PCR-amplified using primers ODS438 (5'-ctcaggaaatcttctctgt-3') and ODS439 (5'-tttcttttcagcatagcttt-3'). All positions are given with respect to corresponding +1 transcription start sites. In each case, prior to the PCR reactions, the forward primer was labelled with  $\gamma$ -<sup>32</sup>P-dATP (6000Ci/mmol, ICN) using T4 polynucleotide kinase according to the manufacturer's recommendations (Euromedex, France). Labelled fragments were then purified by a column purification procedure using the PCR Purification kit (Qiagen).

**Electromobility shift assays.** Binding reactions were performed by incubating  $\leq 5$  fmol ( $\leq 0.6$  nM) of the labelled DNA fragments with various amounts (0 to 40 nM) of purified Fis protein in a 15  $\mu$ L solution containing 20 mM Tris pH7.5, 1 mM EDTA pH8, 80 mM NaCl, 4% glycerol, 0.5 mM DTT, 0.5 mg/mL BSA and 10  $\mu$ g/mL of sonicated herring sperm DNA. After 20 min of incubation at 37°C, 5  $\mu$ L of loading buffer (0.08% bromophenol blue in 50% sucrose) were added and samples were electrophoresed on 5% polyacrylamide gels (19:1) in 0.5 X Tris-Borate EDTA buffer. The gels were then dried, exposed to Fujifilm Imaging Plates and revealed on a FLA8000 scanner (Fujifilm). The percentage of bound DNA was quantified

with the Image Gauge software (Fujifilm). Three types of control were performed. First, a 5- to 1000-fold excess of cold DNA fragments were added to the binding mixtures. The DNA fragments consist in either a 30 bp-oligonucleotide containing a consensus Fis binding site or the same oligonucleotide but with a mutated Fis site. A supershift assay was also performed by adding the antibody raised against Fis to the binding reactions.

## Figure legend

### Figure 1. The *fis* Mutations Substituted in the Evolving Populations

(A) The structure of the *dusB-fis* operon is shown. The -35 and -10 elements of the promoter [64] are indicated by black boxes, and both coding sequences by white rectangles. The direction and initiation site of transcription are shown by a broken arrow. The 10 mutations substituted in the evolution experiment are indicated both by their localization (arrows) and by the name of the evolving population in which they arose. (B) Immunodetection of the Fis protein in a set of isogenic ancestral strains and in evolved clones. Proteins were extracted in early exponential phase after 1.5 h of growth in LB medium for each of 9 strains: ancestral strains with the ancestral, Ara-1, Ara+1, Ara-3 or in-frame deletion *fis* alleles; a 20,000-generation evolved clone isolated from each of the Ara-1, Ara+1, Ara-3 and Ara-4 populations. 606 is the REL606 ancestor strain. Each *fis* evolved allele (-1, +1, -3) gave similar results in both ancestral and evolved backgrounds and a typical result is given here. -4 is the 20,000-generation evolved clone isolated from population Ara-4. The strain containing an in-frame *fis* deletion ( $\Delta$ ) was used as a control. The anti-RpoA is the internal control antibody to check for equivalent protein amounts loaded in each well. The results shown here represent typical data from one of three experiments. (C) Single-copy *lacZ* fusions were constructed in the chromosome of the REL606 ancestor with either the ancestral *fis* promoter ( $606^{fisP}$ ) or the evolved Ara+1 *fis* promoter ( $606^{fis+1P}$ ). Cultures of both strains were grown in kanamycin-containing LB medium and  $\beta$ -galactosidase specific activities, expressed as nkat/mg of total protein, were measured from samples isolated during exponential phase ( $OD_{600nm}=0.2$ , 1 h of growth), transition to stationary phase ( $OD_{600nm}=1$ , 2.5 h of growth), and stationary phase ( $OD_{600nm}=4$  to 5, 6 h of growth). The data are based on three independent experiments.

### Figure 2. DNA Binding of Ancestral and Evolved Fis Proteins

(A) Binding of the Fis ancestral (Fis Anc), Ara-3 evolved (Fis-3) and Ara-4 evolved (Fis-4) proteins to the *fisP* (top), *rrnBP1* (middle), and *proP* P1 and P2 (bottom) promoter regions *in vitro*. The interaction of the Fis proteins with the transcription regulatory regions was assessed by electrophoretic mobility shift assays. The regulatory sequences were amplified by PCR, with radiolabelled primers, and incubated with increasing concentrations of purified Fis proteins (given in nM) and the complexes were electrophoresed. Free (F) and Fis-bound (B) DNA bands are indicated. Note the different amounts of protein used for the different *fis* alleles. (B) Quantitation of bound DNA. The intensities of the free and bound DNA bands were measured using Image Gauge software (Fujifilm), and the percentage of bound DNA is plotted against Fis concentration (in nM). All experiments were repeated three times and typical curves for each of the three Fis proteins (Fis Anc, Fis-3, Fis-4) using the *fisP* promoter are shown. (C) Dissociation constants (in nM), calculated from the curves in (B) as the amount of Fis yielding half retardation in the electromobility shift assays in (A), are given for each of the three proteins with each of the three promoters.

### Figure 3. DNA Supercoiling and Fitness Effects of a *topA* Evolved Allele in Different Genetic Contexts

(A) DNA supercoiling levels were measured using a reporter plasmid in five different genetic backgrounds. 606 is the REL606 ancestor of the evolution experiment (bearing the *topA* ancestral allele); 606<sup>*topA-1*</sup> is REL606 with the *topA* evolved allele found in population Ara-1 [41]; 1164 is a 2,000-generation clone from population Ara-1 with the *topA* evolved allele; 1158C is a 2,000-generation clone isolated from population Ara+1 carrying the ancestral allele; 1158C<sup>*topA-1*</sup> is the same evolved clone from Ara+1 population but in which we introduced the *topA* evolved allele from population Ara-1. (B) Fitness effects of a topology-related gene (*topA*) evolved allele, moved into the ancestral or evolved Ara+1 genetic backgrounds. Competitions were performed in the same medium used in the long-term evolution experiment. Errors bars are 95% confidence intervals based on 6 replicate competition assays for each genotype. The genotypes are the same as in (A), with 607 being the Ara+ ancestor counterpart of REL606. The competition between 606 and 607 leads to a fitness of 1, showing the absence of discernible effect of the arabinose marker on fitness in these competitions.

**Table 1.** Mutations Found in 12 Evolving Populations after 20,000 Generations of Evolution among the Topology-Related Genes

Population <sup>a</sup>	Gene	Nucleotide Position <sup>b</sup>	Mutation	Amino Acid Change
Ara-1	<i>fis</i>	-4	A to C	NA <sup>c</sup> (RBS) <sup>d</sup>
	<i>topA</i>	+97	C to T	H33 to Y
Ara-2 (M)	<i>topB</i>	+628	C to T	D210 to N
	<i>ydjA</i> ( <i>topB</i> locus)	+270	G to A	no change
	<i>ydjA</i> ( <i>topB</i> locus)	+15	T to C	no change
	<i>topA</i>	+1012	T to C	Y338 to H
	<i>topA</i>	+1526	A to G	Y509 to C
Ara-3	<i>fis</i>	+152	A to C	Y51 to S
	<i>topA</i>	+2315	C to A	T792 to K
Ara-4 (M)	<i>fis</i>	+177	+ A	CTD change
Ara-5	<i>topA</i>	-112	G to T	NA <sup>c</sup> (promoter) <sup>e</sup>
	<i>topA</i>	+599	C to A	A200 to E
Ara-6	<i>dusB</i> ( <i>fis</i> locus)	+899	G to A	A302 to T
Ara+1	<i>fis</i>	-63 <sup>f</sup>	T to G	NA <sup>c</sup> (promoter) <sup>e</sup>
Ara+2	<i>fis</i>	+329	C to A	NA <sup>c</sup> (downstream)
Ara+3 (M)	<i>dusB</i> ( <i>fis</i> locus)	+927	G to A	no change
	<i>parC</i>	+460	G to A	P154 to S
Ara+4	<i>dusB</i> ( <i>fis</i> locus)	+927	+ 9 bp	+ ALE after E314
	<i>dps</i>	+143	IS1 insertion	NA <sup>c</sup>
Ara+5	<i>dusB</i> ( <i>fis</i> locus)	+934	G to A	A315 to T
Ara+6 (M)	<i>dusB</i> ( <i>fis</i> locus)	+904	T to G	F300 to C
	<i>selD</i> ( <i>topB</i> locus)	+960	T to G	no change
	<i>yqiB</i> ( <i>parE</i> locus)	+134	T to G	Y45 to S
	<i>yqiA</i> ( <i>parE</i> locus)	+251	A to C	L84 to R
	<i>hns</i>	-89	A to C	NA <sup>c</sup> (promoter) <sup>e</sup>
	<i>gyrA</i>	+1925	T to G	Q657 to H
	<i>gyrB</i>	+1971	T to G	E662 to A

<sup>a</sup> Populations showing a mutator phenotype [47,48] are indicated by (M).

<sup>b</sup> Nucleotide positions are given relative to the first position of the translational start codon of the corresponding gene.

<sup>c</sup> NA: Not Applicable.

<sup>d</sup> Mutation within the Ribosome Binding Site.

<sup>e</sup> Mutation within the promoter region.

<sup>f</sup> The nucleotide position is given relative to the first position of the translational start codon of *dusB*.

**Table 2.** Number of Mutations in Random and Topology-Related Genes in 12 *E. coli* Populations that Evolved for 20,000 Generations

Population	Random Genes (18,374 bp total)		Topology-Related Genes (23,102 bp total)	
	No. of mutations	Rate per 1000 bp	No. of mutations	Rate per 1000 bp
Ara-1	0	0.000	2	0.087
Ara-2 (M)	0	0.000	5	0.216
Ara-3	0	0.000	2	0.087
Ara-4 (M)	3	0.163	1	0.043
Ara-5	0	0.000	2	0.087
Ara-6	0	0.000	1	0.043
Ara+1	0	0.000	1	0.043
Ara+2	0	0.000	1	0.043
Ara+3 (M)	1	0.054	2	0.087
Ara+4	0	0.000	2	0.087
Ara+5	0	0.000	1	0.043
Ara+6 (M)	2	0.109	7	0.303

Topology-related genes include regulatory regions as well as other ORFs found in the same loci. Populations Ara-2, Ara-4, Ara+3, and Ara+6 became mutators (M), while the others kept the low ancestral mutation rate.

**Table 3.** Candidate Topology and Random Genes Differ in Relative Abundance of Synonymous and Non-Synonymous Mutations

Genes	Synonymous	Non-Synonymous
<i>topA, fis, dusB</i>	1	11
random genes	3	3

**Table 4.** Candidate Topology and Random Genes Differ in Relative Abundance of Point Mutations in Mutator and Non-Mutator Populations.

Genes	Mutator	Non-Mutator
<i>topA, fis, dusB</i>	4	10
random genes	6	0

**Table 5 : The *dusB-fis* Operon in Genome-Sequenced Bacteria**

Bacteria		Total	<i>dusB</i> + <i>fis</i>	<i>fis</i>	<i>dusB</i>	No <i>dusB</i> , no <i>fis</i>	
γ-proteobacteria	Alteromonadales	6	6	0	0	0	
	Chromatiales	1	0	1 <sup>a</sup>	0	0	
	Enterobacteriales	endosymbiont	7	0	3 <sup>b</sup>	0	4
		others	24	24	0	0	0
	Legionellales	<i>Coxiellaceae</i>	1	0	1 <sup>c</sup>	0	0
		<i>Legionellaceae</i>	3	0	3 <sup>cd</sup>	0	0
	Methylococcales, Oceanospirillales, Pasteurellales, Vibrionales	14	14	0	0	0	
	Pseudomonadales	<i>Moraxellaceae</i>	3	0	3 <sup>ce</sup>	0	0
		<i>Pseudomonadaceae</i>	8	8	0	0	0
	Thiotrichales	<i>Francisellaceae</i>	2	0	0	2	0
		<i>Piscirickettsiaceae</i>	1	1	0	0	0
	Xanthomonadales	8	0	8 <sup>f</sup>	0	0	
	Unclassified <sup>g</sup>	1	0	0	0	1	
	Total γ-proteobacteria		79	53	19	2	5
α-proteobacteria		46	1 <sup>h</sup>	0	45 <sup>i</sup>	0	
β-proteobacteria		23	22	0	1	0	
δ-proteobacteria		11	1	1 <sup>j</sup>	9	0	
ε-proteobacteria		8	0	0	8	0	
Others		156	0	0	129	27	

The number of each genome within each bacterial group where both *fis* and *dusB*, only *fis*, only *dusB* and neither gene were detected, is shown.

<sup>a</sup> L11 methyltransferase-encoding gene upstream *fis*.

<sup>b</sup> 2 with RluCD (pseudo-uridine synthase for 23S rRNA)-encoding gene upstream *fis*.

<sup>c</sup> *dusB* far away.

<sup>d</sup> 3 *fis* genes.

<sup>e</sup> 2 with L11 methyltransferase-encoding gene upstream *fis* and one with *Xanthomonadaceae*-like protein-encoding gene.

<sup>f</sup> 7 with *dusB* far away, 1 without *dusB*. All with same protein-encoding gene upstream *fis*.

<sup>g</sup> *Baumannia*.

<sup>h</sup> *Zymomonas mobilis* ZM4.

<sup>i</sup> Presence of *nifR3*.

<sup>j</sup> *dusB* far away, protein NrdG (organic radical activating enzyme)-encoding gene upstream *fis*.

---

**Table 6.** The 17 Genes Sequenced in this Study (23,102bp in total)

---

<b>Gene (upstream ORF)</b>	<b>Number of bp sequenced</b>	<b>Function</b>	<b>Reference</b>
<i>dps</i>	802	Histone-like protein	[98]
<i>fis (dusB)</i>	1,695	Histone-like protein	[30]
<i>gyrA</i>	2,774	Gyrase subunit A	[16]
<i>gyrB</i>	2,673	Gyrase subunit B	[16]
<i>hns</i>	1,160	Histone-like protein	[99]
<i>parC (yqiS)</i>	2,599	Topoisomerase IV subunit A	[100]
<i>parE (yqiA, icc, yqiB, yqiE)</i>	4,652	Topoisomerase IV subunit B	[100]
<i>topA</i>	2,905	Topoisomerase I	[18]
<i>topB (selD, ydjA)</i>	3,842	Topoisomerase III	[16]

---

## References

1. Travers A, Muskhelishvili G (2005) DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nat Rev Microbiol* 3: 157-169.
2. Balke VL, Gralla JD (1987) Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *J Bacteriol* 169: 4499-4506.
3. Reyes-Dominguez Y, Contreras-Ferrat G, Ramirez-Santos J, Membrillo-Hernandez J, Gomez-Eichelmann MC (2003) Plasmid DNA supercoiling and gyrase activity in *Escherichia coli* wild-type and *rpoS* stationary-phase cells. *J Bacteriol* 185: 1097-1100.
4. Weinstein-Fischer D, Elgrably-Weiss M, Altuvia S (2000) *Escherichia coli* response to hydrogen peroxide: a role for DNA supercoiling, topoisomerase I and Fis. *Mol Microbiol* 35: 1413-1420.
5. Higgins CF, Dorman CJ, Stirling DA, Waddell L, Booth IR, May G, Bremer E (1988) A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *Salmonella typhimurium* and *Escherichia coli*. *Cell* 52: 569-584.
6. Goldstein E, Drlica K (1984) Regulation of bacterial DNA supercoiling: plasmid linking numbers vary with growth temperature. *Proc Natl Acad Sci U S A* 81: 4046-4050.
7. Karem K, Foster JW (1993) The influence of DNA topology on the environmental regulation of a pH-regulated locus in *Salmonella typhimurium*. *Mol Microbiol* 10: 75-86.
8. Bhriain NN, Dorman CJ, Higgins CF (1989) An overlap between osmotic and anaerobic stress responses: a potential role for DNA supercoiling in the coordinate regulation of gene expression. *Mol Microbiol* 3: 933-942.
9. O'Croinin T, Carroll RK, Kelly A, Dorman CJ (2006) Roles for DNA supercoiling and the Fis protein in modulating expression of virulence genes during intracellular growth of *Salmonella enterica* serovar *Typhimurium*. *Mol Microbiol* 62: 869-882.
10. Peter BJ, Arsuaga J, Breier AM, Khodursky AB, Brown PO, Cozzarelli NR (2004) Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biol* 5: R87.
11. Dai X, Rothman-Denes LB (1999) DNA structure and transcription. *Curr Opin Microbiol* 2: 126-130.
12. Postow L, Hardy CD, Arsuaga J, Cozzarelli NR (2004) Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev* 18: 1766-1779.
13. Deng S, Stein RA, Higgins NP (2005) Organization of supercoil domains and their reorganization by transcription. *Mol Microbiol* 57: 1511-1521.
14. Valens M, Penaud S, Rossignol M, Cornet F, Boccard F (2004) Macrodome organization of the *Escherichia coli* chromosome. *EMBO J* 23: 4330-4341.
15. Jeong KS, Ahn J, Khodursky AB (2004) Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*. *Genome Biol* 5: R86.
16. Champoux JJ (2001) DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* 70: 369-413.
17. Dorman CJ, Deighan P (2003) Regulation of gene expression by histone-like proteins in bacteria. *Curr Opin Genet Dev* 13: 179-184.
18. Wang JC (1971) Interaction between DNA and an *Escherichia coli* protein omega. *J Mol Biol* 55: 523-533.
19. Zechiedrich EL, Khodursky AB, Bachellier S, Schneider R, Chen D, Lilley DM, Cozzarelli NR (2000) Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. *J Biol Chem* 275: 8103-8113.
20. Gellert M, Mizuuchi K, O'Dea MH, Nash HA (1976) DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc Natl Acad Sci U S A* 73: 3872-3876.
21. Luijsterburg MS, Noom MC, Wuite GJ, Dame RT (2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *J Struct Biol* 156: 262-272.
22. Hardy CD, Cozzarelli NR (2005) A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure. *Mol Microbiol* 57: 1636-1652.
23. Kelly A, Goldberg MD, Carroll RK, Danino V, Hinton JC, Dorman CJ (2004) A global role for Fis in the transcriptional control of metabolism and type III secretion in *Salmonella enterica* serovar *Typhimurium*. *Microbiology* 150: 2037-2053.
24. Johansson J, Balsalobre C, Wang SY, Urbonaviciene J, Jin DJ, Sonden B, Uhlin BE (2000) Nucleoid proteins stimulate stringently controlled bacterial promoters: a link between the cAMP-CRP and the (p)ppGpp regulons in *Escherichia coli*. *Cell* 102: 475-485.

25. Hommais F, Krin E, Laurent-Winter C, Soutourina O, Malpertuy A, Le Caer JP, Danchin A, Bertin P (2001) Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol Microbiol* 40: 20-36.
26. Kar S, Edgar R, Adhya S (2005) Nucleoid remodeling by an altered HU protein: reorganization of the transcription program. *Proc Natl Acad Sci U S A* 102: 16397-16402.
27. Arfin SM, Long AD, Ito ET, Tollerli L, Riehle MM, Paegle ES, Hatfield GW (2000) Global gene expression profiling in *Escherichia coli* K12. The effects of integration host factor. *J Biol Chem* 275: 29672-29684.
28. Muskhelishvili G, Travers A (2003) Transcription factor as a topological homeostat. *Front Biosci* 8: 279-285.
29. Schneider R, Lurz R, Luder G, Tolksdorf C, Travers A, Muskhelishvili G (2001) An architectural role of the *Escherichia coli* chromatin protein FIS in organising DNA. *Nucleic Acids Res* 29: 5107-5114.
30. Finkel SE, Johnson RC (1992) The Fis protein: it's not just for DNA inversion anymore. *Mol Microbiol* 6: 3257-3265.
31. Schneider R, Travers A, Kutateladze T, Muskhelishvili G (1999) A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol Microbiol* 34: 953-964.
32. Schneider R, Travers A, Muskhelishvili G (2000) The expression of the *Escherichia coli* *fis* gene is strongly dependent on the superhelical density of DNA. *Mol Microbiol* 38: 167-175.
33. Talukder AA, Iwata A, Nishimura A, Ueda S, Ishihama A (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J Bacteriol* 181: 6361-6370.
34. Hengge-Aronis R (2002) Stationary phase gene regulation: what makes an *Escherichia coli* promoter sigmaS-selective? *Curr Opin Microbiol* 5: 591-595.
35. Bordes P, Conter A, Morales V, Bouvier J, Kolb A, Gutierrez C (2003) DNA supercoiling contributes to disconnect sigmaS accumulation from sigmaS-dependent transcription in *Escherichia coli*. *Mol Microbiol* 48: 561-571.
36. Hirsch M, Elliott T (2005) Fis regulates transcriptional induction of RpoS in *Salmonella enterica*. *J Bacteriol* 187: 1568-1580.
37. Feldman-Cohen LS, Shao Y, Meinhold D, Miller C, Colon W, Osuna R (2006) Common and variable contributions of Fis residues to high-affinity binding at different DNA sequences. *J Bacteriol* 188: 2081-2095.
38. Schneider DA, Ross W, Gourse RL (2003) Control of rRNA expression in *Escherichia coli*. *Curr Opin Microbiol* 6: 151-156.
39. Gonzalez-Gil G, Bringmann P, Kahmann R (1996) FIS is a regulator of metabolism in *Escherichia coli*. *Mol Microbiol* 22: 21-29.
40. Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* 4: 457-469.
41. Crozat E, Philippe N, Lenski RE, Geiselman J, Schneider D (2005) Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169: 523-532.
42. Lenski RE, Rose M.R., Simpson S.C., Tadler S.C. (1991) Longterm experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *The American naturalist* 138: 1315-1341.
43. Cooper VS, Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407: 736-739.
44. Lenski RE (2004) Phenotypic and Genomic Evolution during a 20,000-Generation Experiment with the Bacterium *Escherichia coli*. *Plant Breed Rev* 24: 225-265.
45. Schneider D, Duperchy E, Depeyrot J, Coursange E, Lenski R, Blot M (2002) Genomic comparisons among *Escherichia coli* strains B, K-12, and O157:H7 using IS elements as molecular markers. *BMC Microbiol* 2: 18.
46. Lenski RE (1991) Quantifying fitness and gene stability in microorganisms. *Biotechnology* 15: 173-192.
47. Sniegowski PD, Gerrish PJ, Lenski RE (1997) Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387: 703-705.
48. Shaver AC, Dombrowski PG, Sweeney JY, Treis T, Zappala RM, Sniegowski PD (2002) Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. *Genetics* 162: 557-566.
49. Lenski RE, Mongold JA, Sniegowski PD, Travisano M, Vasi F, Gerrish PJ, Schmidt TM (1998) Evolution of competitive fitness in experimental populations of *Escherichia coli*: what makes one genotype a better competitor than another? *Antonie Van Leeuwenhoek* 73: 35-47.
50. Vasi F, Travisano M, Lenski RE (1994) Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. *Am Nat* 144: 432-456.

51. Pelosi L, Kuhn L, Guetta D, Garin J, Geiselmann J, Lenski RE, Schneider D (2006) Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics* 173: 1851-1869.
52. Cooper TF, Rozen DE, Lenski RE (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc Natl Acad Sci U S A* 100: 1072-1077.
53. Cooper VS, Schneider D, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *J Bacteriol* 183: 2834-2841.
54. Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE, Blot M (1999) Genomic evolution during a 10,000-generation experiment with bacteria. *Proc Natl Acad Sci U S A* 96: 3807-3812.
55. Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M (2000) Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics* 156: 477-488.
56. Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE (2006) Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc Natl Acad Sci U S A* 103: 9107-9112.
57. Lenski RE, Winkworth CL, Riley MA (2003) Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *J Mol Evol* 56: 498-508.
58. Losos JB, Jackman TR, Larson A, Queiroz K, Rodriguez-Schettino L (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science* 279: 2115-2118.
59. Pigeon D, Chouinard A, Bernatchez L (1997) Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of lake whitefish (*Coregonus clupeaformis* Salmonidae). *Evolution* 51: 196-205.
60. Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L (2000) Rapid evolution of a geographic cline in size in an introduced fly. *Science* 287: 308-309.
61. Crandall KA, Kelsey CR, Imamichi H, Lane HC, Salzman NP (1999) Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol Biol Evol* 16: 372-382.
62. Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS (2000) Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* 406: 64-67.
63. Bishop AC, Xu J, Johnson RC, Schimmel P, Crecy-Lagard V (2002) Identification of the tRNA-dihydrouridine synthase family. *J Biol Chem* 277: 25090-25095.
64. Mallik P, Pratt TS, Beach MB, Bradley MD, Undamatla J, Osuna R (2004) Growth phase-dependent regulation and stringent control of *fis* are conserved processes in enteric bacteria and involve a single promoter (*fis P*) in *Escherichia coli*. *J Bacteriol* 186: 122-135.
65. Walker KA, Atkins CL, Osuna R (1999) Functional determinants of the *Escherichia coli fis* promoter: roles of -35, -10, and transcription initiation regions in the response to stringent control and growth phase-dependent regulation. *J Bacteriol* 181: 1269-1280.
66. Yuan HS, Finkel SE, Feng JA, Kaczor-Grzeskowiak M, Johnson RC, Dickerson RE (1991) The molecular structure of wild-type and a mutant Fis protein: relationship between mutational changes and recombinational enhancer function or DNA binding. *Proc Natl Acad Sci U S A* 88: 9558-9562.
67. Kostrewa D, Granzin J, Koch C, Choe HW, Raghunathan S, Wolf W, Labahn J, Kahmann R, Saenger W (1991) Three-dimensional structure of the *Escherichia coli* DNA-binding protein FIS. *Nature* 349: 178-180.
68. Simons RW, Houman F, Kleckner N (1987) Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene* 53: 85-96.
69. Gosink KK, Gaal T, Bokal AJ, Gourse RL (1996) A positive control mutant of the transcription activator protein FIS. *J Bacteriol* 178: 5182-5187.
70. Pan CQ, Finkel SE, Cramton SE, Feng JA, Sigman DS, Johnson RC (1996) Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J Mol Biol* 264: 675-695.
71. Ross W, Thompson JF, Newlands JT, Gourse RL (1990) *Escherichia coli* Fis protein activates ribosomal RNA transcription *in vitro* and *in vivo*. *EMBO J* 9: 3733-3742.
72. McLeod SM, Aiyar SE, Gourse RL, Johnson RC (2002) The C-terminal domains of the RNA polymerase alpha subunits: contact site with Fis and localization during co-activation with CRP at the *Escherichia coli proP* P2 promoter. *J Mol Biol* 316: 517-529.
73. Xu J, Johnson RC (1995) Fis activates the RpoS-dependent stationary-phase expression of *proP* in *Escherichia coli*. *J Bacteriol* 177: 5222-5231.
74. Fisher RA (1930) *The Genetical Theory of Natural Selection*. Oxford: Oxford University Press.
75. Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
76. Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ (1999) Different trajectories of parallel evolution during viral adaptation. *Science* 285: 422-424.

77. Owens RM, Pritchard G, Skipp P, Hodey M, Connell SR, Nierhaus KH, O'Connor CD (2004) A dedicated translation factor controls the synthesis of the global regulator Fis. *EMBO J* 23: 3375-3385.
78. Free A, Dorman CJ (1994) *Escherichia coli tyrT* gene transcription is sensitive to DNA supercoiling in its native chromosomal context: effect of DNA topoisomerase IV overexpression on *tyrT* promoter function. *Mol Microbiol* 14: 151-161.
79. Figueroa-Bossi N, Guerin M, Rahmouni R, Leng M, Bossi L (1998) The supercoiling sensitivity of a bacterial tRNA promoter parallels its responsiveness to stringent control. *EMBO J* 17: 2359-2367.
80. Cooper VS, Bennett AF, Lenski RE (2001) Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. *Evolution Int J Org Evolution* 55: 889-896.
81. Auner H, Buckle M, Deufel A, Kutateladze T, Lazarus L, Mavathur R, Muskhelishvili G, Pemberton I, Schneider R, Travers A (2003) Mechanism of transcriptional activation by FIS: role of core promoter structure and DNA topology. *J Mol Biol* 331: 331-344.
82. Skoko D, Yan J, Johnson RC, Marko JF (2005) Low-force DNA condensation and discontinuous high-force decondensation reveal a loop-stabilizing function of the protein Fis. *Phys Rev Lett* 95: 208101.
83. Skoko D, Yoo D, Bai H, Schnurr B, Yan J, McLeod SM, Marko JF, Johnson RC (2006) Mechanism of Chromosome Compaction and Looping by the *Escherichia coli* Nucleoid Protein Fis. *J Mol Biol* 364: 777-798.
84. Rochman M, Aviv M, Glaser G, Muskhelishvili G (2002) Promoter protection by a transcription factor acting as a local topological homeostat. *EMBO Rep* 3: 355-360.
85. Morett E, Bork P (1998) Evolution of new protein function: recombinational enhancer Fis originated by horizontal gene transfer from the transcriptional regulator NtrC. *FEBS Lett* 433: 108-112.
86. Elena SF, Whittam TS, Winkworth CL, Riley MA, Lenski RE (2005) Genomic divergence of *Escherichia coli* strains: evidence for horizontal transfer and variation in mutation rates. *Int Microbiol* 8: 271-278.
87. Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 26: 74-86.
88. Moran NA, Mira A (2001) The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol* 2: 12.
89. Cashel M, Gentry VJ, Hernandez VJ, Vinella D (1996) The stringent response. In: *Escherichia coli* and *Salmonella*: cellular and molecular biology. Washington DC: ASM Press. pp. 1458-1496.
90. Lederberg S (1966) Genetics of host-controlled restriction and modification of deoxyribonucleic acid in *Escherichia coli*. *J Bacteriol* 91: 1029-1036.
91. Link AJ, Phillips D, Church GM (1997) Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *J Bacteriol* 179: 6228-6237.
92. Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33: 103-119.
93. Sambrook J, Fritsch JE, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Plainview, NY: Cold Spring Harbor Laboratory Press.
94. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3: 418-426.
95. Powell BS, Rivas MP, Court DL, Nakamura Y, Turnbough CL, Jr. (1994) Rapid confirmation of single copy lambda prophage integration by PCR. *Nucleic Acids Res* 22: 5765-5766.
96. Appleman JA, Ross W, Salomon J, Gourse RL (1998) Activation of *Escherichia coli* rRNA transcription by FIS during a growth cycle. *J Bacteriol* 180: 1525-1532.
97. Xu J, Johnson RC (1997) Activation of RpoS-dependent *proP* P2 transcription by the Fis protein in vitro. *J Mol Biol* 270: 346-359.
98. Almiron M, Link AJ, Furlong D, Kolter R (1992) A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev* 6: 2646-2654.
99. Yamada H, Yoshida T, Tanaka K, Sasakawa C, Mizuno T (1991) Molecular analysis of the *Escherichia coli hns* gene encoding a DNA-binding protein, which preferentially recognizes curved DNA sequences. *Mol Gen Genet* 230: 332-336.
100. Peng H, Marians KJ (1993) *Escherichia coli* topoisomerase IV. Purification, characterization, subunit structure, and subunit interactions. *J Biol Chem* 268: 24481-24490.

RÉSULTATS

PARTIE III

UNE NOUVELLE  
FONCTION DE  
RÉGULATION POUR FIS

Comme nous l'avons vu dans l'introduction, la superhélicité de l'ADN et la protéine Fis jouent un rôle primordial dans l'organisation du chromosome d'*E. coli* ainsi que dans la transcription de nombreux gènes. Des mutations dans les gènes *topA* et *fis* ayant été fixées au cours de l'évolution expérimentale, et démontrées comme étant bénéfiques (voir Résultats, Partie I), nous avons voulu comprendre quels étaient les effets de ces mutations au sein de la cellule. Pour cela, nous avons réalisé des électrophorèses protéiques bidimensionnelles, pour comparer les profils protéiques globaux des souches isogéniques à l'ancêtre portant les allèles évolués *topA* ou *fis* isolés dans la population Ara-1, à celui du clone ancêtre. Quelques protéines dont la quantité est modifiée par ces mutations ont été identifiées, dont la porine majeure OmpF d'*E. coli* B. Des analyses moléculaires et biochimiques plus précises ont ensuite été réalisées.

## I. Matériel et méthodes

### A. Milieux de cultures, souches et plasmides

Les souches analysées par électrophorèse bidimensionnelle ont été cultivées dans le milieu DM250. Il s'agit du même milieu que celui utilisé lors de l'évolution expérimentale (voir Résultats, Partie I), mais avec une concentration 10 fois plus élevée en glucose, ce qui permet l'obtention d'une quantité suffisante de protéines.

Les souches 606, 606<sup>*topA-1*</sup>, 606<sup>*fis-1*</sup>, 606<sup>*topA-1fis-1*</sup> et 606 <sup>$\Delta$ *fis*</sup> utilisées ici ont été décrites auparavant (**Tableau 1**), et dérivent toutes de l'ancêtre REL606. La souche d'*E. coli* K12 utilisée est CF7968 (MG1655 *rph<sup>+</sup>  $\Delta$ lacI-Z*) et nous a été fournie par Michael Cashel. Une délétion de *fis* a également été construite dans cette souche, ce qui a conduit à la souche K12 <sup>$\Delta$ *fis*</sup>. Sa construction a été réalisée par recombinaison homologe dans le chromosome en utilisant le plasmide suicide pKO3 (comme décrit dans la Partie I des Résultats). Les plasmides pRS550 et pRS551, permettant la construction de fusions transcriptionnelles avec le gène rapporteur *lacZ*, ont été décrits par Simons et al. (1987).

## B. Electrophorèses bidimensionnelles

Le principe est d'effectuer une double séparation sur un échantillon protéique, la première se faisant en fonction du point isoélectrique et la seconde en fonction de la masse molaire des protéines.

Des précultures de nuit des souches d'intérêt ont été réalisées dans du milieu DM250, puis diluées au 1/100<sup>ème</sup> dans 1L de milieu DM250, incubées à 37°C avec agitation jusqu'à une  $DO_{600nm}=0,1$  (début de phase exponentielle), et centrifugées. Les culots sont lavés dans 50mM de Tris pH8, resuspendus dans 200 $\mu$ L d'une solution de 50mM Tris pH8, 0,2M DTT, 0,3% SDS et 1mM EDTA, et chauffés à 100°C pendant 5 min. Vingt microlitres d'une solution contenant 50mM Tris pH7,5, 50mM MgCl<sub>2</sub>, 1mg/mL DNaseI et 0,25mg/mL RNase A sont ajoutés aux extraits cellulaires et le tout est incubé pendant 10min sur la glace. Les protéines sont solubilisées par l'ajout de 800 $\mu$ L d'un tampon contenant 10M urée, 4% NP40, 0,1M DTT et 2,2% ampholines avec une gamme de pH allant de 3 à 10 (v/v) (Amersham Pharmacia). Les échantillons sont stockés à -20°C. Les électrophorèses bidimensionnelles en gels de polyacrylamide ont été réalisées avec un système Multiphor II (Amersham Pharmacia) pour l'isoélectrofocalisation et Protean II xi cell (Bio-Rad, Hercules, CA) pour la seconde dimension. Les bandes de première dimension (18 cm, gamme de pH de 4 à 8) sont réhydratées dans les échantillons protéiques (500mg) préparés dans 2M thiourée, 5M urée, 1,6% 3-[(3-cholamidopropyl)-diméthylammonio]-1-propanesulfonate (p/v), 100mM DTT et 0,8% ampholines. L'isoélectrofocalisation est effectuée jusqu'à atteindre l'équilibre de pH (Lelong et Rabilloud, 2003). Les bandes sont ensuite équilibrées pendant 10min dans une solution de Tris pH6,8 contenant 2,5% SDS, 30% glycérol, 6M urée et 50mM DTT. Cette étape est effectuée une seconde fois en remplaçant le DTT par 100mM d'iodoacétamide. Pour la seconde dimension, l'électrophorèse est réalisée comme décrit précédemment (Laemmli, 1970), en utilisant des gels d'acrylamide à 12%, avec les modifications suivantes : les tampons de migration de la cathode et de l'anode sont remplacés par une solution de 50mM Tris, 0,1% SDS et 200mM taurine ou 380mM glycine, respectivement. Les protéines sont ensuite colorées au bleu de Coomassie ou au nitrate d'argent.

