

UNIVERSITE DE LIMOGES

ECOLE DOCTORALE THEMATIQUE Biologie-Santé

Faculté de Médecine

EA3175, Equipe Avenir Inserm, Biologie Cellulaire et Moléculaire des Microorganismes

Thèse N° []

Thèse

pour obtenir le grade de

Docteur de l'Université de Limoges

Discipline : **Biologie-Sciences-Santé**

Présentée et soutenue par

THOMAS JOVÉ

Le 4 février 2010

**REGULATION DE L'EXPRESSION DES GENES DE CASSETTES ET DE L'INTEGRASE
DES INTEGRONS DE MULTIRESISTANCE AUX ANTIBIOTIQUES**

Thèse dirigée par Madame le Professeur **Marie-Cécile Ploy**

Présentée devant le jury composé de :

Professeur **FRANÇOIS DENIS**

Président

Professeur **MARIE-CÉCILE PLOY**

Directeur de thèse

Docteur **MICHAEL CHANDLER**

Rapporteur

Docteur **THIERRY NAAS**

Rapporteur

Docteur **DIDIER MAZEL**

Examineur

Docteur **ERIC PINAUD**

Examineur

A ma grand-mère,

A mes parents,

A ma femme,

A mes enfants.

Remerciements

A Monsieur le Professeur **François Denis**, Professeur des Universités de Bactériologie-Virologie-Hygiène, de me faire l'honneur de présider ce jury et de me donner ainsi l'occasion de vous présenter moi-même mes travaux de thèse.

A Monsieur le Docteur **Michael Chandler**, Directeur de Recherche au Centre National de la Recherche Scientifique (Toulouse) et Monsieur le Docteur **Thierry Naas**, Maître de Conférences à l'Université Paris XI et Praticien Hospitalier à l'Hôpital de Bicêtre d'avoir accepté de juger mon travail en participant à ce jury. Soyez tous deux assurés de mon profond respect pour votre travail.

A Monsieur le Docteur **Didier Mazel**, Directeur de Recherche à l'Institut Pasteur (Paris) pour ses conseils pendant toute ma thèse. Votre participation à ce jury est un très grand honneur. Enfin, merci à Monsieur le Docteur **Eric Pinaud**, Chargé de Recherche au Centre National de la Recherche Scientifique (Limoges) de contribuer à ce jury.

A Madame le Professeur **Marie-Cécile Ploy**, Professeur des Universités et Chef de Service de Bactériologie-Virologie-Hygiène au CHU de Limoges. Pour sa confiance, son honnêteté, sa disponibilité. Le sujet que vous m'avez proposé a été une source inépuisable d'intérêt et de motivation. J'en garde un excellent souvenir et considère avoir profité des meilleures conditions de thèse qui soient.

Mes tous premiers remerciements s'adressent à mes collègues, compagnons de laboratoire, « intégristes » ou non mais sans lesquels ces années de thèse n'auraient pas eu la même saveur : Olivier, pour ton humour, ta brillance, et nos conversations sur les intégrons ; Emilie G., pour ton amitié dans les jours bons comme les mauvais, les rires, la musique, et la façon dont tu gardes ton calme quand tu stresses les bactéries ; Martine, pour ta bienveillance à notre égard et bien sûr les chocolats ; Sébastien C., pour ta décontraction (une réponse en Joules !), tes petites et grandes histoires ; Aurore, pour ton aide, tes résultats, ta motivation; Delphine, pour ton amitié et tes visites (toujours un plaisir) ; Sandra, pour ta méticulosité et tes compétences en protéines (que j'envie !) ; Gaël, pour ta façon de travailler, très motivante ; Sophie R., pour ta bonne humeur, tes conseils ; Fadi pour tes imitations de Gaston Lagaffe. Sébastien H., Aurélie, Jérôme et Stéphanie pour leur amitié, et les autres « chefs » de l'équipe pour leur considération à mon égard, Sophie A. et Fabien. Aux nouveaux doctorants, Thibault et Lucie, je souhaite le meilleur !

Une thèse, c'est aussi des collaborations et des échanges, l'occasion de saluer Laura de l'Université de La Rioja et Alexandra de l'Université d'Aveiro. Je pense aussi à mes stagiaires, en particulier Fahed, dont l'intérêt pour mon travail m'a touché. Je n'oublie pas l'ensemble du personnel du Service de Bactériologie-Virologie-Hygiène : merci pour votre accueil chaleureux et sincère !

J'adresse des remerciements tout particuliers à deux amis, Anne-Gaëlle et Aurélien : votre soutien est très important !

Ma dernière pensée ira à ma famille, pour leur compréhension, leur soutien, leur patience, particulièrement à ma femme Olympe dont la patience n'a d'égale que la tendresse que je lui porte.

Sommaire

Résumé	7
Abstract	9
Index des tableaux	10
Index des figures	12
Liste des annexes	14
Glossaire	15
Préambule	18
Introduction	19
<u>Chapitre 1. Initiation de la transcription chez les procaryotes</u>	19
A/ L'ARN Polymérase	19
B/ Les promoteurs bactériens	23
C/ Déroulement de la transcription chez les procaryotes	26
D/ Régulation de la transcription des gènes bactériens	28
<u>Chapitre 2. Les intégrons</u>	37
A/ Généralités sur les intégrons	37
B/ Diversité et Epidémiologie des intégrons	44
C/ Origine et Evolution des intégrons	50
D/ Dynamique d'échange de cassettes	50
E/ Expression des gènes d'intégrons	53
Présentation du travail de thèse	62
Résultats	62
<u>Partie n°1 : Le polymorphisme du promoteur Pc des intégrons de classe 1 affecte l'activité de l'intégrase IntI1</u>	64
<u>Partie n°2 : Etude épidémiologique de la prévalence des variants de Pc au sein d'une collection de souches d'<i>Escherichia coli</i></u>	65

<u>Partie n°3 : Interférences entre expression des cassettes et expression de l'intégrase chez les intégrons de classe 1</u>	<u>71</u>
<u>Partie n°4 : Expression des gènes de cassettes et de l'intégrase chez les intégrons de classe 2</u>	<u>81</u>
Discussion	90
Conclusions et Perspectives	102
Références Bibliographiques	103
Annexes	120

Résumé

Les intégrons de multirésistance (IM) sont des éléments génétiques jouant un rôle majeur dans la résistance des bactéries aux antibiotiques, principalement les bactéries à Gram négatif. Ils constituent un système naturel de génie génétique permettant la capture, l'expression et la dissémination de gènes de résistance contenus dans des cassettes. Un intégron code une intégrase IntI qui catalyse l'insertion de ces cassettes en aval d'un promoteur Pc. Ce promoteur assure l'expression des gènes de cassettes, à la façon d'un opéron. Il existe plusieurs classes d'IM. Notre travail a porté sur les IM des classes 1 et 2, les plus prévalentes dans les isolats cliniques. Le Pc des IM de classe 1 est situé dans la séquence codante du gène de l'intégrase et existe en plusieurs variants de forces distinctes. En analysant l'ensemble des séquences d'IM de classe 1 disponibles *in silico*, nous avons montré qu'il existe au moins 13 variants du promoteur Pc inégalement distribués, les variants les plus faibles prédominant. Ces tendances ont été confirmées dans une étude épidémiologique d'IM de classe 1 d'une population de souches d'*Escherichia coli* isolées en France ou en Espagne. D'autre part, nous avons montré que plus le variant de Pc est faible, plus l'intégrase est active. Par ailleurs, Pc faisant face au promoteur de l'intégrase PintI1, il existe une interférence entre ces promoteurs.

Enfin, nous nous sommes intéressés aux IM de classe 2 souvent décrits dans différentes études épidémiologiques mais assez peu étudiés pour l'expression de leurs gènes de cassettes et de l'intégrase, cette dernière étant le plus souvent non fonctionnelle. Nous avons caractérisé la région promotrice des gènes de cassettes et de l'intégrase de ces IM et avons montré que dans les rares cas où le gène de l'intégrase code une protéine fonctionnelle, la région promotrice des gènes de cassettes est moins efficace.

Ce système d'équilibre entre expression des gènes de cassette et activité de l'intégrase est probablement une des clés du succès biologique des IM de classe 1 et 2 dans le monde bactérien.

Mots clés : intégrons, promoteurs, régulation de la transcription, résistances, antibiotiques.

Title : Regulation of gene cassettes genes and integrase expression in multiresistant integrons.

Abstract: Multiresistant integrons (MRI) are genetic elements largely involved in the dissemination of antibiotic resistance among Gram-negative bacteria. They act as natural genetic tools able to capture, express and disseminate antibiotic resistance genes embedded within cassettes. They typically consist of a gene (*intI*) encoding an integrase that catalyzes the gene cassette movement by site-specific recombination, a recombination site (*attI*) and a promoter (Pc) responsible for the expression of inserted gene cassettes. Several classes of MRI have been identified and our work focused on the classes 1 and 2 MRI, which are the most prevalent among clinical isolates. In class 1 MRI, Pc is located within the integrase gene. Based on an *in silico* analysis of all the class 1 MRI sequences available in the databases, we showed that there are 13 variants of Pc unequally distributed, the weakest being the most prevalent. We found that these *in silico* trends are consistent with *in vivo* distribution of Pc, as we studied in french or spanish *E.coli* strains. Moreover, we showed that the weakest the Pc variant, the more efficient the encoded IntI1. On the other hand, Pc being face to face with the integrase gene promoter, we showed that there are interferences between these promoters. Lastly, we focused on the class 2 MRI, whose *intI* gene most often encodes a truncated integrase. We characterized the promoter of this gene but also of the two promoters driving the transcription of the cassettes genes inserted in this class. These two promoters are less efficient in the rare MRI 2 that encodes a functional IntI2.

Thus, MRI display equilibrium between cassette gene expression and integrase activity that is probably a key element for the successfulness of these IM in the bacterial world.

Index des Tableaux

- Tableau 1 : Liste et caractéristiques des sept facteurs σ d'*Escherichia coli* identifiés à ce jour.
- Tableau 2 : Liste et caractéristiques des sept facteurs de transcription généraux d'*Escherichia coli*
- Tableau 3 : Pourcentages d'identité entre les séquences en acides aminés des intégrases d'intégrons de multi résistances
- Tableau 4 : Séquence des promoteurs Pc de plusieurs classes d'intégrons
- Tableau 5 : Séquence des variants des promoteurs Pc et P2 des IM1
- Tableau 6 : Séquence des variants des promoteurs Pc des IM 3
- Tableau 7 : Séquence des promoteurs putatifs de cassettes de gènes.
- Tableau 8 : Séquence des promoteurs PintI des gènes d'intégrase d'intégrons
- Tableau 9 : Réseaux des cassettes de gènes des IM 1 en fonction de l'origine géographique et clinique de la souche.
- Tableau 10 : Réseaux des cassettes de gènes des IM 1 en fonction de l'origine géographique et clinique de la souche.
- Tableau 11 : Prévalence des variants de Pc et de P2 parmi les 89 IM 1 en fonction de leur groupe phylogénétique
- Tableau 12 : Distribution des variants de Pc/P2 pour les réseaux de cassettes identifiés plusieurs fois.
- Tableau 13 : Souches bactériennes utilisées lors de ce travail
- Tableau 14 : Plasmides utilisées lors de ce travail
- Tableau 15 : Oligonucléotides utilisés dans ce travail

- Tableau 16: Polymorphisme du promoteur PintI1
- Tableau 17: Souches bactériennes et plasmides utilisés dans ce travail
- Tableau 18: Oligonucléotides utilisés dans ce travail
- Tableau 19: Correspondance entre le variant de Pc et la séquence en acides aminés de l'intégrase IntI1

Index des figures

- Figure 1 : Représentation schématique de l'ARN polymérase bactérienne sous forme holoenzyme.
- Figure 2 : Classification des facteurs σ bactériens
- Figure 3: Organisation d'un opéron et d'un promoteur bactérien
- Figure 4: Représentation schématique des interactions de l'ARN polymérase bactérienne avec un promoteur à σ^{70} .
- Figure 5: Mécanismes d'activations des promoteurs bactériens par des facteurs de transcription.
- Figure 6: Mécanismes de répression des promoteurs bactériens par des facteurs de transcription.
- Figure 7: Représentation schématique de l'organisation de clusters de promoteurs.
- Figure 8: Mécanismes d'interférences transcriptionnels entre promoteurs convergents
- Figure 9: Structure générale de la plate-forme fonctionnelle d'un intégron
- Figure 10: Organisation et séquence des sites *attI* des intégrons de multirésistance
- Figure 11: Structure d'une cassette de gène sous sa forme libre ou insérée au sein d'un intégron
- Figure 12: Structure d'un intégron complexe : In60
- Figure 13: Structure des différents IM 2 caractérisés à ce jour
- Figure 14: Structure des différents IM 3 caractérisés à ce jour
- Figure 15: Dynamique des mouvements de cassettes d'intégrons
- Figure 16: Structure d'ORF11 selon la nature de la première cassette.
- Figure 17: Efficacité du variant du promoteur P2 identifié dans ce travail
- Figure 18: Principe de la 5' Rapid Amplification of cDNA Ends (5'RACE).

- Figure 19: Caractérisation du promoteur PintI1 du gène de l'intégrase des IM 1
- Figure 20: Influence du promoteur P2 sur la séquence et l'activité du promoteur PintI1
- Figure 21: Influence du promoteur Pc sur l'activité du promoteur PintI1
- Figure 22: Influence du promoteur PintI1 sur l'activité du promoteur PcW
- Figure 23: Structure des IM 2
- Figure 24: Expression et régulation du gène *intI2* des IM 2
- Figure 25: Caractérisation des promoteurs des gènes des cassettes d'IM 2.
- Figure 26: Caractérisation du polymorphisme de PcA et Pc2B'
- Figure 27: Distribution des variants de Pc en fonction de la nature de la première cassette de l'IM 1 ou de l'organisme hôte
- Figure 28: Modèle de co-évolution des variants de Pc et de l'intégrase IntI1

Liste des annexes

- Contribution à une base de données dédiée aux intégrons : INTEGRALL
- Moura, A., M. Soares, C. Pereira, N. Leitao, I. Henriques et A. Correia. 2009.
« INTEGRALL: a database and search engine for integrons, integrases and gene cassettes». *Bioinformatics*. vol. 25, no 8, p. 1096-1098.

Glossaire

Abréviations	Définitions
3'CS	3' Conserved Region
5'CS	5' Conserved Region
5'RACE	5' Rapid Amplification of cDNA Ends (Amplification Rapide des Extrêmités d'ADNc)
A	Adénine/Adénosine
aa	Acide Aminé
ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire
ARN	Acide ribonucléique
ARNi	ARN interférent
ARNm	ARN messenger
ARNnc	ARN non codant
ARNr	ARN ribosomique
ARNt	ARN de transfert
ARNtm	ARN transfert-message
C	Cytidine/Cytosine
CR	Common Region
D	Adénine, Guanine ou Thymidine
dNTP	Désoxyribonucléotide Triphosphate
ddNTP	Didesoxyribonucléotides Triphosphate
E	ARN polymérase (apoenzyme)

Eσ	ARN polymérase (holoenzyme)
ECF	Extracytoplasmic Function
<i>E.coli</i>	<i>Escherichia coli</i>
FT	Facteur de Transcription
G	Guanine/Guanosine
H	Cytidine , Adénine ou Thymidine
ICE	<u>I</u> ntegrative <u>C</u> onjugative <u>E</u> lement
IM	Intégrons de Multirésistance
IM n	Intégrons de Multirésistance de classe n
INSD	<u>I</u> nternational <u>N</u> ucleotide <u>S</u> equence <u>D</u> atabase
IS	<u>I</u> nsertion <u>S</u> equence (Séquence d'Insertion)
IT	<u>I</u> nterférence <u>T</u> ranscriptionnelle
kDa	Kilo Dalton
M	Adénine ou Cytidine
Min.	Minutes
N	Adénine, Cytidine, Guanine ou Thymidine
ng	Nanogramme
nts	Nucléotides
ORF	<u>O</u> pen <u>R</u> eadin <u>G</u> <u>F</u> rame (Cadre Ouvert de Lecture)
Pb	Paire(s) de bases
PCR	<u>P</u> olymerase <u>C</u> hain <u>R</u> eaction (Réaction en Chaîne de la Polymérase)
<i>P. aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
ppGpp	Guanosine 3'-5' biphosphate
R	Adénine ou Guanine

RBS	<u>R</u> ibosome <u>B</u> inding <u>S</u> ite (Site de Liaison des Ribosomes)
sec	Secondes
SI	Super-intégron
T	Thymidine
TIR	Translation Initiation Region (Région d'initiation de traduction)
<i>V. cholerae</i>	<i>Vibrio cholerae</i>
W	Adénine ou Thymidine
X-gal	5-bromo-4-chloro-3-indoyl- β -Galactopyranoside
Y	Cytidine or Thymidine

Préambule

La découverte de la pénicilline en 1928 a été la première étape de la caractérisation de nombreuses molécules, appelées antibiotiques, capables à faible concentration de tuer ou inhiber la croissance d'un large spectre de micro-organismes. Les antibiotiques ont bouleversé l'histoire de la médecine en guérissant des infections autrefois fatales. Cependant, les bactéries ont su s'adapter en mettant en œuvre différents mécanismes de résistance (modification, substitution ou altération de la cible intracellulaire, modification ou expulsion de l'antibiotique). Ces bactéries résistantes ont ensuite disséminé sous l'influence de l'utilisation croissante des antibiotiques en médecine humaine. Par ailleurs, les antibiotiques ont aussi souvent été employés comme additif alimentaire pour stimuler la croissance des animaux, ce qui a augmenté la pression de sélection. Enfin, les bactéries possèdent différents supports génétiques très élaborés qui permettent le transfert des gènes de résistances entre bactéries parfois éloignées sur le plan phylogénique favorisant ainsi la dissémination de la résistance aux antibiotiques. Parmi les éléments génétiques mobiles, les plasmides et les transposons ont été largement étudiés. Les plasmides sont des molécules d'ADN indépendantes du chromosome et capables d'autoréplication. Ils peuvent être présents en multicopies chez une bactérie et des plasmides différents peuvent cohabiter au sein de la même bactérie. Chacun peut porter plusieurs gènes de résistance aux antibiotiques conférant ainsi en bloc une multirésistance. Les transposons sont des séquences d'ADN capables de promouvoir leur propre transfert d'un réplicon à l'autre (chromosome ou plasmide). Les transposons peuvent s'intégrer au chromosome de la bactérie. Ces éléments mobiles peuvent servir de véhicule aux gènes de résistance et permettre leur dissémination.

En 1989, Hatch Stokes et Ruth Hall ont décrit un nouvel élément baptisé « intégron » qui constitue un système original de capture et d'expression de gènes sous forme de cassettes. Aujourd'hui, les intégrons sont considérés comme des acteurs majeurs de la dissémination des résistances dans le monde bactérien.

Introduction

Chapitre 1 : Initiation de la transcription chez les procaryotes

La transcription est un processus biologique qui consiste en la copie de régions précises de l'ADN en molécules d'ARN appelés transcrits. Tous les ARN cellulaires sont synthétisés par transcription et il en existe deux types :

- Les ARN messagers (ARNm) relaient l'information génétique codée par les gènes qui sera ensuite, lors du processus de traduction, convertie en protéines.
- Les ARN non codants (ARNnc) regroupent des transcrits impliqués dans d'autres fonctions, notamment dans la traduction (ARNt, ARNr, ARNtm) ou des fonctions de régulations (ARN interférent, ARN antisens).

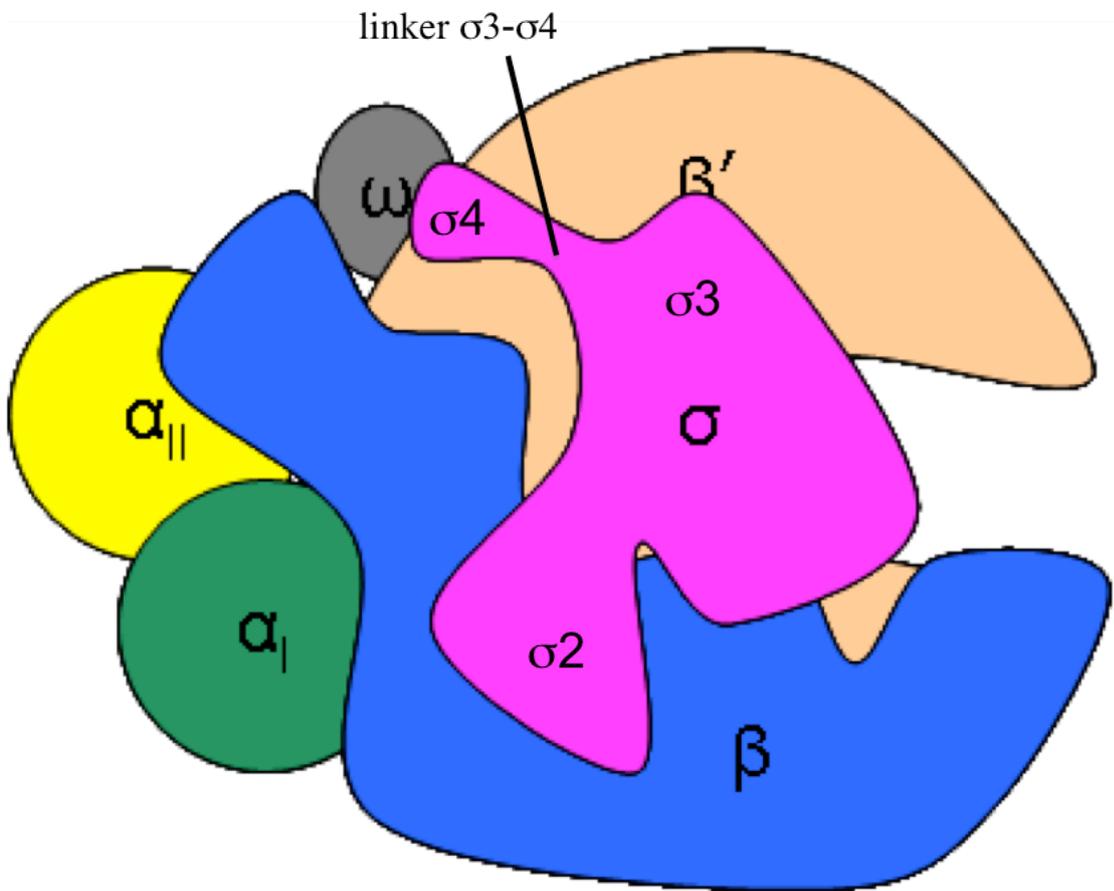


Figure 1 : Représentation schématique de l'ARN polymérase bactérienne sous forme holoenzyme.

Chaque sous unité formant l'holoenzyme (α_I et α_{II} , β , β' , σ et ω) est représentée d'une couleur propre, et son nom est indiquée.

Les positions des régions conservées du facteur σ ($\sigma 2$ à $\sigma 4$) sont précisées.

A. L'ARN polymérase

1. Structure de l'ARN polymérase

Tous les ARN d'une cellule procaryote sont synthétisés par une même ARN polymérase. Celle-ci existe d'abord sous la forme d'une apoenzyme « core » (E) de structure pentamérique $\alpha_2\beta\beta'\omega$ d'environ 450 kDa. Elle présente une conformation tri-dimensionnelle en « pince de crabe » née de l'assemblage des sous unités β et β' qui portent les résidus catalytiques (Borukhov et Nudler, 2008) (Figure 1). La sous unité α participe à cette dimérisation, et la sous unité ω en facilite le repliement telle une protéine chaperone.

Ce complexe enzymatique n'est pas fonctionnel pour initier la transcription ; en effet, s'il présente une certaine affinité pour les molécules d'ADN, il n'y distingue pas les séquences spécifiques des promoteurs. L'association réversible d'une sixième sous unité σ (ou facteur σ) sur l'enzyme « core » crée une holoenzyme ($E\sigma$) dont l'affinité générale pour l'ADN est réduite au profit d'une forte spécificité pour les promoteurs. Seule l'enzyme $E\sigma$ se fixe aux promoteurs d'où elle initie la transcription.

2. Le facteur σ

a- Définition

Le facteur σ est une sous unité de l'ARN polymérase essentielle à la transcription, et dont la taille varie de 20 à 70 kDa. En se liant de façon réversible à l'enzyme « core », le facteur σ diminue son affinité globale pour l'ADN au profit d'une affinité sélective bien plus forte pour les promoteurs. Les bactéries contiennent un nombre variable de facteurs σ , de 1 chez *Mycoplasma genitalium* à 65 chez *Streptomyces coelicolor*. Un seul assure la transcription des gènes dits « de ménage », qui sont essentiels à la survie cellulaire et sont exprimés de façon constitutive. En réponse à de nouvelles conditions de croissance, d'autres facteurs σ , qualifiés d'alternatifs se lient à l'apoenzyme à la place du facteur σ principal. Ils forment ainsi de

nouvelles holoenzymes $E\sigma$ capables de transcrire les gènes requis dans ces nouvelles conditions.

b- Structure

Deux grandes familles structurales de facteurs σ ont été identifiées à ce jour : la famille du facteur σ^{70} et celle du facteur σ^{54} . Le facteur σ^{70} (70 kDa) est composé de quatre domaines conservés numérotés $\sigma 1$ à $\sigma 4$ et d'un « linker » de 33 résidus reliant les domaines $\sigma 3$ et $\sigma 4$ entre-eux (Paget et Helmann, 2003). Les domaines $\sigma 2$ à $\sigma 4$ sont exposés à la surface de la protéine et incluent chacun une hélice qui interagit avec des régions spécifiques du promoteur. La région $\sigma 1$ est moins bien conservée au sein des facteurs σ et son rôle dans la transcription est peu connu. La structure du facteur σ^{54} est différente et lui permet de s'associer à ses promoteurs spécifiques en absence de l'enzyme « core » (Wigneshweraraj *et al.*, 2008).

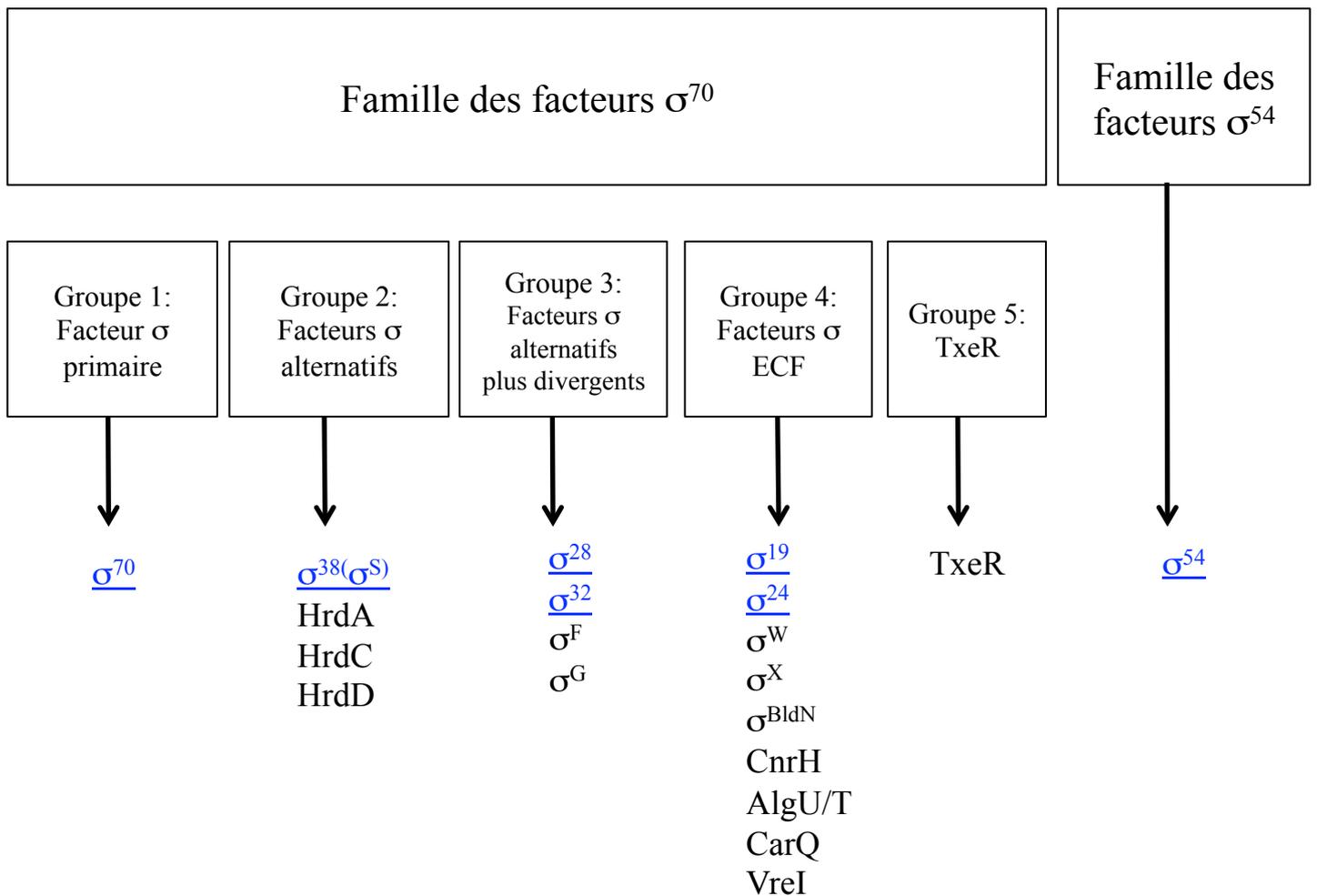


Figure 2: Classification des facteurs σ bactériens

Deux familles structurales de facteur σ se distinguent, la famille des facteurs σ^{70} se divise en 5 groupes. Des membres de chaque groupe sont indiqués, dont les sept qui ont été caractérisés à ce jour chez *E.coli* (leurs noms sont soulignés)

c- Diversité

La grande majorité des facteurs σ décrits appartiennent à la famille structurale de σ^{70} . Sur la base de leur divergence avec σ^{70} , on en distingue quatre groupes principaux (Helmann, 2002) (Figure 2) :

- (i) Le **Groupe 1** compte l'ensemble des facteurs σ principaux, soit σ^{70} chez *Escherichia coli* (*E.coli*) et ses orthologues, qui assurent la transcription des gènes essentiels dits « de ménage ». La synthèse de ces facteurs n'est pas régulée.
- (ii) Le **Groupe 2** comprend les facteurs σ alternatifs les plus apparentés à σ^{70} . Ainsi, le facteur σ^{38} (plus connu sous le nom de σ^S) est si proche de σ^{70} qu'un grand nombre de promoteurs sont reconnus et transcrits par les deux facteurs σ . Ces facteurs sont davantage produits en conditions de stress nutritif, induisant par exemple l'entrée en phase stationnaire de croissance.
- (iii) Le **Groupe 3** rassemble des facteurs σ alternatifs dont la structure est plus éloignée de σ^{70} , puisqu'ils ne présentent pas les régions $\sigma 1$ et $\sigma 3$. Ils sont impliqués dans la réponse aux chocs thermiques (σ^{32}), la biogénèse des flagelles (σ^{28}) ou la sporulation (σ^F , σ^G chez *Bacillus*). Ces facteurs sont très régulés, au niveau de leur synthèse, par protéolyse ou des inhibiteurs « anti- σ ».
- (iv) Le **Groupe 4** regroupe de nombreux facteurs σ (aussi appelés ECF pour ExtraCytoplasmic Function) Ils reconnaissent un petit nombre de promoteurs et interviennent dans la transcription de gènes impliqués dans un certain nombre de processus physiologiques : homéostasie (σ^{24} , σ^X), teneur en fer ferrique (σ^{19}), virulence (VreI), résistance et perméabilité aux antibiotiques (σ^W), adaptation à une haute salinité (σ^M) ou différenciation cellulaire (σ^{BldN}) (Llamas *et al.*, 2009).

Facteur σ	Gène	Fonctions	Séquence consensus du promoteur	Références
σ^{70}	<i>rpoD</i>	Facteur σ majeur en phase exponentielle de croissance	TTGACA-N ₁₇ -TATAAT	Hawley et McClure, 1983
σ^{38}	<i>rpoS</i>	Facteur σ majeur en phase stationnaire de croissance, réponse à des chocs oxydatifs, osmotiques, acides, thermiques ou UV. Gènes de virulence chez <i>Salmonella</i> et <i>Yersinia</i>	TTGACA-N ₁₆ -CTATACT	Lacour, Kolb et Landini, 2003
σ^{32}	<i>rpoH</i>	Transcription des protéines de choc thermique induit par les stress cytoplasmiques	CTTGA-N ₁₄ -GNCCCCATNT	Wang et deHaseth, 2003
σ^{28}	<i>fliA</i>	Transcription des gènes tardifs de flagelle	CTAAA-N ₁₅ -GCCGATAA	Koo <i>et al.</i> , 2009
σ^{24}	<i>rpoE</i>	Réponse au stress périplasmique, maintien de l'enveloppe cellulaire	GGAACCT-N ₁₅₋₁₉ -GTC(A/T) ¹ AA	Rhodus <i>et al.</i> , 2006
σ^{19}	<i>fecI</i>	Captage et transport du fer	GAAAAT-N ₁₅ -TGTCCCT ²	Enz <i>et al.</i> , 2003
σ^{54}	<i>rpoN</i>	Métabolisme de l'azote, modification des ARN, synthèse des flagelles, réponse à l'infection phagique	CTGGNA-N ₆ -TTGCA	Kelly, Ferguson et Hoover, 2000

Tableau 1 : Liste et caractéristiques des sept facteurs σ d'*Escherichia coli* identifiés à ce jour.

N : A, C, G ou T.

¹ Les bases A et T sont également représentées à cette position.

² Séquence du seul promoteur connu pour être transcrit par le facteur σ^{19}

De façon intéressante, alors que la plupart des protéobactéries possèdent moins de 5 facteurs σ ECF, 19 ont été identifiés chez *Pseudomonas aeruginosa* (Potvin, Sanschagrín et Levesque, 2008).

- (v) Le **Groupe 5** a été établi récemment pour prendre en compte des protéines jouant un rôle analogue au facteur σ^{70} mais ne partageant que peu d'identité. Ils sont impliqués dans la synthèse de toxines (TxrR) ou de bactériocines (TetR).

d- Les facteurs σ d'*Escherichia coli*

Sept facteurs σ , décrits dans le tableau 1, coexistent chez *E. coli*, le facteur σ^{70} jouant le rôle de facteur principal.

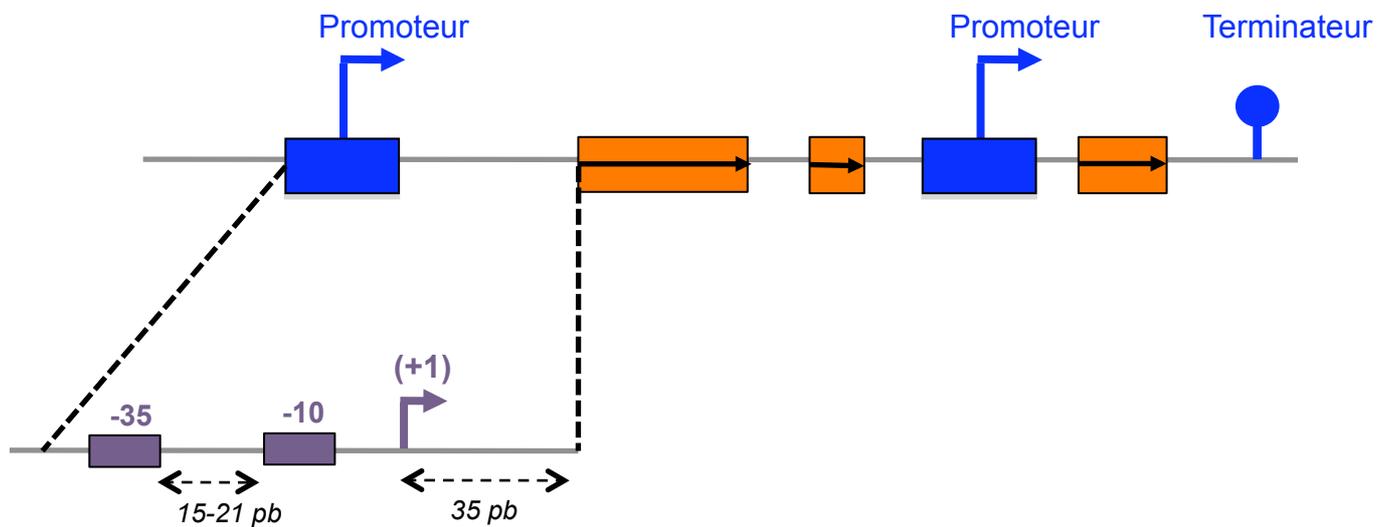


Figure 3: Organisation d'un opéron et d'un promoteur bactérien.

Les ORF de l'opéron, représentés par des rectangles oranges, sont transcrits depuis un même promoteur, symbolisé par une flèche cassée bleue. La transcription se termine au niveau du termineur (symbole ). Cette organisation n'exclut pas la présence d'autres promoteurs actifs pour l'expression des gènes.

L'organisation du promoteur est détaillée: les éléments -35 et -10 sont représentés par des rectangles violets, le (+1) de transcription par une flèche cassée violette. Les distances entre ces éléments, ainsi qu'entre le (+1) de transcription et le début de l'ORF sont indiquées.

B. Les promoteurs bactériens

Un promoteur est la région en amont d'une séquence codante indispensable pour initier, localiser et orienter sa transcription. Il est spécifiquement reconnu et lié par l'ARN polymérase. Un seul promoteur peut contrôler la transcription de plusieurs gènes (organisation en opéron, Figure 3) et réciproquement, plusieurs promoteurs peuvent contribuer à la transcription d'un seul gène.

1. Structure des promoteurs bactériens

Un promoteur se termine par le site (+1) d'initiation de transcription, qui correspond à la première base copiée (ou transcrite) en ARN. En général, il s'agit d'un A ou d'un G, plus rarement un C ou un T. Il est situé en moyenne à 35 paires de bases (pb) en amont du codon d'initiation (ou codon START) du gène transcrit, rarement à plus de 100 pb (sauf pour les cas des gènes transcrits en opéron)(Shultzaberger *et al.*, 2007).

Les promoteurs reconnus par les facteurs σ de la famille σ^{70} sont organisés en deux éléments fonctionnels majeurs initialement nommés « -35 » et « -10 » en raison de leur position par rapport au site (+1) (Figure 3).

La séquence type (ou séquence consensus) de ces éléments a été bien établie pour les promoteurs reconnus par le facteur σ^{70} d'*E. coli* : TTGACA pour l'élément -35, TATAAT pour l'élément -10, avec un espacement entre ces deux éléments pouvant varier de 15 à 21 pb, mais qui est généralement de 16 à 18 pb (Hawley et McClure, 1983 ; Shultzaberger *et al.*, 2007) (Figure 3).

L'élément -35 consensus est légèrement différent chez *P. aeruginosa*: TTGACC (Potvin, Sanschagrin et Levesque, 2008). Quelques données suggèrent également une divergence de ces signaux chez *Acinetobacter* (Kishan et Hillen, 1990 ; Segal et Elisha, 1999).

2. Efficacité et force des promoteurs

a- Règle principale

Tous les promoteurs ne sont pas également compétents pour initier la transcription d'un gène. Globalement, un promoteur est d'autant plus efficace (ou fort) qu'il présente d'identité avec les séquences consensus. La variation d'une base par rapport au consensus affecte d'autant l'efficacité du promoteur qu'elle est située à une position fortement conservée.

b- Cas des promoteurs à σ^{70}

Les deux premières bases (« TA ») et le dernier « T » de l'élément -10 d'un promoteur à σ^{70} sont particulièrement conservés et jouent un rôle crucial pendant l'initiation de la transcription (Schroeder, Choi et DeHaseth, 2007) ; la mutation de ces seules bases peut suffire à inactiver totalement un promoteur à σ^{70} . Moyle et coll. ont établi, pour chaque position des éléments -35 et -10, la hiérarchie des bases en fonction de la force du promoteur généré (Moyle, Waldburger et Susskind, 1991). Cependant, au delà des éléments -35 et -10, d'autres régions contribuent à la force des promoteurs de σ^{70} :

- La taille et la composition de l'espaceur entre les éléments -35 et -10 (« espaceur »). Bien qu'il ne joue pas un rôle direct dans la transcription, l'espaceur est un élément clé de l'architecture du promoteur. Ainsi, les promoteurs fonctionnant avec des espaceurs de 15 ou 21 pb ou qui sont riches en G/C sont peu efficaces (Hook-Barnard et Hinton, 2009). Pour autant, il n'existe pas de séquence consensus pour cette région, très variable selon les promoteurs.
- Environ 20% des promoteurs caractérisés d'*E.coli* présentent un motif TGN (N : A, T, C ou G) immédiatement en amont de l'élément -10 (Burr *et al.*, 2000). Cette région, connue sous le nom de « -10 étendue », peut compenser l'absence d'une région -35 canonique ou augmenter jusqu'à 20 fois la force d'un promoteur (Kumar *et al.*, 1993).

Un promoteur atypique fonctionnant sans élément -10 mais avec un élément -35 et un motif TGN a également été décrit (Hook-Barnard, Johnson et Hinton, 2006).

- Les promoteurs bactériens les plus forts, comme ceux pour la synthèse des ARN ribosomiaux, présentent une région très riche en A/T juste en amont de leur élément -35 (consensus : AAAWWTWTTTTNNNAAANNN, W : A ou T, N : A, C, T ou G). Cet élément nommé « UP » peut augmenter jusqu'à 100 fois la force d'un promoteur (Estrem *et al.*, 1998).
- Enfin, une étude récente suggère que la base située juste en aval de l'élément -10 (position -6) participe aussi à l'initiation de la transcription sans que la base préférentielle à cette position n'ait été identifiée (Haugen *et al.*, 2008).

c- Cas des promoteurs des autres facteurs σ

Les séquences consensus des promoteurs des autres facteurs σ d'*E.coli* sont représentées dans le tableau 1.

3. Polymorphisme des promoteurs bactériens

Le promoteur d'un même gène peut, selon les allèles ou les souches d'une même espèce bactérienne, présenter des mutations dans ses éléments fonctionnels. La plupart du temps, ces variations modifient son efficacité et donc le niveau d'expression du ou des gènes dont il assure l'expression. Par exemple, chez *E.coli*, des variations du promoteur du gène *ampC*, qui code une céphalosporinase, sont à l'origine d'une augmentation du spectre de résistance vis à vis des β -lactamines (Caroff *et al.*, 1999). Le niveau de résistance des souches est alors directement lié à la séquence du promoteur. De la même façon, les gènes *bla*_{TEM} codant des β -lactamases peuvent être transcrits par quatre promoteurs différents dont les forces sont variables (Lartigue *et al.*, 2002).

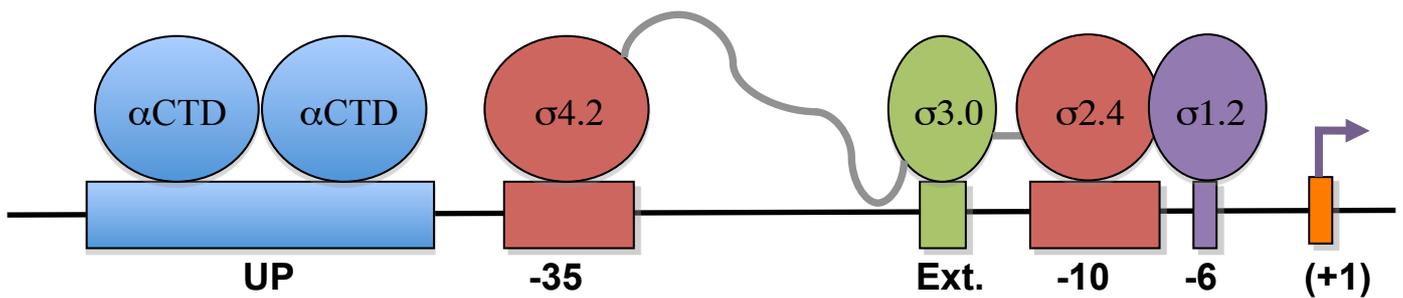


Figure 4: Représentation schématique des interactions de l'ARN polymérase

bactérienne avec un promoteur à σ^{70} .

Les éléments fonctionnels du promoteur bactérien (UP, élément -35, Ext., élément -10 et base -6) sont représentés par des cadres de la même couleur que les régions de l'ARN polymérase (cercles) avec lesquelles ils interagissent. « Ext » = Extension de l'élément -10, ou motif « TGN » immédiatement en amont de l'élément -10;

« α CTD » = région C-terminale de la sous-unité α ; « $\sigma^{4.2}$, $\sigma^{3.0}$, $\sigma^{2.4}$ et $\sigma^{1.2}$ » désignent les régions conservées du facteur σ^{70} . D'après Haugen et al., 2008.

D'autres gènes, en dehors de ceux codant des résistances aux antibiotiques, présentent des variations de leurs promoteurs, dans l'élément -10 (de Vin *et al.*, 2005) ou dans l'espaceur (Kulaeva *et al.*, 1998 ; Rauzier *et al.*, 1999).

Enfin, les promoteurs de certains gènes peuvent être modifiés par l'intégration d'une séquence d'insertion (IS), le plus souvent parce que cette IS fournit un élément -35 alternatif, plus rarement un promoteur complet (Pour revue : (Depardieu *et al.*, 2007)). Par exemple, le promoteur du gène *ampC* est 20 fois plus fort suite à l'insertion de l'IS2, car celle-ci amène un élément -35 mieux positionné par rapport à l'élément -10 que ne l'est l'élément -35 original (17 pb au lieu de 16 pb)(Jaurin et Normark, 1983).

C. Déroulement de la transcription chez les procaryotes

La transcription est communément décrite en 3 étapes : initiation, élongation du transcrit, et terminaison. L'initiation de la transcription, étape limitante et la plus contrôlée du processus, est détaillée ici.

1. Fixation de l'ARN polymérase sur le promoteur

L'ARN polymérase se lie à un promoteur via des contacts spécifiques entre les nucléotides les plus conservés des éléments -35 et -10 et les régions $\sigma_{4.2}$ et $\sigma_{2.4}$ du facteur σ , respectivement (Figure 4). Les promoteurs avec une région -10 étendue contactent aussi la région $\sigma_{3.0}$; ceux avec un élément UP, les sous unités α de l'ARN polymérase (Figure 4). De plus, la base située juste en aval de l'élément -10 interagit avec la région $\sigma_{1.2}$ (Figure 4). L'ARN polymérase fixée sur le promoteur est un « complexe fermé » stable qui recouvre les 55 premières bases en amont du (+1) de transcription.

2. Isomérisation de l'enzyme sur le promoteur

Le contact de l'ARN polymérase avec la base à la position -11 entraîne une série de changements de conformations aboutissant à l'ouverture du duplex ADN en une « bulle » de transcription qui s'étend des bases -11 à +2. Le complexe, nommé « ouvert » à ce stade, est stabilisé en raison d'une zone de contact prolongée. Ces isomérisations irréversibles constituent l'étape limitante de la transcription, d'où l'importance de la nature de la base -11.

3. Transcription abortive et sortie du promoteur

Dans un premier temps, l'ARN polymérase incorpore les nucléotides (nts) complémentaires aux bases +1 et +2 et synthétise d'abord un petit transcrit qui, lorsqu'il atteint une taille d'environ 9-11 nts, déstabilise le complexe, entraînant la dissociation de l'enzyme (clairance) et l'élagage du facteur σ : l'enzyme qui poursuit la transcription est donc un complexe d'élongation pentamérique $\alpha_2\beta\beta'\omega$.

