



**HAL**  
open science

# Caractérisation de structures cytoplasmiques pseudo-nucléaires en utilisant un système de visualisation de locus génomique, le système ANCHOR

Philippe Valenti

► **To cite this version:**

Philippe Valenti. Caractérisation de structures cytoplasmiques pseudo-nucléaires en utilisant un système de visualisation de locus génomique, le système ANCHOR. Génétique. Université Paul Sabatier - Toulouse III, 2022. Français. NNT : 2022TOU30283 . tel-04137690

**HAL Id: tel-04137690**

**<https://theses.hal.science/tel-04137690v1>**

Submitted on 22 Jun 2023

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

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



# THÈSE

**En vue de l'obtention du  
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE  
Délivré par l'Université Toulouse 3 - Paul Sabatier**

---

**Présentée et soutenue par  
Philippe VALENTI**

Le 15 décembre 2022

**Caractérisation de structures cytoplasmiques pseudo-nucléaires  
en utilisant un système de visualisation de locus génomique, le  
système ANCHOR**

---

Ecole doctorale : **BSB - Biologie, Santé, Biotechnologies**

Spécialité : **BIOLOGIE DU DEVELOPPEMENT**

Unité de recherche :

**MCD - Molecular, Cellular and Developmental Biology Unit**

Thèse dirigée par  
**Cedric POLESELLO**

Jury

Mme Caroline MEDIONI, Rapporteur  
M. Julien COLOMBANI, Rapporteur  
M. Frédéric PONTVIANNE, Rapporteur  
Mme Mounia LAGHA, Examinatrice  
Mme Kerstin BYSTRICKY, Examinatrice  
M. David CRIBBS, Examineur  
M. Cédric POLESELLO, Directeur de thèse  
Mme Catherine SOULA, Présidente



<b>Table des matières</b> .....	<b>3</b>
<b>Résumés</b> .....	<b>6</b>
Résumé de thèse en français .....	6
Résumé de thèse en anglais .....	7
Résumé grand public .....	8
<b>Abréviations</b> .....	<b>9</b>
<b>Avant-propos</b> .....	<b>10</b>
<b>INTRODUCTION</b> .....	<b>53</b>
<b>Chapitre I : Organisation du génome eucaryote</b> .....	<b>54</b>
<b>1- Le noyau</b> .....	<b>55</b>
<b>2- Les chromosomes</b> .....	<b>56</b>
2-1 Les différentes couleurs de la chromatine .....	59
2-2 Les différentes marques d'histones .....	62
2-2-1 L'acétylation des histones .....	62
2-2-2 La méthylation des histones .....	63
2-3 Les éléments de cis-régulation .....	64
2-3-1 Les enhanceurs .....	64
2-3-2 Les insulateurs .....	65
<b>3- Shavenbaby, un acteur déterminant de la formation des trichomes dans l'épiderme</b> .....	<b>67</b>
3-1 Organisation moléculaire du locus ovo/svb .....	68
3-2 Svb gouverne la différenciation des trichomes dans l'épiderme .....	69
3-3 Régulation de l'expression du gène <i>svb</i> .....	70
3-4 Svb coordonne un réseau génique directement impliqué dans la formation des trichomes .....	72
3-5 Les smORFs peptides Pri : de nouveaux régulateurs de la morphogenèse épidermique .....	74
3-6 Régulation temporelle de la différenciation épidermique .....	76
<b>Chapitre II : Visualisation de l'ADN</b> .....	<b>79</b>
<b>1- Dans des cellules fixées</b> .....	<b>80</b>
1-1 Visualisation de l'ADN en microscopie optique .....	80
1-2 Visualisation de l'ADN par rayonnement .....	84
1-3 Visualisation de l'ADN par DAPI .....	84
1-4 Visualisation de l'ADN par hybridation fluorescente <i>in situ</i> (FISH) .....	85
<b>2- Visualisation de loci génomiques uniques et ARNm en cellules vivantes</b> .....	<b>89</b>
2-1 La Green Fluorescent Protein ou GFP .....	89

2-2 Les Systèmes Fluorescents Répresseur/Opérateur : FROS .....	91
2-3 Autres .....	93
<b>3- Organisation et Ségrégation des génomes bactériens .....</b>	<b>96</b>
3-1 Organisation des génomes bactériens .....	97
3-1-1 Organisation en chromosomes et plasmides .....	97
3-1-2 Organisation du chromosome bactérien .....	100
3-1-3 Les protéines associées au nucléoïde .....	106
3-1-4 Le super-enroulement négatif .....	110
3-2 Ségrégation des génomes bactériens .....	111
3-2-1 Ségrégation par diffusion passive .....	113
3-2-2 Ségrégation active .....	113
3-2-2-1 Ségrégation des origines de réplication .....	116
3-2-2-1a Les systèmes de partition de type II et III .....	118
3-2-2-1b Les systèmes de partition de type I : le système ParABS .....	124
3-2-2-2 Ségrégation de l'ensemble du chromosome .....	145
3-2-2-3 Terminaison de la ségrégation .....	147
<b>4- Le système ANCHOR .....</b>	<b>148</b>
4-1 Principe et applications du système de visualisation de locus génomique en temps réel : le système ANCHOR .....	148
4-2 Le système ANCHOR et les autres systèmes permettant de visualiser des loci génomiques en temps réel .....	156
<b>RESULTATS .....</b>	<b>160</b>
<b>Chapitre III : Visualisation de loci génomiques en temps réel dans des tissus d'un organisme entier, la drosophile .....</b>	<b>161</b>
<b>1- Les différents systèmes ANCHOR .....</b>	<b>164</b>
<b>2- Le système ANCHOR est-il toxique pour le développement de la drosophile ? .....</b>	<b>166</b>
<b>3- Le système ANCHOR est-il fonctionnel ? .....</b>	<b>169</b>
3-1 Sélection des pilotes permettant l'expression de la protéine ParB dans les glandes salivaires .....	169
3-2 Un site <i>parS</i> est suffisant pour générer un foyer fluorescent dans le noyau .....	170
3-3 Spécificité de fixation de la protéine ParB2 .....	172
3-4 Visualisation par la protéine ParB2 de deux loci génomiques différents .....	174
3-5 Visualisation de deux loci séparés par dix kilobases .....	175
<b>4- Le système ANCHOR dans l'épiderme embryonnaire et dans d'autres tissus de la drosophile .....</b>	<b>177</b>

4-1 Chez l'adulte, dans des cellules polyploïdes .....	177
4-2 Dans des cellules diploïdes .....	179
4-2-1 Dans les disques imaginaux .....	179
4-2-2 Dans l'épiderme embryonnaire de la drosophile .....	181
4-3 La fixation des protéines ParB sur leurs séquences <i>parS</i> modifie-t-elle l'activité de régions régulatrices ? .....	184
<b>5- Discussions et Perspectives .....</b>	<b>188</b>
<b>Chapitre IV : Caractérisation de structures cytoplasmiques pseudo-nucléaires</b>	
<b>    dans les cellules S2 .....</b>	<b>193</b>
<b>    1- Développement du système ANCHOR en cellules S2 .....</b>	<b>194</b>
1-1 Distribution intracellulaire des protéines ParB2 en présence des séquences <i>parS</i> .....	194
1-2 Effet du système ANCHOR sur l'activité transcriptionnelle .....	195
1.3 Effet du système ANCHOR sur les séquences insulatrices .....	199
<b>    2- Structures cytoplasmiques pseudo-nucléaires en cellules S2 .....</b>	<b>203</b>
2.1 Identification des structures pseudo-nucléaires cytoplasmiques .....	203
2.2 Le système ANCHOR est-il responsable de la formation des structures pseudo-nucléaires ? .....	205
2.3 Caractérisation des structures pseudo-nucléaires .....	209
2-4 Peut-il avoir une activité transcriptionnelle au niveau de ces structures cytoplasmiques pseudo-nucléaires ? .....	213
<b>    3- Discussions et perspectives .....</b>	<b>216</b>
3-1 Mécanismes d'import des plasmides dans le noyau .....	216
3-2 Formation des structures pseudo-nucléaires .....	218
3-3 Le système ANCHOR, Shavenbaby et accessibilité chromatinienne .....	219
<b>Annexes .....</b>	<b>223</b>
<b>Annexe 1 : Liste des plasmides construits .....</b>	<b>224</b>
<b>Annexe 2 : Liste des lignées transgéniques construites .....</b>	<b>225</b>
<b>Annexe 3 : Matériels et méthodes .....</b>	<b>226</b>
<b>Annexe4 : Liste des publications obtenues pendant la thèse .....</b>	<b>236</b>
<b>Références bibliographiques .....</b>	<b>237</b>

# Résumés

## Résumé de la thèse en français

La structure chromatinienne possède une organisation dynamique qui joue un rôle clé dans le contrôle de la destinée cellulaire et du développement. L'accessibilité de la chromatine est déterminante pour permettre la fixation des facteurs de transcription à l'ADN, contrôler l'expression des gènes et mettre en œuvre des programmes transcriptionnels spécifiques. Pour étudier la dynamique de la chromatine en temps réel et compléter les analyses structurales à l'échelle du génome (HiC ...), différents systèmes de localisation de loci génomiques ont été développés. Ils sont basés sur la visualisation, en microscopie photonique, de protéines de liaison à l'ADN fusionnées à des molécules fluorescentes mais ils peuvent perturber l'environnement chromatinien ou l'expression génique.

A partir du complexe de partition des génomes bactériens, un système de visualisation de loci génomiques en temps réel, le système ANCHOR, a été développé et a la particularité et l'avantage de permettre la visualisation de loci génomiques de manière peu intrusive. Il est constitué des protéines procaryotes ParB qui se lient à une séquence spécifique de 16 pb, *parS*. Une fois liée à la séquence *parS*, la protéine ParB fusionnée à une étiquette fluorescente se répand sur l'ADN environnant, créant un amas de molécules suffisamment grand pour être observé par microscopie à fluorescence.

Au cours de ma thèse, j'ai mis en place le système ANCHOR chez la drosophile. Après avoir montré que l'expression ectopique de la protéine ParB n'affecte pas la viabilité des cellules ou de l'organisme, j'ai démontré que dans un organisme eucaryote supérieur, le système ANCHOR permet la visualisation de loci génomiques spécifiques. De plus, l'utilisation du système ANCHOR n'interfère pas avec l'activité des éléments cis-régulateurs situés à proximité des sites d'insertion de *parS*, ni avec leur activité transcriptionnelle. Lors de transfections transitoires dans des cellules S2 de drosophile, la puissance de ce système nous a permis de suivre la localisation subcellulaire des vecteurs possédant le système ParB/*parS* dans le noyau et plus étonnamment dans le cytoplasme. J'ai montré que les plasmides présents dans le cytoplasme semblent adopter une structure pseudo-nucléaire : présence d'éléments de la membrane nucléaire (lamine), d'histones et d'activité transcriptionnelle.

En conclusion, mon travail de thèse m'a permis de mettre en place le système ANCHOR dans un organisme multicellulaire, la drosophile, et de confirmer l'efficacité de ce système pour visualiser en temps réel des loci génomiques, ouvrant ainsi de nouvelles perspectives de recherche.

## Résumé de thèse en anglais

Chromatin structure has a dynamic organization that plays a key role in the control of cell fate and development. Chromatin accessibility is critical for the binding of transcription factors to DNA, control of gene expression and the implementation of specific transcriptional programs. To study chromatin dynamics in real time and to complement genome-wide structural analyses (HiC...), different genomic loci localization systems have been developed. They are based on the visualization of DNA-binding proteins fused to fluorescent molecules under light microscopy, but they can disrupt the chromatin environment or gene expression.

Based on the bacterial genome partition complex, a real-time genomic loci visualization system, the ANCHOR system, has been developed and has the particularity and advantage of allowing the visualization of genomic loci in a non-intrusive way. It consists of prokaryotic ParB proteins that bind to a specific 16 bp sequence, *parS*. Once bound to the *parS* sequence, the ParB protein fused to a fluorescent tag spreads over the surrounding DNA, creating a cluster of molecules large enough to be observed by fluorescence microscopy.

During my thesis, I set up the ANCHOR system in *Drosophila*. After showing that ectopic expression of the ParB protein does not affect cell or organism viability, I demonstrated that in a higher eukaryotic organism, the ANCHOR system allows the visualization of specific genomic loci. Furthermore, the use of the ANCHOR system does not interfere with the activity of cis-regulatory elements located near *parS* insertion sites, nor with their transcriptional activity. In transient transfections in *Drosophila* S2 cells, the power of this system allowed us to follow the subcellular localization of vectors possessing the ParB/*parS* system in the nucleus and more surprisingly in the cytoplasm. I showed that plasmids present in the cytoplasm seem to adopt a pseudo-nuclear structure: presence of nuclear membrane elements (laminin), histones and transcriptional activity.

In conclusion, my thesis work allowed me to set up the ANCHOR system in a multicellular organism, *Drosophila*, and to confirm the efficiency of this system to visualize genomic loci in real time, thus opening new research perspectives.



## **Résumé grand public.**

Afin d'étudier la dynamique chromatinienne, support de l'information génétique, il est nécessaire de disposer d'outils permettant de suivre des régions génomiques. Le système ANCHOR, dérivé des bactéries, est constitué de protéines fusionnées à une molécule fluorescente qui se fixent, en grand nombre sur des séquences nucléotidiques spécifiques introduites à un locus donné. Cela permet la formation d'un foyer de fluorescence nucléaire, observable en microscopie.

Durant ma thèse, j'ai développé le système ANCHOR chez la drosophile et montré qu'il est parfaitement compatible avec la viabilité et la régulation transcriptionnelle. La puissance du système ANCHOR m'a permis de suivre le comportement d'ADNs plasmidiques, que l'on transfecte dans des cellules en culture. Ils sont présents sous forme d'agrégats dans le noyau et aussi dans le cytoplasme et, étonnamment, s'organisent en structure pseudo-nucléaires.

Mes travaux ont permis de montrer l'efficacité du système ANCHOR pour visualiser des loci génomiques au sein des organismes eucaryotes supérieurs, ouvrant ainsi de nouvelles perspectives de recherche dans ces organismes.

## Abréviations

A	adénine	HAT	histone acétyl transférase
Å	Angström	HCL	acide chlorydrique
aa	acide aminé	HCMV	cytomégalo virus humain
Ac	groupe acétylé	HDAC	histone déacétylase
ADN	acide désoxyribonucléotide	HDR	High Density chromosomal Region
ADNr	acide désoxyribonucléotide ribosomique	HiC	Capture de conformation chromosomique à haute résolution
AMP	adénosine monophosphate	HKMT	lysine méthyltransférases
Ala	alanine	H-NS	Histone-like Nucleoid-Structuring protein
ARN	acide ribonucléotide	HP1	Heterochromatin protein 1
ARNm	acide ribonucléotide messenger	HR	recombinaison homologue
ARNt	acide ribonucléotide de transfert	HTH	hélice-tour-hélice
ATP	adénosine triphosphate	H-U	Heat-Unstable protein
βGal	béta-galactosidase	IHF	integration Host Factor
bp	paire de base	ISH	hybridation in situ
BSA	bovine serum albumin	K	lysine
C	cytosine	kb	kilo bae
CBP	centromere binding protein	kDa	kilo dalton
CDP	cytidine diphosphate	lncRNA	longs ARN non codants
CDS	séquence codante	Mb	megabase
CIDs	Domaines d'Interaction Chromosomiques	MCP	MS2 coat protein
ChiP-seq	Chromatine Immuno-Précipitation suivi d'un séquençage	MRG15	MORF-related gene 15
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats	MSD	Mean Square Displacement
CRM	région de regulation en cis	NAP	nucleopid associated protein
CTD	domaine C-terminal	NLS	nucleair localization signal
CTP	cytidine triphosphate	nsDNA	ADN non spécifique
D1	D1 chromosomal protein	NTD	domaine N-terminal
DAPI	4'-6-Diamidino-2-Phenylindole	PALM	Photo-Activated Localization Microscopy
dATP	désoxyadénosine triphosphate	PcG	groupe Polycomb
dCTP	désoxycytidine triphosphate	PCR	polymerase chain reaction
Df31	decondensation factor 31	PRMT	arginine methyltransferases
dGTP	désoxyguanosine triphosphate	pri	plished-rice
dTTP	désoxythymidine triphosphate	ptc	patched
ddATP	di-désoxyadénosine triphosphate	PWM	matrice poids position
ddCTP	di-désoxycytidine triphosphate	R	arginine
ddGTP	di-désoxyguanosine triphosphate	RE	reticulum endoplasmique
ddTTP	di-désoxythymidine triphosphate	RNase	ribonucléase
DBD	DNA binding domain	scFv	fragments variables simple chaîne
DNase	désoxyribonucléase	sha	shavenoïd
DSB	cassures double brin	SIR2	Sirtuin 2
dyl	<i>dsky-like</i>	SAM	S-adenosylméthionine
E. coli	<i>Escherichia coli</i>	smORF	small Open Reading Frame
EcR	récepteur à l'ecdysone	snRNP	petites ribonucléoprotéine
EGF	epidermal growth factor	SMC	Structural Maintenance of Chromosome
EM	microscopie électronique	SMLM	Single Molecule Localization Microscopy
FIS	Factor for Inversion Stimulation	SUUR	Suppressor of Under-Replication
FISH	hybridation fluorescente in situ	svb	<i>shavenbaby</i>
FRAP	Fluorescence Recovery After Photobleachin	T	thymine
FROS	Fluorescents Répresseur/Opérateur	TAD	domaines associés topologiquement
G	guanine	tal	tarsal-les
GFP	Green Fluorescent Protein	TALE	Transcription Activator Like Effector
GTP	guanosine triphosphate	TF	facteur de transcription
H1	histone 1	TFBS	sites de liaison de facteurs de transcription
H3K4me2	histone 3 diméthylée sur la lysine 4	TIRFM	Total Internal Reflection Fluorescence Micoscopy
H3K9me2	histone 3 diméthylée sur la lysine 9	TMV	virus de la mosaïque du tabac
H3K27me3	histone 3 diméthylée sur la lysine 9	UTR	région non traduite
H3K79me3	histone 3 diméthylée sur la lysine 27	Wg	wingless

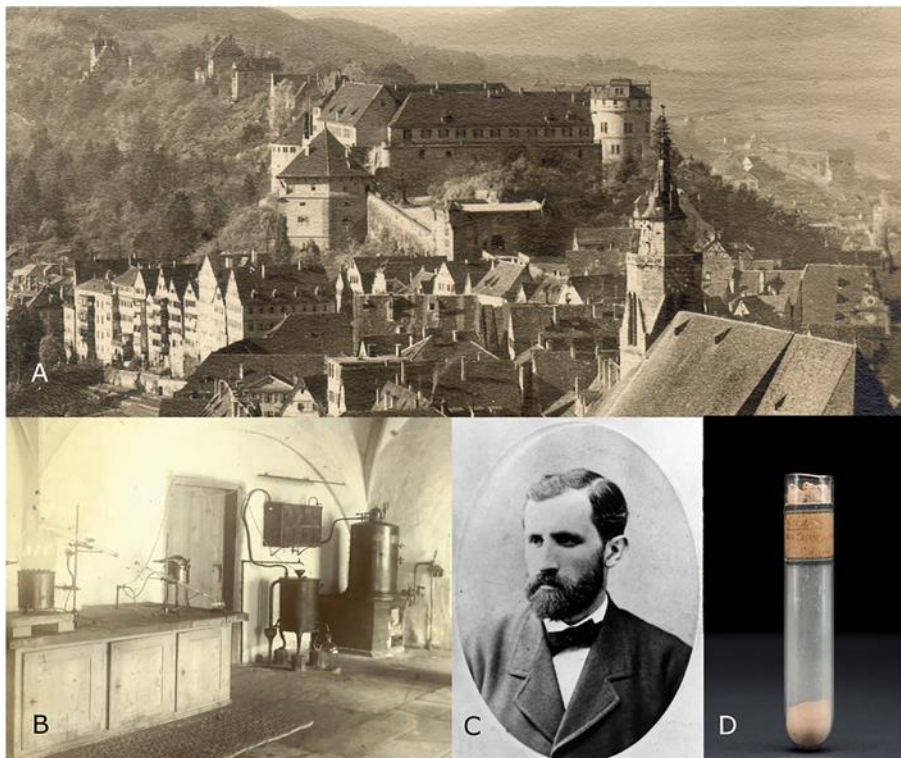
## **Avant-propos ...**

# De la découverte de l'ADN à la localisation de locus génétique

## ou ma petite histoire de l'ADN

### Tout commence, il y a 150 ans ...

L'histoire de l'acide désoxyribonucléotide (ADN) commence il y a 150 ans avec sa découverte par Friedrich Miescher, un médecin suisse. En 1869, Miescher, effectue la première purification de l'ADN. Il isole à partir de noyaux de leucocytes provenant de pus de bandages chirurgicaux, une substance acide, riche en phosphate qu'il nomma « **nucléine** » car isolée à partir de noyaux (**fig. 1**). Il montre que cette substance n'est pas digérée par une protéase, la pepsine et que ce n'est pas une protéine. A la suite d'autres tests sur la « nucléine » isolée à partir de sperme de saumon, Miescher confirme la présence d'éléments trouvés dans les molécules organiques : carbone, hydrogène, oxygène et azote. Contrairement aux protéines, il note l'absence de soufre dans la « nucléine » mais de large quantité de phosphore,



**Figure 1: Friedrich Miescher découvre la "nucléine" comme nouveau composant biochimique des cellules.**  
**A)** Image contemporaine du château de Tübingen. **B)** A l'époque, l'ancienne cuisine du château servait de laboratoire biochimique où **C)** Friedrich Miescher a découvert la "Nucléine". **D)** A côté de lui, la fiole contenant la Nucléine "historique" dérivée du sperme de saumon par Miescher en 1871.  
(à partir de Thess, 2021).

représentant 22,5 % de la masse totale (un chiffre très proche de la proportion réelle de 22,9 %) et présent sous forme d'acide phosphorique. Il confirme les propriétés acides de la nucléine et qu'il devrait s'agir d'un « acide avec 3 voire 4 bases acides » de grand poids moléculaire car peu diffusible (Dahm, 2005).