### C. Analyse des électrophorèses bidimensionnelles

Le profil protéique global des 5 souches (606, 606<sup>topA-1</sup>, 606<sup>fis-1</sup>, 606<sup>topA-1fis-1</sup>, 606<sup>Δfis</sup>) a été réalisé en triplicats à partir de 3 cultures indépendantes de chaque souche. Chaque série de 5 souches a été effectuée le même jour, avec les mêmes solutions, pour donner un maximum de reproductibilité. Chaque profil a été comparé à celui de l'ancêtre, et seules les protéines dont la quantité variait dans le même sens pour chacun des 3 réplicats ont été analysées. Les spots correspondants ont été excisés à partir de gels colorés au bleu de Coomassie, oxydés par 7% de H<sub>2</sub>O<sub>2</sub>, et soumis à une digestion tryptique. Les peptides ont été analysés par MALDI-TOF-MS (matrix-assisted laser desorption ionization-time of flight-mass spectrometry) par resuspension dans une solution matrice (acide α-cyano-4-hydroxycinnamique à mi-saturation dans 60% d'acétonitrile et 0,1% d'acide trifluoroacétique). L'analyse est ensuite faite avec un spectromètre de masse MALDI-TOF (Autoflex, Bruker Daltonik, Bremen, Allemagne) en mode extraction avec une gamme de masse de 0 à 4200 Da. Une recherche consécutive automatique dans la base de données Swissprot Trembl a été effectuée à l'aide du logiciel Mascot. Lorsqu'aucun résultat n'a été obtenu par cette méthode, une analyse par nano-LC-TMS/MS (nano-liquid chromatography-tandem mass spectrometry) a été réalisée. Les peptides ont alors été extraits avec de l'acide formique et de l'acétonitrile et injectés dans un système CapLC (Waters, Milford, MA) nanoLC directement couplé à un spectromètre de masse QTOF Ultima (Waters). L'acquisition des données est faite avec le logiciel MassLynx (Waters), et l'identification des protéines est alors effectuée de la même façon que précédemment. Toutes les analyses par spectrométrie de masse ont été réalisées par Lauriane Kühn dans le cadre d'une collaboration avec le laboratoire Chimie des Protéines (DRDC-CI, ERM0201 CEA/INSERM/UJF, CEA, Grenoble) dirigé par Jérôme Garin.

### D. Fusions transcriptionnelles

La transcription des gènes *ompF*, *ompC*, *ompR*, *micF* et *greB* a été analysée par construction de fusions transcriptionnelles des promoteurs de chacun de ces gènes avec le gène rapporteur *lacZ*, codant la β-galactosidase. Les gènes *micF* et *greB* ont été inclus dans cette analyse car leur région promotrice est adjacente et divergente de celle de *ompC* et *ompR*,

respectivement. Les fragments contenant les différentes régions promotrices ont été amplifiés par PCR, puis clonés dans le plasmide pCRII-TOPO (Invitrogen). Pour *ompF*, le fragment amplifié s'étend des positions -422 à +148 (amorces ODS451 : 5'-gcagggacgatcactgccag-3' et ODS453 : 5'-agtcaagcaatctatttgca-3') et l'ADN chromosomique d'*E. coli* B ou K12 a été utilisé. Pour *ompC*, le fragment s'étend des positions -211 à +110 (amorces ODS456 : 5'-accaggagggacagtacttt-3' et ODS457 : 5'-cagtgtgtcaataacttaag-3') et a été amplifié à partir d'ADN chromosomique d'*E. coli* K12. Pour *ompR*, le fragment s'étend des positions -116 à +199 (amorces ODS396 : 5'-cggtgagataacgttccagcag-3' et ODS455 : 5'-gggcgttttcatctcgttga-3') et a été amplifié à partir d'ADN chromosomique d'*E. coli* B. Pour le promoteur de *micF*, le fragment cloné s'étend des positions -304 à +110 et les amorces utilisées sont ODS456 et ODS460 (5'-acagagattcaggatccgaatgacggtaa-3') et a été amplifié à partir d'ADN chromosomique d'*E. coli* K12. Toutes les positions nucléotidiques sont données par rapport au site +1 majeur d'initiation de la transcription pour chacun des gènes (Huang et al., 1992 ; Batchelor et al., 2005 ; Ferrario et al., 1995). La séquence de chacun des fragments a ensuite été vérifiée par séquençage, puis les fragments contenant les promoteurs des gènes *ompF*, *ompC* et *micF* ont été clonés dans le plasmide pRS551 (Simons et al., 1987). La région promotrice de *ompR* a quant à elle été clonée à la fois dans pRS551 et pRS550, dont le site multiple de clonage est inversé par rapport à pRS551, afin d'obtenir les fusions transcriptionnelles *ompR-lacZ* et *greB-lacZ*. (Les gènes *ompR* et *greB* sont divergents). Les fusions transcriptionnelles ont ensuite été réalisées de la même façon que précédemment (voir Résultats, Partie II, Materials and Methods) par intégration du bactériophage  $\lambda$ RS45 dans le chromosome de la souche d'intérêt et sélection et criblage des bactéries lysogènes sur les phénotypes de résistance à la kanamycine, dont le gène conférant la résistance est porté par pRS550, pRS551 et  $\lambda$ RS45, et de couleur bleue sur milieu contenant de la kanamycine et du Xgal.

## E. Retards sur gels

La fixation de Fis sur les régions promotrices de *ompF*, *ompC* et *ompR* a été étudiée par des expériences de retards sur gel. Tous les fragments d'ADN utilisés recouvrent les régions promotrices ainsi que les sites d'initiation de la transcription des gènes. Les mêmes fragments

que ceux clonés pour la construction des fusions transcriptionnelles (voir paragraphe précédent) ont été amplifiés par PCR. Les marquages radioactifs, ainsi que les conditions utilisées lors des expériences de retards sur gel, sont les mêmes que ceux décrits précédemment (voir Résultats, Partie II, Materials and Methods). Les contrôles pour tester la spécificité des retards observés sont également similaires à ceux décrits et utilisés précédemment.

## F. Empreintes à la DNase I

Les mêmes fragments d'ADN que ceux utilisés pour les retards sur gel ont été employés pour réaliser les empreintes à la DNase I, ainsi que la protéine Fis ancestrale purifiée comme décrit dans la Partie II des résultats (Materials and Methods). Pour chaque fragment d'ADN, les empreintes à la DNase I ont systématiquement été réalisées avec chacun des deux brins marqués, correspondant respectivement aux brins codant et non codant. Les marquages ont été réalisés par la T4 nucléotide kinase en marquant avec du  $\gamma$ -<sup>32</sup>P dATP (Amersham) l'extrémité 5' de chacune des deux amorces utilisées pour amplifier chacun des fragments d'ADN. Pour *ompF*, en plus du fragment obtenu à l'aide des amorces ODS451 et ODS453, deux autres fragments ont été utilisés, obtenus à l'aide des amorces ODS453-ODS458 (5'-ctaaacggaaatnttgcgt-3') et ODS452-ODS459 (5'-cagaaacaaaatnttccgtag-3'), pour confirmer l'empreinte à la DNase I (**Figure 34**). La quantité de fragment radiomarké dans une réaction d'empreinte à la DNase I est ici de 30cpm par tube ; les rendements de la réaction de radiomarkage ainsi que de la purification des produits PCR radiomarkés (PCR purification kit, Qiagen) n'étant pas toujours constants, la dilution appropriée pour obtenir 30cpm/tube a été systématiquement déterminée, et la quantité d'ADN présent dans le volume utilisé calculée. La quantité de protéine Fis à ajouter dans chaque échantillon a ainsi été déterminée afin d'obtenir le même rapport molaire que lors des expériences de retards sur gels. C'est ce rapport molaire qui est donné dans les figures.

Les complexes Fis/ADN ont été réalisés dans les mêmes conditions que pour les retards sur gel (voir Résultats Partie II, Materials and Methods), dans le même tampon et dans un volume total de 15 $\mu$ L. Après 20min d'incubation à 37°C, les échantillons sont traités par 0,75ng de DNase I (Worthington) pendant 35s à température ambiante. La réaction est arrêtée

par l'ajout de 50 $\mu$ L de phénol pur pH8, puis 180 $\mu$ L d'une solution d'arrêt (12,5mM EDTA pH8, 0,4M acétate de sodium et 250 $\mu$ g ADN de sperme de hareng) sont ajoutés. L'ADN est ensuite précipité puis resuspendu dans 5 $\mu$ L de bleu de formamide (10mM EDTA pH8, 5mg bleu de bromophénol et xylène cyanol dans 10mL de formamide), chauffé à 90°C pendant 3min et déposé sur un gel de polyacrylamide à 8% (TBE 1X, 1,5M urée). Afin de déterminer les positions des sites de fixation de Fis, une réaction de séquençage G+A de Maxam et Gilbert (Maxam et Gilbert, 1977), réalisée sur le même fragment radiomarqué, est déposée en parallèle sur le gel. L'électrophorèse est réalisée dans un tampon TBE 0,5X et, après un temps de migration adéquat, le gel est séché 1h30 à 80°C, et révélé de la même manière que les gels de retards.

## II. Résultats

### A. Modification des profils protéiques

Les profils protéiques de 4 souches ancestrales isogéniques portant les allèles évolués de *topA* et *fis* identifiés dans la population Ara-1, seuls ou en combinaison, ou une délétion de *fis* (606<sup>*topA-1*</sup>, 606<sup>*fis-1*</sup>, 606<sup>*topA-1fis-1*</sup>, 606 <sup>$\Delta$ *fis*</sup>) ont été comparés à celui du clone ancêtre par électrophorèses bidimensionnelles. Trois profils protéiques ont été réalisés à partir de trois cultures indépendantes de chacune des souches dans du milieu DM250, les protéines ayant été extraites en phase exponentielle. Seules les protéines dont la quantité variait de manière similaire au sein des 3 réplicats ont été identifiées par MALDI-TOF-MS ou LC-MS. La liste de ces protéines est donnée dans le **Tableau 2**.

Trois protéines varient ici de manière systématique soit dans les 4 souches, soit dans les 3 modifiées pour *fis*: PAL, OmpA et OmpF. Ces deux dernières protéines sont des porines, OmpF étant la porine principale présente chez *E. coli* B. Chez *E. coli* K12, les deux porines majeures sont OmpF et OmpC. Cette dernière est absente chez les dérivés de la souche *E. coli* B utilisés ici, suite à la délétion d'une partie du gène *ompC* (Schneider et al., 2002). Nous avons choisi de nous focaliser sur les analyses des porines, et notamment de OmpF. En effet, la

composition de la membrane externe pourrait jouer un rôle important au cours de cette évolution expérimentale. Nous avons donc étudié l'expression du gène *ompF* dans nos souches ancestrales isogéniques portant les allèles ancestraux ou évolués de *topA* et *fis*. Avant de détailler les expériences effectuées, une brève description des porines d'*E. coli* et de leur régulation est donnée.

## **B. Les porines d'*E. coli***

Trois porines majeures ont été décrites chez *E. coli* : OmpC, OmpF et OmpA. Les porines sont des protéines insérées dans la membrane externe, qui forment des pores permettant l'entrée passive de petits composants. Elles sont impliquées dans le transport de molécules, la discrimination se faisant essentiellement sur la base de leur taille. Les deux principales, OmpC et OmpF, se différencient par la taille des pores formés, OmpF formant des pores de 1,16nm, alors que ceux formés par OmpC sont plus petits (1,08nm) et rendent ainsi plus difficile la diffusion des molécules (Nikaido et Rosenberg, 1983). La quantité de chacune de ces porines est fonction de la composition du milieu dans lequel se trouve la cellule, et leur régulation se fait de façon opposée. Dans l'intestin par exemple, la concentration en nutriments, mais aussi en composés toxiques comme les sels biliaires, est élevée : dans ces conditions, la quantité de porine OmpC est plus élevée que celle de OmpF. Dans un milieu plus dilué, l'inverse se produit (Pratt et Silhavy, 1995). Dans des conditions de stress osmotique, OmpC est également majoritaire par rapport à OmpF. Ces changements d'expression sont dus à différentes voies de régulation des deux porines, et les mécanismes de régulation ont été particulièrement bien étudiés dans le cas d'un stress osmotique.

La régulation majeure exercée au niveau transcriptionnel passe par l'intermédiaire du système à deux composants OmpR/EnvZ (Hall et Silhavy, 1981). La protéine senseur EnvZ, une histidine kinase, est insérée dans la membrane externe et détecte les variations d'osmolarité de l'environnement. Ses activités de phosphorylation et de déphosphorylation exercées sur le régulateur de réponses OmpR permettent de réguler l'expression de OmpC et OmpF. En effet, suite à des variations d'osmolarité, il y a autophosphorylation de EnvZ sur le résidu H243, puis transfert du groupement phosphate sur le résidu D55 de OmpR, qui régule la transcription des deux porines (Norioka et al., 1986), la phosphorylation de OmpR

accroissant sa capacité à se fixer sur les régions promotrices de *ompF* et *ompC* (Aiba et Mizuno, 1990). A faible osmolarité, la quantité de OmpR phosphorylé (OmpR-P) est faible, et OmpR-P se fixe alors de façon coopérative sur trois sites à haute affinité, appelés F1 à F3, localisés en amont du promoteur de *ompF* (**Figure 29**). Ceci entraîne l'activation de la transcription de *ompF*, probablement grâce à une interaction avec l'ARN polymérase (Pratt et Silhavy, 1994). En revanche, cette concentration de OmpR-P n'est pas suffisante pour occuper les trois sites C1 à C3 de la région promotrice de *ompC* et entraîne donc une transcription à un niveau basal de ce gène. Lors d'un choc osmotique, il y a augmentation de la concentration en OmpR-P suite à l'activité de phosphorylation de EnvZ. OmpR-P se fixe alors sur les trois sites C1 à C3 de moindre affinité en amont du promoteur de *ompC*, ce qui active sa transcription, mais aussi sur un quatrième site d'affinité moindre appelé F4, localisé en amont des sites F1 à F3 de *ompF*. Ceci entraînerait le recouvrement des quatre sites en amont du promoteur et la formation d'une boucle d'ADN, ce qui inhiberait la transcription de *ompF* (**Figure 29B** ; Bergstrom et al., 1998 ; Mattison et al., 2002). La transcription des gènes *ompF* et *ompC* est également régulée par un second système à deux composants CpxA-CpxR (Batchelor et al., 2005) et par la protéine de type histone IHF qui se fixe sur deux sites en amont de *ompF*, ce qui entraîne une inhibition de la transcription (Ramani et al., 1992), qui est levée par OmpR-P. Des analyses génétiques ont également révélé que Lrp (Ferrario et al., 1995), CRP (Scott et Harwood, 1980) et H-NS (Suzuki et al., 1996) pouvaient jouer un rôle dans la régulation de *ompF* et *ompC*.

L'expression des porines est également régulée de façon post-transcriptionnelle par de petits ARNs, MicF, MicC et MicA, dont la séquence est complémentaire de l'extrémité 5' non traduite des ARNm de OmpF, OmpC et OmpA, respectivement (Delilhas et Forst, 2001 ; Chen et al., 2004 ; Udekwu et al., 2005). Ainsi, l'hybridation de ces trois ARNs au niveau des ARNm entraîne l'inhibition de la traduction de OmpF, OmpC et OmpA. De plus, le gène *micF* est transcrit de façon divergente à *ompC* (**Figure 29A**), et est activé de la même façon en cas de stress osmotique. Ceci a pour effet une inhibition très rapide de OmpF dans ces conditions (Ramani et al., 1994). La régulation de *ompF* par H-NS (Suzuki et al., 1996) et Lrp (Ferrario et al., 1995) est réalisée par l'intermédiaire de *micF*.

Chez *E. coli* B, une région correspondant à la totalité de *micF* et à une partie de *ompC* est délétée (Schneider et al., 2002). Ainsi, OmpF est la seule porine majoritaire présente chez *E. coli* B et sa régulation négative par *micF* est absente.

Au vu des résultats obtenus par électrophorèse bidimensionnelle, nous nous sommes d'abord intéressés à l'effet des mutations *topA* et *fis* sur la transcription de *ompF* chez *E. coli* B dans nos souches isogéniques dérivées du clone ancêtre.

### **C. Effets des mutations *topA* et *fis* sur la transcription de *ompF***

#### **1. La transcription de *ompF* dépend de la superhélicité de l'ADN et de Fis**

Une fusion transcriptionnelle a été construite pour le promoteur du gène *ompF* d'*E. coli* B (*ompFB*) avec le gène rapporteur *lacZ*, codant la  $\beta$ -galactosidase. Cette fusion a été introduite à l'aide du bactériophage  $\lambda$ RS45 (Simons et al., 1987) dans le chromosome du clone ancêtre, ainsi que des souches isogéniques dérivées 606<sup>*topA-1*</sup>, 606<sup>*fis-1*</sup> et 606 <sup>$\Delta$ *fis*</sup>. Des mesures d'activité  $\beta$ -galactosidase ont alors été effectuées en fonction de la phase de croissance et de la concentration en NaCl, et chacune des expériences a été réalisée à partir de 3 cultures indépendantes.

Les prélèvements réalisés pour chaque souche au cours du temps (**Figure 30**), montrent que la transcription de *ompF* chez *E. coli* B est quasiment constitutive, avec une légère baisse en fin de phase exponentielle. Une diminution de la transcription est observée dans la souche 606<sup>*topA-1*</sup>. Cette mutation entraînant une augmentation de superhélicité de l'ADN, ceci pourrait indiquer un rôle potentiel de la superhélicité dans l'activité de ce promoteur, ce qui a été suggéré auparavant (Graeme-Cook et al., 1989). Cette baisse de transcription est observée pendant toute la croissance cellulaire, avec cependant un effet plus marqué (diminution de la transcription de 3 fois) en phase exponentielle. Une baisse de transcription de l'ordre de 3 fois est également observée suite à une diminution de quantité ou à une disparition de Fis (souches 606<sup>*fis-1*</sup> et 606 <sup>$\Delta$ *fis*</sup>, respectivement). Ces résultats confirment les profils protéiques. Ainsi, Fis activerait la transcription de *ompF* de façon directe ou indirecte. Cette régulation est spécifique de la phase exponentielle et n'est plus observée en phase stationnaire. Ceci est cohérent avec le profil d'expression de Fis, protéine très abondante en phase exponentielle mais quasiment absente en phase stationnaire (voir Introduction Partie I, I.B.3).

La transcription du promoteur *ompFB* a également été mesurée, grâce à la même fusion transcriptionnelle, dans les souches 606 et 606<sup>Δfis</sup> en fonction de la concentration en NaCl, c'est-à-dire en absence et en présence d'un stress osmotique (**Figure 31**). Trois concentrations en NaCl ont été utilisées : 0,086M, correspondant à la concentration habituelle du milieu LB, 0,2M et 0,5M. Deux prélèvements ont été effectués pour chaque souche et chaque concentration, en début de phase exponentielle (DO<sub>600nm</sub>=0,2) et en phase stationnaire (DO<sub>600nm</sub>=4). Bien que le stress osmotique soit connu pour entraîner l'inhibition de la transcription de *ompF* chez *E. coli* K12 et ses dérivés (Barron et al., 1986), cette inhibition n'a pas été observée ici chez *E. coli* B (**Figure 31**). Une légère augmentation de la transcription de *ompF* semble même se produire en augmentant la concentration en NaCl. Ces résultats devraient être confirmés en introduisant dans nos expériences une souche *E. coli* K12, ainsi qu'un dérivé délété du gène *fis* (ces travaux sont en cours). D'autre part, la validité du stress osmotique devra également être réalisée par utilisation d'un milieu minimum additionné ou non de saccharose.

La régulation de *ompFB* par Fis révèle quelques différences lors d'un stress osmotique (**Figure 31**). En milieu LB classique (0,086M NaCl), l'activation de la transcription de *ompFB* par Fis a été retrouvée, uniquement en phase exponentielle. En revanche, lorsque la concentration en NaCl augmente, l'activation de la transcription de *ompFB* par Fis est détectée aussi bien en phase exponentielle qu'en phase stationnaire. De plus amples expériences seraient ici nécessaires afin de comprendre le rôle joué par Fis dans la régulation de *ompFB* en fonction du stress osmotique. Notamment, l'expression de Fis en fonction de la concentration en NaCl devrait être analysée. Ces expériences sont actuellement en cours au laboratoire.

## 2. L'activation de *ompF* par Fis est directe

Afin de savoir si le mécanisme d'activation de *ompF* par Fis était direct ou indirect, des retards sur gels ont été réalisés en utilisant d'une part la protéine Fis purifiée à partir de la souche ancêtre (voir Résultats, Partie II) et d'autre part la région promotrice de *ompF*, amplifiée également à partir de la souche ancêtre (**Figure 32**). Un retard est visible dès 0,2nM de Fis et augmente jusqu'à une concentration en protéine de 40nM (**Figure 32A**). De multiples

bandes retardées sont observées, ce qui pourrait indiquer la présence de plusieurs sites de fixation de Fis dans la région promotrice de *ompF*. La constante de dissociation, calculée de la même façon que dans la Partie II des résultats (Materials and Methods) est de 2nM. Trois types de contrôles pour tester la spécificité de la fixation ont été effectués. Le premier a consisté à ajouter un anticorps anti-Fis au mélange ADN-Fis. Un retard de migration supplémentaire (supershift) a été observé (résultat non montré). Le second a consisté à remplacer la protéine Fis purifiée par une fraction plus tardive de la purification sur SP-sépharose par FPLC (voir Résultats Partie II, Materials and Methods). Aucun retard de migration n'a alors été constaté (résultat non montré). Enfin, le troisième contrôle a consisté à ajouter deux types d'oligonucléotides non marqués radioactivement, en quantités croissantes, au mélange ADN-Fis (**Figure 32B**). Le premier type contient une séquence consensus de fixation de Fis (Finkel et Johnson, 1992), ici le site I à haute affinité présent dans la région promotrice du gène *fis* lui-même ; dans ce cas, le retard de migration observé diminue rapidement avec l'augmentation de la concentration de l'oligonucléotide. Le deuxième type correspond à la même séquence d'ADN mais pour laquelle le site de fixation de Fis a été mutagénisé. Aucune différence de retard de migration n'a alors été observée, quelle que soit la concentration en oligonucléotide. Tous ces contrôles démontrent que la fixation de Fis sur le promoteur *ompF* est spécifique. Ces résultats montrent également que Fis active la transcription de *ompF* de manière directe, en se fixant sur la région promotrice du gène.

### 3. Détermination des sites de fixation de Fis sur *ompF*

Afin de déterminer précisément les sites de fixation de Fis dans la région promotrice de *ompF*, nous avons réalisé des empreintes à la DNase I sur cette région protégée par Fis (**Figure 33**). Des quantités croissantes de Fis ont été employées, correspondant aux rapports molaires utilisés dans les expériences de retards sur gel. Quatre couples d'amorces ont été utilisés (**Figure 34**) pour amplifier par PCR quatre fragments d'ADN de la région promotrice de *ompF* B, qui ont ensuite permis de réaliser des expériences d'empreintes à la DNase I (**Figure 33**). Les empreintes à la DNase I révèlent 5 zones de protection par Fis dans la région promotrice de *ompF*, au sein desquelles sont observés des sites d'hypersensibilité (**Figure 33**) caractéristiques de la courbure engendrée par la fixation de Fis sur l'ADN, ce qui exposerait particulièrement

quelques bases localisées entre les deux domaines C-terminaux de Fis (Slany et Kersten, 1992). Cinq sites de fixation de Fis, I à V, ont ainsi été mis en évidence (**Figure 34**). Le premier est situé en amont de la boîte -35 du promoteur, chevauchant les sites F1, F2 et F3 de fixation de OmpR, ainsi qu'un des sites de fixation d'IHF. L'étendue de ce premier site, des positions -45 à -118, laisse supposer la fixation de plusieurs dimères de Fis, bien qu'un seul site proche de la séquence consensus ait pu être identifié. La localisation de ce site pourrait également suggérer des interactions entre Fis et OmpR au niveau de la régulation de *ompF*. Les quatre autres sites de fixation de Fis s'étendent des positions -130 à -363 et sont tous localisés entre les sites F1 et F4 de fixation de OmpR. Le centre des différents sites de fixation de Fis est espacé, respectivement de 69pb (I à II), 82pb (II à III), 60pb (III à IV) et 54pb (IV à V). L'éloignement de ces sites pourrait suggérer que les dimères de Fis puissent se fixer pour la plupart du même côté de la double hélice d'ADN et y induire une forte courbure qui entraînerait l'activation de la transcription (Newlands et al., 1992). *In vivo*, cette région pourrait déjà présenter une courbure intrinsèque importante, puisqu'elle est riche en AT, à laquelle viendrait s'ajouter celle induite par la fixation de Fis.

Nous avons ici démontré que Fis activait la transcription du gène *ompF* d'*E. coli* B en se fixant sur 5 sites présents dans la région promotrice. Bien que la régulation des porines soit relativement bien connue chez *E. coli* K12, et étudiée depuis plus de 30 ans maintenant, aucune régulation par Fis des gènes de porines, que ce soit *ompF* ou *ompC*, ou des gènes de régulation *ompR* et *envZ*, n'a été répertoriée, ce qui peut paraître surprenant. Ceci pourrait être lié au fait que la régulation que nous venons de mettre en évidence soit spécifique d'*E. coli* B. Nous avons donc réalisé les mêmes expériences dans l'isolat d'*E. coli* le plus utilisé en laboratoire : K12. Pour ce faire, nous avons utilisé la souche CF7968, un dérivé Lac<sup>-</sup> de la souche MG1655, dont la séquence du génome est disponible (Blattner et al., 1997). Au préalable, nous avons construit, toujours en utilisant le plasmide suicide pKO3, une délétion en phase du gène *fis* dans la souche CF7968 (appelée K12), ce qui nous a permis d'obtenir une souche isogénique CF7968<sup>Δfis</sup> (appelée K12<sup>Δfis</sup>). La régulation des gènes *ompF* et *ompR* a été comparée dans les 4 souches : B, K12, ainsi que leurs dérivés délétés du gène *fis*. L'utilisation d'un contexte génétique K12 nous a également permis d'inclure l'analyse de la régulation du

gène *ompC*, absent chez les dérivés d'*E. coli* B. La régulation de chacun de ces gènes a été analysée en fonction de la phase de croissance, en phase exponentielle et en phase stationnaire.

## **D. Régulation des gènes *ompF*, *ompC* et *ompR* chez *E. coli* B et K12**

Les trois mêmes types d'analyses ont été réalisés pour les régions promotrices de chacun des trois gènes : des fusions transcriptionnelles avec le gène *lacZ*, des retards sur gel pour analyser la fixation de Fis, et des empreintes à la DNase I pour identifier les sites de fixation potentiels.

### **1. Régulation du gène *ompFK12***

Les fusions transcriptionnelles du promoteur *ompFK12* ont été construites comme précédemment (voir paragraphe I.D. de cette partie) par clonage dans pRS551 d'un produit PCR correspondant à la région promotrice de *ompF* mais obtenu cette fois à partir de l'ADN génomique de K12, transfert sur l'ADN de  $\lambda$ RS45 et introduction dans le chromosome par lysogénéisation des souches K12 et K12 $\Delta$ *fis*. Le promoteur *ompFK12* contient trois mutations par rapport à celui d'*E. coli* B (**Figure 34**), localisées dans les sites de fixation F2 et F4 de OmpR et entre les boîtes -35 et -10 du promoteur.

Les cultures des deux souches K12 et K12 $\Delta$ *fis* portant les fusions transcriptionnelles ont été réalisées en triplicats dans du milieu LB. Des prélèvements ont été effectués au cours de la croissance et des mesures d'activités spécifiques  $\beta$ -galactosidase ont été réalisées (**Figure 35**). Contrairement au promoteur *ompFB* dont la transcription était quasiment constitutive tout au long de la croissance (**Figure 30**), la transcription de *ompFK12* présente un niveau de base en phase exponentielle de croissance et est activée en phase stationnaire (**Figure 35**) pour atteindre un niveau d'expression semblable à celui de *ompFB*. Une seconde différence a été observée entre *ompFK12* et *ompFB* : Fis ne semble pas réguler la transcription de *ompFK12*, ce qui pourrait expliquer l'absence d'activation du promoteur *ompFK12* en phase exponentielle par rapport à *ompFB* (**Figures 30 et 35**).

Une hypothèse qui permettrait d'expliquer en partie ce résultat serait que Fis ne se fixe pas sur le promoteur de K12, malgré de faibles différences entre les séquences des promoteurs *ompFB* et K12. Des expériences de retard sur gel ont donc été effectuées dans les mêmes conditions que celle décrites dans le paragraphe II.C.2., mais cette fois avec un fragment d'ADN correspondant au promoteur *ompFK12* (obtenu par PCR avec les mêmes amorces que précédemment). La fixation de Fis sur *ompFK12* est strictement équivalente à celle de Fis sur *ompFB*, avec une constante de dissociation de 1,5nM (résultats non montrés). Ce n'est donc pas une différence de liaison à l'ADN qui explique les différences de régulation observées entre *ompFB* et *ompFK12*.

Afin de comprendre les différences de régulation entre *ompFB* et *ompFK12*, nous avons mesuré l'activité des deux promoteurs en introduisant chacune des fusions transcriptionnelles *ompFB-lacZ* et *ompFK12-lacZ* dans chacune des souches *E. coli* B et K12. Des cultures des deux souches portant chacune des deux fusions transcriptionnelles ont été réalisées en triplicat en milieu LB. Des prélèvements ont été effectués au cours de la croissance et les activités spécifiques  $\beta$ -galactosidase ont été mesurées (**Figure 36**). Les deux résultats précédents ont été retrouvés : l'activité constitutive du promoteur *ompFB* dans *E. coli* B et l'activation du promoteur *ompFK12* dans *E. coli* K12 en phase stationnaire et à des niveaux équivalents à ceux de *ompFB*. En revanche, la fusion transcriptionnelle *ompFK12-lacZ* exprimée dans *E. coli* B révèle maintenant une expression de base, et donc une absence d'activation, en phase exponentielle, alors qu'une activation est toujours observée en phase stationnaire (**Figure 36**). Une légère diminution de la transcription se produit par la suite. Ainsi, il semblerait que les trois changements nucléotidiques entre les promoteurs *ompFB* et K12 pourraient être à l'origine des différences de transcription observées en phase exponentielle. Ces trois modifications nucléotidiques (ou l'une ou deux d'entre elles) chez *E. coli* B pourraient être à la base de l'activation de la transcription par Fis. Nous sommes actuellement en train de tester l'activation par Fis de *ompFK12* dans le contexte génétique d'*E. coli* B. Cependant, ces modifications de séquences nucléotidiques entre *ompFB* et *ompFK12* ne semblent pas être suffisantes pour expliquer les différences de régulation observées. En effet, la fusion transcriptionnelle *ompFB-lacZ* exprimée dans *E. coli* K12 révèle le même comportement que *ompFK12-lacZ*, c'est-à-dire une expression de base en phase exponentielle suivie d'une

activation en phase stationnaire (**Figure 36**). Ceci indique, en plus des différences de séquences nucléotidiques entre les promoteurs *ompFB* et K12, des différences de régulation entre les deux isolats d'*E. coli*. En particulier, des différences de séquences ont été identifiées dans les gènes *ompR* et *envZ* entre les deux isolats, ce qui pourrait expliquer une partie des différences de régulation observées. Ceci sera abordé dans la partie Discussion de ce chapitre.

Nous avons donc montré ici que la protéine Fis se fixait sur le promoteur de *ompF* de K12, sans pour autant avoir d'effet détectable sur sa transcription dans les conditions testées. Ceci est différent du promoteur *ompFB* qui est activé par Fis en phase exponentielle. Nous avons également démontré que deux conditions étaient nécessaires à cette activation de *ompFB* par Fis : la séquence promotrice issue de la souche B, ainsi que le contexte génomique d'*E. coli* B. Il est donc probable que les différences de régulation observées soient fonction de plusieurs facteurs, dont par exemple des différences au niveau du système à deux composantes OmpR/EnvZ.

## 2. Régulation des gènes *ompC* et *ompR*

Des fusions transcriptionnelles *ompC-lacZ* et *ompR-lacZ* ont été construites de la même façon que celle décrite précédemment pour *ompF*, avec les fragments d'ADN contenant les régions promotrices de *ompC* et *ompR* provenant respectivement des souches *E. coli* K12 et B.

Les fusions transcriptionnelles ont été introduites au site d'attachement de l'ADN du bactériophage  $\lambda$  dans les souches 606, 606 $\Delta_{fis}$ , K12 et K12 $\Delta_{fis}$ . Des prélèvements ont été effectués au cours de la croissance en milieu LB et les activités spécifiques  $\beta$ -galactosidase ont été mesurées (**Figure 37**). Ces mesures montrent trois résultats majeurs : i) l'activité du promoteur *ompC* est environ 3 fois plus élevée que celle du promoteur *ompR*, quelle que soit la phase de croissance ; ii) les deux promoteurs présentent une activité de base en phase exponentielle de croissance et sont activés d'un facteur 2 en phase stationnaire ; iii) aucune régulation par Fis de ces deux promoteurs n'a été observée.

Des expériences de retard sur gel ont été réalisées pour analyser la fixation éventuelle de Fis sur les promoteurs *ompC* et *ompR*. Les fragments d'ADN utilisés sont les mêmes que ceux clonés pour les fusions transcriptionnelles et les conditions expérimentales ainsi que les

contrôles sont les mêmes que ceux décrits précédemment (**Figure 38**). Les mêmes observations que dans le cas du promoteur *ompF* peuvent être faites : la première bande retardée est observée dès 0,2nM de Fis, et la quantité d'ADN retardé, ainsi que le nombre de bandes, augmentent avec la concentration en Fis, jusqu'à 20nM. Les constantes de dissociation calculées dans ces conditions sont d'environ 1,5nM pour chacun des promoteurs, ce qui indique une forte fixation de Fis. Pour localiser avec précision les sites de fixation, des expériences d'empreintes à la DNase I ont été réalisées sur ces deux promoteurs, et ce dans les mêmes conditions que pour le promoteur *ompF* (**Figure 39**). Ces expériences ont mis en évidence des sites de fixation de Fis dans les régions régulatrices de *ompR* et *ompC* (**Figure 40**).

La région promotrice de *ompR* présente cinq sites de fixation spécifique de Fis, contenant chacun une séquence proche de la séquence consensus de fixation de la protéine. Le profil de fixation de Fis est relativement atypique, avec trois différences importantes par rapport à ce qui avait été observé dans la région promotrice de *ompF* (**Figure 34 et 40A**) : i) le site I recouvre le codon ATG d'initiation de la traduction de *ompR* ; ii) les sites II et III sont situés sur une séquence correspondant à la région 5' non traduite de l'ARNm de *ompR* ; iii) le site IV recouvre le site d'initiation majeur de la transcription (Huang et al., 1992). L'absence d'effet de Fis sur la transcription de *ompR* malgré la présence de ces sites de fixation sera abordée dans la Discussion de cette partie. De la même façon, le site V recouvre une région localisée juste en amont du codon GTG d'initiation de la traduction de *greB* (**Figure 40A**).

En ce qui concerne la région promotrice de *ompC*, le profil de fixation de Fis est similaire à celui observé pour *ompR* (**Figure 40B**), avec le site I de fixation de Fis dans la région 5' non traduite de l'ARNm de *ompC* et proche du site d'initiation majeur de la transcription, T1 (Ferrario et al., 1995 ; Yoshida et al., 2006), le site II recouvrant la région -35 de ce promoteur. Chacun de ces sites présente une séquence consensus de fixation de Fis. En amont de ces deux sites, une région d'environ 100pb, appelée site III, est protégée des coupures par la DNase I, mais elle ne contient qu'une seule séquence consensus de fixation de Fis. Cette région recouvre en grande partie les sites de fixation C1 à C3 de OmpR, le site III étant localisé en amont.