4. Elongation et terminaison de la transcription

La synthèse du transcrit se poursuit à une vitesse d'environ 40 nts/sec, ponctuée de pauses réversibles. La bulle de transcription (environ 17 pb) migre en parallèle pour que la polymérase puisse accéder au brin matrice. La transcription se termine quand le complexe ternaire ARN polymérase-ADN-ARN est suffisamment déstabilisé. Deux types de séquences conduisent à la terminaison de la transcription :

- la séquence *rut* transcrite est un site de liaison pour une hélicase qui sépare ADN et ARN, entraînant la dissociation du complexe (Richardson, 2002).
- une séquence palindromique riche en G/C immédiatement suivie d'une série de T : le transcrit de cette région forme une tige-boucle refermée par un duplex hybride A :: U instable, une situation dans laquelle le complexe ternaire se dissocie (Yarnell et Roberts, 1999).

D. Régulation de la transcription des gènes bactériens

Chaque étape de l'expression d'un gène est susceptible d'être régulée. En effet, la cellule bactérienne orchestre un réglage fin du niveau d'expression de la plupart de ses gènes de façon à ce que leur produit soit disponible en quantités et au moment requis. Ces régulations jouent également un rôle majeur dans l'aptitude de la cellule à s'adapter à un changement d'environnement (« stress »).

Dans ce chapitre, nous présenterons les différents modes de régulations notamment celles liées à l'initiation de la transcription ou à la proximité de plusieurs promoteurs, situation fréquente chez les intégrons (voir chapitre 2).

1. Régulation de l'initiation de la transcription

a- Modulation de la disponibilité en ARN polymérase fonctionnelles

Les gènes codant les sous unités de l'ARN polymérase sont exprimés constitutivement ou bien autorégulés pour en maintenir une teneur constante (Dykhorn, St Pierre et Linn, 1996). En revanche, l'assemblage de l'ARN polymérase active peut être perturbé par des facteurs protéiques dits anti- σ qui se lient spécifiquement sur un type de facteur σ libre, l'empêchant ainsi de se lier à l'enzyme core et inhibant donc indirectement la transcription des gènes dont l'expression dépend de ce facteur σ . (Studemann *et al.*, 2003). Ce mode d'action permet de modifier le profil d'expression en bloc de nombreux gènes, par exemple en réponse à un stimulus environnemental. Les genres bactériens pourvus de nombreux facteurs σ alternatifs (*Bacillus*, *Streptomyces*) codent même des facteurs « anti-anti- σ » dirigés contre les facteurs anti- σ pour restaurer l'expression des gènes inhibés (Najafi, Willis et Yudkin, 1995).

Facteur de transcription	Acronyme	Effet	Cibles	Séquence consensus du site de liaison
<u>A</u> naerobic <u>r</u> espiratory <u>c</u> ontrol	ArcA	Activateur/Répresseur	Opérons du métabolisme central. Gènes du métabolisme anaérobie (activateur), Gènes du métabolisme aérobie (répresseur)	WGTTAATTAW
<u>C</u> yclic AMP <u>R</u> eceptor <u>P</u> rotein	CRP	Activateur/Répresseur	Dizaines de gènes dont ceux des opérons arabinose, lactose et galactose	TGTGA-N ₆ -TCANA
<u>F</u> actor for <u>I</u> nversion <u>S</u> timulation	FIS	Activateur/Répresseur	Gènes d'adhésion, de colonisation, de virulence (activateur) Gènes de protéines du nucléoïde (répresseur)	Aucune
<u>F</u> umarate and <u>N</u> itrate réduction <u>R</u> egulator	FNR	Activateur/Répresseur	Gènes des systèmes respiratoires anaérobies fumarate, nitrate, nitrite, DMSO (activateur) Gènes du système respiratoire aérobie (répresseur)	TTGAT-N ₄ -ATCAA
<u>I</u> ntegration <u>H</u> ost <u>F</u> actor	IHF	Activateur/Répresseur	Gènes <i>fis</i> , de transposition du bactériophage λ , de division cellulaire, de la réplication, de la réparation ADN, gènes ribosomaux.	WATCAANNNTTR
<u>L</u> eucine-responsive <u>R</u> egulator Protein	Lrp	Activateur/Répresseur	10% des gènes d' <i>E.coli</i> , gènes du métabolisme des acides aminés et de synthèse des pili, gène <i>lrp</i> (rétrocontrôle)	YAGHAWATTWTDCTR
<u>N</u> itrate <u>R</u> egulation	NarL	Activateur/Répresseur	Gènes de catabolisme du nitrate et du nitrite (activateur) ; Gènes des autres systèmes de respiration anaérobie et de fermentation (répresseur)	TACYYMT

Tableau 2 : Liste et caractéristiques des sept facteurs de transcription généraux d'*E.coli*.

M : A ou C ; R : A ou G ; W : A ou T ; Y : C ou T ; D : A, G ou T ; H : A, C ou T ; N : A, C, G ou T

D'après Kim et al., 1999 (ArcA) ; Busby et Ebright, 1999 (CRP) ; Muskhelishvili et Travers, 2003 (FIS) ; Green et Paget, 2004 (FNR) ; Hales, Gumpert et Gardner, 1994 (IHF) ; Brinkman et al., 2003 (Lrp) et Unden et Bongaerts, 1997 (NarL).

b- Influence de la topologie de l'ADN

Le chromosome bactérien est maintenu compacté dans la cellule du fait de surenroulements. Une douzaine de protéines dites du nucléoïde interviennent dans cette compaction, un peu à la façon des histones chez les organismes eucaryotes ; les mieux connues sont H-NS, HU, IHF, Dps, StpA et Lrp (Pour revue : (Dorman, 2009)). Le niveau de surenroulement, ou topologie, de la molécule d'ADN influence de nombreux processus physiologiques dont la réplication et la transcription. Pendant la transcription, l'accès aux bases est assuré par une topoisomérase qui « déroule » localement l'ADN. D'autre part, chaque variation de la topologie peut modifier la capacité de l'ARN polymérase à se fixer au promoteur ou à évoluer en complexe ouvert, et donc affecter la transcription. La structure du promoteur conditionne l'évolution de sa force en réponse à un changement local de la topologie (Pour revue : (Travers et Muskhelishvili, 2005)).

c- Régulations par des facteurs de transcription

Plus de 300 gènes du chromosome d'*E.coli* codent des protéines qui se fixent aux promoteurs et modifient leur niveau d'activité (Perez-Rueda et Collado-Vides, 2000). Ces molécules, appelées facteurs de transcription (FT), régulent un nombre variable de gènes, de 1 à plusieurs centaines. Sept FT dits généraux, décrits dans le tableau 2, réguleraient jusqu'à 50% des gènes d'*E.coli*. Ils assurent ainsi la coordination de l'expression des gènes d'une même voie métabolique ou d'un système respiratoire. Les FT interviennent sur l'initiation de la transcription d'un gène en se fixant sur un site opérateur qui en général est situé à proximité du promoteur ou le chevauche. A titre d'exemple, les séquences consensus des opérateurs des sept FT généraux d'*E.coli* sont présentées dans le tableau 2.

Dans ce chapitre, nous décrirons brièvement les principaux modes d'actions des FT.

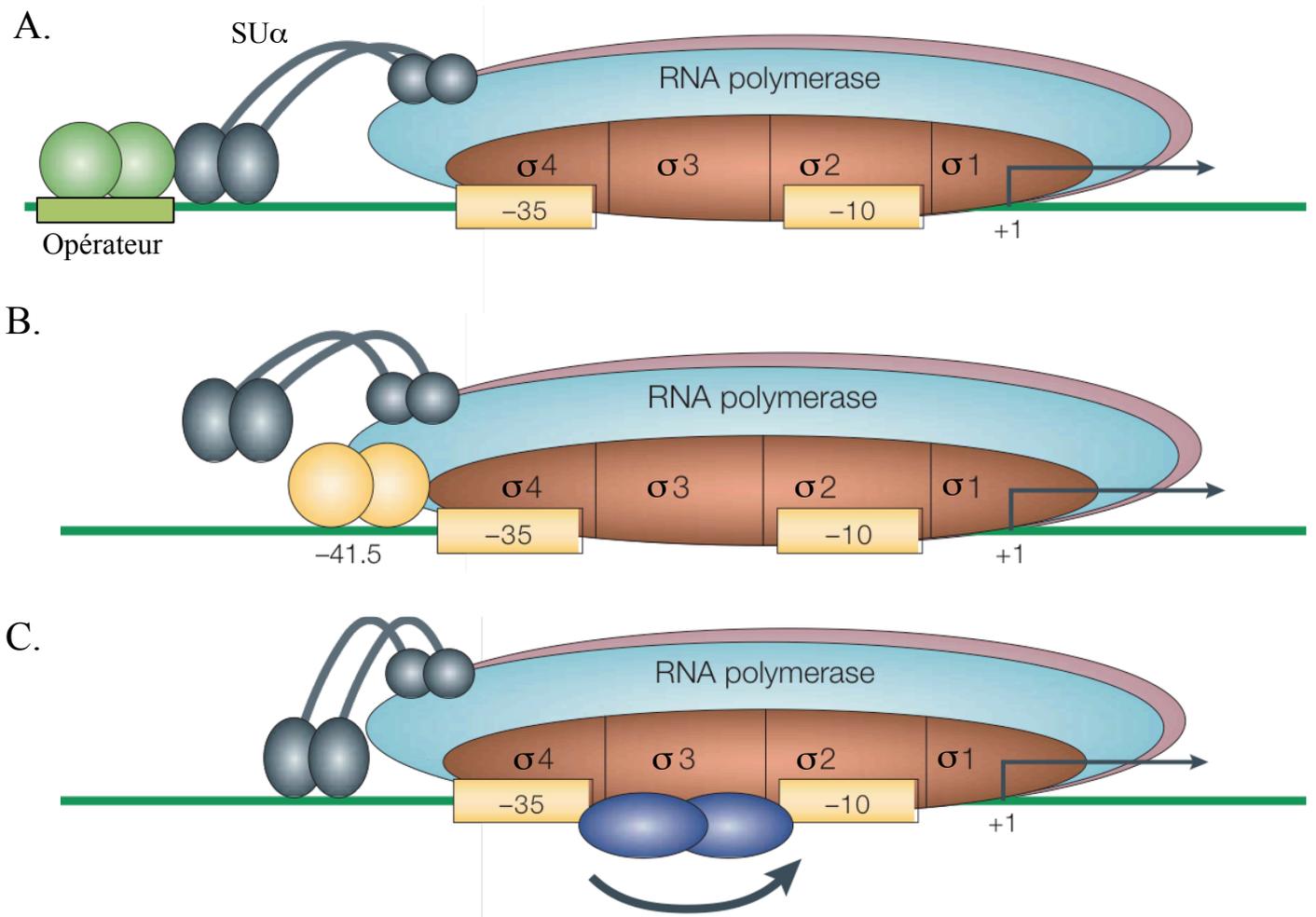


Figure 5: Mécanismes d'activations des promoteurs bactériens par des facteurs de transcription.

A. Les activateurs de classe 1 (double cercle vert) se lient en dimère sur leur opérateur et interagissent avec la sous-unité α de l'ARN polymérase ($SU\alpha$), ce qui facilite le recrutement de l'ARN polymérase au niveau du promoteur

B. Les activateurs de classe 2 (double cercle jaune) se lient sur un site immédiatement en amont ou chevauchent l'élément -35 et recrutent des ARN polymérase via des interactions avec le facteur σ .

C. Les activateurs de classe 3 (double cercle bleu) se lient au promoteur pour modifier sa conformation, ce qui augmente son efficacité. D'après Browning et Busby, 2004

- Facteurs de transcriptions positifs (activateurs)

Les activateurs influencent positivement l'initiation de la transcription :

- En favorisant le recrutement de l'ARN polymérase via des interactions avec sa sous unité α (activateurs de classe 1 ; Figure 5A). Les opérateurs de ces activateurs sont en amont de l'élément -35 du promoteur régulé. La régulation du promoteur *lac* par le FT CRP est le paradigme des activations de classe 1 (Ebright, 1993).
- En favorisant le recrutement de l'ARN polymérase via des interactions avec la région σ_4 du facteur σ (activateurs de classe 2 ; Figure 5B). Son opérateur chevauche alors l'élément -35 du promoteur (par exemple : CI sur le promoteur PRM du bactériophage λ) (Nickels *et al.*, 2002).
- En modifiant l'architecture du promoteur régulé (activateurs de classe 3). Les FT de la famille de MerR, par exemple, se lient à un opérateur situé dans l'espaceur d'un promoteur σ^{70} , ce qui corrige l'espacement entre les éléments -35 et -10 (Pour revue : (Brown *et al.*, 2003) ; Figure 5C).

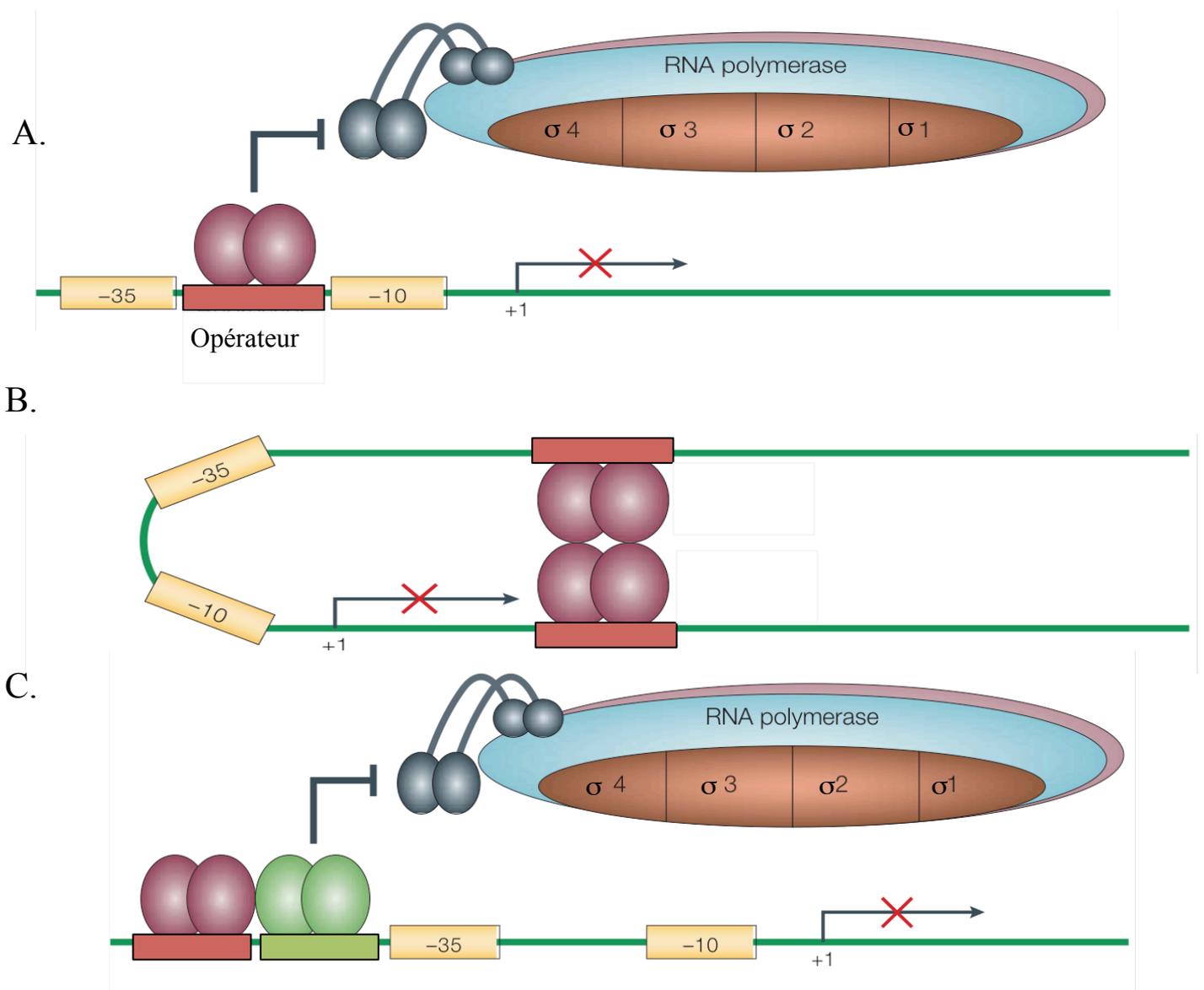


Figure 6: Mécanismes de répression des promoteurs bactériens par des facteurs de transcription.

A. La liaison de répresseurs (double ovale rose) sur le promoteur empêche le recrutement d'ARN polymérase sur le promoteur.

B. Les répresseurs masquent le promoteur à l'ARN polymérase quand ils sont liés à leur opérateurs.

C. La liaison de répresseurs (double ovale rose) empêchent un FT activateur (double ovale vert) de fonctionner sur le promoteur.

D'après Browning et Busby, 2004

- Facteurs de transcriptions négatifs (répresseurs)

D'autres FT agissent en réprimant l'expression des gènes :

- La liaison de certains FT sur leur opérateur empêche par encombrement stérique la fixation de l'ARN polymérase sur le promoteur (Figure 6A). C'est ainsi par exemple que la protéine LexA réprime, dans les conditions optimales de croissance, l'expression de gènes impliqués dans la réponse au stress oxydatif (Quinones *et al.*, 2006).
- D'autres FT modifient la conformation de l'ADN lorsqu'ils sont liés à leur opérateur, de sorte à masquer le promoteur à l'ARN polymérase (Figure 6B). Par exemple, des molécules du FT GalR liées à des opérateurs en aval du promoteur *galP1* interagissent avec d'autres molécules de GalR liées à des opérateurs en amont : cette interaction entraîne une courbure de l'ADN qui isole le promoteur dans une boucle, le rendant moins accessible aux ARN polymérase (Roy *et al.*, 2004).
- Un FT répresseur peut également empêcher l'activation d'un promoteur par un FT activateur (Figure 6C). Par exemple, l'interaction du FT répresseur CytR avec le FT activateur CRP lié à un promoteur empêche le fonctionnement de ce dernier (Roy *et al.*, 2004).

Des promoteurs qui possèdent plusieurs opérateurs peuvent être régulés par plusieurs FT agissant de façon coopérative, synergique ou antagoniste. Ces situations complexes permettent d'adapter le niveau d'expression d'un gène en réponse à un ou plusieurs signaux environnementaux.

2. Inhibition compétitive de l'ARN polymérase : les alarmones

En réponse à un changement de l'environnement, la cellule bactérienne peut être amenée à modifier complètement son métabolisme via une réorganisation générale de l'expression de ses gènes. Elle opère via de petites molécules ligands nommées alarmones qui sont des régulateurs pléiotropiques de transcription. La mieux connue est la guanosine 3'-5' biphosphate (ppGpp), synthétisée en situation de carence en acide aminé, dans le cadre de la réponse stringente. La ppGpp se fixe directement sur les ARN polymérase au stade de « complexe ouvert » au cours de l'initiation de la transcription. Cette interaction a un effet négatif sur l'activité des promoteurs à σ^{70} , en particulier ceux des opérons ribosomiques mais un effet positif sur ceux des gènes de réponse aux stress (Srivatsan et Wang, 2008). De plus, le relargage des ARN polymérase σ^{70} rend l'enzyme core plus disponible aux facteurs σ alternatifs σ^{38} , σ^{54} et σ^{32} et stimule ainsi l'expression des gènes de réponse aux stress nutritionnels ou thermiques. Cette redistribution des ARN polymérase sur les promoteurs, combinée à d'autres effets inhibiteurs de ppGpp sur la réplication et la traduction, conduit à limiter la croissance de la population au profit du maintien de la survie (Nystrom, 2004).

L'ARN 6S est un second type d'alarmone mis en évidence plus récemment : cette molécule agit comme un inhibiteur compétitif de la transcription en se liant à la région $\sigma 4.2$ du facteur σ^{70} à la place de l'élément -35 du promoteur (Cavanagh *et al.*, 2008 ; Klocko et Wassarman, 2009). Trotochaud et Wassarman ont remarqué que souvent les promoteurs sensibles à l'ARN 6S présentent une région -10 étendue (Minakhin et Severinov, 2003), ce qui porterait à une centaine le nombre de promoteurs d'*E. coli* sensibles à l'ARN 6S (Mitchell *et al.*, 2003).

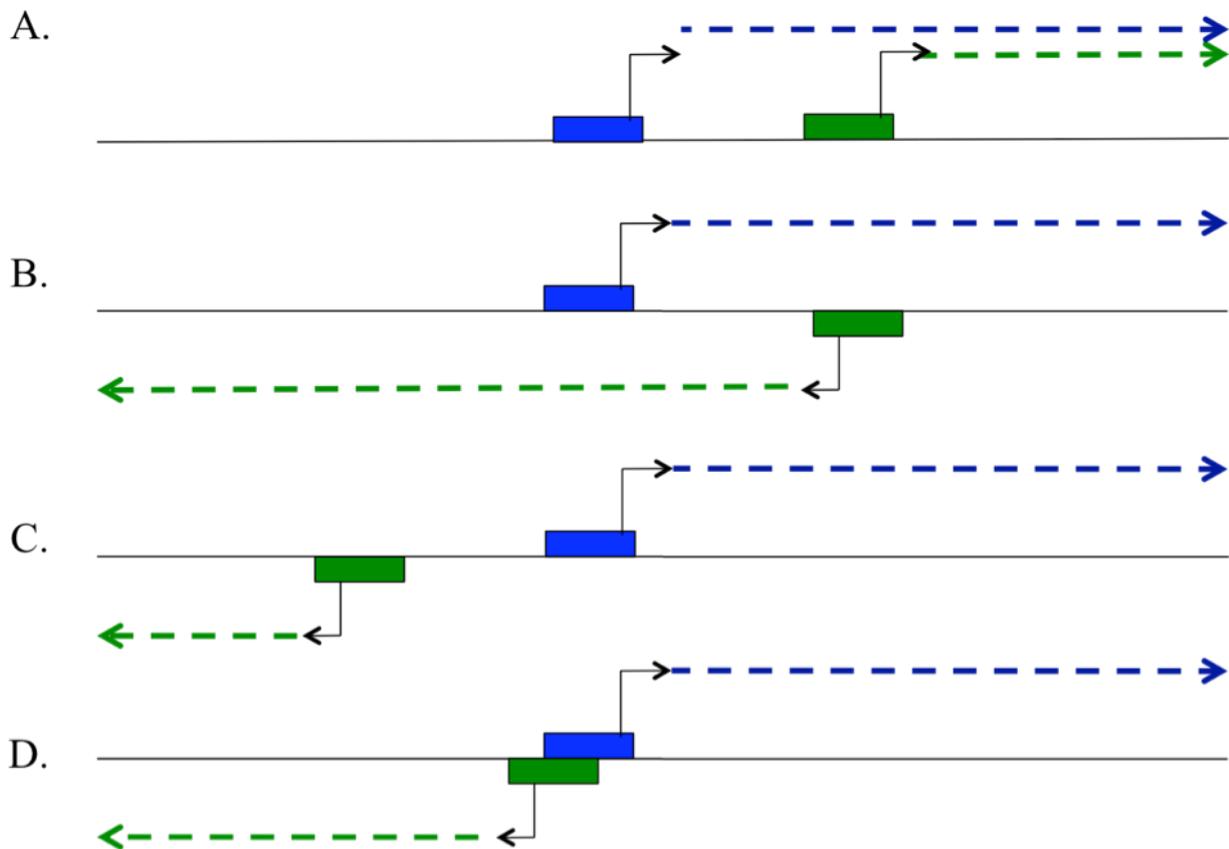


Figure 7: Représentation schématique de l'organisation d'un cluster de deux promoteurs:

A. *En tandem*

B. *Convergent*

C. *Divergent*

D. *Divergent et chevauchants*

Les promoteurs sont représentés par des cadres bleus ou verts, le sens de transcription et la position du (+1) par une flèche cassée. Des flèches pointillées symbolisent les transcrits issus des promoteurs de même couleur. D'après Shearwin, Callen et Egan, 2005.

D'autre part, les promoteurs transcrits par le facteur σ^{38} , bien que très similaires aux promoteurs à σ^{70} , sont quant à eux régulés positivement par l'ARN 6S. Le mécanisme moléculaire par lequel s'opère cette activation n'est pas documenté (Wassarman, 2007).

3. Régulation de la transcription par interférences transcriptionnelles

L'interférence transcriptionnelle (IT) est définie comme l'influence négative d'un flux transcriptionnel sur un autre *en cis*, c'est à dire sur la même molécule d'ADN (Shearwin, Callen et Egan, 2005). Il s'agit d'un phénomène asymétrique qui résulte de l'existence de plusieurs promoteurs physiquement proches, pour lesquels l'activité du promoteur le plus faible est influencée par celle du plus fort.

a- Organisation en clusters des promoteurs bactériens

Le génome d'*E.coli* compte 4462 promoteurs expérimentalement décrits ou prédits *in silico*, soit en moyenne 1 tous les 1000 pb. En fait, ils ne sont pas répartis de façon homogène sur les molécules ADN, si bien que certains peuvent être physiquement très proches. Deux promoteurs espacés de 40 à 200 pb peuvent être orientés soit dans le même sens (en tandem, Figure 7A) soit de façon différente : ils peuvent alors se faire face (promoteurs convergents, Figure 7B) ou non (promoteurs divergents, Figure 7C). Des promoteurs en tandem, convergents ou divergents mais dont les (+1) de transcription sont distants de 40 pb ou moins sont considérés comme chevauchants, en raison de l'encombrement stérique des ARN polymérases qui y sont recrutées (Figure 7D).

b- Promoteurs en tandem

Près de 7% des promoteurs d'*E.coli* sont organisés en tandem, soit 166 paires dont les (+1) de transcription sont distants d'au moins 70 pb (Sneppen *et al.*, 2005). Cette organisation est également courante chez les éléments extra-chromosomiques : plasmides, séquences d'insertions (IS), prophages, etc.

En général, les deux promoteurs sont différemment utilisés selon les conditions de croissance. Ce mécanisme est bien décrit pour le cas des opérons ribosomiaux d'*E.coli*, transcrits par deux promoteurs P1 et P2, ce dernier étant 120 pb en aval de P1 : dans un milieu de croissance riche, P1 est plus efficace que P2 et assure seul l'expression des gènes; quand le milieu s'appauvrit, P1 est inhibé et c'est P2 qui transcrit les gènes, mais plus faiblement, conduisant à une production d'ARNr moindre, ce qui contribue à ralentir la croissance cellulaire (Dennis, Ehrenberg et Bremer, 2004). En général, au moins un des deux promoteurs est transcrit par un facteur σ alternatif (Ramos, Marques et Timmis, 1997). Des situations plus complexes où trois voire quatre promoteurs en tandem participent simultanément à l'expression d'un gène ont été décrites (Govantes, Albrecht et Gunsalus, 2000) (Nasser, Rochman et Muskhelishvili, 2002). D'autre fois, des promoteurs en tandem se chevauchent mais fonctionnent simultanément (Ikenaka *et al.*, 1986). Des IT entre promoteurs en tandem ont été observées, notamment par occlusion du promoteur en aval par les ARN polymérases en cours d'élongation issues du promoteur en amont (Adhya et Gottesman, 1982 ; Hausler et Somerville, 1979). L'interférence est renforcée si les ARN polymérases font des pauses au niveau du promoteur en aval, du fait par exemples d'opérateur adjacents (Palmer *et al.*, 2009). D'autre part, à l'approche du promoteur en aval, une fraction des complexes en cours d'élongation issus du promoteur en amont peut être convertie en complexe abortifs (Kubori et Shimamoto, 1997).

c- Promoteurs divergents

Quand deux promoteurs sont dos à dos mais peu éloignés, des sites de liaisons pour leurs éventuels régulateurs transcriptionnels (ou opérateurs) peuvent être communs. Par conséquent, deux promoteurs divergents peuvent partager des régulations communes. Ce processus assure par exemple une régulation coordonnée de deux opérons d'une même voie métabolique (Lopez-Sanchez *et al.*, 2009), ou la répression simultanée des promoteurs de lyse et de lysogénie du prophage λ (Court, Oppenheim et Adhya, 2007).

Un des gènes divergents peut même coder le régulateur de l'autre : ainsi, les gènes *ampC*, qui codent une céphalosporinase, et le gène *ampR* sont transcrits depuis deux promoteurs divergents très peu éloignés; *ampR* code un régulateur positif qui se lie dans la région intergénique et stimule l'expression du gène *ampC*, mais pas sa propre expression (Kong *et al.*, 2005). Des modèles plus complexes avec plusieurs régulateurs communs ou un rétrocontrôle du gène régulateur ont aussi été décrits (voir notamment celui de l'induction du prophage CTX ϕ , (Waldor et Friedman, 2005)).

L'activité intrinsèque d'un promoteur peut profiter au promoteur divergent : en effet, le trafic des ARN polymérase induit des surenroulements négatifs en amont, ce qui facilite la formation de complexes ouverts de transcription (Chen et Lilley, 1999).

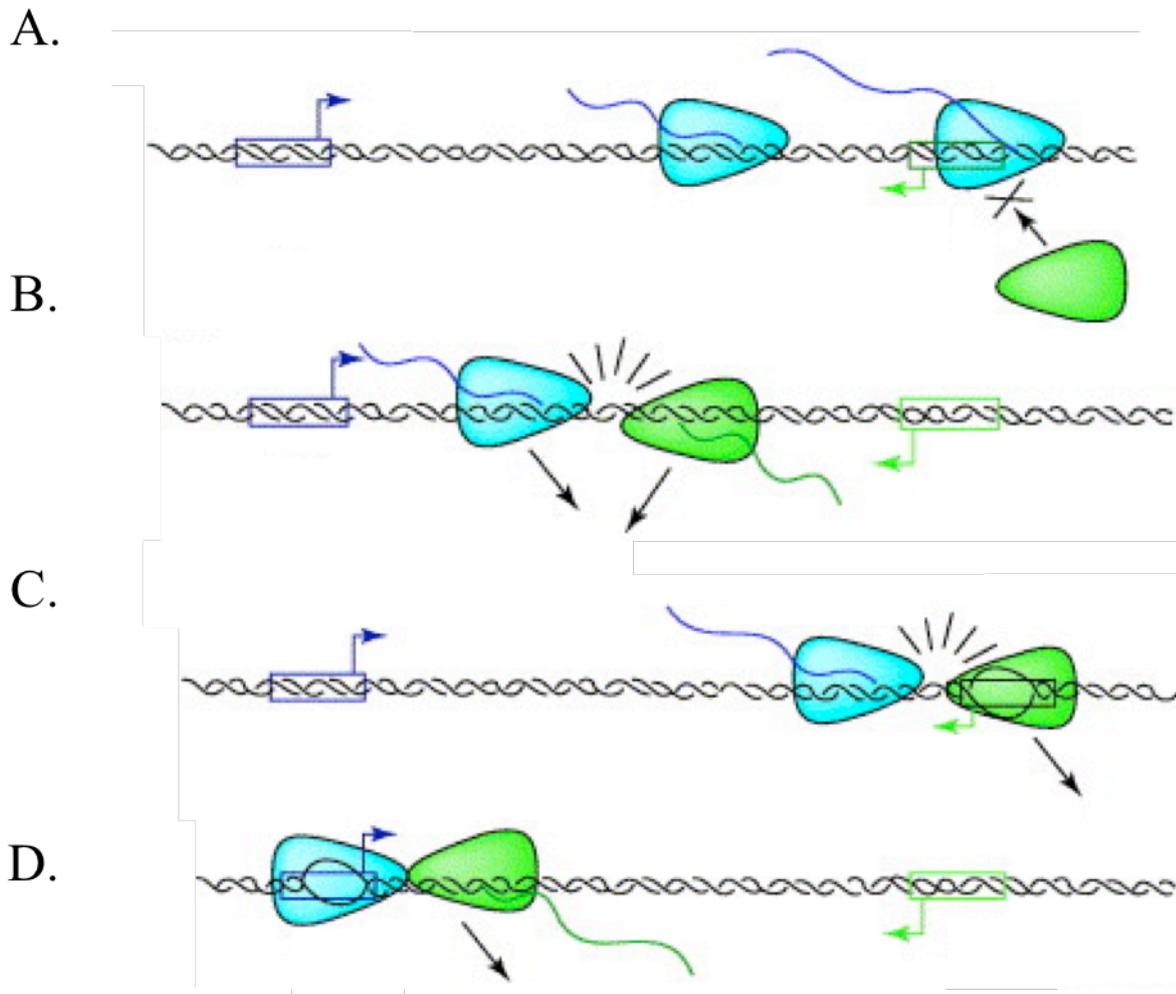


Figure 8: Mécanismes d'interférences transcriptionnelles entre promoteurs convergents

Les ARN polymérases sous forme holoenzyme sont représentés par des symboles  de la même couleur que le promoteur (flèche cassée) qui les a recrutées. Les transcrits en cours de synthèse sont représentés par une vague de la couleur de l'ARN polymérase qui les synthétise.

A. Occlusion du promoteur: la circulation des ARN polymérases issues du promoteur bleu empêche le recrutement de molécules sur le promoteur vert.

B. Collision de complexes d'élongation qui déstabilise les deux complexes.

C. « Sitting Duck »: La collision entre un complexe d'élongation issu du promoteur bleu et une ARN polymérase en cours d'isomérisation sur le promoteur vert déstabilise ce dernier.

D. « Roadblock »: L'enzyme liée au promoteur bleu est déjà en complexe ouvert et fait barrage au complexe d'élongation issue du promoteur vert, le déstabilisant. D'après Shearwin, Callen et Egan, 2005.

d- Promoteurs convergents

Deux promoteurs non chevauchants qui se font face et dont les (+1) de transcription sont éloignés de moins de 200 pb sont dits convergents (54 paires dans le génome d'*E.coli* (Shearwin, Callen et Egan, 2005)).

Ils génèrent des flux transcriptionnels concourants qui peuvent interférer de plusieurs façons : par occlusion du promoteur « faible » ou bien par des collisions entre les ARN polymérases (Figure 8A). L'issue de ces collisions dépend de l'étape de la transcription de ces molécules :

(i) la rencontre de deux molécules en cours d'élongation entraîne la dissociation des deux ARN polymérases (Figure 8B)

(ii) une ARN polymérase en cours d'isomérisation sur un promoteur sera déstabilisée par l'arrivée d'une autre en élongation, issue du promoteur convergent (mécanisme de « Sitting Duck », Figure 8C) (Callen, Shearwin et Egan, 2004). En revanche, si elle est déjà en complexe ouvert, plus stable, c'est l'ARN polymérase en élongation qui est déstabilisée (Figure 8D).

D'autre part, les promoteurs convergents génèrent *en cis* des ARN antiparallèles, ou ARN antisens, qui peuvent donc s'apparier ; de nombreux ARN antisens produits *en cis* inhibent l'expression d'un gène, soit en masquant aux ribosomes la région d'initiation de traduction de ce gène (Arthur *et al.*, 2003) soit parce que le duplex ARN est dégradé par la RNaseIII (Silvaggi, Perkins et Losick, 2005). De rares régulateurs *en cis* ont un effet positif sur l'expression du gène car ils stabilisent le transcrite du gène régulé (Tramonti, De Canio et De Biase, 2008).

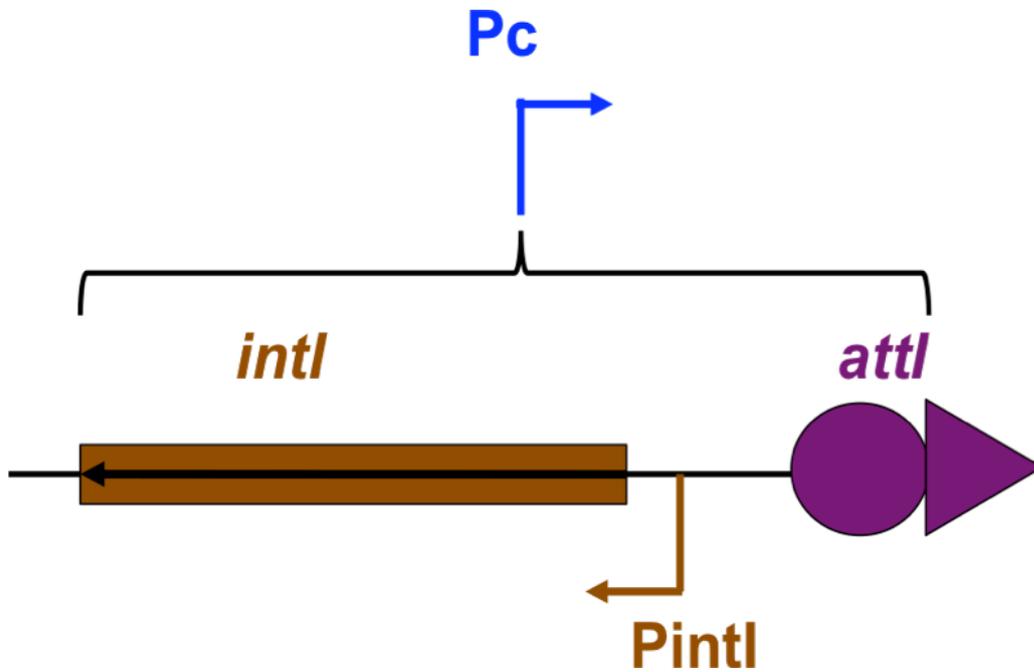


Figure 9: Structure générale de la plate-forme fonctionnelle d'un intégron

La définition moléculaire des intégrons se limite à trois éléments: un gène *intl* codant une intégrase (rectangle marron, dans lequel une flèche indique le sens de transcription), un site de recombinaison *attI* (symbole ) et un promoteur *Pc* (flèche cassée bleue) orienté dans le sens inverse du gène de l'intégrase. La position du promoteur *Pc* est variable selon les classes d'intégrons. Le gène *intl* est sous contrôle du promoteur *Pintl* (flèche cassée marron).

Chapitre 2 : Les intégrons

Les intégrons sont des éléments génétiques décrits à la fin des années 1980. Ils forment un système original de capture, d'expression et de dissémination de gènes sous forme de cassettes. On distingue deux grands types d'intégrons:

- Les intégrons de multirésistance (IM) sont portés par des plasmides et/ou des transposons ; ils sont connus pour leur rôle majeur dans la dissémination des gènes de résistances aux antibiotiques.
- Les super-intégrons (SI) sont strictement chromosomiques et jouent un rôle plus large en tant que réservoir de gènes à haute valeur adaptative.

A. Généralités sur les intégrons

1. Définition des intégrons

Les intégrons sont définis comme l'association physique d'un gène *intI* codant une intégrase, d'un site de recombinaison *attI* et d'un promoteur *Pc* orienté dans le sens inverse du gène *intI* (Stokes et Hall, 1989). Ces trois éléments constituent une plate-forme fonctionnelle très conservée (Figure 9).

Intégrase	IntI1	IntI2*	IntI3	IntI4	IntI5
IntI1	100%	46%	59%	40%	41%
IntI2*		100%	47%	54%	44%
IntI3			100%	44%	44%
IntI4				100%	40%
IntI5					100%

Tableau 3 : Pourcentages d'identité entre les séquences en acides aminés des intégrases d'intégrons de multi résistances.

Les numéros d'accès GenBank (collection protéique) utilisés sont : IntI1, AAB59081 ; IntI3, BAA08929 ; IntI4 : AAK95987 ; IntI5 : CAC35342.

La séquence d'IntI2* correspond à la séquence nucléotidique du gène *intI2* du fichier GenBank AJ001816 traduite au delà du codon STOP précoce. D'après Collis et al., 2002.

a- Le gène *intI* et la protéine IntI

Le gène *intI* code une recombinase à tyrosine spécifique de site. Les enzymes de cette famille catalysent des recombinaisons entre des sites d'attachement situés sur une même ou sur deux molécules d'ADN et interviennent notamment dans des processus de lysogénie, de transposition ou d'inversion de phase (Nunes-Duby *et al.*, 1998). La protéine IntI s'apparente plus particulièrement aux recombinases XerC/XerD qui résolvent les dimères de chromosomes formés pendant la réplication (Mazel, 2006). Des dizaines de classes d'intégrons ont été définies sur la base de la séquence de leurs intégrases IntI, les intégrases codées partageant de 40 à 59% d'identité protéique (Collis *et al.*, 2002 ; Holmes *et al.*, 2003 ; Nield *et al.*, 2001 ; Stokes *et al.*, 2001) (Tableau 3). La majorité des classes correspondent à des SI, cinq à des IM (IM 1 à 5). L'appartenance des intégrons de classe 6, 7 et 8 à un type ou l'autre d'intégron n'a pas été établie. Les IM 1 à 3 sont les classes d'intégrons les plus prévalentes en microbiologie clinique et les mieux caractérisées. L'intégrase de classe 1 est une protéine de 337 acides aminés (aa). L'intégrase de classe 2 existe le plus souvent sous une forme non fonctionnelle, réduite à 178 aa du fait d'un codon STOP ochre prématuré dans le gène *intI2*. La suppression de ce codon conduit à une protéine active IntI2* de 325 aa présentant 46% d'identité avec IntI1 (Hansson *et al.*, 2002). L'intégrase de classe 3 (325 aa) partage 59% d'identité avec IntI1 et 46% avec IntI2* (Collis *et al.*, 2002).

Les intégrases des IM 4 et 5, initialement décrits chez des espèces du genre *Vibrio*, sont probablement fonctionnelles mais n'ont jamais été testées (Tableau 3). La capacité des intégrases IntIA de SI à recombinaison des cassettes n'a été montrée que pour les SI de *Vibrio cholerae* (*V.cholerae*) (Rowe-Magnus *et al.*, 2001), de *Nitrosomonas europaea* (Leon et Roy, 2003) et de *Pseudomonas stutzeri* (Coleman et Holmes, 2005 ; Holmes *et al.*, 2003).

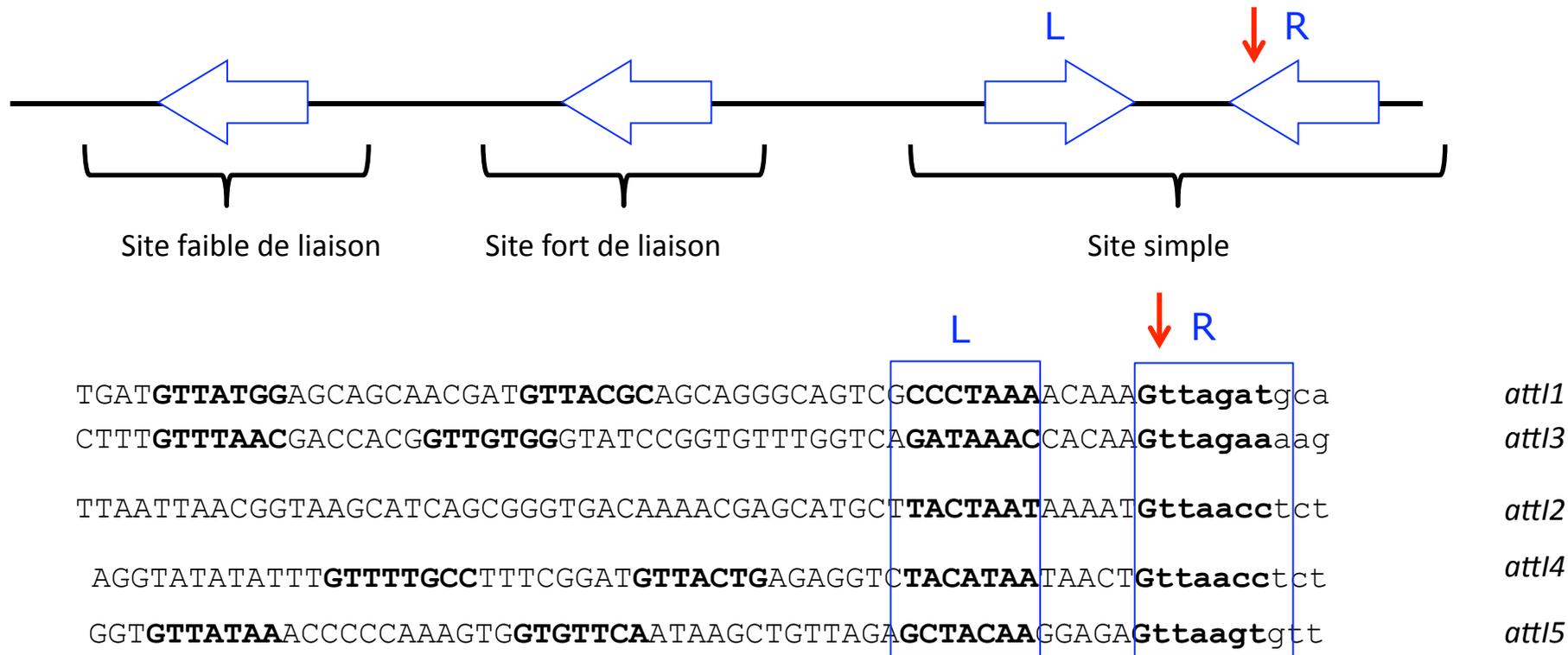


Figure 10: Organisation des sites attI des intégrons de multirésistance

Les sites attI sont organisés en quatre sites « core » (flèches horizontales) définissant trois sites de liaison (fort, faible et simple) pour IntI. Une flèche rouge indique le point de recombinaison. Les séquences des sites attI sont détaillées, les sites « cores » étant indiqués en gras. Seuls les sites « core » d'attI1 et attI3 ont été déterminés expérimentalement. Une flèche rouge indique le point de recombinaison.

D'après Collis et Hall, 2004

b- Le site *attI*

Le site *attI* est un domaine d'attachement spécifique de la protéine IntI qui se termine au point de recombinaison des cassettes (Collis et Hall, 2004 ; Partridge *et al.*, 2000). La séquence du site *attI* varie fortement d'une classe à l'autre à l'exception d'un motif de 7 pb nommé boîte R, de consensus GTTRRRY (R : Purine ; Y : Pyrimidine) où a lieu la recombinaison des cassettes (entre le G et le 1^{er} T) (Figure 10). La boîte R forme avec une boîte L située plus en amont un site « core » de liaison essentiel à l'activité d'IntI1, Le site *attI* comprend deux autres sites de fixation d'IntI1 qui sont des répétitions directes portant le motif GTTR (Figure 10). Le premier est un site fort indispensable à l'activité d'IntI1, l'autre en améliore l'efficacité. Certains IM 1 présentent une duplication du site fort. La séquence du site *attI* indispensable pour la recombinaison a été délimitée à environ 60 pb pour les IM 1 et 3 (Figure 10). Nous avons fait le choix dans ce manuscrit de nommer « région *attI* » l'ensemble de la séquence comprise entre la séquence codante du gène *intI* et le début de la 1^{ère} cassette. Sa taille varie de 115 à 311 pb selon les classes d'IM.

Classe d'intégron	Localisation du promoteur Pc		Références
	Dans la séquence codante d' <i>intl</i>	En amont de la séquence codante d' <i>intl</i>	
IM 1	TTGACA-N₁₇-TAAACT*	TTGTTA-N₁₇-TACAGT⁽¹⁾	Collis et Hall, 1995
IM 2	---	GTGACA-N ₁₉ -TAAAAT	Fichier GenBank L10818
		TTGTAT-N ₁₈ -TTTAAT ; TTTTAA-N ₁₆ -TAAAAT	Fichier GenBank AM261760
		TTGCAA-N ₁₈ -TATTCT	Hansson <i>et al.</i> , 2002
IM 3	TAGACA-N₁₇-TAGGCT	---	Collis <i>et al.</i> , 2002
IM 4	<u>Aucun évident</u>	<u>TTGATT-N₁₇-TATATA</u>	Ce travail
IM 5	<u>TATAGG-N₁₇-TAAGCT</u>	TTTACA-N ₁₇ -AACTAT	Fichier GenBank AJ277063
SIP _{StQ}	TTGAGC-N ₁₆ -TCTGAT	---	Coleman et Holmes, 2005
Classe 6	---	TATAAA- N ₁₇ -TATACT*	Nield <i>et al.</i> , 2001
Classe 7	---	TCGAGG- N ₁₇ -TAAAAT	Nield <i>et al.</i> , 2001
Classe 8	TTGCAG- N ₁₆ -TAGCAT	CTGTAA- N ₁₇ -AAGAAT	Nield <i>et al.</i> , 2001
InTde35405	---	TTGCAG- N ₁₇ -TATTGG; TTGCAA- N ₁₇ -TATAGT*	Coleman <i>et al.</i> , 2004

Tableau 4 : Séquence des promoteurs Pc de plusieurs classes d'intégrons

Les séquences des promoteurs caractérisés sont écrites en gras, les autres correspondent à des propositions de la littérature, ou de ce travail (ces dernières sont soulignées).