### Chromosome, chromatine et acides nucléiques...

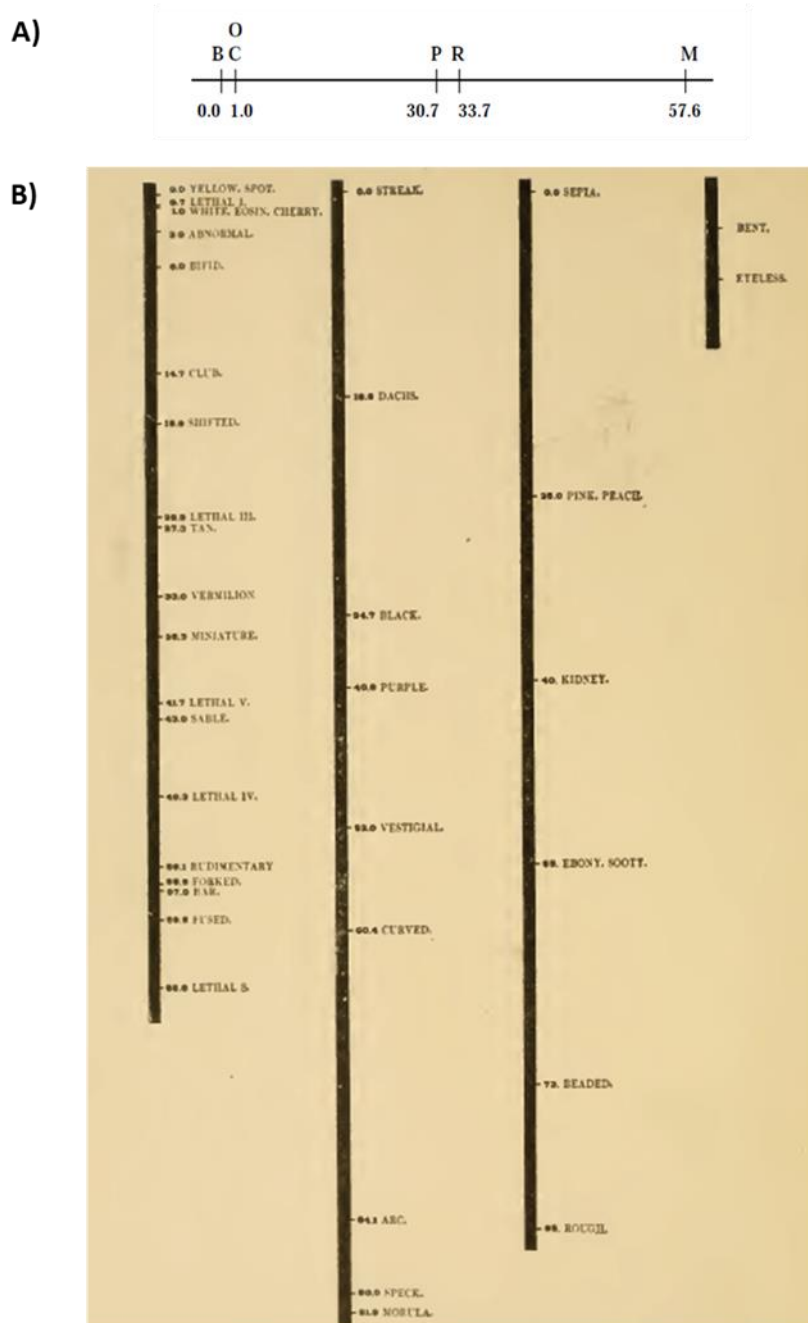
Quelques années plus tard, en 1878, Albrecht Kossel, un biochimiste allemand, confirme la localisation de la nucléine dans le noyau et identifie les composants de la nucléine : les bases purines et pyrimidiques, un sucre et l'acide phosphorique. Il identifia la nucléine comme un composant de la **chromatine** avec des protéines comme les **histones** qu'il a découvertes en 1884 (Dahm, 2008). La chromatine (du grec chrôma=couleur), substance acide qui capte les colorants basiques, a été décrite par Walther Flemming à la suite d'observation, à l'aide de microscopes, de structure dans les noyaux qui se laissent colorer. Flemming, focalisant ses travaux sur le comportement des noyaux lors de la division cellulaire, observe que la chromatine se transforme en filaments lors de la division cellulaire qu'il nomme **mitose** (du grec mitos = filament). Il donnera **chromosome** comme nom à ces filaments ([chroma=couleur ; soma = corps] = corps coloré) (Olins, 2003 ; Dahm, 2005, 2008). En 1889 Richard Altmann renomme la nucléine en **acide nucléique**. En 1902, Walter Sutton et Theodor Boveri développent indépendamment la « Théorie Chromosomique de l'Hérédité ». En effet ils proposent que les chromosomes portent les facteurs héréditaires, en accord avec les lois de Mendel qui ont été redécouvertes indépendamment par trois scientifiques, Hugo de Vries, Carl Erich Correns et Erich Von Tschermak (Olins, 2003 ; Dahm, 2005). Sutton, en étudiant la méiose chez la sauterelle et Boveri, le développement embryonnaire de l'oursin, arrivent aux mêmes conclusions. Les chromosomes demeurent organisés et dans des structures individuelles à travers le processus de division, qu'ils vont par paire, avec un membre provenant de chaque parent. Au cours du processus de la méiose, le nombre de chromosomes est réduit de moitié dans les spermatozoïdes et les ovules, le nombre original étant rétabli dans l'œuf fécondé, pendant la reproduction. Ce processus était en accord avec l'idée de ségrégation de Mendel. En 1902, Sutton suggère que "l'association des chromosomes paternels et maternels par paires et leur séparation ultérieure lors de la division réductrice...

peuvent constituer la base physique de la loi mendélienne de l'hérédité" (Sutton, 1902, 1903 ; Baltzer, 1964 ; Genome News Network, a, b).

### **Et Thomas H. Morgan arriva : naissance de la génétique expérimentale.**

La « Théorie Chromosomique de l'Hérédité » de Sutton et Boveri fut controversée jusqu'à sa démonstration en 1915 par les travaux de Thomas Hunt Morgan chez la mouche *Drosophila melanogaster*. En effet, Morgan confirme les lois mendéliennes de l'hérédité et l'hypothèse selon laquelle les facteurs de l'hérédité ou **gènes**, mot utilisé par Wilhelm Johannsen en 1909 pour parler des facteurs de l'hérédité (Dahm, 2005), sont situés sur les chromosomes. Les travaux de Thomas Hunt Morgan et de ces collaborateurs, notamment Alfred Sturtevant, portent sur la transmission de caractères et mutations chez la drosophile. Ils s'intéressent à la mutation de la couleur rouge de l'œil de la drosophile en blanc. Morgan appelle « white » le gène portant ce caractère. Cette façon de nommer les gènes chez les drosophilistes perdure, en effet, il est coutume de donner à un gène le nom d'après le phénotype de leurs allèles mutants. En étudiant la ségrégation de ce caractère « yeux blancs », ils s'aperçoivent que ce caractère « yeux blancs » est récessif par rapport au caractère « yeux rouges » et qu'il est porté par un chromosome sexuel (Morgan, 1910). Ainsi la transmission de caractères pouvait être liée au sexe mais également que des gènes pouvaient être portés par des chromosomes non liés au sexe. Morgan établit la théorie de l'hérédité liée au sexe qui va servir de base à la suite de ces travaux. Chaque chromosome de la drosophile contient un groupe d'unités physiques réelles correspondant aux gènes. Ces derniers sont disposés linéairement le long du chromosome. Ainsi les gènes sont alignés sur les chromosomes dans un ordre bien défini tel un collier de perles sur un fil. Autrement dit, les chromosomes sont le support des gènes. L'analyse de la transmission de caractères de milliers de drosophiles permet à Morgan et à ces étudiants de montrer la présence de milliers de gènes sur les chromosomes, correspondant tous à des caractères héréditaires distincts. En se basant sur la recombinaison chromosomique, Morgan et ces collaborateurs établissent des cartes chromosomiques donnant la position des gènes sur les chromosomes. Alfred Sturtevant, alors étudiant, construit en 1913 la première carte génétique du chromosome X contenant les gènes *yellow*, *white*, *vermillon*, *miniature*, et *rudimentary*, correctement positionnés sauf pour *miniature* et *rudimentary* qui sont inversés. En 1915, Thomas Hunt Morgan publie "The Mechanism of

Mendelian Heredity” où le frontispice représente une carte génétique des quatre chromosomes avec une quarantaine de gènes dont *miniature* et *rudimentary* qui sont correctement localisés (fig. 2 ; Morgan, 1915 ; Sturtevant, 1913). Les travaux de Morgan eurent des conséquences majeures en ouvrant la voie à la génétique expérimentale moderne.



**Figure 2: Première carte génétique.**

**A)** Première carte chromosomique construite en 1913 par Alfred Sturtevant; B= *yellow*; C = *white*; O = *eosin eyes*; P = *vermillion*; R = *rudimentary*; M = *miniature* (Sturtevant, 1913).

**B)** Frontispice du livre de Thomas Hunt Morgan “The Mecanism of Mendelian Heredity” publié en 1915 représentant une carte génétique des quatre chromosomes de la drosophile (Morgan, 1915).

Ce début du vingtième siècle a révélé de grandes avancées dans le domaine émergent de la génétique mais a été largement dépourvu dans la compréhension de la structure de la chromatine et malgré ces avancées, il demeurait une question qui restait sans réponse : quelle est la nature chimique du support de l'hérédité ?

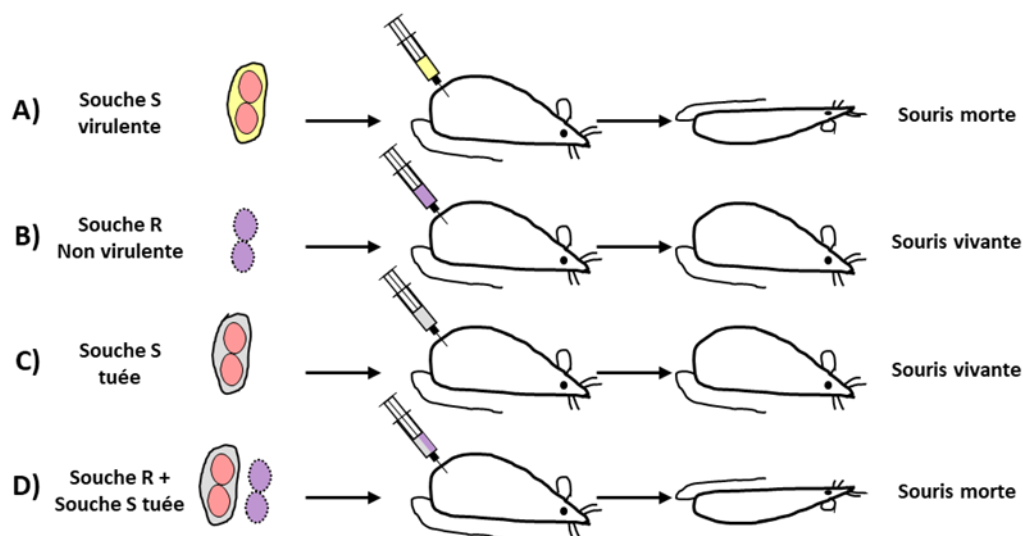
### La théorie tétranucléotidique ou comment une théorie peut bloquer l'avancée des connaissances.

Pendant la première moitié du XXème siècle, il y a peu d'intérêt pour l'ADN en partie à cause la **théorie tétranucléotidique** proposée par Phoebus Levene en 1910 (Levene, 1919). Ce biochimiste qui découvre le ribose et le désoxyribose respectivement en 1909 et 1929, montre que l'ADN est composé de 4 bases azotées (adénine (A), cytosine (C), guanine (G) et thymine (T)), chacune reliée à un sucre et à un phosphate dans un ordre précis : phosphate – sucre – base pour former une unité qu'il appelle **nucléotide**. Il pense que la structure de l'acide nucléique est organisée en un tétranucléotide répétitif composé de ces 4 nucléotides reliés entre eux par les groupes phosphates, pour constituer le squelette de la molécule. Ainsi Phoebus Levene considère que les 4 bases sont dans des proportions équimolaires ce qui signifie que l'ADN contiendrait exactement la même quantité de chacune des bases (Teich 1975 ; Dahm, 2008). La communauté scientifique, comme d'ailleurs Phoebus Levene lui-même, est sceptique quant au fait que la « simplicité » de la molécule d'ADN puisse être le support de l'information génétique, capable de véhiculer une information complexe. Les scientifiques considèrent que le support de l'hérédité est plus probablement véhiculé par les protéines qui offrent une immense diversité avec leur multitude de combinaisons possibles à partir de leur code à 20 éléments plutôt que l'ADN qui est constitué de l'agencement de seulement 4 nucléotides différents, trop peu, pensent-ils, pour stocker toute l'information génétique. Ses idées sur la structure de l'ADN sont erronées mais ses travaux constituent une base essentielle pour les travaux ultérieurs qui permettront de déterminer la structure de l'ADN.



### Quel le support pour l'hérédité ? ADN ou protéine ?

La première étape vers l'identification du support de l'hérédité est issue des travaux sur les pneumocoques par un médecin bactériologue anglais, Frederick Griffith, en 1928, au cours desquels il découvre le phénomène de « transformation » bactérienne. Lorsqu'il co-injecte à des souris des pneumocoques non pathogènes (R) de type antigénique II et des pneumocoques virulents (S) de type antigénique III mais préalablement tués par la chaleur, il constate la mort des souris. Le sang récupéré de ces animaux produit des pneumocoques de type III. Les caractères de virulence et de spécificité biologique de type III de la souche bactérienne S ont été transférés à la souche bactérienne R (Griffith, 1928 ; Avery, 1944). Cette transformation est liée à la présence d'un « principe transformant » ou « principe actif » provenant d'extraits bactériens tués par la chaleur (**fig. 3**). C'est la première fois que des caractères génétiques sont transmis à l'aide d'une « substance chimique » contenue dans des bactéries mortes (Berche, 2016). La nature chimique du « principe actif » mis en évidence

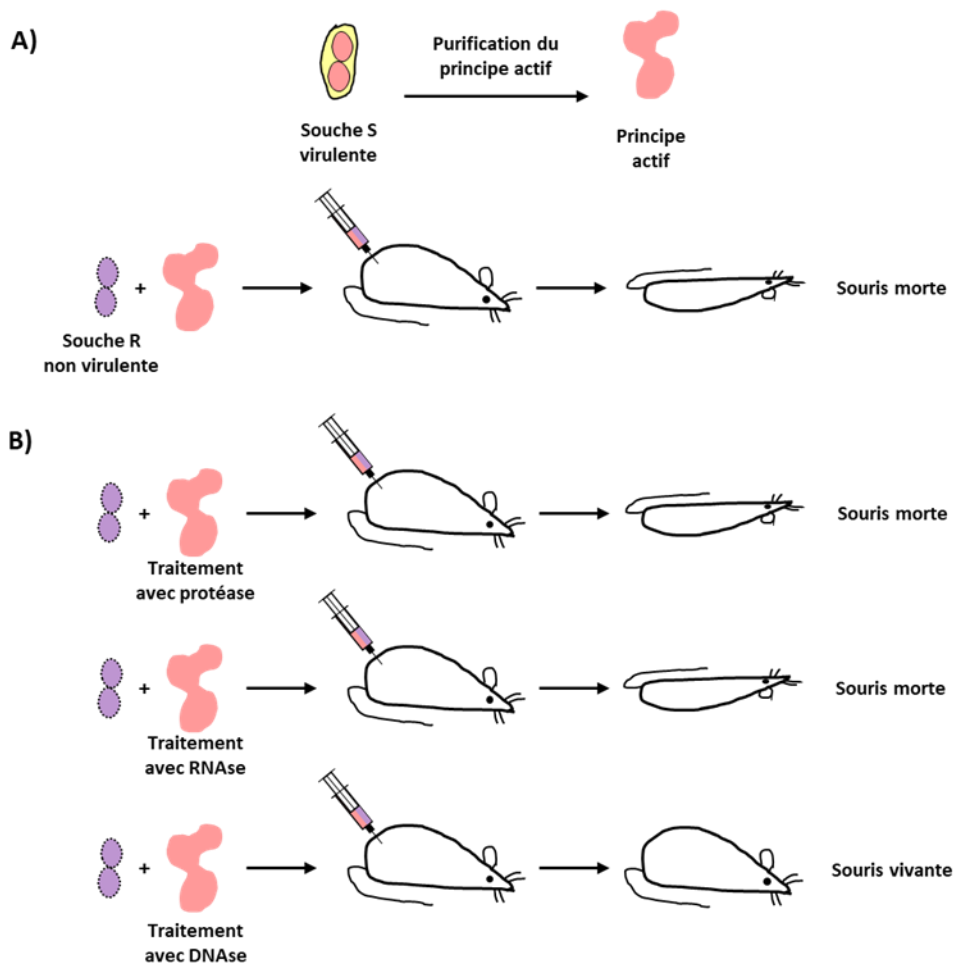


**Figure 3: Expérience de Griffith, 1928.**

- A) L'injection de la souche de pneumocoques S virulente entraîne la mort de la souris.
- B) Les pneumocoques de la souche R, non pathogènes, injectés à la souris n'entraînent pas la mort.
- C) L'injection de pneumocoques S, tués par la chaleur, et donc non pathogènes, n'entraîne pas la mort de la souris.
- D) La co-injection de pneumocoques de la souche S tués par la chaleur + de pneumocoques R vivants entraîne la mort de la souris mettant en évidence la présence d'un "principe transformant ou actif" provenant des pneumocoques de la souche S tués par la chaleur.

par Griffith ne sera élucidée qu'une dizaine d'années plus tard, en 1944, par Avery, McLeod et McCarty (Avery, 1944). Ils basent leurs recherches sur les expériences de Griffith et effectuent une série de purification à partir d'extrait de pneumocoques et isolent une solution pure de

cet agent transformant. L'analyse chimique révèle que cette substance a la même composition que l'ADN (Maguin, 2021). De plus, les traitements de cet agent avec une ribonucléase (RNase) purifiée et avec différentes protéases n'abolissent pas son pouvoir transformant. Par contre un traitement de cette substance par des extraits dégradant l'ADN (Avery, 1944) ou par une désoxyribonucléase (DNase) purifiée (McCarty, 1946a, b) inactive l'activité de transformation (**fig. 4**). Dans sa publication, Avery conclut: « **The evidence presented supports the belief that a nucleic acid of the desoxyribose type is the fundamental unit of the transforming principle of Pneumococcus Type III.** » (Avery, 1944 ; Maguin,2021).



**Figure 4: Expérience d'Avery, McLeod et McCarthy, 1944.**

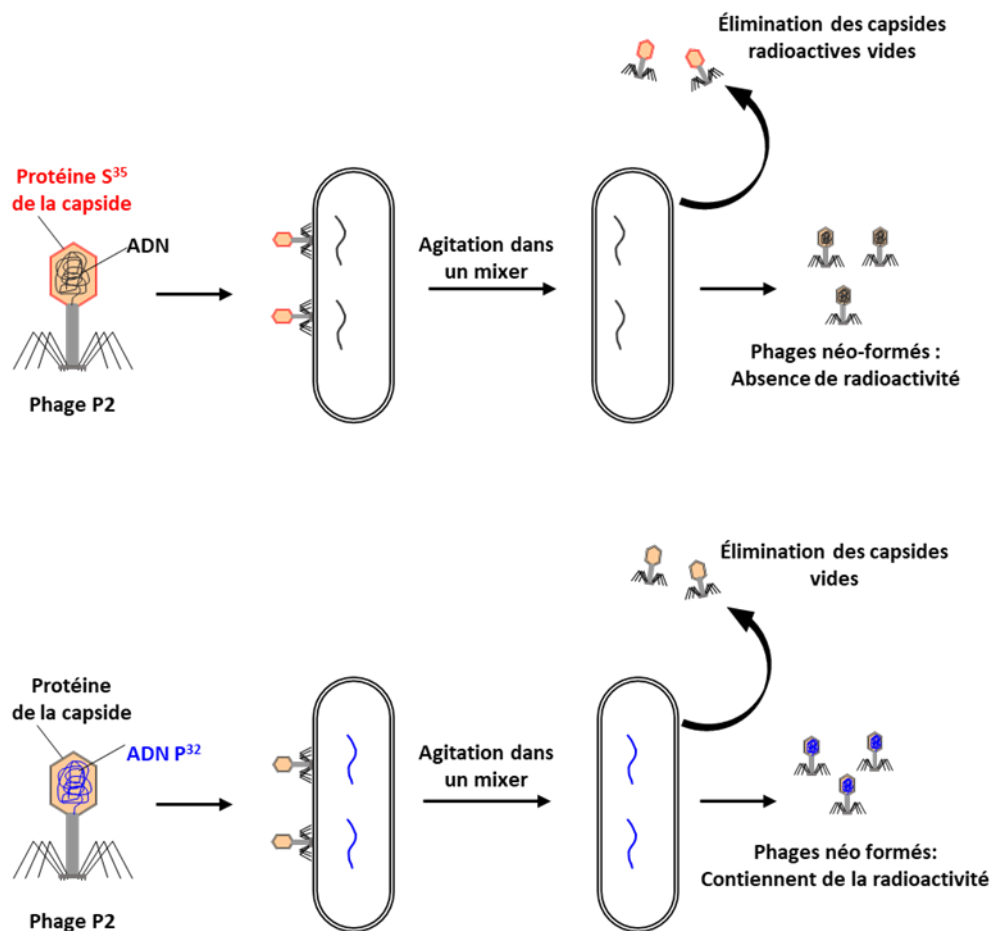
**A)** Avery & coll purifient à partir d'une souche virulente S de pneumocoques le principe actif. Ce principe actif est capable de transformer une souche non virulente R en une souche virulente capable de tuer une souris

**B)** Ce principe transformant est testé avec différentes enzymes pour déterminer ses propriétés chimiques. Les traitements avec des protéases ou des RNases n'abolissent pas la capacité de transformation de ce principe actif. Par contre, après un traitement avec une DNase capable de dégrader l'ADN, la principe actif n'est plus capable de transformer la souche non virulente en souche virulente. Ce résultat démontre que l'ADN est le principal composant de ce principe actif et qu'il contient l'information génétique nécessaire pour transformer une souche R non virulente en souche S virulente.