## E. Régulation de *ompF* au cours de l'évolution expérimentale

Les résultats présentés dans ce paragraphe sont préliminaires. Les protéines FIS évoluées Ara-3 et Ara-4 (voir Résultats Partie II, Materials and Methods) ont été utilisées lors d'expériences de retards sur gel avec le promoteur *ompFB*, dans les mêmes conditions que celles décrites précédemment. Aucune fixation n'a alors été observée (résultats non montrés), suggérant une absence d'activation de *ompF* en phase exponentielle dans les clones évolués. Ceci reste cependant à confirmer par l'analyse des fusions transcriptionnelles *ompFB-lacZ* dans les souches ancestrales isogéniques portant les différents allèles évolués de *fis*, et dans les clones évolués isolés au cours du temps dans les différentes populations d'évolution. Un premier résultat nous indique toutefois que les mutations apparues dans le gène *fis* résultent en une absence d'activation du promoteur *ompFB* par la protéine. En effet, la fusion transcriptionnelle *ompFB-lacZ* introduite dans 606<sup>*fis*-1</sup> montre la même diminution d'activité par rapport à 606 que lorsqu'elle est introduite dans 606 <sup>$\Delta$ *fis*</sup> (**Figure 30**), montrant ainsi qu'une diminution de 3 fois de la quantité de protéine a le même effet sur la transcription de *ompFB* qu'une absence totale de Fis.

De plus, des différences de régulation au niveau de *ompF* se sont probablement produites au cours des 20 000 générations d'évolution expérimentale. En effet, l'estimation des quantités de OmpF par extraction des protéines membranaires et électrophorèse en gel SDS-PAGE, a révélé que les allèles *fis* évolués introduits dans le contexte génétique du clone ancêtre entraînaient une diminution de la quantité de OmpF. Ces allèles évolués diminuant la quantité ou l'activité de Fis, ces résultats confirment le mécanisme d'activation de Fis sur l'expression de *ompF*. En revanche, dans le contexte génétique des clones évolués isolés des différentes populations d'évolution, aucune diminution de la quantité de OmpF n'a été détectée malgré la présence des allèles évolués de *fis*. Ceci signifie que la régulation de *ompF* par Fis a été éliminée probablement à la suite d'autres mutations substituées dans ces populations. Ces phénomènes sont actuellement en cours d'étude.

### III. Discussion

Les profils protéomiques des souches ancestrales portant les allèles ancestraux ou évolués des gènes *fis* et/ou *topA*, ont révélé que l'expression d'un certain nombre de gènes codant des protéines ancrées dans la membrane externe (PAL, OmpA, OmpF), et de *surA* impliqué dans la régulation des protéines membranaires (Lazar et Kolter, 1996), était modifiée par la présence des allèles évolués. Nous nous sommes focalisés au cours de ce travail sur les gènes codant les porines majeures d'*E. coli*. Dans un premier temps, nous avons pu montrer que l'allèle évolué de *topA*, qui diminue l'activité de la TopoI et entraîne ainsi une augmentation de superhélicité de l'ADN, conduisait à une diminution de la transcription de *ompF*. Ceci avait déjà été observé chez *E. coli* K12 (Graeme-Cook et al., 1989), et ce phénomène pourrait également contribuer à la répression de la transcription de *ompF* lors d'un choc osmotique. En effet, le stress osmotique est connu pour augmenter de façon transitoire la superhélicité de l'ADN (Higgins et al., 1988). Nous avons pu également démontrer que la transcription de *ompF* chez *E. coli* B était dépendante d'une activation par Fis, mais uniquement au cours de la phase exponentielle de croissance. Le profil d'expression de Fis est tout à fait compatible avec cette activation, Fis étant abondante en phase exponentielle et quasiment absente en phase stationnaire. Il peut paraître surprenant que cette régulation de *ompF* par Fis n'ait jamais été détectée auparavant, et nous pensons que ceci est dû au fait que cette activation est spécifique d'*E. coli* B. Elle n'a en effet pas été observée dans une souche d'*E. coli* K12. Contrairement à la situation chez *E. coli* B où la transcription de *ompF* est constitutive, la transcription de *ompF* chez *E. coli* K12 est indépendante de Fis et, après un niveau de base en phase exponentielle, elle est activée en phase stationnaire. L'absence de forte transcription de *ompFK12* en phase exponentielle pourrait être liée à l'absence d'activation par Fis. Le mécanisme permettant une activation de la transcription de *ompFK12* et le maintien d'une transcription élevée de *ompFB* en phase stationnaire est inconnu.

Quelles pourraient être les bases moléculaires de cette différence de régulation de *ompF* entre *E. coli* B et K12 ? L'analyse de chacune des fusions transcriptionnelles *ompFB* et *ompFK12* dans chacune des deux souches a révélé qu'une forte transcription en phase exponentielle ne pouvait se faire que si deux conditions étaient réunies : avoir le promoteur de *ompFB* et le contexte génétique d'*E. coli* B.

En ce qui concerne le promoteur *ompFB*, trois différences nucléotidiques existent par rapport à *ompFK12* (**Figure 34**), deux d'entre elles étant situées au sein des sites de fixation F2 et F4 de OmpR. Dans notre hypothèse, c'est la mutation située dans le site F2, correspondant à une délétion d'un résidu T (**Figure 34**), qui serait un des facteurs cruciaux de la différence de régulation entre *E. coli* B et K12. Toujours selon notre hypothèse, Fis serait responsable de la transcription de *ompFB* en phase exponentielle. Il a récemment été montré que l'activation du promoteur *ompF* par OmpR se faisait par un modèle « galopant » (Yoshida et al., 2006). Ces auteurs ont démontré que chaque site F1 à F4 de fixation de OmpR était divisé en deux sous-sites a et b, chacun fixant une molécule de OmpR (**Figure 41**). Ainsi, six molécules de OmpR au total peuvent se fixer sur les sites F1 à F3. Le modèle « galopant » implique un ordre de fixation séquentiel, avec une première molécule de OmpR se fixant d'abord sur F1b, ce qui permet la fixation d'une seconde molécule de OmpR sur F1a, puis d'une troisième sur F2b, d'une quatrième sur F2a, etc... (**Figure 41**). L'activation de *ompF* a lieu à partir du recouvrement des sites F1 et F2 par OmpR (Bergstrom et al., 1998). Selon notre hypothèse, la délétion d'une base se produisant chez *E. coli* B au sein du site F2a entraînerait un déficit de reconnaissance de ce sous-site par OmpR, ce qui bloquerait le « galop ». Notre hypothèse est que cette mutation affecterait gravement la capacité de OmpR à activer ce promoteur. Cette même délétion d'un nucléotide chez *E. coli* B a également pour conséquence de rapprocher la séquence du site I de fixation de Fis que nous avons déterminé (**Figure 34**) de la séquence consensus de fixation de Fis (11pb conservées sur 13 au lieu de 10). Ceci améliorerait donc la capacité de liaison de Fis à cette région, et l'activation par Fis se substituerait à celle par OmpR. Cependant, nos résultats indiquent que ceci ne serait pas suffisant pour permettre l'activation du promoteur *ompFB* par Fis, puisque, dans le contexte génétique d'*E. coli* K12, celui-ci n'est pas activé par Fis (**Figure 36**). Ces résultats suggèrent en effet une régulation différentielle sur le promoteur *ompFB* entre *E. coli* B et K12. Nous suggérons que cette différence pourrait provenir des protéines OmpR ou EnvZ elles-mêmes. En effet, de multiples modifications nucléotidiques existent pour les gènes *ompR* et *envZ* entre *E. coli* B et K12, et trois d'entre elles conduisent à des changements non-synonymes d'acides aminés ; deux au niveau de OmpR (K6N et A130T), et un au niveau d'EnvZ (P41L). La protéine senseur EnvZ est l'histidine kinase qui phosphoryle le régulateur transcriptionnel OmpR et le résidu P41 n'est pas connu pour être impliqué dans ce processus. Les résidus K6 et A130 de OmpR ne sont pas

non plus connus pour être impliqués dans l'activité de la protéine. En revanche, le résidu A130 est situé dans la région reliant les domaines N-terminal et C-terminal et permettant la flexibilité de la protéine (Kenney et al., 1995). La protéine OmpR modifiée pourrait montrer un déficit de fixation à l'ADN qui favoriserait la liaison de Fis, qui se fixerait alors sur la région promotrice de *ompFB* et activerait la transcription en phase exponentielle. Ces hypothèses sont actuellement testées d'une part par des approches génétiques en introduisant les mutations *ompF* et *ompR* présentes chez *E. coli* B dans le chromosome d'*E. coli* K12, et réciproquement. La transcription de *ompF* sera alors analysée dans ces différents contextes génétiques. D'autre part, des approches biochimiques pourront être utilisées, notamment par purification des protéines OmpR d'*E. coli* B et K12 et par des expériences de retard sur gel avec ces protéines et la protéine Fis, en utilisant les promoteurs de *ompFB* ou K12.

Ces différences entre les deux souches d'*E. coli* entraîneraient donc une expression constitutive de *ompF* chez *E. coli* B. OmpC et OmpF étant les porines majeures d'*E. coli*, la perte de OmpC chez *E. coli* B aurait ainsi pu être compensée par ces mutations dans le promoteur de *ompF* et dans *ompR* et *envZ*. Cette compensation entraînerait la régulation d'un gène de porine par Fis, qui est un régulateur central chez les bactéries, permettant l'adaptation bactérienne à de multiples conditions de l'environnement (état nutritionnel, stress oxydatif...).

Enfin, nous avons également montré que Fis se fixait sur les régions promotrices des gènes *ompC* et *ompR*, sans toutefois réguler leur transcription de façon détectable. Trois hypothèses pourraient expliquer ce résultat. Tout d'abord, il est possible que les conditions testées ne soient pas celles dans lesquelles Fis intervient. Il est par exemple connu que Fis active *topA*, mais cela n'a pour l'instant été détecté que dans des conditions de stress oxydatif (Weinstein-Fischer et al., 2000). N'ayant testé qu'une seule condition, la croissance en milieu LB, il est possible que Fis régule ces promoteurs dans d'autres conditions. La seconde hypothèse serait que la régulation par Fis ait lieu sur les gènes transcrits de façon divergente à *ompC* et *ompR*, respectivement *micF* et *greB*. Les fusions transcriptionnelles de ces régions promotrices avec *lacZ* ont été réalisées. La fusion contenant le promoteur de *greB* ne montre pas de régulation par Fis au cours de la croissance en milieu LB, et celle contenant le promoteur de *micF* est en cours d'étude. La troisième hypothèse serait que Fis régule *ompC* et *ompR* au niveau post-transcriptionnel. A notre connaissance, une telle fonction n'a jamais été

montrée pour Fis. Cependant, la protéine Fis présente de fortes homologies avec le domaine C-terminal de NtrC (Morett et Bork, 1998, et Résultats, Partie II, Discussion), un régulateur de réponse qui fait partie d'un système à deux composants impliqués dans la réponse cellulaire aux carences en azote (Keener et Kustu, 1988). La partie C-terminale de NtrC permet sa fixation à l'ADN et donc l'activation transcriptionnelle de certains gènes. Cependant, il a aussi été montré pour cette protéine un rôle d'activateur de la traduction de *NifR3*, dont le gène est situé en amont de *ntrC* au sein du même opéron (Cullen et al., 1998). Des sites de fixation de NtrC ont été mis en évidence dans la région promotrice de *nifR3*, situés à la position -81 et autour du +1 de transcription, ce dernier site recouvrant également la boîte -10 du promoteur. Les auteurs ont alors suggéré une activation de la traduction par NtrC, soit par une interaction avec l'ARN polymérase, qui ralentirait l'étape d'élongation de la transcription et favoriserait le recrutement des ribosomes et ainsi la traduction, soit par une interaction directe avec l'ARNm de *nifR3* qui favoriserait également la traduction.

Cette situation correspond à celle observée pour les régions promotrices de *ompR* et *ompC*, où les sites de fixation de Fis recouvrent le +1 de transcription de *ompR* ainsi que la boîte -10 (site IV). Trois autres sites se situent également entre le +1 et l'ATG qui pourraient jouer sur la traduction de *ompR* (**Figure 40A**). De même pour *ompC*, les sites de fixation de Fis (**Figure 40B**) recouvrent une partie de la région 5' non traduite, ainsi que la boîte -35 du promoteur. Un rôle de Fis dans le contrôle de la traduction des deux protéines pourrait donc être imaginé. De plus, des interactions entre Fis et l'ARN polymérase ont été mises en évidence (Bokal et al., 1997). Cette hypothèse est actuellement testée au laboratoire.

# DISCUSSION GÉNÉRALE

Les résultats obtenus au cours de ce travail nous ont permis de révéler l'importance du rôle joué par la superhélicité de l'ADN dans les processus évolutifs. En effet, des changements parallèles et reproductibles de superhélicité ont été mesurés à trois niveaux.

Premièrement, nos résultats démontrent une association forte entre l'augmentation de superhélicité négative de l'ADN et celle du fitness. Un parallélisme phénotypique important a été mis en évidence avec une augmentation de superhélicité de l'ADN dans la majorité des populations. De plus, cette augmentation est souvent détectée pendant les 2000 premières générations, période de forte augmentation du fitness (Lenski et Travisano, 1994).

Deuxièmement, ce parallélisme phénotypique s'explique par un parallélisme génétique exceptionnel : tous les gènes impliqués dans la régulation de la superhélicité de l'ADN ne sont pas la cible de la sélection naturelle, seuls trois d'entre eux le sont : *topA*, *fis* et *dusB*. Bien que le mécanisme par lequel l'augmentation de superhélicité confère un avantage sélectif à ces cellules soit encore inconnu, il est probablement lié à une ou plusieurs fonctions associées aux protéines Fis, DusB et à la TopoI.

Troisièmement, une forte divergence allélique est observée pour un même gène dans les différentes populations. Quatre allèles évolués différents du gène *fis*, isolés dans quatre populations indépendantes affectent tous les niveaux de régulation du gène : transcriptionnel, traductionnel et l'activité de liaison à l'ADN de la protéine. Malgré cette divergence au niveau mutationnel, les différents allèles conduisent à des conséquences similaires : une diminution soit de la quantité, soit de l'activité de la protéine de type histone Fis.

Au cours de l'analyse de cette évolution expérimentale et des allèles évolués du gène *fis*, nous avons également pu mettre en évidence une nouvelle fonction de Fis : la régulation de l'expression des porines au niveau transcriptionnel chez *E. coli* B, et peut-être au niveau traductionnel chez *E. coli* K12.

De façon plus générale, ces travaux ont permis de montrer l'originalité et l'intérêt de ces stratégies d'évolution expérimentale. Les mutations isolées ont la particularité d'avoir été sélectionnées au cours de l'évolution bactérienne pour l'effet bénéfique qu'elles confèrent. Ceci est donc totalement différent des approches génétiques classiques où une délétion ou une modification de gènes est réalisée pour analyser leur fonction dans des environnements donnés ou en réponse à certains stress. En particulier, des fonctions nouvelles peuvent être

mises en évidence, par exemple ici l'implication de Fis dans le contrôle de la composition protéique de la membrane externe ou l'implication de *dusB* dans les processus évolutifs, avec une connexion probable avec Fis. Une analyse plus détaillée de ces mutations est nécessaire pour caractériser ces fonctions. D'autre part, les résultats obtenus dans ce travail, ainsi que précédemment (Woods et al., 2006), ont permis de montrer que les stratégies d'évolution expérimentale pouvaient constituer un excellent outil de prédiction. Ceci n'est évidemment possible que par la présence de plusieurs populations indépendantes. Sur la base des séquences de gènes, il est possible de prédire par des tests statistiques lesquels peuvent être des cibles de la sélection naturelle. Nous avons d'ailleurs utilisé cet outil de prédiction afin d'analyser des mutations isolées dans le gène *dusB* dont la fonction n'est pas connue précisément. Nous pouvons prédire que ces mutations sont bénéfiques, et que ce gène est un bon candidat pour intervenir dans la superhélicité de l'ADN, ces prédictions étant actuellement testées expérimentalement.

L'adaptation des bactéries au cours de cette stratégie d'évolution expérimentale se caractérise par deux processus fondamentaux, que nous allons maintenant discuter : le parallélisme évolutif et l'implication de la superhélicité de l'ADN et, de façon plus générale, des réseaux de régulation globale. De plus, ces travaux peuvent également soulever deux questions importantes, que nous discuterons également : par quels mécanismes une augmentation de superhélicité de l'ADN peut-elle être bénéfique ? Qu'en est-il de la généralisation de ces phénomènes dans d'autres conditions ?

## **I. Parallélisme de l'évolution**

Le parallélisme évolutif au cours de notre stratégie d'évolution expérimentale n'est pas observé uniquement au niveau du degré de superhélicité de l'ADN. Au niveau phénotypique tout d'abord, le fitness ainsi que le volume cellulaire augmentent de façon similaire au sein des 12 populations (Lenski et Travisano, 1994 ; Lenski, 2004). D'autre part, des diminutions, voire des pertes de fonctions cataboliques caractérisent les différentes populations (Cooper et Lenski, 2000 ; Cooper et al., 2001 ; Pelosi et al., 2006). Les profils globaux de transcription (Cooper et al., 2003) et d'expression des protéines (Pelosi et al., 2006) révèlent également des

changements parallèles de l'expression des gènes. D'autre part, à chaque fois que cela a été analysé, les changements phénotypiques parallèles sont associés à un parallélisme génétique, les mutations responsables des modifications étant situées dans les mêmes gènes dans la majorité des 12 populations. Ainsi, des mutations dans l'opéron *rbs* (dans 12 populations), et les gènes *malT* (8 populations) et *spoT* (8 populations) sont à l'origine de réductions parallèles des capacités cataboliques et de changements parallèles d'expression des gènes (Cooper et al., 2001 ; 2003 ; Pelosi et al., 2006). Comme dans le cas du gène *fts*, le parallélisme génétique s'accompagne d'une grande divergence allélique et il est fort probable qu'elle conduise également à un fort parallélisme moléculaire.

Un fort degré de parallélisme a également été observé lors d'autres expériences d'évolution, y compris dans des conditions différentes ou en utilisant des organismes différents. En effet, des mutations sont sélectionnées de façon répétée au sein des mêmes gènes, et certains gènes sont même des cibles fréquentes de la sélection naturelle dans différentes stratégies d'évolution expérimentale. D'autres expériences d'évolution ont employé *E. coli* comme organisme modèle, en propageant ces bactéries soit dans des conditions de carences prolongées, soit dans des conditions de cultures continues en chémostat où la nourriture est apportée continuellement. Dans la première expérience, des modifications parallèles des mêmes phénotypes permettent l'adaptation aux carences. En l'occurrence, les bactéries accroissent leurs capacités à utiliser des acides aminés comme source de carbone (Zinser et Kolter, 1999). Des mutations se produisant en parallèle dans les mêmes gènes (impliqués dans le catabolisme des acides aminés, dans *rpoS* codant le facteur  $\sigma^S$  de la phase stationnaire ou dans *Irp* codant un régulateur de type histone) sont responsables de ces changements phénotypiques (Zambrano et al., 1993 ; Zinser et Kolter, 2000). Dans des conditions de cultures continues en chémostat, des modifications parallèles au niveau du transport du glucose à travers la membrane ont été détectées dans des populations indépendantes d'*E. coli* (Ferenci, 2001). A nouveau, un fort parallélisme génétique en est responsable, avec notamment des mutations dans *rpoS* (Seeto et al., 2004) et *malT* (Manché et al., 1999). De plus, l'étude de l'évolution de souches pathogènes d'*E. coli* montre que les lignées anciennes ont acquis les mêmes facteurs de virulence de façon indépendante et parallèle, incluant un flot de pathogénicité ainsi que des gènes de toxines (Reid et al., 2000).

D'autres expériences d'évolution ont utilisé d'autres organismes qu'*E. coli*, et le parallélisme phénotypique et génétique de l'évolution a été retrouvé à tous les niveaux de complexité des organismes : depuis des lignées indépendantes de virus adaptées à des températures élevées, où la moitié des substitutions nucléotidiques sont identiques (Bull et al., 1997), en passant par des populations indépendantes de levure adaptées à des transitions glucose-galactose dans lesquelles le même gène répresseur de la voie d'utilisation du galactose a été désactivé (Segre et al., 2006), jusqu'à des poissons vivants dans des lacs différents où des modifications parallèles de phénotypes (métabolisme énergétique et déplacement) sont dues à un parallélisme génétique (augmentation de l'expression des gènes codant la cristalline- $\gamma$  et la parvalbumine- $\beta$  notamment) (Derome et al., 2006). Dans ce dernier cas, il ne s'agit plus d'évolution expérimentale, mais d'études réalisées dans des environnements naturels. De telles conclusions ont également été obtenues pour d'autres eucaryotes, comme par exemple la drosophile (Sucena et al., 2003). Le parallélisme est donc un point commun à tous les types d'évolutions, en laboratoire ou dans la nature. Les systèmes d'évolution expérimentale sont donc des outils tout à fait appropriés pour comprendre les mécanismes sous-jacents de l'adaptation d'un organisme à un environnement donné.

## II. Plasticité des réseaux de régulation globale

Toutes les mutations conférant une augmentation du fitness, et mises en évidence à ce jour dans la stratégie d'évolution expérimentale, affectent les gènes ou opérons suivants : *rbs* (Cooper et al., 2001), *spoT* (Cooper et al., 2003), *topA*, *fis*, *dusB* (ce travail), *malT* (Pelosi et al., 2006), *nadR*, *pbpA-rodA*, *pykF*, *hokB-sokB* (Schneider et al., 2000a ; Woods et al., 2006). Près d'un tiers de ces gènes (*topA*, *fis*, *spoT*) codent des régulateurs globaux de l'expression des gènes, intervenant dans deux des plus importants réseaux de régulation génétique : la superhélicité de l'ADN et la réponse stringente. La plasticité des réseaux de régulation globale est donc un paramètre crucial permettant l'adaptation des bactéries à leur environnement. Au sein de la population modèle Ara-1, les mutations substituées dans les gènes *topA* et *spoT* apparaissent très tôt (avant 500 et 1500 générations, respectivement) et confèrent les augmentations de fitness les plus importantes identifiées à ce jour (13% et 10%,

respectivement) (Cooper et al., 2003 ; Crozat et al., 2005). La modification de ces deux réseaux de régulation fondamentaux de la cellule doit forcément engendrer des effets pleïotropes, et il est improbable que tous ces effets soient bénéfiques. Certains d'entre eux sont sans doute neutres, voire faiblement néfastes, mais compensés par des effets bénéfiques plus importants. Il est alors possible d'imaginer qu'une fois ces mutations fixées, la cellule puisse continuer à s'adapter en compensant les effets néfastes par d'autres mutations plus tardives. Un tel effet de compensation pourrait expliquer le bénéfice conféré par les mutations du gène *fis*, dont nous avons démontré qu'elles provoquaient une diminution de la quantité ou de l'activité de Fis. En effet, nous avons observé une augmentation de la quantité de Fis chez des clones évolués isolés à 2000 générations, c'est-à-dire avant l'apparition des mutations dans *fis*. Cette augmentation de la quantité de Fis pourrait être néfaste et être liée à une autre mutation fixée avant 2000 générations. Ainsi, les mutations *fis*, toutes fixées vers 10 000 générations, permettraient de compenser ce « défaut ». Pour essayer de confirmer cette hypothèse, nous avons introduit par recombinaison homologue dans le chromosome de la souche ancêtre une mutation affectant la boîte -35 du promoteur de *fis* (le résidu T en position -34 est modifié en G) et résultant en une augmentation de 18 fois de la transcription de *fis* (Walker et al., 1999). Nous avons vérifié cette augmentation par Western blot et mesuré le fitness de cette souche par des expériences de compétitions, qui ont révélé une diminution de fitness. Ainsi, les mutations *fis* pourraient compenser les effets néfastes liés à une augmentation de Fis en diminuant la concentration ou l'activité de Fis. Un tel phénomène de compensation d'effets pleïotropes a été observé dans une autre étude d'évolution expérimentale, au cours de laquelle 7 populations indépendantes d'*E. coli* ont été adaptées pendant 600 à 1000 générations à deux conditions différentes : du milieu minimum contenant du glycérol ou du lactate comme seule source de carbone (Fong et al., 2005). L'analyse du profil de transcription de ces 14 populations au cours de l'évolution révèle que l'expression de nombreux gènes est modifiée en début d'expérience en réponse aux nouvelles conditions de croissance. Ceci est rapidement suivi d'un retour de l'expression de la plupart de ces gènes au niveau initial. Ainsi, un ensemble relativement restreint de gènes est exprimé différenciellement dans les clones évolués par rapport à l'ancêtre, et un nombre important de changements compensatoires d'expression des gènes est observé au cours de l'évolution (Fong et al., 2005).

La plasticité des réseaux de régulation est également visible chez les isolats naturels. En effet, l'étude de 41 isolats cliniques de *Salmonella* sérotype Typhi montre que 15 isolats présentent des mutations de types insertions, délétions et mutations non-sens dans le gène *rpoS*, codant le facteur  $\sigma^S$  de réponse aux stress (Hengge-Aronis, 2002) requis pour la virulence chez la souris (Robbe-Saule et al., 2003). Ce régulateur est également très variable chez différents isolats naturels d'*E. coli* qui présentent une variation de l'activité de RpoS corrélée avec des différences de résistance à la pression osmotique (Robey et al., 2001 ; King et al., 2004). Les isolats de laboratoire et naturels d'*E. coli* présentent des niveaux relatifs différents des deux facteurs sigma RpoS et RpoD, ce qui affecte la transcription globale et aboutit à une variabilité des capacités métaboliques et de résistance à divers stress (King et al., 2004). Ces deux facteurs sigma entrent en compétition pour la liaison à l'ARN polymérase, la liaison de RpoD améliorant la croissance, celle de RpoS favorisant la résistance aux stress (King et al., 2004). La modification de ce réseau de régulation majeur dont le pivot est RpoS permet donc l'adaptation à des environnements différents.

La complexité et la plasticité des réseaux de régulation pourraient ainsi avoir été sélectionnées au cours de l'évolution, de par leur capacité à accroître le spectre évolutif des organismes. Dans notre expérience d'évolution, la plasticité des réseaux de régulation permet une adaptation rapide par de grandes modifications de l'expression des gènes, puis les effets bénéfiques sont conservés, et les effets néfastes seraient compensés par des mutations ultérieures. Ceci constitue non seulement un premier indice sur le mécanisme d'adaptation des bactéries à leur environnement, mais aussi un phénotype supplémentaire révélant une évolution parallèle au sein des 12 populations.

### **III. Les mécanismes moléculaires de l'adaptation**

Au cours de l'évolution expérimentale, deux catégories de mutations bénéfiques ont été sélectionnées. La première touche des gènes de structure ou des régulateurs d'un petit nombre de gènes, ce qui conduit à des phénotypes relativement clairs et évidents. Par exemple, les délétions de l'opéron *rbs* provoquent la perte de la capacité à utiliser le ribose (Cooper et al., 2001), et les mutations dans le gène *malT* diminuent la capacité des bactéries à utiliser le

maltose comme source de carbone (Pelosi et al., 2006). La deuxième catégorie de mutations affecte des gènes de régulation globale, comme la superhélicité de l'ADN ou la réponse stringente. Les effets de ces mutations sont beaucoup plus difficiles à appréhender, pour plusieurs raisons. Tout d'abord, les effets pleïotropes de ces mutations peuvent rendre les effets bénéfiques majeurs difficiles à caractériser. Ensuite, les mutations sélectionnées ont des effets subtils. En effet, elles ne correspondent pas à des délétions ou des inactivations de fonction, mais entraînent des variations plus faibles : par exemple, les mutations du gène *fis* dans les populations Ara-1 et Ara+1 diminuent la quantité de protéine de 3 fois. De plus, les travaux effectués au cours de cette thèse et de celle de N. Philippe (Philippe, 2006) suggèrent que le bénéfice apporté par les mutations isolées n'est pas directement lié aux fonctions connues de ces gènes. Par exemple, les mutations bénéfiques affectant le gène *spoT*, codant une des protéines majeures responsable du métabolisme (synthèse et dégradation) de (p)ppGpp, l'effecteur de la réponse stringente (Cashel et al., 1996), ne modifient aucunement son niveau dans la cellule, ni ses taux de synthèse ou de dégradation (Philippe, Crozat, Lenski et Schneider, manuscrit en préparation). De même, Fis étant connue pour réprimer la transcription des gènes codant la gyrase (Schneider et al., 1999), la diminution de quantité ou d'activité de Fis observée au cours de l'évolution pourrait conduire à une augmentation de la quantité de gyrase, ce qui permettrait d'expliquer la seconde augmentation de superhélicité observée à 20 000 générations dans la population Ara-1 (Crozat et al., 2005). Cependant, les expériences de Western blot effectuées n'ont pas permis de détecter une telle modification suite aux mutations du gène *fis* (résultats non montrés).

L'augmentation de superhélicité de l'ADN pourrait être bénéfique par augmentation de la transcription des ARN stables qui est dépendante d'un niveau élevé de superhélicité négative (Ross et al., 1990 ; Muskhelishvili et Travers, 2003). Ceci permettrait alors une croissance plus rapide et pourrait ainsi conduire à une augmentation du fitness. Cependant, des fusions transcriptionnelles entre le promoteur P1 de l'opéron d'ARN ribosomiques *rrnB* avec le gène rapporteur *lacZ*, introduites dans le chromosome des souches ancestrales isogéniques 606, 606<sup>*fis-1*</sup>, 606<sup>*topA-1*</sup>, et 606<sup>*topA-1fis-1*</sup>, n'ont révélé aucune différence significative dans l'expression de ce promoteur. Nous avons également quantifié les contenus cellulaires en ADN, ARN et protéines de ces quatre souches, ce qui n'a également révélé aucune différence.

Le bénéfice conféré par les mutations *fis* et *topA* ne réside donc pas dans une différence phénotypique évidente.

La superhélicité de l'ADN est connue pour favoriser la transcription de gènes nécessaires à la croissance, comme les opérons d'ARN ribosomiques, notamment lors de l'initiation de la transcription. Ces gènes sont caractérisés par la présence d'un discriminateur riche en GC situé entre la boîte -10 du promoteur et le site +1 d'initiation de la transcription. Ce discriminateur constitue une barrière énergétique qu'il faut surmonter pour l'initiation de la transcription de ces gènes. L'augmentation de superhélicité pourrait ainsi améliorer le taux d'initiation de la transcription de certains de ces gènes. Ceci pourrait conférer un avantage sélectif important d'une part au niveau de la phase de croissance, mais aussi à la reprise de la croissance lors des transferts journaliers des cellules dans du milieu frais.

Nous avons montré que le bénéfice apporté par une augmentation de la superhélicité de l'ADN était associé à des mutations apparues non pas dans l'ensemble des gènes régulant la superhélicité, mais spécifiquement dans *topA*, *fis* et *dusB*. Ainsi, des fonctions spécifiques de ces trois gènes pourraient avoir été soumises à pression de sélection au cours de cette évolution expérimentale. Les allèles évolués des gènes *topA* et *fis* pourraient entraîner des modifications du profil d'expression globale, dont certaines pourraient conférer une augmentation du fitness. En effet, des mutations du gène *topA* sont connues pour affecter les profils protéomiques chez les bactéries (Steck et al., 1993), et Fis est connue pour réguler la transcription d'un grand nombre de gènes (Kelly et al., 2004). Nous avons démontré au cours de ces travaux que les mutations bénéfiques dans *topA* et *fis* provoquaient une diminution de l'activité de la TopoI ainsi que de Fis. En ce qui concerne Fis, elle est capable de se fixer de façon spécifique sur des sites à haute affinité, mais également de manière aspécifique. De plus, l'affinité de Fis pour ses différents sites de fixation est très variable (Muskhelishvili et Travers, 2003). La diminution de la quantité ou de l'activité de Fis ne pourrait donc affecter qu'un certain nombre de sites à plus faible affinité et de ce fait ne modifier la transcription que de cette catégorie de gènes dépendants de Fis. La transcription des gènes d'opérons d'ARN ribosomiques ne serait ainsi pas affectée.

Enfin, le bénéfice apporté par ces mutations pourrait être lié à un effet sur la structure globale du chromosome. En effet, Fis est impliquée dans la formation de domaines topologiques et constitue une barrière à ces domaines (Hardy et Cozzarelli, 2005). La

diminution de quantité ou d'activité de liaison à l'ADN de Fis pourrait diminuer le nombre de barrières topologiques et, dans les conditions de l'évolution expérimentale, ceci pourrait avoir un effet bénéfique en accélérant les processus de réplication du chromosome en limitant les pauses du complexe de réplication dues à la présence de ces barrières (Postow et al., 2001).

L'augmentation de superhélicité de l'ADN via les allèles évolués de *topA* pourrait également favoriser une structure chromosomique permettant une meilleure ségrégation des chromosomes-fils au moment de la division cellulaire. En effet, lors d'une augmentation de superhélicité, l'efficacité de décaténation des chromosomes-fils augmente de manière exponentielle avec la longueur de l'ADN (Alexandrov et al., 1999).

#### **IV. La superhélicité de l'ADN dans la nature**

Nous avons observé une augmentation systématique de la superhélicité dans 10 des 12 populations d'évolution, ces modifications étant bénéfiques dans ces conditions. Afin d'étendre ces résultats à d'autres environnements, des expériences de compétitions ont été réalisées dans différents milieux avec les souches ancêtres portant les allèles évolués de *fis* et *topA* identifiés dans la population Ara-1 (Résultats Partie II). Quelles que soient les conditions (milieu riche, source de carbone différente ou basse température), les deux mutations se sont révélées bénéfiques. Il semblerait donc qu'une augmentation de superhélicité soit bénéfique dans une grande variété d'environnements. Les bactéries, dans leur environnement naturel, sont confrontées à des fluctuations des conditions externes (carences nutritionnelles, apports nutritifs, stress de température, oxydatif, osmotique...). Cela se traduit par une forte diversité des réseaux de régulation dans les isolats naturels, par exemple des niveaux relatifs de RpoS et RpoD, ce qui conduit alors à des compromis évolutifs entre croissance et survie (King et al., 2004). L'implication évidente de la superhélicité de l'ADN dans les processus évolutifs pourrait se traduire de la même façon dans les isolats naturels par un polymorphisme important du niveau de superhélicité. Cela permettrait alors des compromis évolutifs et une adaptation appropriée à différents types d'environnements.

## V. Perspectives

Le but principal de cette expérience d'évolution en laboratoire étant de comprendre les mécanismes d'adaptation des bactéries à leur environnement, il est crucial de disséquer les différentes étapes de l'évolution, en essayant de comprendre les effets des différentes mutations bénéfiques les unes par rapport aux autres. L'histoire évolutive d'une population peut notamment être reconstruite. Par exemple dans notre population modèle, une dizaine de mutations conférant une augmentation du fitness ont été caractérisées. Un des projets du laboratoire est de les introduire une par une, dans leur ordre d'apparition, dans la souche ancêtre, et en parallèle, d'introduire chacune de ces mutations séparément dans l'ancêtre. Leur effet pourra alors être analysé à deux niveaux : dans le contexte génétique de l'ancêtre, et dans le contexte génétique où elle est apparue.

Dans le cas particulier des mutations *fis* et *topA*, l'étude de leurs effets sur les domaines topologiques et la mise en place de barrières de ces domaines devrait être envisagée. Ceci pourrait par exemple se faire de deux façons, qui ont été décrites dans l'introduction, soit par l'étude des événements de recombinaison entre sites *res* (Higgins et al., 1996), soit par l'étude du profil de transcription de gènes sensibles à la superhélicité avant et après coupure du chromosome par une enzyme de restriction (Postow et al., 2004). L'effet des mutations *fis* et *topA* sur la réplication et la ségrégation des chromosomes pourrait également être étudié.