Les promoteurs dont la séquence est suivie d'un astérisque (*) présentent un motif TGN-10 qui caractérise la région « -10 étendue ».

N : A, C, G ou T

⁽¹⁾ Ce promoteur (P2) est occasionnellement présent dans les IM 1.

c- La région promotrice

La plate-forme fonctionnelle des intégrons héberge un promoteur fonctionnel Pc, caractérisé chez les IM 1 et 3 uniquement. Il se situe dans la séquence codante du gène *intI* et il en existe plusieurs variants qui diffèrent par des substitutions dans les éléments -35 et -10 de Pc (Collis et Hall, 1995 ; Collis *et al.*, 2002 ; Correia *et al.*, 2003 ; Hansson *et al.*, 2002 ; Levesque *et al.*, 1994). Aucun Pc de SI n'a été caractérisé précisément, mais il a été localisé au début du gène *intI* du SI de *Pseudomonas stutzeri* (Tableau 4) (Coleman et Holmes, 2005). En revanche, le Pc des IM 2 serait dans la région *attI*, et plusieurs séquences ont été proposées (Hansson *et al.*, 2002) (Tableau 4). Enfin, des séquences promotrices putatives ont été proposées pour le Pc des IM 5, des intégrons de classe 6, 7 et 8 ainsi que pour l'intégron de *Treponema denticola* (Tableau 4). Ce dernier est particulier, puisque le gène de l'intégrase est orienté dans le sens des cassettes (Coleman *et al.*, 2004). Un certain nombre de promoteurs Pc connus ou suspectés présentent le motif « TGN » qui caractérise les régions « -10 étendue » (séquences marquées d'un astérisque dans le tableau 4).

2. Les cassettes de gènes

a- Définition des cassettes de gène

L'intégrase IntI est capable d'insérer en aval du site *attI* un ou plusieurs gènes sous forme de cassettes. Une cassette de gène est un élément mobile non autonome et non répliatif qui associe généralement un seul cadre ouvert de lecture à un site de recombinaison *attC* (Recchia et Hall, 1995). Les cassettes peuvent exister sous forme linéaire, insérée au sein d'un intégron ou transitoirement à l'état libre sous forme circulaire (Collis et Hall, 1992).

Le nombre de cassettes varie de 0 à 8 au sein des IM (Bissonnette et Roy, 1992 ; Naas *et al.*, 2001), de 40 à plus de 200 au sein des SI (Mazel, 2006). De plus, une même cassette peut être présente en deux, voire trois copies dans un même réseau de cassettes. D'autre part, les réseaux de cassettes peuvent être interrompus par des IS insérées au niveau des sites *attC* (Post et Hall, 2009). De rares cassettes sont insérées en dehors des intégrons, à des sites de recombinaison dits secondaires (Francia *et al.*, 1993; Recchia *et al.*, 1995; Hansson *et al.*, 1997; Ramirez *et al.*, 2008; Frumerie *et al.*, 2009; Zong, Partridge et Iredell, 2009).

b- Les gènes contenus dans les cassettes

A ce jour, plus de 130 cassettes différentes (moins de 98% d'identité nucléotidique) portant des gènes de résistance aux antibiotiques ont été identifiées dans les IM, la quasi-totalité des familles d'antibiotiques étant représentés : aminosides, β -lactamines, erythromycine, fosfomycine, lincosamides, phénicol, quinolones, rifampicine, streptothricine et triméthoprime (Partridge *et al.*, 2009). D'autres cassettes sont impliquées dans la résistance à des antiseptiques de type ammonium quaternaires ou à des espèces oxygénées réactives (*msrA*, *msrB*) (Gillings *et al.*, 2009). Les cassettes portant un cadre ouvert de lecture de fonction inconnue sont minoritaires dans les IM. Enfin, de rares cassettes dépourvues de gènes (cassettes cryptiques) ou de sites *attC* complet ont été identifiées (Partridge, Collis et Hall, 2002 ; Partridge *et al.*, 2009).

Un grand nombre de gènes des cassettes de SI sont impliqués dans d'autres rôles que la résistance aux antibiotiques, tels que le métabolisme ou la virulence de la cellule. L'analyse des cassettes des SI a montré la présence de gènes à haute valeur adaptative, dont certains impliqués dans la virulence ou le métabolisme de l'hôte. Cependant, contrairement aux IM, les gènes de fonction inconnue prédominent (67% pour les SI de *Vibrio*) et ceux susceptibles de coder des résistances aux antibiotiques sont plus rares (Boucher *et al.*, 2007).

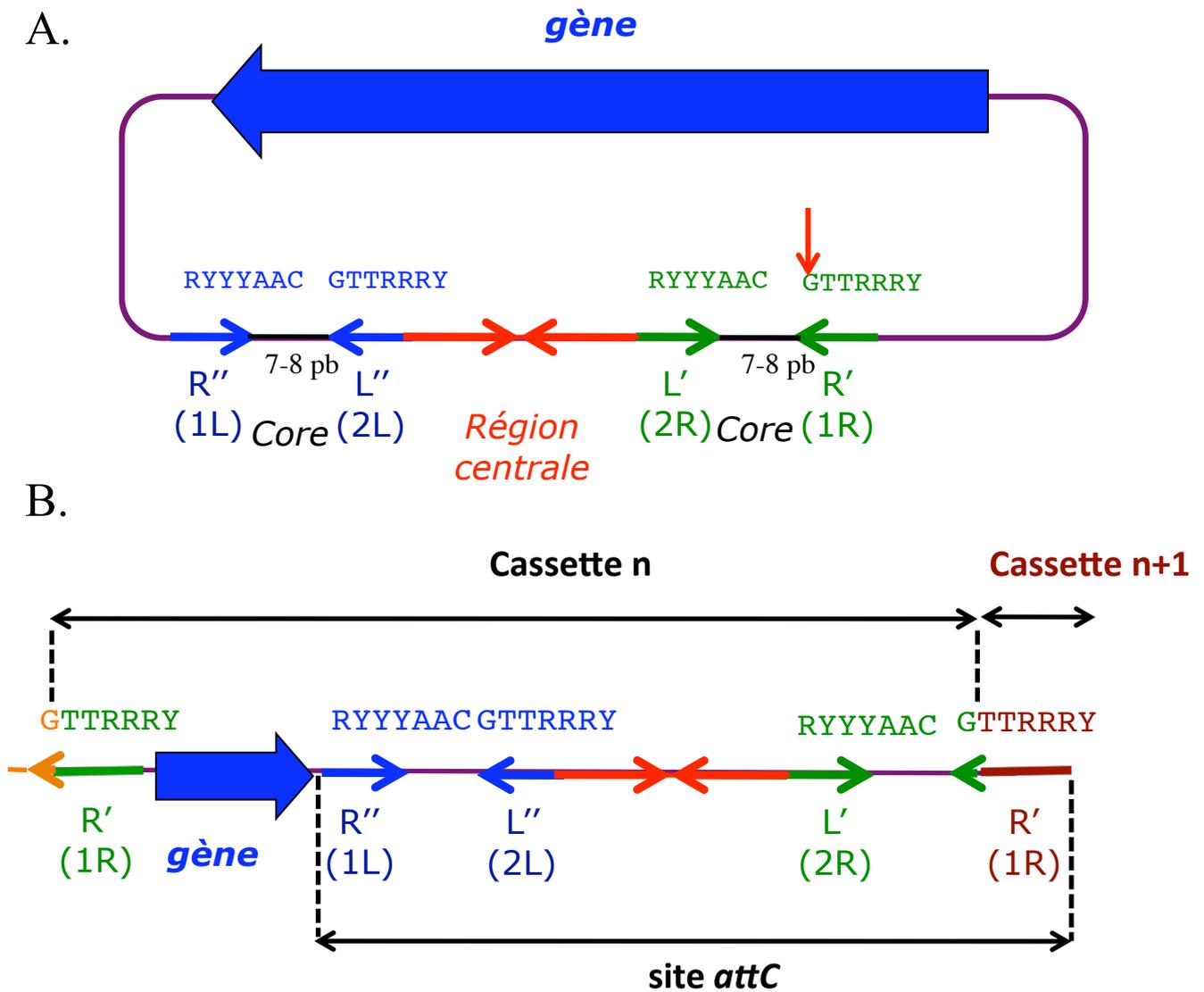


Figure 11: Structure d'une cassette de gène sous forme libre ou insérée au sein d'un intégron.

A. Schéma d'une cassette de gène sous sa forme libre, circulaire. Une cassette est formée de deux éléments: la séquence codante d'un gène (flèche bleue), et un site de recombinaison *attC*. Ce dernier s'organise en deux sites « core » (en bleu et vert) séparés par une région centrale (en rouge). Chaque site core est une paire de répétitions inversées, dont les séquences consensus sont précisées (flèches simples) séparées par 7 ou 8 pb. Une flèche verticale rouge indique le point de recombinaison.

B. Schéma d'une cassette linéaire au sein d'un intégron. Les éléments appartenant à la cassette en amont et à celle en aval sont représentés en orange et marron, respectivement. Les limites des sites *attC* et de la cassette sont précisées: la cassette se termine par le G du site R' et commence par le T du même site. D'après Partridge et al., 2009.

A ce jour, la seule cassette de SI connue pour conférer effectivement une résistance à un antibiotique est la cassette *catB9* (résistance au chloramphénicol).

Enfin, des cassettes codant des systèmes toxine-antitoxine ont également été repérées dans les SI de *V.fischeri* et *V.cholerae* (Rowe-Magnus *et al.*, 2003).

c- Les sites attC

Chaque cassette possède un site de recombinaison *attC* reconnu par les intégrases. Un site *attC* est composé de deux sites « core » L et R espacés par une région centrale de taille et de séquence très variable. Chaque site core est une paire de motifs de 7 ou 8 pb (R''/L'' et R'/L') qui sont les complémentaires inverses l'un de l'autre (Stokes *et al.*, 1997) et la région centrale forme un palindrome imparfait (Figure 11A). Cette organisation permet aux sites *attC* d'adopter, sous forme simple brin, une conformation en « tige-boucle » dont la tige est formée à la base par les duplex R''-R' et L''-L' et la boucle par la région centrale (Hall, Brookes et Stokes, 1991 ; Rowe-Magnus *et al.*, 2003). Une base supplémentaire dans le motif L'' sans équivalent dans L' permet à l'intégrase de reconnaître le brin à recombinaison (Bouvier *et al.*, 2009). La protéine IntI se lie aux sites core et core inverse et catalyse la recombinaison au niveau du point de clivage des cassettes situé entre le G et le premier T du motif R'. Aussi, une fois qu'elle est insérée en *attI*, la cassette commence par ce T et se termine par ce G du motif R' (Figure 11B). Dans le réseau, chaque gène de cassette est donc suivi d'un site *attC* fonctionnels « hybride » formé avec le site R' de la cassette suivante. La variabilité de la région centrale explique la diversité des sites *attC*, dont la taille varie de 57 à 141 pb. Chez les IM, le degré de similarité entre les sites *attC* de deux cassettes est indépendant de celui de leurs gènes : ainsi, deux gènes très proches peuvent être associés à des sites *attC* très différents (par exemple, *blaIMP-8* et *blaIMP-9*) et à l'inverse certains gènes plus éloignés présentent des sites *attC* quasi-identiques (par exemple, *aadA1a* et *aadA7*). En revanche, les sites *attC* des cassettes d'un même SI sont très conservés.

3. Nomenclature

a- Nomenclature des cassettes et des sites *attC*

La plupart des cassettes sont désignés sous le nom de leur gène. Pour les cassettes de fonction inconnue, l'acronyme *orf* est souvent utilisé, suivi d'une lettre (*orfA*, *orfX*, *orfX'*). Récemment, le terme de *gcu* (pour *gene cassette of unknown function*) a été proposé (Partridge *et al.*, 2009).

Les sites *attC* des IM portent le nom des cassettes (ex : *attC_{aadA7}*), ceux des SI portent un nom évoquant l'espèce hébergeant le SI : par exemple, VCR (*Vibrio cholerae repeats*) pour les cassettes du SI de *V. cholerae*.

b- Nomenclature des intégrons et des intégrases

Les intégrases et les sites *attI* sont numérotés selon les classes d'intégrons pour les IM (IntI1, *attI2*, etc.) où de l'espèce hôte pour les SI (IntIVch pour le SI de *V. cholerae*, IntI1Gsu pour celui de *Geobacter sulfurreducens*, etc.). Le seuil de divergence d'IntI suffisant pour définir une nouvelle classe d'intégrons n'a pas été fixé. De plus, il n'existe pas de nomenclature permettant d'unifier l'ensemble des intégrons d'une même classe. Par exemple, si des IM 1 ont été numérotés de façon quasi continue de In0 à In120 sur la base du réseau de cassettes de gènes, d'autres ont été numérotés autrement, comme In2345 qui renvoie au numéro de la souche (Llanes *et al.*, 2006). En outre, plusieurs IM 1 distincts sont parfois numérotés de la même façon, tel In71 (Castanheira *et al.*, 2007 ; Lagatolla *et al.*, 2006).

Certains IM 2 et tous les IM 3 ont été numérotés (In2-0 à In2-12 et In3-1 à In3-4) (Ramirez, Pineiro et Centron, 2009 ; Xu, Davies et Miao, 2007).

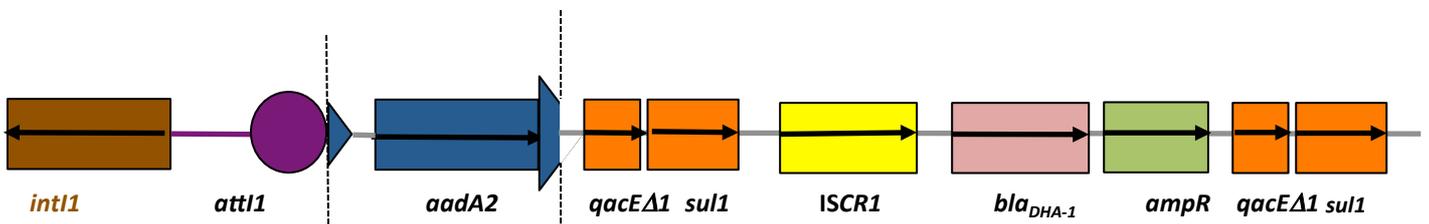


Figure 12 : Structure d'un intégron complexe: In60

La région attI1 est représenté par un fragment violet terminé par un cercle, les sites de recombinaison de la cassette aadA2 par un triangle, les séquences codantes par des rectangles, en marron pour le gène de l'intégrase intI1, en bleu pour celui de la cassette aadA2, en orange pour les gènes des régions 3'CS, en jaune pour le gène de l'ISCR1 qui caractérise les intégrons complexes, en vert ou rose pour les autres gènes de la structure. Leurs sens de transcription sont indiqués par des flèches. D'après (Toleman, Bennett et Walsh, 2006).

B- Diversité et épidémiologie des intégrons

L'analyse de 603 génomes partiellement ou totalement séquencés a montré que 56 (9%) d'entre eux présentent des intégrases d'intégrons putatives (Boucher *et al.*, 2007). Le panel des bactéries porteuses d'intégrons est très large, incluant des espèces très éloignées sur le plan phylogénétique (protéobactéries, spirochètes, cyanobactéries, actinobactéries, etc.) notamment pour les SI. Le but de ce chapitre est de détailler les particularités des principales classes d'intégrons, au delà des divergences dans la séquence de leur 5'CS.

1. Intégrons de multirésistance

a- Les IM de classe 1

Initialement décrits comme associés au transposon Tn21 (Stokes et Hall, 1989), de nombreux IM 1 ont été caractérisés suite aux études sur les bactéries multirésistantes aux antibiotiques, car ils prévalent dans les isolats cliniques. Ainsi, pas moins de 1500 entrées plus ou moins complètes de la GenBank traitent de cette classe d'IM.

Le réseau de cassettes d'un IM de classe 1 est très fréquemment suivie en 3' de tout ou partie d'une région nommée 3'CS (3' Conserved Segment) et comprenant trois gènes :

- le gène *qacEΔ1*, une version tronquée et de moindre efficacité du gène *qacE* codant une pompe d'efflux conférant une résistance aux ammoniums quaternaires.
- Le gène *sulI* code une dihydroptéroate synthétase essentielle à la synthèse des folates et qui confère une résistance aux sulfamides
- La séquence codante *orf5* codant un peptide de fonction inconnue.

Un certain nombre d'IM 1 ont la particularité de posséder une duplication de cette région 3'CS séparée de la première par une région commune nommée CR (Common Region) ainsi que d'un ou plusieurs gènes de résistances aux antibiotiques non contenus dans ces cassettes (Toleman, Bennett et Walsh, 2006) (Figure 12) .

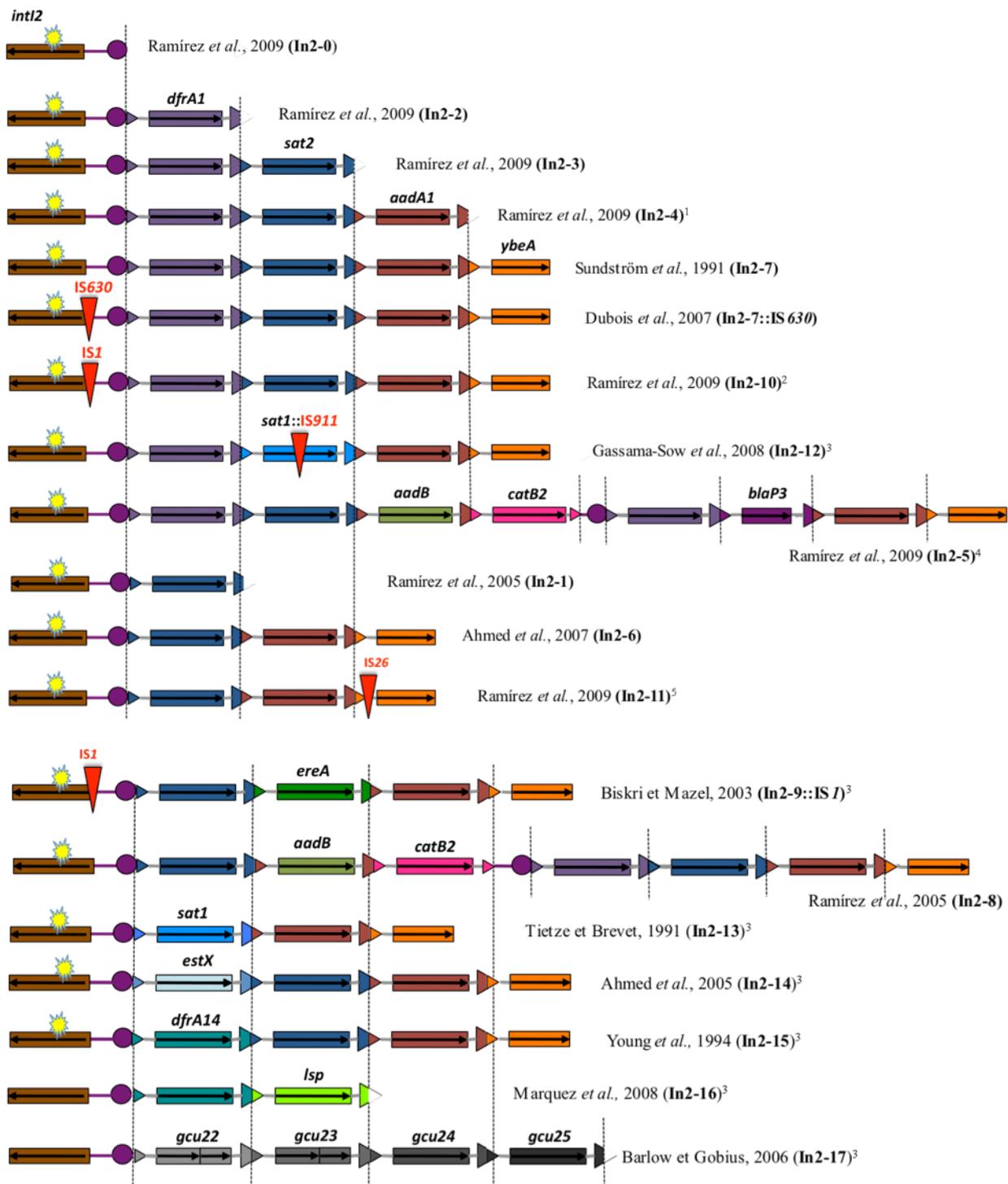


Figure 13: Structure des différents IM 2 caractérisés à ce jour

Les références sont indiquées pour chaque IM. Chaque gène est représenté par un rectangle de couleur différente, une flèche précise le sens dans lequel il est transcrit. Une étoile indique quand le gène *intI2* est tronqué. Les sites *attC* sont symbolisés par des triangles de la même couleur que le gène de la cassette correspondante, la région *attI2* par un segment violet (il se termine par le site de recombinaison d'*attI2*, symbole ). Des lignes pointillées verticales délimitent les cassettes. Les cassettes *gcu22* et *gcu23* contiennent 2 ORF chacune. ¹ Ce réseau de cassettes a également été retrouvé suivi de la 3'CS des IM 1 (Ploy *et al.*, 2000) ² Les auteurs ont attribué ce nom à cet IM 2 (In2-7::IS1) alors que le réseau de cassettes est identique à celui de In2-7. ³ Ces IM 2 ont été numérotés ici en continuant la numérotation utilisée dans (Ramírez *et al.*, 2009.) ⁴ Ce réseau a été nommé à la fois In2-5 et In2-9. ⁵ Les auteurs ont attribué ce nom à cet IM 2 (In2-6::IS26) alors que le réseau de cassettes est identique à celui de In2-6.

La région CR contient une séquence codante longtemps désignée *orf513* et qui a été caractérisée récemment comme un nouveau type d'IS, nommé *ISCR1* apparentée aux *IS91* (Toleman, Bennett et Walsh, 2006). L'*ISCR1* fournit en outre un promoteur actif pour l'expression des gènes en aval (Rodriguez-Martinez *et al.*, 2006). Ces intégrons sont désignés intégrons complexe ou « sul1-type » intégrons

b- Les IM de classe 2 (IM 2)

Les IM 2 sont souvent associés au transposon Tn7 ou des transposons apparentés tels que Tn1825, Tn1826 ou Tn4132 (Sundstrom, Roy et Skold, 1991 ; Tietze, Brevet et Tschape, 1987 ; Young, Qumsieh et McIntosh, 1994). Le réseau de cassettes des IM 2 est généralement très conservé avec 4 cassettes: *dfrA1*, codant une résistance au triméthoprim, *sat2*, codant une résistance à la streptothricine, *aadA1*, codant un résistance à la streptomycine et la spectinomycine et d'une cassette putative connue sous le nom de *orfX* qui en fait correspond à un gène *ybeA* associé à un site *attC* tronqué (Ramirez, Pineiro et Centron, 2009). Le gène *intI2* est le plus souvent interrompu par un codon STOP ochre et code donc une protéine IntI2 raccourcie à 178 aa, qui n'est pas fonctionnelle pour la recombinaison des cassettes. Hansson *et coll.* ont proposé un rôle de régulateur pour ce polypeptide (Hansson *et al.*, 2002). L'inactivité d'IntI2 explique, au moins en partie, la faible diversité des réseaux de cassettes des IM 2. Toutefois, ces dernières années, des IM 2 avec des réseaux de cassettes différents ont été décrits (Ahmed, Hussein et Shimamoto, 2007 ; Ahmed, Nakano et Shimamoto, 2005 ; Barlow et Gobius, 2006 ; Biskri et Mazel, 2003 ; Dubois *et al.*, 2007 ; Gassama Sow *et al.*, 2008 ; Marquez *et al.*, 2008 ; Ramirez, Pineiro et Centron, 2009 ; Ramirez, Quiroga et Centron, 2005 ; Ramirez *et al.*, 2005 ; van Essen-Zandbergen *et al.*, 2007) (Figure 13). Ces IM 2 pourraient avoir été créés suite à des recombinaisons catalysées par une intégrase d'IM 1

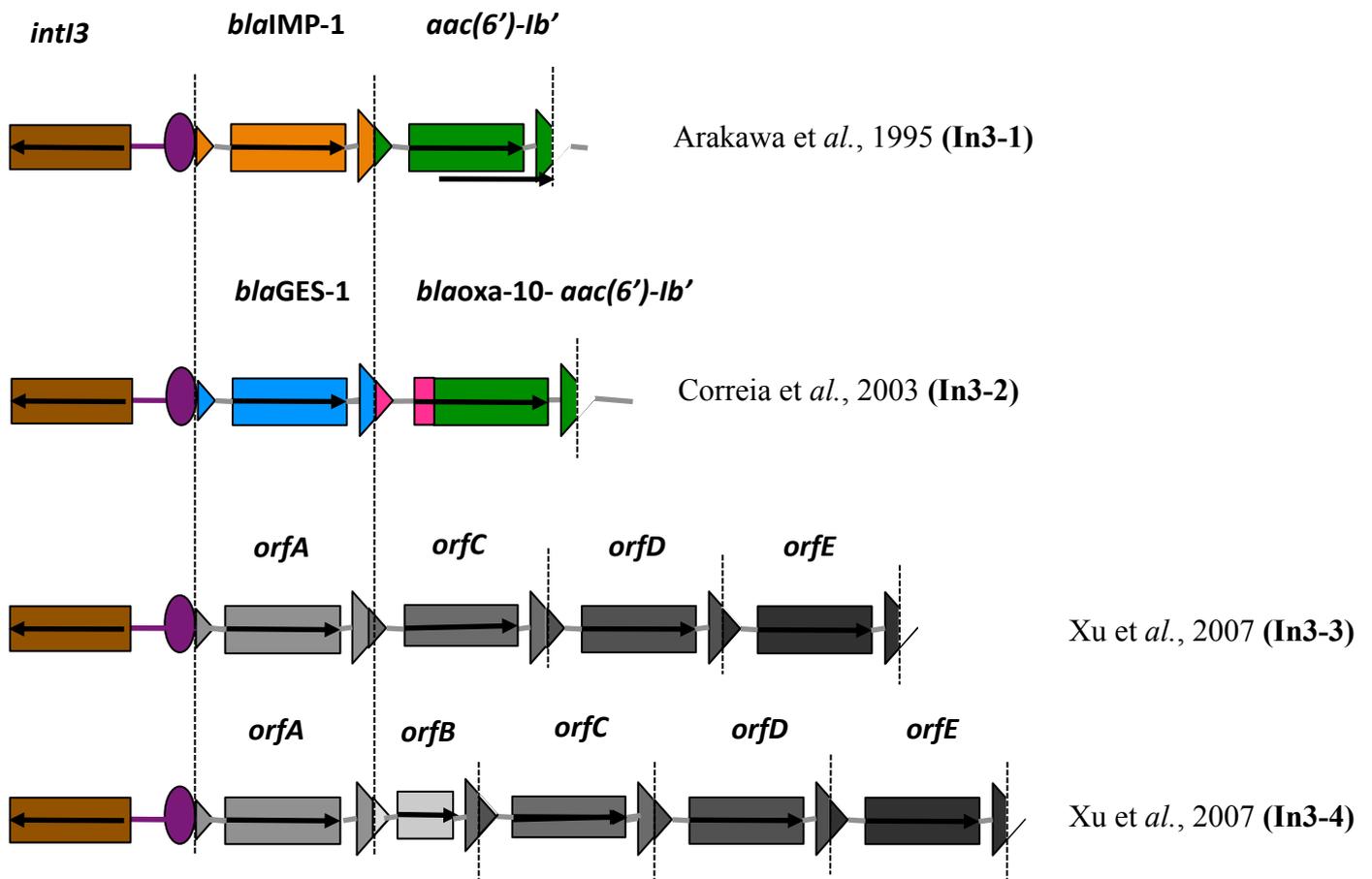


Figure 14: Structure des différents IM 3 caractérisés à ce jour.

Les références sont indiquées pour chaque intégron. Chaque gène est représenté sous la forme d'un rectangle de couleur différente, dans lequel une flèche précise le sens où il est transcrit. Les sites attC des cassettes sont symbolisés par des triangles de la même couleur que le gène de la cassette, la région attI3 par un segment violet (il se termine par le site de recombinaison d'attI3, le cercle plein violet). Des lignes pointillées verticales délimitent les cassettes. Le schéma n'est pas à l'échelle. Le gène de la cassette blaOXA-10/aac(6')-Ib est une fusion traductionnelle entre ces deux gènes.

car certaines souches hébergent simultanément ces deux classes d' IM (White, McIver et Rawlinson, 2001) . De plus, deux allèles d'*intI2* dont le codon STOP est remplacé par un codon d'acide aminé (codant la glutamine) ont été caractérisés récemment (Barlow et Gobius, 2006 ; Marquez *et al.*, 2008), dont l'un code une intégrase fonctionnelle pour la recombinaison des cassettes (Barlow et Gobius, 2006 ; Marquez *et al.*, 2008). D'autres allèles potentiellement fonctionnel d'*intI2* ont été repérés dans des échantillons environnementaux mais les IM 2 correspondants n'ont pas été caractérisés (Rodriguez-Minguela *et al.*, 2009). Certains IM 2 présentant des IS insérées dans la région *attI2* ou le réseau de cassettes ont été décrits (Dubois *et al.*, 2007 ; Gassama Sow *et al.*, 2008) (Figure 13). Enfin, un IM 2 « hybride » dont le réseau de cassettes est suivi de la 3'CS typique des IM 1 a été découvert (Ploy *et al.*, 2000). Il est probablement le produit de recombinaisons homologues entre des cassettes identiques portées par des intégrons de classe distincte.

c- Les IM de classe 3 (IM 3)

La structure des IM 3 est très proche de celle des IM 1 : le gène *intI3* code une protéine fortement apparentée à *IntI1* (59% d'identité) (Collis et Hall, 2004). De plus, il héberge le promoteur *Pc* à la position correspondante à celle du *Pc* dans *intI1* (Collis *et al.*, 2002 ; Correia *et al.*, 2003).

Seuls quatre IM 3 ont été caractérisés à ce jour (Arakawa *et al.*, 1995 ; Correia *et al.*, 2003 ; Xu, Davies et Miao, 2007) (Figure 14). En effet, leur prévalence est faible, et souvent nulle, au sein des isolats cliniques (Shibata *et al.*, 2003). D'ailleurs, deux de ces intégrons ont été caractérisés chez des souches environnementales isolées d'eaux usées, et les gènes de leur cassettes codent des fonctions inconnues. Dans une étude récente, des IM 3 ont été détectés dans six d'échantillons issus d'une station d'épuration (Rodriguez-Minguela *et al.*, 2009). Ces résultats suggèrent que les IM 3 joueraient un rôle analogue aux IM 1 mais en dehors du contexte clinique.

d- Les IM de classe 4 et 5 (IM 4 et IM 5)

Deux autres classes d'IM ont été identifiées via leur implication dans la résistance au triméthoprimé chez deux espèces de *Vibrio*. Le seul IM 4 décrit est un composant d'un élément conjugatif intégratif (ICE) localisé sur le chromosome de *V.cholerae*. Il héberge cinq cassettes dont quatre contiennent des gènes de fonction inconnue (Hochhut *et al.*, 2001). Deux IM de classe 5 ont été trouvés : le premier est sur un plasmide de *Vibrio salmonicida* et héberge 7 cassettes dont 6 codent des fonctions inconnues : le second est sur le chromosome I d'une protéobactérie marine (Non publié, fichier GenBank CR954246).

e- Epidémiologie des IM

La plupart des études épidémiologiques sur les IM porte sur des souches cliniques multirésistantes aux antibiotiques. Aussi, les fortes prévalences observées pourraient être biaisées par ces sélections (Recchia et Hall, 1997). Le plus souvent, les IM 1 prédominent, les IM 2 sont plus rares, les IM 3 n'étant que peu recherchés. Des études récentes suggèrent que les IM 2 et surtout les IM 3 seraient plus fréquents dans les échantillons issus de l'environnement, avec des cassettes de gènes dont le rôle serait autre que la résistance aux antibiotiques. (Rodriguez-Minguela *et al.*, 2009)

- Prévalences générales des IM

Les chiffres des prévalences d'intégrons en microbiologie clinique sont très variables selon les études et les méthodologies employées. Ils varient de 16% (Enne *et al.*, 2001 ; Schmitz *et al.*, 2001) à près de 50% chez des souches d'entérobactéries (White, McIver et Rawlinson, 2001), essentiellement des IM 1.

Cependant, la prévalence des IM, en particulier des IM 1, est probablement sous-estimée. En effet, les auteurs utilisent généralement des techniques de PCR avec soit une amorce dans la plate-forme fonctionnelle et une autre dans la région « 3'CS », soit avec les deux amorces

dans le gène *intI1*. Or certains IM 1 qui ne contiennent pas de 3'CS ne peuvent être détectés de cette manière.

De plus, ces approches PCR ne renseignent pas sur le nombre d'IM 1 distincts dans la souche. En effet, une même souche peut porter plusieurs IM 1 (Fournier *et al.*, 2006 ; White, McIver et Rawlinson, 2001). Il est important de noter que quasiment tous les travaux publiés ont recherché les intégrons sur les souches isolées de différents prélèvements et non directement sur les prélèvements.

Les IM ont été initialement décrits pour leur implication dans la dissémination des résistances aux antibiotiques dans le contexte clinique, mais ils sont aussi retrouvés chez des souches animales ou environnementales. Ainsi, des IM ont été détectés chez des souches isolées d'animaux domestiques (Goldstein *et al.*, 2001), d'élevage (Vali *et al.*, 2007) ou sauvages (Sunde, 2005), ou de produits de la viande bovine, ovine, porcine ou aviaire (Sunde, 2005). Les prévalences observées varient de 25 à 52% (toutes classes d'IM confondues) et sont plus fortes si les animaux vivent à proximité des hommes.

Les IM de souches environnementales proviennent de cours d'eau naturels (Rosser et Young, 1999), de bassins artificiels (Jacobs et Chenia, 2007), ou d'eaux usées (Ferreira da Silva *et al.*, 2007), avec des prévalences fortes, de 35 à 68% selon le type d'échantillon.

- Les bactéries hôtes

La majorité des IM décrits sont hébergés par des bactéries à Gram négatif. Ainsi, des IM 1 ont été détectés chez plus de 87 espèces de protéobactéries dont la quasi-totalité des espèces d'entérobactéries (Sallen *et al.*, 1995), *Pseudomonas* (Levesque *et al.*, 1995), *V. cholerae* (Dalsgaard *et al.*, 1999), *Acinetobacter* (Seward, 1999), mais également des α -protéobactéries telles que *Ochrobactrum* (Fichier GenBank FJ950723, Li *et al.*, non publié) ou des ϵ -protéobactéries (*Campylobacter*, (Gibreel et Skold, 1998)).

Les descriptions d' IM 1 chez des bactéries à Gram positif sont plus rares et concernent essentiellement le genre *Corynebacterium* (Nesvera, Hochmannova et Patek, 1998) et *Staphylococcus* (Xu *et al.*, 2008). Les IM 2 ont été décrits principalement chez des γ -protéobactéries, le genre *Shigella* étant un hôte fréquent (Ahmed *et al.*, 2006 ; Gassama-Sow *et al.*, 2006) et plus récemment chez une bactérie à Gram positif, *Aerococcus viridans*, d'origine animale (Byrne-Bailey *et al.*, 2009). Enfin, si les IM 3 semblent limités dans le contexte clinique, ils pourraient être plus fréquents chez les β -protéobactéries environnementales (Xu, Davies et Miao, 2007).

2. Les super-intégrons (SI)

Les SI se distinguent des IM par trois caractères (Mazel, 2006 ; Mazel *et al.*, 1998) :

- (i) Les SI sont toujours chromosomiques alors que les IM sont véhiculés par des plasmides ou des transposons.
- (ii) Le nombre de cassettes de gènes est le plus souvent élevé, notamment chez les espèces de *Vibrio*, où il varie de 72 à plus de 200.
- (iii) Les sites *attC* des cassettes d'un même SI sont très conservés, tandis qu'ils varient fortement en taille et en séquence entre les cassettes d'un même IM.

Des SI ont été identifiés dans le génome d'un large panel de bactéries incluant γ -protéobactéries (*Vibrio*, *Xanthomonas*, *Pseudomonas*), β -protéobactéries (*Nitrosomonas*), δ -protéobactéries (*Geobacter*) mais également des Planctomycètes (*Rhodopirellula*, *Candidatus*) et une spirochète (*Treponema*) (Mazel, 2006).

C- Origine et évolution des intégrons

Les intégrons existaient déjà bien avant l'introduction des antibiotiques, comme en témoigne la présence d'un SI sur le chromosome d'une souche de *Vibrio metschnikovii* datant de 1888 (Mazel *et al.*, 1998). De plus, la comparaison des phylogénies des gènes *intI* des SI avec celle des génomes de leurs organismes hôtes montrent une coévolution ancienne de ces structures, couplée à des transferts horizontaux de gènes (Mazel, 2006).

L'origine des cassettes n'est pas connue. L'absence de promoteur dans ces éléments laisse penser qu'elles sont issues d'éléments mobiles dont la mobilité inclut une étape de transcription (rétroéléments : rétrotransposon, introns). Récemment, Leon et Roy ont proposé un modèle où les cassettes sont formées par recombinaison homologue entre deux copies d'introns, l'une étant liée à un gène, l'autre à un site *attC* (Leon et Roy, 2009).

Enfin, les SI sont considérés comme les réservoirs de cassettes pour les IM. Cette hypothèse est fondée sur :

- la capacité des *IntI* de IM à mobiliser les cassettes de SI et les insérer dans un IM
- l'identité entre certains sites *attC* d'IM avec ceux des SI.

Un IM présente alors des cassettes provenant des SI de plusieurs espèces, identifiables grâce à la séquence caractéristique de leurs sites *attC*.

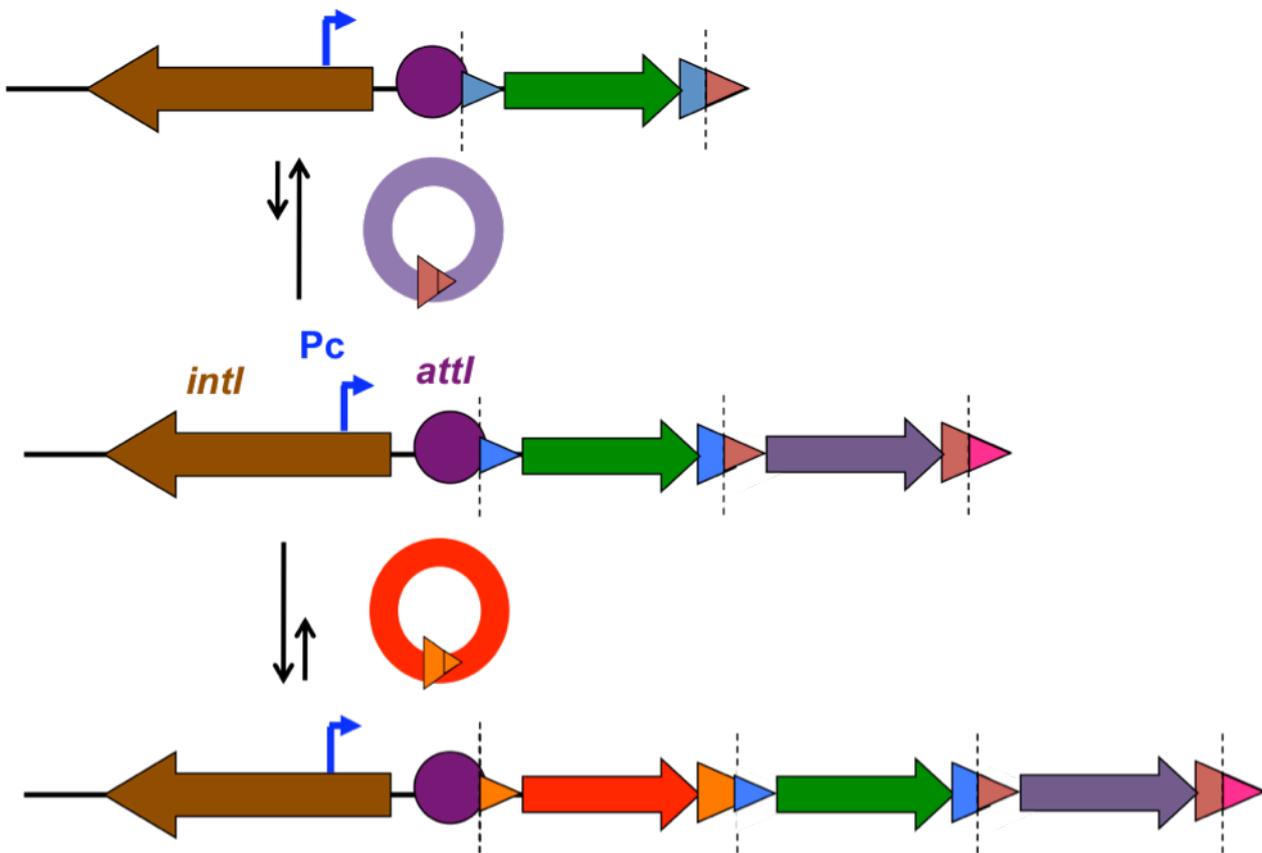


Figure 15: *Dynamique des mouvements de cassettes d'intégrons*

L'intégron de départ (au milieu de la figure) évolue préférentiellement par insertion de nouvelles cassettes en attI (ce qui génère l'intégron du bas) ou par excision des cassettes qui ne sont pas en première position (ce qui génère l'intégron du haut). Ces réactions sont catalysées par l'intégrase IntI. La taille des flèches indique la préférence du sens de la recombinaison.

D- Dynamique d'échange des cassettes

1. Modèle d'échange des cassettes de gènes

Les intégrases d'intégrons peuvent catalyser des recombinaisons aux sites *attI* et *attC* selon trois réactions avec des efficacités variables : *attC* x *attC*, *attI* x *attC* et *attI* x *attI*. La recombinaison entre les sites *attI* est significativement moins efficace que celle entre les sites *attC* et la réaction *attI* x *attC* est la préférée d'entre toutes (Collis *et al.*, 2001).

Par conséquent, l'insertion d'une nouvelle cassette dans l'intégron se fait préférentiellement au site *attI* plutôt qu'au centre du réseau (Figure 15). Selon une étude récente, les recombinaisons préférentielles varieraient selon la concentration intracellulaire en IntI et la phase de croissance. (Shearer et Summers, 2009).

Des sites secondaires de consensus GWTMW (W : A ou T ; M : A ou C) peuvent également être recombinaisonnés avec un site *attC* ou un site *attI*, mais beaucoup moins efficacement dans les deux cas (Collis *et al.*, 2001 ; Ramirez *et al.*, 2008).

2. Mécanisme de recombinaisons

a- Relation structure-fonction des IntI

La structure 3D de l'intégrase du SI de *V. cholerae* liée à son substrat *attC* a été résolue et peut être inférée aux autres IntI (MacDonald *et al.*, 2006).

Les protéines IntI présentent les caractéristiques structurales communes aux recombinases à tyrosine. Le domaine N-terminal contient quatre hélices α organisées en motifs hélice-boucle-hélice qui sont impliquées dans la liaison au site de recombinaison et les interactions entre les protéines. La partie C-terminale contient les régions hautement conservées caractéristiques des recombinases à tyrosine qui abritent les résidus essentiels à l'activité de recombinaison (Nunes-Duby *et al.*, 1998).

Les intégrases d'intégrons se distinguent des autres recombinases à tyrosine par la présence d'un domaine additionnel qui inclut une hélice supplémentaire I₂ intervenant dans la formation de la synapse de recombinaison.

b- Mécanisme de recombinaison

Des expériences *in vivo* ont montré que les enzymes IntI emploient un mécanisme atypique de recombinaison du fait de la nature simple brin du site *attC* recombinaison (Bouvier, Demarre et Mazel, 2005). Deux molécules d'IntI1 se lient sur un site *attI* double brin ou sur le brin « bottom » d'un site *attC*. Ce dernier peut être généré au passage d'une fourche de réplication, ou à la suite de lésions de l'ADN, comme après un stress oxydatif par exemple. De plus, il est structuré en épingle à cheveux grâce à sa séquence palindromique. Dubois *et coll.* ont proposé que la formation de cet épingle à cheveux requiert la liaison préalable d'une ou plusieurs molécules d'IntI sur le site *attC* double brin (Dubois *et al.*, 2007).

c- Spécificité des protéines IntI envers les sites de recombinaison

Chaque intégrase IntI présente une certaine spécificité pour son site *attI* mais reconnaît tous les sites *attC* (Mazel, 2006). La protéine IntI1 se montre toutefois la plus tolérante à cet égard et peut catalyser des recombinaisons avec *attI2* ou *attI3* (Biskri *et al.*, 2005 ; Collis et Hall, 2004).

E- Expression des gènes d'intégrons

La spécificité des intégrons réside dans leur capacité à véhiculer des gènes de résistance aux antibiotiques tout en assurant leur expression.

Ces deux fonctions requièrent la présence de signaux d'initiation de transcription et de traduction pour les gènes des cassettes et pour le gène de l'intégrase.

1. Expression des gènes de cassettes

L'expression des phénotypes de résistance codés par les gènes des cassettes requiert la présence de signaux pour initier la transcription du gène (promoteurs) ainsi que la traduction des ARN correspondants (TIR). Si un certain nombre de cassettes possède déjà leur propre TIR, la grande majorité est dépourvue de promoteur interne (Hanau-Bercot *et al.*, 2002)). Les gènes de cassettes sont transcrits à partir du promoteur commun Pc, comme dans un opéron (Stokes et Hall, 1989). Le promoteur Pc n'a été caractérisé que pour les IM 1 et 3, et est situé dans la séquence codante du gène *intI* (Collis et Hall, 1995 ; Collis *et al.*, 2002).

a- Le promoteur Pc des IM 1

Chez les IM 1, le promoteur Pc présente une organisation typique des promoteurs σ^{70} : un élément -35 séparé par 17 pb d'un élément -10 avec pour (+1) de transcription une base « A » 8 pb en aval. Pc est situé dans la séquence codante du gène *intII*, l'élément -10 se terminant au niveau du codon 30, soit à 84 pb du codon START (Collis et Hall, 1995).

A.

Promoteur	Élément -35	Espaceur (pb)	Élément -10	Référence ou numéro d'accès
Consensus σ^{70}	TTGACA	17+/-1	TATAAT	Hawley et McClure, 1983
PcS (Strong)	TTGACA	17	TAAACT	U12441
PcW (Weak)	TGGACA	17	TAAGCT	U49101
PcH1 (Hybrid 1)	TGGACA	17	TAAACT	M95287
PcH2 (Hybrid 2)	TTGACA	17	TAAGCT	U13880
PcSS (Super-Strong)	TTGATA	17	TAAACT	Brizio <i>et al.</i> , 2006
PcIn42	TTGGCA	17	TAAACT	AJ243491
PcIn116	TTGACA	17	TGAACT	AJ621187
PcPUO	TCGACA	17	TAAACT	S68049

B.

P2	TTGTTA	17	TACAGT	Collis et Hall, 1995
P2m1 (mutée 1)	TTGTTA	17	GACAGT	Naas <i>et al.</i> , 2001
P2m2 (mutée 2)	TTGTTA	17	TACACA	Lindstedt <i>et al.</i> , 2003

Tableau 5 : Séquence des variants des promoteurs Pc et P2 des IM 1.

La séquence consensus des promoteurs σ^{70} est rappelée dans la première ligne.