Cependant, malgré ces résultats solides, une partie de la communauté scientifique reste sceptique quant au rôle de l'ADN comme support de l'information et émettent la possibilité que des traces d'impuretés dans leur extrait purifié pourrait être le véritable agent transformateur, Avery et ses collègues ne l'excluant pas non plus (Avery, 1944 ; Maguin, 2021). Il faut attendre quelques années, en 1952, pour que les expériences menées par Hershey et Chase confirment que l'ADN est le support de l'information génétique. Ces 2 scientifiques mènent des expériences sur le bactériophage T2. Ils utilisent un marquage différentiel de l'ADN et des protéines du phage avec des isotopes radioactifs : soit du phosphore radioactif ( $P^{32}$ ) est incorporé à l'ADN, soit du soufre radioactif ( $S^{35}$ ) est incorporé aux protéines. Des bactéries *Escherichia coli* sont infectées par ces phages marqués soit au  $P^{32}$  pour suivre l'ADN soit au  $S^{35}$  pour suivre les protéines. Ils observent que les protéines marquées au  $S^{35}$  ne pénètrent pas dans la bactérie et ne sont pas retrouvées dans la descendance des phages alors que l'ADN marqué au  $P^{32}$  se trouvent dans les bactéries : cette fraction nucléique est seule responsable de la reproduction des phages (Hershey, 1952 ; Berche, 2016). Ainsi, cette célèbre expérience de Hershey et Chase, qui est une des toutes premières où des isotopes radioactifs sont utilisés pour suivre des molécules, montre que seul l'ADN de phage et non leur coque protéique permet à ces bactériophages d'infecter des bactéries et de se multiplier (**fig. 5**). Ces résultats confirment ceux d'Avery. A l'époque ce fut un séisme conceptuel qui substitue l'ADN aux protéines comme support de l'information génétique et replace l'ADN au centre de la recherche scientifique. Les expériences d'Avery et de Hershey peuvent être considérées comme la naissance de la biologie moléculaire (Berche, 2016 ; Maguin, 2021).

Les années 1940-50 sont des années marquées par des expériences fondatrices dans l'évolution des connaissances sur le support de l'hérédité. En 1944, le physicien Erwin Schrödinger publie un petit livre « What is life ? » où il introduit l'idée d'un « cristal aperiodique » c'est-à-dire où l'information génétique serait contenue dans une structure codée, aussi rudimentaire qu'un code binaire, très complexe, basée sur une association correctement ordonnée d'atomes, dotée de stabilité chimique, permettant de stocker et transmettre d'énormes quantités d'informations pour permettre l'hérédité. La vision et les idées de Schrödinger sont une source d'inspiration et suscitent l'enthousiasme de nombreux

scientifiques, notamment de Watson et Crick, pour rechercher et identifier les molécules du vivant permettant le traitement de ces informations (Berche, 2016 ; Bruch, 2019).



**Figure 5: Expérience d’Hershey & Chase, 1952.**

Hershey et Chase infectent des bactéries avec un phage P2 qui injecte son ADN à l’intérieur des bactéries tout en laissant sa capsid à l’extérieur. Ils infectent des bactéries avec des phages qui sont marqués alternativement avec du soufre  $S^{35}$  incorporé aux protéines de la capsid ou avec du phosphore  $P^{32}$  incorporé à l’ADN. Après infection, ils agitent dans un mixer afin de séparer les capsides vides de phages des bactéries puis ils remettent en culture les bactéries infectées. Les phages produits à partir de bactéries infectées par des phages avec la radioactivité incorporée dans l’ADN contiennent de la radioactivité ce qui n’est pas le cas lorsque la radioactivité est incorporée dans les protéines de la capsid phagique.

### La règle de Chargaff.

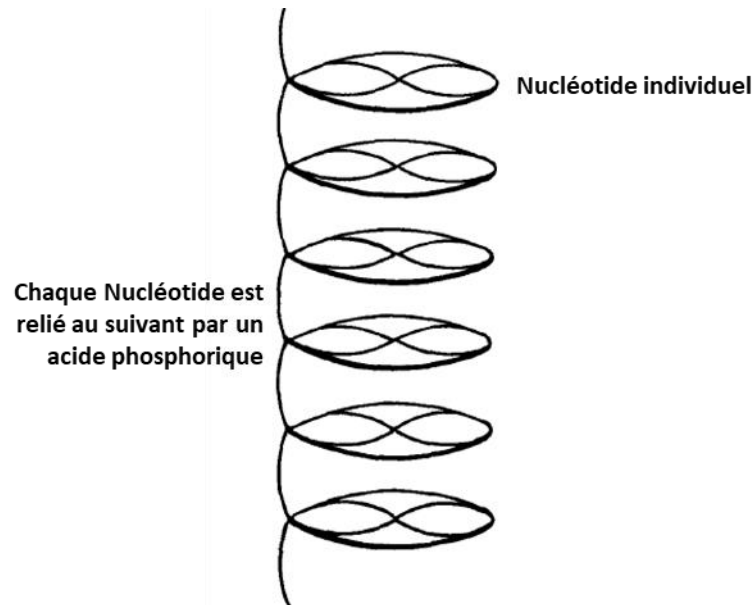
Parallèlement aux travaux d’Avery et d’Hershey faisant de l’ADN le support de l’information génétique, des recherches sur la structure et la composition chimique de l’ADN sont menées, notamment par Erwin Chargaff, un biochimiste des lipoprotéines, qui, à la lecture des travaux d’Avery décide d’orienter ses recherches sur les acides nucléiques. Grand bien lui en a pris,

puisque ces travaux l'ont amené à 2 découvertes majeures qui ont joué un rôle essentiel dans la compréhension de la structure en double hélice de l'ADN : 1) la composition en bases puriques et pyrimidiques de l'ADN varie d'une espèce à l'autre en particulier dans les quantités relatives des bases A, C, G, T, mais cette composition est constante pour une espèce donnée, 2) dans n'importe quelle cellule ou organisme, il y a autant de bases puriques que pyrimidiques et les rapports A/T et G/C sont égaux à 1 (Chargaff, 1950a, b, 1951, 1952). Ces découvertes ont été nommées « **les règles de Chargaff** » et ont mis à mal la théorie tétranucléotide de Levene.

### **Les physiciens à la rescousse pour la compréhension de la structure de l'ADN.**

La compréhension de la structure de l'ADN viendra en partie des physiciens et des données de cristallographie par rayons X, une technique d'analyse fondée sur la diffraction des rayons X par la matière notamment quand elle est sous forme de cristaux. La diffraction des rayons X par des cristaux a été découverte par le physicien allemand Max von Laue ce qui lui valut le prix Nobel de Physique en 1914. Cette technique a été adaptée pour l'analyse de cristaux de substances minérales mais rapidement elle a été utilisée pour l'étude des macro-molécules biologiques dans les années 1920 et des protéines dans les années 1930 par John Desmond Bernal et William T. Astbury (Amar-Costesec). Les premières photographies de l'acide thymonucléique (dénomination de l'ADN utilisée à cette époque) par cristallographie au rayons X ont été réalisées par William T. Astbury et Florence O. Bell, en 1937 où après l'analyse des données ils concluent « ... **the spacing of 3,34 Å along the fibre axis corresponds to that of a close succession of flat or flattish nucleotides standing out perpendicularly to the long axis of the molecule to form a relatively rigid structure ...** » (Astbury, 1938a). Les clichés indiquent une structure régulière où le plan des bases est perpendiculaire au grand axe de la molécule d'acide thymonucléique avec une distance de 3,4 Å (soit 0,34nm) entre 2 bases soit une valeur très proche de la valeur réelle dans la forme B de l'ADN (0,32nm), confirmant les conclusions d'une étude de la viscosité et de la double réfraction du flux d'acide thymonucléique en solution aqueuse par Signer et Hammarsten (Signer, 1938). Astbury et Bell proposent un modèle d'ADN monocaténaire comme une colonne de nucléotides, espacés les uns des autres de 3,4 Å, constitués de bases puriques ou pyrimidiques liées à un sucre, chaque nucléotide étant relié par l'intermédiaire d'un acide phosphorique : cette colonne de

nucléotide a été appelée la « pile de pennies » d'Astbury et Bell (**fig. 6**) (Astbury, 1938b ; Hall, 2011).



**Figure 6: Représentation de la “pile de pennies” d’Astbury et Bell, 1938.**

Astbury et Bell se représentaient l’acide thymonucléique ou ADN comme une colonne de nucleotides constituant l’unité de l’acide thymonucléique. Chaque nucleotide étant relié au suivant par un acide phosphorique et séparé les uns des autres par 3,4 Å. (à partir de Astbury, 1938).

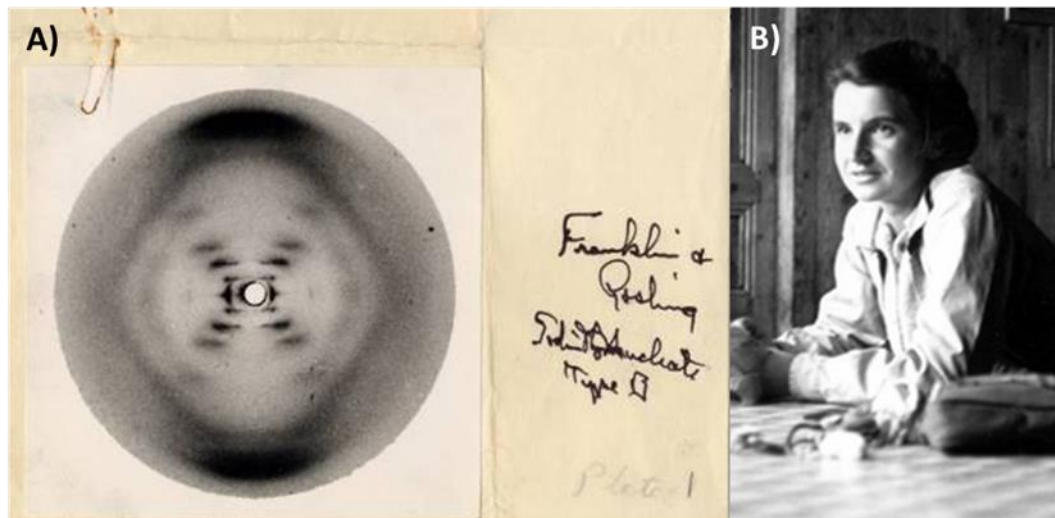
Les années qui ont suivi les résultats d’Astbury et Bell n’ont pas amené de données expérimentales supplémentaires notamment de photographies aux rayons X de qualité supérieure à celles d’Astbury et Bell. Il faudra attendre le début des années 50 pour voir de nombreuses études portant sur la structure de l’acide désoxyribonucléique provenant notamment de physiciens, cristallographes... Pourtant une des avancées sur la compréhension de la structure de l’acide désoxyribonucléique sera amenée par le chimiste Linus Pauling qui s’intéresse à la structure des protéines en utilisant la diffraction des rayons X et qui propose une forme hélicoïdale pour la chaîne polypeptidique comme structure secondaire des protéines : **l’hélice alpha** (Pauling, 1951). S’appuyant sur les travaux d’Astbury et Bell, et faisant un parallèle avec la configuration des chaînes de polypeptides (enchaînement d’unités, même distance entre 2 unités...), il suppose que la structure de l’ADN peut être considérée comme une structure hélicoïdale (Pauling, 1953a). Astbury et Bell avaient également fait ce parallèle entre chaîne polypeptidique et acide thymonucléique (Astbury, 1938a). Pauling voit

l'ADN comme une molécule géante, cylindrique avec une section approximativement circulaire d'un diamètre de  $15 \pm 5 \text{ \AA}$  (Pauling, 1953a).

Sven Furberg, biologiste et cristallographe norvégien, est le premier à proposer une structure hélicoïdale et monocaténaire de l'ADN où les bases purines et pyrimidiques sont parallèles entre elles mais perpendiculaires à l'axe de l'hélice tandis que les sucres et la liaison phosphodiester sont parallèles à l'axe. De plus les bases sont à l'intérieur de l'hélice tandis que les phosphates sont à l'extérieur (Furberg, 1952). Pauling se base sur l'analyse de photographies au rayon X de thymonucléate de sodium (l'ADN) et notamment sur celles d'Astbury et Bell. Il considère que ces photographies sont de moins bonne qualité que celles d'Astbury et Bell (« ... **Our own preparations have given photographs somewhat inferior to those of Astbury and Bell...** » Pauling, 1953a). Il propose comme modèle pour l'ADN, une structure constituée de 3 chaînes polynucléotidiques hélicoïdales entremêlées, et contrairement à Furberg, avec les groupes phosphates près de l'axe central et les bases nucléiques orientées vers l'extérieur (Pauling, 1953a, 1953b). Robert Fraser suggère, comme Pauling, une structure pour les acides nucléiques impliquant trois chaînes hélicoïdales entremêlées mais à la différence de Pauling, Fraser rejoint Furberg pour la localisation des groupes phosphates à la surface de la molécule avec les bases dirigées vers l'axe (Fraser, 1953). Malgré ces nombreux modèles, aucun ne fait l'unanimité principalement dû à l'absence de données de rayons X convaincantes.

### **De la photo 51 de Rosalind Franklin à la découverte de la double hélice de l'ADN : ou comment sans faire une seule expérience on décroche le prix Nobel !!**

Parallèlement à ces travaux, Maurice Wilkins et Rosalind Franklin, au King's College à Londres, travaillent sur la définition spectrale par diffraction des rayons X des formes non-hydratées (forme A) et hydratées (forme B) de l'ADN. A partir de précipité de thymonucléate de sodium, Wilkins obtient des fibres uniformément orientées. Cette uniformisation d'orientation des fibres est importante car elle influe sur la qualité du cliché de diffraction. Ainsi Rosalind Franklin et Raymond Gosling, doctorants à l'époque, obtiennent des clichés de diffraction des formes A et B, de bien meilleure qualité, et notamment la remarquable et célèbre **photo 51** de la forme B (**fig. 7** ; Franklin, 1953).



**Figure 7: La "Photo 51"!**

**A)** Cliché de la diffraction de rayons X sur des cristaux d'ADN de type B de thymus de veau, réalisé par Franklin et Gosling, le 2 mai 1952

<http://scarc.library.oregonstate.edu/coll/pauling/dna/pictures/sci9.001.5-large.html>

**B)** Rosalind Franklin (1920-1958).

Ainsi toutes les pièces du puzzle sont réunies pour élucider la structure moléculaire de l'ADN :

1) la composition chimique de l'ADN et son agencement (bases azotées-sucre-groupements phosphates) ont été déterminés notamment par Levene dans les années 1920 puis confirmés par Astbury (Levene, 1919 ; Astbury, 1938a, b).

2) les clichés de diffraction de rayons X de Rosalind Franklin et Raymond Gosling (photo 51 !) faisant suite à ceux d'Astbury et Bell et aux travaux de Pauling sur l'hélice alpha, indiquent une structure hélicoïdale de deux molécules co-axiales de l'ADN, comportant une succession d'unités répétitives (les bases) espacées de 3,4 Å avec une période de rotation de 34 Å soit 10 nucléotides par tour d'une hélice de 10 Å de rayon et où les groupes phosphates sont à l'extérieur de l'unité structurale avec les groupes sucres-bases tournés vers l'intérieur, vers l'axe de l'hélice, et avec le plan des bases perpendiculaire à l'axe de la molécule (Astbury, 1938a, b; Pauling, 1953a, b; Franklin, 1953; Wilkins, 1953).

3) « les règles de Chargaff » sur la composition en bases puriques et pyrimidiques de l'ADN qui est constante pour une espèce donnée mais varie d'une espèce à l'autre et qu'il y a autant d'adénine que de thymine et autant de cytosine que de guanine.

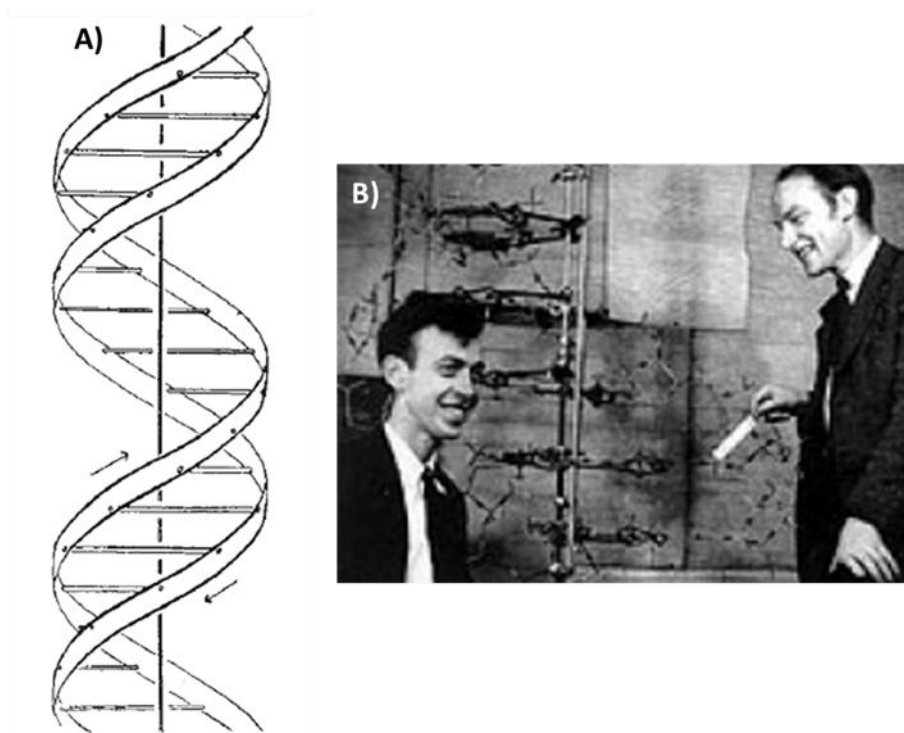
Nous sommes au début de 1953 et à partir de ces pièces du puzzle, une compétition va s'installer notamment entre d'un côté Watson et Crick (laboratoire Cavendish à Cambridge)



et de l'autre Wilkins et Franklin (King's College à Londres) pour établir un modèle permettant de reconstituer la structure tridimensionnelle de l'ADN.

Tandis que Rosalind Franklin poursuit ces travaux afin d'obtenir des données objectives, Watson et Crick essaient de reconstituer la structure de l'acide désoxyribonucléique en construisant, notamment à l'aide de formes découpées dans du carton représentant les quatre bases, des modèles qui proposent une structure répondant à l'ensemble des données cristallographiques et biochimiques disponibles.

Ils parviennent à la solution du problème sans avoir fait une seule expérience de diffraction aux rayons X et ils publient dans le numéro du 25 avril 1953 du journal *Nature*, sous forme d'un court article d'une seule page, leur modèle de double hélice représentant la structure moléculaire de l'ADN (fig. 8 ; Watson, 1953a).



**Figure 8: La double hélice de l'ADN.**

**A)** Schéma représentant la double hélice de l'ADN publié dans *Nature* le 25 avril 1953 par James Watson et Francis Crick.

**B)** Watson et Crick devant un modèle de la fameuse double hélice.  
(à partir de Kahn, 2003).

Ayant eu accès aux clichés et données de diffraction obtenues par Rosalind Franklin, ils observent une forme en X, caractéristique d'une structure hélicoïdale. Ils comprennent qu'il s'agit d'une double hélice. La symétrie des clichés de cristallographie leur indique que la

structure possède un axe d'ordre 2 et que chaque chaîne va dans un sens opposé. Contrairement au modèle de Pauling, ils considèrent que les bases sont à l'intérieur de l'hélice tandis que les phosphates sont à l'extérieur, comme dans les modèles de Fraser (Fraser, 1953) ou de Furberg (Furberg, 1952). Pour Watson et Crick, le modèle de Pauling n'est pas « satisfaisant », car les clichés ont été obtenus à partir du sel de l'acide désoxyribonucléique ce qui signifie que les groupes hydroxydes des phosphates ont perdu leur hydrogène et qu'ils possèdent une charge négative. Dans le modèle de Pauling, les groupes phosphates se trouvant à l'intérieur, se repousseraient mutuellement.

Mais ce qui est génial et révolutionnaire dans leur modèle, c'est la façon dont les 2 chaînes se tiennent ensemble par les bases purines et pyrimidines. Ils ont pour cela bénéficié des informations fournies par le physico-chimiste Jerry Donohue sur la forme tautomérique des bases : la forme tautomérique la plus plausible est avec la configuration céto plutôt que enol. En prenant compte du diamètre de 20 Å de la molécule d'ADN, de la distance inter-atomique et de la forme tautomérique céto des bases, Watson, construisant des maquettes avec les formes céto des bases trouvent que les seules possibilités d'appariements spécifiques des bases sont une purine avec une pyrimidine c'est-à-dire adénine (purine) avec thymine (pyrimidine) et guanine (purine) avec cytosine (pyrimidine). Les 2 chaînes se tiennent par les liaisons hydrogènes entre leurs bases complémentaires se trouvant sur chacune des 2 chaînes: 2 liaisons hydrogènes entre adénine et thymine ainsi qu'entre cytosine et guanine (c'est la seule erreur de leur modèle puisque qu'entre guanine et cytosine il y a 3 liaisons hydrogènes). Ils introduisent aussi la notion de complémentarité puisque si « ... la séquence des bases sur une chaîne est donnée alors la séquence sur l'autre chaîne est automatiquement déterminée... ». Cette notion est confortée par les « Règles de Chargaff » qui prônent l'équivalence des proportions des bases : autant de purines que de pyrimidines, autant d'adénine que de thymine et de guanine que cytosine (Chargaff, 1950a, b, 1951, 1952).