Concernant la protéine Fis, son rôle sur l'expression des porines reste à élucider, en tentant par exemple d'identifier les conditions dans lesquelles cette protéine régule les gènes *ompF*, *ompC* et *ompR* chez *E. coli* K12, ainsi qu'un rôle éventuel dans la traduction des ARNm. La seconde découverte liée à Fis est son rôle commun potentiel avec DusB. La construction d'une des mutations isolées dans le gène *dusB* est en cours, et son rôle sur le fitness et la superhélicité de l'ADN sera étudié. De futures études pourraient également porter sur l'influence des mutations localisées dans *dusB* sur la traduction de l'ARNm de *fis*.

Enfin, la topologie de l'ADN dans des isolats naturels pourrait être étudiée, afin tout d'abord de voir comment le niveau de superhélicité varie entre différents isolats, et ensuite de tester si une augmentation de superhélicité serait bénéfique, dans des isolats pathogènes ou dans différentes conditions de croissance par exemple.

RÉFÉRENCES  
BIBLIOGRAPHIQUES

- Adams DE, Shekhtman EM, Zechiedrich EL, Schmid MB, Cozzarelli NR. **1992**. The role of topoisomerase IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. *Cell* 71:277-288.
- Ahumada A, Tse-Dinh YC. **1998**. The Zn(II) binding motifs of *Escherichia coli* DNA topoisomerase I is part of a high-affinity DNA binding domain. *Biochem.Biophys.Res.Commun.* 251:509-514.
- Aiba H, Mizuno T. **1990**. Phosphorylation of a bacterial activator protein, OmpR, by a protein kinase, EnvZ, stimulates the transcription of the *ompF* and *ompC* genes in *Escherichia coli*. *FEBS Lett.* 261:19-22.
- Aiyar SE, McLeod SM, Ross W, Hirvonen CA, Thomas MS, Johnson RC, Gourse RL. **2002**. Architecture of Fis-activated transcription complexes at the *Escherichia coli* *rrnB* P1 and *rrnE* P1 promoters. *J.Mol.Biol.* 316:501-516.
- Alexandrov AI, Cozzarelli NR, Holmes VF, Khodursky AB, Peter BJ, Postow L, Rybenkov V, Vologodskii AV. **1999**. Mechanisms of separation of the complementary strands of DNA during replication. *Genetica* 106:131-140.
- Ali BM, Amit R, Braslavsky I, Oppenheim AB, Gileadi O, Stavans J. **2001**. Compaction of single DNA molecules induced by binding of integration host factor (IHF). *Proc.Natl.Acad.Sci.U.S.A* 98:10658-10663.
- Altuvia S, Almiron M, Huisman G, Kolter R, Storz G. **1994**. The *dps* promoter is activated by OxyR during growth and by IHF and sigma S in stationary phase. *Mol.Microbiol.* 13:265-272.
- Aravind L, Leipe DD, Koonin EV. **1998**. Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* 26:4205-4213.
- Ashizawa Y, Yokochi T, Ogata Y, Shobuike Y, Kato J, Ikeda H. **1999**. Mechanism of DNA gyrase-mediated illegitimate recombination: characterization of *Escherichia coli* *gyrA* mutations that confer hyper-recombination phenotype. *J.Mol.Biol.* 289:447-458.
- Asoh S, Matsuzawa H, Ishino F, Strominger JL, Matsushashi M, Ohta T. **1986**. Nucleotide sequence of the *pbpA* gene and characteristics of the deduced amino acid sequence of penicillin-binding protein 2 of *Escherichia coli* K12. *Eur.J.Biochem.* 160:231-238.
- Auner H, Buckle M, Deufel A, Kutateladze T, Lazarus L, Mavathur R, Muskhelishvili G, Pemberton I, Schneider R, Travers A. **2003**. Mechanism of transcriptional activation by FIS: role of core promoter structure and DNA topology. *J.Mol.Biol.* 331:331-344.
- Aussel L, Barre FX, Aroyo M, Stasiak A, Stasiak AZ, Sherratt D. **2002**. FtsK is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell* 108:195-205.
- Badaut C, Williams R, Arluison V, Bouffartigues E, Robert B, Buc H, Rimsky S. **2002**. The degree of oligomerization of the H-NS nucleoid structuring protein is related to specific binding to DNA. *J.Biol.Chem.* 277:41657-41666.
- Balke VL, Gralla JD. **1987**. Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *J.Bacteriol.* 169:4499-4506.
- Ball CA, Johnson RC. **1991a**. Efficient excision of phage lambda from the *Escherichia coli* chromosome requires the Fis protein. *J.Bacteriol.* 173:4027-4031.
- Ball CA, Johnson RC. **1991b**. Multiple effects of Fis on integration and the control of lysogeny in phage lambda. *J.Bacteriol.* 173:4032-4038.
- Ball CA, Osuna R, Ferguson KC, Johnson RC. **1992**. Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. *J.Bacteriol.* 174:8043-8056.
- Baquero MR, Bouzon M, Varea J, Moreno F. **1995**. *sbmC*, a stationary-phase induced SOS *Escherichia coli* gene, whose product protects cells from the DNA replication inhibitor microcin B17. *Mol.Microbiol.* 18:301-311.
- Barron A, May G, Bremer E, Villarejo M. **1986**. Regulation of envelope protein composition during adaptation to osmotic stress in *Escherichia coli*. *J.Bacteriol.* 167:433-438.
- Bartlett MS, Gaal T, Ross W, Gourse RL. **2000**. Regulation of rRNA transcription is remarkably robust: FIS compensates for altered nucleoside triphosphate sensing by mutant RNA polymerases at *Escherichia coli* *rrn* P1 promoters. *J.Bacteriol.* 182:1969-1977.

- Batchelor E, Walthers D, Kenney LJ, Goulian M. **2005**. The *Escherichia coli* CpxA-CpxR envelope stress response system regulates expression of the porins *ompF* and *ompC*. *J.Bacteriol.* 187:5723-5731.
- Bergerat A, de Massy B, Gabelle D, Varoutas PC, Nicolas A, Forterre P. **1997**. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* 386:414-417.
- Bergstrom LC, Qin L, Harlocker SL, Egger LA, Inouye M. **1998**. Hierarchical and co-operative binding of OmpR to a fusion construct containing the *ompC* and *ompF* upstream regulatory sequences of *Escherichia coli*. *Genes Cells* 3:777-788.
- Bhriain NN, Dorman CJ. **1993**. Isolation and characterization of a *topA* mutant of *Shigella flexneri*. *Mol.Microbiol.* 7:351-358.
- Bishop AC, Xu J, Johnson RC, Schimmel P, Crecy-Lagard V. **2002**. Identification of the tRNA-dihydrouridine synthase family. *J.Biol.Chem.* 277:25090-25095.
- Blattner FR, Plunkett G, III, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. **1997**. The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453-1474.
- Bokal AJ, Ross W, Gaal T, Johnson RC, Gourse RL. **1997**. Molecular anatomy of a transcription activation patch: FIS-RNA polymerase interactions at the *Escherichia coli* *rrnBP1* promoter. *EMBO J.* 16:154-162.
- Bordes P, Conter A, Morales V, Bouvier J, Kolb A, Gutierrez C. **2003**. DNA supercoiling contributes to disconnect sigmaS accumulation from sigmaS-dependent transcription in *Escherichia coli*. *Mol.Microbiol.* 48:561-571.
- Bosch L, Nilsson L, Vijgenboom E, Verbeek H. **1990**. FIS-dependent trans-activation of tRNA and rRNA operons of *Escherichia coli*. *Biochim.Biophys.Acta* 1050:293-301.
- Boye E, Lyngstadaas A, Lobner-Olesen A, Skarstad K, Wold S. **1993**. Regulation of DNA replication in *Escherichia coli*. In: *DNA replication and the cell cycle*. Berlin: Springer. 15-26.
- Brown PO, Cozzarelli NR. **1979**. A sign inversion mechanism for enzymatic supercoiling of DNA. *Science* 206:1081-1083.
- Brown PO, Cozzarelli NR. **1981**. Catenation and knotting of duplex DNA by type 1 topoisomerases: a mechanistic parallel with type 2 topoisomerases. *Proc.Natl.Acad.Sci.U.S.A* 78:843-847.
- Browning DF, Beatty CM, Sanstad EA, Gunn KE, Busby SJ, Wolfe AJ. **2004**. Modulation of CRP-dependent transcription at the *Escherichia coli* *acsP2* promoter by nucleoprotein complexes: anti-activation by the nucleoid proteins FIS and IHF. *Mol.Microbiol.* 51:241-254.
- Browning DF, Grainger DC, Beatty CM, Wolfe AJ, Cole JA, Busby SJ. **2005**. Integration of three signals at the *Escherichia coli* *nrf* promoter: a role for Fis protein in catabolite repression. *Mol.Microbiol.* 57:496-510.
- Browning DF, Cole JA, Busby SJ. **2000**. Suppression of FNR-dependent transcription activation at the *Escherichia coli* *nir* promoter by Fis, IHF and H-NS: modulation of transcription initiation by a complex nucleo-protein assembly. *Mol.Microbiol.* 37:1258-1269.
- Bull JJ, Badgett MR, Wichman HA, Huelsenbeck JP, Hillis DM, Gulati A, Ho C, Molineux IJ. **1997**. Exceptional convergent evolution in a virus. *Genetics* 147:1497-1507.
- Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N, Davey M, Parkinson J, Greenblatt J, Emili A. **2005**. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 433:531-537.
- Cabrera JE, Jin DJ. **2003**. The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. *Mol.Microbiol.* 50:1493-1505.
- Cashel M, Gentry VJ, Hernandez VJ, Vinella D. **1996**. The stringent response. In: *Escherichia coli and Salmonella: cellular and molecular biology*. Washington DC: ASM Press. 1458-1496.
- Champoux JJ. **2001**. DNA topoisomerases: structure, function, and mechanism. *Annu.Rev.Biochem.* 70:369-413.
- Chatterji M, Nagaraja V. **2002**. GyrI: a counter-defensive strategy against proteinaceous inhibitors of DNA gyrase. *EMBO Rep.* 3:261-267.
- Chen S, Zhang A, Blyn LB, Storz G. **2004**. MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. *J.Bacteriol.* 186:6689-6697.
- Chen SJ, Wang JC. **1998**. Identification of active site residues in *Escherichia coli* DNA topoisomerase I. *J.Biol.Chem.* 273:6050-6056.

- Cheng B, Zhu CX, Ji C, Ahumada A, Tse-Dinh YC. **2003**. Direct interaction between *Escherichia coli* RNA polymerase and the zinc ribbon domains of DNA topoisomerase I. *J.Biol.Chem.* 278:30705-30710.
- Cheng YS, Yang WZ, Johnson RC, Yuan HS. **2000**. Structural analysis of the transcriptional activation on Fis: crystal structures of six Fis mutants with different activation properties. *J.Mol.Biol.* 302:1139-1151.
- Cheung KJ, Badarinarayana V, Selinger DW, Janse D, Church GM. **2003**. A microarray-based antibiotic screen identifies a regulatory role for supercoiling in the osmotic stress response of *Escherichia coli*. *Genome Res.* 13:206-215.
- Cooper TF, Rozen DE, Lenski RE. **2003**. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc.Natl.Acad.Sci.U.S.A* 100:1072-1077.
- Cooper VS, Lenski RE. **2000**. The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407:736-739.
- Cooper VS, Schneider D, Blot M, Lenski RE. **2001**. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *J.Bacteriol.* 183:2834-2841.
- Corbett KD, Berger JM. **2003**. Structure of the topoisomerase VI-B subunit: implications for type II topoisomerase mechanism and evolution. *EMBO J.* 22:151-163.
- Corbett KD, Berger JM. **2004**. Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases. *Annu.Rev.Biophys.Biomol.Struct.* 33:95-118.
- Corbett KD, Shultzaberger RK, Berger JM. **2004**. The C-terminal domain of DNA gyrase A adopts a DNA-bending beta-pinwheel fold. *Proc.Natl.Acad.Sci.U.S.A* 101:7293-7298.
- Crellin P, Sewitz S, Chalmers R. **2004**. DNA looping and catalysis; the IHF-folded arm of Tn10 promotes conformational changes and hairpin resolution. *Mol.Cell* 13:537-547.
- Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D. **2005**. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169:523-532.
- Cullen PJ, Bowman WC, Hartnett DF, Reilly SC, Kranz RG. **1998**. Translational activation by an NtrC enhancer-binding protein. *J.Mol.Biol.* 278:903-914.
- Cunningham CW, Jeng K, Husti J, Badgett M, Molineux IJ, Hillis DM, Bull JJ. **1997**. Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. *Mol.Biol.Evol.* 14:113-116.
- Cusick ME, Belfort M. **1998**. Domain structure and RNA annealing activity of the *Escherichia coli* regulatory protein StpA. *Mol.Microbiol.* 28:847-857.
- Dame RT. **2005**. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol.Microbiol.* 56:858-870.
- Dame RT, Goosen N. **2002**. HU: promoting or counteracting DNA compaction? *FEBS Lett.* 529:151-156.
- Dame RT, Wyman C, Goosen N. **2000**. H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res.* 28:3504-3510.
- Dame RT, Wyman C, Goosen N. **2001**. Structural basis for preferential binding of H-NS to curved DNA. *Biochimie* 83:231-234.
- Dame RT, Wyman C, Wurm R, Wagner R, Goosen N. **2002**. Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the *rrnB* P1. *J.Biol.Chem.* 277:2146-2150.
- Dame RT, van Mameren J, Luijsterburg MS, Mysiak ME, Janicijevic A, Pazdzior G, van der Vliet PC, Wyman C, Wuite GJ. **2005**. Analysis of scanning force microscopy images of protein-induced DNA bending using simulations. *Nucleic Acids Res.* 33:68-74.
- Delihias N, Forst S. **2001**. MicF: an antisense RNA gene involved in response of *Escherichia coli* to global stress factors. *J.Mol.Biol.* 313:1-12.
- Deng S, Stein RA, Higgins NP. **2004**. Transcription-induced barriers to supercoil diffusion in the *Salmonella typhimurium* chromosome. *Proc.Natl.Acad.Sci.U.S.A* 101:3398-3403.
- Deng S, Stein RA, Higgins NP. **2005**. Organization of supercoil domains and their reorganization by transcription. *Mol.Microbiol.* 57:1511-1521.
- Depew RE, Liu LF, Wang JC. **1978**. Interaction between DNA and *Escherichia coli* protein omega. Formation of a complex between single-stranded DNA and omega protein. *J.Biol.Chem.* 253:511-518.
- Derome N, Duchesne P, Bernatchez L. **2006**. Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchill) ecotypes. *Mol.Ecol.* 15:1239-1249.
- DiGate RJ, Marians KJ. **1989**. Molecular cloning and DNA sequence analysis of *Escherichia coli topB*, the gene encoding topoisomerase III. *J.Biol.Chem.* 264:17924-17930.

- Dorman CJ. **2004**. H-NS: a universal regulator for a dynamic genome. *Nat.Rev.Microbiol.* 2:391-400.
- Dorman CJ, Higgins CF. **1987**. Fimbrial phase variation in *Escherichia coli*: dependence on integration host factor and homologies with other site-specific recombinases. *J.Bacteriol.* 169:3840-3843.
- Dorman CJ, Deighan P. **2003**. Regulation of gene expression by histone-like proteins in bacteria. *Curr.Opin.Genet.Dev.* 13:179-184.
- Drlica K, Zhao X. **1997**. DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiol.Mol.Biol.Rev.* 61:377-392.
- Drolet M. **2006**. Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol.Microbiol.* 59:723-730.
- Drolet M, Broccoli S, Rallu F, Hraiky C, Fortin C, Masse E, Baaklini I. **2003**. The problem of hypernegative supercoiling and R-loop formation in transcription. *Front Biosci.* 8:210-221.
- Earnshaw WC, Heck MM. **1985**. Localization of topoisomerase II in mitotic chromosomes. *J.Cell Biol.* 100:1716-1725.
- Endo T, Sasaki N, Tanaka I, Nakata M. **2002**. Compact form of DNA induced by DNA-binding protein HU. *Biochem.Biophys.Res.Comm.* 290:546-551.
- Eriksson PR, Mendiratta G, McLaughlin NB, Wolfsberg TG, Marino-Ramirez L, Pompa TA, Jainerin M, Landsman D, Shen CH, Clark DJ. **2005**. Global regulation by the yeast Spt10 protein is mediated through chromatin structure and the histone upstream activating sequence elements. *Mol.Cell Biol.* 25:9127-9137.
- Espeli O, Marians KJ. **2004**. Untangling intracellular DNA topology. *Mol.Microbiol.* 52:925-931.
- Esposito D, Gerard GF. **2003**. The *Escherichia coli* Fis protein stimulates bacteriophage lambda integrative recombination in vitro. *J.Bacteriol.* 185:3076-3080.
- Falconi M, Prosseda G, Giangrossi M, Beghetto E, Colonna B. **2001**. Involvement of FIS in the H-NS-mediated regulation of *virF* gene of *Shigella* and enteroinvasive *Escherichia coli*. *Mol.Microbiol.* 42:439-452.
- Fennell-Fezzie R, Gradia SD, Akey D, Berger JM. **2005**. The MukF subunit of *Escherichia coli* condensin: architecture and functional relationship to kleisins. *EMBO J.* 24:1921-1930.
- Ferenci T. **2001**. Hungry bacteria--definition and properties of a nutritional state. *Environ.Microbiol.* 3:605-611.
- Ferrario M, Ernsting BR, Borst DW, Wiese DE, Blumenthal RM, Matthews RG. **1995**. The leucine-responsive regulatory protein of *Escherichia coli* negatively regulates transcription of *ompC* and *micF* and positively regulates translation of *ompF*. *J.Bacteriol.* 177:103-113.
- Filutowicz M, Ross W, Wild J, Gourse RL. **1992**. Involvement of Fis protein in replication of the *Escherichia coli* chromosome. *J.Bacteriol.* 174:398-407.
- Finkel SE, Johnson RC. **1992**. The Fis protein: it's not just for DNA inversion anymore. *Mol.Microbiol.* 6:3257-3265.
- Fong SS, Joyce AR, Palsson BO. **2005**. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res.* 15:1365-1372.
- Franco RJ, Drlica K. **1988**. DNA gyrase on the bacterial chromosome. Oxolinic acid-induced DNA cleavage in the *dnaA-gyrB* region. *J.Mol.Biol.* 201:229-233.
- Franco RJ, Drlica K. **1989**. Gyrase inhibitors can increase *gyrA* expression and DNA supercoiling. *J.Bacteriol.* 171:6573-6579.
- Free A, Dorman CJ. **1995**. Coupling of *Escherichia coli hns* mRNA levels to DNA synthesis by autoregulation: implications for growth phase control. *Mol.Microbiol.* 18:101-113.
- Free A, Porter ME, Deighan P, Dorman CJ. **2001**. Requirement for the molecular adapter function of StpA at the *Escherichia coli bgl* promoter depends upon the level of truncated H-NS protein. *Mol.Microbiol.* 42:903-917.
- Frenkiel-Krispin D, Ben Avraham I, Englander J, Shimoni E, Wolf SG, Minsky A. **2004**. Nucleoid restructuring in stationary-state bacteria. *Mol.Microbiol.* 51:395-405.
- Friedberg E, Walker GC, Siede W. **1995**. DNA repair and mutagenesis. ASM press, Washington, DC. 376.
- Gellert M, Mizuuchi K, O'Dea MH, Nash HA. **1976**. DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc.Natl.Acad.Sci.U.S.A* 73:3872-3876.

- Georgopoulos CP. **1977**. A new bacterial gene (*groPC*) which affects lambda DNA replication. *Mol.Gen.Genet.* 151:35-39.
- Gerrish PJ, Lenski RE. **1998**. The fate of competing beneficial mutations in an asexual population. *Genetica* 102-103:127-144.
- Gille H, Egan JB, Roth A, Messer W. **1991**. The FIS protein binds and bends the origin of chromosomal DNA replication, *oriC*, of *Escherichia coli*. *Nucleic Acids Res.* 19:4167-4172.
- Gmuender H, Kuratli K, Di Padova K, Gray CP, Keck W, Evers S. **2001**. Gene expression changes triggered by exposure of *Haemophilus influenzae* to novobiocin or ciprofloxacin: combined transcription and translation analysis. *Genome Res.* 11:28-42.
- Goldberg MD, Johnson M, Hinton JC, Williams PH. **2001**. Role of the nucleoid-associated protein Fis in the regulation of virulence properties of enteropathogenic *Escherichia coli*. *Mol.Microbiol.* 41:549-559.
- Gonzalez-Gil G, Bringmann P, Kahmann R. **1996**. FIS is a regulator of metabolism in *Escherichia coli*. *Mol.Microbiol.* 22:21-29.
- Goosen N, van de Putte P. **1995**. The regulation of transcription initiation by integration host factor. *Mol.Microbiol.* 16:1-7.
- Graeme-Cook KA, May G, Bremer E, Higgins CF. **1989**. Osmotic regulation of porin expression: a role for DNA supercoiling. *Mol.Microbiol.* 3:1287-1294.
- Grainger DC, Hurd D, Goldberg MD, Busby SJ. **2006**. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.*
- Grant RA, Filman DJ, Finkel SE, Kolter R, Hogle JM. **1998**. The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat.Struct.Biol.* 5:294-303.
- Grossman AD, Erickson JW, Gross CA. **1984**. The *htpR* gene product of *E. coli* is a sigma factor for heat-shock promoters. *Cell* 38:383-390.
- Hall MN, Silhavy TJ. **1981**. Genetic analysis of the *ompB* locus in *Escherichia coli* K-12. *J.Mol.Biol.* 151:1-15.
- Hanvey JC, Shimizu M, Wells RD. **1988**. Intramolecular DNA triplexes in supercoiled plasmids. *Proc.Natl.Acad.Sci.U.S.A* 85:6292-6296.
- Hardy CD, Cozzarelli NR. **2005**. A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure. *Mol.Microbiol.* 57:1636-1652.
- Hatfield GW, Benham CJ. **2002**. DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu.Rev.Genet.* 36:175-203.
- Heck MM. **1997**. Condensins, cohesins, and chromosome architecture: how to make and break a mitotic chromosome. *Cell* 91:5-8.
- Heggeler-Bordier B, Wahli W, Adrian M, Stasiak A, Dubochet J. **1992**. The apical localization of transcribing RNA polymerases on supercoiled DNA prevents their rotation around the template. *EMBO J.* 11:667-672.
- Hengen PN, Bartram SL, Stewart LE, Schneider TD. **1997**. Information analysis of Fis binding sites. *Nucleic Acids Res.* 25:4994-5002.
- Hengge-Aronis R. **2002**. Stationary phase gene regulation: what makes an *Escherichia coli* promoter sigmaS-selective? *Curr.Opin.Microbiol.* 5:591-595.
- Hiasa H, DiGate RJ, Marians KJ. **1994**. Decatenating activity of *Escherichia coli* DNA gyrase and topoisomerases I and III during *oriC* and pBR322 DNA replication in vitro. *J.Biol.Chem.* 269:2093-2099.
- Hiasa H, Marians KJ. **1996**. Two distinct modes of strand unlinking during theta-type DNA replication. *J.Biol.Chem.* 271:21529-21535.
- Higgins CF, Dorman CJ, Stirling DA, Waddell L, Booth IR, May G, Bremer E. **1988**. A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *Salmonella typhimurium* and *Escherichia coli*. *Cell* 52:569-584.
- Higgins NP, Yang X, Fu Q, Roth JR. **1996**. Surveying a supercoil domain by using the gamma delta resolution system in *Salmonella typhimurium*. *J.Bacteriol.* 178:2825-2835.
- Hiraga S, Niki H, Ogura T, Ichinose C, Mori H, Ezaki B, Jaffe A. **1989**. Chromosome partitioning in *Escherichia coli*: novel mutants producing anucleate cells. *J.Bacteriol.* 171:1496-1505.
- Hirano M, Hirano T. **2006**. Opening closed arms: long-distance activation of SMC ATPase by hinge-DNA interactions. *Mol.Cell* 21:175-186.

- Hirsch M, Elliott T. **2005**. Fis regulates transcriptional induction of RpoS in *Salmonella enterica*. *J.Bacteriol.* 187:1568-1580.
- Hirvonen CA, Ross W, Wozniak CE, Marasco E, Anthony JR, Aiyar SE, Newburn VH, Gourse RL. **2001**. Contributions of UP elements and the transcription factor FIS to expression from the seven *rrn* P1 promoters in *Escherichia coli*. *J.Bacteriol.* 183:6305-6314.
- Hommais F, Krin E, Laurent-Winter C, Soutourina O, Malpertuy A, Le Caer JP, Danchin A, Bertin P. **2001**. Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol.Microbiol.* 40:20-36.
- Huang L, Tsui P, Freundlich M. **1992**. Positive and negative control of *ompB* transcription in *Escherichia coli* by cyclic AMP and the cyclic AMP receptor protein. *J.Bacteriol.* 174:664-670.
- Ikeda H, Aoki K, Naito A. **1982**. Illegitimate recombination mediated *in vitro* by DNA gyrase of *Escherichia coli*: structure of recombinant DNA molecules. *Proc.Natl.Acad.Sci.U.S.A* 79:3724-3728.
- Jeong KS, Ahn J, Khodursky AB. **2004**. Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*. *Genome Biol.* 5:R86.
- Johansson J, Balsalobre C, Wang SY, Urbonaviciene J, Jin DJ, Sonden B, Uhlin BE. **2000**. Nucleoid proteins stimulate stringently controlled bacterial promoters: a link between the cAMP-CRP and the (p)ppGpp regulons in *Escherichia coli*. *Cell* 102:475-485.
- Johnson RC, Bruist MF, Simon MI. **1986**. Host protein requirements for *in vitro* site-specific DNA inversion. *Cell* 46:531-539.
- Johnson RC, Johnson LM, Schmidt JW, Gardner JF. **2005**. The major nucleoid proteins in the structure and function of the *Escherichia coli* chromosome. In: *The bacterial chromosome*. Washington DC: ASM Press. 65-132.
- Jones PG, Krah R, Tafuri SR, Wolffe AP. **1992**. DNA gyrase, CS7.4, and the cold shock response in *Escherichia coli*. *J.Bacteriol.* 174:5798-5802.
- Josephson BL, Fraenkel DG. **1969**. Transketolase mutants of *Escherichia coli*. *J.Bacteriol.* 100:1289-1295.
- Kaidow A, Wachi M, Nakamura J, Magae J, Nagai K. **1995**. Anucleate cell production by *Escherichia coli* delta hns mutant lacking a histone-like protein, H-NS. *J.Bacteriol.* 177:3589-3592.
- Kampranis SC, Bates AD, Maxwell A. **1999**. A model for the mechanism of strand passage by DNA gyrase. *Proc.Natl.Acad.Sci.U.S.A* 96:8414-8419.
- Kar S, Adhya S. **2001**. Recruitment of HU by piggyback: a special role of GalR in repressosome assembly. *Genes Dev.* 15:2273-2281.
- Kar S, Edgar R, Adhya S. **2005**. Nucleoid remodeling by an altered HU protein: reorganization of the transcription program. *Proc.Natl.Acad.Sci.U.S.A* 102:16397-16402.
- Keener J, Kustu S. **1988**. Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NTRB and NTRC of enteric bacteria: roles of the conserved amino-terminal domain of NTRC. *Proc.Natl.Acad.Sci.U.S.A* 85:4976-4980.
- Kelly A, Goldberg MD, Carroll RK, Danino V, Hinton JC, Dorman CJ. **2004**. A global role for Fis in the transcriptional control of metabolism and type III secretion in *Salmonella enterica* serovar *Typhimurium*. *Microbiology* 150:2037-2053.
- Kenney LJ, Bauer MD, Silhavy TJ. **1995**. Phosphorylation-dependent conformational changes in OmpR, an osmoregulatory DNA-binding protein of *Escherichia coli*. *Proc.Natl.Acad.Sci.U.S.A* 92:8866-8870.
- Kido M, Yamanaka K, Mitani T, Niki H, Ogura T, Hiraga S. **1996**. RNase E polypeptides lacking a carboxyl-terminal half suppress a *mukB* mutation in *Escherichia coli*. *J.Bacteriol.* 178:3917-3925.
- King T, Ishihama A, Kori A, Ferenci T. **2004**. A regulatory trade-off as a source of strain variation in the species *Escherichia coli*. *J.Bacteriol.* 186:5614-5620.
- Koch C, Kahmann R. **1986**. Purification and properties of the *Escherichia coli* host factor required for inversion of the G segment in bacteriophage Mu. *J.Biol.Chem.* 261:15673-15678.
- Kornberg RD, Lorch Y. **1999**. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98:285-294.
- Kostrewa D, Granzin J, Koch C, Choe HW, Raghunathan S, Wolf W, Labahn J, Kahmann R, Saenger W. **1991**. Three-dimensional structure of the *Escherichia coli* DNA-binding protein FIS. *Nature* 349:178-180.

- Kramlinger VM, Hiasa H. **2006**. The "GyrA-box" is required for the ability of DNA gyrase to wrap DNA and catalyze the supercoiling reaction. *J.Biol.Chem.* 281:3738-3742.
- Kreuzer KN, Cozzarelli NR. **1979**. *Escherichia coli* mutants thermosensitive for deoxyribonucleic acid gyrase subunit A: effects on deoxyribonucleic acid replication, transcription, and bacteriophage growth. *J.Bacteriol.* 140:424-435.
- Kruse T, Gerdes K. **2005**. Bacterial DNA segregation by the actin-like MreB protein. *Trends Cell Biol.* 15:343-345.
- Kusano S, Ding Q, Fujita N, Ishihama A. **1996**. Promoter selectivity of *Escherichia coli* RNA polymerase E sigma 70 and E sigma 38 holoenzymes. Effect of DNA supercoiling. *J.Biol.Chem.* 271:1998-2004.
- Laemmli UK. **1970**. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685.
- Lazar SW, Kolter R. **1996**. SurA assists the folding of *Escherichia coli* outer membrane proteins. *J.Bacteriol.* 178:1770-1773.
- Lazarus LR, Travers AA. **1993**. The *Escherichia coli* FIS protein is not required for the activation of *tyrT* transcription on entry into exponential growth. *EMBO J.* 12:2483-2494.
- Lelong C, Rabilloud T. **2003**. Prokaryotic proteomics. In: *Methods and Tools in Biosciences and Medicine: Prokaryotic Genomics*. Basel, Switzerland/Boston/Berlin: Birkhäuser Verlag. 157-171.
- Lenski RE. **2004**. Phenotypic and Genomic Evolution during a 20,000-Generation Experiment with the Bacterium *Escherichia coli*. *Plant Breed.Rev.* 24:225-265.
- Lenski RE, Rose M.R., Simpson S.C., Tadler S.C. **1991**. Long term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *The American naturalist* 138:1315-1341.
- Lenski RE, Travisano M. **1994**. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc.Natl.Acad.Sci.U.S.A* 91:6808-6814.
- Lenski RE, Winkworth CL, Riley MA. **2003**. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *J.Mol.Evol.* 56:498-508.
- Lesley SA, Jovanovich SB, Tse-Dinh YC, Burgess RR. **1990**. Identification of a heat shock promoter in the *topA* gene of *Escherichia coli*. *J.Bacteriol.* 172:6871-6874.
- Levine C, Hiasa H, Marians KJ. **1998**. DNA gyrase and topoisomerase IV: biochemical activities, physiological roles during chromosome replication, and drug sensitivities. *Biochim.Biophys.Acta* 1400:29-43.
- Lima CD, Wang JC, Mondragon A. **1994**. Three-dimensional structure of the 67K N-terminal fragment of *E. coli* DNA topoisomerase I. *Nature* 367:138-146.
- Lindow JC, Britton RA, Grossman AD. **2002**. Structural maintenance of chromosomes protein of *Bacillus subtilis* affects supercoiling *in vivo*. *J.Bacteriol.* 184:5317-5322.
- Link AJ, Phillips D, Church GM. **1997**. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *J.Bacteriol.* 179:6228-6237.
- Liu LF, Wang JC. **1987**. Supercoiling of the DNA template during transcription. *Proc.Natl.Acad.Sci.U.S.A* 84:7024-7027.
- Lockshon D, Morris DR. **1985**. Sites of reaction of *Escherichia coli* DNA gyrase on pBR322 *in vivo* as revealed by oxolinic acid-induced plasmid linearization. *J.Mol.Biol.* 181:63-74.
- Lopez CR, Yang S, Deibler RW, Ray SA, Pennington JM, DiGate RJ, Hastings PJ, Rosenberg SM, Zechiedrich EL. **2005**. A role for topoisomerase III in a recombination pathway alternative to RuvABC. *Mol.Microbiol.* 58:80-101.
- Losos JB, Jackman TR, Larson A, de Queiroz K, Rodriguez-Schettino L. **1988**. *Science* 279: 2115-2118.
- Lu M, Campbell JL, Boye E, Kleckner N. **1994**. SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* 77:413-426.
- Luijsterburg MS, Noom MC, Wuite GJ, Dame RT. **2006**. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *J.Struct.Biol.* 156:262-272.
- Lynch AS, Wang JC. **1993**. Anchoring of DNA to the bacterial cytoplasmic membrane through cotranscriptional synthesis of polypeptides encoding membrane proteins or proteins for export: a mechanism of plasmid hypernegative supercoiling in mutants deficient in DNA topoisomerase I. *J.Bacteriol.* 175:1645-1655.

- Malik M, Bensaid A, Rouviere-Yaniv J, Drlica K. **1996**. Histone-like protein HU and bacterial DNA topology: suppression of an HU deficiency by gyrase mutations. *J.Mol.Biol.* 256:66-76.
- Mallik P, Pratt TS, Beach MB, Bradley MD, Undamatla J, Osuna R. **2004**. Growth phase-dependent regulation and stringent control of *fis* are conserved processes in enteric bacteria and involve a single promoter (*fisP*) in *Escherichia coli*. *J.Bacteriol.* 186:122-135.
- Manché K, Notley-McRobb L, Ferenci T. **1999**. Mutational adaptation of *Escherichia coli* to glucose limitation involves distinct evolutionary pathways in aerobic and oxygen-limited environments. *Genetics* 153:5-12.
- Mangan MW, Lucchini S, Danino V, O'Croinin T, Hinton JC, Dorman CJ. **2006**. The integration host factor (IHF) integrates stationary-phase and virulence gene expression in *Salmonella enterica serovar Typhimurium*. *Mol.Microbiol.* 59:1831-1847.
- Masse E, Drolet M. **1999**. *Escherichia coli* DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. *J.Biol.Chem.* 274:16659-16664.
- Matoba K, Yamazoe M, Mayanagi K, Morikawa K, Hiraga S. **2005**. Comparison of MukB homodimer versus MukBEF complex molecular architectures by electron microscopy reveals a higher-order multimerization. *Biochem.Biophys.Res.Commun.* 333:694-702.
- Mattison K, Oropeza R, Byers N, Kenney LJ. **2002**. A phosphorylation site mutant of OmpR reveals different binding conformations at *ompF* and *ompC*. *J.Mol.Biol.* 315:497-511.
- Maxam AM, Gilbert W. **1977**. A new method for sequencing DNA. *Proc.Natl.Acad.Sci.U.S.A* 74:560-564.
- McCann MP, Kidwell JP, Matin A. **1991**. The putative sigma factor KatF has a central role in development of starvation-mediated general resistance in *Escherichia coli*. *J.Bacteriol.* 173:4188-4194.
- McLeod SM, Aiyar SE, Gourse RL, Johnson RC. **2002**. The C-terminal domains of the RNA polymerase alpha subunits: contact site with Fis and localization during co-activation with CRP at the *Escherichia coli proPP2* promoter. *J.Mol.Biol.* 316:517-529.
- Meddows TR, Savory AP, Grove JI, Moore T, Lloyd RG. **2005**. RecN protein and transcription factor DksA combine to promote faithful recombinational repair of DNA double-strand breaks. *Mol.Microbiol.* 57:97-110.
- Menzel R, Gellert M. **1983**. Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. *Cell* 34:105-113.
- Menzel R, Gellert M. **1987**. Modulation of transcription by DNA supercoiling: a deletion analysis of the *Escherichia coli gyrA* and *gyrB* promoters. *Proc.Natl.Acad.Sci.U.S.A* 84:4185-4189.
- Minsky A, Shimoni E, Frenkiel-Krispin D. **2002**. Stress, order and survival. *Nat.Rev.Mol.Cell Biol.* 3:50-60.
- Miyamoto S, Kashiwagi K, Ito K, Watanabe S, Igarashi K. **1993**. Estimation of polyamine distribution and polyamine stimulation of protein synthesis in *Escherichia coli*. *Arch.Biochem.Biophys.* 300:63-68.
- Mongold JA, Lenski RE. **1996**. Experimental rejection of a nonadaptive explanation for increased cell size in *Escherichia coli*. *J.Bacteriol.* 178:5333-5334.
- Morett E, Bork P. **1998**. Evolution of new protein function: recombinational enhancer Fis originated by horizontal gene transfer from the transcriptional regulator NtrC. *FEBS Lett.* 433:108-112.
- Morrison A, Cozzarelli NR. **1979**. Site-specific cleavage of DNA by *E. coli* DNA gyrase. *Cell* 17:175-184.
- Moulin L, Rahmouni AR, Boccard F. **2005**. Topological insulators inhibit diffusion of transcription-induced positive supercoils in the chromosome of *Escherichia coli*. *Mol.Microbiol.* 55:601-610.
- Muskhelishvili G, Buckle M, Heumann H, Kahmann R, Travers AA. **1997**. FIS activates sequential steps during transcription initiation at a stable RNA promoter. *EMBO J.* 16:3655-3665.
- Muskhelishvili G, Travers AA, Heumann H, Kahmann R. **1995**. FIS and RNA polymerase holoenzyme form a specific nucleoprotein complex at a stable RNA promoter. *EMBO J.* 14:1446-1452.
- Muskhelishvili G, Travers A. **2003**. Transcription factor as a topological homeostat. *Front Biosci.* 8:279-285.
- Nakanishi A, Oshida T, Matsushita T, Imajoh-Ohmi S, Ohnuki T. **1998**. Identification of DNA gyrase inhibitor (GyrI) in *Escherichia coli*. *J.Biol.Chem.* 273:1933-1938.
- Newlands JT, Josaitis CA, Ross W, Gourse RL. **1992**. Both Fis-dependent and factor-independent upstream activation of the *rnmB*P1 promoter are face of the helix dependent. *Nucleic Acids Res.* 20:719-726.
- Nichols MD, DeAngelis K, Keck JL, Berger JM. **1999**. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. *EMBO J.* 18:6177-6188.