A. Séquences des variants de Pc. Les bases qui diffèrent du variant PcS sont soulignées

B. Séquences des variants de P2. Les bases qui diffèrent de la forme P2 usuelle sont soulignées

Plusieurs variants de Pc ont été identifiés, sur la base de variations dans les éléments -35 et -10. A ce jour, huit ont été décrits mais seulement cinq étudiés pour leur force (Brizio *et al.*, 2006 ; Houang *et al.*, 2003 ; Levesque *et al.*, 1994 ; Power *et al.*, 2005 ; Riccio *et al.*, 2000) (Tableau 5A).

Globalement, les forces relatives de ces variants varient en fonction de leur degré d'identité avec la séquence consensus pour les promoteurs à σ^{70} établie chez *E.coli*. Le variant le plus fort, PcS (Strong) est de 20 à 30 fois plus efficace que le PcW (Wweak) ; les variants PcH1 (Hybride 1) et PcH2 (Hybride 2), qui combinent les éléments -35 et -10 de PcS et PcW, sont de force intermédiaire, PcH2 étant le plus efficace, suggérant que c'est la mutation de l'élément -35 (TTGACA → TGGACA) qui est la plus délétère pour l'activité de Pc (Collis et Hall, 1995 ; Papagiannitsis, Tzouvelekis et Miriagou, 2009).

Brizio *et coll.* ont caractérisé un variant de Pc (désigné ici PcSS pour Super-Strong) qui serait plus efficace que PcS bien que moins proche de la séquence consensus que ne l'est le variant PcS (Brizio *et al.*, 2006).

Les variations dans la séquence du promoteur Pc déterminent également des changements de structure primaire de l'intégrase IntI1. Ainsi, le résidu 39 d'IntI1 est une histidine lorsque Pc est PcW ou PcH1, mais une asparagine lorsque Pc est PcS ou PcH2. L'impact du polymorphisme de Pc sur sa force et sur l'activité de l'intégrase IntI1 fait l'objet de l'étude du premier travail décrit dans ce manuscrit.

✓ *Epidémiologie et distribution des variants de Pc*

Les quelques études épidémiologiques traitant de la prévalence des variants de Pc indiquent une forte prédominance du PcW, en particulier chez les intégrons de *Salmonella* (Lindstedt *et al.*, 2003 ; Schmitz *et al.*, 2001).

Selon les études, il est suivi soit du PcS dans une étude menée sur plusieurs espèces d'entérobactéries (Schmitz *et al.*, 2001)), soit du PcH1 dans une autre ciblant un groupe d'isolats d'*E. coli* (Cocchi *et al.*, 2007). Aucun de ces articles ne mentionne la présence d'un des quatre autres variants de Pc, qui semblent extrêmement rares.

Dans le travail de Schmitz *et coll.*, la proportion d'intégrons portant des variants de Pc forts augmente d'autant plus que les souches ont été isolées récemment : 14% (n=14) en 1993, 24% (n=25) en 1996, 46% (n=50) en 1999 (Schmitz *et al.*, 2001).

Ces études de prévalence ne tiennent pas compte des réseaux de cassettes des IM 1 analysés. Or deux IM 1 ayant le même réseau de cassettes peuvent être associés à un même variant de Pc (Schmitz *et al.*, 2001) ou des variants distincts (Peters *et al.*, 2001; A.P. Solomenny, 2009)).

✓ *Diversité de Pc au delà des variants*

Nous avons vu qu'outre les éléments -35 et -10, d'autres régions d'un promoteur contribuent à son efficacité. Par conséquent, la définition initiale des variants de Pc basée uniquement sur ces éléments ne suffit pas à décrire son polymorphisme. Ainsi, un certain nombre d'intégrons avec PcW présentent une mutation de la deuxième base en amont de l'hexamère -10, communément un résidu C, en G, ce qui augmente de cinq fois son efficacité, probablement parce que cela crée une région -10 étendue fonctionnelle (Nesvera, Hochmannova et Patek, 1998).

a- Le promoteur P2

Certains IM 1 présentent un second promoteur P2 qui contribue à l'expression des cassettes, localisé dans *attII*, à 119 pb en aval de Pc (Stokes et Hall, 1989). Ce promoteur est créé via l'insertion de trois résidus guanine, ce qui corrige l'espacement entre des motifs -35 et -10 potentiels et leur permet ainsi de fonctionner comme tel. P2 est le plus souvent associé au PcW, plus rarement au PcH1 ou PcS (Papagiannitsis, Tzouvelekis et Miriagou, 2009).

	Élément -35	Espaceur (pb)	Élément -10	Références
Consensus σ^{70}	TTGACA	17+/-1	TATAAT	Hawley et McClure, 1983
Variant de Pc				
Variant 1	TAGACA	17	TAGGCT	In3-1 (Arakawa <i>et al.</i> , 1995) In3-3, In3-4 (Xu, Davies et Miao, 2007)
Variant 2	TAGACA	17	TAGG <u>A</u> T	In3-2 (Correia <i>et al.</i> , 2003)

Tableau 6: Séquence des variants des promoteurs Pc des IM 3.

La séquence consensus des promoteurs σ^{70} est rappelée dans la première ligne.

La base qui diffère entre les deux variants de Pc est soulignée.

Il permet d'augmenter d'autant plus le niveau de transcription des cassettes que le variant de Pc est faible (Papagiannitsis, Tzouveleki et Miriagou, 2009 ; Pournaras *et al.*, 2005), se révélant parfois indispensable pour que la première cassette exprime le phénotype codé par son gène (Tenover *et al.*, 1988). Deux autres variants de P2 ont été identifiés, mais leurs forces respectives n'ont jamais été déterminées (Lindstedt *et al.*, 2003 ; Naas *et al.*, 2001) (Tableau 5B). La seule étude épidémiologique sur le promoteur P2 laisse penser qu'il pourrait être actif chez 20% des IM 1, du moins chez *Salmonella* (Lindstedt *et al.*, 2003).

c- Le promoteur Pc des IM 3

Le promoteur Pc des IM 3 est situé dans la séquence codante du gène *intI3*, à la position correspondant du Pc des IM 1 dans *intI1*. Dans l'IM 3 numéroté In3-2, Pc diffère par une substitution dans l'élément -10 (Tableau 6) (Correia *et al.*, 2003). Ce variant est plus proche de la séquence consensus établie pour les promoteurs bactériens à σ^{70} et devrait donc a priori être plus fort que le Pc des autres IM 3, mais cela n'a pas été montré expérimentalement.

A.

	Élément -35	Espaceur(pb)	Élément -10	Références ou numéro d'accès
Consensus σ^{70}	TTGACA	17+/-1	TATAAT	Hawley et McClure, 1983
<i>Cassettes</i>				
<i>qacH</i>	TTGGCA	19	TATCCT	Ploy, Courvalin et Lambert, 1998
<i>cmlA1</i>	TTGCAG	17	TACGAT*	Bissonnette <i>et al.</i> , 1991
	TTGAAA	16	TTCAAT³	Bissonnette <i>et al.</i> , 1991
	TTGTTA	17	TGAGAT	Bissonnette <i>et al.</i> , 1991
<i>cmlA2</i>	TCGCGG¹	17	TACGAT*	Naas <i>et al.</i> , 2001
<i>ereA1</i>	TGGCGA	17	TACAGC	Biskri et Mazel, 2003
<i>ccdAB²</i>	TTGTTA	17	TATACT	Rowe-Magnus <i>et al.</i> , 2003

B.

<i>qacG</i>	CTGTTT	18	TAAAAT	Laraki <i>et al.</i> , 1999
<i>cmlA4</i>	TTGAAA	16	TTCAAT	Poirel <i>et al.</i> , 2000
	TTGTTA	17	TGAGAT	Poirel <i>et al.</i> , 2000
<i>cmlA10</i>	TCGCTG	17	TACGAT*	AJ867811
<i>catB2</i>	TGGTCT	17	TATGGT	Parent et Roy, 1992
<i>qnrVC1</i>	TTGAGA	17	TAGTCT*	Fonseca <i>et al.</i> , 2008
<i>blaN29</i>	TAGGCA	18	TTTTAT*	Ito et Hirano, 1997
<i>blaP3 (CARB-4)</i>	TTGAAT	14	TATAAG*	Sanschagrin, Bejaoui et Levesque, 1998
	TTGAAT¹	16	TAAGAT	Ce travail
<i>blaP1 (CARB-2)</i>	ATTATG	18	TAAAAT	Petroni <i>et al.</i> , 2004
<i>blaP2 (CARB-8)</i>	ATTATG	18	TAAAAT	Petroni <i>et al.</i> , 2004
<i>CARB-9²</i>	ATTATG	18	TAAGAT	Petroni <i>et al.</i> , 2004
<i>CARB-6²</i>	ATTATG	18	TAAAAT	Petroni <i>et al.</i> , 2004
<i>CARB-7²</i>	ATTTCA	18	TAAGAT	Petroni <i>et al.</i> , 2004

Tableau 7 : Séquence des promoteurs putatifs de cassettes de gènes

La séquence consensus des promoteurs σ^{70} est rappelée dans la première ligne. Les bases identiques à la séquence consensus sont en gras.

Les promoteurs dont la séquence de l'élément -10 est suivie d'un astérisque (*) présentent un motif TGN-10 qui caractérise la région « -10 étendue ». N : A, C, G ou T

A. Séquences de promoteurs internes de cassettes pour lesquels l'activité a été démontrée in vitro

Les gènes de ces cassettes sont transcrits indépendamment de Pc.

¹ La séquence de l'élément -35 est TCGCGG dans l'article, mais TCGCAG dans le fichier GenBank associé (Fichier GenBank AF205943)

² Cette cassette est exclusivement associée à un SI.

B. Séquences des promoteurs putatifs proposés pour d'autres cassettes.

La présence d'un promoteur actif dans ces cassettes n'a pas été vérifiée.

¹ Ce promoteur est une proposition alternative à celle des auteurs.

² Cette cassette est exclusivement associée à un SI.

d- Les promoteurs internes aux cassettes

La grande majorité des cassettes présente un espacement insuffisant entre le site de recombinaison et le début de leur gène pour abriter un promoteur fonctionnel.

Cependant, les cassettes *qacH*, *cmlA1*, *cmlA2* et *ereA1*, possèdent leur propre promoteur (Biskri et Mazel, 2003 ; Bissonnette *et al.*, 1991 ; Naas *et al.*, 2001 ; Ploy, Courvalin et Lambert, 1998 ; Rowe-Magnus *et al.*, 2003). C'est également le cas des deux gènes de la cassette *ccdAB* qui codent une toxine (un inhibiteur de gyrase) et son antidote dans le SI de *V. fischeri* (Rowe-Magnus *et al.*, 2003). Des promoteurs potentiels ont été proposés pour ces différents gènes, mais aucun n'a été caractérisé à ce jour (Tableau 7A). Pour d'autres cassettes, des régions promotrices bien conservées ont été repérées en amont des gènes, mais leur activité potentielle n'a pas été testée (Tableau 7B) (Fonseca *et al.*, 2008 ; Laraki *et al.*, 1999 ; Parent et Roy, 1992 ; Poirel *et al.*, 2000 ; Sanschagrin, Bejaoui et Levesque, 1998). Notre analyse des promoteurs putatifs internes aux cassettes a montré qu'un certain nombre présentent une région «-10 étendue » non identifiée par les auteurs (Tableau 7A et B).

Enfin, des études transcriptomiques suggèrent que la plupart des gènes de cassettes du SI de *V.cholerae* sont davantage exprimés en phase stationnaire, en réponse à un stress ou une forte densité cellulaire, probablement via la participation de facteurs σ alternatifs (Yildiz *et al.*, 2004).

e- Autres promoteurs pour les gènes de cassettes

L'accumulation de séquences d'IM, en particulier d'IM 1, a révélé l'existence d'autres promoteurs actifs pour l'expression des cassettes de gènes situés ni dans la plate-forme fonctionnelle, ni au sein même de la cassette.

✓ *Les pseudocassettes d'expression*

D'autres gènes de cassettes sont transcrits à partir de promoteurs situés en amont des cassettes. Ainsi, la plupart des cassettes *blaOxa10* et *blaOxa10-aadA1* sont précédées d'une cassette cryptique dépourvue de gène et d'une partie du site *attC* mais contenant un promoteur actif (Naas *et al.*, 2001). De même, un fragment de 101 pb contenant le promoteur de *qacEΔ1* est présent en amont de plusieurs cassettes, dont toutes les *aac(6')-Ib-cr*, *aac(3)-VI* et *aadA16* (Partridge *et al.*, 2009). Ce promoteur suffit au gène de la cassette *aac(3)-VI* pour qu'il confère le phénotype de résistance correspondant (Rather *et al.*, 1993).

✓ *Expression des cassettes depuis un promoteur dans une IS*

Certains intégrons présentent des IS qui procurent des séquences promotrices efficaces pour la transcription des cassettes. L'expression des gènes des cassettes *blaVEB* insérées en première position des IM 1 est augmentée lorsqu'elle est précédée de l'IS1999. Cet effet est lié à un promoteur Pout de l' IS qui coopère avec le promoteur Pc (Aubert, Naas et Nordmann, 2003). Il a également été proposé que l'IS10 insérée entre *attI1* et *blaOxa-2* dans l'intégron de pB10 soit responsable de l'expression du gène *blaOxa-2* (Szczepanowski *et al.*, 2004). En effet, les IS de cette famille ont un promoteur Pout orienté vers l'extérieur de l'IS et connu pour activer certains gènes (Pour revue, (Mahillon et Chandler, 1998)).

d- Traduction des gènes de cassettes

✓ Signaux d'initiation de traduction

La plupart des gènes de cassettes possèdent une bonne région d'initiation de traduction (TIR) potentielle, constituée d'un site de liaison des ribosomes (ou RBS pour Ribosome Binding Site ; de consensus AAGGA) espacé de manière optimale d'un codon d'initiation (Hanau-Bercot *et al.*, 2002)

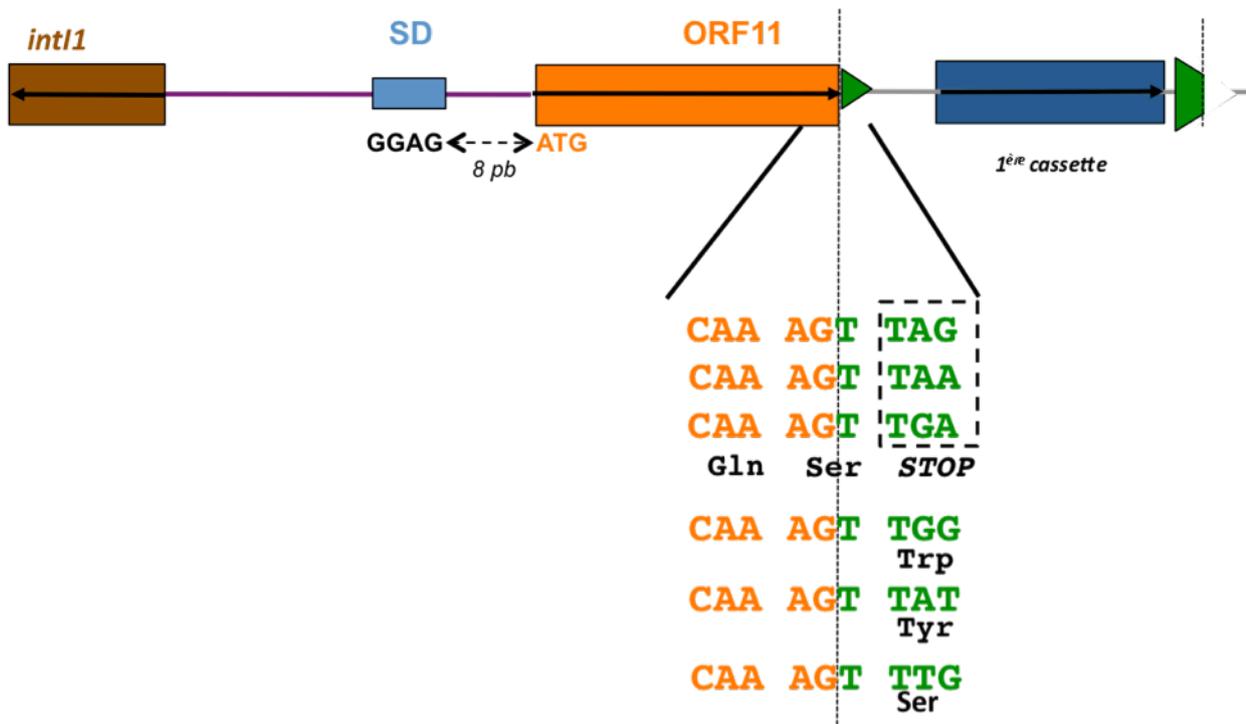


Figure 16: Structure d'ORF 11 selon la nature de la première cassette.

La structure générale de l'IM 1 est rappelée avec les mêmes symboles que précédemment. L'ORF11 (rectangle orange), dont le sens de transcription est rappelé par une flèche, commence dans la région attII et se termine dans la première cassette, le plus souvent grâce à un codon STOP situé dans la boîte R' du site « core » (petit triangle vert). Les différents codons possibles (codons STOP ou codant) à ce niveau sont indiqués. Cette ORF est précédée d'une bonne TIR, constituée d'une RBS (« SD », rectangle bleu), située 8 pb en amont d'ORF11.

✓ Traduction des cassettes en première position des IM 1

De nombreux IM 1 présentent un cadre ouvert de lecture (ORF) qui commence à l'extrémité 3' de la région *attII* et se termine par un codon STOP apporté par la première cassette (Figure 16). Cette ORF appelée ORF11 est précédée d'une TIR efficace (GGAG-N8-ATG) et assure une traduction du gène de la première cassette. Hanau-Bercot et coll. ont proposé un modèle de couplage traductionnel dans lequel une fraction des ribosomes recrutés par la TIR d'ORF11, après avoir traduit ORF11, réinitie une traduction au niveau du gène de la première cassette (Hanau-Bercot *et al.*, 2002).

La taille du peptide codé par ORF11, généralement 12 aa, dépend de la première cassette (Figure 16). Elle peut être réduite à 8 aa si l'ORF est mutée ou rallongée jusqu'à 21 aa en fonction de la position du codon STOP dans la cassette.

D'autre part, certains IM 1 présentent une duplication des 19 premières pb d'ORF11. Généralement, cela entraîne la formation d'une ORF18 fusionnée au gène de la première cassette, (Casin *et al.*, 1998).

✓ Influence de la position de la cassette

Collis et coll. ont constaté que le niveau d'expression d'un même gène de cassette est d'autant plus réduit que la cassette était éloignée de Pc. Ces auteurs proposent qu'en formant des structures tige-boucle, les sites *attC* constituent des atténuateurs de transcription (Collis et Hall, 1995). En fait, la présence d'une série de T immédiatement en aval de la tige-boucle est requise pour former un atténuateur de transcription (Yarnell et Roberts, 1999), ce dont sont dépourvus les sites *attC*. D'ailleurs, Jacquier *et coll.* ont récemment montré que l'influence des sites *attC* s'exerce au niveau de la traduction du gène, et non de la transcription : les tiges-boucles seraient des barrières physiques affectant, à des amplitudes variables, le transit des ribosomes le long des transcrits (Jacquier *et al.*, 2009).

Promoteur	Élément -35	Espaceur (pb)	Élément -10	Référence ou numéro d'accès
Consensus σ^{70}	TTGACA	17+/-1	TATAAT	Hawley et McClure, 1983
Classe d'intégron				
IM 1	TTGCTG	17	TAGACT	Guerin <i>et al.</i> , 2009
IM 2	CAGGCA	17	TAGAAT	L10818
IM 3	TTGAAA	17	CATACT	Collis <i>et al.</i> , 2002
	CGGATA	17	TAAACA	Collis <i>et al.</i> , 2002
IM 4	<u>TAGTTA</u>	<u>17</u>	<u>TAATTT</u>	Ce travail
IM 5	TTGGAC	17	CTATAT ¹	Guerin <i>et al.</i> , 2009
	<u>TGGACA</u>	<u>17</u>	<u>TATATA</u>	Ce travail
SIV _{ch}	TTGACA	17	TACTGT	Guerin <i>et al.</i> , 2009
SIV _{me}	TTGACA	17	CACTGT	Guerin <i>et al.</i> , 2009
Classe 6	TGGTTT	17	TATAAT	Nield <i>et al.</i> , 2001
Classe 7	TTGAAA	18	TAGATT	Nield <i>et al.</i> , 2001
Classe 8	TCCATA	17	AATAAT	Nield <i>et al.</i> , 2001
	<u>TTGATT</u>	<u>17</u>	<u>TACTGT</u>	Ce travail

Tableau 8 : Séquence des promoteurs P_{intI} des gènes d'intégrase d'intégrons.

La séquence consensus des promoteurs σ^{70} est rappelée dans la première ligne.

Seul le promoteur des IM 1 a été caractérisé, les autres séquences correspondent à des propositions non expérimentalement vérifiées, soit des auteurs, soit de ce travail (séquences soulignées). Deux promoteurs ont été proposés pour les IM 3.

¹ Cette séquence présente en outre une région « -10 étendue » non observée par les auteurs.

Modulation du niveau d'expression

Le niveau d'expression d'un gène d'une cassette insérée dans un intégron dépend donc de plusieurs paramètres, dont :

- (i) La présence et la séquence du promoteur Pc et/ou P2
- (ii) La présence d'un promoteur additionnel (interne à la cassette ou dans une autre structure en amont)
- (iii) La présence de signaux de traductions internes
- (iv) Le nombre et la nature des cassettes en amont

Par ailleurs, il est aussi influencé par le nombre de copies (i) de la cassette au sein d'un intégron, (ii) de l'intégron au sein du support génétique qui l'héberge.

2. Expression du gène de l'intégrase des intégrons

Le promoteur *PintI* contrôlant l'expression du gène *intI* n'a été décrit que pour les IM 1 (Guerin *et al.*, 2009). Ce promoteur *PintI1* (TTGCTG-N17-TAGACT) est localisé 22 pb en amont du codon START d'*intI1* ; il est très conservé par rapport à la séquence consensus des promoteurs σ^{70} . D'autre part, il fait face au promoteur Pc, et chevauche le promoteur P2, une situation propice aux interférences transcriptionnelles. Des séquences de promoteurs potentiels ont été proposées pour les gènes *intI* d'autres IM et quelques SI (Tableau 8).

Au niveau traductionnel, la présence d'une RBS n'a été étudiée pour aucune classe d'intégron : il est possible de repérer des RBS proches du consensus, mais seulement pour les gènes *intI* des IM 2 et de quelques SI. De plus, un site de fixation pour le répresseur LexA chevauche la région promoteur chez la plupart des IM et de nombreux SI (Guerin *et al.*, 2009). La protéine LexA réprime un ensemble de gènes impliqués dans la réponse SOS, mise en place à la suite de tout stress entraînant des dommages à l'ADN.

Il a été montré que la transcription des gènes *intI1* des IM 1 et *VchintIA* du SI de *V.cholerae* est LexA-dépendante et régulée par la réponse SOS. Leur expression peut donc être induite par certains antibiotiques connus pour provoquer un stress oxydatif (quinolones, β -lactamines, triméthoprimine).

Résultats

Présentation du travail de thèse

Nos recherches ont porté sur l'expression des gènes de résistance aux antibiotiques et de l'intégrase au sein des intégrons de multirésistance (IM). Les IM sont des systèmes élaborés de capture, d'expression et de dissémination de gènes de résistance contenus dans des cassettes, chez les bactéries à Gram négatif. Ils codent une enzyme, l'intégrase, qui catalyse l'insertion/excision des cassettes dans l'intégron et fournissent aux cassettes insérées un promoteur Pc pour la transcription de leur gène, comme dans un opéron. Plusieurs classes d'IM ont été définies sur la base de la séquence de l'intégrase, les IM de classe 1 et 2 (IM 1 et IM 2) étant les plus prévalents. Le promoteur Pc des IM 1 existe en plusieurs variants qui diffèrent par la séquence des éléments -35 et -10 et par leur force. Ce promoteur Pc est localisé dans la séquence codante du gène de l'intégrase, si bien que les variations de Pc modifient également la séquence de l'intégrase. Certains IM 1 présentent en outre un second promoteur P2, en aval de Pc, qui participe à l'expression des gènes de cassettes. Bien que des centaines d'IM 1 aient été caractérisées ces dernières années, la diversité et la distribution des séquences promotrices Pc/P2 ne sont pas connues.

C'est pourquoi, dans la **première partie** de ce travail, à partir de l'analyse de l'ensemble des séquences d'IM 1 disponibles *in silico*, nous avons caractérisé le polymorphisme de la région promotrice Pc/P2 et son impact sur le niveau d'expression des cassettes mais aussi sur l'efficacité de l'intégrase. Ce travail fait l'objet d'une publication dans la revue *Plos Genetics*. Nous avons ensuite réalisé une étude, présentée dans la **seconde partie**, chez des souches cliniques pour vérifier si la répartition des variants de Pc et P2 *in vivo* était comparable à celle que nous avons déterminée précédemment *in silico*.

Enfin, le promoteur Pc et celui contrôlant l'expression du gène de l'intégrase PintI1 se faisant face chez les IM 1, nous avons étudié si ces deux promoteurs interfèrent ; les résultats de ce travail sont détaillés dans la **troisième partie**.

Par ailleurs, nous avons, dans la **quatrième partie**, étendu nos travaux aux IM 2 assez peu étudiés bien que fréquemment décrits chez des souches humaines et animales. Nous avons caractérisé les promoteurs du gène de l'intégrase et des gènes de cassettes de cette classe d'IM.

Les travaux présentés dans les parties 2 et 4 sont en cours d'écriture sous forme d'articles pour publication dans des revues internationales. Le travail présenté dans la partie 3 doit être poursuivi afin de disposer de résultats complémentaires avant publication.

Partie n° 1

**Le polymorphisme du promoteur Pc des intégrons de classe
1 affecte l'activité de l'intégrase IntI1**

Le séquençage de nombreux IM 1 a révélé un polymorphisme du promoteur Pc chez cette classe d'intégrons. A ce jour, 8 variants de Pc ont été décrits uniquement sur la base de différences de séquences dans les éléments -35 et -10 (Brizio *et al.*, 2006 ; Levesque *et al.*, 1994 ; Power *et al.*, 2005 ; Riccio *et al.*, 2000). Leur efficacité est variable mais n'a jamais été testée dans un même essai (Levesque *et al.*, 1994). De plus, le promoteur Pc est localisé dans la séquence codante du gène *intI1*, si bien que la séquence de la protéine IntI1 diffère selon les variants de Pc.

Dans ce travail, nous avons analysé *in silico* la diversité et la distribution des variants du promoteur Pc des IM 1 et de la conséquence de ce polymorphisme sur l'activité de l'intégrase IntI1. Les efficacités relatives des variants de Pc ont été déterminées à l'aide d'un même essai enzymatique, dans des fusions transcriptionnelles entre d'une part un fragment d'IM 1 contenant à la fois Pc et P2 et d'autre part un gène *lacZ* dépourvu de promoteur dans le vecteur rapporteur pSU38 Δ tot*lacZ* que nous avons construit. De façon surprenante, ce sont les variants de Pc les moins efficaces pour la transcription des gènes de cassettes qui prédominent, les plus forts étant plus rares (Tableau 1 et Figure 2A de l'article). Cependant, dans deux tiers des IM 1 ayant le promoteur faible PcW, l'activité de ce promoteur est compensée par la présence d'une région -10 étendue ou d'un second promoteur actif P2 (Tableau 2 et Figures 2B et 2C de l'article). Ce dernier, en revanche, n'augmente pas voire diminue le niveau d'expression globale quand il est associé à un variant fort de Pc (Figure 2C de l'article). De plus, il en existe d'autres variants. Globalement, nous avons montré qu'il existait au moins 20 séquences pour la région promotrice Pc-P2, dont nous avons étudié les forces relatives dans un même essai enzymatique.

D'autre part, nous avons étudié l'influence de la séquence (et donc du variant) de Pc sur la structure et l'activité d'IntI1 : les cinq variants principaux de Pc déterminent trois allèles d'IntI1 du fait de variations aux positions 32 et 39 (Tableau 3 de l'article).

En utilisant des tests d'activités de recombinaisons déjà décrits (Guerin *et al.*, 2009; Bouvier, Demarre et Mazel, 2005), nous avons montré que les variants faibles de Pc déterminaient des IntI1 beaucoup plus efficaces pour l'excision des cassettes. En revanche, toutes les IntI1 présentent une activité d'intégration similaire (Figure 3 de l'article). Ces résultats permettent de proposer un rôle pour les résidus Arginine 32 et Histidine 39 (muté(s) quand le Pc est un des variants forts) dans la liaison d'IntI1 avec les sites de recombinaison *attC*. Ainsi, si la plupart des IM 1 sont peu efficaces pour transcrire les gènes de cassettes, ils codent une intégrase plus processive qui facilite l'apparition d'un réarrangement adapté, dans lequel la cassette requise est insérée au site *attI1*, là où elle sera plus fortement exprimée.

Enfin, nous proposons que la pression de sélection antibiotique a contribué à convertir le variant « ancestral » de Pc, PcW, vers des séquences de Pc qui permettent à la fois une meilleure transcription des gènes de cassettes et leurs stabilisations au sein de l'intégron.

Inverse correlation between promoter strength and excision activity in class 1 integrons.

Jové T, Da Re S, Denis F, Mazel D, Ploy MC.

Université de Limoges, Faculté de Médecine, EA3175, Limoges, France.

PLoS Genet. 2010 Jan;6(1):e1000793. Epub 2010 Jan 8.

DOI: [doi : 10.1371/journal.pgen.1000793](https://doi.org/10.1371/journal.pgen.1000793)

Abstract

Class 1 integrons are widespread genetic elements that allow bacteria to capture and express gene cassettes that are usually promoterless. These integrons play a major role in the dissemination of antibiotic resistance among Gram-negative bacteria. They typically consist of a gene (*intI*) encoding an integrase (that catalyzes the gene cassette movement by site-specific recombination), a recombination site (*attI1*), and a promoter (*Pc*) responsible for the expression of inserted gene cassettes. The *Pc* promoter can occasionally be combined with a second promoter designated *P2*, and several *Pc* variants with different strengths have been described, although their relative distribution is not known. The *Pc* promoter in class 1 integrons is located within the *intI1* coding sequence. The *Pc* polymorphism affects the amino acid sequence of *IntI1* and the effect of this feature on the integrase recombination activity has not previously been investigated. We therefore conducted an extensive *in silico* study of class 1 integron sequences in order to assess the distribution of *Pc* variants. We also measured these promoters' strength by means of transcriptional reporter gene fusion experiments and estimated the excision and integration activities of the different *IntI1* variants. We found that there are currently 13 *Pc* variants, leading to 10 *IntI1* variants, that have a highly uneven distribution. There are five main *Pc-P2* combinations, corresponding to five promoter strengths, and three main integrases displaying similar integration activity but very different excision efficiency. Promoter strength correlates with integrase excision activity: the weaker the promoter, the stronger the integrase. The tight relationship between the aptitude of class 1 integrons to recombine cassettes and express gene cassettes may be a key to understanding the short-term evolution of integrons. Dissemination of integron-driven drug resistance is therefore more complex than previously thought.

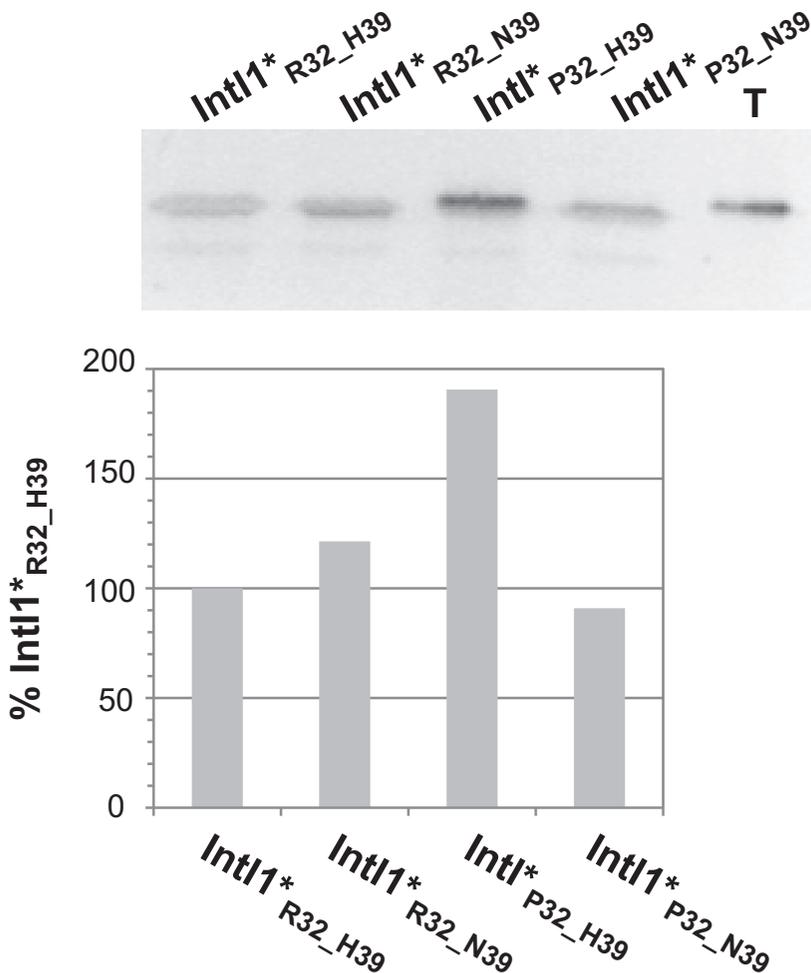


Figure S1: Expression level of the different *Int11* variants

The expression levels were estimated by western blot analysis from cell cultures. The *Int11* variants were expressed from the pBad-*int11** vectors induced by arabinose, the variants being indicated under the graph. T: purified *Int11*_{R32_H39}.

Integrase protein quantification

The amount of IntI1* variants in cells was estimated by Western blot. *E. coli strain* MG1656(pSU38-*attI*) cells expressing one of the IntI1* variants from pBad-*intI1** were grown over-day in LB broth supplemented with ampicillin, kanamycin and 1% glucose, at 37°C, then diluted 1:100 in fresh LB broth supplemented with ampicillin, kanamycin and 0.2% arabinose and grown at 37°C overnight. Cells were pelleted, resuspended in gel loading buffer, boiled, and subjected to 12% SDS-PAGE. Proteins were transferred onto PVDF membranes and the membranes were probed with polyclonal anti-IntI1 peptide antibodies. The immunoblots were revealed with the WesternDotTM625 detection system (Molecular Probes) and the signals were quantified with the quantity One® software (Biorad). Results are expressed as a percentage of IntI1*_{R32;H39}.

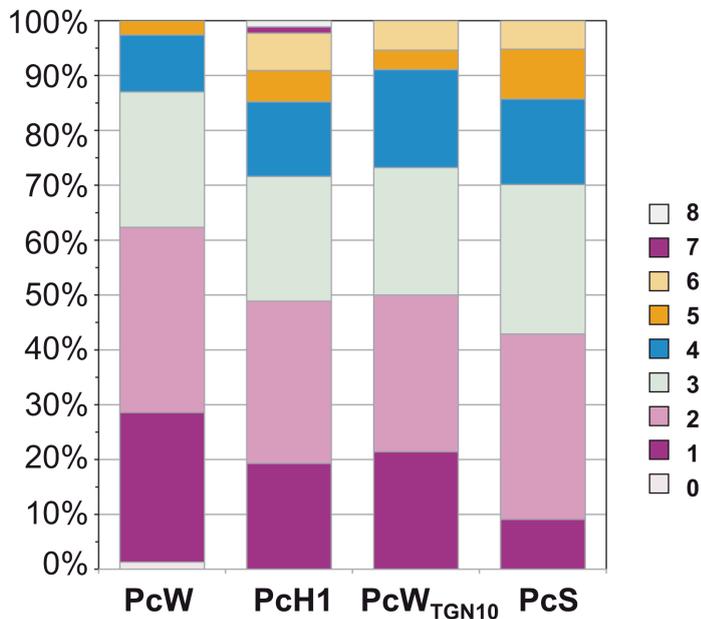


Figure S2: Distribution of the number of gene cassettes in arrays depending on the Pc variant. The number of gene cassettes was determined for the 321 integrons analyzed. The results were sorted according to the Pc variant controlling the cassette array transcription: PcW (n=77), PcH1 (n=88), PcW_{TGN-10} (n=56) and PcS (n=77).

Table S1: List and characteristics of the 321 analyzed integrons

Pc variant	P2 form	Number of Gene Cassettes	Accession Number
PcW		0	U49101, M73819
PcW		1	L06157
PcW		1	D43625
PcW		1	AJ420072
PcW		1	FJ457611
PcW		1	AY257539
PcW		1	X12868
PcW		1	X58425
PcW		1	EU598463
PcW		1	DQ352176
PcW		1	AB281182
PcW		1	BX248359
PcW		1	X53635
PcW		1	AF191564
PcW		1	EU327990
PcW		2	EU327991
PcW		2	AJ877225
PcW		2	EF614235
PcW		2	DQ767903
PcW		2	EU503121
PcW		2	AF024602
PcW		2	AJ289190
PcW		2	AF313471
PcW		2	CP000645
PcW		2	AM261282
PcW		2	Z50802
PcW		2	AJ628353

PcW		2	CP001232
PcW		2	EF592570
PcW		2	DQ372711
PcW		2	AJ863570
PcW		2	AJ564903
PcW		3	EU886977
PcW		3	EU675686
PcW		3	DQ914960
PcW		3	L36547
PcW		3	EU886979
PcW		3	AF313472
PcW		3	AY214164
PcW		3	EU487199
PcW		3	Z21672
PcW		3	DQ393569
PcW		3	AJ698325
PcW		3	EF027105
PcW		3	DQ236170
PcW		3	AY115475
PcW		3	EF577408
PcW		3	EU327989
PcW		4	AJ971344
PcW		4	DQ522237
PcW		4	DQ831140
PcW		4	AY536743
PcW		5	EU327987
PcW	P2	1	X12870
PcW	P2	1	CP000603
PcW	P2	1	X64368

PcW	P2	1	EF660563
PcW	P2	1	EF190326
PcW	P2	1	AF329699
PcW	P2	1	U81136
PcW	P2	2	AJ878850
PcW	P2	2	AF231133
PcW	P2	2	AY136758
PcW	P2	2	AF315786
PcW	P2	2	EU052800
PcW	P2	2	AF174129
PcW	P2	2	AF326777
PcW	P2	2	AJ746361
PcW	P2	2	AY289608
PcW	P2	3	AY758206
PcW	P2	3	AM055748
PcW	P2	3	AF347074
PcW	P2	4	EF636461
PcW	P2	4	AJ584652
PcW	P2	4	AJ969235
PcW	P2	5	AF156486
PcW	P2m2	4	[1]

PcW _{TGN-10}		1	CT025832
PcW _{TGN-10}		1	AF261825
PcW _{TGN-10}		1	AF071555
PcW _{TGN-10}		1	AF486817
PcW _{TGN-10}		1	AF071555
PcW _{TGN-10}		1	EU847537
PcW _{TGN-10}		1	FJ183463
PcW _{TGN-10}		1	AB263754

PcW _{TGN-10}	1	EU847537
PcW _{TGN-10}	1	AB207867
PcW _{TGN-10}	1	AM886293
PcW _{TGN-10}	1	EF427691
PcW _{TGN-10}	2	AY963803
PcW _{TGN-10}	2	AY522431
PcW _{TGN-10}	2	AM295980
PcW _{TGN-10}	2	EF408254
PcW _{TGN-10}	2	AB207867
PcW _{TGN-10}	2	AY069972
PcW _{TGN-10}	2	X17477
PcW _{TGN-10}	2	CT025832
PcW _{TGN-10}	2	AY146989
PcW _{TGN-10}	2	AB366440
PcW _{TGN-10}	2	AB469046
PcW _{TGN-10}	2	EF118171
PcW _{TGN-10}	2	AF133699
PcW _{TGN-10}	2	DQ023222
PcW _{TGN-10}	3	DQ219465
PcW _{TGN-10}	3	DQ315789
PcW _{TGN-10}	3	DQ522234
PcW _{TGN-10}	3	DQ322593
PcW _{TGN-10}	3	AY522923
PcW _{TGN-10}	3	AM040708
PcW _{TGN-10}	3	DQ902344
PcW _{TGN-10}	3	FJ384365
PcW _{TGN-10}	3	AF322577
PcW _{TGN-10}	3	AJ867812
PcW _{TGN-10}	3	AJ511268

PcW _{TGN-10}		3	AM392427
PcW _{TGN-10}		3	DQ342344
PcW _{TGN-10}		4	EU886980
PcW _{TGN-10}		4	AY943084
PcW _{TGN-10}		4	AJ871915
PcW _{TGN-10}		4	EF051038
PcW _{TGN-10}		4	AY920928
PcW _{TGN-10}		4	AJ609296
PcW _{TGN-10}		4	EF125010
PcW _{TGN-10}		4	AF078527
PcW _{TGN-10}		4	Y18050
PcW _{TGN-10}		4	AY560837
PcW _{TGN-10}		5	AJ971342
PcW _{TGN-10}		5	AJ634050
PcW _{TGN-10}		6	AY907717
PcW _{TGN-10}		6	EF051037
PcW _{TGN-10}		6	CT025832
PcW _{TGN-10}	P2	2	AJ009819/AJ310778
PcW _{TGN-10}	P2m1	2	DQ315788
PcW _{TAN-10}		2	AB040994
<hr/>			
PcH1		1	FJ594767
PcH1		1	AF439785
PcH1		1	DQ247972
PcH1		1	AJ517791
PcH1		1	AF458081
PcH1		1	DQ108615
PcH1		1	FJ223605
PcH1		1	AJ517790
PcH1		1	EU715254

PcH1	1	U31119
PcH1	1	FJ167861
PcH1	1	EU491959
PcH1	1	CP000744
PcH1	1	AJ586617
PcH1	1	AM778842
PcH1	2	AF355189
PcH1	2	AB366442
PcH1	2	AY033653
PcH1	2	AF043381
PcH1	2	AJ561197
PcH1	2	AJ639924
PcH1	2	AY509004
PcH1	2	Z83311
PcH1	2	FJ183470
PcH1	2	AY259085
PcH1	2	AY878718
PcH1	2	AM886293
PcH1	2	AM746675
PcH1	2	CP000971
PcH1	2	DQ125241
PcH1	2	AB113580
PcH1	2	AB116260
PcH1	2	AJ628135
PcH1	2	AJ550807
PcH1	2	EF184215
PcH1	2	EU591509
PcH1	2	AF315351
PcH1	2	EF577407

PcH1	2	DQ522235
PcH1	3	EU886981
PcH1	3	AY775051
PcH1	3	AY878717
PcH1	3	AJ812570
PcH1	3	AY740681
PcH1	3	FJ594765
PcH1	3	AY536742
PcH1	3	EU678897
PcH1	3	EF382672
PcH1	3	AY551331
PcH1	3	DQ647028
PcH1	3	AB104852
PcH1	3	AF294653
PcH1	3	AJ786649
PcH1	3	EF207719
PcH1	3	EF207717
PcH1	3	DQ522236
PcH1	3	AY030343
PcH1	3	AY866525
PcH1	3	AY460181
PcH1	4	EF415651
PcH1	4	AY686225
PcH1	4	EU588392
PcH1	4	AJ620678
PcH1	4	AB188812
PcH1	4	DQ089809
PcH1	4	DQ839391
PcH1	4	M95287

PcH1		4	AJ867811
PcH1		4	AJ969237
PcH1		4	AJ969236
PcH1		4	EF522838
PcH1		5	AY260546
PcH1		5	AY444814
PcH1		5	EU219534
PcH1		5	AJ223604
PcH1		5	AY507153
PcH1		6	CP001125
PcH1		6	FJ160769
PcH1		6	FJ196385
PcH1		6	EU834941
PcH1		6	AY816216
PcH1		6	DQ393784
PcH1		7	EF184216
PcH1		8	AF047479
PcH1	P2	2	AM991977
PcH1	P2	1	AY522431
PcH1	P2	1	EU266532
PcH1	P2m1	2	EF184217
<hr/>			
PcH1 _{TTN-10}		2	U36276
PcH1 _{TTN-10}		2	AF302086
<hr/>			
PcS		1	AY123253
PcS		1	X74412
PcS		1	EU219534
PcS		1	AJ870926
PcS		1	AJ628423
PcS		1	EF690696

PcS	1	AY781413
PcS	2	AY458224
PcS	2	EU543272
PcS	2	AY162283
PcS	2	FJ374756
PcS	2	L06822 (X68227)
PcS	2	AB469045
PcS	2	EF370423
PcS	2	DQ861641
PcS	2	AY463797
PcS	2	DQ517526
PcS	2	CP000604
PcS	2	AY232671
PcS	2	AF458082
PcS	2	EF592571
PcS	2	BR000038
PcS	2	X72585, U67194
PcS	2	DQ274503
PcS	2	DQ288156
PcS	2	U59183
PcS	2	EU091084
PcS	2	AF458080
PcS	2	CP000863
PcS	2	EF690695
PcS	2	AY648125
PcS	2	AJ628983
PcS	2	AJ515707
PcS	3	AF263519
PcS	3	EU014166

PcS	3	FJ207466
PcS	3	AY220520
PcS	3	EF375699
PcS	3	DQ153217
PcS	3	DQ153218
PcS	3	EF660562
PcS	3	EU118148
PcS	3	EU118149
PcS	3	AY816215
PcS	3	AM993098
PcS	3	DQ310703
PcS	3	AJ311891
PcS	3	AY294333
PcS	3	U37105
PcS	3	AM183120
PcS	3	EF093145
PcS	3	AM988778
PcS	3	EU014075
PcS	4	AF550679
PcS	4	U90945
PcS	4	AF034958
PcS	4	AY065966
PcS	4	AF263520
PcS	4	U12338/X15852
PcS	4	AJ310480
PcS	4	DQ112355
PcS	4	AJ870988
PcS	4	AF369871
PcS	4	AF418284

PcS		4	EF467306
PcS		5	EF207718
PcS		5	EU117233
PcS		5	CP001182
PcS		5	CT025832
PcS		5	AJ009820
PcS		5	DQ522239
PcS		5	AJ704863
PcS		6	AY029772
PcS		6	AY509004
PcS		6	EF206349
PcS		6	EF138817
PcS	P2	3	DQ143913
<hr/>			
PcS _{TGN-10}		3	AJ640197
PcH2		1	AB061794
PcH2		1	S45954
PcH2		2	CP000650
PcH2		2	AM180753
PcH2		3	AJ744860
PcH2		4	DQ522233
PcH2		4	AJ704863
PcH2		4	AJ278514/5
PcH2		5	U13880
<hr/>			
PcH2 _{TGN-10}		1	AM234698
PcH2 _{TGN-10}		1	DQ519091
PcH2 _{TGN-10}		2	AY038837
PcH2 _{TGN-10}		4	FJ594766
PcH2 _{TGN-10}		4	AF445082
<hr/>			
PcPUO		4	AF288045

PcPUO	1	S68049
PcSS	1	[2]
PcIn116	3	AJ621187
PcIn42	3	AJ243491

1. Lindstedt BA, Heir E, Nygard I, Kapperud G (2003) Characterization of class I integrons in clinical strains of *Salmonella enterica* subsp. *enterica* serovars Typhimurium and Enteritidis from Norwegian hospitals. *J Med Microbiol* 52: 141-149.
2. Brizio A, Conceicao T, Pimentel M, Da Silva G, Duarte A (2006) High-level expression of IMP-5 carbapenemase owing to point mutation in the -35 promoter region of class 1 integron among *Pseudomonas aeruginosa* clinical isolates. *Int J Antimicrob Agents* 27: 27-31.