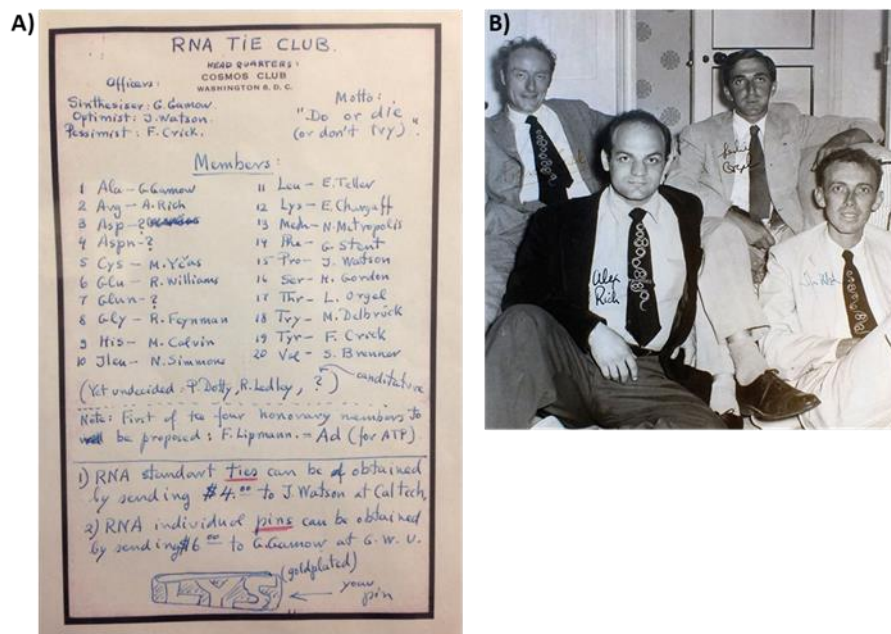
Ils terminent leur article en notant l'importance biologique de la structure en double hélice où « ... l'appariement spécifique que nous avons noté suggère immédiatement un mécanisme possible de reproduction (réplication) pour le matériel génétique ... » (Watson, 1953a). En 1953, ils publient deux autres articles où ils détaillent la structure de la double hélice notamment les liaisons hydrogènes spécifiques entre les couples de bases adénine/thymine et guanine/cytosine et le rôle que doit avoir l'ADN dans une cellule (Watson,

1953b, c). Ils pensent que le matériel génétique doit remplir deux fonctions. Il doit se répliquer, se dupliquer par lui-même et exercer une influence spécifique sur la cellule. Cette spécificité s'exprime par la séquence précise des bases qui constitue un code qui porte l'information génétique (Watson, 1953a, b). A partir de l'appariement spécifique où chaque chaîne est complémentaire de l'autre chaîne, Watson et Crick suggèrent un mécanisme de réplication semi-conservatif du matériel génétique où avant la duplication, les ponts hydrogènes se cassent, les deux chaînes se déroulent et se séparent, chaque chaîne servant de modèle pour le recrutement de nucléotides libres qui peuvent se lier entre eux et former de nouveaux ponts hydrogènes avec leurs nucléotides complémentaires permettant de produire une nouvelle chaîne. Après réplication, il y a deux paires de chaînes où il n'y en avait qu'une auparavant avec une séquence des paires de bases identiques (Watson, 1953b, c).

Ainsi, sans avoir fait une seule expérience de diffraction aux rayons X, Watson et Crick ont élucidé la structure de l'ADN à l'aide de raisonnements théoriques. Dans le même numéro du 25 avril 1953 de *Nature*, et à la suite de l'article de Watson et Crick, sont publiées les données expérimentales de Maurice Wilkins (Wilkins, 1953) et de Rosalind Franklin et Raymond Gosling (Franklin, 1953) ayant servi à Watson et Crick pour établir leur modèle (Franklin, 1953; Victor, 2012; Watson, 1953a; Wilkins, 1953).

En 1954, James Watson et George Gamow, un physicien russo-américain, créent un club de « gentlemen » scientifiques, « The RNA Tie Club » afin d'échanger des idées sur la structure de l'ARN et des protéines et de résoudre l'énigme du code génétique afin de comprendre comment l'information génétique contenue dans l'ADN est transmise aux protéines via l'ARN (**fig. 9**). Ce club comprend 20 membres « sélectionnés », parmi lesquels figurent Francis Crick, Erwin Chargaff, Sydney Brenner.... Chacun des membres est désigné par le nom d'un acide aminé d'où seulement 20 membres plus quatre autres avec le nom d'une des bases de l'ADN. Ainsi Watson est « Pro », Crick est « Tyr ». Chaque membre du « RNA Tie Club » reçoit une cravate noire en laine tricotée sur laquelle était dessinée une hélice d'ARN verte et jaune, moyennant la somme de 4\$ ainsi qu'une épingle à cravate avec l'abréviation de l'acide aminé le désignant. Le club se réunit 2 à 3 fois par an dans une atmosphère cordiale et amicale permettant de nouer des liens et amitiés étroites au sein de cette élite scientifique. En dehors de ces réunions, les membres correspondaient, s'adressant pré-tirages d'articles, lettres, suggérant de nouveaux concepts ou idées. Ce club s'avère être un terrain propice aux idées

créatives. La devise du club est « Do or die : or don't try ». Parmi les membres de ce club, six sont « nobélisés » dont James Watson et Francis Crick (Strauss, 2019).



**Figure 9: The "RNA Tie Club".**

**A)** Lettre fondatrice du "RNA Tie Club", club fondé en 1954 par James Watson et George Gamow.

<https://twitter.com/DrKimHarris/status/639768596735000576/photo/1>

**B)** Réunion du "RNA Tie Club", Cambridge, Angleterre, 1955: sur la photo de gauche à droite, Francis Crick, Alexander Rich, Leslie E. Orgel et James Watson. Crick, Rich et Watson portent la fameuse cravate où est dessinée une hélice d'ARN ainsi qu'une épingle à cravate portant l'abréviation de l'acide aminé correspondant à leur nom.

<http://scarc.library.oregonstate.edu/coll/pauling/dna/pictures/portrait-tieclub.html>

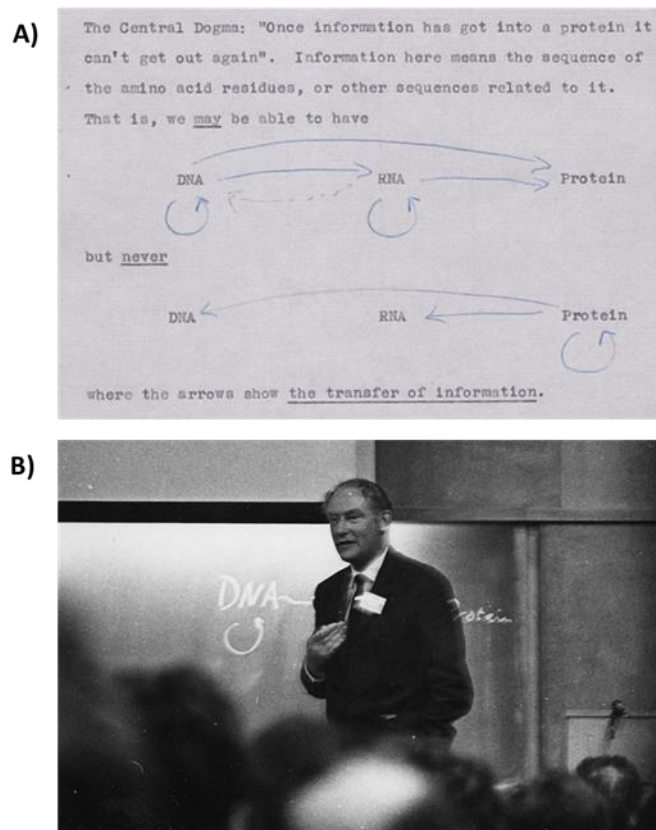
En effet, en 1962, James Watson, Francis Crick ainsi que Maurice Wilkins reçoivent le prix Nobel de Médecine et de Physiologie pour leur découverte de la structure en double hélice de l'ADN. Rosalind Franklin, dont la qualité des clichés de diffraction aux rayons X de l'ADN a permis à Watson et Crick d'élucider la structure de l'acide désoxyribonucléique, décède en 1958 d'un cancer, et n'est pas associée à ce prix !

### Le dogme central de la biologie moléculaire.

En 1956, dans une note personnelle portant sur une réflexion sur la synthèse protéique (Crick, 1956), Crick émet une hypothèse sur le transfert de l'information dans la cellule qu'il intitule « **The Central Dogma** : Once information has got into a protein it can't get out again. » : l'information génétique part de l'ADN (les gènes) vers les protéines par le biais d'un

intermédiaire, l'ARN. Et cette information, une fois transmise dans une protéine, ne peut plus en sortir !

En 1957 au cours d'une conférence donnée à l'*University College* de Londres dans le cadre d'un symposium intitulé « The Biological Replication of Macromolecules », Crick expose pour la première fois ses idées sur le transfert de l'information à l'intérieur de la cellule. Avec Watson, dans leur second article sur la structure de l'acide désoxyribonucléique (Watson, 1953b), ils avaient déjà introduit ce concept « d'information génétique » portée par un code qui est constitué par la séquence précise des bases. Au cours de son exposé, Crick souligne l'existence d'un lien entre la séquence des bases des acides nucléiques et celle des acides aminés dans une protéine, autrement dit entre gène et protéine. Il appelle ce lien, « the flow of information » (Cobb, 2017). Il résume sa définition du Dogme Central par un schéma (**fig 10 A**) qui ne sera jamais publié tel quel, bien que Crick l'utilise régulièrement lors de ces séminaires (**fig. 10 B**). Ce schéma illustre les transferts d'information possibles entre les trois



**Figure 10: The Central Dogma.**

**A)** Premier schéma du « Dogme Central » de Crick, tirée d'une note non publiée rédigée en 1956.

**B)** Intervention de Crick lors du Cold Spring Harbor Symposium, en 1963 : sur le tableau derrière lui, il a dessiné le schéma du Dogme Central.

(à partir de Cobb, 2017).

familles de polymères, ADN, ARN et protéines qui peuvent être divisées en trois groupes. Le premier groupe rassemble ceux pour lesquels des évidences directes ou indirectes existent :

- ADN --> ADN (réplication)
- ADN --> ARN (première étape de la synthèse de protéine)
- ARN --> protéine (2ème étape de la synthèse de protéine)
- ARN --> ARN (réplication de virus à ARN)

Le deuxième groupe indique les transferts que Crick pense être possibles même s'il n'a pas d'évidence à cette époque :

- ARN --> ADN (structurellement possible même si pas de preuve en 1956)
- ADN --> protéine

Enfin le troisième groupe que Crick considère comme impossible (manque de preuve ou absence de mécanisme biochimique) :

- protéine --> protéine
- protéine --> ARN
- protéine --> ADN.

Par ce groupe, Crick signifie que l'information une fois partie de l'ADN pour aller aux protéines via un intermédiaire, l'ARN, ne peut pas ressortir des protéines pour retourner au code génétique : ceci est le fondement du Dogme Central (Cobb, 2017 ; Crick, 1956).

Crick détaillera et apportera des précisions dans son article « Central Dogma of molecular biology » paru dans *Nature* le 8 août 1970 (Crick, 1970) faisant suite à un article critique sur le Dogme Central, non signé, publié deux mois avant et considérant « ... que le Dogme Central, énoncé par Crick en 1958 et qui est depuis la clef de voûte de la biologie moléculaire depuis lors, est susceptible de s'avérer être une simplification excessive considérable ... » (Anonymous, *Nature*, 1970). Cet article fait suite à la découverte par les équipes de Howard Temin et David Baltimore, simultanément et indépendamment, d'une ADN polymérase dépendant de l'ARN dans différents types de virus tumoraux à ARN, le virus du sarcome de Rus et le Virus de la leucémie de la souris de Rauscher (R-MLV). Ainsi le transfert de

l'information de l'ARN vers l'ADN est possible : l'ARN viral infectant est transformé grâce à cette ADN polymérase virale en une copie d'ADN qui sert de modèle pour synthétiser l'ARN viral (Baltimore, 1970 ; Temin, 1970). Irrité par la publication de cet article anonyme, Crick défend avec insistance qu'il n'a jamais soutenu que le transfert ARN --> ADN était impossible. Dans son article Nature de 1970, il propose une nouvelle classification en trois groupes appelés transferts généraux, transferts spéciaux et transferts inconnus

- Transferts généraux qui ont lieu dans toutes les cellules :

- ADN --> ADN
- ADN --> ARN
- ARN --> protéine

- Transferts spéciaux « ... qui n'ont pas lieu dans la plupart des cellules mais dans des circonstances spéciales... » :

- ARN --> ARN
- ARN --> ADN
- ADN --> protéine

Les deux premiers se rapportent lors d'infection de cellules par des virus quant au dernier, Crick admet qu'il n'y a pas d'évidence.

- transferts inconnus que le Dogme Central postule qu'ils ne se produisent jamais :

- protéine --> protéine
- protéine --> ADN
- protéine --> ARN

Ainsi les arguments de Crick sont toujours valables dans la mesure où la synthèse protéique dépend de la séquence d'acide nucléique et une fois que l'information génétique a intégré la protéine, elle ne peut pas modifier la séquence d'ADN (Cobb, 2017 ; Crick,1970).

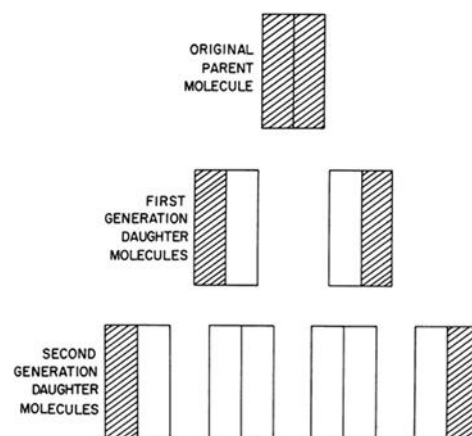
L'utilisation par Crick du mot « dogme » a pu porter à confusion, semer le trouble. En effet, un dogme est une affirmation considérée comme fondamentale, incontestable, intangible. Un dogme n'est pas spéculatif. Crick en énonçant son « dogme » a proposé une relation entre les trois familles de polymères, ADN, ARN et protéines, basée à l'époque en 1956, sur certaines évidences expérimentales mais aussi sur des hypothèses et des spéculations.

La découverte de la structure en double hélice de l'ADN par Watson et Crick en 1953 ainsi que l'énoncé du dogme central de la biologie moléculaire par Crick marquent un tournant majeur dans l'histoire de la biologie moléculaire et sont à l'origine des avancées dans la compréhension de la transmission de l'information dans les systèmes vivants.

Mais un certain nombre de questions demeurent sans réponse : comment l'ADN se réplique-t-il ? Quel est ce code génétique, pressenti par Schrödinger, permettant de contenir toute l'information permettant l'hérédité et comment est-il décrypté pour transférer l'information contenue dans les gènes aux protéines ? Comment une information située dans une chaîne très longue, l'ADN, constituée par un alphabet de 4 lettres, peut produire des chaînes plus ou moins longues, les protéines, constituées par un alphabet de 20 lettres, les acides aminés ?

### La réplication.

En 1958, Matthew Meselson et Franklin Stahl confirment l'hypothèse de Watson et Crick sur la réplication de l'ADN comme un processus semi-conservatif. Ils font croître des bactéries dans un milieu nutritif contenant l'isotope lourd  $N^{15}$  de l'azote puis les bactéries sont transférées dans un milieu avec l'isotope naturel  $N^{14}$  de l'azote. Ils suivent l'incorporation des différentes sources d'azotes dans l'ADN puis la distribution des ADNs au cours des différentes



**Figure 11: Représentation schématique du modèle de réplication semi-conservatif de Meselson et Stahl.**

Suite à l'incorporation d'azote lourd  $N^{15}$ , Meselson et Stahl suivent sa répartition au cours des divisions successives. L'azote de chaque molécule d'ADN est divisé équitablement entre 2 sous-unités. À la duplication suivante, chaque molécule fille reçoit une de ces sous-unités. Les sous-unités sont conservées au cours des duplications successives.

(à partir de Meselson, 1958).



divisions. Les conclusions de leurs expériences (**fig. 11**) rejoignent et confirment les hypothèses de Watson et Crick d'un modèle de **réplication semi-conservatif** (Meselson, 1958).

### **Les ribosomes, lieu de synthèse des protéines.**

A cette période, fin des années cinquante, il est généralement admis que les acides nucléiques contrôlent la synthèse des protéines et plus spécifiquement que l'ordre des acides aminés dans une chaîne polypeptidique est déterminé par l'ordre des nucléotides de l'acide nucléique. Crick est convaincu que la synthèse protéique n'implique pas directement l'ADN mais requiert l'ARN et qu'elle a lieu dans le cytoplasme. La notion que l'ARN est impliqué dans la synthèse des protéines remonte aux expériences pionnières de Brachet et Caspersson dans les années 40 qui ont montré que l'ARN est principalement dans le cytoplasme des cellules où la synthèse protéique a lieu et que dans les cellules synthétisant activement des protéines les niveaux d'ARN augmentent (Boivin, 1947 ; Cobb, 2015 ; Watson, 1963). En 1955, Brachet et d'autres constatent que des cellules sans noyau continuent à produire des protéines pendant plusieurs jours en absence d'ADN (Brachet, 1955). Claude puis Brachet montrent que l'acide ribonucléique rentre dans la composition de granules cytoplasmiques, **les microsomes**, qui représentent le lieu de la synthèse des protéines (Boivin, 1947 ; Claude, 1943). En 1958, lors d'un symposium de la société de Biophysique, le nom de **ribosome** est suggéré pour remplacer microsome, particule microsomale ou encore particule ribonucléoprotéique (Genuth, 2018). Zamecnik, utilisant des acides aminés radioactifs, affirme que ces particules ribonucléoprotéiques cytoplasmiques sont le site d'incorporation initiale des acides aminés libres dans les protéines et cette incorporation dans les ribosomes est une étape irréversible dans la synthèse protéique qui a lieu dans le cytoplasme sans intervention directe de l'ADN localisé dans le noyau (Littlefield, 1955). De nombreux travaux ont permis de déterminer la structure des ribosomes qui sont constitués de 2 sous unités, une grande et une petite : 30S et 50S pour les bactéries et 40S et 60S pour les mammifères. Ces unités sont constituées d'ARN, **les ARNs ribosomiques**, qui ne sont pas sensibles à l'action des ribonucléases car protégés dans les ribosomes, ainsi que de très nombreuses protéines n'ayant qu'un rôle structural (à l'époque plus d'une vingtaine avait été dénombrées à la suite d'électrophorèse sur starch-gel) pour constituer un complexe ribonucléoprotéique. Ces ribosomes ont la même

composition en sous-unités quel que soit la source cellulaire (Watson, 1963). En 1958, Crick propose un modèle de la circulation de l'information génétique qui est connu comme l'hypothèse « un gène – un ribosome – une protéine » où chaque ribosome porte l'information génétique pour coder une seule protéine. Ce modèle suggère que la cellule contient des milliers de ribosomes différents, chacun étant affecté à la production d'une seule protéine (Genuth, 2018). L'ARN présent dans le ribosome serait l'image cytoplasmique de l'ADN dans le noyau.

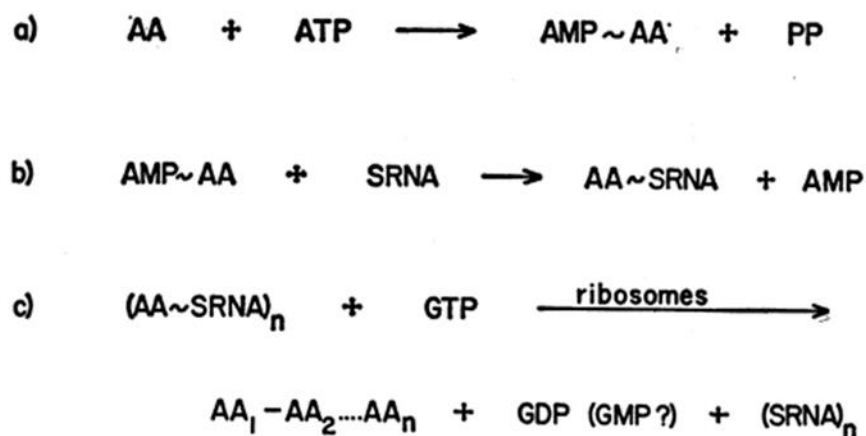
### **Le début de l'*in vitro* : les premiers systèmes acellulaires.**

Parallèlement à ces travaux et pour pouvoir étudier le processus de synthèse protéique, Paul Zamecnik met au point un système *in vitro* de synthèse des protéines « sans-cellule » à partir d'extraits d'une bactérie du colon, *Escherichia coli* contenant ADN, ARN messagers, ribosomes, enzymes ... et complétés par une source d'énergie, principalement de l'ATP. Ce système permet d'incorporer rapidement des acides aminés dans une protéine et cette incorporation peut être suivie par l'utilisation d'acides aminés radioactifs (principalement du carbone 14). Grâce à ce système acellulaire, Paul Zamecnik montre que la synthèse de protéine est possible en dehors de toute cellule. Bien que ce système ne soit capable de traduire que de l'ARNm endogène, il fournit de nouvelles approches pour étudier les mécanismes moléculaires de synthèse des protéines (Hoagland, 1958; Lamborg, 1960 ; Littlefield, 1955 ; Matthaei, 1961, Nirenberg, 1963).

### **Découverte des ARNs de transfert.**

Tissières, utilisant ce système mis au point par Zamecnik, montre que l'incorporation des acides aminés, donc la synthèse protéique, s'effectue au niveau des ribosomes 70S (Tissières, 1960). Zamecnik et ses collaborateurs cherchent à comprendre comment l'information contenue dans l'ARN est décryptée par les acides aminés pour former les protéines. Ils découvrent un peu par accident que les acides aminés sont accrochés à l'ARN avant qu'ils ne soient liés ensemble pour former une protéine. A leur surprise, l'ARN qui lie les acides aminés ne fait pas partie de la machinerie ribosomale mais il est localisé dans la fraction soluble du cytoplasme et consiste en de relatives petites molécules qu'ils appellent **ARNs solubles** et sont

nommés quelques temps plus tard **ARN de transfert** (ARNt). Chaque acide aminé possède son propre ARNt qui lui est spécifique. Ces composés amino-acyl-ARNs de transferts servent d'intermédiaires pour la formation de liaisons peptidiques au niveau des ribosomes. Ces ARNs de transfert peuvent être isolés et purifiés. S'ils sont incubés avec des ribosomes, les acides aminés se séparent des ARNs de transferts au niveau des ribosomes pour se lier, aux autres acides aminés et former une nouvelle protéine. Hoagland et Zamecnik font deux importantes découvertes sur la formation de ces amino-acyl-ARNs de transferts : 1) tout d'abord l'acide aminé doit être initialement activé en présence d'adénosine triphosphate (ATP) pour former un complexe à haute énergie constitué par l'acide aminé et l'adénosine monophosphate (AMP), 2) l'acide aminé activé est transféré sur son ARN de transfert où il se fixe à une de ses extrémités (**fig. 12**). Berg et d'autres ont montré que c'est la même enzyme qui active l'acide aminé et catalyse la réaction de transfert sur l'ARN. Il y a 20 enzymes différentes, chacune spécifique pour un acide aminé donné. L'enzyme impliquée dans cette réaction sera identifiée comme **l'acido-acyl-ARNt synthétase** (Berg, 1958 ; Hoagland, 1957, 1958, 1959).



**Figure 12: Étapes enzymatiques de la formation de la liaison AA-protéine.**

Dans une revue publiée en 1963 sur le rôle de l'ADN dans la synthèse protéique, Watson représente les différentes étapes pour la liaison de l'acide aminé à la protéine en cours de synthèse. L'acide aminé doit être initialement activé (étape **a**) pour pouvoir être transféré sur son ARNt spécifique (sARN) (étape **b**) où dans les ribosomes il sera intégré à la protéine en cours d'élaboration (étape **c**). (à partir de Watson, 1963).