- Nikaido H, Rosenberg EY. **1983**. Porin channels in *Escherichia coli*: studies with liposomes reconstituted from purified proteins. *J.Bacteriol.* 153:241-252.
- Niki H, Yamaichi Y, Hiraga S. **2000**. Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev.* 14:212-223.
- Ninnemann O, Koch C, Kahmann R. **1992**. The *Escherichia coli* *fis* promoter is subject to stringent control and autoregulation. *EMBO J.* 11:1075-1083.
- Norioka S, Ramakrishnan G, Ikenaka K, Inouye M. **1986**. Interaction of a transcriptional activator, OmpR, with reciprocally osmoregulated genes, *ompF* and *ompC*, of *Escherichia coli*. *J.Biol.Chem.* 261:17113-17119.
- Nosil P, Crespi BJ, Sandoval CP. **2002**. Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* 417:440-443.
- Nurse P, Levine C, Hassing H, Mariani KJ. **2003**. Topoisomerase III can serve as the cellular decatenase in *Escherichia coli*. *J.Biol.Chem.* 278:8653-8660.
- Oh TJ, Jung IL, Kim IG. **2001**. The *Escherichia coli* SOS gene *sbmC* is regulated by H-NS and RpoS during the SOS induction and stationary growth phase. *Biochem.Biophys.Res.Commun.* 288:1052-1058.
- Orphanides G, Maxwell A. **1994**. Evidence for a conformational change in the DNA gyrase-DNA complex from hydroxyl radical footprinting. *Nucleic Acids Res.* 22:1567-1575.
- Osheroff N. **1987**. Role of the divalent cation in topoisomerase II mediated reactions. *Biochemistry* 26:6402-6406.
- Osuna R, Finkel SE, Johnson RC. **1991**. Identification of two functional regions in Fis: the N-terminus is required to promote Hin-mediated DNA inversion but not lambda excision. *EMBO J.* 10:1593-1603.
- Owen-Hughes TA, Pavitt GD, Santos DS, Sidebotham JM, Hulton CS, Hinton JC, Higgins CF. **1992**. The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression. *Cell* 71:255-265.
- Owens RM, Pritchard G, Skipp P, Hodey M, Connell SR, Nierhaus KH, O'Connor CD. **2004**. A dedicated translation factor controls the synthesis of the global regulator Fis. *EMBO J.* 23:3375-3385.
- Pan CQ, Finkel SE, Cramton SE, Feng JA, Sigman DS, Johnson RC. **1996**. Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J.Mol.Biol.* 264:675-695.
- Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE, Blot M. **1999**. Genomic evolution during a 10,000-generation experiment with bacteria. *Proc.Natl.Acad.Sci.U.S.A* 96:3807-3812.
- Parekh BS, Sheridan SD, Hatfield GW. **1996**. Effects of integration host factor and DNA supercoiling on transcription from the *ilvPG* promoter of *Escherichia coli*. *J.Biol.Chem.* 271:20258-20264.
- Pato ML, Howe MM, Higgins NP. **1990**. 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* 87:8716-8720.
- Paul BJ, Barker MM, Ross W, Schneider DA, Webb C, Foster JW, Gourse RL. **2004**. DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* 118:311-322.
- Pedersen K, Gerdes K. **1999**. Multiple hok genes on the chromosome of *Escherichia coli*. *Mol.Microbiol.* 32:1090-1102.
- Pelosi L, Kuhn L, Guetta D, Garin J, Geiselmann J, Lenski RE, Schneider D. **2006**. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics* 173:1851-1869.
- Perederina A, Svetlov V, Vassilyeva MN, Tahirov TH, Yokoyama S, Artsimovitch I, Vassilyev DG. **2004**. Regulation through the secondary channel--structural framework for ppGpp-DksA synergism during transcription. *Cell* 118:297-309.
- Perry K, Mondragon A. **2003**. Structure of a complex between *E. coli* DNA topoisomerase I and single-stranded DNA. *Structure.* 11:1349-1358.
- Peter BJ, Arsuaga J, Breier AM, Khodursky AB, Brown PO, Cozzarelli NR. **2004**. Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biol.* 5:R87.
- Pettijohn DE. **1996**. The nucleoid. In: *Escherichia coli and Salmonella*. Washington DC: ASM Press. 158-166.
- Philippe, N. **2006**. Mécanismes moléculaires de l'adaptation au cours de 20 000 générations d'évolution expérimentale chez *Escherichia coli*, Doctorat de l'Université Joseph Fourier, Grenoble I.

- Ponce E, Flores N, Martinez A, Valle F, Bolivar F. **1995**. Cloning of the two pyruvate kinase isoenzyme structural genes from *Escherichia coli*: the relative roles of these enzymes in pyruvate biosynthesis. *J.Bacteriol.* 177:5719-5722.
- Postow L, Crisona NJ, Peter BJ, Hardy CD, Cozzarelli NR. **2001**. Topological challenges to DNA replication: conformations at the fork. *Proc.Natl.Acad.Sci.U.S.A* 98:8219-8226.
- Postow L, Hardy CD, Arsuaga J, Cozzarelli NR. **2004**. Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev.* 18:1766-1779.
- Pratt LA, Silhavy TJ. **1994**. OmpR mutants specifically defective for transcriptional activation. *J.Mol.Biol.* 243:579-594.
- Pratt LA, Silhavy TJ. **1995**. Identification of base pairs important for OmpR-DNA interaction. *Mol.Microbiol.* 17:565-573.
- Pratt TS, Steiner T, Feldman LS, Walker KA, Osuna R. **1997**. Deletion analysis of the *fis* promoter region in *Escherichia coli*: antagonistic effects of integration host factor and Fis. *J.Bacteriol.* 179:6367-6377.
- Pruss GJ, Manes SH, Drlica K. **1982**. *Escherichia coli* DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. *Cell*31:35-42.
- Pruss GJ, Drlica K. **1989**. DNA supercoiling and prokaryotic transcription. *Cell*56:521-523.
- Qi H, Menzel R, Tse-Dinh YC. **1997**. Regulation of *Escherichia coli topA* gene transcription: involvement of a sigmaS-dependent promoter. *J.Mol.Biol.* 267:481-489.
- Raji A, Zabel DJ, Laufer CS, Depew RE. **1985**. Genetic analysis of mutations that compensate for loss of *Escherichia coli* DNA topoisomerase I. *J.Bacteriol.* 162:1173-1179.
- Ramani N, Huang L, Freundlich M. **1992**. *In vitro* interactions of integration host factor with the *ompF* promoter-regulatory region of *Escherichia coli*. *Mol.Gen.Genet.* 231:248-255.
- Ramani N, Hedeshian M, Freundlich M. **1994**. *micF* antisense RNA has a major role in osmoregulation of OmpF in *Escherichia coli*. *J.Bacteriol.* 176:5005-5010.
- Ramstein J, Hervouet N, Coste F, Zelwer C, Oberto J, Castaing B. **2003**. Evidence of a thermal unfolding dimeric intermediate for the *Escherichia coli* histone-like HU proteins: thermodynamics and structure. *J.Mol.Biol.* 331:101-121.
- Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS. **2000**. Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* 406:64-67.
- Reizer J, Reizer A, Bairoch A, Saier MHJr. **1993**. A diverse transketolase family that includes the RecP protein of *Streptococcus pneumoniae*, a protein implicated in genetic recombination. *Res.Microbiol.* 144:341-347.
- Reyes-Dominguez Y, Contreras-Ferrat G, Ramirez-Santos J, Membrillo-Hernandez J, Gomez-Eichelmann MC. **2003**. Plasmid DNA supercoiling and gyrase activity in *Escherichia coli* wild-type and *rpoS* stationary-phase cells. *J.Bacteriol.* 185:1097-1100.
- Richardson SM, Higgins CF, Lilley DM. **1984**. The genetic control of DNA supercoiling in *Salmonella typhimurium*. *EMBO J.* 3:1745-1752.
- Robbe-Saule V, Algorta G, Rouilhac I, Norel F. **2003**. Characterization of the RpoS status of clinical isolates of *Salmonella enterica*. *Appl.Environ.Microbiol.* 69:4352-4358.
- Robey M, Benito A, Hutson RH, Pascual C, Park SF, Mackey BM. **2001**. Variation in resistance to high hydrostatic pressure and *rpoS* heterogeneity in natural isolates of *Escherichia coli* O157:H7. *Appl.Environ.Microbiol.* 67:4901-4907.
- Ross W, Thompson JF, Newlands JT, Gourse RL. **1990**. *Escherichia coli* Fis protein activates ribosomal RNA transcription *in vitro* and *in vivo*. *EMBO J.* 9:3733-3742.
- Rouviere-Yaniv J, Gros F. **1975**. Characterization of a novel, low-molecular-weight DNA-binding protein from *Escherichia coli*. *Proc.Natl.Acad.Sci.U.S.A* 72:3428-3432.
- Rowland SJ, Boocock MR, Stark WM. **2006**. DNA bending in the Sin recombination synapse: functional replacement of HU by IHF. *Mol.Microbiol.* 59:1730-1743.
- Rundle HD, Nagel L, Boughman JW, Schluter D. **2000**. *Science* 287:306-308.
- Ruthenburg AJ, Graybosch DM, Huetsch JC, Verdine GL. **2005**. A superhelical spiral in the *Escherichia coli* DNA gyrase A C-terminal domain imparts unidirectional supercoiling bias. *J.Biol.Chem.* 280:26177-26184.

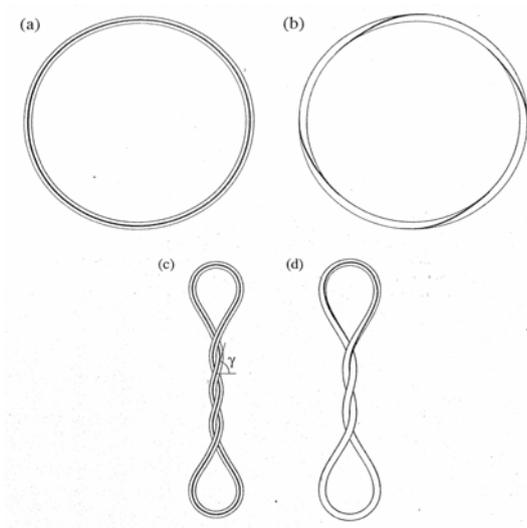
- Ryan VT, Grimwade JE, Camara JE, Crooke E, Leonard AC. **2004**. *Escherichia coli* prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. *Mol.Microbiol.* 51:1347-1359.
- Ryan VT, Grimwade JE, Nievera CJ, Leonard AC. **2002**. IHF and HU stimulate assembly of pre-replication complexes at *Escherichia coli oriC* by two different mechanisms. *Mol.Microbiol.* 46:113-124.
- Safo MK, Yang WZ, Corselli L, Cramton SE, Yuan HS, Johnson RC. **1997**. The transactivation region of the Fis protein that controls site-specific DNA inversion contains extended mobile beta-hairpin arms. *EMBO J.* 16:6860-6873.
- Sawitzke JA, Austin S. **2000**. Suppression of chromosome segregation defects of *Escherichia coli muk* mutants by mutations in topoisomerase I. *Proc.Natl.Acad.Sci.U.S.A* 97:1671-1676.
- Schechter LM, Jain S, Akbar S, Lee CA. **2003**. The small nucleoid-binding proteins H-NS, HU, and Fis affect *hilA* expression in *Salmonella enterica serovar Typhimurium*. *Infect.Immun.* 71:5432-5435.
- Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M. **2000a**. Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics* 156:477-488.
- Schneider D, Duperchy E, Depeyrot J, Coursange E, Lenski R, Blot M. **2002**. Genomic comparisons among *Escherichia coli* strains B, K-12, and O157:H7 using IS elements as molecular markers. *BMC.Microbiol.* 2:18.
- Schneider R, Travers A, Kutateladze T, Muskhelishvili G. **1999**. A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol.Microbiol.* 34:953-964.
- Schneider R, Travers A, Muskhelishvili G. **2000b**. The expression of the *Escherichia coli fis* gene is strongly dependent on the superhelical density of DNA. *Mol.Microbiol.* 38:167-175.
- Schneider R, Lurz R, Luder G, Tolksdorf C, Travers A, Muskhelishvili G. **2001**. An architectural role of the *Escherichia coli* chromatin protein FIS in organising DNA. *Nucleic Acids Res.* 29:5107-5114.
- Schoeffler AJ, Berger JM. **2005**. Recent advances in understanding structure-function relationships in the type II topoisomerase mechanism. *Biochem.Soc.Trans.* 33:1465-1470.
- Schultz SC, Shields GC, Steitz TA. **1991**. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* 253:1001-1007.
- Scott NW, Harwood CR. **1980**. Studies on the influence of the cyclic AMP system on major outer membrane proteins of *Escherichia coli* K12. *FEMS Microbiol.Lett.* 9:95-98.
- Seeto S, Notley-McRobb L, Ferenci T. **2004**. The multifactorial influences of RpoS, Mlc and cAMP on ptsG expression under glucose-limited and anaerobic conditions. *Res.Microbiol.* 155:211-215.
- Segre AV, Murray AW, Leu JY. **2006**. High-Resolution Mutation Mapping Reveals Parallel Experimental Evolution in Yeast. *PLoS.Biol.* 4:e256.
- Shaver AC, Dombrowski PG, Sweeney JY, Treis T, Zappala RM, Sniegowski PD. **2002**. Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. *Genetics* 162:557-566.
- Sheridan SD, Benham CJ, Hatfield GW. **1998**. Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced DNA duplex destabilization in an upstream activating sequence. *J.Biol.Chem.* 273:21298-21308.
- Sherratt DJ. **2003**. Bacterial chromosome dynamics. *Science* 301:780-785.
- Shindo H, Furubayashi A, Shimizu M, Miyake M, Imamoto F. **1992**. Preferential binding of *E.coli* histone-like protein HU alpha to negatively supercoiled DNA. *Nucleic Acids Res.* 20:1553-1558.
- Simons RW, Houman F, Kleckner N. **1987**. Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene* 53:85-96.
- Sinden RR, Pettijohn DE. **1981**. Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling. *Proc.Natl.Acad.Sci.U.S.A* 78:224-228.
- Slany RK, Kersten H. **1992**. The promoter of the *tgt/sec* operon in *Escherichia coli* is preceded by an upstream activation sequence that contains a high affinity FIS binding site. *Nucleic Acids Res.* 20:4193-4198.
- Sniegowski PD, Gerrish PJ, Lenski RE. **1997**. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387:703-705.
- Sokolsky TD, Baker TA. **2003**. DNA gyrase requirements distinguish the alternate pathways of Mu transposition. *Mol.Microbiol.* 47:397-409.

- Spassky A, Rimsky S, Garreau H, Buc H. **1984**. H1a, an *E. coli* DNA-binding protein which accumulates in stationary phase, strongly compacts DNA *in vitro*. *Nucleic Acids Res.* 12:5321-5340.
- Staczek P, Higgins NP. **1998**. Gyrase and Topo IV modulate chromosome domain size in vivo. *Mol.Microbiol.* 29:1435-1448.
- Stark WM, Boocock MR. **1995**. Gatecrashers at the catalytic party. *Trends Genet.* 11:121-123.
- Steck TR, Franco RJ, Wang JY, Drlica K. **1993**. Topoisomerase mutations affect the relative abundance of many *Escherichia coli* proteins. *Mol.Microbiol.* 10:473-481.
- Stein RA, Deng S, Higgins NP. **2005**. Measuring chromosome dynamics on different time scales using resolvases with varying half-lives. *Mol.Microbiol.* 56:1049-1061.
- Stewart L, Redinbo MR, Qiu X, Hol WG, Champoux JJ. **1998**. A model for the mechanism of human topoisomerase I. *Science* 279:1534-1541.
- Stivers JT, Harris TK, Mildvan AS. **1997**. Vaccinia DNA topoisomerase I: evidence supporting a free rotation mechanism for DNA supercoil relaxation. *Biochemistry* 36:5212-5222.
- Straney R, Krah R, Menzel R. **1994**. Mutations in the -10 TATAAT sequence of the *gyrA* promoter affect both promoter strength and sensitivity to DNA supercoiling. *J.Bacteriol.* 176:5999-6006.
- Stupina VA, Wang JC. **2005**. Viability of *Escherichia coli topA* mutants lacking DNA topoisomerase I. *J.Biol.Chem.* 280:355-360.
- Sucena E, Delon I, Jones I, Payre F, Stern DL. **2003**. Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature* 424:935-938.
- Suzuki T, Ueguchi C, Mizuno T. **1996**. H-NS regulates OmpF expression through *micF* antisense RNA in *Escherichia coli*. *J.Bacteriol.* 178:3650-3653.
- Talukder AA, Iwata A, Nishimura A, Ueda S, Ishihama A. **1999**. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J.Bacteriol.* 181:6361-6370.
- Thompson JF, Moitose d, V, Koch C, Kahmann R, Landy A. **1987**. Cellular factors couple recombination with growth phase: characterization of a new component in the lambda site-specific recombination pathway. *Cell*50:901-908.
- Tkachenko AG, Nesterova LY. **2003**. Polyamines as modulators of gene expression under oxidative stress in *Escherichia coli*. *Biochemistry (Mosc.)* 68:850-856.
- Travers A, Muskhelishvili G. **1998**. DNA microloops and microdomains: a general mechanism for transcription activation by torsional transmission. *J.Mol.Biol.* 279:1027-1043.
- Travers A, Muskhelishvili G. **2005**. Bacterial chromatin. *Curr.Opin.Genet.Dev.* 15:507-514.
- Tritz GJ, Chandler JL. **1973**. Recognition of a gene involved in the regulation of nicotinamide adenine dinucleotide biosynthesis. *J.Bacteriol.* 114:128-136.
- Tse-Dinh YC. **1985**. Regulation of the *Escherichia coli* DNA topoisomerase I gene by DNA supercoiling. *Nucleic Acids Res.* 13:4751-4763.
- Tse YC, Kirkegaard K, Wang JC. **1980**. Covalent bonds between protein and DNA. Formation of phosphotyrosine linkage between certain DNA topoisomerases and DNA. *J.Biol.Chem.* 255:5560-5565.
- Tupper AE, Owen-Hughes TA, Ussery DW, Santos DS, Ferguson DJ, Sidebotham JM, Hinton JC, Higgins CF. **1994**. The chromatin-associated protein H-NS alters DNA topology in vitro. *EMBO J.* 13:258-268.
- Udekwu KI, Darfeuille F, Vogel J, Reimegard J, Holmqvist E, Wagner EG. **2005**. Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes Dev.* 19:2355-2366.
- Ueda O, Yoshimura F. **2003**. Transposon-induced norfloxacin-sensitive mutants of *Bacteroides thetaiotaomicron*. *Microbiol.Immunol.* 47:17-25.
- Ussery D, Larsen TS, Wilkes KT, Friis C, Worning P, Krogh A, Brunak S. **2001**. Genome organisation and chromatin structure in *Escherichia coli*. *Biochimie* 83:201-212.
- Valens M, Penaud S, Rossignol M, Cornet F, Bocard F. **2004**. Macrodome organization of the *Escherichia coli* chromosome. *EMBO J.* 23:4330-4341.
- van Noort J, Verbrugge S, Goosen N, Dekker C, Dame RT. **2004**. Dual architectural roles of HU: formation of flexible hinges and rigid filaments. *Proc.Natl.Acad.Sci.U.S.A* 101:6969-6974.
- van Workum M, van Dooren SJ, Oldenburg N, Molenaar D, Jensen PR, Snoep JL, Westerhoff HV. **1996**. DNA supercoiling depends on the phosphorylation potential in *Escherichia coli*. *Mol.Microbiol.* 20:351-360.

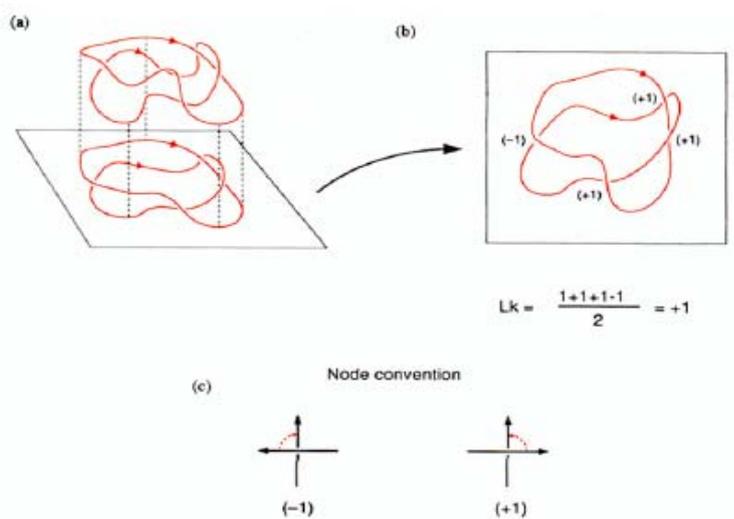
- Vasi F, Travisano M, Lenski RE. **1994**. Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. *Am.Nat.* 144:432-456.
- Walker KA, Atkins CL, Osuna R. **1999**. Functional determinants of the *Escherichia coli* *fis* promoter: roles of -35, -10, and transcription initiation regions in the response to stringent control and growth phase-dependent regulation. *J.Bacteriol.* 181:1269-1280.
- Walker KA, Mallik P, Pratt TS, Osuna R. **2004**. The *Escherichia coli* *fis* promoter is regulated by changes in the levels of its transcription initiation nucleotide CTP. *J.Biol.Chem.* 279:50818-50828.
- Wang Q, Mordukhova EA, Edwards AL, Rybenkov VV. **2006**. Chromosome condensation in the absence of the non-SMC subunits of MukBEF. *J.Bacteriol.* 188:4431-4441.
- Weinstein-Fischer D, Elgrably-Weiss M, Altuvia S. **2000**. *Escherichia coli* response to hydrogen peroxide: a role for DNA supercoiling, topoisomerase I and Fis. *Mol.Microbiol.* 35:1413-1420.
- Weitao T, Nordstrom K, Dasgupta S. **1999**. Mutual suppression of *mukB* and *seqA* phenotypes might arise from their opposing influences on the *Escherichia coli* nucleoid structure. *Mol.Microbiol.* 34:157-168.
- Weitao T, Nordstrom K, Dasgupta S. **2000**. *Escherichia coli* cell cycle control genes affect chromosome superhelicity. *EMBO Rep.* 1:494-499.
- West SC. **1997**. Processing of recombination intermediates by the RuvABC proteins. *Annu.Rev.Genet.* 31:213-244.
- Wijker JE, Jensen PR, Snoep JL, Vaz GA, Guiral M, Jongsma AP, de Waal A, Hoving S, van Dooren S, van der Weijden CC. **1995**. Energy, control and DNA structure in the living cell. *Biophys.Chem.* 55:153-165.
- Willenbrock H, Ussery DW. **2004**. Chromatin architecture and gene expression in *Escherichia coli*. *Genome Biol.* 5:252.
- Wilson RL, Libby SJ, Freet AM, Boddicker JD, Fahlen TF, Jones BD. **2001**. Fis, a DNA nucleoid-associated protein, is involved in *Salmonella typhimurium* SPI-1 invasion gene expression. *Mol.Microbiol.* 39:79-88.
- Wolf SG, Frenkiel D, Arad T, Finkel SE, Kolter R, Minsky A. **1999**. DNA protection by stress-induced biocrystallization. *Nature* 400:83-85.
- Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. **2006**. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc.Natl.Acad.Sci.U.S.A* 103:9107-9112.
- Worcel A, Burgi E. **1972**. On the structure of the folded chromosome of *Escherichia coli*. *J.Mol.Biol.* 71:127-147.
- Wu CW, Goldthwait DA. **1969**. Studies of nucleotide binding to the ribonucleic acid polymerase by equilibrium dialysis. *Biochemistry* 8:4450-4464.
- Xu J, Johnson RC. **1995a**. *aldB*, an RpoS-dependent gene in *Escherichia coli* encoding an aldehyde dehydrogenase that is repressed by Fis and activated by Crp. *J.Bacteriol.* 177:3166-3175.
- Xu J, Johnson RC. **1995b**. Identification of genes negatively regulated by Fis: Fis and RpoS comodulate growth-phase-dependent gene expression in *Escherichia coli*. *J.Bacteriol.* 177:938-947.
- Yamazoe M, Onogi T, Sunako Y, Niki H, Yamanaka K, Ichimura T, Hiraga S. **1999**. Complex formation of MukB, MukE and MukF proteins involved in chromosome partitioning in *Escherichia coli*. *EMBO J.* 18:5873-5884.
- Yang SW, Nash HA. **1995**. Comparison of protein binding to DNA *in vivo* and *in vitro*: defining an effective intracellular target. *EMBO J.* 14:6292-6300.
- Yang Y, Ames GF. **1988**. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. *Proc.Natl.Acad.Sci.U.S.A* 85:8850-8854.
- Yoshida M, Kashiwagi K, Shigemasa A, Taniguchi S, Yamamoto K, Makinoshima H, Ishihama A, Igarashi K. **2004**. A unifying model for the role of polyamines in bacterial cell growth, the polyamine modulon. *J.Biol.Chem.* 279:46008-46013.
- Yoshida T, Qin L, Egger LA, Inouye M. **2006**. Transcription regulation of *ompF* and *ompC* by a single transcription factor, OmpR. *J.Biol.Chem.* 281:17114-17123.
- Yuan HS, Finkel SE, Feng JA, Kaczor-Grzeskowiak M, Johnson RC, Dickerson RE. **1991**. The molecular structure of wild-type and a mutant Fis protein: relationship between mutational changes and recombinational enhancer function or DNA binding. *Proc.Natl.Acad.Sci.U.S.A* 88:9558-9562.

- Zacharias M, Goring HU, Wagner R. **1992**. Analysis of the Fis-dependent and Fis-independent transcription activation mechanisms of the *Escherichia coli* ribosomal RNA P1 promoter. *Biochemistry* 31:2621-2628.
- Zambrano MM, Siegle DA, Almiron M, Tormo A, Kolter R. **1993**. Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. *Science* 259:1757-1760.
- Zechiedrich EL, Khodursky AB, Cozzarelli NR. **1997**. Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in *Escherichia coli*. *Genes Dev.* 11:2580-2592.
- Zechiedrich EL, Khodursky AB, Bachellier S, Schneider R, Chen D, Lilley DM, Cozzarelli NR. **2000**. Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. *J.Biol.Chem.* 275:8103-8113.
- Zhang A, Rimsky S, Reaban ME, Buc H, Belfort M. **1996**. *Escherichia coli* protein analogs StpA and H-NS: regulatory loops, similar and disparate effects on nucleic acid dynamics. *EMBO J.* 15:1340-1349.
- Zhao G, Ceci P, Ilari A, Giangiacomo L, Laue TM, Chiancone E, Chasteen ND. **2002**. Iron and hydrogen peroxide detoxification properties of DNA-binding protein from starved cells. A ferritin-like DNA-binding protein of *Escherichia coli*. *J.Biol.Chem.* 277:27689-27696.
- Zhu Q, Pongpech P, DiGate RJ. **2001**. Type I topoisomerase activity is required for proper chromosomal segregation in *Escherichia coli*. *Proc.Natl.Acad.Sci.U.S.A* 98:9766-9771.
- Zinser ER, Kolter R. **1999**. Mutations enhancing amino acid catabolism confer a growth advantage in stationary phase. *J.Bacteriol.* 181:5800-5807.
- Zinser ER, Kolter R. **2000**. Prolonged stationary-phase incubation selects for lrp mutations in *Escherichia coli* K-12. *J.Bacteriol.* 182:4361-4365.

**A**



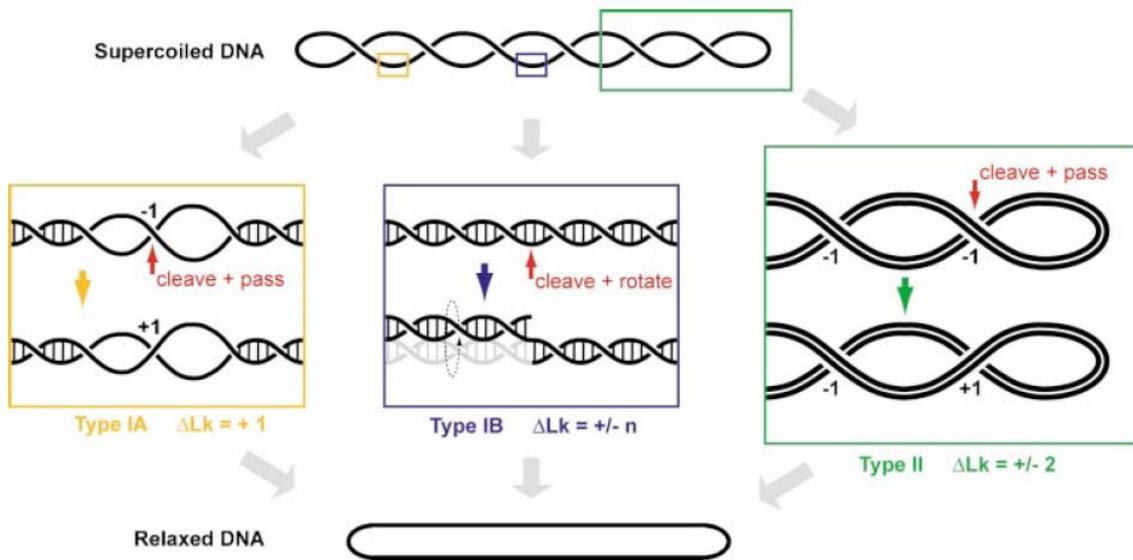
**B**



**Figure 0 : Les paramètres de définition de la superhélicité de l'ADN**

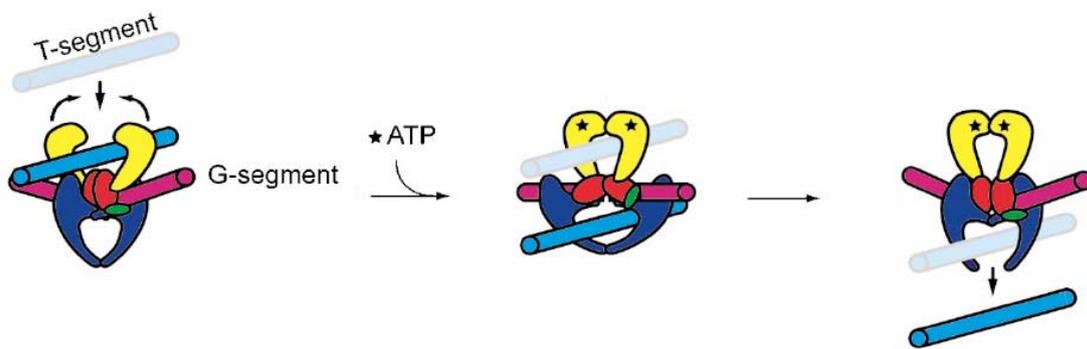
**A** : Exemple de superenroulement d'un plasmide. En (a) le plasmide est sous la forme la plus relâchée possible. En (b), un torsillement  $Tw$  y a été introduit ;  $\Delta Lk = \Delta Tw = -4$ , le plasmide étant maintenu dans un plan. En (c), le plasmide présente cette fois-ci un vrillage  $Wr$  ;  $\Delta Lk = \Delta Wr = -4$ . En (d), une conformation de moindre énergie, combinant les deux paramètres, est adoptée naturellement par le plasmide ;  $\Delta Lk = -4 = \Delta Tw + \Delta Wr \approx [-1] + [-3]$

**B** : Le nombre de liens  $Lk$  est calculé en projetant les courbes formées par un plasmide (a) (par exemple) sur un plan. (b) Un sens est attribué à chaque courbe, et le nombre de supertours est calculé en superposant une courbe sur l'autre et en donnant à chaque croisement une valeur de + ou -1, suivant la convention établie en (c).  $Lk$  est égal à la moitié de la somme des nombres de tours.



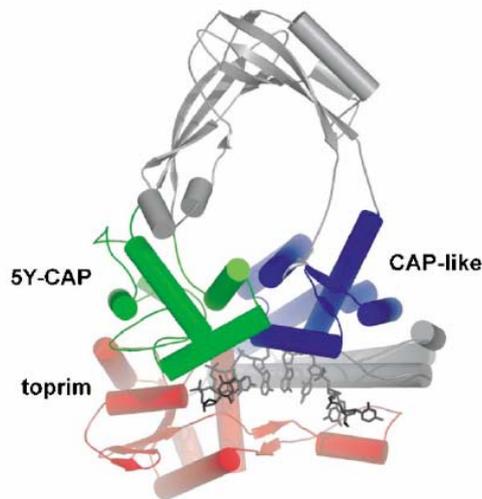
**Figure 1 : Activité des différentes Topos (Corbett, Berger, 2004)**

Les topoisomérases de type IA (en jaune) transfèrent un brin d'ADN à travers le second brin de l'hélice, modifiant le superenroulement (Lk) par étapes de 1. Les topoisomérases de type IB (en bleu) sont capables de laisser le brin d'ADN clivé effectuer une rotation autour du brin non clivé, modifiant Lk de plusieurs unités, positivement ou négativement. Les topoisomérases de type II (en vert) transfèrent une hélice à travers une autre hélice, changeant Lk par étapes de 2.