Partie n° 2

**Etude épidémiologique de la prévalence des variants de Pc
au sein d'une collection de souches d'*Escherichia coli***

Dans le premier travail, nous avons montré que la plupart des IM 1 présente des variants faibles de Pc (PcW, PcH1). Notre analyse, bien qu'effectuée sur un grand nombre d'IM 1 (n= 321), était cependant basée sur des données disponibles *in silico* et ne tenait pas compte de la fréquence de chaque IM 1 dans la réalité épidémiologique. Aussi, pour vérifier si les fréquences issues de l'analyse *in silico* sont cohérentes avec des données épidémiologiques, nous avons effectué une étude, en collaboration avec l'équipe de Carmen Torres (Área de Bioquímica y Biología Molecular, Departamento de Agricultura y Alimentación, Universidad de La Rioja, Logroño), sur la répartition des promoteurs de cassettes des IM 1 au sein d'une collection de souches d'*E. coli* isolées en France et en Espagne. Les rares études traitant de l'épidémiologie des variants de Pc chez des souches cliniques indiquaient la prédominance du variant le moins efficace, PcW au sein des *Enterobacteriaceae* (Cocchi *et al.*, 2007 ; Lindstedt *et al.*, 2003 ; Schmitz *et al.*, 2001). Cependant, les auteurs de ces travaux ne tenaient pas compte de la présence ou non d'une région -10 étendue, un élément dont nous avons montré qu'il augmente largement la force de PcW. D'autre part, la prévalence des variants de P2 est totalement inconnue.

Matériel et Méthodes

I. Souches bactériennes et milieux de culture

86 souches d'*E.coli* chez lesquelles un fragment du gène *intI1* avait été détecté préalablement ont été incluses dans cette étude ; elles ont été isolées en France ou en Espagne d'écouvillonnages rectaux de volontaires sains ou d'échantillons cliniques de patients hospitalisés. Les cellules ont été cultivées en milieu BHI (Brain-Heart Infusion).

II. Caractérisation des réseaux de cassettes de gènes des IM 1

Les réseaux des cassettes de gènes des IM 1 des souches d'*E.coli* ont été caractérisés comme décrit précédemment, par des méthodes de PCR et de séquençage avec des oligonucléotides situés dans les régions 5'CS, 3'CS ou les gènes de cassettes (Saenz *et al.*, 2004).

III. Caractérisation de la région promotrice Pc/P2 des IM 1

L'ADN génomique des souches étudiées a été extrait à l'aide du kit QIAamp DNA mini kit (Qiagen, Courtaboeuf) à partir d'une culture de nuit dans 5ml de BHI à 37°C. Un fragment de 691 pb a été amplifié par PCR avec les oligonucléotides *intI1-LC9* (5' - CCA AGC TCT CGG GTA ACA TC - 3') et *ORF11R* (5'-GCG GGA TCC AGG TAA CTT TGT TTT AGG GCG ACT GCC CT -3') et 1 unité d'ADN polymérase fidèle PhusionTM High-Fidelity (Finnzymes, Finlande) selon les conditions d'amplification suivantes : 30 sec. de dénaturation à 98°C suivis de 30 cycles de 5 sec. à 98°C, 10 sec. à 64°C et 30 sec. à 72°C et une extension finale de 5 min. à 72°C.

Les produits de PCR ont été analysés sur gel d'agarose, purifiés à l'aide du kit Wizard® SV gel and PCR Clean Up System (Promega, Madison) et séquencés en utilisant l'oligonucléotide intI1-LC9 à l'aide d'un automate ABI Prism 3310 (Applied Biosystems).

IV. Détermination des groupes phylogénétiques

Les quatre groupes phylogénétiques majeurs (A, B1, B2 et D) ont été déterminés pour toutes les souches d'*E.coli* par la méthode basée sur une PCR triplex (Clermont, Bonacorsi et Bingen, 2000).

V. Détermination de la force du nouveau variant de P2 identifié

La force relative du nouveau variant du promoteur P2 identifié a été estimée dans une fusion transcriptionnelle entre un fragment contenant à la fois Pc et P2 et un gène *lacZ* dépourvu de promoteur cloné dans le vecteur rapporteur pSU38 Δ tot*lacZ* déjà décrit dans le premier travail (Jové *et al.*, 2010). Le plasmide pPcW-P2m3 a été construit à partir de pPcW-P2 par PCR assemblage avec pour oligonucléotides internes P2mut3L (5'-ctg ttt ttt tgg ggT ACA ATC TAt gcc tcg ggc atc caa g-3') et P2mut3R (5'-ccc gag gca TAG ATT GTA ccc caa aaa aac agt cat aac-3'). Le plasmide pPcW-P2m3 a été transformé dans la souche *E.coli* MC1061 pour estimation de la force du nouveau variant de P2 par essais β -galactosidase, réalisés comme précédemment (Jové *et al.*, 2010). Cinq essais différents ont été effectués.

Origine géographique des souches Origine clinique des souches	Espagne		France		Total
	Volontaire sain	Clinique	Volontaire sain	Clinique	
Réseau de cassettes de gènes					
<i>aadA</i>	0	4	2	4	10
<i>aadB</i> + <i>aadA1</i> + <i>cmlA1</i>	0	1	1	0	2
<i>blaOXA-1</i> + <i>aadA1</i>	0	1	1	1	3
<i>dfrA1</i>	1	2	1	0	4
<i>dfrA1</i> + <i>aadA1</i>	0	3	6	10	19
<i>dfrA5</i>	1	3	5	0	9
<i>dfrA5</i> + <i>ereA2</i>	0	2	0	0	2
<i>dfrA7</i>	1	1	0	0	2
<i>dfrA12</i> + <i>orfF</i> + <i>aadA2</i>	0	1	0	2	3
<i>dfrA12</i> + <i>orfF</i> + <i>aadA2</i> + <i>cmlA1</i> + <i>aadA1</i> ::IS440	1	0	0	0	1
<i>dfrA12</i> + <i>orfF</i> + <i>aadA2</i> + <i>cmlA1</i> + <i>aadA1</i> + <i>qacH</i> :: IS10	0	1	0	0	1
<i>dfrA12</i> + <i>orfF</i> + <i>aadA2</i> + <i>cmlA1</i> + <i>aadA1</i> + <i>qacH</i> :: IS440	0	2	0	0	2
<i>dfrA14</i>	0	0	0	1	1
<i>dfrA15</i> + <i>aadA1</i>	0	1	0	0	1
<i>dfrA16</i> + <i>aadA2</i>	1	1	0	1	3
<i>dfrA17</i>	0	3	0	1	4
<i>dfrA17</i> + <i>aadA5</i>	0	3	3	4	10
<i>dfrA25</i>	0	0	0	1	1
<i>dfrB1</i> + <i>aadA11</i>	0	0	0	1	1
<i>estX</i> + <i>psp</i> + <i>aadA2</i> + <i>cmlA</i> + <i>aadA1</i> + <i>qacH</i> ::IS440	0	2	0	0	2
<i>estX</i> + <i>psp</i> + <i>aadA2</i> + <i>cmlAΔ</i> ::IS1294 + <i>aadA1</i> + <i>qacH</i> ::IS440	1	0	0	0	1
Aucune cassette	0	1	0	1	2
Non déterminé	1	3	1	0	5
TOTAL	7	35	20	27	89

Tableau 9: Réseaux des cassettes de gènes des IM 1 en fonction de l'origine géographique et clinique de la souche.

Le nombre d'IM 1 présentant chacun des réseaux est indiqué

Configuration Pc-P2	Total	Espagne (n=43)				France (n=46)				Total selon l'origine (%)			
		Volontaires sains (n=7)		Cliniques (n=36)		Volontaires sains (n=20)		Cliniques (n=26)		Espagne	France	Volontaires sains	Clinique
		Nombre	%	Nombre	%	Nombre	%	Nombre	%				
PcW	40,4%	3	42,9	15	41,7	11	55	7	26,9	41,9	39,1	51,9	35,5
PcH1	33,7%	3	42,9	11	30,6	5	25	11	42,3	32,6	34,8	29,6	35,5
PcW + P2	16,9%	-	-	6	16,7	4 ^a	20	5	19,2	14	19,6	14,8	17,7
PcW_{TGN-10}	6,7%	1	14,3	3	8,3	-	-	2	7,7	9,3	4,3	3,7	8,1
PcS	2,3%	-	-	1	2,7	-	-	1	3,8	2,3	2,2	0	1,6

Tableau 10 : Prévalence des variants de Pc et de P2 parmi les 89 IM 1 détectés dans 86 souches d'E.coli isolées en France ou en Espagne

^a Dans un de ces quatre IM 1, la séquence de P2 est mutée et définit un nouveau variant nommé P2m3.

Résultats

1. Diversité des réseaux de cassettes de gènes des IM 1

89 IM 1 ont été identifiés, trois souches hébergant 2 IM 1. Les 21 réseaux de cassettes distincts qui ont pu être identifiés chez 82 IM 1 sont décrits dans le tableau 9. De plus, 2 des 7 IM 1 restants n'avaient aucune cassette insérée, et pour les cinq autres, nous n'avons pas réussi à identifier les cassettes. Les cassettes les plus fréquemment retrouvées sont les cassettes *dfr* de résistance au triméthoprimé et *aadA* de résistance à la streptomycine et spectinomycine, comme cela a été décrit dans divers études épidémiologiques sur les intégrons ((Iversen *et al.*, 2003 ; Soufi *et al.*, 2009)). La répartition est équivalente pour les souches françaises et espagnoles. Cependant, des IM 1 avec 6 réseaux de cassettes n'ont été retrouvés qu'en Espagne et tous présentent un gène tronqué (*qacH* ou *cmlA*) par une IS (Tableau 9). Enfin, le réseau de cassettes *dfrB1+aadA11* n'ayant jamais été décrit dans un intégron, sa séquence a été déposée dans la GenBank (Numéro d'accès GQ900749).

2. Prévalence des variants des promoteurs Pc et P2

Quatre variants de Pc ont été retrouvés dans les 89 IM 1 de l'étude (Tableau 10): PcW, PcH1, PcW_{TGN-10} et PcS, les variants les plus faibles étant les plus prévalents : PcW (57,3%) et PcH1 (33,7%) et les variants plus efficaces plus rares : PcW_{TGN-10} (6,7%) et PcS (2,3%). Cependant, presque un tiers des IM 1 ayant PcW présentaient aussi un P2 (PcW+P2, 16,9% des 89 IM1), une situation dans laquelle le niveau de transcription des gènes est plus élevée (Tableau 10). En fait, plus le variant de Pc est efficace pour la transcription, moins il est fréquent dans notre collection d'IM 1 d' *E.coli*.

Variants de Pc/P2	Groupes phylogénétiques				Total
	A	B1	B2	D	
PcW	16 (50%)	3 (27%)	9 (36%)	8 (38%)	36 (40%)
PcH1	6 (19%)	6 (55%)	8 (32%)	10 (48%)	30 (34%)
PcW+P2	6 (19%)	0 (0%)	6 (24%)	3 (14%)	15 (17%)
PcW_{TGN-10}	3 (9%)	1 (9%)	2 (8%)	0 (0%)	6 (7%)
PcS	1 (3%)	1 (9%)	0 (0%)	0 (0%)	2 (2%)
Total	32 (100%)	11 (100%)	25 (100%)	21 (100%)	89 (100%)

Tableau 11 : Prévalence des variants de Pc et de P2 parmi les 89 IM 1 en fonction de leur groupe phylogénétique.

Le nombre de séquences avec chaque variant est indiqué dans chaque cas, suivi du pourcentage de souches du même groupe.

Réseau de cassettes de gènes	Occurrence	Nombre de Pc/P2 distincts	PcW	PcH1	PcW+P2	PcW _{TGN-10}	PcS
<i>aadB + aadA1 + cmlA1</i>	2	1	0	2	0	0	0
<i>blaOXA-1 + aadA1</i>	3	1	0	0	3 ^a	0	0
<i>dfrA1</i>	4	1	4	0	0	0	0
<i>dfrA5</i>	9	1	9	0	0	0	0
<i>dfrA7</i>	2	1	0	0	0	0	0
<i>dfrA12 + orfF + aadA2</i>	3	1	0	0	0	3	0
<i>dfrA17</i>	4	1	0	4	0	0	0
<i>dfrA17 + aadA5</i>	10	1	0	10	0	0	0
<i>aadA</i>	10	2	0	1	9	0	0
<i>dfrA5 + ereA2</i>	2	2	1	1	0	0	0
<i>dfrA12 + orfF + aadA2 + cmlA1 + aadA1 + qacH ::IS440</i>	2	2	0	0	0	1	1
<i>dfrA16+aadA2</i>	3	2	1	2	0	0	0
<i>estX + psp + aadA2 + cmlA + aadA1 + qacH ::IS440</i>	2	2	1	1	0	0	0
<i>dfrA1+aadA1</i>	19	3	14	3	2	0	0
<i>Aucune cassette</i>	2	2	1	1	0	0	0
TOTAL	78						

Tableau 12 : Distribution des variants de Pc/P2 pour les réseaux de cassettes identifiés plusieurs fois.

^a Dans un de ces trois IM 1, la séquence de P2 est mutée et définit un nouveau variant nommé P2m3.

3. Fréquence des variants de Pc et P2 selon l'origine de la souche

Les résultats présentés dans le tableau 10 indiquent une représentation similaire des variants de Pc quelles que soient l'origine géographique de la souche ou le contexte clinique dont elle est issue. Dans les deux cas, les variants sont d'autant plus prévalents qu'ils sont faibles. La seule exception à cette règle concerne les souches françaises d'origine clinique où le variant PcH1 est plus représenté que PcW (Tableau 10). D'autre part, la distribution des variants de Pc/P2 était similaire quel que soit le groupe phylogénétique des souches, excepté pour les souches du groupe B pour lesquelles la proportion de variants PcH1 est plus inhabituellement élevée (Tableau 11).

4. Fréquence des variants de Pc/P2 selon le réseau de cassettes de gènes

Au sein des 89 IM 1, parmi les 21 réseaux de cassettes clairement distincts caractérisés, 14 réseaux ont été trouvés chez plusieurs IM 1, de 2 à 19 fois (Tableau 12). Seulement 8 réseaux étaient toujours associés à la même combinaison de Pc/P2. Chacun des 6 autres peuvent être associés à deux combinaisons Pc/P2 différentes, voir même trois dans un cas. De façon intéressante, le réseau identifié avec 3 formes Pc/P2 distinctes était aussi le plus fréquent. De même, les deux IM 1 dépourvus de cassettes présentaient des séquences de Pc distinctes (Tableau 12).

5. Caractérisation d'une nouvelle forme de P2

Un des IM 1 caractérisés, qui héberge les cassettes *blaOXA-1+aadA1* et dont le variant de Pc est PcW, présentait un promoteur P2 dont l'élément -10 (TACAGT) était mutée en TACAAT. Cette mutation est inédite et définit donc un nouveau variant de P2 nommé ici P2m3. Comme P2m3 est plus conservé que P2 par rapport à la séquence consensus des promoteurs σ^{70} , il est possible qu'il soit aussi plus efficace.

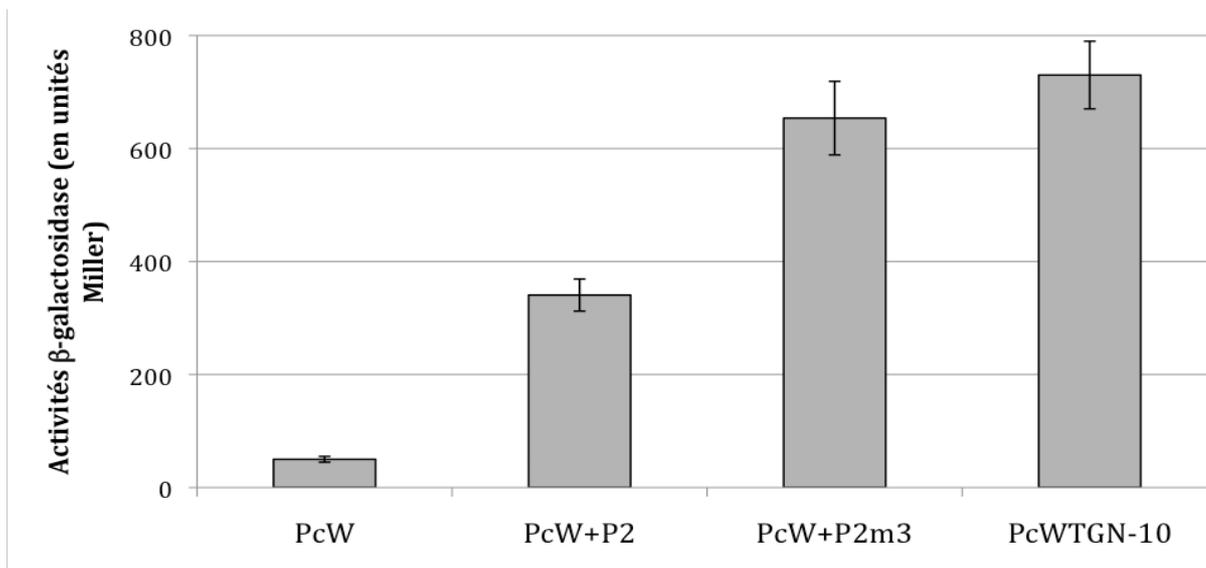


Figure 17: Efficacité du variant du promoteur P2 identifié dans ce travail.

L'activité de la région promotrice Pc-P2 a été mesurée par essais β -galactosidase à partir de fusions transcriptionnelles entre un fragment incluant Pc et P2 et le gène lacZ dépourvu de promoteur. Les barres d'erreurs indiquent les écarts type d'au moins cinq essais indépendants pour chaque plasmide.

Pour vérifier cette hypothèse, nous avons mesuré les activités β -galactosidases obtenues à partir de fusions transcriptionnelles entre la région promotrice PcW-P2 (ou PcW-P2m3) et un gène *lacZ* dépourvu de promoteur. Comme le montre la figure 17, P2m3 est effectivement deux fois plus efficace que P2 et permet un niveau de transcription des gènes de cassettes similaire à celui obtenu à partir du variant fort PcW_{TGN-10}.

Partie n° 3

**Interférences entre l'expression des cassettes et l'expression
de l'intégrase chez les intégrons de classe 1**

Les résultats présentés dans le premier travail décrivaient l'influence du polymorphisme de Pc sur l'activité de l'intégrase IntI1. Les travaux préliminaires de cette 3^{ème} partie s'attachent à déterminer si ce polymorphisme influence également le niveau d'expression de l'intégrase. En effet, comme Pc est situé dans la séquence codante du gène de l'intégrase, il fait face au promoteur PintI1 responsable de l'expression du gène *intI1*. Dans cette configuration, les molécules d'ARN synthétisées depuis les deux promoteurs se chevauchent en partie, une situation propice à des régulations connues sous le nom d'ARN antisens régulateurs en *cis*. De plus, les trafics simultanés d'ARN polymérases issues des deux promoteurs peuvent interférer au détriment de l'expression d'un des deux gènes. D'autre part, nous savons que l'expression de l'intégrase, sous le contrôle d'un répresseur LexA, est induite par la réponse SOS. Notre premier objectif était donc de déterminer si les promoteurs Pc et PintI1 faisaient l'objet de régulations mutuelles et ceci pour chaque variant de Pc. Nous avons également étudié l'influence du promoteur P2 sur l'activité du promoteur de l'intégrase PintI1 ; en effet, l'élément -10 de P2 chevauche à la fois PintI1 et le site de liaison de LexA. Les résultats présentés ici font partie de mon travail de thèse mais aussi de celui d'Emilie Guérin, doctorante dans le laboratoire.

Souches de <i>Escherichia coli</i>	Description/ Caractéristiques	Références
DH5 α	(F ⁻) <i>endA1 supE44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80 Δ lacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r_K⁻ m_K⁺), λ-</i>	Collection du laboratoire
MG1656	K12 sauvage délétée du gène <i>lacZ</i>	Espeli, Moulin et Bocard, 2001
MG1656 Δ LexA	MG1656 délétée du gène <i>lexA</i>	Guerin <i>et al.</i> , 2009

Tableau 13: Souches bactériennes utilisées lors de ce travail.

Référence	Plasmide	Description
Ploy <i>et al.</i> 1998	pAT674	Fragment BamHI de 6.5 kb de l'intégron In40 cloné dans pBGS18 ; héberge un IM 1 dont le variant de Pc est PcS.
Collis et Hall, 1995	pRMH821	Dérivé du plasmide R388 dont le variant de Pc a été changé en PcW
Jové <i>et al.</i> , 2010	pSU38 Δ totlacZ	Vecteur portant la séquence codante du gène <i>lacZ</i> sans promoteur ni signal d'initiation de traduction; [KmR]
Guerin <i>et al.</i> , 2009	pPintI1	Toute la région <i>attI1</i> d'In40 (138 pb) amplifiée avec les oligonucléotides Pint1 et Pint2 clonée dans le plasmide pSU38 Δ totlacZ
Ce travail	pPintI1 mut	pPintI1 muté avec les oligonucléotides PintI1mutL et PintI1mutR de façon à inactiver le promoteur PintI1 (élément -10 TAGACT muté en cgGACg).
Ce travail	pPintI1 ^L	pPintI1 muté avec les oligonucléotides LexAmut2L et LexAmut2R de façon à inactiver le site de liaison de LexA (CTGTACAAAAAACAG en agtTACAAAAAACAG)
Ce travail	pPintI1P2	pPintI1 muté avec les oligonucléotides P2mutL et P2mutR de façon à créer le promoteur P2 (insertion de 3G)
Ce travail	pPintI1P2m	pPintI1P2 muté avec les oligonucléotides P2-12GL et P2-12GR de façon à inactiver le promoteur P2
Ce travail	pPintI1PcW	Toute la région <i>attI1</i> + 150 bases d' <i>intI1</i> (amplifié avec les oligonucléotides int4bBamHI et DORF11EcoRI) de pRMH821 clonée dans pSU38 Δ totlacZ.
Ce travail	pPintI1mutPcW	pPintI1PcW dont le promoteur PintI1 est inactivé (oligonucléotides Pint1mutL et Pint1mutR)
Ce travail	pPintI1PcW ^L	Dérivé de pPintI1PcW dans lequel l'opérateur LexA est muté (oligonucléotides LexAmut2L et LexAMut2R)
Ce travail	pPintI1mutPcW ^L	Dérivé de pPintI1PcW ^L dans lequel PintI1 est inactivé (oligonucléotides Pint1mutL et Pint1mutR)
Jové <i>et al.</i> , 2010	pPcS	Toute la région <i>attI1</i> (138 pb) d'In40 (amplifiée avec les oligonucléotides int4b et DORF11) clonée dans le plasmide pSU38 Δ totlacZ
Ce travail	pPcS ^L	Dérivé de pPcS dans lequel l'opérateur LexA est muté (oligonucléotides LexAmut2L et LexAMut2R)
Jové <i>et al.</i> , 2010	pPcW	Toute la région <i>attI1</i> (138 pb) de pRMH821 (amplifiée avec les oligonucléotides int4b et DORF11) clonée dans le plasmide pSU38 Δ totlacZ
Ce travail	pPcW ^L	Dérivé de pPcW dans lequel l'opérateur LexA est muté (oligonucléotides LexAmut2L et LexAMut2R)
Ce travail	pPcW ^{L*}	Dérivé de pPcW ^L dans lequel PintI1 est inactivé (PintI1mutL et PintI1mutR)

Tableau 14 : Plasmides utilisés lors de ce travail

Matériels et Méthodes

I. Souches bactériennes et plasmides

Les souches bactériennes utilisées dans cette étude sont décrites dans le tableau 13. Les souches d'*E.coli* ont été cultivées en milieu BHI (Brain-Heart Infusion) à 37°C et les antibiotiques utilisés aux concentrations suivantes : Ampicilline (Amp), 100 µg/ml ; Kanamycine (Kan), 25 µg/ml. Le X-gal a été utilisé à une concentration finale de 40 µg/ml.

Les plasmides utilisés dans cette étude sont décrits dans le tableau 14. Les extractions d'ADN plasmidique ont été réalisées à l'aide du kit Wizard® Plus Minipreps DNA Purification System (Promega, Madison).

II. Plasmides et techniques de biologie moléculaire

A. Techniques de base

Les digestions enzymatiques ont été effectuées avec les enzymes *EcoRI* et *BamHI* (de New England Biolabs, Hitchin ou Invitrogen, Cergy-Pontoise) conformément aux indications du fournisseur. La T4 DNA ligase est fournie par Promega, les polymérases *Taq* par New England Biolabs et la Phusion™ par Finnzymes. Les oligonucléotides utilisés dans ce travail, listés dans le tableau 15, ont été synthétisés par Sigma-Aldrich.

B. Réactions d'amplification par PCR

La synthèse des fragments de PCR destinés à être clonés a été réalisée à l'aide de la *Taq* polymérase (NEB) où la Phusion™ (Finnzymes, Finlande) lorsqu'une haute fidélité était requise. Les réactions ont été effectuées selon les recommandations du fournisseur dans les conditions d'amplification suivantes :

Oligonucléotides	Séquence (5'→3')	Localisation
GSP1	ATTCTCCGTGGAACAAACG	pSU38ΔtotlacZ
GSP2	TTACGTTGGTGTAGATGGGC	pSU38ΔtotlacZ
GSP3	TGCATCTGCCAGTTTGAGG	pSU38ΔtotlacZ
Int4b	CCGGAATTCACACCGTGGAACGGATGAAG	<i>int11</i>
Int4bBamHI	CGCGGATCCACACCGTGGAACGGATGAAGG	<i>int11</i>
DORF11	CGCGGATCCATCGTTGCTGCTCCATAACA	<i>att11</i>
DORF11EcoRI	CCGGAATTCATCGTTGCTGCTCCATAACA	<i>att11</i>
LexAmut2L	<u>TGTACAAAAAAACTTCATAACAAGCCATG</u>	<i>att11</i>
LexAmut2R	GGCTTGTTATGAAGTTTTTTTT <u>TGTACAGTCTATG</u>	<i>att11</i>
Pint1	CCGGAATTCCTTTGTTTTAGGGCGACTG	<i>att11</i>
Pint2	GCGGGATCCATGGCTTGTTATGACTGTT	<i>att11</i>
Pint11MutL	GTTTTTTTGTACAGTCCGTGCCTCGGGCATCCAAGC	<i>att11</i> (Pint11)
Pint11MutR	GGATGCCCGAGGCACGGACTGTACAAAAAACAGTC	<i>att11</i> (Pint11)
P2mutL	ATGACTGTTTTTTTGGGG <u>TACAGTCTATGCCTCGGGCATCCAAG</u>	<i>att11</i> (P2)
P2mutR	ATGCCCGAGGCATAG <u>ACTGTACCCCAAAAAACAGTCATAACAA</u>	<i>att11</i> (P2)
P2-12GL	CTGTTTTTTTGGGG <u>GACAGTCTATGCCTCGGGCAT</u>	<i>att11</i> (P2)
P2-12GR	CCCGAGGCATAG <u>ACTGTCCCAAAAAACAGTCA</u>	<i>att11</i> (P2)

Tableau 15: Oligonucléotides utilisés dans ce travail.

Les bases des éléments -10 des promoteurs sont soulignées, celles mutées sont en gras.

- Avec la *Taq* polymérase : 2 min. à 95°C, suivi de 30 cycles de 30s à 95°C, 30s. à une température dépendant du T_m, le plus souvent 55°C ou 52°C et 30s à 2 min. à 72°C selon la taille du fragment attendue ; puis 10 min. à 72°C
- A la Phusion™ : 2 min. à 95°C, suivi de 30 cycles de 15s à 95°C, 15s. à 68°C 15s à 72°C; puis 10 min. à 72°C

Les mutations des séquences d'intérêts (promoteurs, opérateur LexA) ont été réalisées par PCR assemblage, à l'aide d'oligonucléotides internes qui se chevauchent et portent la mutation désirée, et d'oligonucléotides externes permettant d'amplifier le fragment contenant la mutation.

C. Constructions des fusions transcriptionnelles

Les plasmides pPintI1, pPcW, pPcS et leurs dérivés ont été construits en clonant aux sites EcoRI-BamHI de pSU38Δ*totlacZ* (Jové *et al.*, 2010) un produit de PCR de 144 pb (pPintI1 et dérivés) ou 258 pb (pPcW, pPcS et leurs dérivés) amplifié à partir du plasmide pAT674 ou pRMH821, qui hébergent un IM 1 dont le variant de Pc est PcS ou PcW, respectivement (Collis et Hall, 1995 ; Ploy, Courvalin et Lambert, 1998). Tous les plasmides obtenus sont listés dans le tableau 14. Tous les fragments clonés ont été vérifiés par séquençage.

D. Expériences de 5'RACE

- *Extraction des ARN totaux*

L'ARN total a été extrait à l'aide du kit Fast RNA Problue (MP Biomedicals, Illkirch, France) avec l'appareil FastPrep (MP Biomedicals). Le culot d'ARN a été repris dans de l'eau traitée au DEPC. Les ADN contaminants ont été éliminés via deux traitements successifs à la DNase à l'aide du kit DNA-free (Ambion, Courtaboeuf) et leur absence vérifiée par PCR.

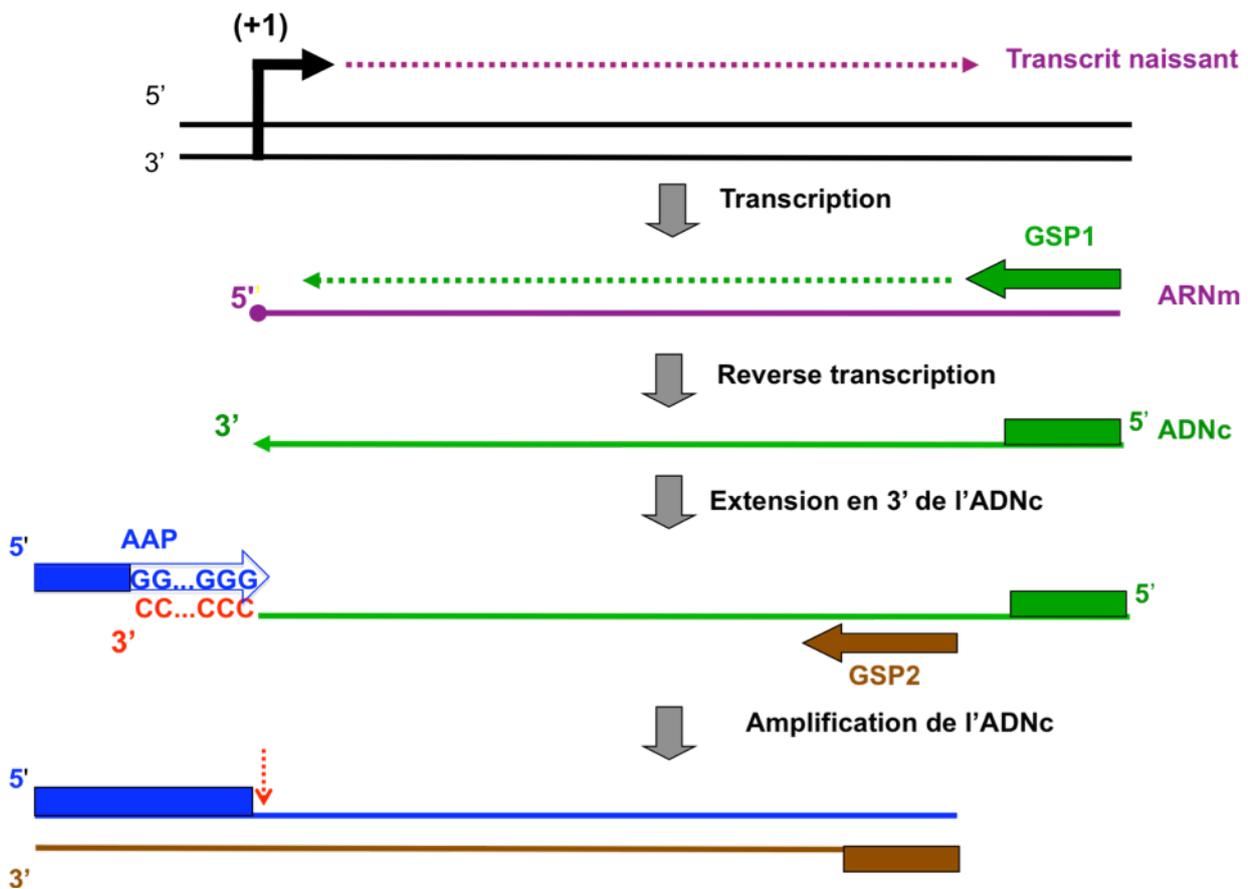


Figure 18: Principe de la 5' Rapid Amplification of cDNA Ends (5'RACE).

La région ADN qui suit le promoteur est indiquée sous sa forme double brin par des traits noirs; l'ARN polymérase cellulaire synthétise des ARN (trait violet) à partir du (+1) de transcription (flèche cassée). L'ARN d'intérêt est converti en ADN complémentaire (ADNc, trait vert) à l'aide d'une amorce « GSP1 » spécifique. Pendant cette étape, un traitement enzymatique détruit les ARN restants. Les ADNc sont étendus en 3' par une série de C (en rouge) qui servent ensuite d'ancrage pour l'amorce AAP fournie dans le kit: la molécule d'ADNc peut donc être amplifiée avec l'amorce AAP (en bleue) et l'amorce GSP2 (en marron). La molécule matrice étant simple brin, GSP2 ne se fixe qu'après synthèse des premières molécules ADN issues de l'amorce AAP. La position exacte du +1 de transcription (fléchée en rouge) peut être déterminée par séquençage du produit de PCR issu de cette amplification.

L'intégrité des ARN extraits a été contrôlée sur gel 0,8%, leur quantité et leur pureté évaluées par dosage spectrophotométrique à 260 puis 280 nm.

- *Principe général de la 5'Rapid Amplification of cDNA Ends (5'RACE)*

Cette technique permet d'identifier le site (+1) d'initiation de transcription d'un promoteur via le séquençage de l'extrémité 5' de l'ARNm correspondant. L'ADNc synthétisé à partir de ce transcrit se termine donc en 3' par le (+1) de transcription (Figure 18). L'ajout en 3' de cet ADNc d'une série de bases C permet de l'amplifier et de le séquencer, le (+1) de transcription étant la base qui précède immédiatement la série de 20 C dans la molécule.

- *Protocole de 5'Rapid Amplification of cDNA Ends (5'RACE)*

Nous avons utilisé le kit 5'RACE System version 2.0 d'Invitrogen (Cergy-Pontoise, France). La synthèse de l'ADNc spécifique de PintI1 a été réalisée selon les conditions du fournisseur, avec l'amorce GSP1 et à partir de l'ARNm de la souche MG1656 Δ lexA transformée avec pPintI1. L'absence d'ADN plasmidique et la présence d'ADNc ont été vérifiées par PCR. L'ADNc a été purifié et étendu en 3' dans les conditions du fournisseur, puis amplifié avec les oligonucléotides AAP (qui s'hybride sur le polyC) et GSP2. Le produit a été identifié sur gel d'agarose à 1,6% puis cloné dans un vecteur pGem®T-Easy selon les conditions du fournisseur (Promega, Madison).

- *Séquencage des clones*

Les clones recombinants ont été séquencés par la méthode dite de Sanger à partir de 6 μ l d'ADN plasmidique, 4 μ l de l'oligonucléotide GSP3 à 1 μ M, 2 μ l de tampon 5x et 3 μ l de mélange BigDye® (Applied Biosystems, Courtaboeuf, France) qui contient les désoxyribonucléotides triphosphate (dNTP), la polymérase et les terminateurs (ddNTP). Nous avons utilisé un automate ABI Prism 3310 (Applied Biosystems).

D. Essais β -galactosidase

Chaque fusion transcriptionnelle a été transformée dans la souche *E. coli* MG1656 pour les mesures d'activités de l'enzyme β -galactosidase. Les essais ont été effectués selon la technique décrite par Miller (Miller, 1992) selon le protocole précédemment décrit dans le premier travail (Jové *et al.*, 2010). Les expériences ont été réalisées au moins 5 fois pour chaque souche.

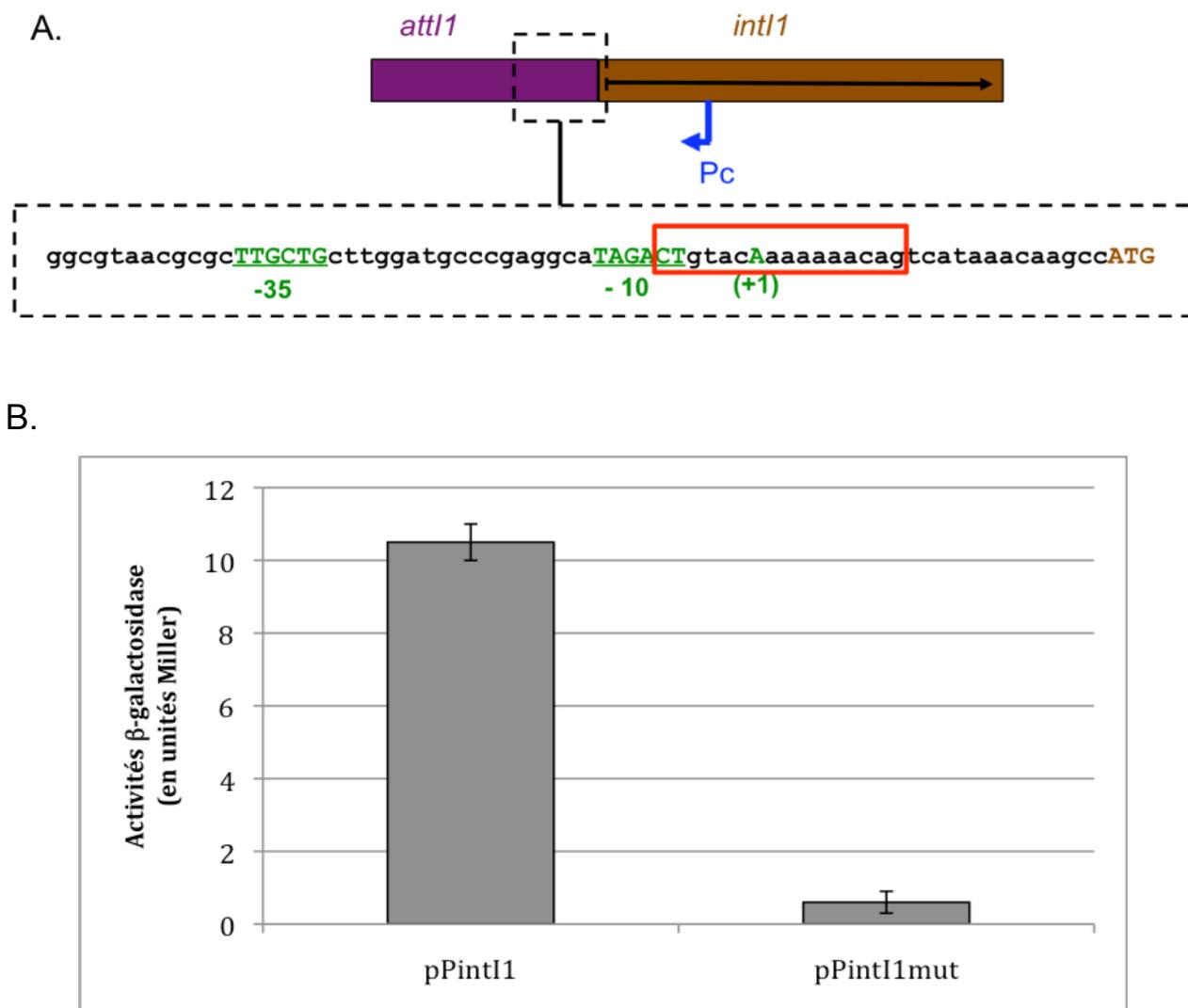


Figure 19 : Caractérisation du promoteur *PintII* du gène de l'intégrase des *IM 1*

A. Séquence du promoteur *PintII*. La structure de la 5'CS des *IM 1* est rappelée et la séquence de la partie d'*attI1* contenant *PintII* est détaillée. Le (+1) de transcription de *PintII* identifié par 5'RACE est écrit en majuscules vertes, ainsi que les éléments -35 et -10 qui en ont été déduits. Le codon *START* du gène *intI1* est écrit en majuscules marrons. La séquence du site de liaison de la protéine *LexA* est encadrée en rouge.

B. Activité du promoteur *PintII*. L'activité du promoteur *PintII* a été mesurée par des essais β-galactosidase à partir de fusions transcriptionnelles entre l'intégralité du site *attI1* et un gène *lacZ* dépourvu de promoteur. L'élément -10 potentiel est muté dans la construction *pPintI1mut*. Les barres d'erreurs indiquent les écarts type d'au moins cinq essais indépendants pour chaque plasmide.

Résultats

1. Caractérisation du promoteur PintI1

Au début de ce travail, aucun promoteur d'intégrase d'intégron n'avait été caractérisé. Pour localiser le promoteur PintI1 des IM 1, nous avons cloné toute la région *attI1* dans le sens du gène *intI1* en fusion transcriptionnelle avec le gène *lacZ* ; le plasmide obtenu pPintI1 a permis, à l'aide d'essais β -galactosidase, de mesurer une activité transcriptionnelle faible mais significativement différente du bruit de fond détecté en absence d'insert. Le (+1) de transcription du promoteur responsable de cette activité a été déterminé par des expériences de 5'RACE à partir du plasmide pPintI1. Les premiers essais ayant échoué, le plasmide pPintI1 a été transformé dans une souche MG1656 Δ *lexA*, dont le gène *lexA* est délété, de façon à augmenter l'expression du gène *intI1*, la répression par LexA étant levée. Nous avons alors obtenu un seul produit de 5' RACE de 350 pb qui a été cloné dans le vecteur pGem®T-Easy. Les séquences de plusieurs clones recombinants ont montré que des transcrits naissaient au niveau d'une base adénine (A) distante de 22 pb du codon START d'*intI1* (Figure 19A). Cette position de (+1) est compatible avec un hexamère situé 5 pb en amont et communément considéré comme l'élément -10 de PintI1 : TAGACT, les bases conservées par rapport au consensus des promoteurs de σ^{70} étant soulignées (Martinez et de la Cruz, 1990 ; Naas *et al.*, 2001).

Pour confirmer le rôle de cette séquence dans la transcription d'*intI1*, ces bases ont été mutées. Les mutations choisies inactivent aussi le site de fixation du répresseur LexA ; ainsi, l'activité de PintI1, si elle a lieu, n'est pas réprimée. L'absence d'activité β -galactosidase avec ce plasmide recombinant pPintI1mut confirme le rôle d'élément -10 pour cette séquence (Figure 19B).

Promoteur	Élément -35	Espaceur (pb)	Élément -10	Référence ou numéro d'accès
Consensus σ^{70}	TTGACA	17	TATAAT	Hawley et McClure, 1983
Variant de PintI1				
PintI1	TTGCTG	17	TAGACT	Guerin <i>et al.</i> , 2009
PintI1m1 (muté 1)	TTGCTG	17	TA <u>I</u> GAC	FM897214, AM988778 , Walsh <i>et al.</i> , 2003
PintI1m2 (muté 2)	TTGCTG	17	TA <u>A</u> CTG	AJ628983
PintI1m3 (muté 3)	TT <u>A</u> GCT	19	TAGACT	(Zolg et Hanggi, 1981)

Tableau 16: Polymorphisme du promoteur PintI1.

La séquence consensus des promoteurs σ^{70} est rappelée dans la première ligne.

Les bases qui diffèrent du PintI1 non muté sont soulignées.

En outre, cette séquence est située 17 pb en aval d'une séquence d'un élément -35 potentiel (TTGCTG).

Ces résultats suggèrent fortement que la transcription du gène *intI1* est sous le contrôle d'un unique promoteur PintI1 que nous avons caractérisé et de séquence :TTGCTG-N₁₇-TAGACT.

2. Polymorphisme du promoteur PintI1

L'analyse de l'ensemble des séquences de IM 1 disponibles dans les banques de données nucléotidiques menée dans le cadre du premier travail nous a permis de repérer de nouvelles versions de PintI1 : 3 IM 1 distincts présentent en effet une insertion d'une base T au sein de l'élément -10 (tableau 16) (Numéro d'accès GenBank : FM897214, AM988778, (Walsh *et al.*, 2003)).

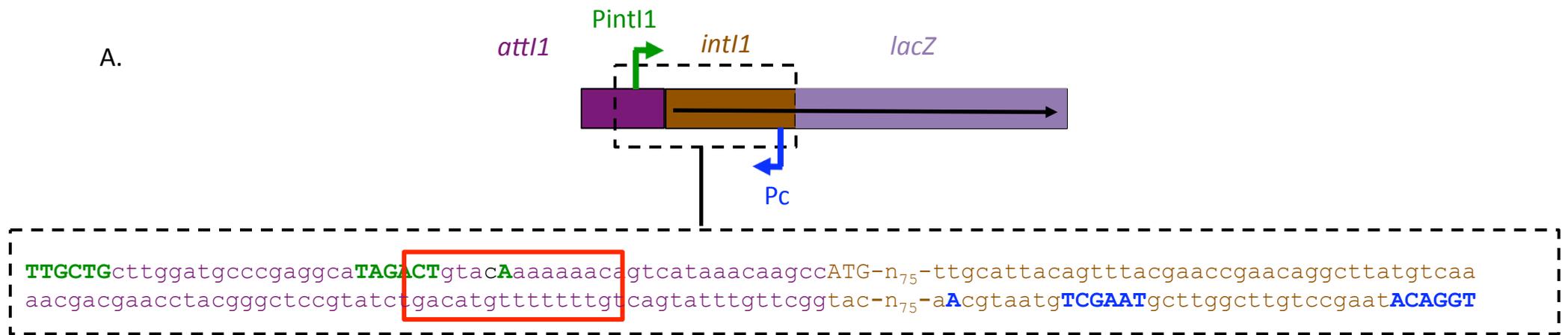
Cette mutation conserve le site de liaison du répresseur LexA de PintI1 et ne devrait donc *a priori* pas modifier l'expression d'*intI1* dans les conditions de croissance habituelles. Elle convertit néanmoins l'élément -10 de PintI1 en un élément potentiellement moins efficace pour initier la transcription notamment en raison de la perte en dernière position de l'élément -10 d'un T très conservé au sein des promoteurs à σ^{70} . Deux autres séquences avec des PintI1 mutés existent dans les bases de données, mais qui n'apparaissent qu'une seule fois chacune, laissant penser qu'il pourrait s'agir d'erreur de séquences (tableau 16) (Numéro d'accès GenBank : V00252 et AJ628983) ; c'est pourquoi nous ne les discutons pas ici.

3. Influence de l'activité du promoteur P2 sur l'expression du gène *intI1*

Dans le premier travail, nous avons montré que dans près de 10% des IM 1, il existe dans la région *attI1*, un promoteur P2 qui participe à la transcription des gènes de cassettes. Dans ce cas, l'élément -10 de P2 chevauche totalement celui du promoteur *PintI1* (Figure 20A). Ces IM 1 n'ont *a priori* plus de site de liaison fonctionnel pour le répresseur LexA, en raison d'un espacement trop élevé (13 pb au lieu de 10) entre les triplets terminaux CTG et CAG.

En principe, une altération du site de liaison de LexA induit l'activité de *PintI1*, comme il a été montré récemment (Guerin *et al.*, 2009). Pourtant, lorsque P2 est présent, nous n'avons mesuré aucune activité β -galactosidase significative (Figure 20B) avec le plasmide *pPintI1P2*. L'inactivation de P2 dans *pPintI1P2m* restaure un niveau d'expression comparable à celui mesuré avec le plasmide *pPintI1^L* dans lequel le site LexA est inactivé (Figure 20B).

Ces résultats suggèrent que, même s'il empêche la protéine LexA de réprimer le promoteur *PintI1*, le promoteur P2 actif perturbe l'expression du gène de l'intégrase à partir du promoteur *PintI1*.



B.