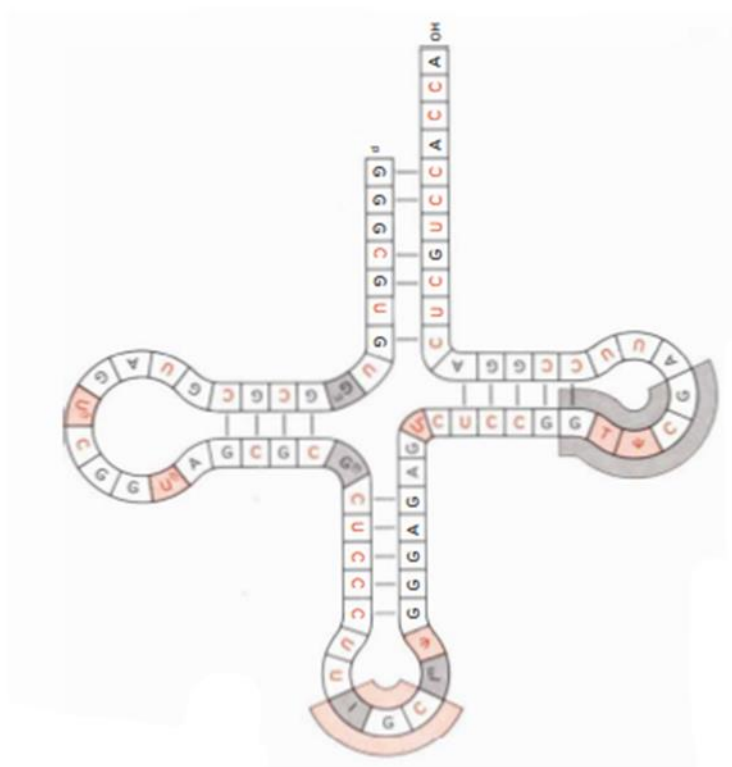
En 1962, François Chapeville et ses collègues ont établi l'importance biologique d'une amino-acylation précise de l'ARN de transfert pour une traduction correcte du code génétique c'est-à-dire qu'une amino-acylation incorrecte d'un ARN de transfert entraînerait l'incorporation erronée d'un acide aminé dans une protéine avec les conséquences que l'on peut imaginer.

En effet, François Chapeville et ses collègues ont montré que les propriétés de codage de l'ARN de transfert ne sont pas déterminées par l'acide aminé qu'il porte mais par l'interaction de l'ARNt-amino-acylé avec l'ARN messager. A partir d'ARN de transfert de cystéine purifié, la cystéine est convertie en alanine par une réduction de son groupement thiol par le nickel de Raney. Ainsi cet ARNt modifié porte l'alanine mais avec une propriété de codage caractéristique de la cystéine. Chapeville & al utilisent un polymère poly-UG qui porte le code pour la cystéine et non celui de l'alanine et ils constatent l'incorporation de l'alanine. Ainsi, la matrice poly-UG reconnaît la spécificité de l'ARN de transfert plutôt que celle de l'acide aminé. La cystéine fixée à son ARN de transfert peut être modifiée en un autre acide aminé mais les propriétés de codage demeurent inchangées (Chapeville, 1962 ; Geigé, 2006).

### **Première séquence nucléotidique.**

En raison du rôle important des ARNs de transfert dans la synthèse protéique et des questions concernant les mécanismes par lesquels les acides nucléiques transfèrent l'information, Holley et ses collaborateurs ont entrepris de déterminer la structure des ARNs de transfert. Pour purifier les ARNs de transfert, ils utilisent une technique de fractionnement à contre-courant qui est basée sur le fait que des molécules similaires de structure différente présentant des solubilités légèrement différentes peuvent se répartir entre deux liquides non miscibles. Cette technique du contre-courant peut être mécanisée de façon à ce que le mélange de molécules est divisé des centaines ou des milliers de fois, tandis que les solvants non miscibles s'écoulent les uns après les autres l'un l'autre dans un schéma à contre-courant. Par cette technique, il était relativement facile de séparer les différents ARNs de transfert, cependant cela engendre une instabilité des ARNs due à la contamination par des nucléases. Après avoir adapté et amélioré cette technique, ils ont purifié trois ARNs de transfert spécifiques pour l'alanine, la valine et la tyrosine (Apgar, 1962 ; Holley, 1966). Les analyses des digestions de ces ARNs de transfert par la ribonucléase pancréatique ont indiqué que ces trois ARNs de transfert pour l'alanine, la valine et la tyrosine ont des structures sensiblement différentes que ce soit au niveau de la composition en nucléotides ou du profil de digestion obtenu par chromatographie (Holley, 1961). Ils ont déterminé la séquence nucléotidique de l'ARN de transfert de l'alanine par l'analyse des fragments obtenus par la digestion complète de l'ARN avec deux ribonucléases. La première ribonucléase est une ribonucléase

pancréatique qui coupent après les nucléotides pyrimidiques qui sont la cytosine et l'uracile. La seconde est la ribonucléase takadiastase T1 qui clive après une guanine. La combinaison des résultats obtenus après digestion par ces deux ribonucléases ont permis de fournir suffisamment d'information pour établir la séquence nucléotidique complète de l'ARN de transfert de l'alanine. Ainsi en 1965, Robert Holley et ces collaborateurs déterminent pour la première fois la structure chimique d'un acide nucléique : après 7 ans de travail, ils obtiennent la première séquence nucléotidique de l'ARN de transfert de l'alanine qui consiste en une seule chaîne d'ARN composée de 77 nucléotides dont 9 nucléotides non usuels (Holley, 1965a, b, c, 1966). Quel travail et quelle persévérance !!!! Dans son article de 1966 dans *Scientific American*, Holley propose différents modèles représentant comment la chaîne linéaire de la molécule de l'ARN de transfert pour l'alanine peut être repliée. Un des modèles proposés en forme de feuille de trèfle ressemble à l'organisation de la structure 2D de l'ARN de transfert que l'on représente maintenant (**fig. 13**) (Holley, 1966).



**Figure 13: Représentation en feuille de trèfle de l'ARNt-Ala par Holley.**

Dans un article d'Holley dans *Scientific American*, publié en 1966, la séquence de l'ARNt de l'alanine est notamment représentée sous forme de feuille de trèfle. Le triplet C-G-I, entouré en rose, est présumé être l'anticodon.

(à partir de Holley, 1966).

Dès 1955, suite à une discussion avec Sydney Brenner, Crick dans une note adressée aux membres du « RNA Tie Club » et jamais publiée, avait formulé « the adaptor hypothesis » où il prédit l'existence de molécules adaptatrices et d'enzymes pour chacun des vingt acides aminés: les acides aminés n'interagissent pas directement avec l'ARN mais avec des petites molécules adaptatrices spécifiques, l'ARN soluble ou ARN de transfert, à l'aide d'enzymes spécifiques et cette combinaison acide aminé/ARN soluble serait capable d'interagir sélectivement avec les surfaces de liaison hydrogène fournies par les bases puriques et pyrimidiques de l'ARN. Il suggère qu'il y aurait vingt adaptateurs différents, un spécifique pour chaque acide aminé ainsi que vingt enzymes spécifiques pour assembler l'acide aminé à son adaptateur (Crick, 1955).

### **Découverte des ARNs messagers : comment l'information génétique circule de l'ADN aux protéines et les premiers principes de la régulation génétique.**

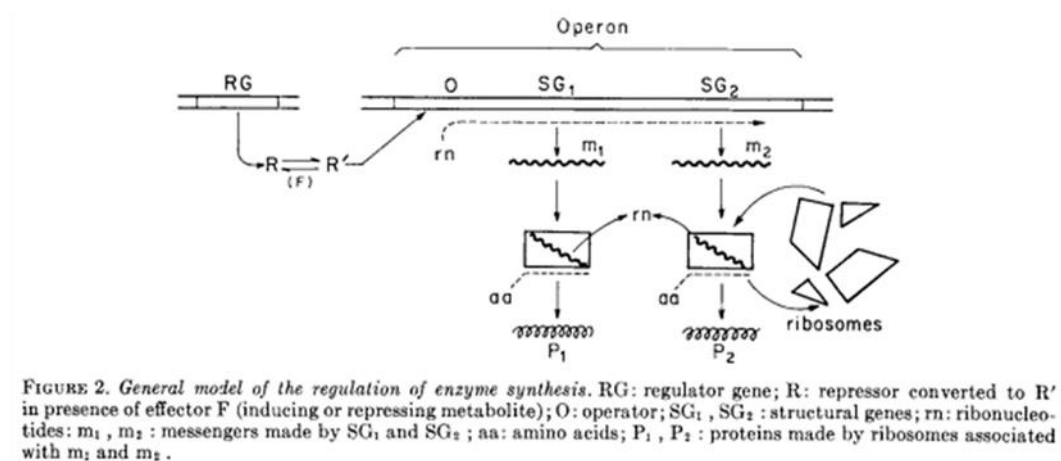
En ce début des années 1960, de nombreuses preuves expérimentales suggèrent que l'information génétique nécessaire pour la structure des protéines est codée par l'acide désoxyribonucléique qui se trouve dans le noyau alors que la synthèse des protéines se déroule dans le cytoplasme au niveau de particules nucléoprotéiques, les ribosomes. L'information génétique étant portée par l'ADN dans le noyau et les protéines n'étant pas synthétisées directement sur l'ADN mais dans le cytoplasme, l'existence d'un support de l'information intermédiaire est nécessaire (Brenner, 1961). A cette époque le rôle de cet intermédiaire est dévolu à l'ARN contenu dans le ribosome comme le stipule l'Hypothèse de Crick « un gène – un ribosome – une protéine ». Cependant très rapidement cette hypothèse est tombée en désuétude lorsque des caractérisations biochimiques du ribosome ont révélé qu'il n'y avait pas de différences majeures de taille dans les ARN ribosomiaux (Crick, 1959; Genuth, 2018).

Les premières expériences suggérant un ARN intermédiaire dans la synthèse protéique proviennent de Volkin et Astrachan qui ont utilisé un système basé sur l'infection de bactéries par des phages, en l'occurrence la bactérie *Escherichia coli* infectée par des bactériophages T2. Seymour Cohen et d'autres avaient précédemment montré que lors de l'infection de la bactérie *Escherichia coli* par un phage T2, la synthèse des protéines de la bactérie est arrêtée

au profit de la synthèse des protéines du phage. Après infection, les protéines néo-synthétisées sont déterminées génétiquement par l'ADN du phage, un grand nombre de nouvelles activités enzymatiques apparaissent et jusqu'à 60% de la synthèse protéique peut être imputable pour les protéines de la tête du phage. Ce système d'infection par un phage est un modèle idéal pour observer la synthèse de nouvelles protéines suite à l'introduction d'ADN de phage et également l'apparition d'ARN phage spécifique (Cohen, 1949; Brenner, 1961; Gros, 1961). En utilisant l'incorporation d'isotope radioactif, notamment le P<sup>32</sup>, Volkin et Astrachan ont montré qu'après l'infection par le bactériophage une fraction mineure de l'ARN a un turn-over rapide et cette fraction d'ARN a une composition nucléotidique qui correspond à celle de l'ADN du phage et est sensiblement différente de celle de l'ADN de la bactérie. Volkin et Astrachan donnèrent le nom de « DNA-like RNA » à l'ARN composant cette fraction mineure (Volkin, 1956, 1958, 2001). Il était clair que cette fraction d'ARN était directement associée à la synthèse protéique mais Volkin et Astrachan n'ont pas été capables de déterminer le mécanisme de cette association (Volkin, 2001). Cet « ADN-like RNA » mis en évidence par ces 2 chercheurs est un candidat sérieux pour jouer le rôle d'intermédiaire dans la transmission de l'information entre l'ADN et les protéines.

A l'Institut Pasteur, Jacques Monod, François Jacob et François Gros essayent d'appréhender comment s'effectue la lecture du message génétique et plus particulièrement la régulation de l'expression génétique en étudiant comment une bactérie dans un milieu de culture va utiliser préférentiellement un sucre plutôt qu'un autre. Ils mettent en évidence un mécanisme d'adaptation montrant qu'en fonction des conditions de culture, des gènes peuvent être activés ou réprimés. Jacques Monod et François Jacob étudient l'activité du gène contrôlant la synthèse d'une enzyme, la  $\beta$ -galactosidase qui métabolise le lactose en glucose et galactose. En absence de lactose ou en présence de glucose, la synthèse de cette enzyme est inhibée. Ils introduisent le concept d'opéron ainsi que la notion de « gènes structuraux » codant pour des protéines effectrices et de « gènes régulateurs » synthétisant des répresseurs ou activateurs de l'activité des gènes structuraux. Pour Monod et Jacob, le gène de la  $\beta$ -galactosidase fait partie d'un groupe de gènes dont l'activité transcriptionnelle de plusieurs de ces gènes structuraux est initiée à un seul point d'initiation du DNA et est coordonnée par un simple « opérateur ». Ces gènes dont l'activité est ainsi coordonnée forme un « opéron » (Jacob, 1961a, b). Leurs travaux les ont conduits à la conclusion que l'expression des gènes codant

pour ces protéines est contrôlée par des protéines régulatrices dont l'activité dépend de la présence du sucre (**fig. 14**). Jacob et Monod s'interrogent comment l'information contenue dans les gènes circule jusqu'aux ribosomes, lieu de synthèse des protéines. Ils émettent l'hypothèse que la transmission de cette information soit effectuée par un intermédiaire, un ARN copie de la séquence d'ADN du gène, qu'ils qualifient de messenger (Actualité, Institut Pasteur, 2021). Peu de temps après, en 1961, cet intermédiaire est mis en évidence conjointement par 2 groupes de chercheurs, comprenant notamment les pasteurien François Jacob et François Gros en séjour aux Etats Unis dans les laboratoires respectifs de Matthew Meselson et James Watson et qui publièrent, à la suite dans *Nature*, la nature exacte de cet intermédiaire : **l'ARN messenger** (ARNm) (Brenner, 1961 ; Gros, 1961). Ayant connaissance des travaux de Volkin et Astrachan (Volkin, 1956, 1958), ces 2 laboratoires entreprennent la détection de cet intermédiaire messenger par des incorporations très courtes de phosphore radioactif, le  $P^{32}$ , suivies par des ultracentrifugations contre un gradient linéaire de saccharose permettant d'analyser la distribution des fractions d'ARN. Le laboratoire de Meselson effectue leurs expériences sur des bactéries infectées par un phage tandis que celui de Watson sur des bactéries non-infectées. Ils excluent les ARNs ribosomiques comme intermédiaire. Ce sont des ARNs qui sont très stables et qui sont d'une remarquable homogénéité que ce soit en taille ou en composition nucléotidique ce qui ne reflète ni la variété de taille des chaînes polypeptidiques ni la variation dans la composition nucléotidique observée dans les ADNs de



**Figure 14: Modèle de l'opéron lactose.**

Dans ce modèle, Jacob et Monod présentent de nombreuses avancées dans la compréhension du transfert de l'information et sur la régulation de l'activité génétique. Ils présentent les différentes étapes de la synthèse protéique et introduisent les notions de protéines régulatrices. (à partir de Jacob, 1961).



différentes espèces de bactéries. L'ARN ribosomique et l'ARN de transfert représente 95% des ARNs d'*Escherichia coli* (respectivement 80-85% et 10-15%). Ainsi l'ARN messager ne représenterait que quelques pourcentages. Si l'ARN messager a un turn-over rapide, comme le suggère les expériences originales d'Hershey, alors une grande part de la fraction d'ARN néo-synthétisé doit être le messager et en regardant la synthèse d'ARN durant un temps très court alors la plupart de l'ARN nouvellement synthétisé devra être le messager même si cette fraction ne représente que quelques pourcents.

Les deux laboratoires arrivent aux mêmes conclusions : ils ont mis en évidence un ARN messager à courte durée de vie qui se fixe sur les ribosomes des bactéries infectées ou pas. Cet ARN messager est physiquement distinct des ARNs ribosomiques ou de transfert. Les ARNs messagers de bactéries non-infectées se comportent comme les ARNs spécifiques de phage après une infection, que ce soit au niveau des constantes de sédimentation ou de leur liaison aux ribosomes à des fortes concentrations de magnésium. Ces ARNs néo-synthétisés sont localisés dans les ribosomes existants et sont hétérogènes en taille. De plus, dans les bactéries infectées par des phages, les ARN messagers ont une composition nucléotidique qui correspond à celle de l'ADN de phage et seraient une simple copie de gène.

Ainsi ces deux laboratoires séparés par quelques milliers de kilomètres, l'un sur la côte est et l'autre sur la côte ouest des Etats Unis, réalisèrent, sans se concerter, des expériences très similaires qui aboutirent aux mêmes conclusions sur le transfert de l'information génétique : l'ARN messager est bien l'intermédiaire entre le gène et la protéine. L'ARN ribosomal n'est pas l'intermédiaire qui porte l'information du gène à la protéine et les ribosomes sont des structures non spécialisées qui reçoivent l'information génétique à partir du gène sous la forme d'un intermédiaire instable, l'ARN messager. Il existe autant de types d'ARNs messagers qu'il existe de types de gènes et de protéines distinctes (Brenner, 1961 ; Gros, 1961).

Les découvertes de Jacques Monod et François Jacob associés à André Lwoff sur le contrôle génétique de la synthèse des enzymes et des virus leur valurent, en 1965, l'attribution du prix Nobel de physiologie ou de médecine. Les trois chercheurs de l'Institut Pasteur ont ainsi découvert que nos gènes ne sont pas exprimés de manière constante au fil du temps, mais qu'ils sont régulés – c'est-à-dire activés ou réprimés – très finement, pour répondre aux besoins de notre organisme. Les travaux de ces trois pasteurien ont abouti à des découvertes

fondatrices pour la biologie moléculaire moderne et ont ouvert la voie aux études sur le contrôle de l'expression des gènes.

### La problématique du codage de l'information génétique.

Le problème du codage dans les systèmes biologiques pour transmettre l'information génétique contenue dans les gènes jusqu'aux protéines peut être abordé sous deux aspects : 1) le mécanisme par lequel l'information contenue dans le matériel génétique, c'est à dire les gènes, est lue et traduite pour produire une protéine fonctionnelle, 2) la correspondance entre l'information contenue dans l'ADN sous forme d'une longue chaîne de nucléotides et une information sous forme de chaîne plus ou moins longue d'acides aminés formant les protéines ou comment un alphabet de 4 lettres peut être traduit en un alphabet de 20 lettres : décrypter le code génétique.

Au début des années 1960, par rapport à ce premier aspect, un certain nombre de découvertes, d'avancées technologiques permettent d'appréhender les mécanismes de transfert de l'information des gènes aux protéines : que ce soit le développement de système de synthèse protéique acellulaire et d'ARN permettant d'expérimenter les prédictions théoriques, le mécanisme de réplication, de transcription (découverte des ARN messagers, de diverses enzymes, ...) de traduction (découverte des ARNs de transfert, des ribosomes, ...), les premiers pas sur la compréhension de la régulation de l'expression des gènes ... En ce qui concerne le deuxième aspect, le décryptage du code génétique, de nombreux travaux basés uniquement sur une réflexion théorique ont été menés, dans les années précédentes, pour répondre à un certain nombre de questions afin de définir la nature du code génétique : **combien de nucléotides déterminent un acide aminé ? comment le langage de l'ADN écrit dans une séquence de quatre nucléotides peut être traduit en un langage de protéine avec vingt acides aminés ? Le code est-il chevauchant ou non ?** A la suite de ces réflexions menées notamment par Brenner et Crick, le consensus consistait en un code qui est défini par 20 triplets, chaque triplet codant pour un acide aminé, dans un mode non-chevauchant. En 1960, ces triplets ont été nommés **codon** par Brenner (Brenner, 1957 ; Crick, 1957 ; Szymanski, 2017).

### **Overlapping, or not overlapping, that is the question.**

L'évidence expérimentale de la nature non-chevauchante du code génétique est fournie en 1960 par des expérimentations sur le virus de la mosaïque du tabac (TMV), virus à ARN simple brin, par les groupes de Wittmann et Fraenkel-Conrat. Wittmann traite l'ARN du virus de la mosaïque du tabac (TMV-RNA) par l'acide nitreux qui convertit certaines bases en d'autres, par exemple l'uracile en cytosine afin d'obtenir des mutations dans l'ARN du virus et d'observer les conséquences sur la séquence protéique. Il constate qu'une simple mutation dans le TMV-RNA entraîne la substitution d'un seul acide aminé dans la séquence de la protéine virale. Les travaux de Wittmann et Fraenkel-Conrat permettent d'affirmer que le code n'est pas chevauchant car en effet si le code était chevauchant, la mutation d'un seul nucléotide aurait entraîné la substitution de deux ou trois acides aminés (Tsugita, 1960 ; Wittmann, 1960 ; Crick, 1961).

### **Premières réponses aux questions sur le code génétique ...**

En 1961, Crick rapporte des expériences de génétique sur la région rII du bactériophage T4. Il traite ce phage avec l'acridine qui possède des propriétés d'intercalation au niveau des molécules d'ADN et qui en s'incorporant dans l'ADN du phage T4 ajoute ou supprime une ou plusieurs bases. La région rII possède 2 gènes adjacents A et B. Ce phage croit sur des bactéries *Escherichia coli* (*E. coli*) B ou K12 mais la perte de fonction d'un des gènes empêche ce phage de pousser sur *E. coli* K12. L'analyse de ces expériences de recombinaison de divers mutants de ce phage, notamment de triples mutants poussant sur *E. coli* K12 et associée à d'autres travaux de scientifiques, suggère à Crick que le code génétique pourrait avoir les caractéristiques suivantes. 1) un groupe de 3 bases code pour un acide aminé, 2) se basant sur les travaux de Wittmann et Fraenkel-Conrat le code n'est pas de type chevauchant. 3) la séquence des bases est lue à partir d'un point de départ fixe. Si ce point de départ est déplacé d'une base, alors la lecture des triplets est décalée et devient incorrecte, 4) le code est dégénéré c'est-à-dire qu'il y a plus d'un triplet qui code pour un acide aminé.