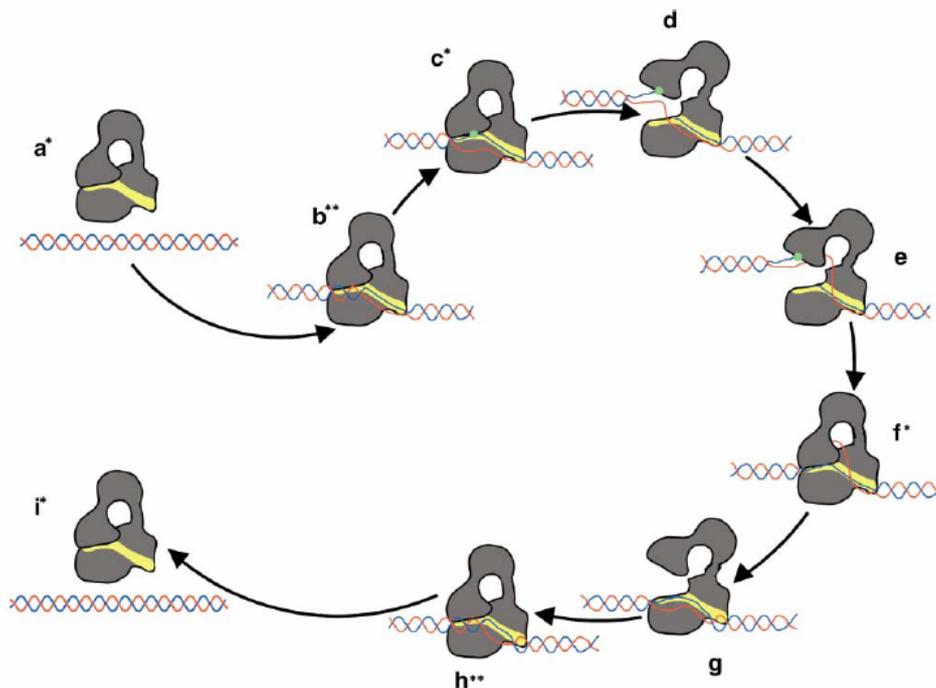
**Figure 2 : Mécanisme d'action d'une Topo de type IIA (Corbett, Berger, 2004)**

Les différents domaines sont colorés comme suit : en vert, le domaine 5Y-CAP, en rouge, le domaine toprim, en jaune, les domaines GHKL et transducteur, les hélices d'ADN étant matérialisées en rose (hélice G) et bleu (hélice T). Un premier segment d'ADN est d'abord lié (hélice G) puis l'entrée du deuxième entraîne la dimérisation des domaines GHKL. Les molécules d'ATP (étoiles) sont alors hydrolysées et le changement conformationnel qui s'en suit amène à la coupure de la première hélice d'ADN et au transfert de la seconde dans les « portes » (bleu foncé), qui s'ouvrent enfin pour libérer l'hélice T.



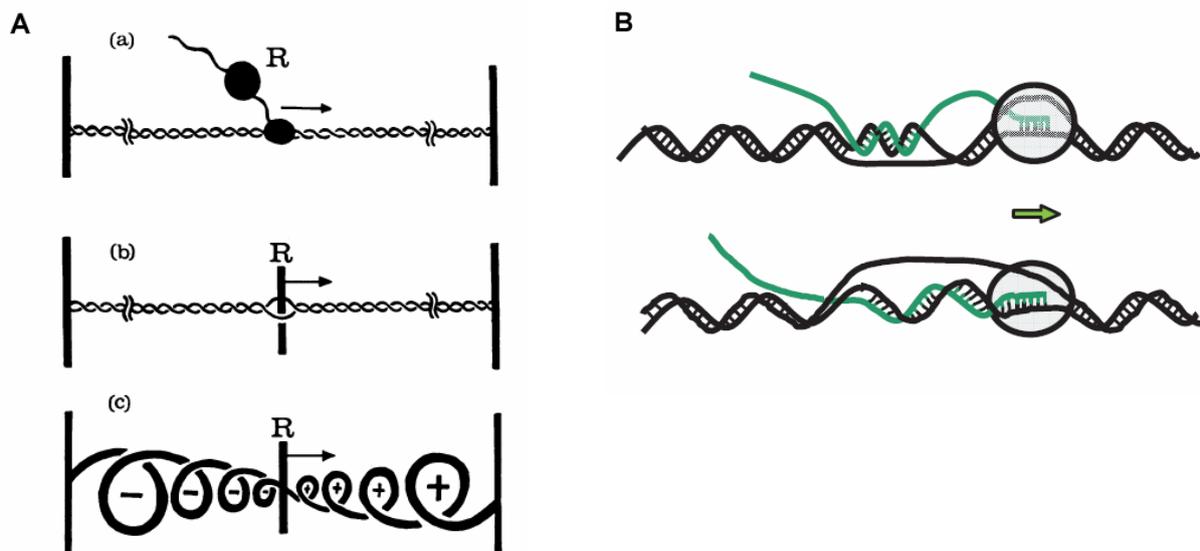
**Figure 3 : Structure d'une Topo de type IA (Corbett, Berger, 2004)**

La structure de la TopoIII d'*E. coli*, très similaire à celle de la TopoI, est montrée ici. Les différents domaines conférant leur activité aux topoisomérases de type IA sont colorés comme suit : en vert, le domaine 5Y-CAP ; en bleu, le domaine de type CAP, similaire au 5Y-CAP mais sans la tyrosine catalytique et en rouge le domaine toprim. L'enzyme adopte une forme toroïdale.



**Figure 4 : Mécanisme d'action de la TopoI (Perry, Mondragon, 2003)**

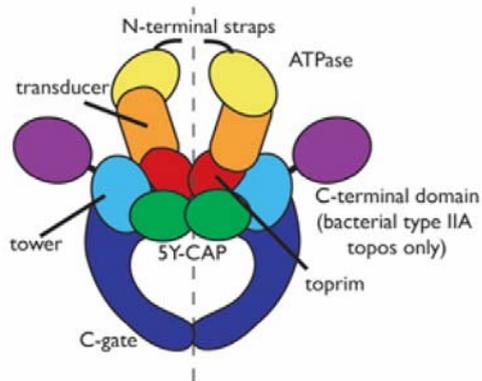
Les différentes étapes du relâchement de l'ADN par la TopoI sont montrées ici. L'enzyme est schématisée en gris avec le site d'interaction avec l'ADN en jaune et la double hélice d'ADN en bleu et rouge. Pour le détail des étapes, voir texte.



**Figure 5 : Modèle de Liu et Wang (1987) et boucles R (Drolet, 2006)**

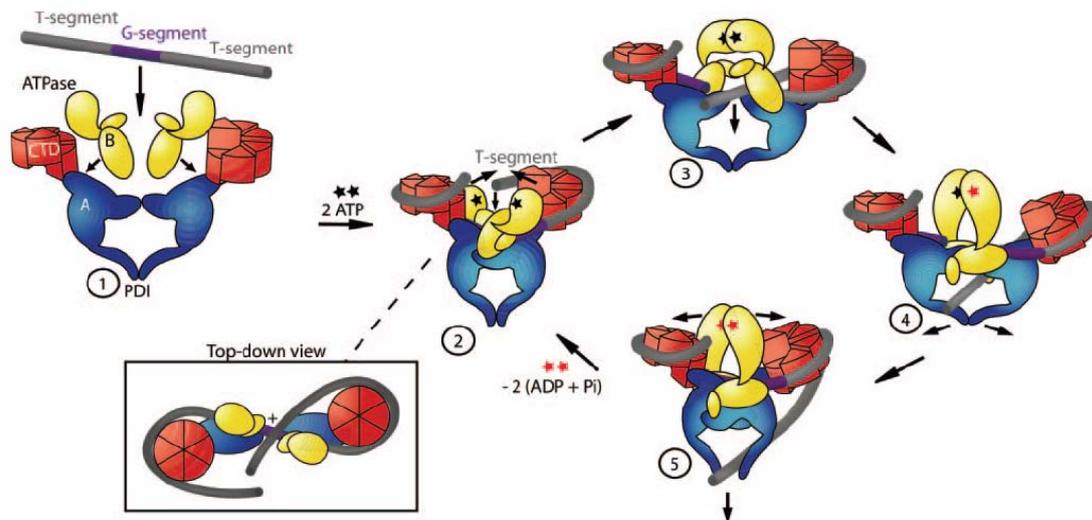
**A.** Modèle de Liu et Wang. En (a), une machinerie transcriptionnelle (R), incluant l'ARN polymérase, l'ARN synthétisé et des protéines liées à cet ARN, est représentée, progressant suivant la direction de la flèche. Les extrémités de l'ADN sont ancrées à une structure plus large schématisée par les barres verticales, représentant ainsi un domaine topologique. (b) Une autre vue de cette machinerie, schématisée comme un séparateur au sein de l'hélice d'ADN. (c) Quand R se déplace de gauche à droite, l'ADN devient superenroulé positivement en amont et superenroulé négativement en aval.

**B.** Schéma de formation de boucles R (Drolet, 2006). Les boucles R peuvent apparaître dans certaines conditions (ADN hyper enroulé, basse température). Ces boucles sont formées par l'hybridation de l'ARN nouvellement synthétisé avec l'ADN matrice, et empêche ainsi la réhybridation de la double hélice d'ADN. Ces structures peuvent soit empêcher la transcription en bloquant l'ARN polymérase, soit inhiber la traduction de nouvelles protéines par la dégradation de l'ARNm hybridé à l'ADN.



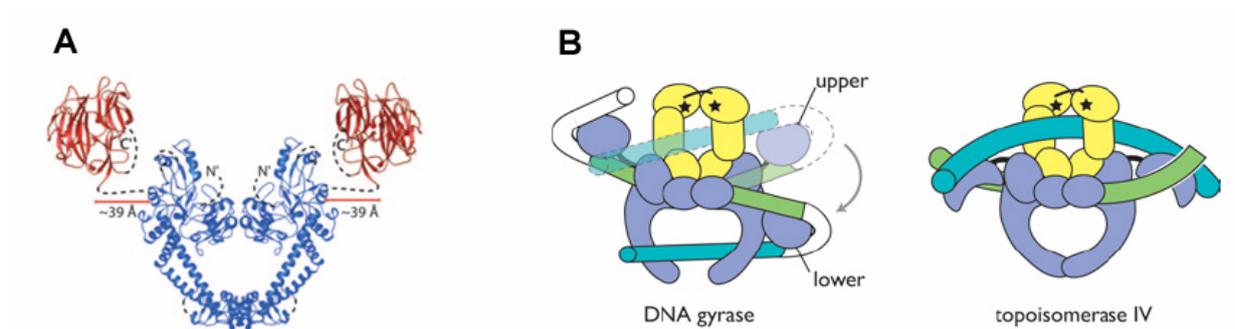
**Figure 6 : Les différents domaines des Topos de type IIA (Schoeffler, Berger, 2005)**

Les domaines GHKL formant l'ATPase sont montrés en jaune avec en orange les domaines transducteurs ; la « porte » est formée par deux domaines 5Y-CAP en vert, et les deux domaines toprim aidant au clivage de l'ADN en rouge. Les domaines formant les « C-gate » sont montrés en bleu foncé et les domaines C-terminaux en violet.



**Figure 7 : Mécanisme d'action des Topos de type II, ici de la gyrase (Ruthenburg et al., 2005)**

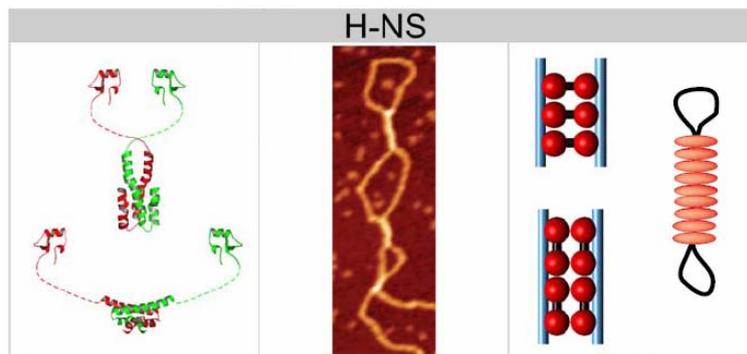
**(1)** Un segment d'ADN se lie à la sous-unité GyrB (en jaune). Les domaines C-terminaux des sous-unités GyrA (en rouge) le lient et l'enroulent de façon à présenter une des extrémités de la même hélice aux sous-unités GyrB (segment T) **(2, et vue de dessus)**. Elle est capturée par la dimérisation des domaines ATPase (étoile noire) en conformation de supertour positif, et son passage dans le domaine N-terminal des sous-unités GyrA (en bleu) induit la formation de 2 supertours négatifs **(3)**. Le segment G est ensuite libéré en même temps que deux molécules d'ADP et de Pi.



**Figure 8 : Rôle de la sous-unité GyrA de la gyrase d'*E. coli* dans la modification de la superhélicité de l'ADN**

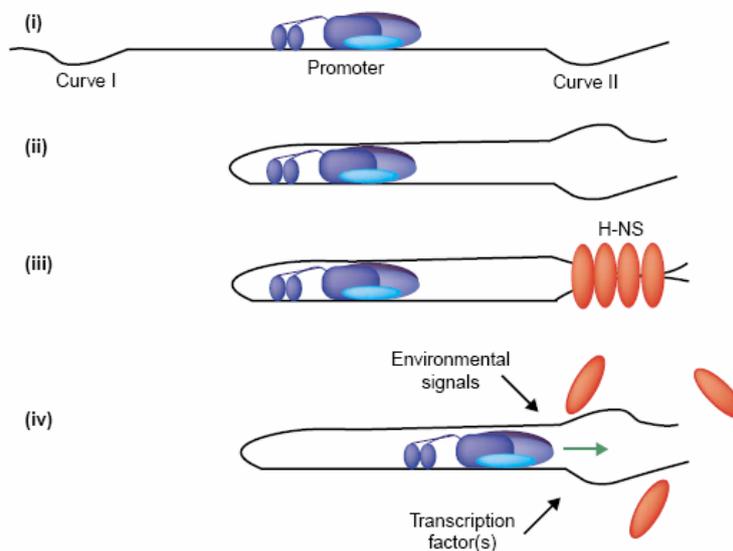
**A.** Les domaines N-terminaux d'un dimère de sous-unités GyrA [résidus 30 à 522] sont montrés ici en bleu, et les domaines C-terminaux en rouge [résidus 535 à 841]. La distance maximale séparant les domaines NTD et CTD est de 39 Å ; cette distance pourrait permettre la mobilité des extrémités C-terminales, entraînant le transfert de l'hélice T et la diminution du paramètre Lk par incréments de 2. [Ruthenburg et al., 2005]

**B.** Différence entre les CTD de GyrA de la gyrase (à gauche, en bleu) et les CTD de ParC de la TopoIV (à droite, en bleu). Les sous-unités GyrB et ParE sont représentées en jaune, à gauche et à droite, respectivement. En ce qui concerne la gyrase (à gauche), l'ADN est enroulé autour des CTD de GyrA, et un mouvement de l'un deux provoquerait le transfert de la double hélice T dans la cavité formée par les portes C. La fixation de l'ADN autour des CTD de ParC (à droite) serait moins forte, ou bien leur mouvement serait moins coordonné. Ils présentent néanmoins une double hélice courbée dont la conformation est prévalente dans les caténanes, qui forment le substrat majeur de la TopoIV [Schoeffler, Berger, 2005].



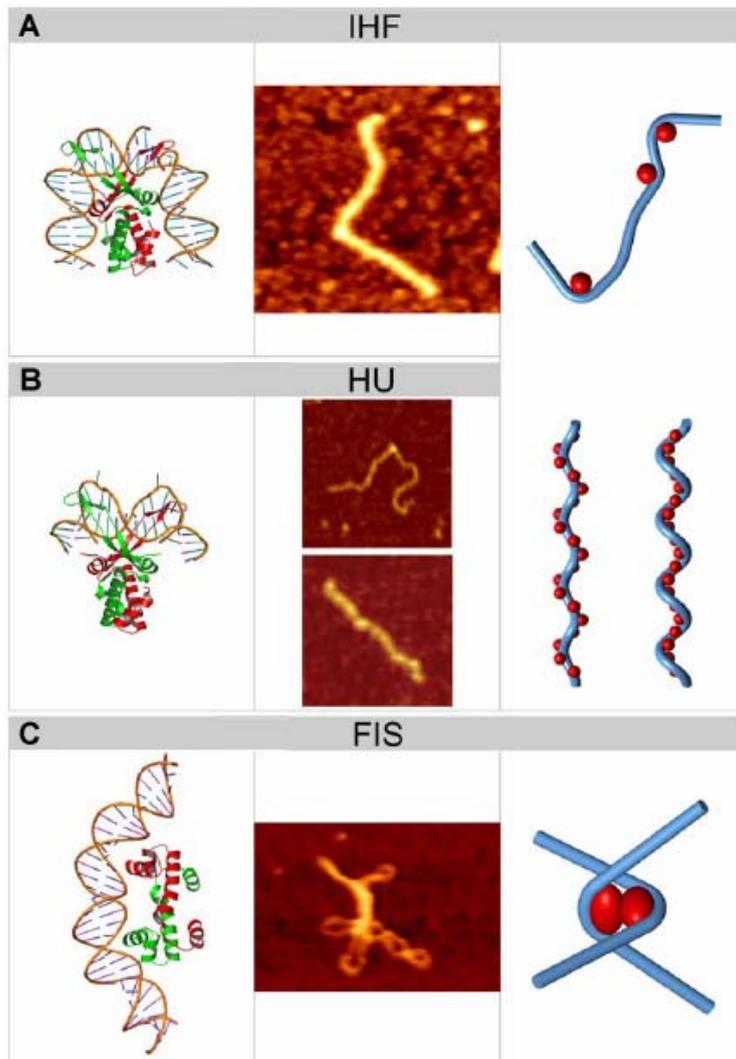
**Figure 9 : Effets sur l'ADN de la protéine H-NS (Luijsterburg et al., 2006)**

A gauche, modèles représentant les dimères de H-NS sous leur forme active. Au centre, image par microscopie à force atomique montrant un plasmide incubé avec la protéine H-NS. A droite, différentes structures formées par H-NS fixée à l'ADN.



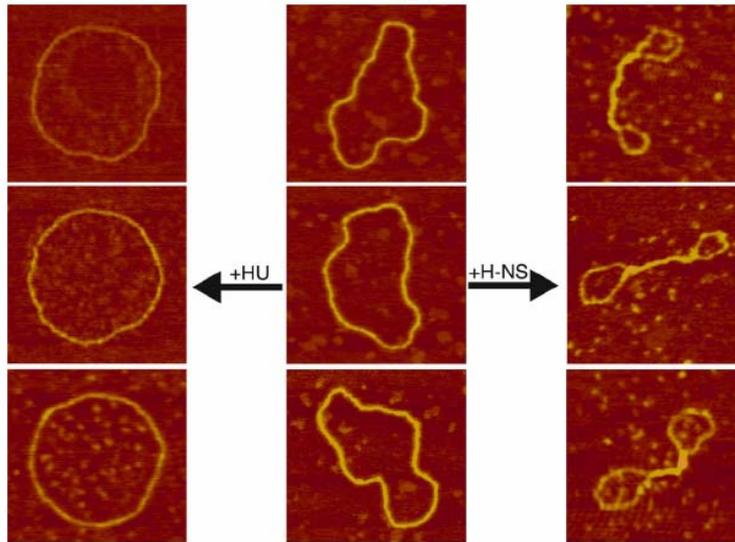
**Figure 10 : Mécanisme de répression de la transcription par H-NS (Dorman, Deighan, 2003 ; Dame et al., 2002)**

L'ARN polymérase est représentée en bleu et H-NS en rouge. **(i)** Liaison de l'ARN polymérase sur un promoteur flanqué de régions courbées. **(ii)** Enroulement de l'ADN autour de l'ARN polymérase, ce qui rapproche les courbures intrinsèques de l'ADN. **(iii)** La fixation de plusieurs molécules de H-NS sur ses sites entraîne le blocage de l'ARN polymérase. **(iv)** Le complexe formé par H-NS est séparé par un changement de l'environnement ou l'intervention de facteurs de transcription.



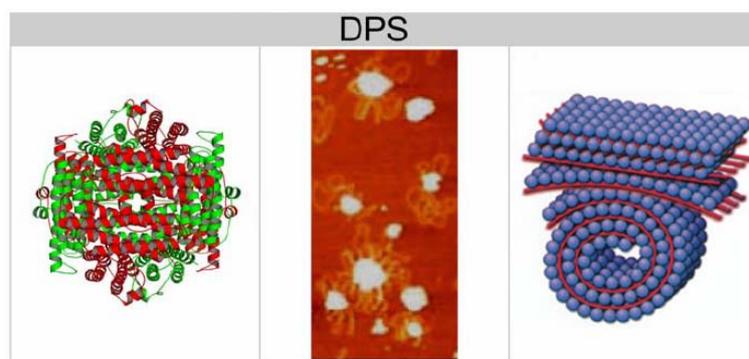
**Figure 11 : Effet sur l'ADN des protéines courbant l'ADN (Luijsterburg et al., 2006)**

**A**, IHF ; **B**, HU ; **C**, Fis. A gauche, modèles représentant les oligomères liés à l'ADN. Au centre, image par microscopie à force atomique, montrant un plasmide incubé avec la protéine. A droite, modèles de liaison des protéines (en rouge) à deux hélices d'ADN (en bleu). Pour HU, on observe 2 formes de fixation à l'ADN : en faible concentration, HU courbe l'ADN ; à hautes concentrations, il forme des filaments rigides.



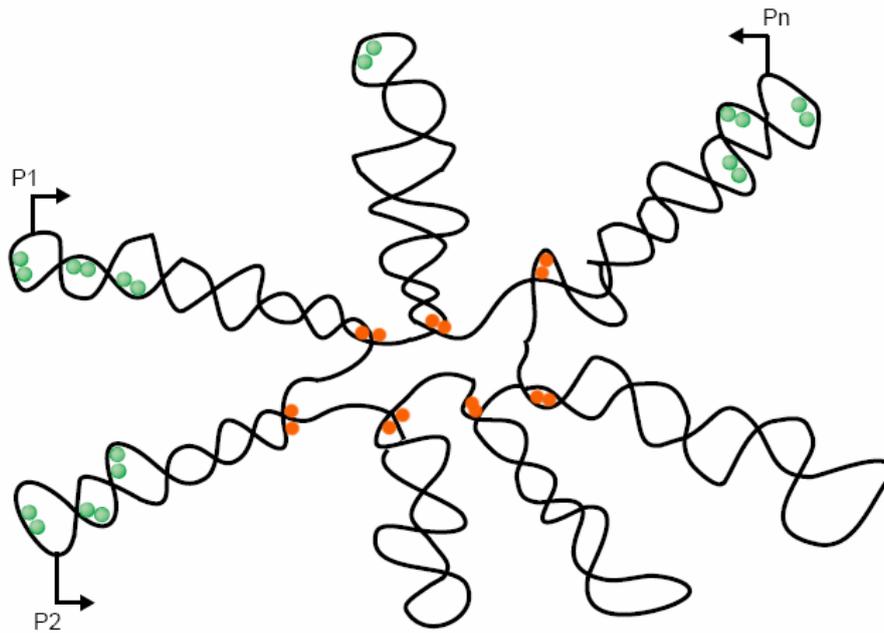
**Figure 12 : Effet antagoniste de HU et H-NS sur un plasmide (Dame, Goosen, 2002)**

A gauche, le plasmide pUC19 est incubé avec HU [1 dimère/9pb] ; au centre, le plasmide pUC19 relâché ; à droite, le plasmide pUC19 incubé avec H-NS [1 dimère/12pb].



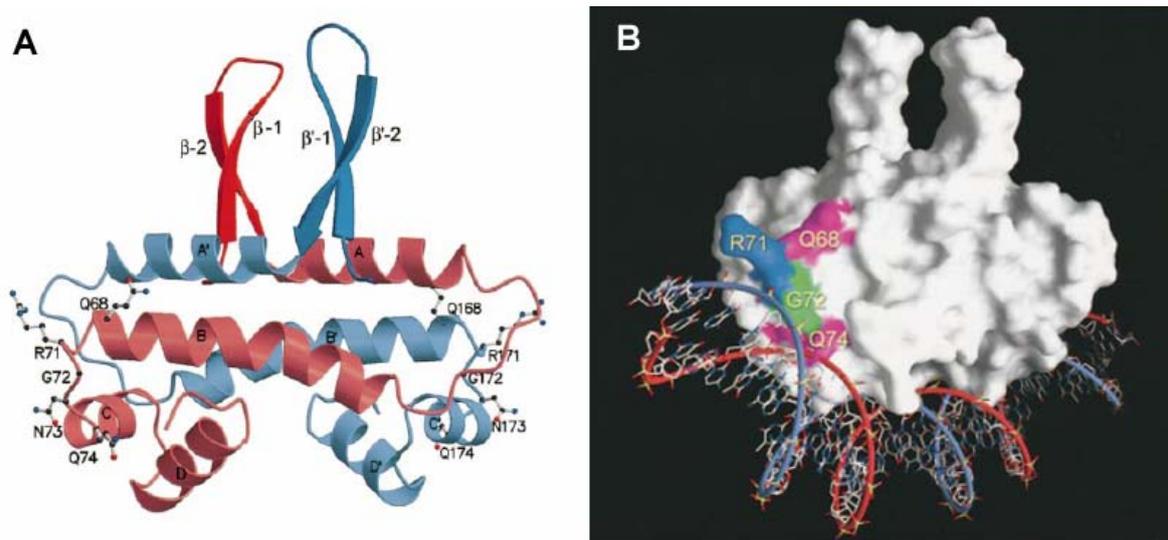
**Figure 13 : Protection de l'ADN par Dps (Luijsterburg et al., 2006)**

A gauche, modèle représentant les 12 monomères formant une sphère. Au centre, image par microscopie à force atomique montrant de l'ADN incubé avec Dps. A droite, modèle de protection de l'ADN (en rouge) par des dodécamères associés entre eux (en bleu) (Frenkiel-Krispin et al., 2004).



**Figure 14 : Effet de la protéine Fis sur la structure du nucléoïde et la transcription (Dorman, Deighan, 2003)**

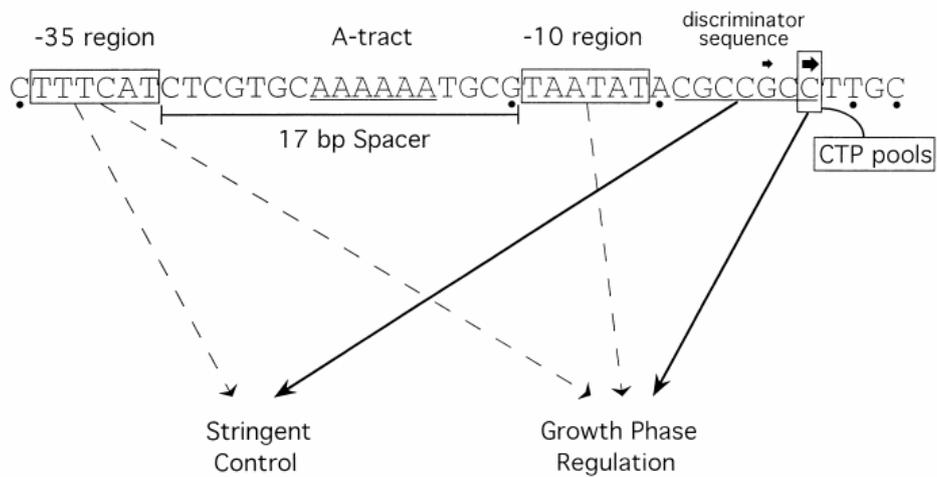
Fis agit sur l'ADN à deux niveaux : global (sphères oranges), en favorisant la formation de domaines topologiques et de boucles, et local (sphères vertes), en courbant l'ADN au niveau des promoteurs, ce qui favorise la transcription par interaction avec l'ARN polymérase.



**Figure 15 : Structure d'un dimère de Fis (Cheng et al., 2000)**

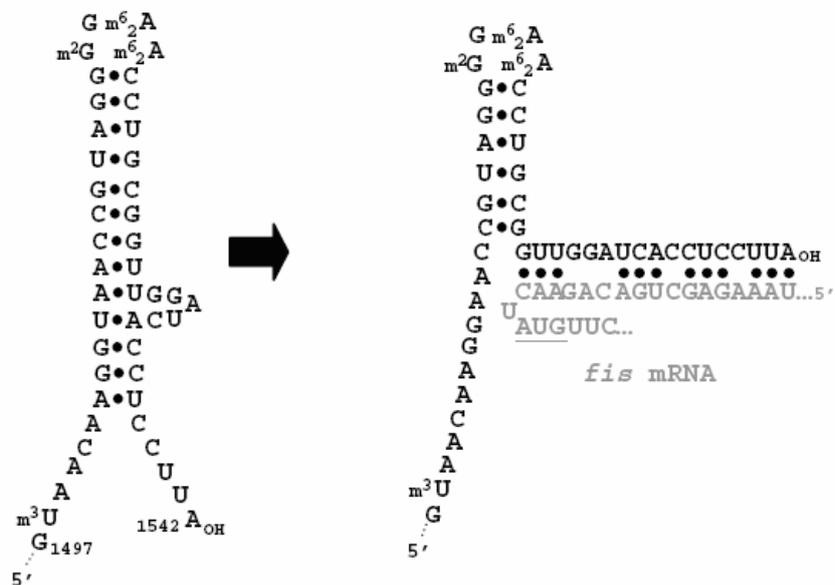
**A.** Modèle d'un dimère de Fis. Les structures des cristaux obtenus à partir de la protéine Fis sauvage (résidus 26 à 98) et d'un mutant K36E (résidus 10 à 25) ont été combinées pour donner un modèle plus complet. Les résidus 26 à 98 (bleu et rouge plus clairs) forment 4 hélices A à D (résidus 26 à 98) et A' à D' (résidus 126 à 198). Les résidus 10 à 25 forment deux feuillets  $\beta$  arrangés en épingle à cheveux. Les résidus 68 à 74 montrés sont ceux impliqués dans l'interaction avec l'ARN polymérase sur les promoteurs *rnnB* P1 ou *proPP2*.

**B.** Modèle représentant la surface de Fis en interaction avec l'ADN, courbé. Les résidus 68 à 74 sont montrés en couleur.



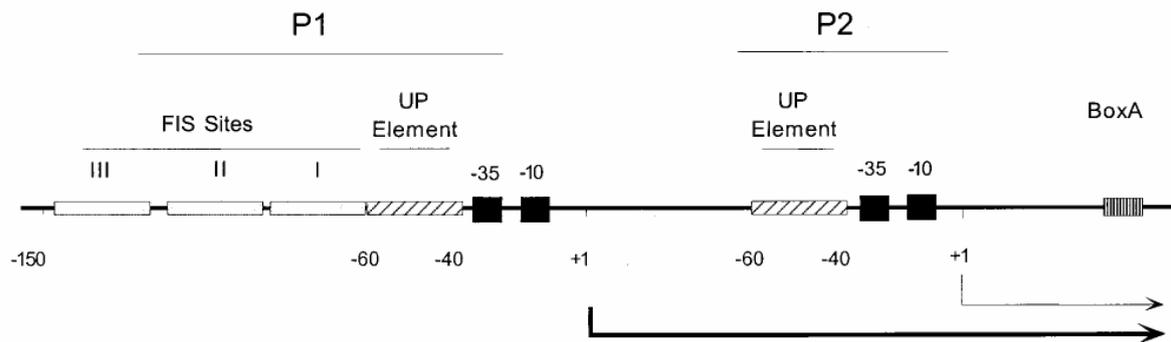
**Figure 16 : Régulation du promoteur de *fis* (Walker et al., 1999)**

Les séquences encadrées sont les régions -35, -10 et le site majeur d'initiation de la transcription, au dessus duquel une flèche en gras souligne la force de ce site. La séquence discriminatrice riche en GC et la région riche en A sont soulignées ; les bases marquées d'un point sont importantes pour l'activité du promoteur. Les flèches pleines indiquent les régions nécessaires à la régulation par la réponse stringente et au cours de la phase de croissance ; celles en pointillés indiquent des régions avec des effets mineurs dans l'un ou l'autre des processus. La connexion avec la concentration en CTP est aussi mise en valeur.



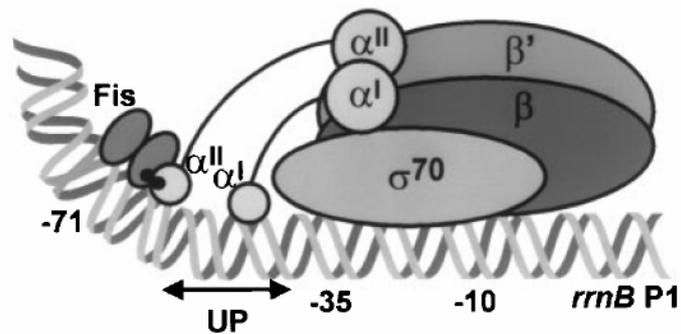
**Figure 17 : Complémentarité entre l'ARNr 16S et l'extrémité 5' non traduite de l'ARNm *fis* (Owens et al., 2004)**

Les conformations prédites de l'extrémité 3' de l'ARNr 16S sont représentées en absence ou en présence de l'ARNm de *fis* (en gris). Le codon d'initiation de la traduction est souligné. Une interaction particulièrement forte existe entre l'extrémité 3' de l'ARNr 16S et l'ARNm de *fis*. Celle-ci empêcherait, en absence de BipA, le démarrage de la traduction.



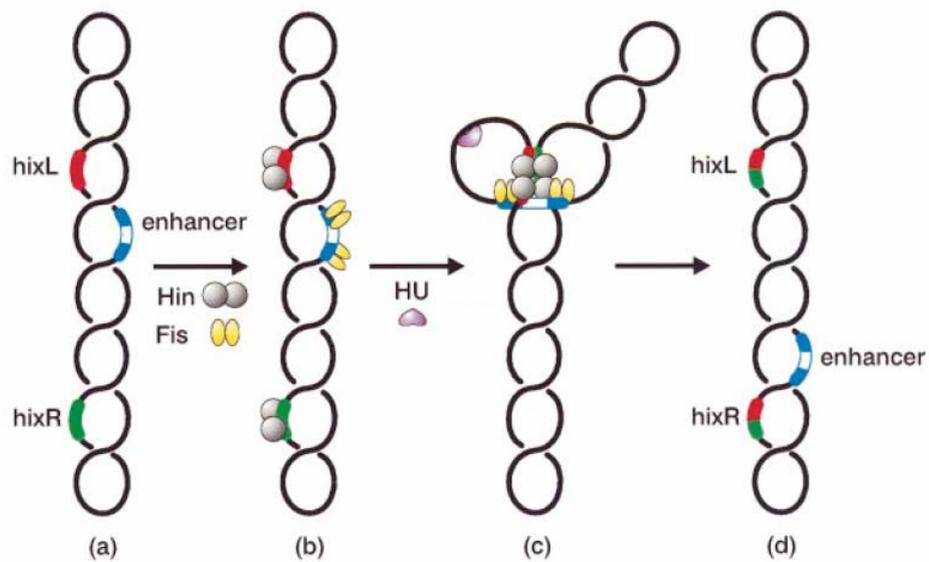
**Figure 18 : Région promotrice de *rrnB* (Bartlett et al., 2000)**

Pour chacun des deux promoteurs, le +1 de transcription, les boîtes -10 et -35, les séquences en amont (UP elements) et les sites de fixation de Fis sont indiqués.



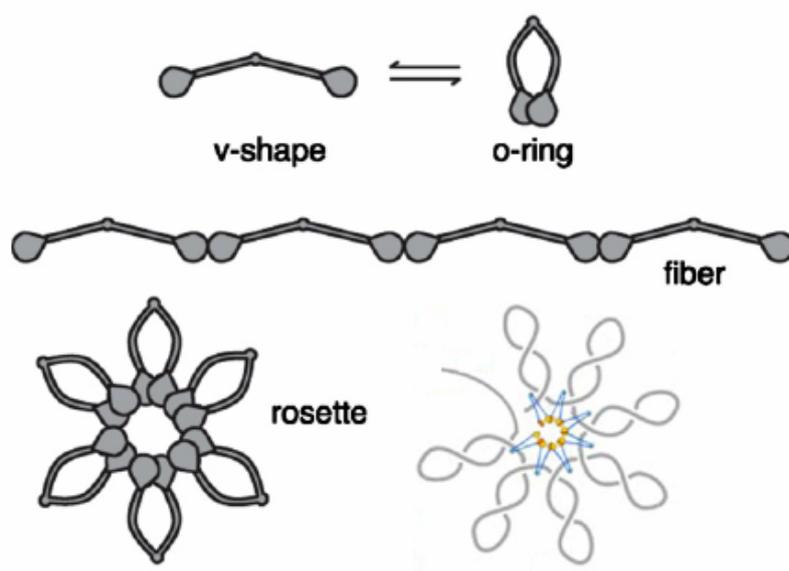
**Figure 19 : Modèle d'interaction de Fis avec l'ARN polymérase sur le promoteur *rrnBP1* (McLeod et al., 2002)**

La transcription est majoritairement activée par la fixation de Fis sur le site I, centré à -71. Les deux CTD des sous-unités  $\alpha$  se lient à l'ADN, l'un en contact avec Fis et l'autre sur l'élément en amont du promoteur (UP).