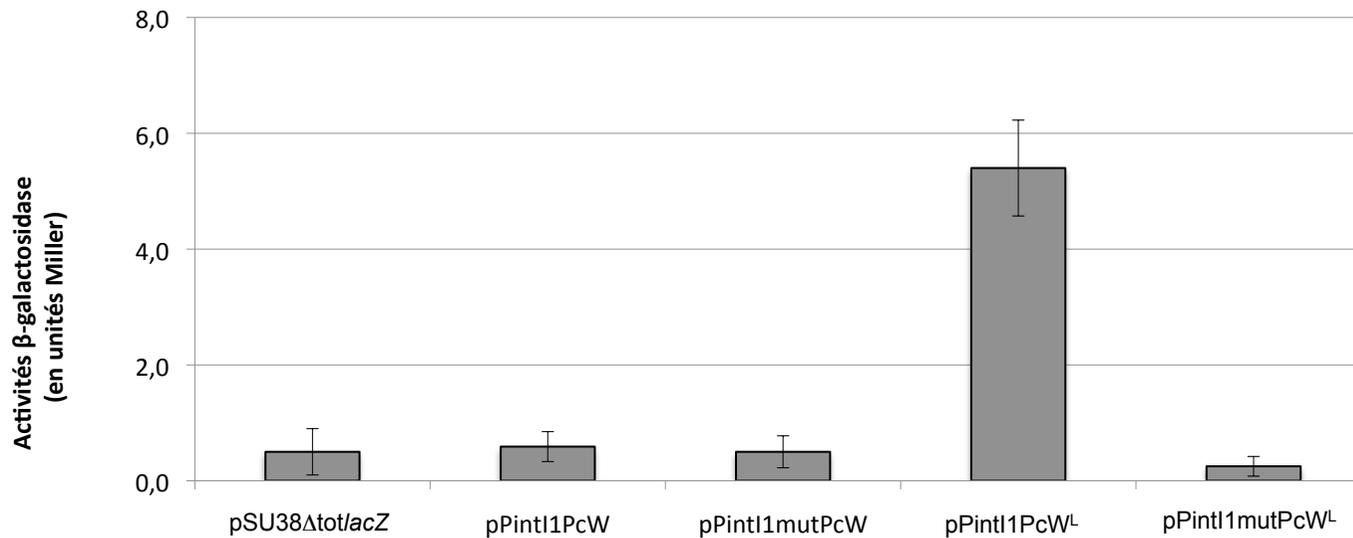


Figure 21: Influence du promoteur Pc sur l'activité du promoteur Pint11

A. Représentation schématique de la fusion traductionnelle utilisée pour étudier l'influence de Pc sur Pint11. Une partie de la séquence du fragment cloné est détaillée (dans le cas de pPint11PcW). Les éléments -35, -10 et les sites d'initiation de transcription des promoteurs sont écrits en gras majuscules (en vert pour Pint11, en bleu pour Pc). Le codon d'initiation d'int11 est écrit en majuscule et le site de liaison du répresseur LexA encadré en rouge. N: A, C, G ou T.

B. Activité du promoteur Pint11 selon la séquence de Pc ou du site de liaison du répresseur LexA. L'activité du promoteur Pint11 a été mesurée par des essais β -galactosidase à partir de la fusion traductionnelle représentée en A ou des dérivées.. Les barres d'erreurs indiquent les écarts type d'au moins cinq essais indépendants pour chaque plasmide.

4. Interférence entre l'activité des promoteurs Pc et PintI1

L'arrangement en vis-à-vis des promoteurs Pc et PintI1 nous a conduit à déterminer si leurs activités interfèrent.

✓ Influence de l'activité de Pc sur l'expression de l'intégrase

Pour étudier l'éventuelle influence de la présence du promoteur Pc sur l'activité du promoteur PintI1, le niveau d'expression d'*intI1* a été évalué par des essais β -galactosidase à partir d'une fusion traductionnelle qui inclut les 50 premiers codons du gène *intI1*, de façon à inclure le promoteur Pc (Figure 21A). Nous avons choisi de travailler d'abord avec le variant PcW car, d'après nos résultats précédents, c'est le plus courant au sein des IM 1. Avec le plasmide obtenu, pPintI1PcW, l'activité β -galactosidase sous le contrôle de PintI1 était très faible, à peine distinct du bruit de fond du vecteur sans insert ou d'un plasmide dérivé dans lequel PintI1 est muté (Figure 21B, pPintI1mutPcW), et inférieure à l'activité de base de PintI1 quand il est réprimé par LexA (pPintI1, Figure 20). Lorsque l'activité du promoteur PintI1 est constitutive via la mutation de l'opérateur de LexA (Figure 21A), il y a une augmentation de l'activité du promoteur PintI1 même en présence de PcW (Figure 21B, pPintI1PcW^L) qui est perdue si PintI1 est inactif (Figure 21B, pPintI1mutPcW^L). Cependant, l'activité β -galactosidase contrôlée par PintI1 est moins forte qu'en l'absence de PcW (Figure 20B, pPintI1^L).

Ces résultats suggèrent que le promoteur PcW n'empêche pas l'induction de l'expression d'*intI1* par la réponse SOS mais diminue le niveau d'expression à partir de PintI1.

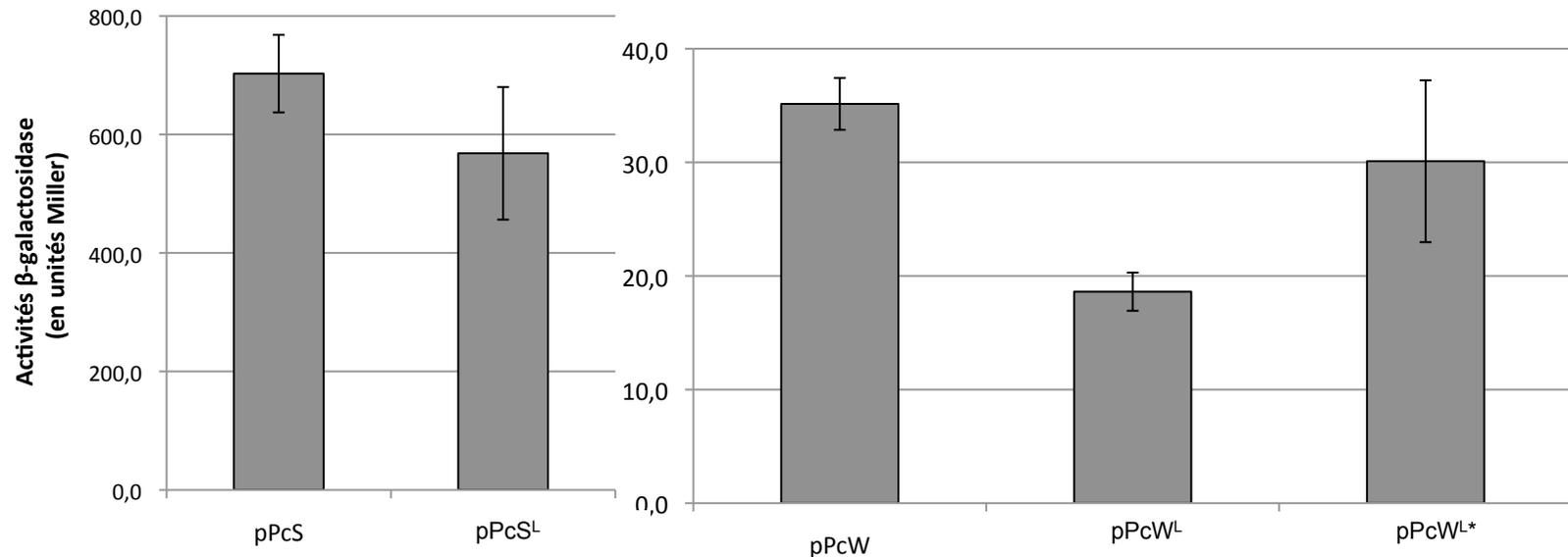
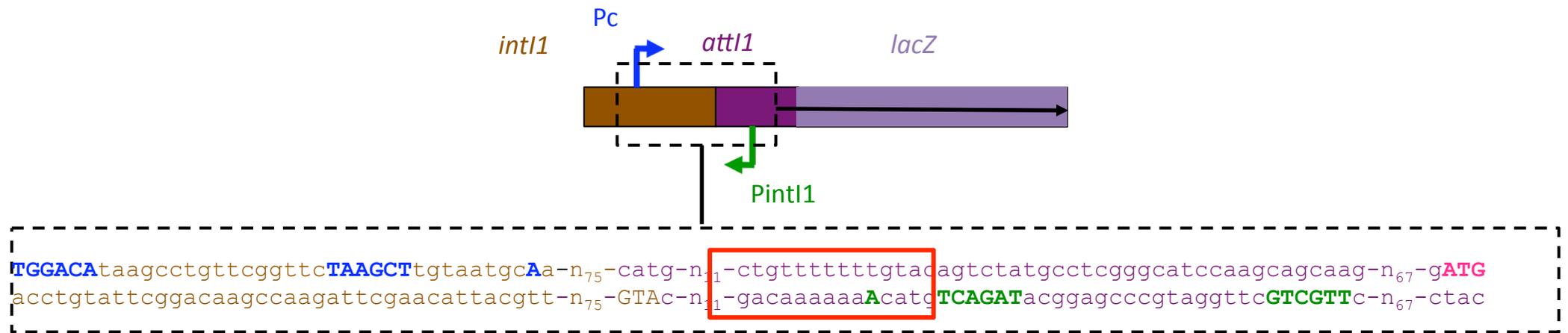


Figure 22: Influence du promoteur *Pint11* sur l'activité du promoteur *PcW*

A. Représentation schématique de la fusion transcriptionnelle utilisée pour étudier l'influence de *Pc* sur *Pint11*. Une partie de la séquence du fragment cloné est détaillée (dans le cas de *pPcW*). Les éléments -35, -10 et les sites d'initiation de transcription des promoteurs sont écrits en gras majuscules (en vert pour *Pint11*, en bleu pour *Pc*). Le codon d'initiation d'*int11* est écrit en majuscule, celui utilisé pour initier la traduction de *lacZ* en rose, et le site de liaison du répresseur *LexA* est encadré en rouge. N: A, C, G ou T.

B. Activité du promoteur *Pint11* selon la séquence de *Pc* ou du site de liaison du répresseur *LexA*. L'activité du promoteur *Pint11* a été mesurée par des essais β -galactosidase à partir de la fusion traductionnelle représentée en A ou des dérivées. Les barres d'erreurs indiquent les écarts type d'au moins cinq essais indépendants pour chaque plasmide.

✓ Influence de l'activité de PintI1 sur l'expression des cassettes

L'incidence de l'activité du promoteur PintI1 sur la force du promoteur Pc a été déterminée avec des fusions transcriptionnelles *Pc-lacZ* similaires à celles utilisées dans le premier travail (Figure 22A). Nous avons étudié l'influence de deux variants de Pc, l'un fort PcS et l'autre faible PcW. Lorsque PintI1 est réprimé par LexA, PcS est 20 fois plus fort que PcW (pPcS Vs pPcW) (Figure 22B). La mutation de l'opérateur de LexA, qui induit le promoteur PintI1, n'a aucun effet mesurable dans nos conditions, sur l'activité de PcS (Figure 22B, pPcS^L) mais réduit de moitié celle de PcW (Figure 22B, pPcW^L). Cette inhibition est supprimée lorsque le promoteur PintI1 est inactivé (Figure 22B, pPcW^{L*}) indiquant qu'elle n'était due qu'à ce promoteur.

Ces résultats montrent que l'interférence entre Pc et PintI1 est fonction du variant (et donc de la force) de Pc ; ainsi, l'induction de l'activité de PintI1 réduit celle de Pc seulement s'il s'agit du variant faible PcW, le variant fort PcS y étant insensible. L'activité de PintI1 est quant à elle aussi diminuée par la présence de PcW. Les constructions testant l'influence de PcS sur PintI1 sont en cours ; il est probable que cette diminution soit plus prononcée qu'avec le variant faible PcW.

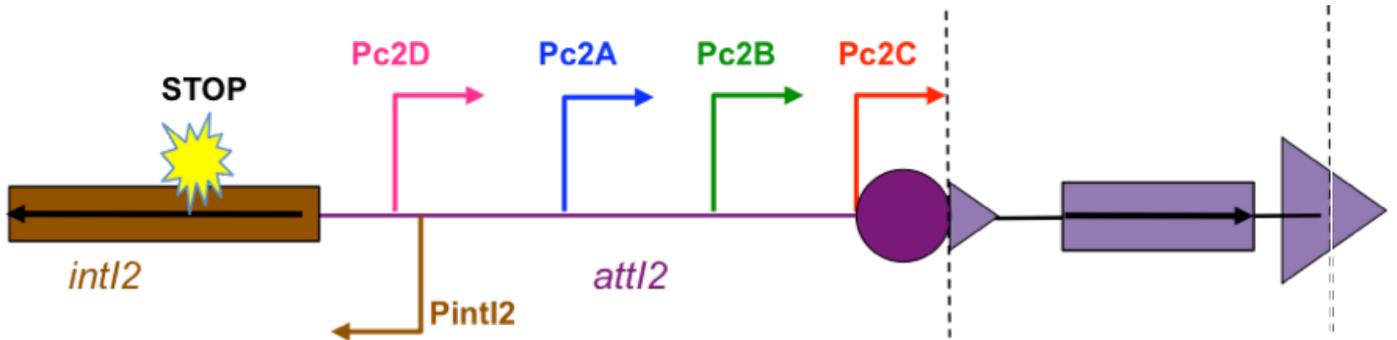
D'autre part, nous avons montré que le promoteur P2 quand il est présent interfère aussi avec PintI1 inhibant complètement son activité alors que l'expression de l'intégrase devrait être constitutive, le site de fixation de LexA étant modifié.

L'ensemble de ces résultats montrent que, chez les IM 1, l'expression du gène de l'intégrase et des gènes de cassettes sont inversement liées et dépendent du variant de Pc.

Partie n° 4

**Expression des gènes de cassettes et de l'intégrase des
intégrons de classe 2**

A.



B.

```

cataaaacgctccttgtcTTGCAActgtctgctatacagccTATTCTagctgggatttaaaaaagtgcctgctgtttttacg
GTAttttgcgAGGAAcagaacgttgacagacggatatgtcgggaTAAGATcgaccctaattttttcACGGACggacaaaaatgc
                                     -35          Pc2D          -10
cctagagatgcttgtttaccggtagagTTTTAAtttaatgctaataaatTAAAATggtatgagttctttgggtgagataatgtg
g gatctctacgaacaaatggccatctcaaattaatctacgatttatttaattttacaatactcaagaaaccactctattacac
                                     -35          Pc2A          -10
catcgtgcaagcaggatagacggcatgcacgatttghtaataacagagtgtcTTGTATttttaagaaagtctaTTTAATacaagt
gtagcacgttcgtcctatctgccgtacgtgctaacattattgtctcacagaacataaaaatttctttcagataaattatgttca
                                     -35          Pc2B          -10
gattatattaattacggtaagcatcagcggGTGACAaaacgagcatgcttactaaTAAAATg
c taatataattaatgccattcgtagtcgccactgttttgctcgtacgaatgattatttac
                                     -35          Pc2C          -10

```

Figure 23: Structure des intégrons de classe 2

A. Représentation schématique d'un intégron de classe 2 (IM 2). Le gène *intI2* (rectangle marron) est le plus souvent interrompu par un codon STOP précoce (étoile). La région *attI2* (en violet) est suivie d'un nombre variable de cassettes de gènes, dont une est représentée ici en mauve (le gène sous forme de rectangle, le site *attC* d'un triangle). Les promoteurs *Pc2* et *PintI2* putatifs de la littérature sont représentés par des flèches cassées. Le sens de transcription des gènes est indiqué par une flèche.

B. Détail de la séquence de la région *attI2*. Les séquences des éléments -35 et -10 des promoteurs *Pc2* et *PintI* représentés en A sont écrites en majuscule et en gras, celle de l'élément -10 de *Pc2B* est doublement soulignée. La séquence RBS putative et le codon START d'*intI2* sont écrits en majuscule et en gras, le site de fixation potentiel de l'opérateur *LexA* est encadré en rouge. Séquence tirée du fichier GenBank numéro d'accès AJ001816 (Tn7)

Nos études sur les interactions entre expression des cassettes et de l'intégrase chez les IM 1 nous ont conduit à considérer la situation des IM de classe 2, la seconde classe la plus prévalente de IM au sein des isolats cliniques. Les IM 2 ont la particularité d'avoir un gène *intI2* interrompu par un codon STOP rendant l'intégrase non fonctionnelle (Hansson et al., 2002). Probablement en cette raison, le réseau des cassettes est très stable chez les IM 2, même si des IM 2 avec des réseaux de cassettes plus variés ont été décrits ces dernières années. Le promoteur Pc des IM 2 n'a pas été caractérisé. Une étude suggère toutefois qu'il est situé dans la région *attI* et non dans le gène *intI* comme c'est le cas chez les IM 1 et 3. En effet, la présence de la région *attI2* suffit pour que les gènes de cassettes de l'IM 2 de Tn7 soient exprimés (Hansson et al., 2002). Quatre promoteurs Pc putatifs situés dans *attI2* ont été proposés dans la littérature (Hansson et al., 2002 ; Simonsen, Chen et Levinson, 1983) ; ils sont nommés ici Pc2A à Pc2D, leurs positions et leurs séquences sont précisées dans la figure 23. Le promoteur PintI2 du gène *intI2* n'a pas été caractérisé non plus mais une séquence a été proposée pour ce promoteur (Fichier GenBank L10818). Dans ce travail, nous avons identifié précisément le promoteur du gène de l'intégrase des IM 2 et mis en évidence la coexistence dans *attI2* de deux promoteurs constitutifs qui contribuent à l'expression des cassettes de gènes des IM 2 ; de plus, nous avons montré que le site de fixation potentiel de LexA des IM 2 n'est pas fonctionnel.

Références	Souches bactériennes	Description/ Caractéristiques
Collection du laboratoire	<i>E.coli</i> DH5 α	(F ⁻) <i>endA1 supE44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80 ΔlacZAM15 Δ(lacZYA-argF)U169, hsdR17(r_K⁻ m_K⁺), λ-</i>
Collection du laboratoire	<i>E.coli</i> MC1061	<i>hsdR2 hsdM hsdS araD139 Δ(ara-leu)7697Δ(lac)X74 galE15 galk16 rpsL (Str) mcrA mcrB1</i>
Gassama Sow <i>et al.</i> , 2008	<i>S. sonnei</i> Dak 0898-14	Souche isolée au Sénégal dont le chromosome héberge un intégron de classe 2
Références	Plasmides	Description
Jové <i>et al.</i> , 2010	pSU38 Δ totlacZ	Vecteur portant la séquence codante du gène <i>lacZ</i> sans promoteur ni signal d'initiation de traduction; [KmR]
Ce travail	pPintI2	Produit de PCR Pint2L x Pint2R PCR à partir de l'ADN total de l'isolat Dak 0898-14 cloné dans pSU38 Δ totlacZ. Contient toute la région <i>attI2</i> (310 pb)
Ce travail	pPintI2-2	pPintI2 muté avec les oligonucléotides PintI2mutL et PintI2mutR pour inactiver PintI2 (élément -10 TAGAAT muté en cgGAcg).
Ce travail	pPintI2-3	Produit de PCR Pint2Lbis x Pint2R amplifié à partir de l'ADN total de l'isolat Dak 0898-14 cloné dans pSU38 Δ totlacZ. Contient les 97 pb en amont du gène <i>intI2</i>
Ce travail	pPintI2-4	pPintI2 muté avec les oligonucléotides C2-LexAmut2L et C2-LexAMut2R pour inactiver l'opérateur putatif de LexA
Ce travail	pPc2	Produit de PCR Pc2L x Pc2R amplifié à partir de l'ADN total de l'isolat Dak 0898-14 cloné dans pSU38 Δ totlacZ. Contient toute la région <i>attI2</i> et le début du gène <i>dfrA1</i> .
Ce travail	pPc2-2	Produit de PCR Pc2IntI2L x Pc2R amplifié à partir de l'ADN total de l'isolat Dak 0898-14 cloné dans pSU38 Δ totlacZ. Contient toute la région <i>attI2</i> , une partie du gène <i>intI1</i> et le début du gène <i>dfrA1</i> .
Ce travail	pPc2-3	Produit de PCR Pc2IntI2L x Pc2DR amplifié à partir de l'ADN total de l'isolat Dak 0898-14 cloné dans pSU38 Δ totlacZ. Contient un fragment de <i>attI2</i> limité au Pc2D et une partie du gène <i>intI2</i>
Ce travail	pPc2A	Produit de PCR Pc2AL x Pc2AR amplifié à partir de l'ADN total de l'isolat Dak 0898-14 cloné dans pSU38 Δ totlacZ. Promoteur Pc2A cloné isolément
Ce travail	pPc2B	Produit de PCR Pc2BL x Pc2BR amplifié à partir de l'ADN total de l'isolat Dak 0898-14 cloné dans pSU38 Δ totlacZ. Promoteur Pc2B cloné isolément
Ce travail	pPc2Amut	pPc2A muté avec les oligonucléotides Pc2AmutL et Pc2AmutR de façon à inactiver Pc2A
Ce travail	pPc2Bmut	pPc2B muté avec les oligonucléotides Pc2BmutL et Pc2BmutR de façon à inactiver Pc2B
Ce travail	pPc2B-Pc2B'mut	pPc2B muté avec les oligonucléotides Pc2Bmut2L et Pc2Bmut2R de façon à inactiver Pc2B'
Ce travail	pPc2-Pc2Amut	pPc2 muté avec les oligonucléotides Pc2AmutL et Pc2AmutR de façon à inactiver Pc2A
Ce travail	pPc2-Pc2B'mut	pPc2 muté avec les oligonucléotides Pc2Bmut2L et Pc2Bmut2R de façon à inactiver Pc2B'
Ce travail	pPc2-Pc2AmutPc2B'mut	pPc2-Pc2Amut muté avec les oligonucléotides Pc2BmutL et Pc2BmutR de façon à inactiver Pc2A et Pc2B' concomitamment
Ce travail	pPc2-Pc2Av2Pc2B'v2	pPc2 muté avec les oligonucléotides Pc2Amut2L/Pc2Amut2R et Pc2B'mut2L/Pc2B'mut2R de façon à créer le deuxième variant de Pc2A et de Pc2B'
Ce travail	pPc2Av2	pPc2A muté avec les oligonucléotides Pc2Amut2L et Pc2Amut2R de façon à créer le deuxième variant de Pc2A
Ce travail	pPc2B'v2	pPc2B muté avec les oligonucléotides Pc2Bmut3L et Pc2Bmut3R de façon à créer le deuxième variant de Pc2B
Ce travail	pPc2C	Produit de PCR Pc2CL x Pc2CR amplifié à partir de l'ADN total de l'isolat Dak 0898-14 cloné dans pSU38 Δ totlacZ. Promoteur Pc2C cloné isolément
Ce travail	pPc2D	Produit de PCR Pc2DL x Pc2DR amplifié à partir de l'ADN total de l'isolat Dak 0898-14 cloné dans pSU38 Δ totlacZ. Promoteur Pc2D cloné isolément

Tableau 17: Souches bactériennes et plasmides utilisés dans ce travail

Matériels et Méthodes

I. Souches bactériennes et milieux de culture

Les souches bactériennes utilisées dans cette étude sont décrites dans le tableau 17. Les cellules ont été cultivées en milieu BHI (Brain-Heart Infusion) à 37°C et, complémenté lorsque cela est nécessaire en kanamycine à 25µg/ml final. Le X-gal a été utilisé à une concentration finale de 40 µg/ml.

II. Plasmides et techniques de biologie moléculaire

A. Techniques de base

Les plasmides utilisés dans cette étude sont décrits dans le tableau 17. Les extractions d'ADN plasmidiques ont été réalisées à l'aide du kit Wizard® Plus Minipreps DNA Purification System (Promega, Madison). Les oligonucléotides utilisés dans ce travail, listés dans le tableau 18, ont été synthétisés par Sigma-Aldrich.

B. Réactions d'amplification par PCR

La synthèse des fragments de PCR destinés à être clonés a été réalisée comme décrit dans la partie n°3. Les mutations des séquences d'intérêts (promoteurs, opérateur LexA) ont été réalisées par PCR assemblage.

C. Expériences de 5'RACE

Les extractions d'ARN et expériences de 5'RACE ont été réalisées comme décrit dans la partie n°3. Les amorces GSP1, GSP2 et GSP3 utilisées sont décrites dans le tableau 18.

Nom	Séquence (5'→3')
C2-LexAmut2L	AGGCTGTATAGGCAGAACTTTGCAAGACAAGGAG
C2-LexAmut2R	CTTGTCTTGCAAAGTTCTGCCTATACAGCCTATTC
GSP1	ATTCTCCGTGGAACAAACG
GSP2	TTACGTTGGTGTAGATGGGC
GSP3	TGCATCTGCCAGTTTGAGG
Pint2L	CCGGAATTCATTTTATTAGTAAGCATGCTC
Pint2Lbis	CCGGAATTCGTAAACAAGCATCTCTAGGCG
Pint2R	GCGGGATCCATAAAACGCTCCTTGTCTTGCAAC
Pintf2mutL	TTAAATCCCAGC CGGACG AGGCTGTATAGGC
Pintf2mutR	CTATACAGCCT CGTCCG GCTGGGATTTAAAAAAGTG
Pc2L	CCGAATTCTCGCATATCCGTGCGTATAGA
Pc2R	CGCGGATCCACTCCATTCTTCGATATAGCTAC
Pc2Intf2L	CCGGAATCCAACGGCTGTTGTAAAAACC
Pc2AL	CCGGAATTCGTGCCTGTTTTTTACGCCTA
Pc2AR	GCGGGATCCCATGGCTGTTTCTATCTCACCCAAAGAACTC
Pc2BL	CCGGAATCCAGGATAGACGGCATGCACG
Pc2BR	GCGGGATCCATGGCTGTTTCTATGCTTACCG TAAATTAATA TAATC
Pc2AmutL	GCTAAATAAAT CGAGCG GTTATGAGTTCTTTGGGTG
Pc2AmutR	AAGAACTCATAAC CGCTCG ATTTATTTAGCAT
Pc2BmutL	AAGAAAGTCTA GTTCCG CACAAGTGATTATATTAATTAAC
Pc2BmutR	ATAATCACTTGT GCGAACT AGACTTTCTTTAAAAATAC
Pc2Bmut2L	AAGAAAGTCTA TTCGAT GCAAGTGATTATATTAATTAAC
Pc2Bmut2R	ATAATCACTTGC ATCGAAT AGACTTTCTTTAAAAATAC
Pc2DR	GCGGGATCCATGGCTGTTTCTTAAACAAGCATCTCTAGGCG
Pc2CL	CGCGAATTCACAAGTGATTATATTAATTAACGG
Pc2CmutL	ACGAGCATGCTTACTA ACGAAAT GTTAACCTCTGAGG
Pc2CmutR	GAGGTTAAC ATTTCG TTAGTAAGCATGCTCGTTTTGTC
Pc2Bmut3L	GTAATAACAGAGTGTCT TATAT TTTTAAAGAAAG
Pc2Bmut3R	TTAAAA ATATA AGACTCTGTTATTACAAATCGTG
Pc2Amut2L	GCTAAATAAAT TAAAGT GTTATGAGTTCTTTGGGTG
Pc2Amut2R	AAGAACTCATAAC ACTTTA ATTTATTTAGCAT

Tableau 18: Oligonucléotides utilisés dans ce travail.

Les éléments -35 et -10 des promoteurs étudiés sont soulignés, les bases mutées sont écrites en gras

D. Constructions des fusions transcriptionnelles

Les plasmides pPc2, pPintI2 et leurs dérivés ont été construits en clonant aux sites EcoRI-BamHI de pSU38 Δ totlacZ (Jové *et al.*, 2010) un produit de PCR de 423 pb (pPc2 et dérivés), 114 pb (pPc2A et dérivés), 119 pb (pPc2B et dérivés) ou 318 pb (pPintI2 et dérivés) amplifié à partir de l'ADN génomique d'une souche de *Shigella sonnei* hébergeant un IM 2 précédemment décrit au laboratoire (Gassama Sow *et al.*, 2008). Tous les plasmides obtenus sont listés dans le tableau 17. Dans le plasmide pPc2, une fusion traductionnelle *dfrA1-lacZ*, l'addition d'un site BamHI dans la séquence codante de *dfrA1* nous a conduit à ajouter une mutation silencieuse (GTT en GTG, les deux codons codant une valine) dans le 14^{ème} codon de *dfrA1*. Tous les fragments clonés ont été vérifiés par séquençage.

III. Essais β -galactosidases

Chaque fusion transcriptionnelle a été transformée dans la souche *E. coli* MC1061 pour les mesures d'activités de l'enzyme β -galactosidase. Les essais ont été effectués comme décrit dans la partie n°3 et au moins 5 fois pour chaque souche.

IV. Expériences de retard sur gel

La surexpression et la purification de la protéine LexA ont été réalisées comme décrit précédemment (Da Re *et al.*, 2009). Les sondes ADN ont été obtenues par PCR avec les oligonucléotides Pint1 et Pint2 en utilisant le plasmide pAT674 comme matrice pour la sonde PintI1, et avec les oligonucléotides Pint2L et Pint2R et l'ADN génomique de *Shigella sonnei* Dak 0898/12-14 comme matrice pour la sonde PintI2). Les fragments de PCR ont été ensuite radiomarqués en 5' avec du [γ ³²P]ATP (Amersham, Saclay, France) en utilisant la T4 DNA polynucleotide kinase (Promega, Charbonnières, France) selon la technique classique (Sambrook, Maniatis et Fritsh, 1989).

Les expériences de retard ont été réalisées comme décrit dans (Abella *et al.*, 2004) en utilisant différentes quantités de protéine LexA purifiée (0, 20, 200 ou 2000 ng) et 40 ng de sonde marquée dans le mélange réactionnel. Un excès de sonde non marquée (15 fois) a été utilisé pour les expériences de compétition. Les échantillons ont été séparés dans un gel de Tris-glycine-EDTA polyacrylamide natif 5%, le gel séché est exposé sur la nuit dans un écran phosphoimager (Perkin-Elmer, Courtaboeuf, France). Les images ont été digitalisées avec un scanner Cyclone.

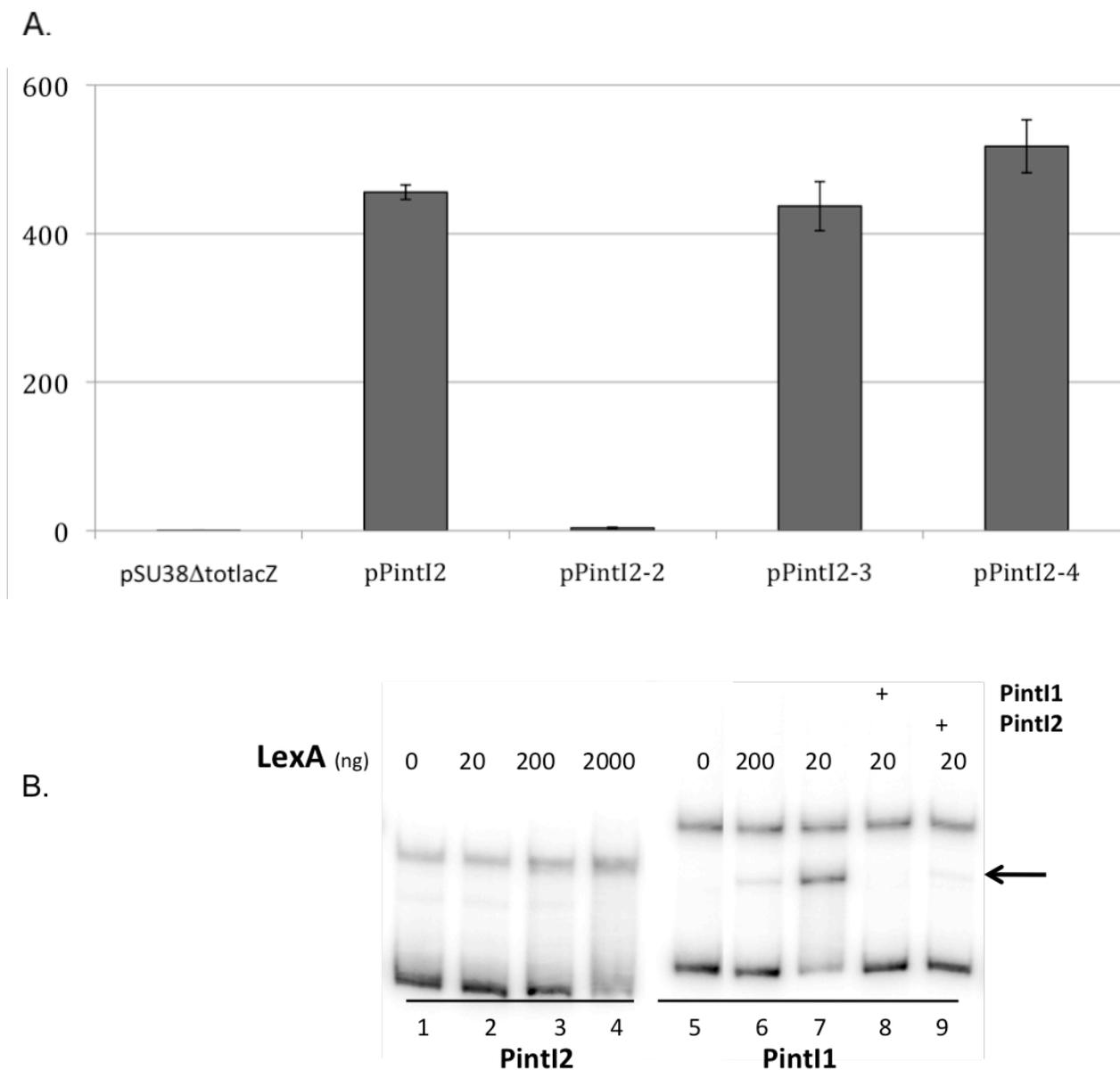


Figure 24: Expression et régulation du gène *intI2* des IM 2

A. Caractérisation et régulation du promoteur *PintI2*, d'après les mesures d'activités β -galactosidase effectuées avec les plasmides *pPintI2* et ses dérivées : *pPintI1-2* (*PintI1* muté en -10), *pPintI2-3* (*PintI2* cloné isolément), *pPintI2-4* (Opérateur *LexA* potentiel muté).

Les barres d'erreurs indiquent les écarts types d'au moins cinq essais pour chaque plasmide.

B. Expériences de retard sur gel avec le site *att2* natif, en présence ou en absence de la protéine *LexA* purifiée (colonne 1 à 4, « *PintI2* », les quantités de protéines sont indiquées en ng). Des contrôles positifs de la liaison de cette protéine à l'ADN dans nos conditions ont été réalisés (colonnes 5 à 7) ainsi que des essais de compétition avec les sondes froides *PintI1* (colonne 8) ou *PintI2* (colonne 9). La position du complexe retardé est indiquée par une flèche.

Résultats

1. Expression du gène de l'intégrase des IM 2

a- Caractérisation du promoteur PintI2

Nous avons testé la capacité de la région *attI2* à contrôler l'expression d'un gène *lacZ* dépourvu de promoteur dans une fusion transcriptionnelle PintI2. Nous mesurons alors un signal β -galactosidase relativement fort comparé au bruit de fond détecté avec un vecteur sans insert, indiquant qu'il existe en effet dans *attI2* un promoteur fonctionnel pour *intI2* (Figure 24A). Grâce à des expériences de 5'RACE, nous avons identifié précisément le (+1) de transcription comme étant un A situé 33 pb en amont du codon START d'*intI2* et à une distance compatible avec les éléments -35 CAGGCA et -10 TAGAAT potentiels déjà proposés dans la littérature (Figure 23B). La fonctionnalité de ce promoteur a été confirmée après que la mutation, dans pPintI2-2, des bases les plus conservées de l'élément -10 par rapport au consensus des promoteurs à σ^{70} , supprime toute activité β -galactosidase mesurable (Figure 24A). D'autre part, le clonage d'un fragment de 97 pb d'*attI2* contenant ce promoteur, pPintI2-3, suffit à conférer le niveau d'activité mesuré en présence de toute la région *attI2*. L'ensemble de ces résultats montrent que la transcription du gène *intI2* est sous le contrôle d'un unique promoteur PintI2 : CAGGCA-N₁₇-TAGAAT.

b- Le système SOS n'induit pas l'expression du gène *intI2*

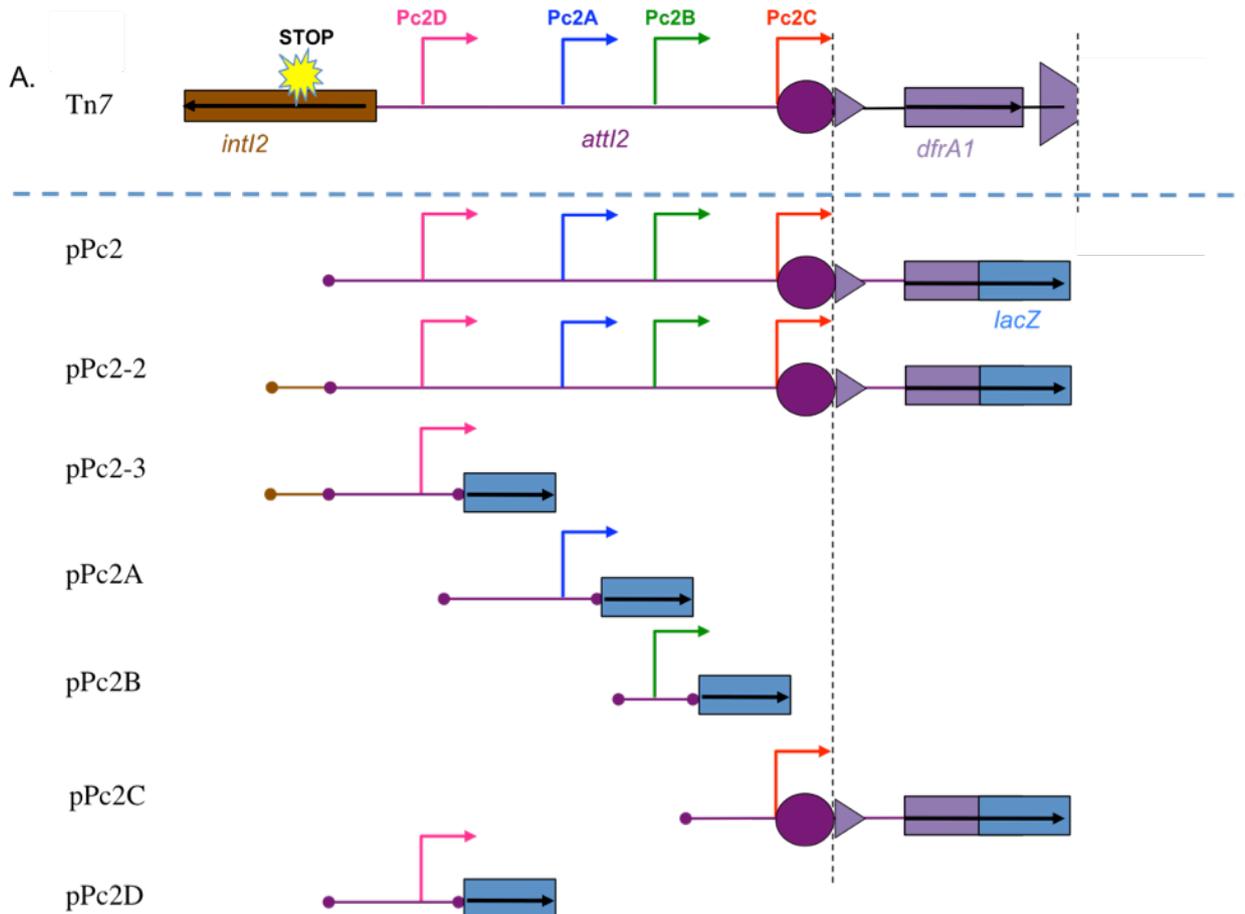
La transcription des gènes *intI* des IM 1 et du SI de *V. cholerae* est réprimée par la protéine LexA et induite par la réponse SOS (Guerin *et al.*, 2009). Cette régulation est liée à la présence dans les régions *attI* de la plupart des classes d'intégrons, d'opérateurs spécifiques de LexA, de consensus CTGT-N₈-ACAG, avec N₈ une séquence riche en A/T (Butala, Zgur-Bertok et Busby, 2009). L'opérateur LexA putatif des IM 2 (CTGTATAGGCAGACAG) chevauche le promoteur PintI2 que nous venons d'identifier (Figure 23B).

Pour vérifier que l'expression d'*intI2* est également induite par la réponse SOS, nous avons muté les bases clés de l'opérateur de LexA pour construire pPintI2-4. De façon surprenante, ces mutations n'ont pas entraîné d'augmentation dans l'activité détectée comparée à la situation non mutée, suggérant que l'expression d'*intI2* n'est pas réprimée par la protéine LexA *in vivo*. Pour le vérifier, nous avons procédé à des essais de retard sur gel avec la protéine LexA purifiée et un produit de PCR contenant la région *attI2*. L'absence de retard spécifique indique la protéine LexA n'est pas capable de se lier dans la région *attI* des IM 2 (Figure 24B). Ces résultats montrent donc que malgré la présence d'une séquence conservée d'opérateur de LexA à proximité du promoteur du gène *intI2*, la réponse SOS n'induit pas la transcription du gène de l'intégrase des IM 2.

2. Caractérisation du promoteur Pc des gènes de cassettes

a- L'expression des cassettes ne dépend que de *attI2*

Le promoteur Pc des IM 2 était supposé être dans la région *attI2* (Hansson *et al.*, 2002). Pour confirmer cette hypothèse, nous avons évalué l'activité du gène *lacZ* lorsqu'il est en fusion traductionnelle avec un fragment de 417 pb qui inclut toute la région *attI2* et les 14 premiers codons de la première cassette *dfrA1*. La forte activité détectée indique la présence d'un promoteur fonctionnel dans la région *attI2* (Figure 25, pPc2). Cependant, elle n'exclut pas l'existence d'un second promoteur actif dans la séquence codante d'*intI2*, une situation déjà décrite chez certains IM 1 (Jové *et al.*, 2010). D'ailleurs, il existe dans *intI2* une séquence bien conservée par rapport au promoteur type σ^{70} (séquence : TGGCTA-N₁₇-TAAGCT, avec en plus une région -10 étendue) et donc susceptible de participer à l'expression des cassettes de gènes des IM 2. En fait, un dérivé de pPc2, pPc2-2, qui en plus d'*attI2* inclut une partie d'*intI2* (220 pb de plus) génère la même activité β -galactosidase et, en revanche, aucune activité n'a été détectée dans pPc2-3, qui inclut cette région d'*intI2* et seulement une partie



B.

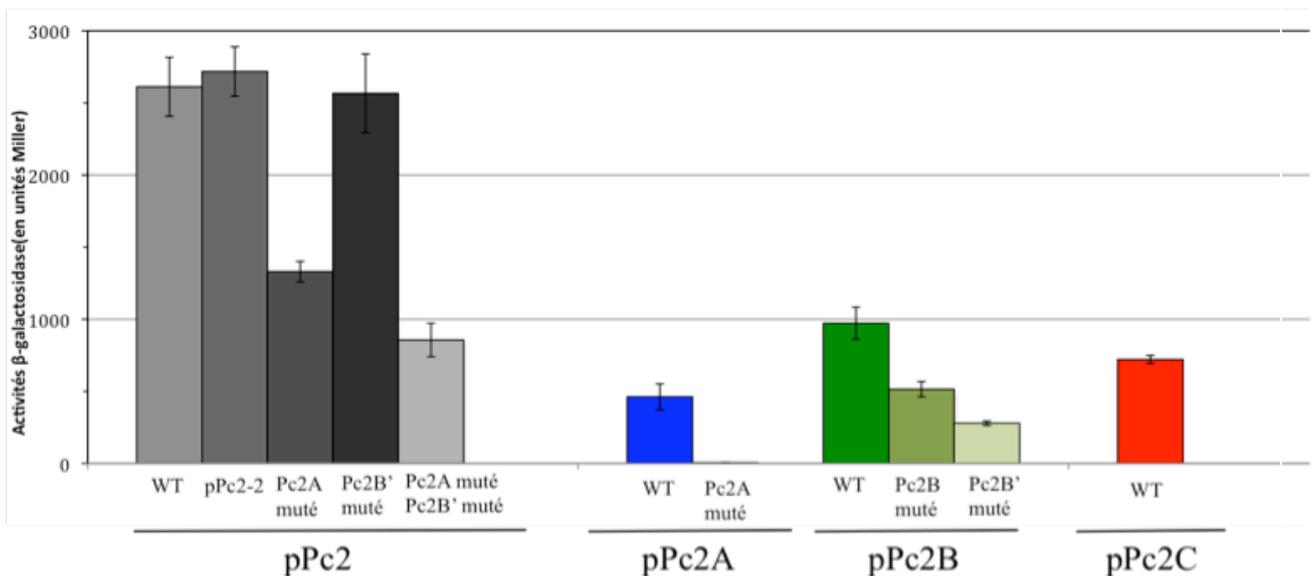


Figure 25: Caractérisation des promoteurs des gènes des cassettes d'IM 2.

*A. Représentation schématique des fusions transcriptionnelles utilisées. Tn7 correspond à la structure d'un IM 2 classique. B. Force des promoteurs selon les constructions et/ou les mutations introduites, estimée par la mesure de l'activité β -galactosidase. Les barres d'erreurs indiquent les écarts type d'au moins cinq essais indépendants pour chaque plasmide. L'activité issue de **pPc2D** (1 unité Miller) n'est pas représentée. WT : Wild-type (Sauvage)*

limitée d'*attI2* (Figure 25 A et B). Cette région du gène *intI2* ne fonctionne donc pas comme promoteur pour les gènes de cassettes.

b- Caractérisation des promoteurs des gènes de cassettes

Pour localiser précisément le promoteur Pc des IM 2, des expériences de 5'RACE ont été réalisées à partir du plasmide pPc2. Deux produits ont été obtenus, suggérant la coexistence de deux types de transcrits compatibles avec deux des quatre promoteurs proposés dans la littérature : Pc2A (TTTTAA-N₁₇-TAAAAT) et Pc2B (TTGTAT-N₁₆-TTTAAT) (Fichier GenBank AM261760).

Ces deux promoteurs ont été étudiés en les clonant séparément dans les fusions transcriptionnelles pPc2A et pPc2B (Figure 25A). Ces deux plasmides génèrent des activités β -galactosidase élevées, indiquant la présence d'un promoteur fonctionnel dans chaque fragment (Figure 25B). De plus, lorsque les éléments -10 de ces promoteurs ont été mutés aux positions clés des promoteurs à σ^{70} , l'activité de pPc2A est annulée (Figure 25) tandis que celle de pPc2B est seulement réduite de moitié (Figure 25). De façon à inactiver totalement Pc2B, l'élément -35 a également été muté mais nous n'avons obtenu aucune modification de l'activité (résultats non montrés). Ces résultats suggèrent que l'élément -10 de Pc2B pourrait en fait ne pas être exactement « TTTAAT » mais commencer par le TA interne de cet hexamère (souligné); la séquence de ce promoteur appelé dorénavant Pc2B' serait alors TTGTAT-N₁₈-TAATAC. D'ailleurs, la mutation de ce TA réduit l'activité de pPc2B' de 70% (Figure 25). Deux promoteurs actifs pour la transcription du gène *lacZ* coexisteraient donc dans la région *attI2* : Pc2A et Pc2B'.

c- Contribution de Pc2A et Pc2B' dans l'expression des cassettes

La présence de deux promoteurs actifs en tandem dans la région *attI2* interpelle sur le degré de participation de chaque promoteur dans le niveau de transcription des cassettes. L'activité mesurée avec un plasmide dérivé de pPc2 où Pc2A est inactivé est environ la moitié de l'activité de pPc2 (Figure 25B). En revanche, si c'est l'efficacité de Pc2B' qui est réduite, l'activité obtenue reste similaire à celle de pPc2 (Figure 25B). De même, l'introduction de la mutation de Pc2B' dans le plasmide dérivé de pPc2 où Pc2A est déjà inactivé n'a aucun effet sur le niveau d'expression mesuré (Figure 25B).