### Le casse du siècle : le craquage du code génétique.

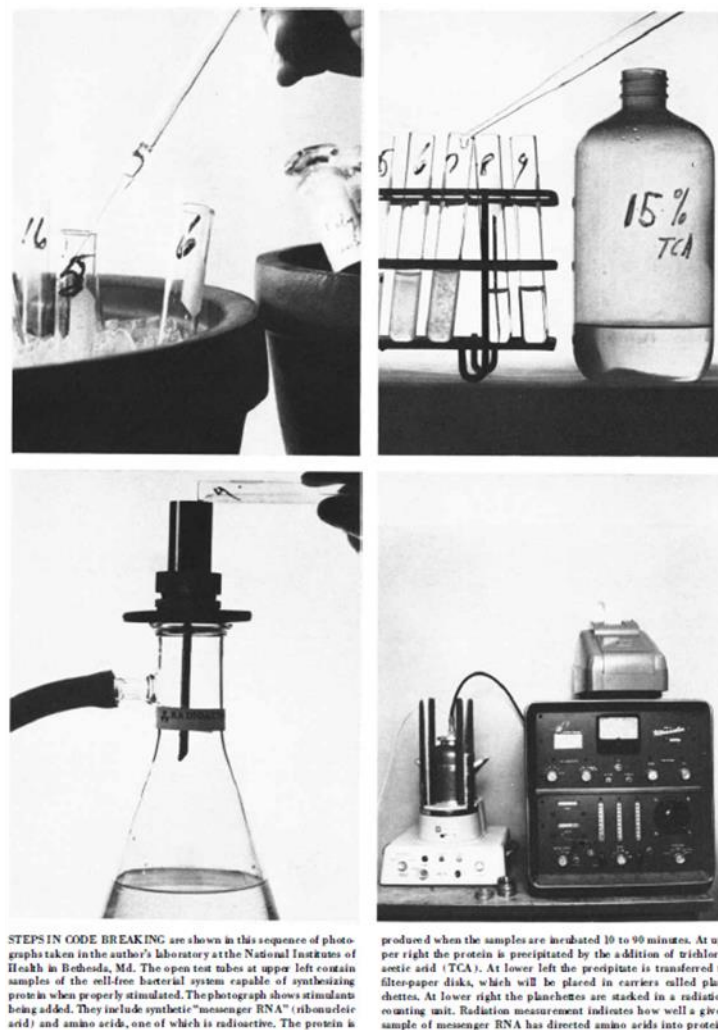
Heinrich Matthaei et Marshall Nirenberg s'intéressent à la synthèse protéique et notamment au rôle de l'ARN messager en utilisant un système acellulaire mis au point par Paul Zamecnik (Lamborg, 1960). Se basant notamment sur les travaux de Zamecnik et ceux de Tissières (Tissières, 1961), Nirenberg et Matthaei traitent leurs extraits acellulaires par une désoxyribonucléase (DNase), enzyme qui dégrade spécifiquement l'ADN, et constatent que la synthèse protéique n'est pas inhibée immédiatement mais une fois que tout l'ARN messager est déplété. Mais avec un tel système acellulaire, de très faible quantité de protéines sont produites à partir de l'ARN endogène. L'ajout d'ARN messager permet de stimuler la synthèse protéique. Pour contrecarrer ces problèmes, ils décident de préparer un ARN synthétique contenant seulement de l'uracile en utilisant une enzyme, la polynucléotide phosphorylase, découverte en 1955 par Marianne Grunberg-Manago dans le laboratoire de Severo Ochoa, qui permet de synthétiser des polymères en absence de matrice d'ADN (Grunberg-Manago, 1955). Elle permet de lier des bases entre elles. Un polynucléotide contenant seulement de l'uracile, l'acide polyuridinique ou poly-U, est obtenu et ajouté à leur système acellulaire en présence d'un mélange de 20 acides aminés dont un acide aminé est marqué avec du carbone 14. Ils obtiennent un polymère composé uniquement de phénylalanine. Leurs résultats indiquent qu'un polynucléotide constitué que d'uracile contient toute l'information nécessaire pour la synthèse d'une protéine contenant uniquement l'acide aminé phénylalanine. Ce résultat leur fournit la preuve que le code pour la phénylalanine est une séquence composée uniquement d'uracile. Ainsi, un ou plusieurs nucléotides d'uracile code pour la phénylalanine. Le polynucléotide poly-U fonctionne comme un ARN messager.

#### **C'est le premier craquage du code génétique !**

Matthaei et Nirenberg confirment leur découverte avec l'utilisation d'un polynucléotide constitué uniquement de cytosine qui code pour un polymère composé uniquement de proline. Le code pour la proline est constitué d'une ou plusieurs cytosines. C'est le début du déchiffrement du code génétique. Nous sommes en 1961 (Nirenberg, 1961a, b, 1963)! Ainsi ce système acellulaire capable de synthétiser une protéine à partir de préparation d'ARN synthétique fournit un moyen simple pour décrypter l'ensemble du code génétique.

## La table du code génétique : depuis 1965 on utilise toujours la même.

La plupart des codons du code génétique ont été identifiés par les groupes de Nirenberg et de Khorana, utilisant un système de synthèse protéique acellulaire mais chacun développant sa propre approche. Nirenberg a développé une technique de liaison aux ribosomes qui consiste en l'utilisation d'ARN trinuécléotidiques synthétiques qui sont capables de se lier aux ribosomes et qui jouent le rôle de codon pour permettre la liaison de l'ARN de transfert spécifique marqué par un acide aminé radioactivement. Après fixation de l'ARN de transfert, l'acide aminé radioactif reste fixé aux ribosomes. L'ensemble est déposé sur un filtre puis lavé pour éliminer les ARN de transfert non fixés et finalement la radioactivité est déterminée. Dans l'article de Nirenberg, publié dans *American Scientific* en 1963, une séquence de photographies illustre cette technique (**fig. 15**) (Nirenberg, 1963, 1964). Ainsi Nirenberg a pu



**Figure 15: Série de photographies illustrant les techniques utilisées pour identifier les premiers codons.**  
(à partir de Nurenberg, 1963).

identifier, entre autres, les codons pour la phénylalanine, la lysine et la proline. De plus, il a remarqué que les triplets phosphorylés en 5' sont plus actifs comme échantillon que ceux sans phosphate en 5' (Nirenberg, 1964). Nirenberg a effectué les mêmes expériences avec des ARNs dinucléotidiques mais il constate que seuls les trinucléotides, et non les dinucléotides, dirigent spécifiquement la liaison aux ribosomes des ARNs de transfert. Cette observation fournit une preuve expérimentale directe que **le code pour un acide aminé est un triplet**, confortant les hypothèses et déductions de travaux précédents notamment ceux de Crick (Crick, 1961 ; Nirenberg, 1964).

Parallèlement, Khorana a développé une méthode basée sur l'utilisation de polynucléotides synthétiques contenant des répétitions de séquence de di ou tri –nucléotides qu'il utilise comme ARN messenger dans un système d'incorporation d'acide aminé acellulaire. L'utilisation de séquence ApApGp répétée dirige la synthèse d'homopolypeptides de lysine, acide glutamique et d'arginine. L'utilisation de polynucléotides, polyUC, polyUG, polyAC, polyAG permet la synthèse, respectivement, de copolypeptides de sérine et de leucine, de valine et de cystéine, de thréonine et d'histidine et d'arginine et d'acide glutamique. Combinant son approche et celle de Nirenberg, Khorana présente, dans son article de 1965 publié dans *PNAS*, la première table du code génétique (**fig. 16**). La méthode de présentation des séquences du code génétique a été imaginée par Crick qui autorisa Khorana à la publier. C'est d'ailleurs

CODON ASSIGNMENTS FROM POLYPEPTIDE SYNTHESIS AND/OR STIMULATION OF  
AMINOACYL-SRNA BINDING TO RIBOSOMES\*

1st	U	C	2nd		G	3rd
	<u>Phe</u>	<u>Ser</u>		<u>Tyr</u>	<u>Cys</u>	U
U	<u>Phe</u>	<u>Ser</u>		<u>Tyr</u>	<u>Cys</u>	C
	<u>Leu</u>	<u>Ser</u>			<u>Try</u>	A
	<u>Leu</u>	<u>Pro</u>		<u>His</u>	<u>Arg</u>	G
C	<u>Leu</u>	<u>Pro</u>		<u>His</u>	<u>Arg</u>	U
	<u>Leu</u>	<u>Pro</u>		<u>Gln</u>	<u>Arg</u>	C
	<u>Ileu</u>	<u>Thr</u>		<u>Gln</u>	<u>Arg</u>	A
A	<u>Ileu</u>	<u>Thr</u>		<u>Asn</u>	<u>Ser</u>	G
	<u>Met</u>	<u>Thr</u>		<u>Asn</u>	<u>Ser</u>	U
	<u>Met</u>	<u>Thr</u>		<u>Lys</u>	<u>Arg</u>	C
	<u>Met</u>	<u>Thr</u>		<u>Lys</u>	<u>Arg</u>	A
	<u>Val</u>	<u>Ala</u>		<u>Lys</u>	<u>Arg</u>	G
G	<u>Val</u>	<u>Ala</u>		<u>Asp</u>	<u>Gly</u>	U
	<u>Val</u>	<u>Ala</u>		<u>Asp</u>	<u>Gly</u>	C
	<u>Val</u>	<u>Ala</u>		<u>Glu</u>	<u>Gly</u>	A
	<u>Val</u>	<u>Ala</u>		<u>Glu</u>	<u>Gly</u>	G

\* This method of presentation of the codon sequences originated with Dr. F. H. C. Crick (private communication).

**Figure 16: la première table du code génétique.**

Cette méthode de représentation de la table du code génétique a été imaginée par Francis Crick qui autorisa Khorana à la publier. Il manque quelques codons notamment les 3 codons stop. En 1967, sera identifié le dernier codon manquant.

(à partir de Söll, 1965).

toujours cette représentation que l'on utilise encore aujourd'hui, (Söll, 1965). La table du code génétique est juste et presque complète, manque essentiellement les 3 codons stop. La même année, en 1965, Brenner identifie les codons UAA et UAG comme des codons non-sens servant de signaux de terminaison de la traduction (Brenner, 1965). Le dernier codon à être identifié sera le troisième codon stop, UGA, deux plus tard, toujours par Brenner (Brenner, 1967). En 1966, Thach et ses collaborateurs identifient le codon AUG comme le codon initiateur de la synthèse polypeptidique qui s'effectue en direction de la partie 3' (Thach, 1966).

En 1968, pour « l'interprétation du code génétique et ses fonctions dans la synthèse protéique », Nirenberg et Khorana ont reçu le prix Nobel de médecine ou de physiologie. Ils partagent ce prix avec Robert Holley pour son travail sur les ARNs de transfert.

### En 15 ans, que d'avancée ...

En 1952, à la suite des travaux d'Hershey et de Chase, confirmant ceux de d'Avery, MacLeod et McCarthy, l'ADN est considéré comme étant le support de l'hérédité. L'année suivante en 1953, la structure moléculaire de l'ADN est élucidée et le 25 avril 1953, Watson et Crick publient leur modèle de la double hélice de l'ADN. Bien que cette découverte fit sensation, les mécanismes mis en œuvre pour permettre la synthèse protéique à partir de l'information contenue dans l'ADN ne sont pas élucidés. Une boîte noire demeure entre l'ADN et les protéines.



En quinze ans, depuis la découverte de la double hélice, de nombreuses réponses ont été apportées sur la compréhension des mécanismes fondamentaux dans la transmission de l'information génétique depuis l'ADN en passant par l'ARN et jusqu'à la synthèse des protéines. Les premières notions sur la régulation génique ont également été énoncées. Ainsi au cours de ces 15 années, on a pu comprendre la nature chimique du matériel génétique et comment sa structure moléculaire peut contenir des instructions codées qui peuvent être

"lues" par la machinerie de la cellule responsable de la synthèse des protéines. Ces avancées ont permis de percer cette boîte noire où la lumière de la connaissance a pu pénétrer mettant en évidence l'avancées des connaissances. En 1966, Crick publie un article dans *Scientific American* où il dresse un état de l'art sur le transfert de l'information génétique et dans lequel il considère que **"The story of the genetic code is now essentially complete. One can trace the transmission of the coded message from its original site in the genetic material to the finished protein molecule."** (Crick, 1966). Ainsi en cette fin des années soixante, les principaux principes fondamentaux ont été découverts qui peuvent être schématiquement résumés par le fameux "Dogme Central" émis par Crick.



Si on veut les énumérer rapidement, façon catalogue :

- répllication de l'ADN.
- découverte des différents ARNs : ARN messenger, ARN de transfert et ARN ribosomiques ainsi que de leurs rôles respectifs.
- décryptage du code génétique où comment un alphabet composé de quatre nucléotides peut être traduit en un alphabet de 20 acides aminés à partir de l'ARNm qui est lu par groupe de 3 nucléotides, ou codons, et de manière non chevauchante, constituant soixante-quatre codons différents.
- tableau du code génétique complet avec les trois codons stop plus le codon d'initiation ATG où la formulation de présentation du tableau imaginée par Crick est toujours celle que l'on utilise aujourd'hui.
- synthèse des protéines au niveau des ribosomes où s'effectue la lecture de l'ARN messenger : pour chaque codon correspond un acide aminé qui, transporté par son ARN de transfert spécifique, est ajouté au suivant sur la chaîne polypeptidique naissante. L'ARNt reconnaît son codon spécifique grâce à sa séquence complémentaire ou anti-codon, dans une direction anti-parallèle.
- convention d'écriture que ce soit au niveau des acides nucléiques où la séquence se lit de gauche à droite en commençant le groupe phosphate en 5' et se terminant par un



groupement hydroxyde -OH en 3' ou au niveau protéique où la séquence commence à gauche par la partie N-terminal avec son groupement amine -NH<sub>2</sub> et se termine par la partie C-terminal et son groupement carboxylique COOH.

- découverte de très nombreuses enzymes impliquées dans ces processus, comme l' amino-acyl-ARNt synthétase, la polynucléotide phosphorylase, l'ADN polymérase I, l'ARN polymérase ...

Evidemment beaucoup de choses restent à découvrir ou à approfondir mais en ces 15 années depuis la publication de la double hélice de l'ADN, des avancées considérables ont été faites dans la transmission de l'information génétique et ont été les bases des découvertes ultérieures.

Les années soixante-dix vont voir émerger à partir de travaux recherche fondamentale de très nombreux outils qui permettront d'ouvrir de nouveaux horizons.

Dans les années soixante, de nombreux laboratoires travaillent sur la caractérisation du phénomène de restriction. Une population de bactériophages qui se développe sur une population bactérienne donnée, dite A, est en général incapable de se développer sur une population bactérienne d'une espèce voisine, dite B. On dit qu'il y a restriction de la croissance du bactériophage sur l'hôte B. En 1962, Werner Arber et son étudiante Daisy Dussoix découvrent que l'ADN du bactériophage lambda est dégradé après infection dans une bactérie qui restreint la croissance de ce bactériophage. Arber émet l'hypothèse que l'ADN étranger est coupé en morceaux par des enzymes qui reconnaissent des endroits spécifiques sur l'ADN. Une bactérie ne détruit pas son propre ADN parce que ce dernier est modifié et donc protégé au niveau des endroits spécifiques qui sont reconnus par son enzyme. Les travaux d'Arber et Dussoix posent les bases pour la découverte des enzymes de restriction, car l'hypothèse qu'une enzyme de la bactérie hôte coupe l'ADN du bactériophage sur des sites spécifiques, et que l'ADN de l'hôte est protégé par une modification chimique (méthylation) sera confirmé par les recherches de Urs Kuhnlein, Hamilton Smith et Daniel Nathans (Arber, 1969 ; Springer, 1987). En 1970, Hamilton Smith purifie la première enzyme de restriction à partir de la

bactérie *Haemophilus influenzae* (Smith, 1970). En 1971, Daniel Nathans découpe l'ADN du virus simien 40 ou SV40. (Danna, 1971).

En 1972, Paul Berg réalise le premier clonage en insérant dans l'ADN circulaire du virus SV40, un fragment d'ADN contenant des gènes du phage lambda et l'opéron galactose d'*Escherichia coli* (Jackson, 1972).

En 1978, Werner Arber, Daniel Nathans et Hamilton O. Smith ont reçu le prix Nobel de physiologie/médecine pour « la découverte des enzymes de restriction et leur application aux problèmes de génétique moléculaire ».

Arber a compris le phénomène biologique sous-jacent à la présence des enzymes de restrictions chez les bactéries, Smith a isolé la première enzyme de restriction et Nathans a été le premier à l'utiliser. C'est un peu comme si Arber a pensé à l'outil, Smith l'a fabriqué et Nathans l'a utilisé (Springer, 1987).

Ces années soixante-dix voient l'émergence de la biologie moléculaire moderne.

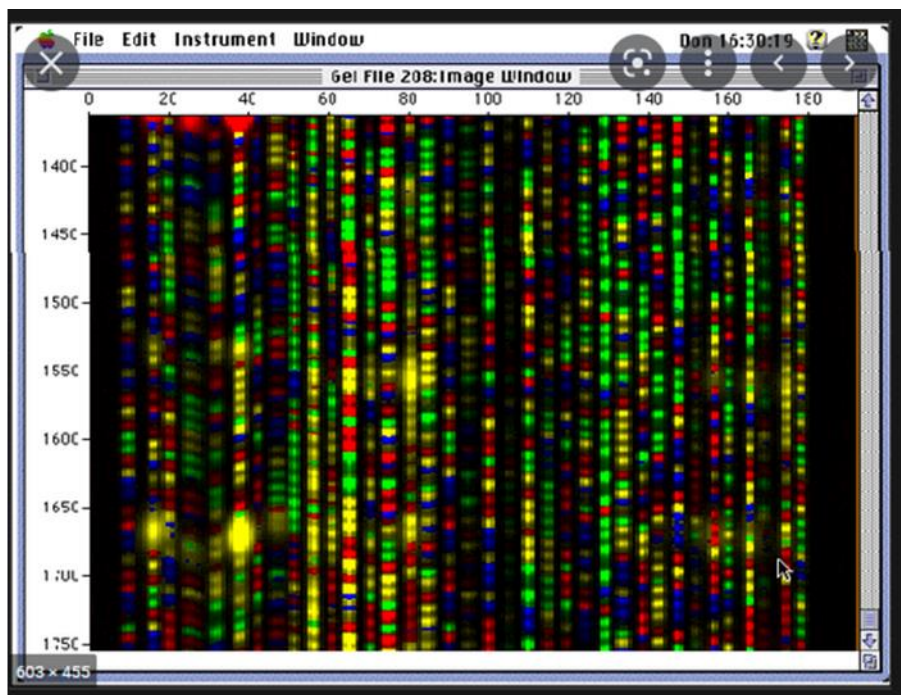
L'année 1977 voit l'arrivée des techniques de séquençage de l'ADN. De manière concomitante, Sanger d'un côté et Maxam et Gilbert de l'autre vont proposer deux techniques pour séquencer les bases constituant l'ADN. Maxam et Gilbert proposent une technique de séquençage fondée sur une dégradation chimique sélective de l'ADN qui utilise les réactivités différentes des quatre bases A, T, G et C, pour réaliser des coupures sélectives (Maxam, 1977) :

- A>G : forte bande en A et faible en G donnera adénine.
- G>A : forte bande en G et faible en A donnera guanine.
- C : donne cytosine.
- C+T : si bande en C et T donne cytosine et si bande qu'en T donne Thymine.

La technique de Sanger (Sanger, 1977) consiste à initier la polymérisation de l'ADN à l'aide d'une amorce oligonucléotide complémentaire à l'ADN à séquencer. Les quatre désoxyribonucléotides (dATP, dCTP, dGTP, dTTP) sont ajoutés, ainsi qu'une faible

concentration de l'un des quatre didésoxyribonucléotides (ddATP, ddCTP, ddGTP ou ddTTP) qui agissent comme des terminateurs de chaîne car une fois incorporés, ils empêchent la poursuite de l'élongation. La réaction est effectuée pour chaque didésoxyribonucléotide. Les différents fragments obtenus sont séparés par électrophorèse. La détection s'effectue en incorporant du soufre radioactif, le  $S^{35}$ . Ainsi en 1978, la séquence nucléotidique de l'ADN du bactériophage G4 est réalisée en utilisant la technique mise au point par Sanger. 5 577 nucléotides sont séquencés (Godson, 1978). Au même moment Sanger et collaborateurs réalisent le séquençage du bactériophage  $\phi$ X174 (Sanger, 1977).

Dans les années 90, la société Applied Biosystems introduit des traceurs fluorescents à la place des marqueurs radioactifs utilisés initialement. Chaque didésoxyribonucléotide a incorporé un marqueur fluorescent spécifique ce qui permet de mélanger les quatre réactions. Les fragments sont séparés par électrophorèse et la détection s'effectue à l'aide d'un laser (**fig. 17**). L'amélioration des techniques de séquençage a déclenché une course des génomes. Le génome d'*Escherichia coli* qui fait plus de 4 millions de bases sera publié en 1997 tout comme celui de *Saccharomyces cerevisiae*, premier génome eucaryote unicellulaire.



**Figure 17: Capture d'écran d'un profil de séquençage, issu d'un séquenceur.**

Chaque couleur correspond à un nucléotide et chaque piste correspond à un échantillon: sur ce gel trente-six échantillons différents sont séquencés.

A la fin des années quatre-vingt-dix, le projet de séquencer le premier génome d'organisme multicellulaire, celui de la drosophile est lancé. Ce projet est réalisé par 2 consortiums : un européen, l'European Drosophila Genome Project, EDGP, qui doit séquencer le chromosome X et un américain, le Berkeley Drosophila Genome Project, le BDGP, qui doit séquencer le chromosome II et III. L'EDGP regroupe une dizaine de laboratoire européen. J'ai eu la chance de participer à cette aventure au sein du laboratoire de David Glover à Dundee durant l'année 1998. Le génome de la drosophile est cloné dans une banque de cosmide où chaque cosmide est séquencé en utilisant la technique du « shotgun » qui consiste à digérer l'ADN par une enzyme de restriction en fragment compris entre 1 et 2kb. Ces fragments sont clonés dans des phages où sont séquencés les fragments insérés. Ainsi, j'ai séquencé, modestement, 3 cosmides soit 150kb. En dehors du côté purement séquençage, j'ai assemblé l'ensemble des séquences obtenues. Ainsi, au fin fond d'une banque de données de l'EMBL doivent dormir mes données de séquence ...