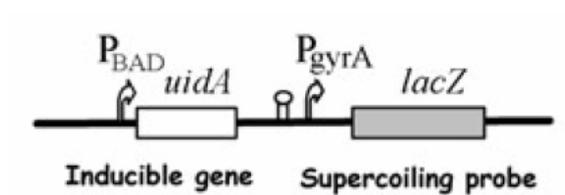
**Figure 20 : Système de recombinaison par Hin (Safo et al., 1997)**

Hin (en gris) se fixe aux sites de recombinaison *hixL* et *hixR* (a), et Fis (en jaune) se lie aux deux domaines « enhancer » sur un substrat superenroulé (b). Le complexe d'inversion est assemblé avec l'aide de HU (c). En d, le fragment d'ADN entre *hixL* et *hixR* a été inversé.



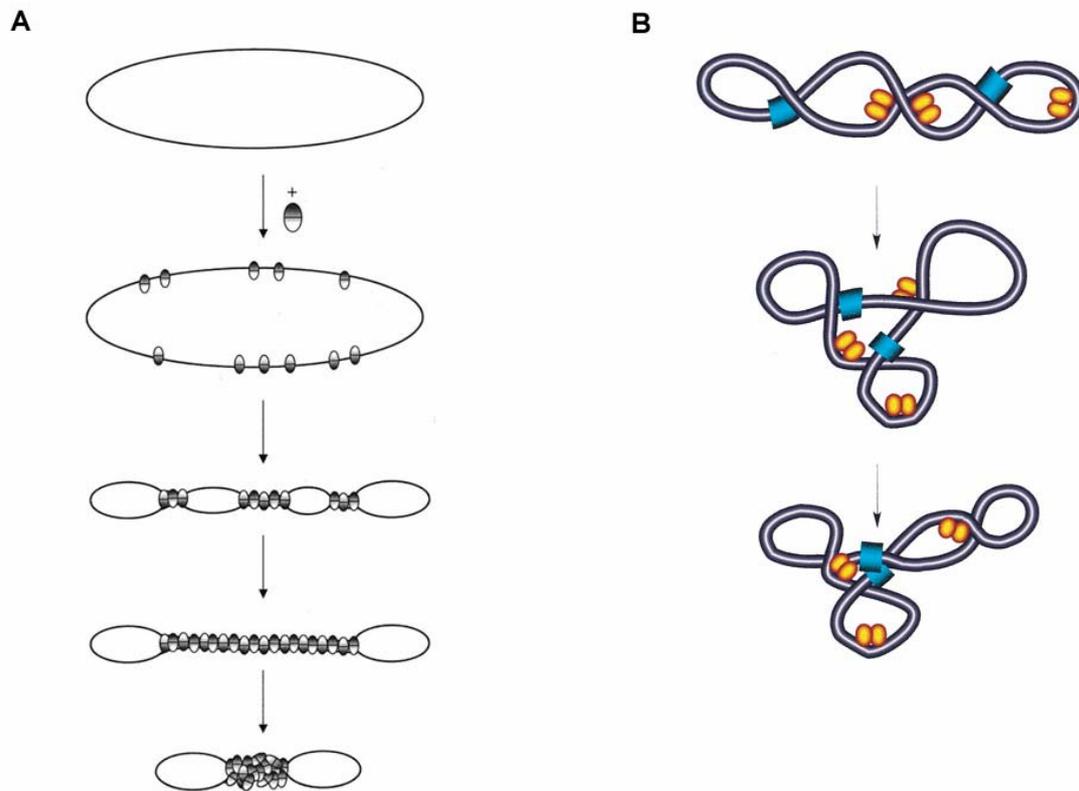
**Figure 21 : Types de structures formées par MukB (Matoba et al., 2005)**

En haut, les deux formes prises par le dimère (forme en V ou anneau) sont montrées. Au milieu, la formation d'une fibre intermoléculaire, et en bas une structure en rosette sont schématisées. Un modèle de compaction de l'ADN (en gris) par les protéines de type SMC (en bleu et jaune), est montré en bas à droite. L'ADN, comme les protéines SMC, prend alors une structure de type rosette.



**Figure 22 : Système rapporteur sensible à la superhélicité de l'ADN (Moulin et al., 2005)**

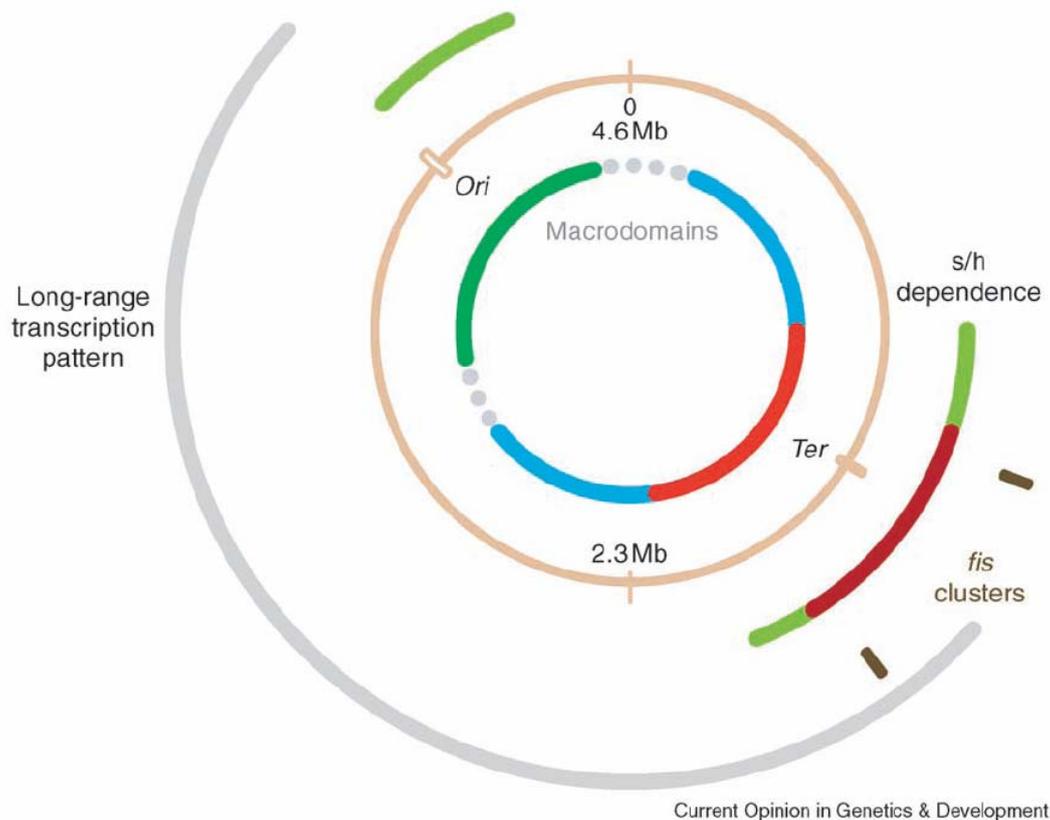
Représentation de la cassette topologique utilisée dans cette étude. L'« inducteur de superenroulement », formé d'un promoteur inductible et du gène rapporteur *uidA* est situé en amont du « rapporteur de superhélicité », formé d'un promoteur sensible au superenroulement, *gyrA*, fusionné à un second gène rapporteur, *lacZ*. Les deux éléments sont séparés par un terminateur fort de la transcription (épingle à cheveux).



**Figure 23 : Modèle de compaction du chromosome par H-NS et Fis**

**A.** Modèle de condensation de l'ADN par H-NS [Dame et al., 2000]. Des dimères d'H-NS (formes ovales) se fixent et polymérisent sur une molécule d'ADN. Quand deux brins d'ADN portant des molécules d'H-NS sont suffisamment proches, les protéines peuvent interagir avec l'autre brin d'ADN, ou entre elles (voir structure de la **Figure 9**), formant ainsi un pontage. Cette réaction est coopérative, la formation des premiers pontages entraînant de nouvelles interactions dans les « bulles » laissées libres. L'oligomérisation de H-NS crée un second niveau de condensation (en bas).

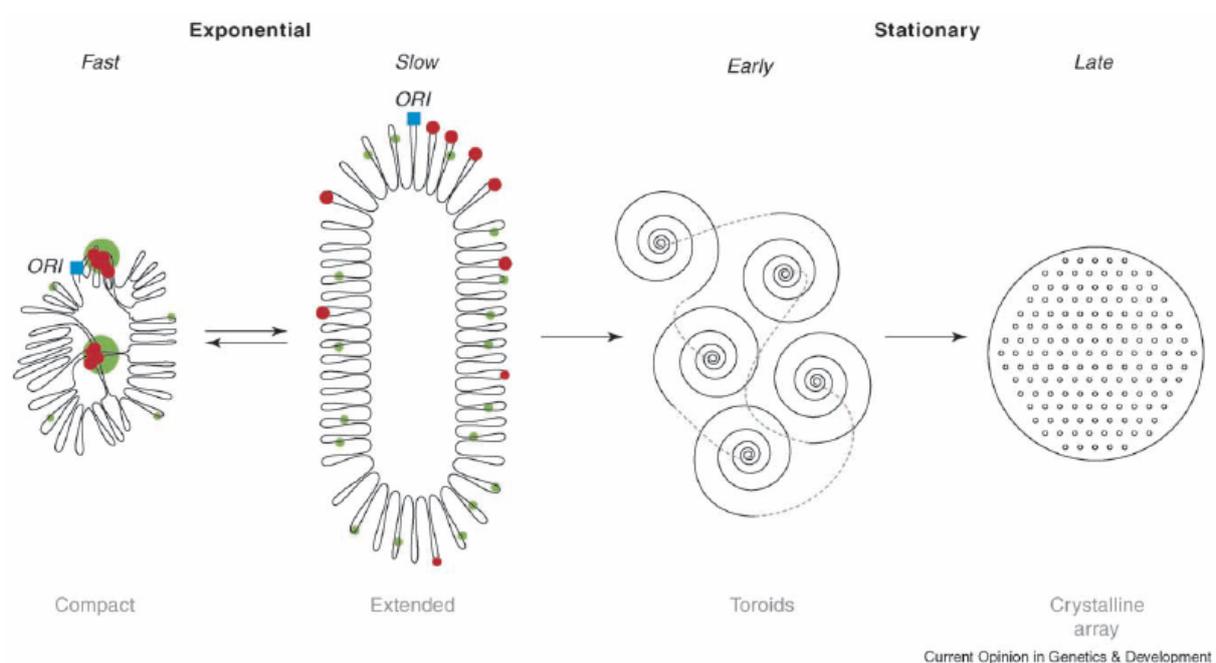
**B.** Modèle de réarrangement de l'ADN par Fis [Schneider et al., 2001]. La fixation et la courbure de l'ADN par des dimères de Fis (en jaune) facilitent les structures formant des branches et des boucles dans l'ADN. Notons que cette réorganisation permet de rapprocher des sites à priori éloignés (en bleu).



**Figure 24 : Structure des macrodomaines d'*E. coli* et leur réponse à la topologie de l'ADN (Travers, Muskhelishvili, 2005)**

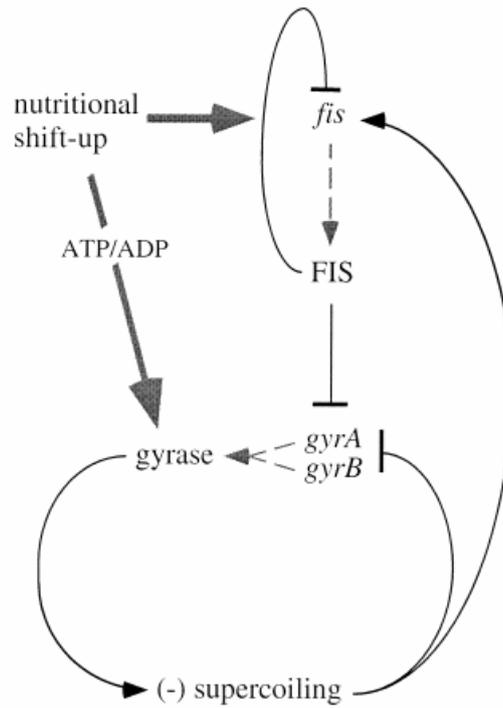
Les macrodomaines [cercle interne] sont représentés d'après Valens *et al.* (2004) : en vert, le macrodomaine entourant *oriC* [*Ori*, l'origine de réplication], en rouge, celui entourant *ter* [*Ter*, le terminus de réplication], en bleu les autres domaines cartographiés, séparés du domaine de *oriC* par des régions moins définies, en pointillés.

Les régions de gènes dont la transcription répond globalement à la superhélicité sont montrées en dehors du chromosome [s/h dependence]. En vert, les gènes répondant à un superenroulement négatif ; en rouge, ceux répondant à un relâchement. Une région riche en sites de fixation de Fis proche de *ter* est montrée [*fis* clusters], et celle d'activité transcriptionnelle intense en période de croissance rapide, en gris [Jeong *et al.*, 2004].



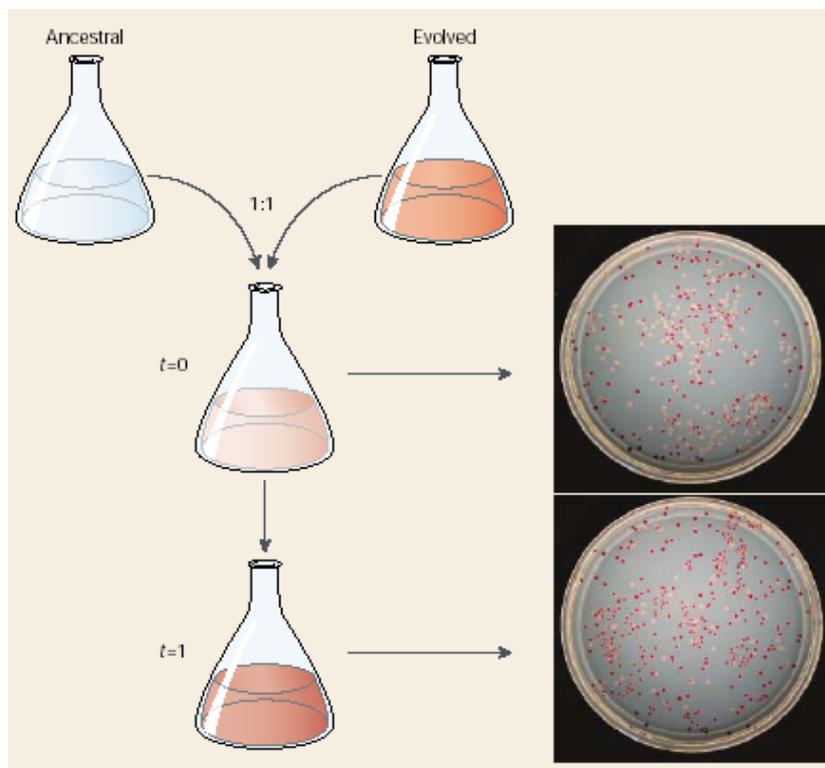
**Figure 25: Réorganisation du chromosome en fonction des conditions de croissance (Travers, Muskhelishvili, 2005)**

Les structures du chromosome formées dans des cellules en phase exponentielle de croissance indiquent que l'organisation des boucles et des sites de transcription dépend fortement du taux de croissance de la cellule. Dans ces conditions de croissance, la structure globale est une forme circulaire ouverte mais des connexions intramoléculaires sont possibles, et celles-ci dépendent fortement de l'environnement dans lequel se trouve la cellule. Les structures formées en phase stationnaire sont basées sur une visualisation en microscopie électronique. En début de phase stationnaire, on observe des structures toroidales, qui deviennent cristallines en phase stationnaire prolongée, notamment sous l'action de Dps (voir paragraphe I.B.1).



**Figure 26 : Boucle de régulation homéostatique de la superhélicité de l'ADN (Schneider et al., 2000b)**

Les effets positifs sont représentés par des flèches, les inhibitions par des lignes terminées par un trait perpendiculaire. Deux boucles d'autorégulation sont ainsi reliées. L'ajout de nutriments pourrait activer directement la gyrase en augmentant le rapport [ATP]/[ADP]. La transcription de *fis* est ainsi activée par l'augmentation de superhélicité, de la concentration en CTP et par la levée du contrôle stringente sur *fisP*. Voir texte pour les détails.

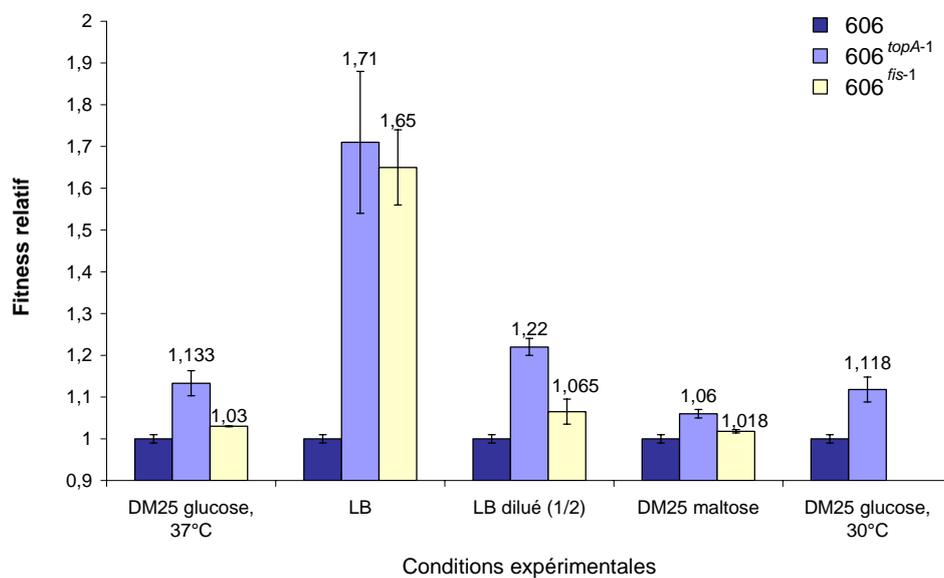


**Fitness clone évolué/clone ancêtre :**

$$W_{e/a} = \frac{\ln (D_1^e \times 100 / D_0^e)}{\ln (D_1^a \times 100 / D_0^a)}$$

**Figure 27 : Détermination du fitness de clones évolués par des expériences de compétitions**

Un clone évolué et l'ancêtre de marqueur Ara opposé sont adaptés pendant 24h au milieu dans lequel le fitness doit être déterminé, le plus souvent du DM25. Des volumes identiques de ces cultures sont ensuite utilisés pour inoculer une même culture. A l'instant du mélange (t=0), un prélèvement est effectué et une dilution adéquate est étalée sur un milieu contenant du tétrazolium et de l'arabinose (TA). Les clones des populations Ara+ y apparaîtront roses, alors que ceux des populations Ara- y apparaîtront rouges. Après 24h, un deuxième prélèvement est effectué et étalé sur TA. Les colonies des deux compétiteurs sont alors dénombrées (D) et le taux de croissance de chaque compétiteur calculé. Le fitness W du clone évolué (e) est le rapport du taux de croissance de ce clone sur le taux de croissance de l'ancêtre (a).

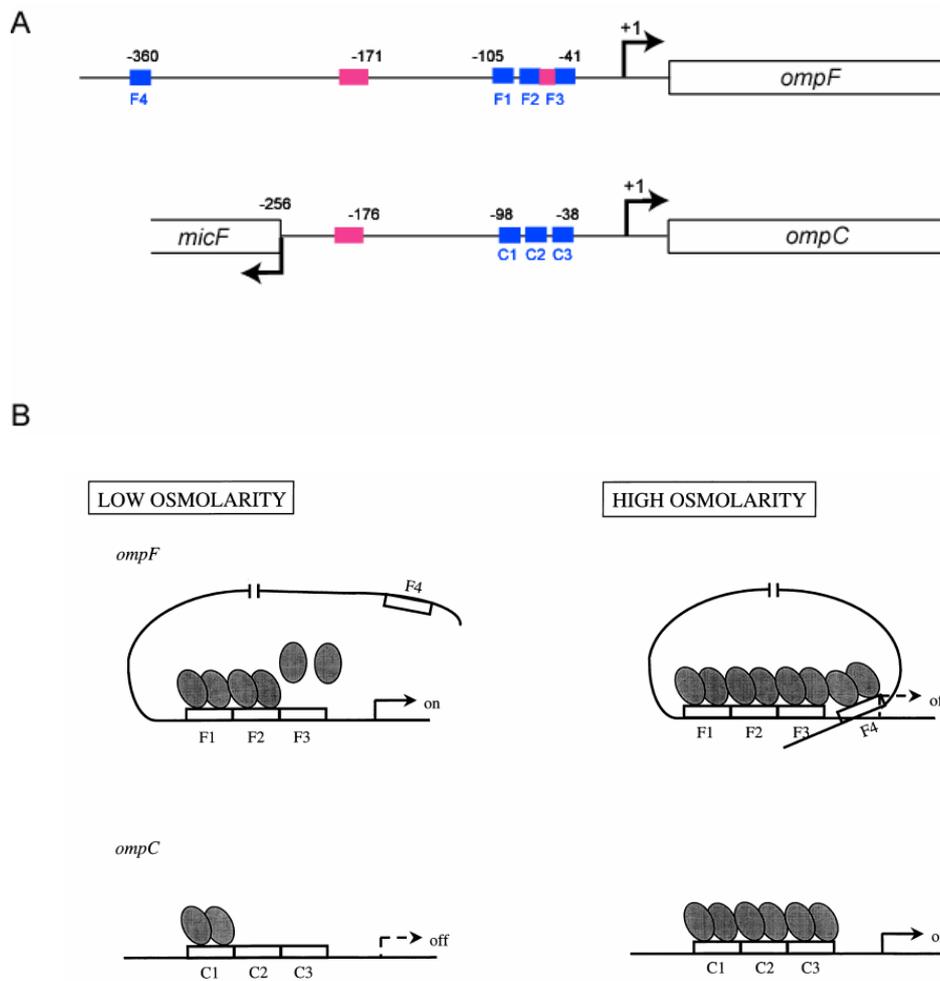


**Figure 28 : Effet sur le fitness des allèles évolués *topA*<sup>1</sup> et *fis*<sup>1</sup> dans le contexte génétique du clone ancêtre**

Les expériences de compétition ont été réalisées dans 5 environnements différents :

- les mêmes conditions que celles de l'évolution expérimentale (DM25 glucose, 37°C),
- du milieu riche LB à 37°C,
- du milieu LB dilué au demi à 37°C
- du milieu minimum DM25 à 37°C, avec du maltose remplaçant le glucose,
- du milieu DM25 glucose, mais à 30°C.

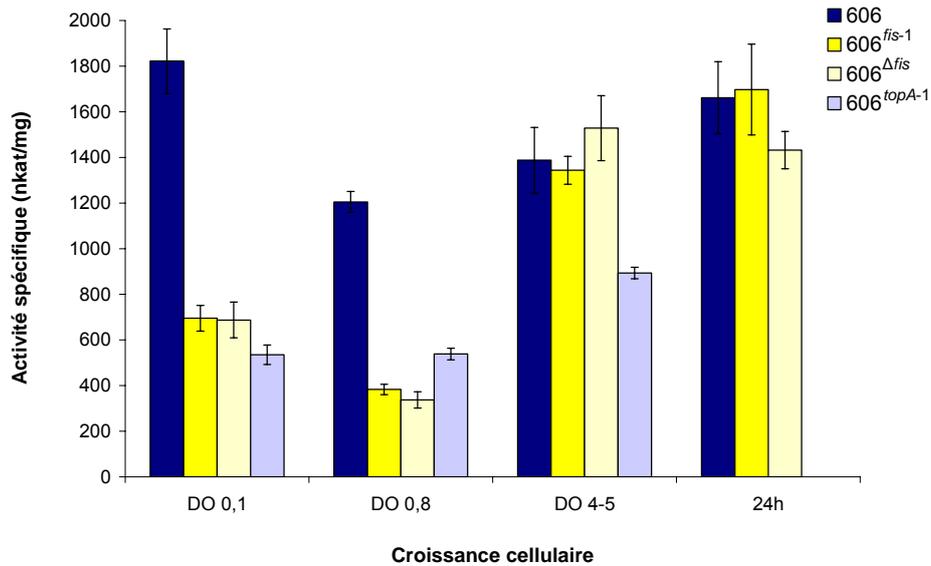
Les barres d'erreurs représentent les intervalles de confiance à 95% sur la base de 6 réplicats pour chaque compétition. Les valeurs de fitness sont données au dessus des barres d'erreurs. Chaque souche ancestrale, portant les allèles ancestraux *topA* et *fis* (606), portant l'allèle évolué *topA* de la population Ara-1 (606<sup>*topA-1*</sup>) et portant l'allèle évolué *fis* de la population Ara-1 (606<sup>*fis-1*</sup>), a été mise en compétition contre le clone ancêtre REL607 de marqueur arabinose opposé.



**Figure 29 : Les régions promotrices des gènes *ompF* et *ompC***

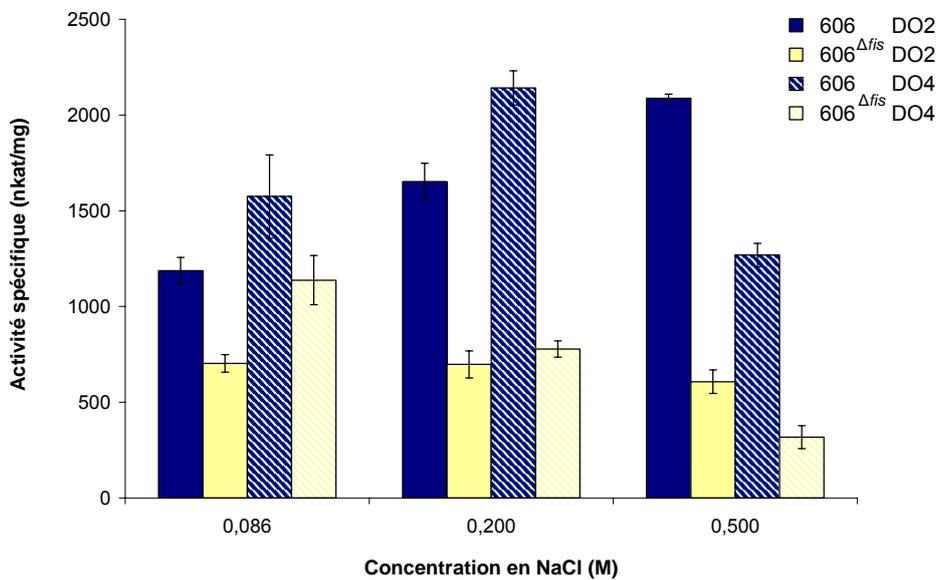
**A.** Régions promotrices des gènes *ompC* et *ompF*. Les gènes sont indiqués en italique, les flèches courbées indiquant les sites d'initiation de la transcription. Les sites de fixation de OmpR-P (F1 à F4 pour *ompF* et C1 à C3 pour *ompC*) sont indiqués en bleu, alors que les sites de fixation d'IHF sont indiqués en rose. Les positions de début et fin de ces différents sites sont indiquées par rapport aux +1 de transcription majoritaires des deux gènes *ompF* et *ompC*. Le gène *micF* est transcrit de façon divergente par rapport à *ompC*, les deux régions promotrices étant chevauchantes.

**B.** Régulation de *ompF* et *ompC* par OmpR. A basse osmolarité, la quantité de OmpR-P est basse. Celle-ci se fixe alors sur les sites F1 à F3 en amont du promoteur de *ompF* et permet son activation, alors que les sites C1 à C3 de la région promotrice de *ompC* ne sont pas couverts, ce qui entraîne une faible transcription à partir du promoteur de *ompC*. A plus forte osmolarité, de plus fortes quantités de OmpR-P sont présentes dans la cellule ; le site F4 de *ompF* est alors couvert par OmpR-P. La formation d'une boucle entraîne alors l'inhibition de la transcription de *ompF*, alors que la concentration d'OmpR-P est maintenant suffisante pour occuper tous les sites C1 à C3, ce qui permet d'activer la transcription de *ompC* (Bergstrom et al., 1998).



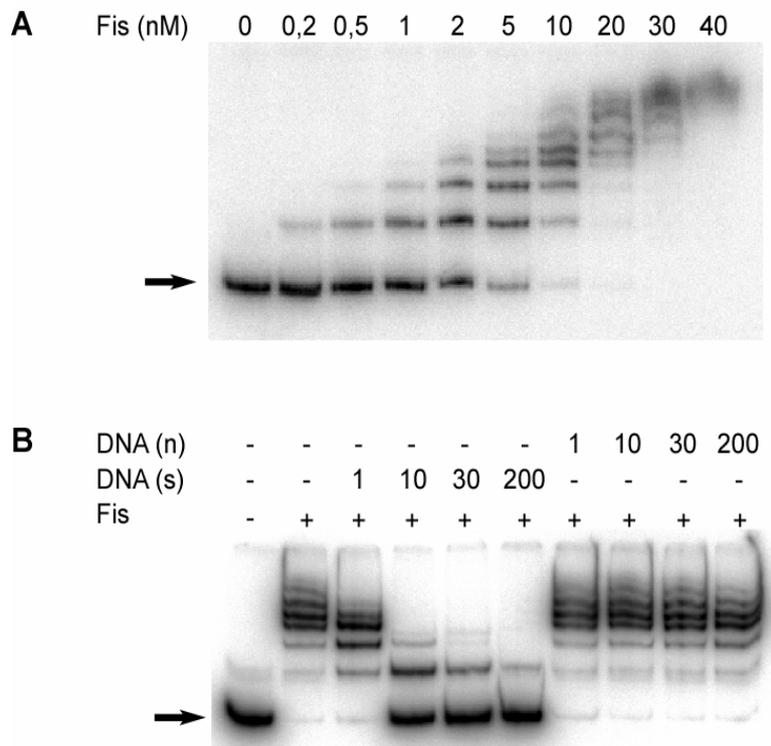
**Figure 30 : Activités spécifiques β-galactosidase correspondant aux fusions *ompFB-lacZ* dans différents contextes génétiques isogéniques du clone ancêtre**

Les activités spécifiques β-galactosidase, données en nkat/mg ont été mesurées après cultures des différentes souches en milieu LB et prélèvements à  $DO_{600nm}$  de 0,1 (début de phase exponentielle, 1h de culture), de 0,8 (fin de phase exponentielle, 2h de culture), de 4-5 (phase stationnaire, 5h de culture), et après 24h de culture. A ce dernier temps, la mesure correspondant à l'activité de *ompF* dans 606<sup>topA-1</sup> n'a pas été réalisée. 606, souche ancêtre REL606 ; 606<sup>fis-1</sup>, 606 portant l'allèle évolué *fis* de la population Ara-1 ; 606<sup>Δfis</sup>, 606 avec une délétion en phase de *fis* ; 606<sup>topA-1</sup>, 606 portant l'allèle évolué *topA* de la population Ara-1. Les erreurs standard ont été calculées à partir de 3 cultures indépendantes.



**Figure 31 : Activités spécifiques β-galactosidase des fusions *ompFB-lacZ* en fonction de la concentration en NaCl**

Les activités spécifiques β-galactosidase, données en nkat/mg ont été mesurées après cultures des différentes souches en milieu LB classique (0,086M NaCl), et en milieu LB avec 0,2 et 0,5M NaCl. Des prélèvements de chaque culture à  $DO_{600nm}$  de 0,2 (phase exponentielle, 1h de culture) et de 4 (phase stationnaire, 5h de culture). 606, souche ancêtre REL606 ; 606<sup>Δfis</sup>, 606 avec une délétion en phase de *fis*. Les erreurs standard ont été calculées à partir de 3 cultures indépendantes.