L'ensemble de ces données suggère que Pc2A contribuerait à hauteur de 50% à la transcription des gènes de cassette et qu'un Pc2B' moins efficace ne réduit pas le niveau d'expression global, et ceci de façon indépendante de la présence de Pc2A.

d- Influence des autres promoteurs Pc putatifs

Nous avons également vérifié que les autres promoteurs Pc putatifs n'ont pas d'incidence sur le niveau de transcription des cassettes. Cloné isolément, Pc2D ne génère aucune activité β -galactosidase (Figure 25, pPc2D) ; la situation est plus compliquée pour Pc2C. En effet, un dérivé de pPc2 dans lequel la région *attI2* est limitée au fragment contenant Pc2C (pPc2C) génère une certaine activité β -galactosidase (Figure 25). Mais une mutation qui réduit de moitié l'efficacité de Pc2C ne réduit pas le niveau d'activité de pPc2 (résultat non montré). Par conséquent, ni Pc2C, ni Pc2D ne contribuent à l'expression des gènes de cassettes des IM 2.

A.

	-35	-10
Pc2A	agag TTTTAA tttaatgctaaataaat TAAAAT gttatgag	
Pc2A-v2	agag TTTTAA tttaatgctaaataaat TAAAGT gttatgag	
Pc2B'	tgct TTGTAT ttttaaagaaagtctatt TAATAC aagtgat	
Pc2B'-v2	tgct TTATAT ttttaaagaaagtctatt TAATAC aagtgat	

B.

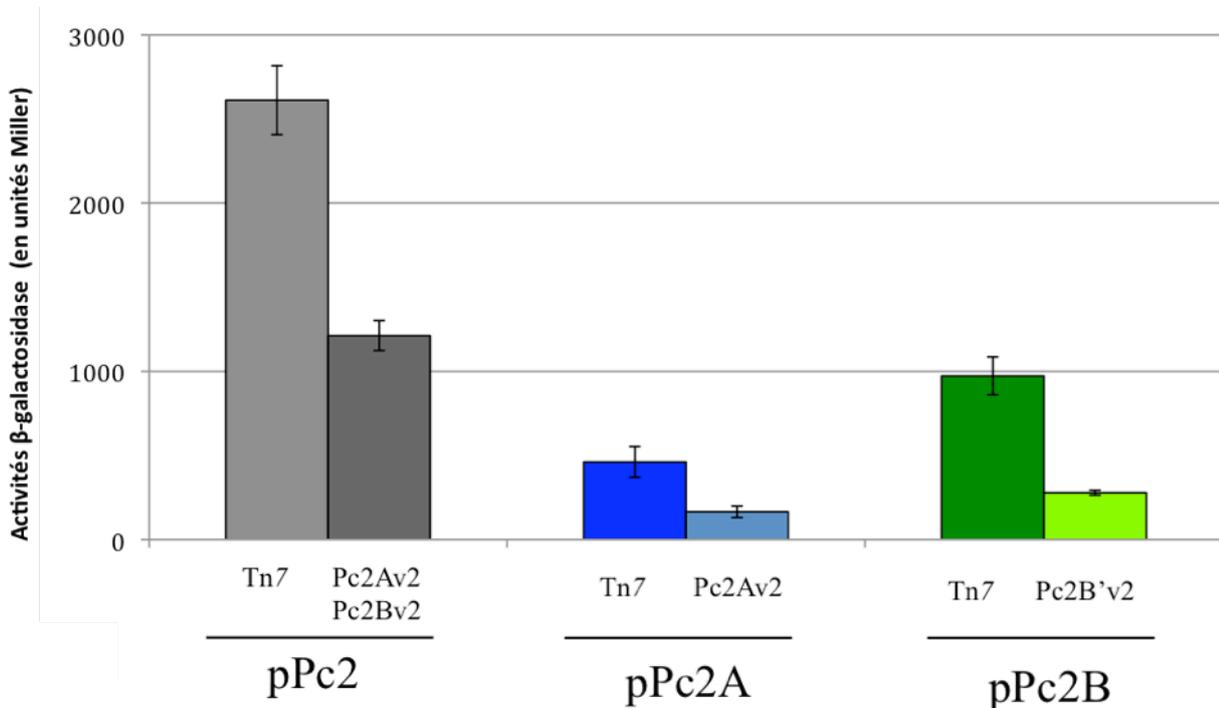


Figure 26: Caractérisation du polymorphisme de Pc2A et Pc2B'

A. Séquence des variants de Pc2A et Pc2B' des IM 2. Les variants PcA-v2 et Pc2B'-v2 sont ceux des IM 2 dont le gène *intI2* code une intégrase active.

B. La force des variants alternatifs de Pc2A et Pc2B' a été comparée au variant « classique » dans des constructions où le Pc est cloné isolément (pPc2A, pPc2B) et grâce à des essais β -galactosidase. L'effet simultané des deux mutations a été testé dans la construction où tout *attI2* dirige l'expression des cassettes (pPc2). Les barres d'erreur représentent les écarts type d'au moins cinq mesures indépendantes pour chaque plasmide. WT: Wild-type (sauvage).

e- Polymorphisme des promoteurs Pc2A et Pc2B'

Un examen de l'ensemble des séquences d'IM 2 disponibles en ligne a révélé que trois séquences présentent des variations à des positions clés des promoteurs Pc2A et/ou Pc2B'.

De façon intéressante, il s'agit des séquences des trois seuls IM 2 connus dont le gène *intI2* code une intégrase fonctionnelle (Fichiers GenBank DQ533990, DQ533991 et EU780012).

Dans les trois cas, Pc2A présente une substitution de l'élément -10 (TAAAAT en TAAAGT, la base mutée étant soulignée) et Pc2B' de l'élément -35 (TTGTTAT en TTATAT). Chacune de ces variations génèrent des variants environ deux fois moins efficaces de Pc2A ou Pc2B et, si elles sont introduites simultanément dans pPc2, elles résultent en une région *attI2* moitié moins efficace pour la transcription des gènes de cassettes (Figure 26). De plus, dans DQ533991, Pc2B' est dupliqué.

Discussion

Les travaux présentés dans ce manuscrit avaient pour objectif d'étudier la régulation de l'expression des gènes de résistance aux antibiotiques et de l'intégrase chez les IM.

1. Polymorphisme du promoteur Pc des IM 1

Des travaux précédents avaient décrit 8 variants de Pc des IM 1, sur la base des séquences des éléments -35 et -10 (Brizio *et al.*, 2006 ; Bunny, Hall et Stokes, 1995 ; Houang *et al.*, 2003 ; Levesque *et al.*, 1994 ; Riccio *et al.*, 2000), mais leurs forces respectives n'avaient jamais été comparées dans un seul essai. Nous avons donc réalisé une étude *in silico* à partir des données issues de la littérature et de l'INSD (International Nucleotide Sequence Database). Nous avons montré qu'il existait bien 8 variants sur la base des séquences des éléments -35 et -10 mais que cette définition était restrictive puisque environ 20% des IM 1 possèdent une région appelée « -10 étendue » connue pour augmenter l'activité de Pc (Mitchell *et al.*, 2003 ; Nesvera, Hochmannova et Patek, 1998). Cette région, créée par la modification de la base -14 de Pc (communément un C, par un G) constitue un nouvel élément de recrutement des ARN polymérases augmentant ainsi l'efficacité de la transcription comme cela a été montré par Murakami et coll. (Murakami *et al.*, 2002).

Par ailleurs, nous avons identifié d'autres mutations dans le promoteur Pc, situées entre l'élément -35 et le (+1) de transcription. La plupart ne sont retrouvées que dans une ou deux séquences et pourraient être liées à des erreurs de séquençage. Cependant, d'autres mutations sont retrouvées chez au moins 2 IM 1. Ainsi, 5 IM 1 distincts ayant le variant PcW présentent une substitution de la base en position -6, communément un G, en A. Cette mutation double l'efficacité de PcW dans nos conditions expérimentales (résultats non montrés). A ce jour, aucune base conservée ou considérée comme optimale pour la fonction d'un promoteur n'a été établie pour cette position. Dans le cas de Pc, cette mutation G-6A convertit une séquence TGTAAT en TATAAT (la base mutée est soulignée), ce qui correspond exactement à la séquence consensus de l'élément -10.

Cependant cette séquence est trop éloignée de l'élément -35 de Pc (22 pb) pour constituer un élément -10 alternatif fort et il n'y a aucun autre élément -35 putatif plus près. Pour expliquer l'augmentation de la force du PcW, nous proposons que cette mutation facilite la dénaturation des brins pendant la formation de la bulle à l'initiation de la transcription.

D'autre part, 10% des IM 1 présentent un promoteur supplémentaire, P2, localisé dans *attII*, 119 pb en aval de Pc (Stokes et Hall, 1989). Ce promoteur est créé via l'insertion de trois résidus guanine, ce qui corrige l'espacement entre des motifs -35 et -10 potentiels et leur permet ainsi de fonctionner comme promoteur.

Tenant compte des différents variants de Pc et de P2, nous avons montré qu'il existe pas moins de 20 combinaisons Pc-P2 possibles chez les IM 1. Nous avons donc souhaité comparer les différentes combinaisons de promoteurs dans un même essai.

Le promoteur PcS est généralement considéré comme un promoteur fort depuis qu'il a été établi que ce variant était 6,5 fois plus fort que le promoteur Ptac (Amann, Brosius et Ptashne, 1983 ; Levesque *et al.*, 1994). Nous avons comparé l'activité β -galactosidase de PcS et celle de Ptac avec la même séquence que celle utilisée par Levesque et coll. et montré que PcS était 100 fois plus faible que Ptac (résultats non montrés), ce qui est davantage compatible avec le fait que les promoteurs bactériens naturels sont tous moins efficaces que le Ptac. Même les meilleurs promoteurs bactériens artificiels sont au mieux deux fois plus forts que Ptac (Kagiya *et al.*, 2005). PcS avait été décrit comme 20 à 30 fois plus efficace que le PcW, celui-ci étant alors appelé Pc faible ; les variants PcH1 et PcH2 étant de force intermédiaire (Collis et Hall, 1995 ; Levesque *et al.*, 1994 ; Papagiannitsis, Tzouvelekis et Miriagou, 2009). Nous avons trouvé pour ces 4 variants PcS, PcW, PcH1, PcH2, des résultats en accord avec les précédentes études. Les variants plus rares PcPUO, PcIn42 et PcIn116 étaient équivalents à PcW ou plus faibles.

Pour le variant PcSS défini initialement comme étant plus fort que PcS, nous avons trouvé une efficacité 12 fois plus faible que PcS ; ceci était cependant attendu car ce promoteur comporte une mutation dans la séquence -35 à une position importante pour l'activité des promoteurs (Brizio *et al.*, 2006 ; Moyle, Waldburger et Susskind, 1991). Enfin, nous avons montré que la présence d'une région « -10 étendue » (motif TGN-10) augmentait effectivement l'efficacité de transcription du variant PcW d'un facteur 15 alors qu'elle n'a que peu d'effet sur des promoteurs plus forts comme PcS et PcH2.

Selon les variants, l'efficacité de Pc varie donc d'une amplitude de 1 à 30 (et même de 1 à 150 si on considère aussi les variants rares).

Enfin, nous avons analysé le niveau d'expression du gène rapporteur *lacZ* en présence du promoteur P2 couplé à différents variants de Pc. Le niveau d'expression résultant du couple Pc/P2 correspondait à celui du promoteur le plus fort, c'est à dire :

- ✓ De P2 seul, quand Pc est faible (PcW, PcH1)
- ✓ De Pc seul, quand Pc est fort (PcS, PcW_{TGN-10})

Ces résultats sont en opposition avec l'hypothèse de Lesvesque et coll., selon laquelle les activités de ces deux promoteurs s'additionnent, mais ces auteurs n'avaient pas vérifié le niveau d'activité lorsque seul P2 est actif (Levesque *et al.*, 1994 ; Papagiannitsis, Tzouveleki et Miriagou, 2009). L'activité de PcS est légèrement mais significativement diminuée en présence du P2. Nous l'expliquons par des interférences entre ces deux promoteurs en tandem.

Il est possible par exemple qu'une petite proportion d'ARN polymérases issues du PcS et en cours d'élongation soit déstabilisée par d'autres enzymes fixées sur le promoteur P2.

Ainsi, l'ensemble de ces données montre que le gène d'une même cassette de résistance aux antibiotiques peut être exprimé à des niveaux variables selon la combinaison de Pc-P2 de l'IM 1. De nombreuses études épidémiologiques ont décrit des intégrons identiques chez des

souches de même espèce non reliées épidémiologiquement ou chez des espèces différentes, mais peu ont étudié la diversité des promoteurs de cassettes. Or, cette donnée est importante pour étudier l'évolution des niveaux de résistance aux antibiotiques que confère une cassette ou un réseau de cassettes, les gènes des cassettes les plus éloignés du Pc étant les moins exprimés. Nous avons donc étudié la distribution des différentes combinaisons de Pc-P2, dans un premier temps à partir des données obtenues *in silico* en considérant que chaque réseau de cassettes détermine un IM 1. Cependant, cette analyse ne tenant pas compte de la fréquence de chaque IM 1 *in vivo*, elle ne correspondait pas à la réalité épidémiologique. C'est pourquoi nous avons aussi étudié les IM 1 de souches cliniques d'*E. coli* isolées de patients sains ou de prélèvements dans des situations pathologiques.

De façon surprenante, dans l'analyse *in silico*, le variant qui prédomine (41,7%) est le variant faible PcW, suivi de PcH1 (28%) et PcS (24,3%), les autres variants étant beaucoup moins fréquents.

Cependant, la faiblesse du variant PcW est très souvent compensée soit par une région « -10 étendue » (17,4%) ou par un P2 (7,2%).

Chez les souches cliniques d'*E.coli*, nous avons trouvé que les variants faibles PcW et PcH1 étaient les plus prévalents, 64% et 34% respectivement alors que le variant fort PcS était plus rare (2,3%). Le variant PcW présentait souvent soit un promoteur P2 (16,9% des 89 IM 1), soit une région « -10 étendue » (6,7%). Ces données montrent comme celles précédemment réalisées chez des entérobactéries (Lindstedt *et al.*, 2003 ; Schmitz *et al.*, 2001) une fréquence élevée d'intégrons ayant le variant PcW. Cependant, très souvent la faiblesse de PcW est compensée par une région -10 étendue, ce qui nous a donc conduit à proposer ce variant PcW_{TGN-10} comme un variant à part entière.

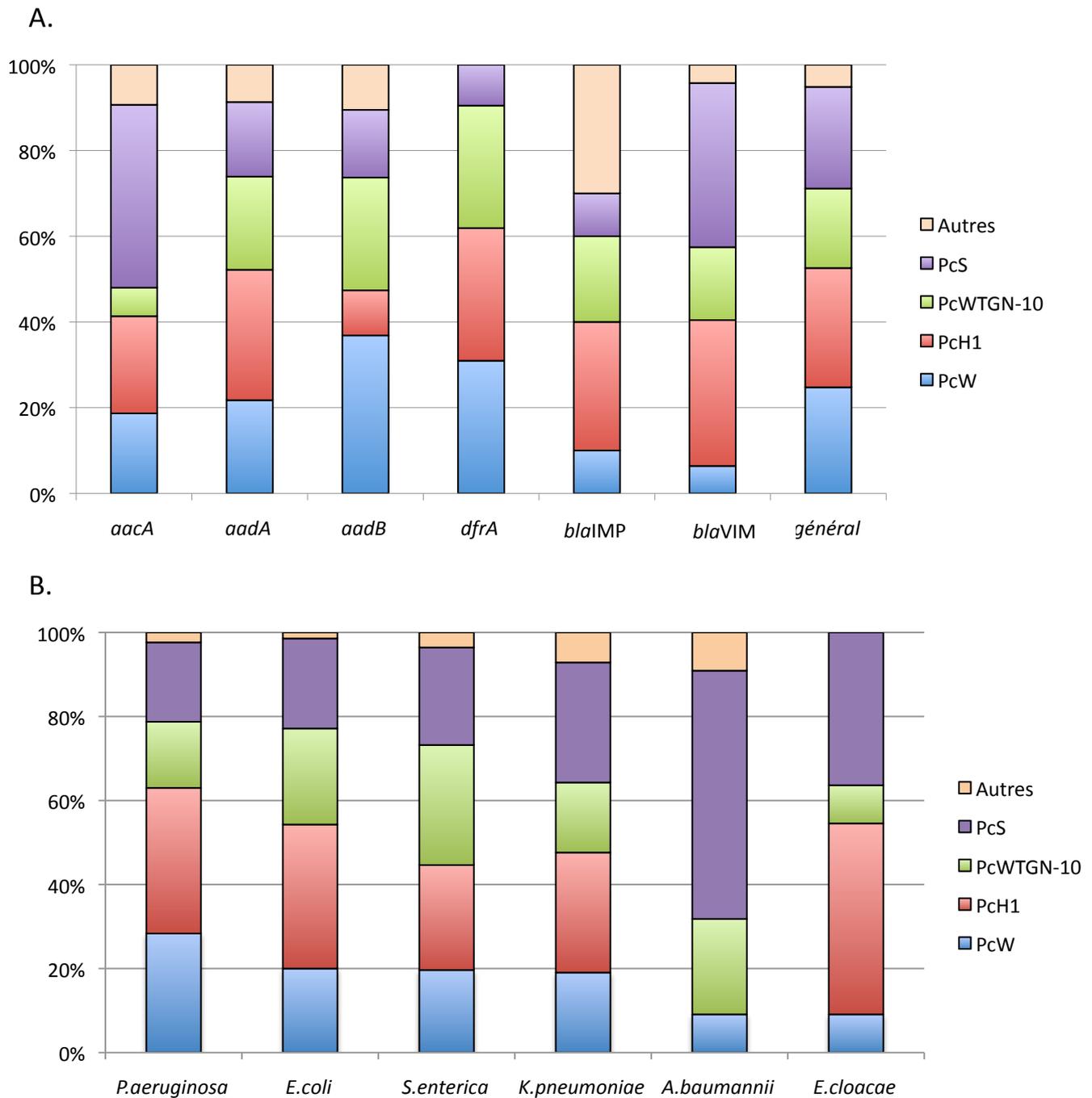


Figure 27: Distribution des variants de Pc en fonction de la nature de la première cassette de l'IM 1 ou de l'organisme hôte

L'analyse est basée sur la collection de 321 IM 1 distincts trouvés in silico.

A. Distribution des variants de Pc selon la nature de la première cassette

Chaque histogramme correspond à un des 7 types de première cassette de réseaux les plus fréquents : aacA (n=75), aadA (n=23), aadB (n=19), dfrA (n=42), blaIMP (n=20), blaVIM (n= 47; la distribution des variants dans la collection entière (n=321) est représenté dans le dernier histogramme (général)

B. Distribution des variants de Pc selon l'organisme hôte de l'IM 1

Chaque histogramme correspond à un des 6 organisme hôte les plus fréquents: P. aeruginosa (n = 127), E.coli (n=70), S.enterica (n=56), K.pneumoniae (n=42), A.baumannii (n=24), E.cloacae (n=11).

La hiérarchie des prévalences des variants de Pc ($PcS < PcW_{TGN-10} < PcW+P2 < PcH1 < PcW$) était l'exacte opposée de celle de leurs forces ($PcS > PcW_{TGN-10} > PcW+P2 > PcH1 > PcW$), et ceci quelle que soit l'origine géographique ou clinique des IM 1.

La forte proportion des variants peu efficaces PcW et PcH1 laisse penser que dans de nombreux IM 1, les gènes de cassettes ne sont pas fortement exprimés. Il se peut que certains gènes de cassettes n'aient pas besoin d'être fortement transcrits pour conférer le phénotype de résistance. Cette idée nous a conduit à rechercher une corrélation entre certains variants de Pc et la nature de la cassette en première position.

La grande diversité des cassettes des IM 1 rend cette analyse délicate mais il est possible de tirer quelques tendances illustrées dans la Figure 27A :

- Les IM 1 dont la première cassette contient un gène *aac(6')* ou *blaVIM* ont plus souvent le variant PcS, au détriment du PcW_{TGN-10} et du PcW, respectivement.
- *A contrario*, ceux avec *dfrA* présentent rarement un PcS mais plus souvent un PcW ou un PcW_{TGN-10}
- D'autre part, *aadB* en première position est plus souvent exprimée à partir d'un PcW
- Enfin, les IM 1 ayant *blaIMP* présentent la plus grande diversité de variants de Pc.

De plus, nous n'observons aucune association préférentielle des Pc forts pour les IM 1 dont le gène de la première cassette possède une TIR efficace ; par exemple, les gènes des cassettes *aadB*, dépourvu de TIR, sont transcrits par un variant faible et les gènes *blaVIM*, dotés d'une TIR proche du consensus, par le variant le plus fort.

D'autre part, un même IM 1 peut exister avec des Pc différents (comme In70 qui peut exister avec PcW, PcW_{TGN-10} , PcH1 ou PcH2), laissant penser que la séquence de Pc peut évoluer indépendamment du réseau de cassettes. Les paramètres qui régissent la mutabilité de Pc ne sont pas connus, le changement de variant n'ayant jamais été observé *in vitro*. Cette

Nom de IntI1	Variant(s) de Pc	aa 31	aa 32	aa 38	aa 39
IntI1 _{R32_H39}	PcW, PcH1	L	R	V	H
IntI1 _{R32_N39}	PcS, PcH2	L	R	V	<u>N</u>
IntI1 _{P32_H39}	PcW _{TGN-10}	L	<u>P</u>	V	H
IntI1 _{P32_N39}	PcS _{TGN-10} , PcH2 _{TGN-10}	L	<u>P</u>	V	<u>N</u>
IntI1 _{L32_H39}	PcW _{TAN-10}	L	<u>L</u>	V	H
IntI1 _{Q32_H39}	PcH1 _{TTN-10}	L	<u>Q</u>	V	H
IntI1 _{R32_I38_N39}	PcSS	L	R	<u>I</u>	<u>N</u>
IntI1 _{R32_A38_N39}	PcIn42	L	R	<u>A</u>	<u>N</u>
IntI1 _{R32_D38_N39}	PcPUO	L	R	<u>D</u>	<u>N</u>
IntI1 _{S31_R32_N39}	PcIn116	<u>S</u>	R	V	<u>N</u>

Tableau 19: Correspondance entre le variant de Pc et la séquence en acide aminés de l'intégrase IntI1

La nature des aa est indiquée en monolettre en suivant la nomenclature IUPAC. Les aa qui diffèrent de ceux de l'IntI1_{R32_H39} sont soulignés.

observation indique également que la nomenclature des IM 1 ne peut pas être basée uniquement sur la nature du réseau de cassettes.

Nous avons aussi étudié la répartition des Pc selon les espèces bactériennes et observé que la distribution des variants de Pc d'IM 1 était similaire quel que soit l'organisme hôte, à l'exception des IM 1 d'*Acinetobacter* où la variété des séquences de Pc est plus importante (Figure 27B).

2. Impact du polymorphisme de Pc sur l'activité de l'intégrase IntI1

La prédominance des variants faibles de Pc chez les souches cliniques est surprenante car elle laisse penser que seraient favorisés les IM 1 dont les gènes de cassettes sont moins fortement transcrits, et donc *a priori* moins à même de conférer une résistance à un antibiotique. Cependant cette sélection n'est peut-être pas liée qu'au seul promoteur Pc. En effet, le promoteur Pc étant localisé dans le gène *intI1*, ses variations de séquence entraînent des modifications d'aa de la protéine IntI1 (Tableau 19).

Notre analyse de l'activité de recombinaison des trois intégrases correspondant aux principaux variants de Pc a montré que l'intégrase codée lorsque le Pc est faible (PcW ou PcH1) était la plus efficace pour le réarrangement des cassettes. La forte prévalence de ces variants suggérerait donc que la pression de sélection antibiotique s'exerce davantage sur l'activité de l'intégrase que sur le niveau d'expression des gènes de cassette. En fait, ces deux fonctionnalités sont liées car une intégrase plus processive pour l'excision des cassettes crée d'autant plus souvent de nouveaux arrangements de cassettes, dont celui (ou ceux) permettant une expression maximale du gène requis, par exemple en première position

Nous avons montré que si les trois IntI1 les plus prévalentes présentent des efficacités similaires pour catalyser les réactions *attI* x *attC*, l'IntI1_{R32_H39} est en revanche beaucoup plus active que les autres pour les réactions *attC* x *attC*. Ces observations permettent de proposer un rôle pour les aa R32 et H39 d'IntI1 dans son activité de recombinaison des cassettes,

vraisemblablement dans la capacité d'IntI1 à se lier aux sites *attC*. Les substitutions R32P et H39N diminuent d'ailleurs beaucoup l'activité d'excision de l'intégrase. Ces aa s'alignent avec les aa d'une hélice α de l'intégrase du super-intégron de *V.cholerae* VchIntIA, hélice impliquée dans la liaison aux sites *attC* (Demarre *et al.*, 2007). L'aa R32 joue aussi probablement un rôle dans la recombinaison *attI* x *attC*, une substitution R32P dans les intégrases IntI1_{R32_H39} et IntI1_{R32_N39} réduisant la fréquence d'intégration. L'aa H39 ne semble pas être impliqué dans cette réaction. Les recombinaisons *attI* x *attC* et *attC* x *attC* peuvent donc impliquer différentes régions de l'intégrase. En effet, Demarre et coll. ont isolé deux mutants de l'intégrase IntI1_{R32_H39}, IntI1_{P109L} et IntI1_{D161G}, qui ont montré des efficacités d'intégration plus élevées (Demarre *et al.*, 2007).

Dans notre travail, les activités d'excision des variants d'IntI1 ont été comparées avec plusieurs sites *attC* mais qui ne représentent pas à eux seuls l'intégralité des structures de ces sites. Or il reste envisageable que les différences d'activités de recombinaison entre les variants d'IntI1 n'aient pas la même amplitude selon la nature du site *attC*. Si c'est le cas, les variations de Pc influenceraient aussi, outre la cinétique, la nature des réarrangements de cassettes et donc l'évolution des IM 1.

3. Interférences entre Pc et PintI des IM 1

Chez les IM 1, le promoteur Pc fait face à celui du gène de l'intégrase (Collis et Hall, 1995), une situation propice à des interférences entre les deux promoteurs. Or le promoteur PintI1 est réprimé par la protéine LexA et est induit quand est déclenchée la réponse SOS, par exemple en présence d'antibiotiques (Guerin *et al.*, 2009). Nous avons montré que la présence du promoteur faible PcW n'empêche pas l'induction SOS-dépendante de la transcription d'*intI1*, mais qu'il diminue le niveau d'expression d'*intI1*.

L'effet des autres variants majeurs de Pc (tous plus forts que PcW) sur le niveau de transcription d'*intI1* n'a pas été encore déterminé pour le moment mais nous supposons qu'il le diminuera d'autant plus que le variant de Pc est fort. Cela suggère que les variants forts de Pc réduiraient non seulement l'activité mais aussi l'expression de l'intégrase, du fait d'interférences entre Pc et PintI1. Dans ces intégrons IM1, l'expression des gènes de cassettes prévaudrait donc sur les mouvements de cassettes.

De même, la présence du promoteur additionnel P2, qui augmente le niveau d'expression des cassettes, supprime le site de liaison pour LexA, le répresseur de PintI1, ce qui théoriquement devrait induire l'expression de l'intégrase. Cependant nous avons montré que l'activité de P2 a un effet délétère sur le niveau de transcription de PintI1. Ainsi, toutes les séquences promotrices de Pc/P2 permettant d'augmenter l'expression des gènes de cassettes auraient un effet négatif sur celle du gène de l'intégrase.

Le mécanisme moléculaire des régulations observées entre les promoteurs PcW, P2 et PintI1 n'a pas été déterminé. Puisque inactiver P2 permet de restaurer l'activité de PintI1, ce n'est pas la seule présence de P2 qui inhibe PintI1, mais son activité. Le chevauchement de P2 et PintI1 entraîne probablement des compétitions stériques pour le recrutement de l'ARN polymérase au profit du promoteur le plus affiné, qui serait donc P2.

L'incidence négative de P2 sur l'expression de l'intégrase explique peut-être pourquoi la configuration PcW+P2 est moins représentée (à la fois *in silico* et *in vivo*) que PcW seul alors qu'il permet une augmentation de la transcription des cassettes.

Plusieurs types de mécanismes pourraient expliquer la régulation entre les promoteurs convergents PcW et PintI1. Cependant, nos données suggèrent qu'il ne s'agit pas d'une régulation par ARN antisens en *cis* car en présence de PcW, un seul des deux promoteurs est affecté par l'activité de l'autre. La quantification des transcrits, issus de Pc et de PintI1 dans les différentes constructions, permettra d'exclure totalement ces mécanismes post-

transcriptionnels de régulation. Nous proposons donc que ce soit l'activité des promoteurs qui interfère. Cette hypothèse est confortée par le fait que seule l'activité du variant faible PcW et pas celle du variant fort PcS est affectée par l'induction de l'activité de PintI1. Des expériences de transcription *in vitro* seraient nécessaires pour identifier le mécanisme précis d'interférence entre ces deux promoteurs.

L'incidence négative de P2 sur l'expression de l'intégrase explique peut-être pourquoi la configuration PcW+P2 est moins représentée (à la fois *in silico* et *in vivo*) que PcW seul alors qu'il permet une augmentation de la transcription des cassettes.

4. Transcription des gènes de cassettes et de l'intégrase des IM 2

L'étude des promoteurs des IM s'est poursuivie par la caractérisation des promoteurs qui contrôlent la transcription des gènes des IM 2. Le gène *intI2* est transcrit à partir d'un promoteur PintI2 qui a été localisé précisément. Ce promoteur chevauche un opérateur putatif de la protéine LexA mais pour lequel nous montrons qu'il est en fait non fonctionnel car incapable de fixer cette protéine. Pourtant, la séquence de cet opérateur putatif contient les éléments les plus importants des opérateurs avérés de LexA, notamment des triplets terminaux « CTG » et « CAG » espacés par 10 pb riches en A/T. La richesse relative en G/C de l'opérateur putatif des IM 2 (4 G/C parmi les 10 pb) existe aussi chez des opérateurs fonctionnels de LexA. En revanche, la position des bases GC au sein de l'opérateur, notamment avec le GCC central, est plus atypique (Figure 23).

Quoi qu'il en soit, la transcription du gène *intI2* n'est pas induite par la réponse SOS comme c'est le cas du gène *intI1* des IM 1, du SI de *V. cholerae* et probablement d'autres classes d'intégrons. Enfin, le gène *intI2* possède une TIR proche du consensus (AGGA-N₇-ATG) qui permet une traduction efficace. Tous ces éléments expliquent que le niveau d'expression du gène *intI2* soit plus élevé que celui d'*intI1* (de 40 fois lorsque la transcription d'*intI1* est réprimée, de 10 fois si elle est induite), bien qu'il code une protéine non fonctionnelle.

L'expression des gènes de cassette, elle, dépend uniquement de la région *attI2* et non d'une séquence localisée dans le gène de l'intégrase comme c'est le cas chez les IM 1 et 3. Le plus surprenant, c'est que deux promoteurs semblent participer à ce processus : Pc2A (TTTTAA-N₁₇-TAAAAT) et Pc2B' (TTGTAT-N₁₈-TAATAC). La contribution de Pc2B' n'est pas claire car, s'il est possible de détecter des transcrits issus de ce promoteur, sa mutation n'a pas d'effet sur le niveau d'expression global des gènes de cassettes. La possibilité que dans ce cas ce soit Pc2A, plus en amont, qui compense la faiblesse a été écartée dans une construction où ce dernier est muté.

La signification biologique de la coexistence de deux promoteurs fonctionnels n'est pas claire ; en général, des promoteurs en tandem sont différemment régulés et/ou transcrits par des facteurs σ différents ; or Pc2A et Pc2B' se comportent tous deux comme des promoteurs à σ^{70} et aucun site de liaison pour un régulateur bien connu n'a peut être identifié. De plus, bien que les promoteurs divergents Pc2A et PintI2 ne sont éloignés que de 36 pb (distance entre les éléments -35), leurs activités n'interfèrent pas. De façon intéressante, les 3 IM 2 codant des intégrases fonctionnelles présentent des variants de Pc moins efficaces qui affectent le niveau d'expression des gènes de cassettes. Ainsi, comme pour les IM 1, il existerait donc aussi, chez les IM 2, une corrélation inverse entre les deux fonctionnalités des intégrons.

5. Evolution des variants de Pc : la recherche de l'équilibre ?

L'ensemble de nos résultats montre une corrélation négative entre les deux fonctionnalités des IM : expression des gènes des cassettes d'une part et activité de l'intégrase de l'autre : les éléments qui augmentent l'expression des cassettes (un variant fort de Pc, un P2 actif) ont un effet négatif sur l'expression de l'intégrase et, dans le cas des Pc fort, sur son activité de recombinaison.

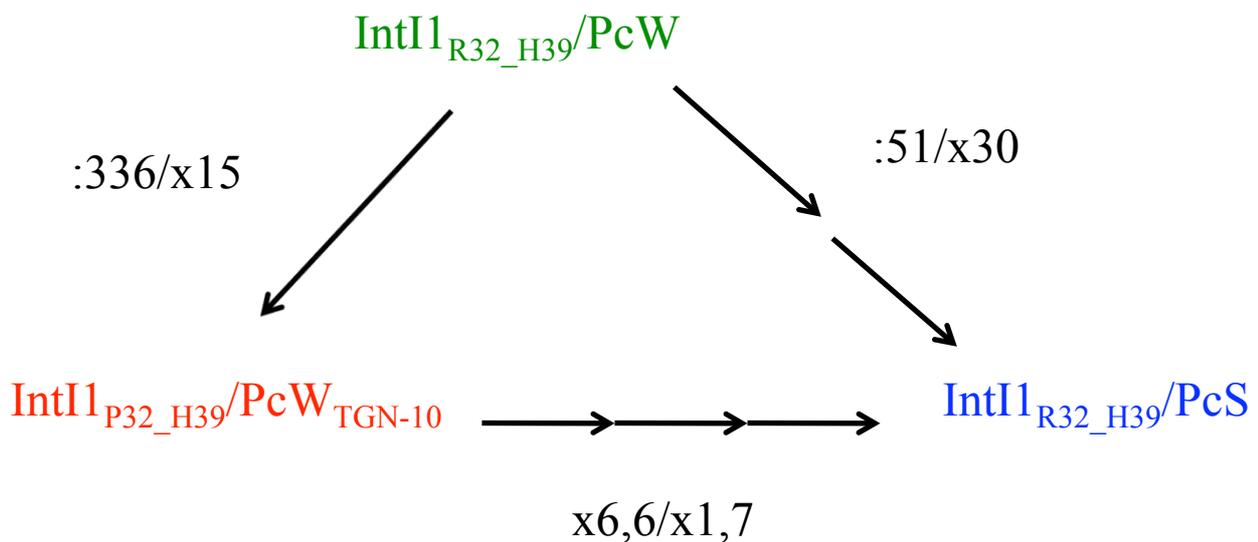


Figure 28: *Modèle de co-évolution des variants de Pc et de l'intégrase IntI1*

Chaque flèche indique la mutation d'une base. A chaque sommet du triangle est placé un couple variant IntI1/variant Pc correspondant. Chaque flèche indique la variation d'une base. L'effet de la transition d'une IntI1 à un autre sur son activité d'excision est indiquée à gauche du « / », celle sur la force du promoteur à droite de ce symbole.

Par exemple, la substitution d'une base suffit à la transition de IntI1_{R32_H39} en IntI1_{P32_H39} (et de PcW en PcW_{TGN-10}), ce qui réduit l'activité d'IntI1 d'un facteur 700 mais augmente la force de PcW d'un facteur 15.

La fonctionnalité qui sera privilégiée dépend du variant de Pc : les variants les plus forts déterminent des IntI1 peu efficaces et réduisent probablement très fortement leur expression ; les variants faibles, eux, permettent une production et activité maximale de l'intégrase mais une expression des cassettes faible et même réduite quand l'expression d'*intI1* est induite.

Les études épidémiologiques de la littérature et ce travail montrent que ce sont les IM 1 avec les Pc faibles qui sont favorisés, ce qui laisse penser qu'*a priori*, la pression de sélection antibiotique s'exerce en premier lieu sur l'activité de l'intégrase. En effet, une intégrase plus processive en excision permettra plus facilement de catalyser des réarrangements de cassettes de façon à ce que la cassette nécessaire pour conférer la résistance soit localisée aux premières positions d'un réseau, au plus près de Pc, là où son niveau d'expression sera le plus élevé. Par conséquent, une intégrase plus active permet indirectement une expression plus forte des gènes de cassettes insérés.

Pourtant, il existe un certain nombre d'IM 1 avec des variants forts de Pc ou de P2 (et donc une activité ou une expression moindre de l'intégrase), mais ils semblent peu prévalents. Il est possible que dans certaines conditions, des promoteurs forts et/ou des réseaux de cassettes stables soient plus avantageux pour les bactéries. Un élément de réponse pourrait être l'association préférentielle de certaines cassettes en première position à des variants forts de Pc, mais nos études ne permettent pas de trancher dans ce sens. D'ailleurs, un certain nombre d'IM 1 peuvent être associés à des variants différents de Pc et P2 ; l'existence de ces IM 1 laisse penser que Pc et P2 peuvent évoluer indépendamment du réseau de cassettes. Cette hypothèse reste plus probable qu'une évolution convergente des réseaux à partir de 2 IM 1 dont les variants de Pc diffèrent. La question de l'évolutivité des promoteurs Pc et P2 n'a jamais été évoquée auparavant. Elle implique qu'il existe un variant originel de Pc.

Or, récemment, des IM 1 localisés sur le chromosome de protéobactéries de l'environnement ont été identifiés et il a été proposé qu'ils soient des ancêtres des IM 1 cliniques actuels (Gillings *et al.*, 2008). Or ces IM 1 ont tous le PcW. C'est pourquoi nous proposons un modèle d'évolution de Pc dans lequel le variant original, PcW, a muté de façon à augmenter l'efficacité du promoteur (Figure 28). D'ailleurs, la mutation d'une seule base de PcW suffit pour créer PcW_{TGN-10}, PcH1 ou PcH2. Le contexte clinique pourrait avoir poussé Pc a évolué vers un variant encore plus fort, PcS.

Conclusions et Perspectives

Les intégrons sont des supports génétiques majeurs de résistances aux antibiotiques assurant à la fois la dissémination et l'expression de gènes de résistance aux antibiotiques contenus dans des cassettes. Nos travaux sur les IM 1, les intégrons les plus prévalents dans le monde clinique, ont décrit l'existence d'un fort polymorphisme de Pc et comment il influence simultanément mais de façon inverse les deux fonctionnalités des IM 1. Chez les IM 2, nous avons montré que Pc prend la forme de deux promoteurs dont il reste à déterminer précisément les (+1) de transcription et la contribution exacte dans l'expression des gènes de cassettes. Chez cette classe d'IM aussi, efficacité des Pc et activité de l'intégrase semblent inversement liées. Les IM 3 n'ont pas été étudiés ici mais l'existence de deux variants de Pc (Correia *et al.*, 2003), chacun déterminant une protéine IntI3 différente, suggère une situation analogue à celle des IM 1. La co-évolution des gènes *intI* et des promoteurs Pc semble donc une règle générale des IM. Chez les IM 1 et 2, elle assure un équilibre entre activité de l'intégrase et expression des gènes de résistance, réduisant d'autant le coût biologique pour la bactérie.

Ce travail pose donc beaucoup de questions sur les choix évolutifs conduisant à différents intégrons. La sélection de tel ou tel variant de Pc et donc d'intégrase chez les IM 1 est-elle liée à des conditions environnementales particulières ? Quel est le rôle de la pression de sélection antibiotique sur cette sélection ? Autant de questions qui ouvrent de nombreuses perspectives de recherche.

Références bibliographiques

Références bibliographiques

- Abella, M., I. Erill, M. Jara, G. Mazon, S. Campoy et J. Barbe. 2004. «Widespread distribution of a *lexA*-regulated DNA damage-inducible multiple gene cassette in the Proteobacteria phylum». *Mol Microbiol.* vol. 54, no 1, p. 212-222.
- Adhya, S., et M. Gottesman. 1982. «Promoter occlusion: transcription through a promoter may inhibit its activity». *Cell.* vol. 29, no 3, p. 939-944.
- Ahmed, A. M., K. Furuta, K. Shimomura, Y. Kasama et T. Shimamoto. 2006. «Genetic characterization of multidrug resistance in *Shigella* spp. from Japan». *J Med Microbiol.* vol. 55, no Pt 12, p. 1685-1691.
- Ahmed, A. M., A. I. Hussein et T. Shimamoto. 2007. «*Proteus mirabilis* clinical isolate harbouring a new variant of *Salmonella* genomic island 1 containing the multiple antibiotic resistance region». *J Antimicrob Chemother.* vol. 59, no 2, p. 184-190.
- Ahmed, A. M., H. Nakano et T. Shimamoto. 2005. «Molecular characterization of integrons in non-typhoid *Salmonella* serovars isolated in Japan: description of an unusual class 2 integron». *J Antimicrob Chemother.* vol. 55, no 3, p. 371-374.
- Amann, E., J. Brosius et M. Ptashne. 1983. «Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*». *Gene.* vol. 25, no 2-3, p. 167-178.
- Arakawa, Y., M. Murakami, K. Suzuki, H. Ito, R. Wacharotayankun, S. Ohsuka, N. Kato et M. Ohta. 1995. «A novel integron-like element carrying the metallo-beta-lactamase gene *blaIMP*». *Antimicrob Agents Chemother.* vol. 39, no 7, p. 1612-1615.
- Arthur, D. C., A. F. Ghetu, M. J. Gubbins, R. A. Edwards, L. S. Frost et J. N. Glover. 2003. «*FinO* is an RNA chaperone that facilitates sense-antisense RNA interactions». *EMBO J.* vol. 22, no 23, p. 6346-6355.
- Aubert, D., T. Naas et P. Nordmann. 2003. «*IS1999* increases expression of the extended-spectrum beta-lactamase VEB-1 in *Pseudomonas aeruginosa*». *J Bacteriol.* vol. 185, no 17, p. 5314-5319.
- Barlow, R. S., et K. S. Gobius. 2006. «Diverse class 2 integrons in bacteria from beef cattle sources». *J Antimicrob Chemother.* vol. 58, no 6, p. 1133-1138.
- Biskri, L., M. Bouvier, A. M. Guerout, S. Boissnard et D. Mazel. 2005. «Comparative study of class 1 integron and *Vibrio cholerae* superintegron integrase activities». *J Bacteriol.* vol. 187, no 5, p. 1740-1750.
- Biskri, L., et D. Mazel. 2003. «Erythromycin esterase gene *ere(A)* is located in a functional gene cassette in an unusual class 2 integron». *Antimicrob Agents Chemother.* vol. 47, no 10, p. 3326-3331.

- Bissonnette, L., S. Champetier, J. P. Buisson et P. H. Roy. 1991. «Characterization of the nonenzymatic chloramphenicol resistance (cmlA) gene of the In4 integron of Tn1696: similarity of the product to transmembrane transport proteins». *J Bacteriol.* vol. 173, no 14, p. 4493-4502.
- Bissonnette, L., et P. H. Roy. 1992. «Characterization of In0 of *Pseudomonas aeruginosa* plasmid pVS1, an ancestor of integrons of multiresistance plasmids and transposons of gram-negative bacteria». *J Bacteriol.* vol. 174, no 4, p. 1248-1257.
- Borukhov, S., et E. Nudler. 2008. «RNA polymerase: the vehicle of transcription». *Trends Microbiol.* vol. 16, no 3, p. 126-134.
- Boucher, Y., M. Labbate, J. E. Koenig et H. W. Stokes. 2007. «Integrons: mobilizable platforms that promote genetic diversity in bacteria». *Trends Microbiol.* vol. 15, no 7, p. 301-309.
- Bouvier, M., G. Demarre et D. Mazel. 2005. «Integron cassette insertion: a recombination process involving a folded single strand substrate». *EMBO J.* vol. 24, no 24, p. 4356-4367.
- Bouvier, M., M. Ducos-Galand, C. Loot, D. Bikard et D. Mazel. 2009. «Structural features of single-stranded integron cassette attC sites and their role in strand selection». *PLoS Genet.* vol. 5, no 9, p. e1000632.
- Brizio, A., T. Conceicao, M. Pimentel, G. Da Silva et A. Duarte. 2006. «High-level expression of IMP-5 carbapenemase owing to point mutation in the -35 promoter region of class 1 integron among *Pseudomonas aeruginosa* clinical isolates». *Int J Antimicrob Agents.* vol. 27, no 1, p. 27-31.
- Brown, N. L., J. V. Stoyanov, S. P. Kidd et J. L. Hobman. 2003. «The MerR family of transcriptional regulators». *FEMS Microbiol Rev.* vol. 27, no 2-3, p. 145-163.
- Bunny, K. L., R. M. Hall et H. W. Stokes. 1995. «New mobile gene cassettes containing an aminoglycoside resistance gene, *aacA7*, and a chloramphenicol resistance gene, *catB3*, in an integron in pBWH301». *Antimicrob Agents Chemother.* vol. 39, no 3, p. 686-693.
- Burr, T., J. Mitchell, A. Kolb, S. Minchin et S. Busby. 2000. «DNA sequence elements located immediately upstream of the -10 hexamer in *Escherichia coli* promoters: a systematic study». *Nucleic Acids Res.* vol. 28, no 9, p. 1864-1870.
- Butala, M., D. Zgur-Bertok et S. J. Busby. 2009. «The bacterial LexA transcriptional repressor». *Cell Mol Life Sci.* vol. 66, no 1, p. 82-93.
- Byrne-Bailey, K. G., W. H. Gaze, P. Kay, A. B. Boxall, P. M. Hawkey et E. M. Wellington. 2009. «Prevalence of sulfonamide resistance genes in bacterial isolates from manured agricultural soils and pig slurry in the United Kingdom». *Antimicrob Agents Chemother.* vol. 53, no 2, p. 696-702.