En 1998, la société Applied Biosystems a développé un séquenceur, l'ABI prism 3700, qui permet de séparer les fragments non pas sur gel mais sur des capillaires permettant d'automatiser le séquençage. L'américain Graig Venter crée la société TIGR (*The Institut for Genome Project*) en 1992. Il a comme projet de séquencer le génome humain. Pour « se faire la main », il décide de séquencer le génome de la drosophile. Disposant de plus de 300 séquenceurs ABI Prism 3700, il séquence en 3 mois le génome de la drosophile qui fait  $1,6 \cdot 10^8$  bases. Il est publié le 24 mars 2000 dans *Science* qui est une issue spécial génome de la drosophile (Adams, 2000). Le consortium européen, l'EDGP, publie dans cette issue une analyse au cours de laquelle j'ai participé, sur l'extrémité du chromosome X (Benos, 2000).

Le séquençage des génomes permet l'émergence de nouvelles disciplines comme la protéomique et des analyses « genome wide » qui permettent d'interroger l'ensemble du génome. De nouveaux mots en -ique apparaissent : génomique, transcriptomique ... ainsi que de nouvelles techniques qui interrogent l'ensemble du génome : transcriptome, RNA-seq, Chip-seq ... La bio-informatique, balbutiante au début des années 2000, devient incontournable de nos jours pour permettre d'analyser la quantité d'informations que génèrent ces nouvelles technologies

Je vais m'arrêter là pour ce long avant-propos sur l'ADN. J'ai voulu raconter ma petite histoire de l'ADN car il a été mon compagnon de paillasse depuis de nombreuses années et à travers cette histoire rendre hommage à ces scientifiques qui souvent avec peu de moyens ont permis de faire avancer la connaissance, le savoir. Je suis toujours fasciné et admiratif de la Nature où à partir d'un œuf fécondé faisant quelques dixièmes de millimètre, une drosophile ou un humain est façonné, où dans une si petite cellule, toute l'information est contenue et s'exprime au bon endroit, au bon moment...

**Ma thèse portera sur la caractérisation de structures cytoplasmiques pseudo-nucléaires en utilisant un système de visualisation génomique en temps réel, le système ANCHOR.**

L'introduction bibliographique sera partagée en deux chapitres : le premier portera sur l'organisation du génome eucaryote et le deuxième sur la visualisation de l'ADN : comment peut-on visualiser ce génome de façon global et s'affiner pour permettre la localisation de locus génomique en cellules vivantes ?

La première partie de mes résultats porteront sur la mise en place et la validation du système ANCHOR chez la drosophile. L'utilisation de ce système ANCHOR m'a permis de révéler, lors de la transfection transitoire de plasmides dans des cellules S2 de drosophile, des structures pseudo-nucléaires cytoplasmiques constituées de plasmides. La deuxième partie de mes résultats me permettra de caractériser l'organisation de ces structures cytoplasmiques pseudo-nucléaires en utilisant le système de visualisation de locus génomique, le système ANCHOR.

# **Introduction**

## **Chapitre I : Organisation du génome eucaryote.**

Pour les organismes uni ou pluricellulaires, la cellule constitue l'unité fonctionnelle fondamentale de tous les organismes vivants. Chez les organismes pluricellulaires, les cellules interagissent entre elles de façon coordonnée afin d'assurer le développement et le fonctionnement des tissus/organes qui les composent. Les cellules ont été historiquement divisées en deux grandes familles :

- les cellules procaryotes (comme les bactéries) pour lesquelles le matériel génétique est constitué d'un chromosome unique et circulaire (nucléoïde) baignant dans le cytoplasme et qui n'est pas séparé physiquement du matériel cytoplasmique.
- les cellules eucaryotes dont le matériel génétique, exception faite de l'ADN mitochondrial, est contenu dans un noyau et isolé du cytoplasme par une membrane nucléaire.

Dans cette partie je vais m'attacher à décrire le noyau, les différents types de chromatine, leur niveau de compaction, les différentes marques d'histones associées ainsi que les systèmes existants permettant de suivre la dynamique chromatinienne.

## 1- Le Noyau.

Une des caractéristiques des organismes eucaryotes est qu'ils se définissent par la présence d'un noyau structuré, comprenant l'ADN et responsable de la protection du génome. L'isolement du matériel génétique est assuré par une double bicouche lipidique séparée par une lumière de 40-50 nm appelée l'enveloppe ou membrane nucléaire. Bien que cette dernière se prolonge au sein du réticulum endoplasmique, elle sert de barrière active entre le cytoplasme et le noyau, en jouant un rôle fondamental dans de nombreux phénomènes physiologiques, comme la régulation du trafic de protéines et autres macromolécules (Dingwall, 1992). La régulation de ce trafic est assurée par les pores nucléaires qui permettent la diffusion de petites molécules et l'import/export actif de macromolécules. L'enveloppe nucléaire est composée de deux feuillets membranaires dits internes et externes. La membrane interne est en liaison étroite avec un réseau de filaments intermédiaires composé de différentes lamines, la **lamina**. Il s'agit d'un réseau protéique fibreux qui sert de point d'ancrage à certains éléments du noyau comme l'ADN, sous une forme compactée appelée chromatine, présent à la périphérie du noyau (van Steensel, 2017). La lamina nucléaire



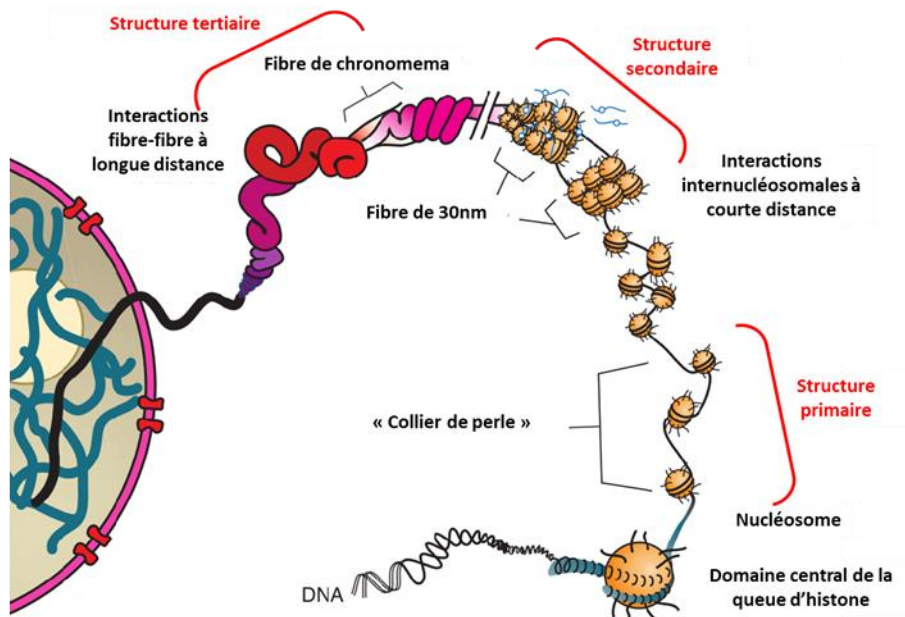
représente donc un support structural pour l'enveloppe nucléaire qui permet de suivre son organisation. La membrane externe est en contact avec le réticulum endoplasmique (RE) présent dans la région périnucléaire du cytoplasme. Une partie du RE est couverte de ribosomes qui assemblent les acides aminés en chaînes protéiques suivant l'information venue du noyau.

Au sein du noyau, des structures d'environ 1 $\mu$ m de diamètre, observables en microscopie optique à fond clair sont identifiables : les nucléoles où se produisent la synthèse, la maturation et l'assemblage des ribosomes (Cerqueira, 2019). Ainsi, les nucléoles sont principalement composés de l'agglomération de gènes codant pour les gènes ribosomiaux, localisés sur différents chromosomes, et des molécules participant à ces processus (petites ribonucléoprotéines (snRNP)...).

## 2- Les chromosomes.

A l'intérieur du noyau, la répartition et la compaction de l'ADN, support du matériel génétique, ne sont pas homogènes. Elles évoluent en fonction des types cellulaires et de leur progression dans le cycle cellulaire. Le challenge majeur pour les cellules eucaryotes est d'intégrer dans le noyau qui mesure quelques microns, une molécule d'ADN pouvant mesurer jusqu'à un deux mètres chez l'Homme (**fig. 18**). La compaction de l'ADN, rendue possible par les nombreuses protéines qui y sont associées, permet de résoudre ce problème (Quénet, 2012).

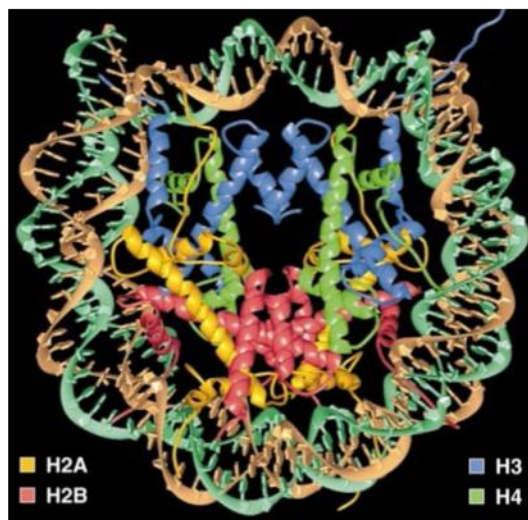
- **Le nucléosome** élément majeur de la chromatine est l'unité de base de la chromatine. Il assure le premier niveau de compaction de l'ADN. Il est constitué de 147 paires de bases d'ADN double-brin enroulé en 1,65 tours autour d'un octamère de protéines d'histones (Luger, 2005). Son cœur comprend 4 types d'histones : H2A, H2B, H3 et H4, organisées en un tétramère de H3-H4 et deux dimères de H2A-H2B (Kornberg, 1999). La chromatine contient une centaine de milliers de nucléosomes reliés entre eux par 20 à 80 paires de bases formant l'ADN internucléosomal et dont l'enchaînement donne à la fibre chromatinienne déroulée un aspect en collier de perles d'environ 10-11 nm de diamètre. Cette structure a été révélée dans les années 70, grâce à des approches de microscopie électronique (**fig. 19** ; Woodcock, 1976).



**Figure 18: Schéma des différentes échelles de compaction de la chromatine.**

Le compactage de l'ADN dans le noyau en interphase (représenté à gauche) se fait par une hiérarchie d'interactions dépendantes des histones qui peuvent être subdivisées en niveaux de structure **primaire**, **secondaire** et **tertiaire**. Les chaînes de nucléosomes constituent l'unité structurelle **primaire**. La formation de fibres de 30 nm par des interactions nucléosome-nucléosome médiées par la queue de l'histone fournit un niveau **secondaire** de compaction, tandis que l'association de fibres individuelles médiée par la queue produit des structures **tertiaires**.

(figure et légende tirée de Horn, 2002).



**Figure 19 : Structure d'un nucléosome.**

Le ruban d'ADN de 146bp est représenté en couleur brune et turquoise, et les huit protéines d'histones sont décrites avec H2A en jaune, H2B en rouge, H3 en bleu et H4 en vert, (figure tirée de Luger, 1997).

- **L’Histone H1**, qui ne fait pas partie des histones nucléosomiques, permet le maintien de la structure nucléosomique en assurant la liaison entre le nucléosome et l’ADN internucléosomique (Thoma, 1979). Elle permet également le rapprochement des nucléosomes de proximité créant ainsi un niveau de compaction supérieur (Hergeth, 2015). La vision standard dépeint une organisation hiérarchique dans laquelle les nucléosomes disposés le long de l’ADN sous forme de « collier de perles », se replient en fibres plus compactes de 30 nm, qui se replient à leur tour en structures d’ordre supérieur de complexité croissante. L’intégralité de ce processus reste encore mal comprise et l’existence même de la fibre de 30nm est remise en question. En effet, de nombreuses études n’ont pas réussi à détecter des fibres de 30 nm dans de nombreux noyaux eucaryotes, où la chromatine semble former des chaînes irrégulièrement pliées avec des caractéristiques en zigzag. Néanmoins, dans certaines cellules différenciées, la présence de fibres de 30 nm bien définies est bien établie, reflétant la dynamique chromatinienne et suggérant un rôle pour de telles structures dans l’inactivation de la transcription (Boopathi, 2020).

Au cours de la division cellulaire, la chromatine se condense à l’extrême pour former des chromosomes. A l’inverse, en interphase, la chromatine est sous une forme décondensée qui occupe une grande surface du noyau. Cependant, dans cette phase du cycle cellulaire, l’état de la chromatine est hétérogène et traditionnellement elle a été divisée en deux catégories reflétant son niveau de compaction : **l’hétérochromatine** et **l’euchromatine**. L’hétérochromatine, constituée en majorité de séquences répétées, est fortement compactée alors que l’euchromatine est plus relaxée (Rosa, 2013). Cependant, cette division historique de l’organisation de la chromatine a été revisitée par des études relativement récentes et subdivisée en cinq catégories en fonction de la composition en protéines associées (complexe Trithorax, Polycomb, etc.) et en types de modifications post-traductionnelles présentes sur la queue des histones (Filion, 2010). Ainsi, la présence des éléments de cis-régulation dans une catégorie chromatinienne et son association avec des modifications post-traductionnelles portées par les histones permettent de prédire leur accessibilité aux facteurs de transcription et donc de contrôler l’expression génique.

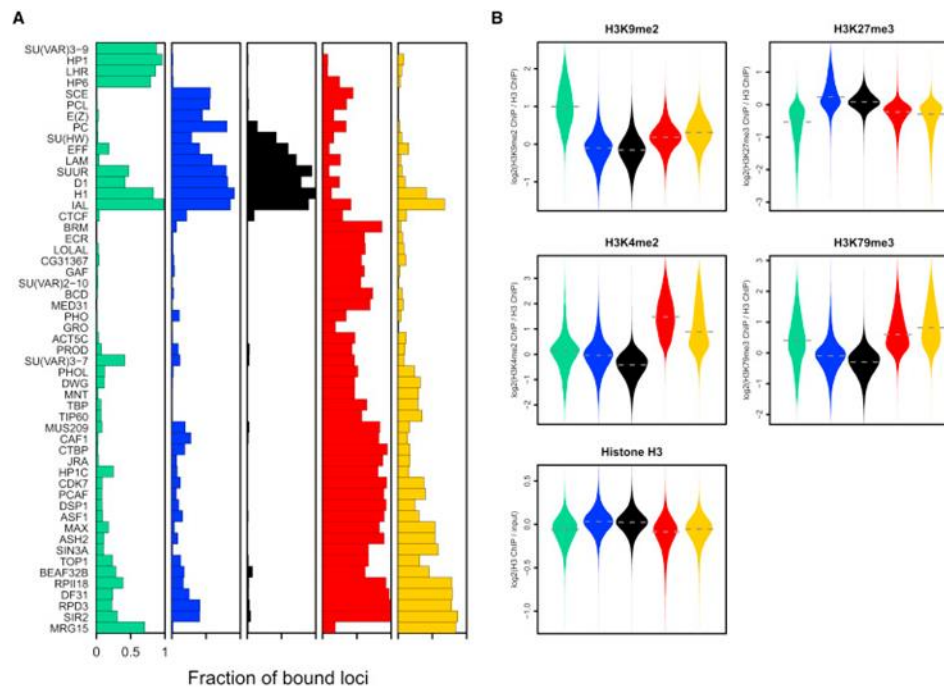
## 2-1 Les différentes couleurs de la chromatine.

En 2010, le laboratoire de van Steensel (Filion, 2010) a entrepris de caractériser les différents états de la chromatine en utilisant les profils de fixation de 56 protéines associées à la chromatine ainsi que ceux de quatre modifications post-traductionnelles des différentes histones: la lysine 4 de l'histone H4 diméthylée (H3K4me2), la lysine 9 de l'histone H3 diméthylée (H3K9me2), la triméthylation de la lysine 27 de l'histone H3 (H3K27me3) ainsi que la triméthylation de la lysine 79 de l'histone H3 (H3K79me3) dans les cellules *Kc167* de drosophile en culture. Les profils de fixation de ces 56 protéines ont permis de définir cinq groupes de chromatine distincts qui constituent les cinq couleurs de la chromatine : **noire**, **bleue**, **verte**, **rouge** et **jaune** (fig 20 ; Filion, 2010). Ces cinq groupes ont été obtenus pour 8428 domaines faisant en moyenne environ 1 à 52 kb mais pouvant atteindre jusqu'à 737 kb pour les plus larges, indiquant que le génome de drosophile est généralement organisé en de larges régions. Parmi les cinq groupes identifiés, trois (**vert**, **bleu** et **noir**) appartiennent à la chromatine « non permissive » et deux (**rouge** et **jaune**) à la chromatine « permissive ». Il est important de retenir que c'est la **combinaison de plusieurs protéines** qui définit chaque groupe et non pas la présence d'une seule protéine (fig. 21).

- **La chromatine noire** couvre 48% du génome et contient 4165 gènes soit 32% des gènes que compte le génome de la drosophile. La présence des quatre protéines suivantes la caractérise : H1, D1, Suppressor of Underreplication (SuUR) et IAL/auroraB. L'histone H1 permet de stabiliser les nucléosomes (Hergeth, 2015), la protéine D1 est essentielle pour le silencing des gènes lors de la différenciation des gamètes (Aulner, 2002), SuUR est impliquée dans l'assemblage de l'hétérochromatine, et IAL/auroraB est impliquée dans la condensation chromosomique en particulier lors de la mitose. La majorité des gènes présents dans cette région, qualifiés de gènes silencieux, n'est pas exprimée ou alors très faiblement (fig. 21).

- **La chromatine bleue** est enrichie en protéines du complexe protéique Polycomb (PcG). Ce complexe protéique est fortement connu pour être impliqué dans la régulation des gènes *via* des modifications post-traductionnelles des histones dans des processus majeurs comme l'inactivation du chromosome X chez le mâle, la plasticité et le renouvellement des cellules souches. Le complexe PcG est fortement décrit comme étant associé à la répression des gènes développementaux. Il participe notamment à la tri-méthylation de la lysine 27 de l'histone H3. (H3K27me3), qui est une modification post-traductionnelle considérée comme répressive. La

chromatine **bleue** est enrichie en H3K27me3 et contient un grand nombre de gènes développementaux réprimés, en accord avec la fonction du complexe Polycomb (fig. 21 ; Kassis, 2017 ; Piunti, 2021).



**Figure 20 : Les 5 types de chromatine sont caractérisés par des combinaisons de protéines et des modifications d'histones distinctes.**

**A)** Fraction de tous les loci génomiques sondés dans chaque type de chromatine qui est liée par chaque protéine. Les loci liés ont été déterminés séparément pour chaque protéine.

**B)** Les niveaux d'histone H3 et de quatre modifications d'histones déterminés par ChIP à l'échelle du génome.

(figure tirée de Filion, 2010).

	Chromatin states	Roles	Histone marks
Permissive	"Yellow" chromatin	Cellular housekeeping / metabolism genes	H3K36me3 (?)
	TrxG-associated (Trithorax)	Developmentally active genes	H3K4me1,3 H3K27ac
Repressive	HP1-associated heterochromatin	Pericentric heterochromatin / terminal differentiation	H3K9me2,3
	PcG-associated (Polycomb)	Developmentally repressed genes	H3K27me3
	"Black" chromatin	Neuronal development	Unknown

**Figure 21 : Les différentes couleurs de la chromatine.**

(figure tirée de Delandre, 2019).

- **La chromatine verte** est la chromatine correspondant à la chromatine associée aux protéines Su(Var)3-9 et HP1. Su(Var)3-9, qui possède une activité HKMT, assure la méthylation de la lysine 9 de l'histone H3 (Schotta, 2003). HP1 participe au « silencing » des régions centromériques. La chromatine verte se caractérise également par une forte présence de di ou tri-méthylation (H3K9me2/me3) des lysines des histones H3 qui sont elles aussi associées au phénomène de « silencing » des gènes. La chromatine verte est donc spécialisée dans la mise sous silence des gènes et plus spécifiquement des régions péri-centriques (**fig. 21**).

- La chromatine « permissive » est divisée en deux catégories de chromatine : **la rouge** et la **jaune**. Ces deux catégories de chromatine contiennent les gènes exprimés et sont enrichies en ARN Polymérase II, H3K4me2 et H3K79me3. Ces deux types de chromatine sont très similaires et partagent de nombreuses protéines communes telles que les HDAC, RPD3 et SIR2, ainsi que la partenaire de RPD3, SIN3A. On retrouve également DF31, décrite comme pouvant décondenser la chromatine *in vivo* et permettant de maintenir accessible l'ADN (Guillebault, 2007), mais aussi Ash2, méthyltransférase de la H3K4, membre de la famille Trithorax (Carbonell, 2013). Pour finir Max facteur de transcription connu pour s'oligo-dimériser avec Myc pour activer la transcription (Gallant, 1996). La chromatine **jaune** et **rouge** se distinguent l'une de l'autre par leur vitesse de réplication, la chromatine **rouge** se répliquant plus rapidement que la **jaune**. Ces deux types de chromatine se différencient aussi en fonction de l'enrichissement en MRG15 et H3K36me3 (**fig. 21**).

La chromatine **jaune** est définie par sa composition en gènes dits « de ménage », fortement traduits et un enrichissement spécifique en H3K36me3 et MGR15. La chromatine **rouge** est une zone de transcription active révélée par la présence de protéines jouant un rôle dans l'activation des gènes développementaux. On y trouve notamment, la sous-unité du complexe Médiateur MED31, le récepteur à l'ecdysone EcR, des facteurs de type GAGA, Brahma et Lola-like qui appartiennent à la famille Trithorax. Cette région de la chromatine est également fortement enrichie en histones H3 mono ou triméthylées sur la lysine 4 (H3K4me1 et H3K4me3) ainsi qu'en histones H3 acétylées sur la lysine 27 (H3K27ac) permettant d'identifier les promoteurs et les enhancers actifs (Creyghton, 2010). La chromatine **rouge** est donc une région où se localise des gènes développementaux actifs (**fig. 21**).

## **2-2 Les différentes marques d'histones.**

Comme évoqué ci-dessus, les histones peuvent être modifiées post-traductionnellement par différentes enzymes (acétylase-désacétylase, méthylase-déméthylase) sur leur chaîne latérale (Bannister, 2011). Ces enzymes sont appelées « writer » si elles ajoutent un groupement fonctionnel ou « eraser » si elles suppriment un groupement existant. A ces premières modifications, se sont ajoutées des modifications consécutives à des phosphorylations, sumoylations ou encore ubiquitinylation (Bannister, 2011). Ces modifications aussi nommées marques des histones peuvent être associées, en fonction de leur enrichissement, à des éléments géniques et des environnements chromatinien spécifiques. Ainsi, la présence et le ratio entre différentes marques d'histones permettent non seulement d'identifier les éléments de cis-régulation mais également les enhanceurs dits actifs et localisés dans l'euchromatine (Cubéñas-Potts, 2017 ; Jiang, 2018).