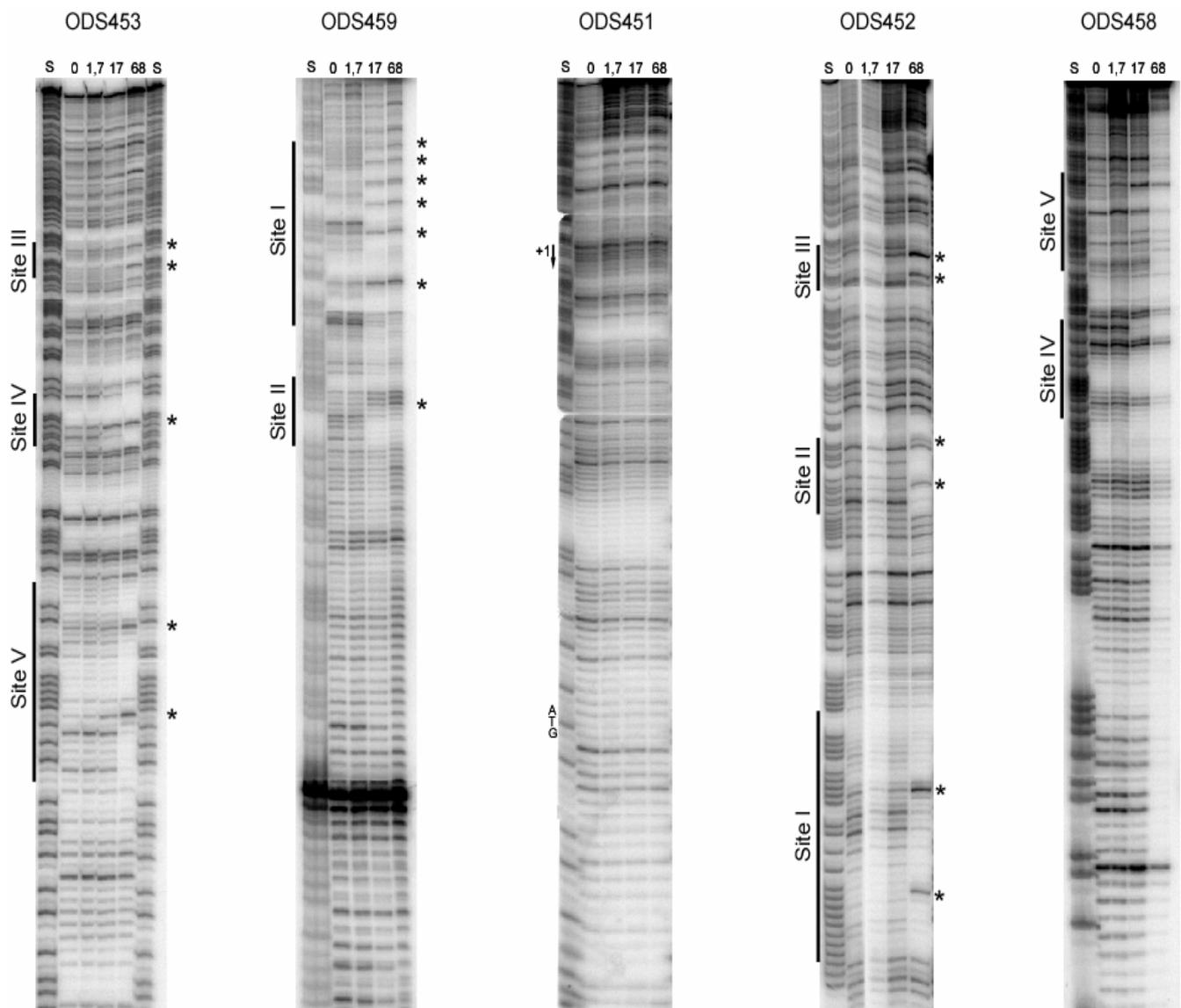


**Figure 32 : Fixation de Fis sur la région promotrice de *ompF***

**A.** Retard de migration de la région promotrice de *ompF* par différentes concentrations de Fis (en nM).

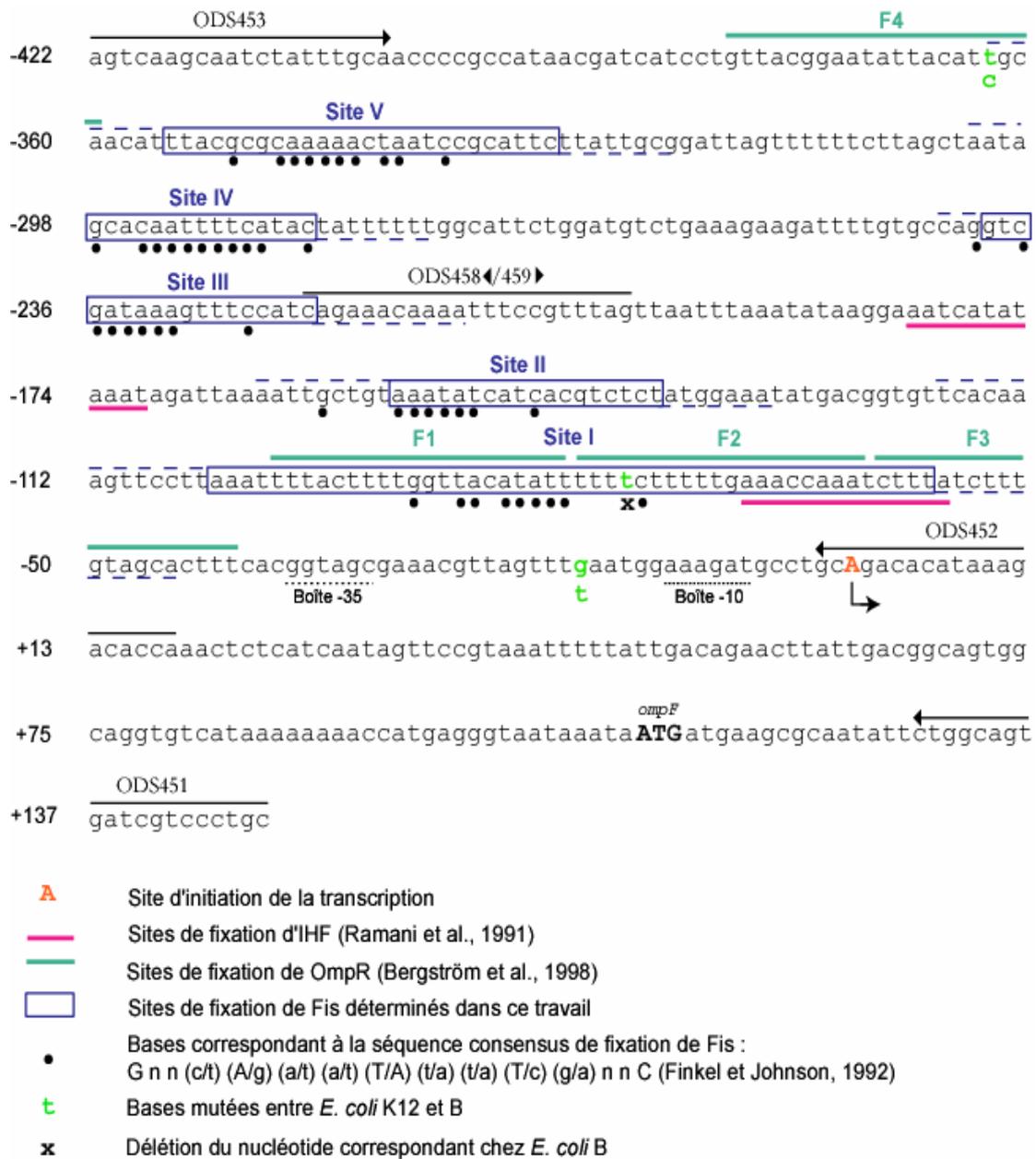
La région promotrice de *ompF* a été amplifiée par PCR, et le fragment obtenu allant des positions -422 à +148 par rapport au site majeur d'initiation de la transcription de *ompF*. Une des amorces utilisées lors de la PCR est radiomarquée. Des concentrations croissantes de Fis [0 à 40nM] ont été incubées en présence de 0,6nM de ce fragment d'ADN. Les réactions ont ensuite été déposées sur gel d'acrylamide 5%. La flèche indique l'ADN non retardé.

**B.** Retard de migration du même fragment *ompF* en présence de 20nM de Fis et de différentes quantités d'ADN compétiteur non marqué (en ng), en excès molaire de 5 à 1000 fois. L'ADN (s) est un oligonucléotide de 30pb contenant le site I de fixation spécifique de Fis sur son propre promoteur. L'ADN (n) est un oligonucléotide de 30pb où ce site de fixation de Fis a été mutagénisé. La flèche indique l'ADN non retardé.



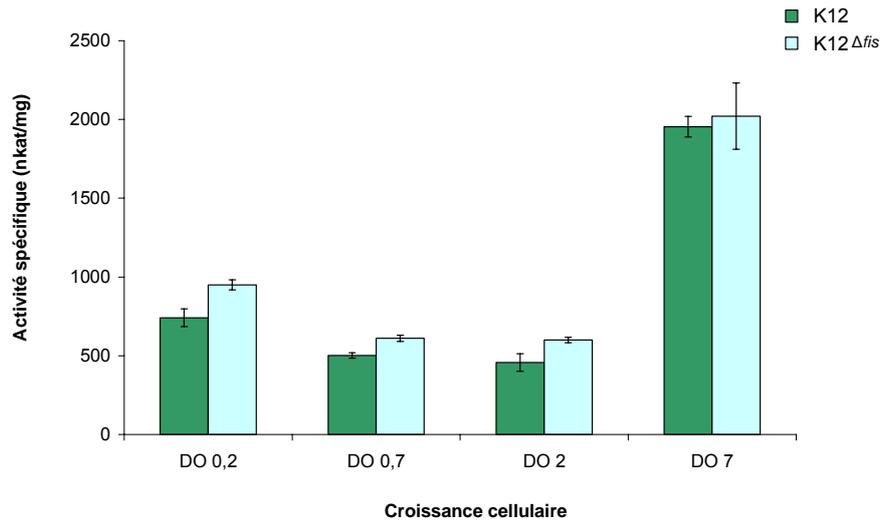
**Figure 33 : Empreintes à la DNase I de la protéine Fis sur la région promotrice de *ompF***

Les amorces radiomarquées sont indiquées en haut de chaque gel et ont été utilisées dans les réactions de PCR suivantes (de gauche à droite) : ODS453-452, ODS459-452, ODS451-453, ODS 452-453 et ODS458-453, permettant d'amplifier des fragments d'ADN allant respectivement des positions -422 à +18, -222 à +18, +148 à -422, +18 à -422 et -200 à -422. Les positions sont données par rapport au site majeur d'initiation de la transcription de *ompF*. La position des amorces sur la séquence est donnée **Figure 34**. Les rapports molaires Fis/ADN sont indiqués pour chaque piste et correspondent à ceux utilisés dans les expériences de retard sur gel (dans les pistes 0, 1, 10 et 40nM, **Figure 32**). Les sites de fixation de Fis déterminés par cette expérience sont indiqués sur le côté et les étoiles indiquent des sites d'hypersensibilité à la DNase I. Le +1 de transcription et l'ATG de *ompF* sont également indiqués. Les positions de chaque site ont été déterminées par une réaction de séquençage G+A par la méthode de Maxam et Gilbert (S).



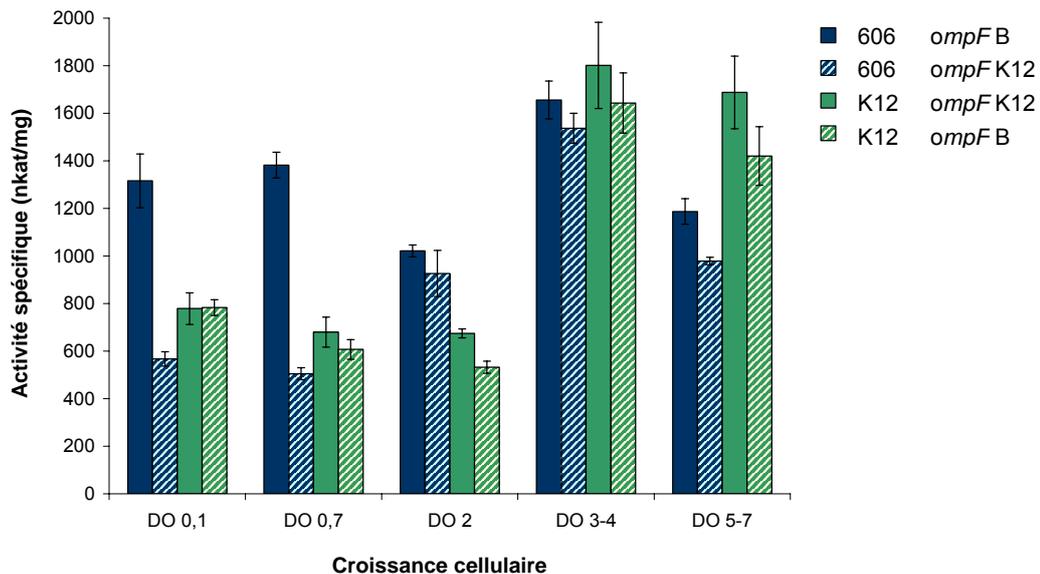
**Figure 34 : Région promotrice de *ompF* et localisation des sites de fixation de Fis**

La séquence de cette région, correspondant à la souche *E. coli* K12 MG1655 (Blattner et al., 1997), est donnée, ainsi que la position des amorces utilisées pour les expériences d'empreintes à la DNase I (les flèches indiquent le sens 5'-3'), les sites de fixation des différents régulateurs (voir légende sur la figure), dont les sites de fixation de Fis déterminés dans ce travail. Les bases correspondant à la séquence consensus de Fis sont indiquées par un point, et les pointillés bleus indiquent les zones protégées sur un seul des deux brins. Le +1 de transcription et l'ATG de *ompF*, ainsi que les boîtes -10 et -35 du promoteur *ompF* sont indiqués. La direction de la transcription est donnée par une flèche courbe. La région promotrice *ompF* chez *E. coli* B présente des différences par rapport à celle d'*E. coli* K12 ; les positions nucléotidiques modifiées chez *E. coli* B sont indiquées en vert, en dessous de la séquence, la croix indiquant une délétion du nucléotide indiqué.



**Figure 35 : Activités spécifiques  $\beta$ -galactosidase des fusions transcriptionnelles *ompFK12-lacZ* dans les contextes isogéniques K12 et K12 $\Delta$ fis**

Les activités spécifiques  $\beta$ -galactosidase, données en nkat/mg, ont été mesurées après cultures des deux souches en milieu LB et des prélèvements à DO<sub>600nm</sub> de 0,2 (début de phase exponentielle, 1h de culture), 0,7 (phase exponentielle, 2h de culture), 2 (transition phase exponentielle - phase stationnaire, 3h de culture) et 7 (phase stationnaire, 6h de culture). K12 est la souche CF7968 et K12 $\Delta$ fis la souche CF7968 portant une délétion en phase de *fis*. Les erreurs standard ont été calculées à partir de 3 cultures indépendantes.



**Figure 36 : Activités spécifiques  $\beta$ -galactosidase des fusions transcriptionnelles *ompFB-lacZ* ou *ompFK12-lacZ* dans les contextes génétiques *E. coli* B ou K12**

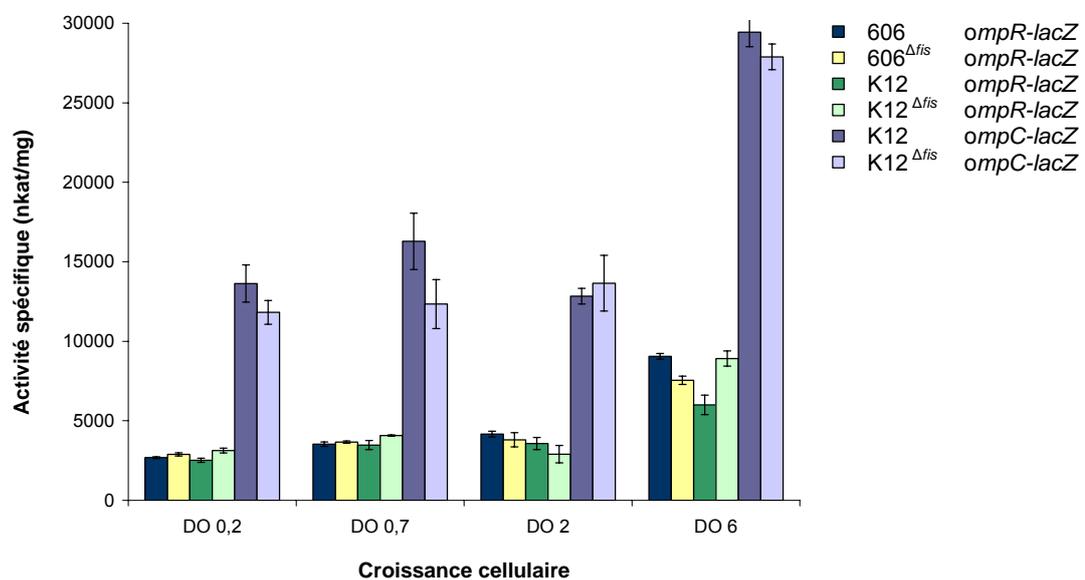
Ces mesures ont été effectuées en introduisant chaque fusion *ompFB* ou K12 - *lacZ* dans chacune des souches *E. coli* B (REL606) ou K12 (CF7968) et les activités spécifiques  $\beta$ -galactosidase, données en nkat/mg, ont été mesurées après des cultures en milieu LB en fonction de la phase de croissance. Les prélèvements ont été effectués à des DO<sub>600nm</sub> de 0,1 (début de phase exponentielle, 1h de culture), 0,7 (phase exponentielle, 2h de culture), 2 (transition phase exponentielle - phase stationnaire, 3h de culture) et 3 et 5 (phase stationnaire, 5h et 9h de culture). Les erreurs standard ont été calculées à partir de 3 cultures indépendantes.

606 *ompFB* : fusion transcriptionnelle *ompFB-lacZ* dans *E. coli* B (REL606)

606 *ompFK12* : fusion transcriptionnelle *ompFK12-lacZ* dans *E. coli* B (REL606)

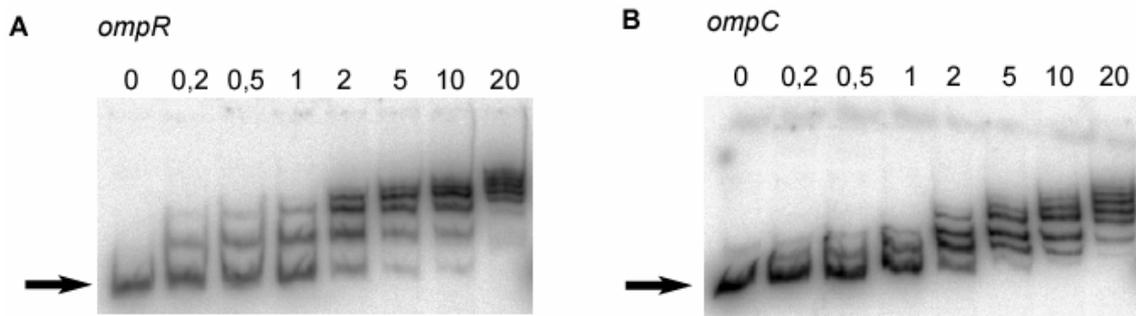
K12 *ompFK12* : fusion transcriptionnelle *ompFK12-lacZ* dans *E. coli* K12 (CF7968)

K12 *ompFB* : fusion transcriptionnelle *ompFB-lacZ* dans *E. coli* K12 (CF7968)



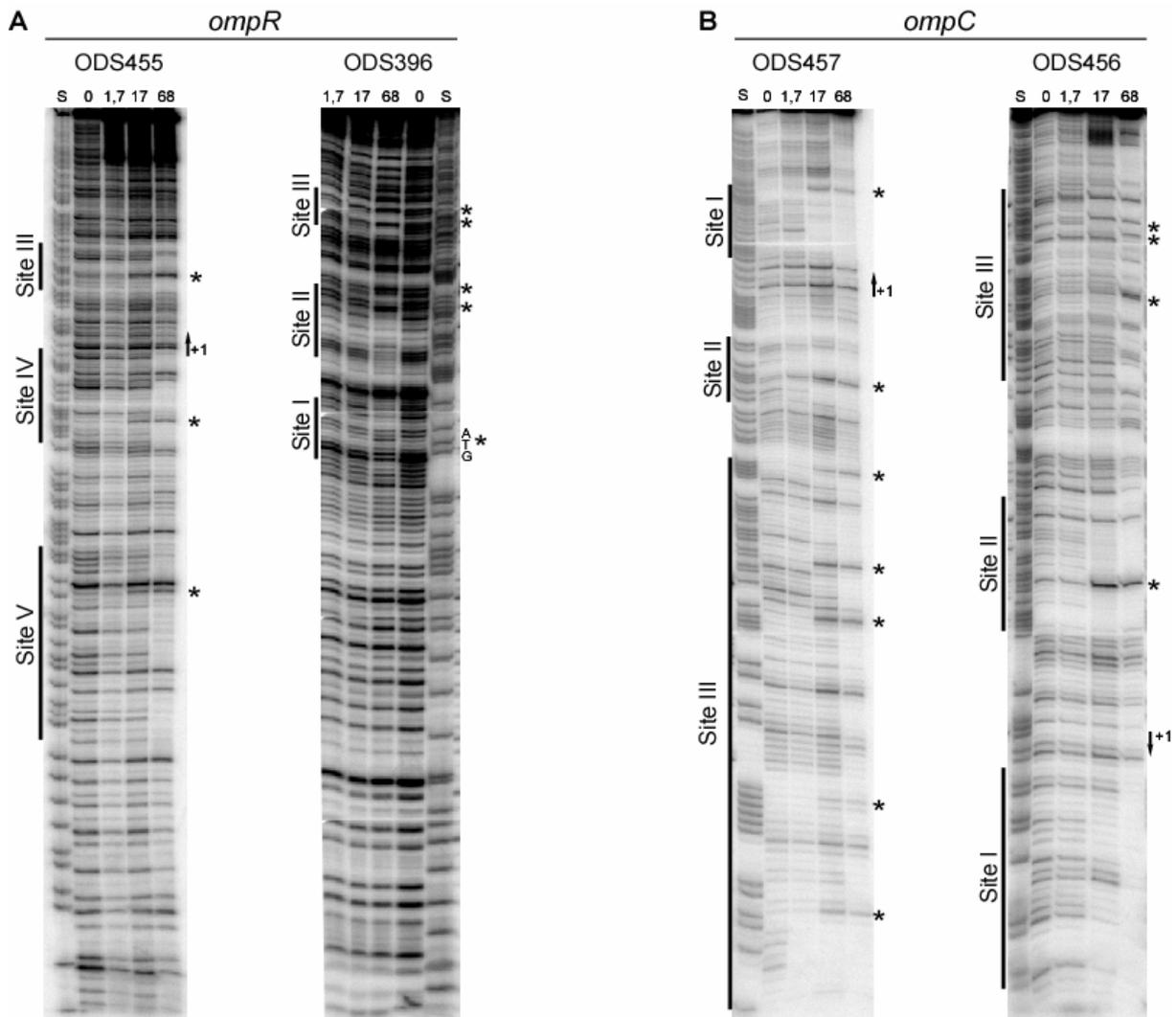
**Figure 37 : Activités spécifiques  $\beta$ -galactosidase des fusions transcriptionnelles *ompR - lacZ* ou *ompC - lacZ* dans les contextes génétiques *E. coli* B et K12**

Les activités spécifiques  $\beta$ -galactosidase, données en nkat/mg, ont été mesurées après cultures des souches en milieu LB. Les prélèvements ont été effectués à  $DO_{600nm}$  de 0,2 (début de phase exponentielle, 1h de culture), 0,7 (phase exponentielle, 2h de culture), 2 (fin de phase exponentielle, 3h de culture) et 6 (phase stationnaire, 6h de culture). Les souches utilisées sont : 606, ancêtre *E. coli* B (REL606) ; 606<sup>Δfis</sup>, REL606 portant une délétion de *fis* ; K12 est la souche CF7968 et K12<sup>Δfis</sup> la souche CF7968 portant une délétion de *fis*. Ces différentes souches portent les fusions transcriptionnelles *ompR - lacZ* ou *ompC - lacZ*. Les erreurs standard ont été calculées à partir de 3 cultures indépendantes.



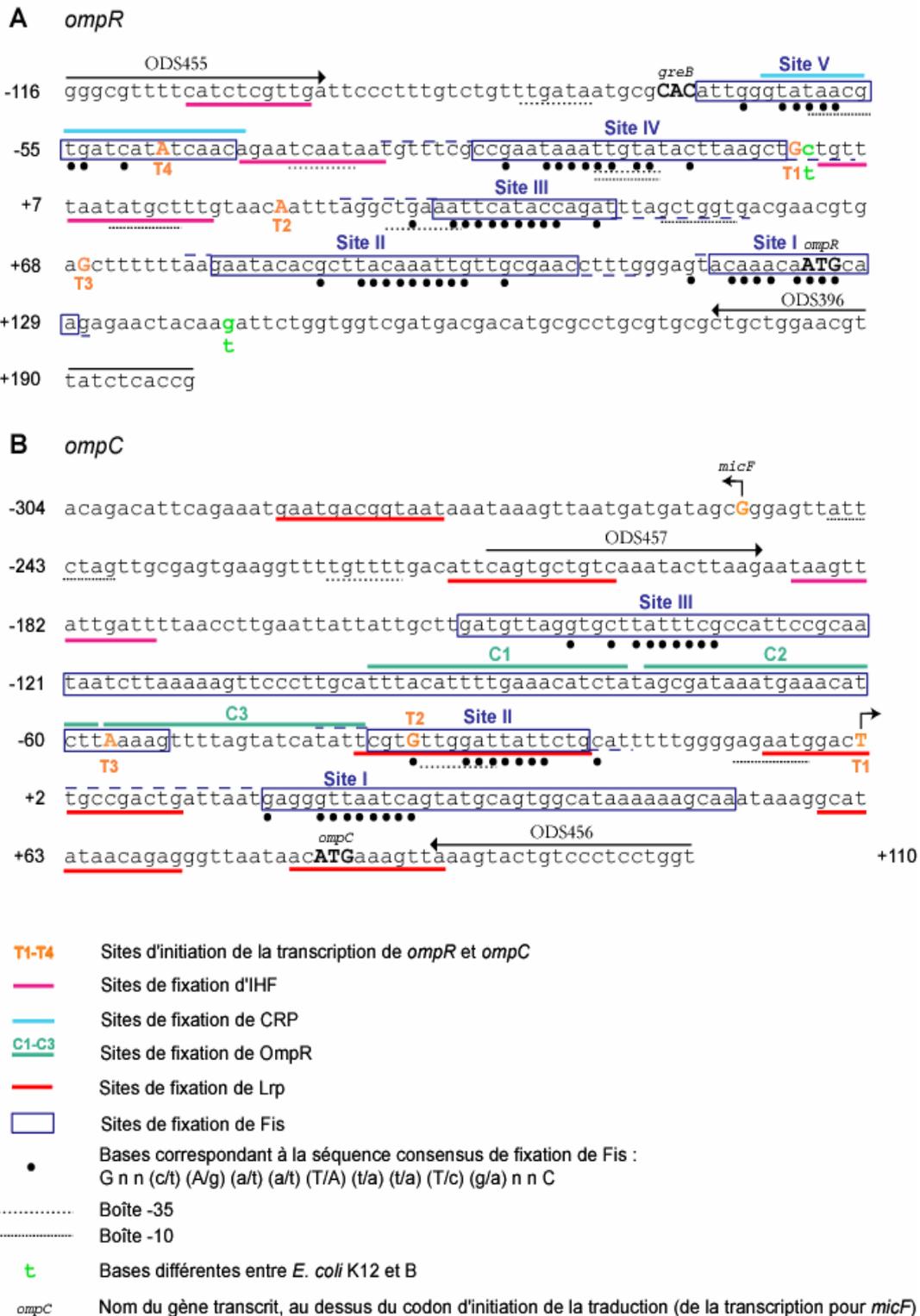
**Figure 38 : Fixation de Fis sur les régions promotrices de *ompR* (A) ou *ompC* (B)**

Des concentrations croissantes de Fis (en nM) ont été incubées avec 0,6nM de fragment radiomarqué correspondant aux régions promotrices de *ompR* (A) ou *ompC* (B). Les expériences ont été réalisées dans les mêmes conditions que pour *ompF*, en effectuant les mêmes contrôles. Les flèches indiquent les fragments non retardés.



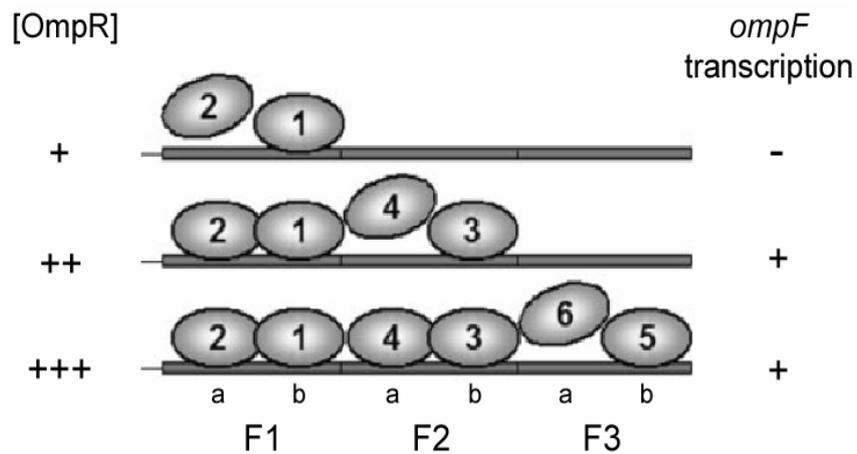
**Figure 39 : Détermination des sites de fixation de Fis sur les régions promotrices de *ompR* ou *ompC* par des expériences d'empreintes à la DNase I**

Les expériences d'empreintes à la DNase I sur les régions promotrices de *ompR* (A) ou *ompC* (B) ont été réalisées dans les mêmes conditions que pour *ompF*, et les différents éléments sont indiqués de la même manière (voir légende de la **Figure 33**). Les fragments d'ADN ont été amplifiés par PCR en utilisant les amorces ODS455 et 396 pour le fragment *ompR* et ODS456 et 457 pour le fragment *ompC*. La localisation des amorces est montrée **Figure 40**. Dans chacun des cas, chacune des amorces a été radiomarquée et des électrophorèses de durées différentes (courte et longue) ont été réalisées afin de visualiser la totalité des deux brins codant et non-codant avec chaque amorce marquée. Ici, seules les électrophorèses dites « courtes » sont montrées.



**Figure 40 : Régions promotrices de *ompR* (A) ou *ompC* (B) et localisation des sites de fixation de Fis**

Les séquences de ces régions, correspondant à la souche d'*E. coli* K12 MG1655 [Blattner et al., 1997], sont données (en **A**, *ompR*, et en **B**, *ompC*), ainsi que la position des amorces utilisées pour les expériences d'empreintes à la DNase I (les flèches indiquent le sens 5'-3'), les sites de fixation des différents régulateurs (voir légende sur la figure), dont les sites de fixation de Fis déterminés dans ce travail. Les bases correspondant à la séquence consensus de Fis sont indiquées par des points. Les +1 de transcription, les codons d'initiation ATG de *ompR* et *ompC* et GTG de *greB*, ainsi que les boîtes -10 et -35 sont indiqués. Les flèches courbes indiquent la direction de la transcription, au niveau des sites majeurs d'initiation. Les nucléotides qui diffèrent chez *E. coli* B sont indiqués en dessous de la séquence, en vert.



**Figure 41 : Fixation de OmpR sur *ompF* (Yoshida et al., 2006)**

Un modèle « galopant » a été proposé pour expliquer l'activation de *ompF* par OmpR-P. OmpR-P, qui présente une plus grande affinité pour les sous-sites b, se fixe d'abord sur F1b. Cette fixation permet à une seconde molécule de OmpR-P de se fixer sur le sous-site a, puis le mécanisme continue jusqu'à recouvrir les trois sites, à la manière d'un galop de cheval. La fixation de quatre molécules de OmpR-P sur les sites F1 et F2 suffit à l'activation de la transcription de *ompF*. Les molécules de OmpR-P sont représentées par des ovals, numérotés de 1 à 6, dans leur ordre séquentiel de fixation.

Souches		Description
Résultats Partie I (Crozat et al., 2005)	Résultats Parties II et III	
606	606	Ancêtre REL606 <i>Escherichia coli</i> B Bc251, T6 <sup>R</sup> , Str <sup>R</sup> , <i>rmIII</i> , Ara-
-	607	Ancêtre REL607, révertant Ara+ de REL606
2K	1164	Clone évolué isolé à 2000 générations dans la population Ara-1
10K	-	Clone évolué isolé à 10 000 générations dans la population Ara-1
20K	-	Clone évolué isolé à 20 000 générations dans la population Ara-1
-	1158C	Clone évolué isolé à 2000 générations dans la population Ara+1
606 <sup>topA<sup>e</sup></sup>	606 <sup>topA-1</sup>	Ancêtre REL606 pour lequel l'allèle évolué <i>topA</i> isolé dans la population Ara-1 a été introduit dans le chromosome
606 <sup>fis<sup>e</sup></sup>	606 <sup>fis-1</sup>	Ancêtre REL606 pour lequel l'allèle évolué <i>fis</i> isolé dans la population Ara-1 a été introduit dans le chromosome
606 <sup>topA<sup>e</sup> fis<sup>e</sup></sup>	606 <sup>topA-1 fis-1</sup>	Ancêtre REL606 pour lequel les allèles évolués <i>topA</i> et <i>fis</i> isolés dans la population Ara-1 ont été introduits dans le chromosome
-	606 <sup>fis-3</sup>	Ancêtre REL606 pour lequel l'allèle évolué <i>fis</i> isolé dans la population Ara-3 a été introduit dans le chromosome
606 <sup>fis<sup>Δ</sup></sup>	606 <sup>Δfis</sup>	Ancêtre REL606 portant une délétion en phase du gène <i>fis</i>
2K <sup>topA<sup>+</sup></sup>	-	Clone évolué 1164 pour lequel l'allèle évolué du gène <i>topA</i> a été remplacé par l'allèle ancestral.
-	1158C <sup>topA-1</sup>	Clone évolué 1158C pour lequel l'allèle ancestral du gène <i>topA</i> a été remplacé par l'allèle évolué isolé dans la population Ara-1
-	606 <sup>fisP</sup>	REL606 λRS45 :: P <i>fis-lacZ</i> (fusion transcriptionnelle du gène <i>lacZ</i> avec le promoteur ancestral de <i>fis</i> intégré dans le chromosome de REL606 au site d'attachement de l'ADN du bactériophage λ)
-	606 <sup>fis-1P</sup>	REL606 λRS45 :: P <i>fis-lacZ</i> (fusion transcriptionnelle du gène <i>lacZ</i> avec le promoteur évolué de <i>fis</i> issu de la population Ara+1)
-	K12	Souche CF7968, <i>Escherichia coli</i> K12 MG1655, Lac-
-	K12 <sup>Δfis</sup>	Souche CF7968 portant une délétion en phase du gène <i>fis</i>

**Tableau 1 : Nom des souches utilisées dans les 3 parties des résultats**

(-) Souche non utilisée dans la partie correspondante des résultats.

Protéine (Fonction)	606 <sup>topA-1</sup>	606 <sup>fis-1</sup>	606 <sup>topA-1 fis-1</sup>	606 <sup>Δfis</sup>
PAL (Peptidoglycan-associated lipoprotein precursor)	+	+	+	+
MetE (Homocystéine méthyltransférase)	+	+	+	=
OmpA (Outer membrane protein A)	=	+	+	+
Aat (Aspartate amino-transferase)	+	=	=	+
OmpF (Outer membrane protein F)	-	-	-	-
Protéine ribosomale 50S L7/L12	=	-	=	-
EF-Tu (Elongation factor Tu)	=	-	=	-
SurA (Précurseur de la chaperone SurA)	=	-	=	-
H-NS (Protéine de type histone H-NS)	=	-	=	-

**Tableau 2 : Modification de l'expression de protéines par les allèles évolués *topA* et *fis* isolés dans la population Ara-1**

Les protéines dont la quantité varie de manière reproductible dans une ou plusieurs souches isogéniques par rapport au clone ancêtre sont indiquées ici. Les souches sont toutes isogéniques entre elles sauf pour les allèles évolués de *topA* et/ou *fis* et sont comparées au clone ancêtre de l'évolution expérimentale.

(+) augmentation de la quantité de protéines dans la souche correspondante,

(-) diminution de la quantité de protéines

(=) pas de changement détecté.

## RÉSUMÉ

Les mécanismes moléculaires de l'adaptation des bactéries à leur environnement ont été étudiés par une stratégie d'évolution expérimentale, initiée dans le laboratoire du Pr. Richard Lenski (MI, USA) en 1988. Douze populations, fondées à partir de la même cellule ancêtre d'*Escherichia coli* ont évolué indépendamment pendant plus de 40 000 générations, par transferts journaliers dans le même milieu minimum limité en glucose. Au cours de cette évolution, le fitness, c'est-à-dire la capacité reproductive des cellules, augmente d'environ 70% dans les 12 populations après 20 000 générations d'évolution. L'objectif de ce travail est d'identifier certaines des mutations conduisant à cette augmentation de fitness, puis de les analyser afin de comprendre les mécanismes à la base de cette évolution darwinienne.

Au cours de ce travail, nous avons montré que l'augmentation du fitness des bactéries était associée à un accroissement de superhélicité de l'ADN, et ce de façon parallèle dans la majorité des 12 populations après 20 000 générations. L'étude d'une population modèle présentant deux modifications successives de superhélicité de l'ADN nous a permis de mettre en évidence deux mutations, une dans le gène *topA* codant la topoisomérase I, et une dans le gène *fis* codant la protéine de type histone Fis. Nous avons pu montrer que ces deux mutations étaient responsables non seulement de l'ensemble des changements de superhélicité de l'ADN dans cette population, mais aussi qu'elles entraînaient une augmentation du fitness. Ces études ont été étendues aux autres populations par séquençage de 9 loci (dont *topA* et *fis*) impliqués dans la modulation de la superhélicité dans un clone évolué issu de chacune des 12 populations. Un degré exceptionnel de parallélisme génétique a été mis en évidence, trois gènes seulement étant les cibles de la sélection naturelle, dont *topA* et *fis*. Ces analyses nous ont également permis d'identifier un troisième gène impliqué dans l'augmentation du fitness : *dusB*, codant une dihydrouridine synthase, et qui pourrait avoir évolué pour remplir une fonction liée à celle de *fis*. Malgré le fort degré de parallélisme génétique observé, une divergence allélique importante a été détectée. L'analyse moléculaire d'allèles évolués du gène *fis* affectant sa transcription, sa traduction ou l'activité de la protéine, nous a cependant permis de montrer que cette divergence allélique conduisait à un niveau de parallélisme moléculaire extrêmement élevé. Cette étude a par ailleurs permis de démontrer l'implication de Fis sur la composition protéique de la membrane externe.

L'évolution de ces 12 populations se caractérise donc par un degré important de parallélisme phénotypique qui s'explique par un degré exceptionnel de parallélisme génétique et moléculaire. L'adaptation des bactéries à leur environnement est réalisée en modifiant un des réseaux majeurs de régulation de l'expression des gènes : la superhélicité de l'ADN.