- Callen, B. P., K. E. Shearwin et J. B. Egan. 2004. «Transcriptional interference between convergent promoters caused by elongation over the promoter». *Mol Cell*. vol. 14, no 5, p. 647-656.
- Caroff, N., E. Espaze, I. Berard, H. Richet et A. Reynaud. 1999. «Mutations in the ampC promoter of *Escherichia coli* isolates resistant to oxyiminocephalosporins without extended spectrum beta-lactamase production». *FEMS Microbiol Lett*. vol. 173, no 2, p. 459-465.
- Casin, I., F. Bordon, P. Bertin, A. Coutrot, I. Podglajen, R. Brasseur et E. Collatz. 1998. «Aminoglycoside 6'-N-acetyltransferase variants of the Ib type with altered substrate profile in clinical isolates of *Enterobacter cloacae* and *Citrobacter freundii*». *Antimicrob Agents Chemother*. vol. 42, no 2, p. 209-215.
- Castanheira, M., H. S. Sader, R. N. Jones, E. Debbia, R. C. Picao et A. C. Gales. 2007. «In71, an *Enterobacter cloacae* blaVIM-1-carrying integron related to In70.2 from Italian *Pseudomonas aeruginosa* isolates: a SENTRY Antimicrobial Surveillance Program report». *Microb Drug Resist*. vol. 13, no 2, p. 130-134.
- Cavanagh, A. T., A. D. Klocko, X. Liu et K. M. Wassarman. 2008. «Promoter specificity for 6S RNA regulation of transcription is determined by core promoter sequences and competition for region 4.2 of sigma70». *Mol Microbiol*. vol. 67, no 6, p. 1242-1256.
- Chen, D., et D. M. Lilley. 1999. «Transcription-induced hypersupercoiling of plasmid DNA». *J Mol Biol*. vol. 285, no 2, p. 443-448.
- Clermont, O., S. Bonacorsi et E. Bingen. 2000. «Rapid and simple determination of the *Escherichia coli* phylogenetic group». *Appl Environ Microbiol*. vol. 66, no 10, p. 4555-4558.
- Cocchi, S., E. Grasselli, M. Gutacker, C. Benagli, M. Convert et J. C. Piffaretti. 2007. «Distribution and characterization of integrons in *Escherichia coli* strains of animal and human origin». *FEMS Immunol Med Microbiol*. vol. 50, no 1, p. 126-132.
- Coleman, N., S. Tetu, N. Wilson et A. Holmes. 2004. «An unusual integron in *Treponema denticola*». *Microbiology*. vol. 150, no Pt 11, p. 3524-3526.
- Coleman, N. V., et A. J. Holmes. 2005. «The native *Pseudomonas stutzeri* strain Q chromosomal integron can capture and express cassette-associated genes». *Microbiology*. vol. 151, no Pt 6, p. 1853-1864.
- Collis, C. M., et R. M. Hall. 1992. «Gene cassettes from the insert region of integrons are excised as covalently closed circles». *Mol Microbiol*. vol. 6, no 19, p. 2875-2885.
- Collis, C. M., et R. M. Hall. 1995. «Expression of antibiotic resistance genes in the integrated cassettes of integrons». *Antimicrob Agents Chemother*. vol. 39, no 1, p. 155-162.
- Collis, C. M., et R. M. Hall. 2004. «Comparison of the structure-activity relationships of the integron-associated recombination sites attI3 and attI1 reveals common features». *Microbiology*. vol. 150, no Pt 5, p. 1591-1601.

- Collis, C. M., M. J. Kim, S. R. Partridge, H. W. Stokes et R. M. Hall. 2002. «Characterization of the class 3 integron and the site-specific recombination system it determines». *J Bacteriol.* vol. 184, no 11, p. 3017-3026.
- Collis, C. M., G. D. Recchia, M. J. Kim, H. W. Stokes et R. M. Hall. 2001. «Efficiency of recombination reactions catalyzed by class 1 integron integrase IntI1». *J Bacteriol.* vol. 183, no 8, p. 2535-2542.
- Correia, M., F. Boavida, F. Grosso, M. J. Salgado, L. M. Lito, J. M. Cristino, S. Mendo et A. Duarte. 2003. «Molecular characterization of a new class 3 integron in *Klebsiella pneumoniae*». *Antimicrob Agents Chemother.* vol. 47, no 9, p. 2838-2843.
- Court, D. L., A. B. Oppenheim et S. L. Adhya. 2007. «A new look at bacteriophage lambda genetic networks». *J Bacteriol.* vol. 189, no 2, p. 298-304.
- Da Re, S., F. Garnier, E. Guerin, S. Campoy, F. Denis et M. C. Ploy. 2009. «The SOS response promotes *qnrB* quinolone-resistance determinant expression». *EMBO Rep.* vol. 10, no 8, p. 929-933.
- Dalsgaard, A., A. Forslund, N. V. Tam, D. X. Vinh et P. D. Cam. 1999. «Cholera in Vietnam: changes in genotypes and emergence of class I integrons containing aminoglycoside resistance gene cassettes in *Vibrio cholerae* O1 strains isolated from 1979 to 1996». *J Clin Microbiol.* vol. 37, no 3, p. 734-741.
- de Vin, F., P. Radstrom, L. Herman et L. De Vuyst. 2005. «Molecular and biochemical analysis of the galactose phenotype of dairy *Streptococcus thermophilus* strains reveals four different fermentation profiles». *Appl Environ Microbiol.* vol. 71, no 7, p. 3659-3667.
- Demarre, G., C. Frumerie, D. N. Gopaul et D. Mazel. 2007. «Identification of key structural determinants of the IntI1 integron integrase that influence attC x attI1 recombination efficiency». *Nucleic Acids Res.* vol. 35, no 19, p. 6475-6489.
- Dennis, P. P., M. Ehrenberg et H. Bremer. 2004. «Control of rRNA synthesis in *Escherichia coli*: a systems biology approach». *Microbiol Mol Biol Rev.* vol. 68, no 4, p. 639-668.
- Depardieu, F., I. Podglajen, R. Leclercq, E. Collatz et P. Courvalin. 2007. «Modes and modulations of antibiotic resistance gene expression». *Clin Microbiol Rev.* vol. 20, no 1, p. 79-114.
- Dorman, C. J. 2009. «Nucleoid-associated proteins and bacterial physiology». *Adv Appl Microbiol.* vol. 67, p. 47-64.
- Dubois, V., M. P. Parizano, C. Arpin, L. Coulangue, M. C. Beziau et C. Quentin. 2007. «High genetic stability of integrons in clinical isolates of *Shigella* spp. of worldwide origin». *Antimicrob Agents Chemother.* vol. 51, no 4, p. 1333-1340.
- Dykxhoorn, D. M., R. St Pierre et T. Linn. 1996. «Synthesis of the beta and beta' subunits of *Escherichia coli* RNA polymerase is autogenously regulated in vivo by both

- transcriptional and translational mechanisms». *Mol Microbiol.* vol. 19, no 3, p. 483-493.
- Ebright, R. H. 1993. «Transcription activation at Class I CAP-dependent promoters». *Mol Microbiol.* vol. 8, no 5, p. 797-802.
- Enne, V. I., D. M. Livermore, P. Stephens et L. M. Hall. 2001. «Persistence of sulphonamide resistance in *Escherichia coli* in the UK despite national prescribing restriction». *Lancet.* vol. 357, no 9265, p. 1325-1328.
- Estrem, S. T., T. Gaal, W. Ross et R. L. Gourse. 1998. «Identification of an UP element consensus sequence for bacterial promoters». *Proc Natl Acad Sci U S A.* vol. 95, no 17, p. 9761-9766.
- Ferreira da Silva, M., I. Vaz-Moreira, M. Gonzalez-Pajuelo, O. C. Nunes et C. M. Manaia. 2007. «Antimicrobial resistance patterns in Enterobacteriaceae isolated from an urban wastewater treatment plant». *FEMS Microbiol Ecol.* vol. 60, no 1, p. 166-176.
- Fonseca, E. L., F. Dos Santos Freitas, V. V. Vieira et A. C. Vicente. 2008. «New qnr gene cassettes associated with superintegron repeats in *Vibrio cholerae* O1». *Emerg Infect Dis.* vol. 14, no 7, p. 1129-1131.
- Fournier, P. E., D. Vallenet, V. Barbe, S. Audic, H. Ogata, L. Poirel, H. Richet, C. Robert, S. Mangenot, C. Abergel, P. Nordmann, J. Weissenbach, D. Raoult et J. M. Claverie. 2006. «Comparative genomics of multidrug resistance in *Acinetobacter baumannii*». *PLoS Genet.* vol. 2, no 1, p. e7.
- Francia, M. V., de la Cruz, F. et J.M. Garcia Lobo. 1993. «Secondary-sites for integration mediated by the *Tn21* integrase». *Mol. Microbiol.* vol. 10, no 4, p. 823-828
- Frumerie, C., M. Ducos-Galand, D. N. Gopaul et D. Mazel. 2009. «The relaxed requirements of the integron cleavage site allow predictable changes in integron target specificity». *Nucleic Acids Res.*
- Gassama Sow, A., M. H. Diallo, M. Gatet, F. Denis, A. Aidara-Kane et M. C. Ploy. 2008. «Description of an unusual class 2 integron in *Shigella sonnei* isolates in Senegal (sub-Saharan Africa)». *J Antimicrob Chemother.* vol. 62, no 4, p. 843-844.
- Gassama-Sow, A., M. H. Diallo, C. S. Boye, B. Garin, J. M. Sire, A. I. Sow et A. Aidara-Kane. 2006. «Class 2 integron-associated antibiotic resistance in *Shigella sonnei* isolates in Dakar, Senegal». *Int J Antimicrob Agents.* vol. 27, no 3, p. 267-270.
- Gibreel, A., et O. Skold. 1998. «High-level resistance to trimethoprim in clinical isolates of *Campylobacter jejuni* by acquisition of foreign genes (*dfr1* and *dfr9*) expressing drug-insensitive dihydrofolate reductases». *Antimicrob Agents Chemother.* vol. 42, no 12, p. 3059-3064.
- Gillings, M., Y. Boucher, M. Labbate, A. Holmes, S. Krishnan, M. Holley et H. W. Stokes. 2008. «The evolution of class 1 integrons and the rise of antibiotic resistance». *J Bacteriol.* vol. 190, no 14, p. 5095-5100.

- Gillings, M. R., M. Labbate, A. Sajjad, N. J. Giguere, M. P. Holley et H. W. Stokes. 2009. «Mobilization of a Tn402-like Class 1 integron with a novel cassette array via flanking MITE-like elements». *Appl Environ Microbiol.*
- Goldstein, C., M. D. Lee, S. Sanchez, C. Hudson, B. Phillips, B. Register, M. Grady, C. Liebert, A. O. Summers, D. G. White et J. J. Maurer. 2001. «Incidence of class 1 and 2 integrases in clinical and commensal bacteria from livestock, companion animals, and exotics». *Antimicrob Agents Chemother.* vol. 45, no 3, p. 723-726.
- Govantes, F., J. A. Albrecht et R. P. Gunsalus. 2000. «Oxygen regulation of the *Escherichia coli* cytochrome d oxidase (cydAB) operon: roles of multiple promoters and the Fnr-1 and Fnr-2 binding sites». *Mol Microbiol.* vol. 37, no 6, p. 1456-1469.
- Guerin, E., G. Cambray, N. Sanchez-Alberola, S. Campoy, I. Erill, S. Da Re, B. Gonzalez-Zorn, J. Barbe, M. C. Ploy et D. Mazel. 2009. «The SOS response controls integron recombination». *Science.* vol. 324, no 5930, p. 1034.
- Hall, R. M., D. E. Brookes et H. W. Stokes. 1991. «Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination cross-over point». *Mol Microbiol.* vol. 5, no 8, p. 1941-1959.
- Hanau-Bercot, B., I. Podglajen, I. Casin et E. Collatz. 2002. «An intrinsic control element for translational initiation in class 1 integrons». *Mol Microbiol.* vol. 44, no 1, p. 119-130.
- Hansson, K., Skold, O. & Sundstrom, L. 1997 «Non-palindromic attL sites of integrons are capable of site-specific recombination with one another and with secondary targets». *Mol. Microbiol.* vol. 26, no 3, p. 441-453.
- Hansson, K., L. Sundstrom, A. Pelletier et P. H. Roy. 2002. «IntI2 integron integrase in Tn7». *J Bacteriol.* vol. 184, no 6, p. 1712-1721.
- Haugen, S. P., W. Ross, M. Manrique et R. L. Gourse. 2008. «Fine structure of the promoter-sigma region 1.2 interaction». *Proc Natl Acad Sci U S A.* vol. 105, no 9, p. 3292-3297.
- Hausler, B., et R. L. Somerville. 1979. «Interaction in vivo between strong closely spaced constitutive promoters». *J Mol Biol.* vol. 127, no 3, p. 353-356.
- Hawley, D. K., et W. R. McClure. 1983. «Compilation and analysis of *Escherichia coli* promoter DNA sequences». *Nucleic Acids Res.* vol. 11, no 8, p. 2237-2255.
- Helmann, J. D. 2002. «The extracytoplasmic function (ECF) sigma factors». *Adv Microb Physiol.* vol. 46, p. 47-110.
- Hochhut, B., Y. Lotfi, D. Mazel, S. M. Faruque, R. Woodgate et M. K. Waldor. 2001. «Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constains». *Antimicrob Agents Chemother.* vol. 45, no 11, p. 2991-3000.

- Holmes, A. J., M. R. Gillings, B. S. Nield, B. C. Mabbutt, K. M. Nevalainen et H. W. Stokes. 2003. «The gene cassette metagenome is a basic resource for bacterial genome evolution». *Environ Microbiol.* vol. 5, no 5, p. 383-394.
- Holmes, A. J., M. P. Holley, A. Mahon, B. Nield, M. Gillings et H. W. Stokes. 2003. «Recombination activity of a distinctive integron-gene cassette system associated with *Pseudomonas stutzeri* populations in soil». *J Bacteriol.* vol. 185, no 3, p. 918-928.
- Hook-Barnard, I. G., et D. M. Hinton. 2009. «The promoter spacer influences transcription initiation via sigma70 region 1.1 of *Escherichia coli* RNA polymerase». *Proc Natl Acad Sci U S A.* vol. 106, no 3, p. 737-742.
- Hook-Barnard, I., X. B. Johnson et D. M. Hinton. 2006. «*Escherichia coli* RNA polymerase recognition of a sigma70-dependent promoter requiring a -35 DNA element and an extended -10 TGn motif». *J Bacteriol.* vol. 188, no 24, p. 8352-8359.
- Houang, E. T., Y. W. Chu, W. S. Lo, K. Y. Chu et A. F. Cheng. 2003. «Epidemiology of rifampin ADP-ribosyltransferase (arr-2) and metallo-beta-lactamase (blaIMP-4) gene cassettes in class 1 integrons in *Acinetobacter* strains isolated from blood cultures in 1997 to 2000». *Antimicrob Agents Chemother.* vol. 47, no 4, p. 1382-1390.
- Ikenaka, K., G. Ramakrishnan, M. Inouye et K. Tsung. 1986. «Regulation of the ompC gene of *Escherichia coli*. Involvement of three tandem promoters». *J Biol Chem.* vol. 261, no 20, p. 9316-9320.
- Iversen, J., D. Sandvang, A. Srijan, P. D. Cam et A. Dalsgaard. 2003. «Characterization of antimicrobial resistance, plasmids, and gene cassettes in *Shigella* spp. from patients in vietnam». *Microb Drug Resist.* vol. 9 Suppl 1, p. S17-24.
- Jacobs, L., et H. Y. Chenia. 2007. «Characterization of integrons and tetracycline resistance determinants in *Aeromonas* spp. isolated from South African aquaculture systems». *Int J Food Microbiol.* vol. 114, no 3, p. 295-306.
- Jacquier, H., C. Zaoui, M. J. Sanson-le Pors, D. Mazel et B. Bercot. 2009. «Translation regulation of integrons gene cassette expression by the attC sites». *Mol Microbiol.* vol. 72, no 6, p. 1475-1486.
- Jaurin, B., et S. Normark. 1983. «Insertion of IS2 creates a novel ampC promoter in *Escherichia coli*». *Cell.* vol. 32, no 3, p. 809-816.
- Jové, T., S. Da Re, F. Denis, D. Mazel et M. C. Ploy. 2010. «Inverse correlation between promoter strength and excision activity in class 1 integrons». *PLoS Genet.* vol 6, no 1
- Kagiya, G., R. Ogawa, M. Hatashita, K. Takagi, T. Kodaki, S. Hiroishi et K. Yamamoto. 2005. «Generation of a strong promoter for *Escherichia coli* from eukaryotic genome DNA». *J Biotechnol.* vol. 115, no 3, p. 239-248.
- Kishan, V., et W. Hillen. 1990. «Molecular cloning, nucleotide sequence, and promoter structure of the *Acinetobacter calcoaceticus* trpFB operon». *J Bacteriol.* vol. 172, no 10, p. 6151-6155.

- Klocko, A. D., et K. M. Wassarman. 2009. «6S RNA binding to Esigma(70) requires a positively charged surface of sigma(70) region 4.2». *Mol Microbiol.* vol. 73, no 2, p. 152-164.
- Kong, K. F., S. R. Jayawardena, S. D. Indulkar, A. Del Puerto, C. L. Koh, N. Hoiby et K. Mathee. 2005. «*Pseudomonas aeruginosa* AmpR is a global transcriptional factor that regulates expression of AmpC and PoxB beta-lactamases, proteases, quorum sensing, and other virulence factors». *Antimicrob Agents Chemother.* vol. 49, no 11, p. 4567-4575.
- Kubori, T., et N. Shimamoto. 1997. «Physical interference between *Escherichia coli* RNA polymerase molecules transcribing in tandem enhances abortive synthesis and misincorporation». *Nucleic Acids Res.* vol. 25, no 13, p. 2640-2647.
- Kulaeva, O. I., E. V. Koonin, J. C. Wootton, A. S. Levine et R. Woodgate. 1998. «Unusual insertion element polymorphisms in the promoter and terminator regions of the *mucAB*-like genes of R471a and R446b». *Mutat Res.* vol. 397, no 2, p. 247-262.
- Kumar, A., R. A. Malloch, N. Fujita, D. A. Smillie, A. Ishihama et R. S. Hayward. 1993. «The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter». *J Mol Biol.* vol. 232, no 2, p. 406-418.
- Lagatolla, C., E. Edalucci, L. Dolzani, M. L. Riccio, F. De Luca, E. Medessi, G. M. Rossolini et E. A. Tonin. 2006. «Molecular evolution of metallo-beta-lactamase-producing *Pseudomonas aeruginosa* in a nosocomial setting of high-level endemicity». *J Clin Microbiol.* vol. 44, no 7, p. 2348-2353.
- Laraki, N., M. Galleni, I. Thamm, M. L. Riccio, G. Amicosante, J. M. Frere et G. M. Rossolini. 1999. «Structure of In31, a *bla*IMP-containing *Pseudomonas aeruginosa* integron phylogenically related to In5, which carries an unusual array of gene cassettes». *Antimicrob Agents Chemother.* vol. 43, no 4, p. 890-901.
- Lartigue, M. F., V. Leflon-Guibout, L. Poirel, P. Nordmann et M. H. Nicolas-Chanoine. 2002. «Promoters P3, Pa/Pb, P4, and P5 upstream from *bla*(TEM) genes and their relationship to beta-lactam resistance». *Antimicrob Agents Chemother.* vol. 46, no 12, p. 4035-4037.
- Leon, G., et P. H. Roy. 2003. «Excision and integration of cassettes by an integron integrase of *Nitrosomonas europaea*». *J Bacteriol.* vol. 185, no 6, p. 2036-2041.
- Leon, G., et P. H. Roy. 2009. «Potential role of group IIC-attC introns in integron cassette formation». *J Bacteriol.* vol. 191, no 19, p. 6040-6051.
- Levesque, C., S. Brassard, J. Lapointe et P. H. Roy. 1994. «Diversity and relative strength of tandem promoters for the antibiotic-resistance genes of several integrons». *Gene.* vol. 142, no 1, p. 49-54.

- Levesque, C., L. Piche, C. Larose et P. H. Roy. 1995. «PCR mapping of integrons reveals several novel combinations of resistance genes». *Antimicrob Agents Chemother.* vol. 39, no 1, p. 185-191.
- Lindstedt, B. A., E. Heir, I. Nygard et G. Kapperud. 2003. «Characterization of class I integrons in clinical strains of *Salmonella enterica* subsp. *enterica* serovars Typhimurium and Enteritidis from Norwegian hospitals». *J Med Microbiol.* vol. 52, no Pt 2, p. 141-149.
- Llamas, M. A., A. van der Sar, B. C. Chu, M. Sparrius, H. J. Vogel et W. Bitter. 2009. «A Novel extracytoplasmic function (ECF) sigma factor regulates virulence in *Pseudomonas aeruginosa*». *PLoS Pathog.* vol. 5, no 9, p. e1000572.
- Llanes, C., C. Neuwirth, F. El Garch, D. Hocquet et P. Plesiat. 2006. «Genetic analysis of a multiresistant strain of *Pseudomonas aeruginosa* producing PER-1 beta-lactamase». *Clin Microbiol Infect.* vol. 12, no 3, p. 270-278.
- Lopez-Sanchez, A., E. Rivas-Marin, O. Martinez-Perez, B. Floriano et E. Santero. 2009. «Coordinated regulation of two divergent promoters through higher-order complex formation by the LysR-type regulator ThnR». *Mol Microbiol.* vol. 73, no 6, p. 1086-1100.
- MacDonald, D., G. Demarre, M. Bouvier, D. Mazel et D. N. Gopaul. 2006. «Structural basis for broad DNA-specificity in integron recombination». *Nature.* vol. 440, no 7088, p. 1157-1162.
- Mahillon, J., et M. Chandler. 1998. «Insertion sequences». *Microbiol Mol Biol Rev.* vol. 62, no 3, p. 725-774.
- Marquez, C., M. Labbate, A. J. Ingold, P. R. Chowdhury, M. S. Ramirez, D. Centron, G. Borthagaray et H. W. Stokes. 2008. «Recovery of a functional class 2 integron from an *Escherichia coli* strain mediating a urinary tract infection». *Antimicrob Agents Chemother.* vol. 52, no 11, p. 4153-4154.
- Martinez, E., et F. de la Cruz. 1990. «Genetic elements involved in Tn21 site-specific integration, a novel mechanism for the dissemination of antibiotic resistance genes». *EMBO J.* vol. 9, no 4, p. 1275-1281.
- Mazel, D. 2006. «Integrons: agents of bacterial evolution». *Nat Rev Microbiol.* vol. 4, no 8, p. 608-620.
- Mazel, D., B. Dychinco, V. A. Webb et J. Davies. 1998. «A distinctive class of integron in the *Vibrio cholerae* genome». *Science.* vol. 280, no 5363, p. 605-608.
- Miller, J. H. 1992. *A short course in bacterial genetics: a laboratory manual and handbook for Escherichia coli and related bacteria.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press, .
- Minakhin, L., et K. Severinov. 2003. «On the role of the *Escherichia coli* RNA polymerase sigma 70 region 4.2 and alpha-subunit C-terminal domains in promoter complex

- formation on the extended -10 galP1 promoter». *J Biol Chem.* vol. 278, no 32, p. 29710-29718.
- Mitchell, J. E., D. Zheng, S. J. Busby et S. D. Minchin. 2003. «Identification and analysis of 'extended -10' promoters in *Escherichia coli*». *Nucleic Acids Res.* vol. 31, no 16, p. 4689-4695.
- Moura, A., M. Soares, C. Pereira, N. Leitao, I. Henriques et A. Correia. 2009. «INTEGRALL: a database and search engine for integrons, integrases and gene cassettes». *Bioinformatics.* vol. 25, no 8, p. 1096-1098.
- Moyle, H., C. Waldburger et M. M. Susskind. 1991. «Hierarchies of base pair preferences in the P22 ant promoter». *J Bacteriol.* vol. 173, no 6, p. 1944-1950.
- Murakami, K. S., S. Masuda, E. A. Campbell, O. Muzzin et S. A. Darst. 2002. «Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex». *Science.* vol. 296, no 5571, p. 1285-1290.
- Naas, T., Y. Mikami, T. Imai, L. Poirel et P. Nordmann. 2001. «Characterization of In53, a class 1 plasmid- and composite transposon-located integron of *Escherichia coli* which carries an unusual array of gene cassettes». *J Bacteriol.* vol. 183, no 1, p. 235-249.
- Najafi, S. M., A. C. Willis et M. D. Yudkin. 1995. «Site of phosphorylation of SpoIIAA, the anti-anti-sigma factor for sporulation-specific sigma F of *Bacillus subtilis*». *J Bacteriol.* vol. 177, no 10, p. 2912-2913.
- Nasser, W., M. Rochman et G. Muskhelishvili. 2002. «Transcriptional regulation of *fis* operon involves a module of multiple coupled promoters». *EMBO J.* vol. 21, no 4, p. 715-724.
- Nesvera, J., J. Hochmannova et M. Patek. 1998. «An integron of class 1 is present on the plasmid pCG4 from gram-positive bacterium *Corynebacterium glutamicum*». *FEMS Microbiol Lett.* vol. 169, no 2, p. 391-395.
- Nickels, B. E., S. L. Dove, K. S. Murakami, S. A. Darst et A. Hochschild. 2002. «Protein-protein and protein-DNA interactions of sigma70 region 4 involved in transcription activation by lambda cI ». *J Mol Biol.* vol. 324, no 1, p. 17-34.
- Nield, B. S., A. J. Holmes, M. R. Gillings, G. D. Recchia, B. C. Mabbutt, K. M. Nevalainen et H. W. Stokes. 2001. «Recovery of new integron classes from environmental DNA». *FEMS Microbiol Lett.* vol. 195, no 1, p. 59-65.
- Nunes-Duby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger et A. Landy. 1998. «Similarities and differences among 105 members of the Int family of site-specific recombinases». *Nucleic Acids Res.* vol. 26, no 2, p. 391-406.
- Nystrom, T. 2004. «Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition?». *Mol Microbiol.* vol. 54, no 4, p. 855-862.

- Paget, M. S., et J. D. Helmann. 2003. «The sigma70 family of sigma factors». *Genome Biol.* vol. 4, no 1, p. 203.
- Palmer, A. C., A. Ahlgren-Berg, J. B. Egan, I. B. Dodd et K. E. Shearwin. 2009. «Potent transcriptional interference by pausing of RNA polymerases over a downstream promoter». *Mol Cell.* vol. 34, no 5, p. 545-555.
- Papagiannitsis, C. C., L. S. Tzouvelekis et V. Miriagou. 2009. «Relative strengths of the class 1 integron promoter hybrid 2 and the combinations of strong and hybrid 1 with an active p2 promoter». *Antimicrob Agents Chemother.* vol. 53, no 1, p. 277-280.
- Parent, R., et P. H. Roy. 1992. «The chloramphenicol acetyltransferase gene of Tn2424: a new breed of cat». *J Bacteriol.* vol. 174, no 9, p. 2891-2897.
- Partridge, S. R., C. M. Collis et R. M. Hall. 2002. «Class 1 integron containing a new gene cassette, aadA10, associated with Tn1404 from R151». *Antimicrob Agents Chemother.* vol. 46, no 8, p. 2400-2408.
- Partridge, S. R., G. D. Recchia, C. Scaramuzzi, C. M. Collis, H. W. Stokes et R. M. Hall. 2000. «Definition of the attI1 site of class 1 integrons». *Microbiology.* vol. 146 (Pt 11), p. 2855-2864.
- Partridge, S. R., G. Tsafnat, E. Coiera et J. R. Iredell. 2009. «Gene cassettes and cassette arrays in mobile resistance integrons». *FEMS Microbiol Rev.* vol. 33, no 4, p. 757-784.
- Perez-Rueda, E., et J. Collado-Vides. 2000. «The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12». *Nucleic Acids Res.* vol. 28, no 8, p. 1838-1847.
- Peters, E. D., M. A. Leverstein-van Hall, A. T. Box, J. Verhoef et A. C. Fluit. 2001. «Novel gene cassettes and integrons». *Antimicrob Agents Chemother.* vol. 45, no 10, p. 2961-2964.
- Ploy, M. C., P. Courvalin et T. Lambert. 1998. «Characterization of In40 of *Enterobacter aerogenes* BM2688, a class 1 integron with two new gene cassettes, cmlA2 and qacF». *Antimicrob Agents Chemother.* vol. 42, no 10, p. 2557-2563.
- Ploy, M. C., F. Denis, P. Courvalin et T. Lambert. 2000. «Molecular characterization of integrons in *Acinetobacter baumannii*: description of a hybrid class 2 integron». *Antimicrob Agents Chemother.* vol. 44, no 10, p. 2684-2688.
- Poirel, L., I. Le Thomas, T. Naas, A. Karim et P. Nordmann. 2000. «Biochemical sequence analyses of GES-1, a novel class A extended-spectrum beta-lactamase, and the class 1 integron In52 from *Klebsiella pneumoniae*». *Antimicrob Agents Chemother.* vol. 44, no 3, p. 622-632.
- Post, V., et R. M. Hall. 2009. «Insertion sequences in the IS1111 family that target the attC recombination sites of integron-associated gene cassettes». *FEMS Microbiol Lett.* vol. 290, no 2, p. 182-187.

- Potvin, E., F. Sanschagrín et R. C. Levesque. 2008. «Sigma factors in *Pseudomonas aeruginosa*». *FEMS Microbiol Rev.* vol. 32, no 1, p. 38-55.
- Pournaras, S., A. Ikonomidis, L. S. Tzouveleki, D. Tokatlidou, N. Spanakis, A. N. Maniatis, N. J. Legakis et A. Tsakris. 2005. «VIM-12, a novel plasmid-mediated metallo-beta-lactamase from *Klebsiella pneumoniae* that resembles a VIM-1/VIM-2 hybrid». *Antimicrob Agents Chemother.* vol. 49, no 12, p. 5153-5156.
- Power, P., M. Galleni, J. Di Conza, J. A. Ayala et G. Gutkind. 2005. «Description of In116, the first blaCTX-M-2-containing complex class 1 integron found in *Morganella morganii* isolates from Buenos Aires, Argentina». *J Antimicrob Chemother.* vol. 55, no 4, p. 461-465.
- Quinones, M., H. H. Kimsey, W. Ross, R. L. Gourse et M. K. Waldor. 2006. «LexA represses CTXphi transcription by blocking access of the alpha C-terminal domain of RNA polymerase to promoter DNA». *J Biol Chem.* vol. 281, no 51, p. 39407-39412.
- Ramirez, M. S., T. R. Parenteau, D. Centron et M. E. Tolmasky. 2008. «Functional characterization of Tn1331 gene cassettes». *J Antimicrob Chemother.* vol. 62, no 4, p. 669-673.
- Ramirez, M. S., S. Pineiro et D. Centron. 2009. «Novel insights about class 2 integrons from experimental and genomic epidemiology». *Antimicrob Agents Chemother.*
- Ramirez, M. S., C. Quiroga et D. Centron. 2005. «Novel rearrangement of a class 2 integron in two non-epidemiologically related isolates of *Acinetobacter baumannii*». *Antimicrob Agents Chemother.* vol. 49, no 12, p. 5179-5181.
- Ramirez, M. S., L. J. Vargas, V. Cagnoni, M. Tokumoto et D. Centron. 2005. «Class 2 integron with a novel cassette array in a *Burkholderia cenocepacia* isolate». *Antimicrob Agents Chemother.* vol. 49, no 10, p. 4418-4420.
- Ramos, J. L., S. Marques et K. N. Timmis. 1997. «Transcriptional control of the *Pseudomonas* TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators». *Annu Rev Microbiol.* vol. 51, p. 341-373.
- Rather, P. N., P. A. Mann, R. Mierzwa, R. S. Hare, G. H. Miller et K. J. Shaw. 1993. «Analysis of the aac(3)-VIa gene encoding a novel 3-N-acetyltransferase». *Antimicrob Agents Chemother.* vol. 37, no 10, p. 2074-2079.
- Rauzier, J., E. Gormley, M. C. Gutierrez, E. Kassa-Kelembho, L. J. Sandall, C. Dupont, B. Gicquel et A. Murray. 1999. «A novel polymorphic genetic locus in members of the *Mycobacterium tuberculosis* complex». *Microbiology.* vol. 145 (Pt 7), p. 1695-1701.
- Recchia, G. D., et R. M. Hall. 1995. «Gene cassettes: a new class of mobile element». *Microbiology.* vol. 141 (Pt 12), p. 3015-3027.
- Recchia, G. D. et R.M Hall. 1995. «Plasmid evolution by acquisition of mobile gene cassettes: plasmid pIE723 contains the aadB gene cassette precisely inserted at a secondary site in the incQ plasmid RSF1010». *Mol Microbiol* vol. 15, no1 179-187.

- Recchia, G. D., et R. M. Hall. 1997. «Origins of the mobile gene cassettes found in integrons». *Trends Microbiol.* vol. 5, no 10, p. 389-394.
- Riccio, M. L., N. Franceschini, L. Boschi, B. Caravelli, G. Cornaglia, R. Fontana, G. Amicosante et G. M. Rossolini. 2000. «Characterization of the metallo-beta-lactamase determinant of *Acinetobacter baumannii* AC-54/97 reveals the existence of bla(IMP) allelic variants carried by gene cassettes of different phylogeny». *Antimicrob Agents Chemother.* vol. 44, no 5, p. 1229-1235.
- Richardson, J. P. 2002. «Rho-dependent termination and ATPases in transcript termination». *Biochim Biophys Acta.* vol. 1577, no 2, p. 251-260.
- Rodriguez-Martinez, J. M., L. Poirel, R. Canton et P. Nordmann. 2006. «Common region CR1 for expression of antibiotic resistance genes». *Antimicrob Agents Chemother.* vol. 50, no 7, p. 2544-2546.
- Rodriguez-Minguela, C. M., J. H. Apajalahti, B. Chai, J. R. Cole et J. M. Tiedje. 2009. «Worldwide prevalence of class 2 integrases outside the clinical setting is associated with human impact». *Appl Environ Microbiol.* vol. 75, no 15, p. 5100-5110.
- Rosser, S. J., et H. K. Young. 1999. «Identification and characterization of class 1 integrons in bacteria from an aquatic environment». *J Antimicrob Chemother.* vol. 44, no 1, p. 11-18.
- Rowe-Magnus, D. A., A. M. Guerout, L. Biskri, P. Bouige et D. Mazel. 2003. «Comparative analysis of superintegrons: engineering extensive genetic diversity in the *Vibrionaceae*». *Genome Res.* vol. 13, no 3, p. 428-442.
- Rowe-Magnus, D. A., A. M. Guerout, P. Ploncard, B. Dychinco, J. Davies et D. Mazel. 2001. «The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons». *Proc Natl Acad Sci U S A.* vol. 98, no 2, p. 652-657.
- Roy, S., S. Semsey, M. Liu, G. N. Gussin et S. Adhya. 2004. «GalR represses galP1 by inhibiting the rate-determining open complex formation through RNA polymerase contact: a GalR negative control mutant». *J Mol Biol.* vol. 344, no 3, p. 609-618.
- Saenz, Y., L. Brinas, E. Dominguez, J. Ruiz, M. Zarazaga, J. Vila et C. Torres. 2004. «Mechanisms of resistance in multiple-antibiotic-resistant *Escherichia coli* strains of human, animal, and food origins». *Antimicrob Agents Chemother.* vol. 48, no 10, p. 3996-4001.
- Sallen, B., A. Rajoharison, S. Desvarenne et C. Mabilat. 1995. «Molecular epidemiology of integron-associated antibiotic resistance genes in clinical isolates of enterobacteriaceae». *Microb Drug Resist.* vol. 1, no 3, p. 195-202.
- Sambrook, J., T. Maniatis et E.F Fritsh. 1989. *Molecular Cloning : a laboratory manual*. New-York: Cold spring harbor laboratory press, cold spring harbor-

- Sanschagrín, F., N. Bejaoui et R. C. Levesque. 1998. «Structure of CARB-4 and AER-1 carbenicillin-hydrolyzing beta-lactamases». *Antimicrob Agents Chemother.* vol. 42, no 8, p. 1966-1972.
- Schmitz, F. J., D. Hafner, R. Geisel, P. Follmann, C. Kirschke, J. Verhoef, K. Kohrer et A. C. Fluit. 2001. «Increased prevalence of class I integrons in *Escherichia coli*, *Klebsiella* species, and *Enterobacter* species isolates over a 7-year period in a German university hospital». *J Clin Microbiol.* vol. 39, no 10, p. 3724-3726.
- Schroeder, L. A., A. J. Choi et P. L. DeHaseth. 2007. «The -11A of promoter DNA and two conserved amino acids in the melting region of sigma70 both directly affect the rate limiting step in formation of the stable RNA polymerase-promoter complex, but they do not necessarily interact». *Nucleic Acids Res.* vol. 35, no 12, p. 4141-4153.
- Segal, H., et B. G. Elisha. 1999. «Characterization of the *Acinetobacter* plasmid, pRAY, and the identification of regulatory sequences upstream of an *aadB* gene cassette on this plasmid». *Plasmid.* vol. 42, no 1, p. 60-66.
- Seward, R. J. 1999. «Detection of integrons in worldwide nosocomial isolates of *Acinetobacter* spp». *Clin Microbiol Infect.* vol. 5, no 6, p. 308-318.
- Shearer, J. E., et A. O. Summers. 2009. «Intracellular steady-state concentration of integron recombination products varies with integrase level and growth phase». *J Mol Biol.* vol. 386, no 2, p. 316-331.
- Shearwin, K. E., B. P. Callen et J. B. Egan. 2005. «Transcriptional interference--a crash course». *Trends Genet.* vol. 21, no 6, p. 339-345.
- Shibata, N., Y. Doi, K. Yamane, T. Yagi, H. Kurokawa, K. Shibayama, H. Kato, K. Kai et Y. Arakawa. 2003. «PCR typing of genetic determinants for metallo-beta-lactamases and integrases carried by gram-negative bacteria isolated in Japan, with focus on the class 3 integron». *J Clin Microbiol.* vol. 41, no 12, p. 5407-5413.
- Shultzaberger, R. K., Z. Chen, K. A. Lewis et T. D. Schneider. 2007. «Anatomy of *Escherichia coli* sigma70 promoters». *Nucleic Acids Res.* vol. 35, no 3, p. 771-788.
- Silvaggi, J. M., J. B. Perkins et R. Losick. 2005. «Small untranslated RNA antitoxin in *Bacillus subtilis*». *J Bacteriol.* vol. 187, no 19, p. 6641-6650.
- Simonsen, C. C., E. Y. Chen et A. D. Levinson. 1983. «Identification of the type I trimethoprim-resistant dihydrofolate reductase specified by the *Escherichia coli* R-plasmid R483: comparison with procaryotic and eucaryotic dihydrofolate reductases». *J Bacteriol.* vol. 155, no 3, p. 1001-1008.
- Sneppen, K., I. B. Dodd, K. E. Shearwin, A. C. Palmer, R. A. Schubert, B. P. Callen et J. B. Egan. 2005. «A mathematical model for transcriptional interference by RNA polymerase traffic in *Escherichia coli*». *J Mol Biol.* vol. 346, no 2, p. 399-409.
- Solomenny, A.P. 2009. «Class 1 integron-associated promoters for gene cassette transcription». *Genetika.* vol. 45, no 7, p. 1001-1003.

- Soufi, L., M. S. Abbassi, Y. Saenz, L. Vinue, S. Somalo, M. Zarazaga, A. Abbas, R. Dbaya, L. Khanfir, A. Ben Hassen, S. Hammami et C. Torres. 2009. «Prevalence and diversity of integrons and associated resistance genes in *Escherichia coli* isolates from poultry meat in Tunisia». *Foodborne Pathog Dis.* vol. 6, no 9, p. 1067-1073.
- Srivatsan, A., et J. D. Wang. 2008. «Control of bacterial transcription, translation and replication by (p)ppGpp». *Curr Opin Microbiol.* vol. 11, no 2, p. 100-105.
- Stokes, H. W., et R. M. Hall. 1989. «A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons». *Mol Microbiol.* vol. 3, no 12, p. 1669-1683.
- Stokes, H. W., A. J. Holmes, B. S. Nield, M. P. Holley, K. M. Nevalainen, B. C. Mabbutt et M. R. Gillings. 2001. «Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA». *Appl Environ Microbiol.* vol. 67, no 11, p. 5240-5246.
- Stokes, H. W., D. B. O'Gorman, G. D. Recchia, M. Parsekhian et R. M. Hall. 1997. «Structure and function of 59-base element recombination sites associated with mobile gene cassettes». *Mol Microbiol.* vol. 26, no 4, p. 731-745.
- Studemann, A., M. Noirclerc-Savoie, E. Klauck, G. Becker, D. Schneider et R. Hengge. 2003. «Sequential recognition of two distinct sites in sigma(S) by the proteolytic targeting factor RssB and ClpX». *EMBO J.* vol. 22, no 16, p. 4111-4120.
- Sunde, M. 2005. «Class I integron with a group II intron detected in an *Escherichia coli* strain from a free-range reindeer». *Antimicrob Agents Chemother.* vol. 49, no 6, p. 2512-2514.
- Sunde, M. 2005. «Prevalence and characterization of class 1 and class 2 integrons in *Escherichia coli* isolated from meat and meat products of Norwegian origin». *J Antimicrob Chemother.* vol. 56, no 6, p. 1019-1024.
- Sundstrom, L., P. H. Roy et O. Skold. 1991. «Site-specific insertion of three structural gene cassettes in transposon Tn7». *J Bacteriol.* vol. 173, no 9, p. 3025-3028.
- Szczepanowski, R., I. Krahn, A. Puhler et A. Schluter. 2004. «Different molecular rearrangements in the integron of the IncP-1 beta resistance plasmid pB10 isolated from a wastewater treatment plant result in elevated beta-lactam resistance levels». *Arch Microbiol.* vol. 182, no 6, p. 429-435.
- Tenover, F. C., D. Filpula, K. L. Phillips et J. J. Plorde. 1988. «Cloning and sequencing of a gene encoding an aminoglycoside 6'-N-acetyltransferase from an R factor of *Citrobacter diversus*». *J Bacteriol.* vol. 170, no 1, p. 471-473.
- Tietze, E., J. Brevet et H. Tschape. 1987. «Relationships among the streptothricin resistance transposons Tn1825 and Tn1826 and the trimethoprim resistance transposon Tn7». *Plasmid.* vol. 18, no 3, p. 246-249.

- Toleman, M. A., P. M. Bennett et T. R. Walsh. 2006. «ISCR elements: novel gene-capturing systems of the 21st century?». *Microbiol Mol Biol Rev.* vol. 70, no 2, p. 296-316.
- Tramonti, A., M. De Canio et D. De Biase. 2008. «GadX/GadW-dependent regulation of the *Escherichia coli* acid fitness island: transcriptional control at the *gadY-gadW* divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites». *Mol Microbiol.* vol. 70, no 4, p. 965-982.
- Travers, A., et G. Muskhelishvili. 2005. «DNA supercoiling - a global transcriptional regulator for enterobacterial growth?». *Nat Rev Microbiol.* vol. 3, no 2, p. 157-169.
- Vali, L., A. Hamouda, D. V. Hoyle, M. C. Pearce, L. H. Whitaker, C. Jenkins, H. I. Knight, A. W. Smith et S. G. Amyes. 2007. «Antibiotic resistance and molecular epidemiology of *Escherichia coli* O26, O103 and O145 shed by two cohorts of Scottish beef cattle». *J Antimicrob Chemother.* vol. 59, no 3, p. 403-410.
- van Essen-Zandbergen, A., H. Smith, K. Veldman et D. Mevius. 2007. «Occurrence and characteristics of class 1, 2 and 3 integrons in *Escherichia coli*, *Salmonella* and *Campylobacter* spp. in the Netherlands». *J Antimicrob Chemother.* vol. 59, no 4, p. 746-750.
- Waldor, M. K., et D. I. Friedman. 2005. «Phage regulatory circuits and virulence gene expression». *Curr Opin Microbiol.* vol. 8, no 4, p. 459-465.
- Walsh, T. R., M. A. Toleman, W. Hryniewicz, P. M. Bennett et R. N. Jones. 2003. «Evolution of an integron carrying *bla*VIM-2 in Eastern Europe: report from the SENTRY Antimicrobial Surveillance Program». *J Antimicrob Chemother.* vol. 52, no 1, p. 116-119.
- Wassarman, K. M. 2007. «6S RNA: a regulator of transcription». *Mol Microbiol.* vol. 65, no 6, p. 1425-1431.
- White, P. A., C. J. McIver et W. D. Rawlinson. 2001. «Integrons and gene cassettes in the enterobacteriaceae». *Antimicrob Agents Chemother.* vol. 45, no 9, p. 2658-2661.
- Wigneshweraraj, S., D. Bose, P. C. Burrows, N. Joly, J. Schumacher, M. Rappas, T. Pape, X. Zhang, P. Stockley, K. Severinov et M. Buck. 2008. «Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor». *Mol Microbiol.* vol. 68, no 3, p. 538-546.
- Xu, H., J. Davies et V. Miao. 2007. «Molecular characterization of class 3 integrons from *Delftia* spp». *J Bacteriol.* vol. 189, no 17, p. 6276-6283.
- Xu, Z., L. Li, M. J. Alam, L. Zhang, S. Yamasaki et L. Shi. 2008. «First confirmation of integron-bearing methicillin-resistant *Staphylococcus aureus*». *Curr Microbiol.* vol. 57, no 3, p. 264-268.
- Yarnell, W. S., et J. W. Roberts. 1999. «Mechanism of intrinsic transcription termination and antitermination». *Science.* vol. 284, no 5414, p. 611-615.

- Yildiz, F. H., X. S. Liu, A. Heydorn et G. K. Schoolnik. 2004. «Molecular analysis of rugosity in a *Vibrio cholerae* O1 El Tor phase variant». *Mol Microbiol.* vol. 53, no 2, p. 497-515.
- Young, H. K., M. J. Qumsieh et M. L. McIntosh. 1994. «Nucleotide sequence and genetic analysis of the type Ib trimethoprim-resistant, Tn4132-encoded dihydrofolate reductase». *J Antimicrob Chemother.* vol. 34, no 5, p. 715-725.
- Zong, Z., S. R. Partridge et J. R. Iredell. 2009. «A blaVEB-1 variant, blaVEB-6, associated with repeated elements in a complex genetic structure». *Antimicrob Agents Chemother.* vol. 53, no 4, p. 1693-1697.

Annexes

Contribution à une base de données dédiée aux intégrons : INTEGRALL

La recherche de l'ensemble des séquences de IM 1 disponibles dans l'INSDB nous a conduit à constituer un fichier de 1351 séquences (au 30/10/09) actualisé quotidiennement.

Notre fichier expose, pour chaque IM 1 et quand elles sont disponibles, les informations suivantes :

- Variant de Pc
- Variant de P2
- Présence ou non de la duplication de 19 pb d'ORF11
- Nom de l'intégron
- Nom du support de l'intégron
- Organisme hôte (genre, espèce)
- Conservation de la région *attI1*
- Conservation d'IntI1 (par rapport à IntI1 d'In0)
- Réseau de cassettes de gène

De plus, selon nos critères, la substitution d'un acide aminé suffit à définir un allèle distinct et potentiellement à considérer un nouveau IM 1. Enfin, un fichier similaire a été créé pour les IM 2 et 3.

Pour valoriser les données recueillies, nous proposons de les partager en ligne via une collaboration avec une banque de données dédiée aux intégrons : INTEGRALL ((Moura *et al.*, 2009) accessible au lien <http://integrall.bio.ua.pt/>).

INTEGRALL est une banque de données en ligne, accessible gratuitement qui compte 5406 entrées (au 8/09/09) relatives aux intégrons. Chaque entrée fait l'objet d'une fiche dans laquelle est détaillée la composition de l'intégron, notamment les gènes du réseau de cassette.

L'utilisateur peut facilement à l'aide de mots clés y retrouver des intégrons via un outil de

recherche convivial. Le site est également doté d'un glossaire et d'un forum de discussion, lieu d'échange d'opinions et ouvert aux suggestions. Enfin, une partie explique les nomenclatures utilisées pour nommer les cassettes de gènes et numéroter les allèles.

La collaboration entre l'équipe d'INTEGRALL et la notre a commencé quand nous avons proposé des corrections (noms d'enzymes, nomenclatures à utiliser, etc.) et la suppression d'entrées redondantes (listées une fois avec le numéro RefSeq, une autre avec le numéro d'accès GenBank). Elle devrait se poursuivre avec le partage de nos données des fichiers décrits ci-dessus.

INTEGRALL: a database and search engine for integrons, integrases and gene cassettes.

Moura A, Soares M, Pereira C, Leitão N, Henriques I, Correia A.

CESAM & Department of Biology, University of Aveiro, Campus Universitário de Santiago, Aveiro, Portugal. amoura@ua.pt

Bioinformatics. 2009 Apr 15;25(8):1096-8. Epub 2009 Feb 19.

[doi:10.1093/bioinformatics/btp105](https://doi.org/10.1093/bioinformatics/btp105)

Abstract

INTEGRALL is a freely available, text-based search system developed with the aim of collecting and organizing information on integrons in a single database. The current release (1.2) contains more than 4800 integron sequences and provides a public genetic repository for sequence data and nomenclature, offering scientists an easy and interactive access to integron's DNA sequences, their molecular arrangements as well as their genetic contexts.