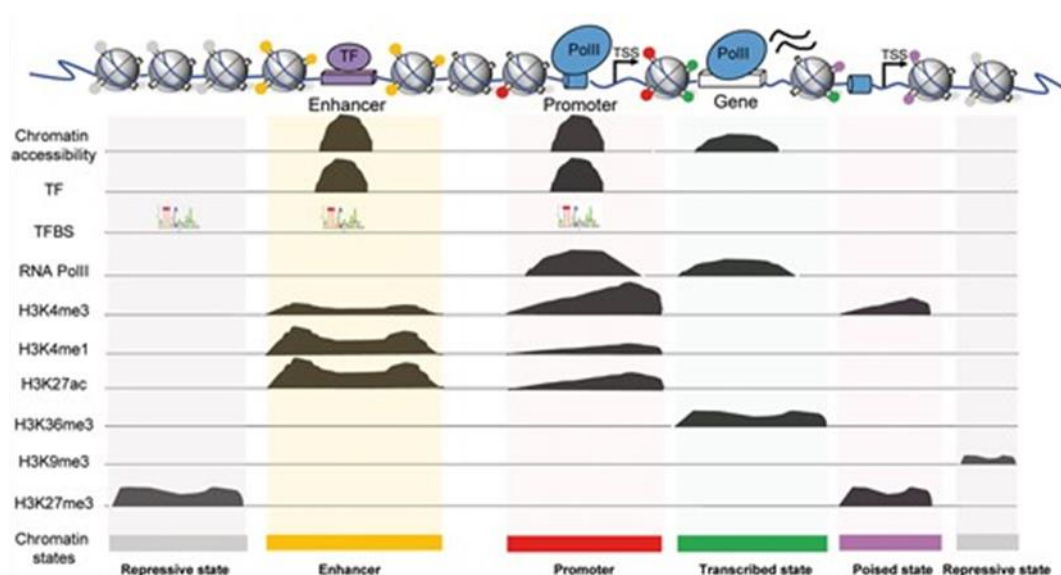
### **2-2-1 L'acétylation des histones.**

L'acétylation des histones est régulée par deux familles d'enzymes aux activités antagonistes : les histones acétyltransférases (HAT) et les histones désacétylases (HDAC). L'acétylation se produit au niveau de la chaîne latérale des lysines (K) des histones H3 et H4, les HATs catalysant le transfert d'un groupe acétyle (Ac) grâce au recrutement du cofacteur CoA. Cette modification modifie la charge de l'histone H3 ou H4 et diminue ainsi son interaction avec la molécule d'ADN. Les HATs sont contenues dans deux familles, les types A et les types B. Les HATs de type B, généralement cytoplasmiques acétylent les histones libres alors que les enzymes de type A, situées en majorité dans le noyau, modifient les histones des nucléosomes. Le type A est lui-même divisé en trois sous-groupes (GNAT, MYST et CBP/p300) en fonction de l'homologie de séquence et de structure conformationnelle des enzymes. Chaque HAT modifie la conformation de résidu différent, par exemple CBP acétyle les lysines 14 et 18 de l'histone H3 alors que HAT1, de la famille des GNAT, acétyle les lysines 5 et 12 de l'histone H4. Elles sont souvent considérées comme des coactivateurs de la transcription (Marmorstein, 2014). Les HDACs, dont la fonction est de retirer un groupement Ac, sont réparties en quatre classes (de I à IV). Par cette modification qui restaure la charge positive

des histones, les HDACs sont généralement considérées comme des répresseurs transcriptionnels (Ho, 2020).

### 2-2-2 La méthylation des histones.

La méthylation est une modification qui se produit sur des résidus arginine et lysine des histones H3. Les lysines (K) peuvent être mono- (me1), di- (me2) ou tri-méthylées (me3) et les arginines (R) mono- et diméthylées. L'ajout d'un groupement méthyle qui contrairement à l'Ac, ne modifie pas la charge des histones, est médié par des lysines méthyltransférases (HKMT) ou par des arginines méthyltransférases (PRMT) en fonction du substrat. Les HKMTs catalysent le transfert du groupement méthyle à partir de S-adénosylméthionine (SAM) sur la lysine de l'histone (Marmorstein, 2014). Les PRMTs sont séparées en quatre types qui se distinguent notamment en fonction de la symétrie du groupement méthyle de SAM ajouté à l'arginine (Lorton, 2019). La méthylation des histones fut longtemps considérée comme une modification irréversible, jusqu'à la découverte de la protéine LSD1, première lysine déméthylase en 2004 (Shi, 2004), puis quelques années après, des arginines déméthylases en 2007 (Chang, 2007). L'analyse de ces marques d'histones et leurs enzymes associées est appelée épigénétique. Il est important de noter ici que l'agencement des diverses modifications post-traductionnelles des histones peut être considéré comme un code.



**Figure 22 : Les marques d'histones comme marqueurs des éléments géniques.**

Les états de la chromatine sont définis par différentes combinaisons de modifications des histones, de liaisons de facteurs de transcription et de l'ARN Pol II. (figure tirée de Jiang, 2018).



En effet, pour une région déterminée, cet agencement permet de donner une indication de l'état de la chromatine, de prédire son activité transcriptionnelle ainsi que sa nature (promoteur...) (**fig. 22 ; Jiang, 2018**). Pendant ma thèse, j'ai notamment utilisé les marques H3K27ac et H3K27me3 pour identifier respectivement, la chromatine ouverte, active en transcription et la chromatine fermée et inactive. La marque H3K27ac permet plus précisément l'identification d'élément actif de modules de cis-régulation (CRM) généralement appelé **enhancers**.

## **2-3 Les éléments de cis-régulation.**

### **2-3-1 Les enhancers.**

Les enhancers sont des séquences clés, bien distinctes des promoteurs. Ils peuvent être localisés en amont ou en aval d'un gène, voire très loin de celui-ci, ou à l'intérieur de la partie transcrite d'un gène comme les régions non traduites de l'ARN (UTR) ou les introns. Les enhancers servent de plateforme pour la fixation des facteurs de transcription (TF) et de ce fait, sont des éléments clés dans la régulation de l'expression génique (Furlong, 2018). Les facteurs de transcription sont des protéines capables de se lier à l'ADN de manière spécifique à une séquence et leur activité est responsable de l'expression du génome dans des cellules données et à un moment précis. Il existe plusieurs types de facteurs de transcription en fonction de leurs domaines de liaison à l'ADN, notamment : Helix-Loop-Helix, Homeobox, Zinc Fingers, etc... Pour réguler l'expression des gènes, chaque TF se lie à un court motif d'ADN spécifique, qui tolère souvent des déviations plus ou moins importantes par rapport à la séquence consensus. Par conséquent, les sites de fixation des TFs sont généralement représentés sous forme de matrices de poids de position (PWM) qui représentent la fréquence des nucléotides à chaque position du consensus, tel que défini par la compilation des sites de liaison connus pour un TF donné (Santolini, 2014). Les sites de fixation des TF sont compris au sein des enhancers qui jouent donc un rôle crucial dans la régulation de l'expression des gènes et plus particulièrement dans celle des gènes dits du développement. En effet, les enhancers intègrent les signaux de divers réseaux géniques et autres voies de signalisation et les traduisent en modèles d'expression hautement spécifiques (Bulger, 2011). En fonction de leur rôle dans la régulation de l'expression génétique, les enhancers ont été initialement séparés

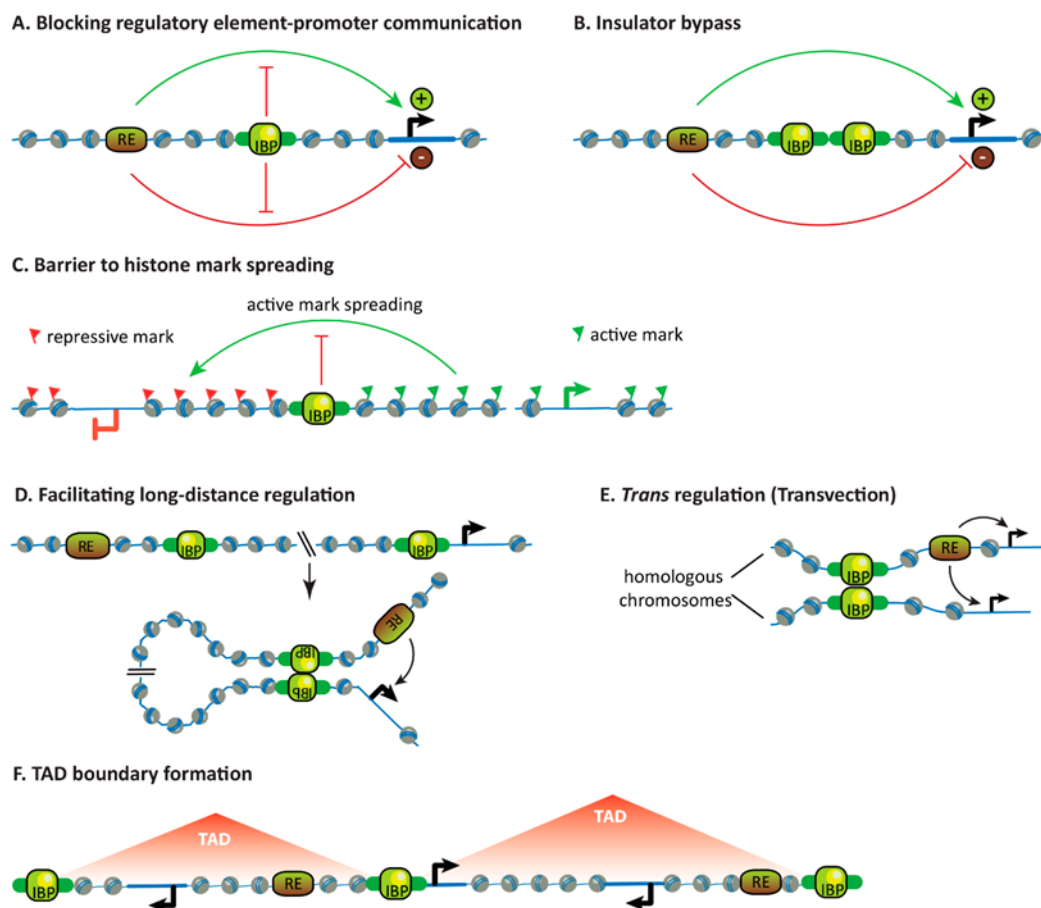
en deux classes. Les « Enhancers » représentent d'abord des régions régulatrices qui activent l'expression du gène cible, tandis que les « Silencers » la répriment, comme observé par des constructions artificielles contenant des éléments régulateurs devant un gène rapporteur, dans des cellules en culture ou chez des animaux transgéniques. Cependant, selon les types de cellules, une même région régulatrice peut agir soit comme amplificateur, soit comme silencieux. Aujourd'hui, le terme "enhancer" est souvent utilisé comme synonyme de CRM. Comme nous le verrons ultérieurement pour le facteur de transcription Shavenbaby (Svb), un gène de développement peut avoir de nombreux enhancers distincts et pléiotropes qui conduisent des expressions spécifiques dans différents tissus (Stern, 2013). Ainsi, les différents enhancers d'un même gène peuvent être considérés comme un réseau de régulation, intégrant différentes informations spatiales et temporelles pour produire un schéma d'expression génétique finement contrôlé (Buchler, 2003). Les enhancers contiennent souvent de multiples sites pour la liaison des TFs. Un tel regroupement de sites de liaison de facteurs de transcription (TFBS) est en effet très répandu dans les enhancers de développement, avec de multiples sites de liaison pour un TF donné (regroupement homotypique) et/ou plusieurs TFs (regroupement hétérotypique) (Aerts, 2012). Les enhancers sont des fragments d'ADN, dont l'activité est indépendante de l'orientation et une des difficultés issues de cette propriété est de restreindre leur activité à un ou plusieurs gènes spécifiques. En effet, si le mécanisme de transcription est le même pour l'ensemble des gènes, ces derniers sont régulés indépendamment les uns des autres. La spécificité de la régulation transcriptionnelle nécessite donc une compartimentation des gènes dans des domaines transcriptionnellement indépendants.

### **2-3-2 Les insulateurs.**

Parmi les éléments régulant cette compartimentation, les insulateurs, éléments de l'ADN, permettent d'isoler des gènes de leur environnement pour assurer une régulation spécifique (**fig. 23** ; Özdemir, 2019).

L'étude du gène Hox Abdominal-B (Abd-B) a donné un premier indice de l'existence d'insulateurs chez la drosophile pour maintenir l'indépendance de régions cis-régulatrices voisines. En effet, une délétion chromosomique spontanée a supprimé la frontière entre deux

enhancers qui dirigent chacun l'expression d'Abd-B dans des segments corporels distincts conduisant à un défaut de segmentation (Gyurkovics, 1990). Ainsi, lorsqu'un transgène est encadré par deux insulateurs et inséré dans un locus, ce dernier n'est pas affecté par la structure chromatinienne environnante (rôle barrière). Par exemple, lorsque les transgènes sont insérés dans une région répressive de l'hétérochromatine, les insulateurs permettent l'expression du gène inclus dans le transgène (Udvardy, 1985). A l'inverse, les insulateurs, placés artificiellement entre un enhancer et un promoteur, peuvent bloquer la communication entre ces éléments et empêcher l'activation du gène (Gaszner, 2006).



**Figure 23 : Propriétés des insulateurs et de leurs protéines de liaison.**

Les insulateurs peuvent bloquer la communication entre les éléments régulateurs et les promoteurs **A)**, soutenir le contournement lorsqu'ils sont appariés **B)**, agir comme barrières à l'étalement des marques d'histones **C)**, faciliter la communication à longue distance entre un élément régulateur et un promoteur en cis **D)** et en trans **E)** et former des frontières de domaines topologiquement associés (TAD) **F)**.

Certaines de ces propriétés n'ont été décrites jusqu'à présent que chez les mouches (**B** et **E**) et d'autres seulement chez les mammifères **F**.

(figure et légende tirées de Özdemir, 2019).

Nombre de travaux initiaux réalisés sur les insulateurs ont été effectués chez la drosophile en utilisant l'insulateur *gypsy*, localisé dans la région 5' du rétrotransposon du même nom.

Lorsque l'élément *gypsy* est transposé entre un gène et ses enhancers, spécifiques d'un tissu, il empêche l'expression du gène uniquement dans ce tissu, suggérant que l'élément *gypsy* isole en quelque sorte le promoteur de ses éléments de cis-régulation (Peifer, 1986; Geyer, 1992). Trois protéines pouvant interagir entre elles, Su(Hw) (Supressor of Hairy wing), Mod(mdg4) et CP190 (Centrosomal Protein 190 Kd) s'associent à cet insulateur. La protéine Su(Hw) que j'utiliserai dans mes travaux, appartient à la famille des facteurs de transcription à doigt de zinc. Le site de liaison de Su(Hw) est suffisant pour le blocage de la communication entre un enhancer et un promoteur (Geyer, 1992), et confère une protection contre les effets de position dans les tests transgéniques (Roseman, 1993).

Dans cette partie, nous avons mis en évidence que l'ADN, compacté et contenu dans le noyau des cellules eucaryotes, est organisé en structures dynamiques, permettant de s'adapter aux contraintes du cycle cellulaire et de l'expression génétique. Les différents états de compaction, reflétés par les différentes couleurs de la chromatine, sont finement régulés par les histones et les modifications post-traductionnelles auxquelles elles sont soumises. La régulation de cette compaction rendant les éléments de cis-régulation (enhancers, promoteurs, insulateurs...) localisés sur l'ADN, plus ou moins accessibles aux facteurs agissant en trans comme les facteurs de transcription, permet de contrôler finement l'expression génique.

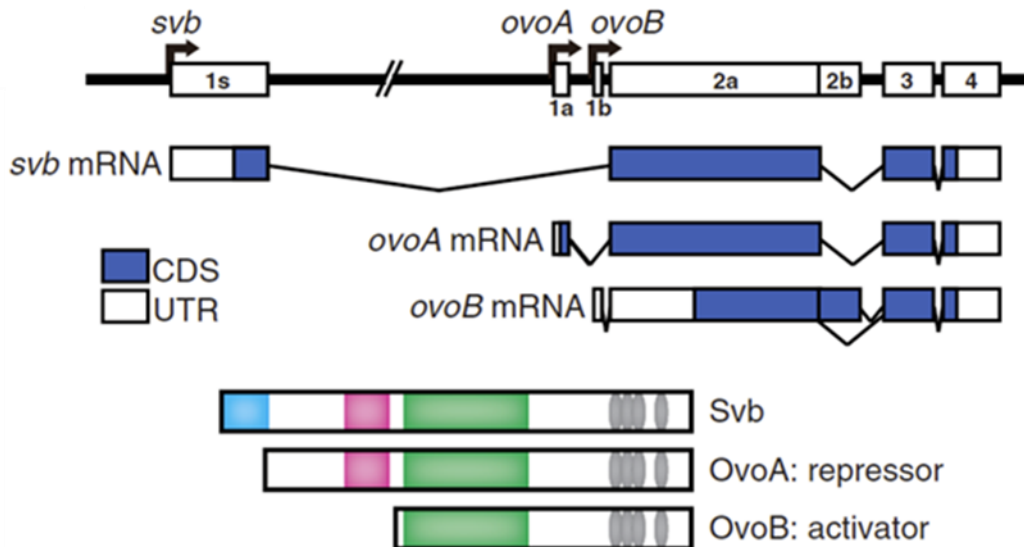
### **3- Shavenbaby, un acteur déterminant de la formation des trichomes dans l'épiderme.**

J'ai réalisé ma thèse dans l'équipe de François Payre qui s'intéresse depuis plusieurs années à comprendre les mécanismes moléculaires impliqués dans la morphogenèse épidermique chez la drosophile. Les travaux de l'équipe ont permis d'identifier le locus *ovo/shavenbaby (svb)*, codant un facteur de transcription essentiel pour le remodelage des cellules et la formation d'extensions cytoplasmiques apicales dans l'épiderme, appelés trichomes. La dissection moléculaire de la fonction de Svb dans la formation des trichomes a révélé qu'il active une

centaine de gènes cibles qui collectivement permettent le remodelage de la surface apicale des cellules.

### 3-1 Organisation moléculaire du locus *ovo/svb*.

Svb est un facteur de transcription à doigt de zinc C2H2 codé à partir du locus *ovo/svb*, qui est indispensable au développement des trichomes et de la lignée germinale femelle (Mével-Ninio, 1991). Deux promoteurs alternatifs conduisent l'expression des isoformes protéiques OvoA et OvoB, spécifiques de la lignée germinale. Un troisième promoteur agit dans les tissus somatiques et permet l'expression d'une isoforme à extrémité N-terminale étendue, SvB. Les trois isoformes partagent un domaine commun de liaison à l'ADN et une région médiant l'activation transcriptionnelle (Mével-Ninio, 1995) (**fig. 24**). OvoB, l'isoforme germinale la plus courte, agit comme un activateur (Andrews, 2000 ; Mével-Ninio, 1995) tandis qu'OvoA possède une région N-terminale supplémentaire, portant un domaine répresseur qui domine l'activité transcriptionnelle de ce TF (Andrews, 2000). Enfin, SvB constitue la forme la plus



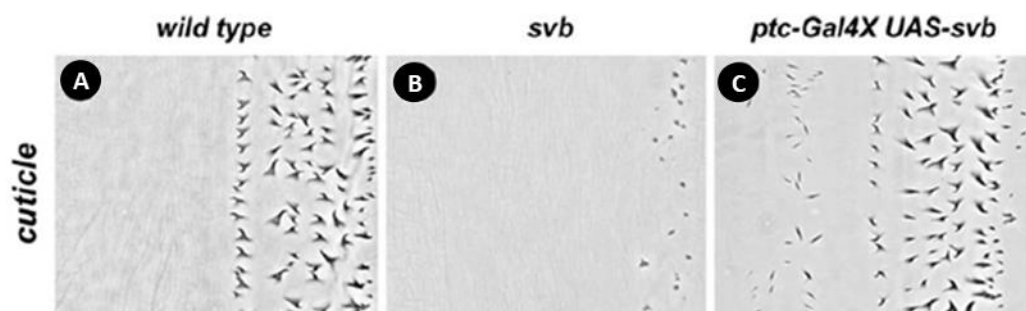
**Figure 24 : Le locus *ovo/shavenbaby* code 3 isoformes différentes.**

Représentation du locus *ovo/svb*, des ARNm transcrits *svb*, *ovoA* et *ovoB* ainsi que des 3 protéines respectives. Les régions codantes sont présentées en **bleu** (CDS) et les régions non codantes en blanc (UTR). Les protéines SvB, OvoA et OvoB comprennent 4 doigts de zinc (ovale gris) et une région activatrice (**vert**) tandis que SvB et OvoA possèdent une région dominante de répression transcriptionnelle (**rouge**). SvB contient spécifiquement en partie N-terminale l'exon 1s (**bleu ciel**). (Kondo et al., 2010).

longue puisqu'elle possède en plus des domaines activateur et répresseur, une extension N-terminale codée à partir de l'exon 15 spécifique au soma (Delon, 2003).

### 3-2 Svb gouverne la différenciation des trichomes dans l'épiderme.

Les premiers mutants *svb* ont été isolés par Nusslein-Volhard et Wieshaus en 1980. Les embryons mutants pour *svb* présentent une cuticule dépourvue de trichomes tandis que des embryons sauvages montrent une alternance de cellules à trichomes et de cellules à surface lisse. Il a été démontré que *svb* est spécifiquement exprimé dans des cellules qui forment des trichomes (Mével-Ninio, 1995). D'autres travaux ont montré que *svb* est à la fois nécessaire et suffisant pour la formation des trichomes et ont élucidé certains aspects de sa régulation, en se concentrant sur les cellules épidermiques ventrales (**fig. 25** ; Payre, 1999). En effet, les individus mutant pour *svb* ne forment pas de trichomes tandis que l'expression ectopique de *svb* dans des cellules lisses conduit à la formation de trichomes. De plus, les cellules qui expriment *svb* dans l'épiderme vont former des denticules. Ainsi *svb* détermine le registre spatial des cellules qui vont former un trichome. La comparaison de l'activité des différentes formes *OvoA*, *OvoB* et *Svb* a permis de mieux comprendre la fonction transcriptionnelle de *Svb* dans les cellules épidermiques. De façon intéressante, l'expression ectopique de la forme *OvoA* (répresseur) dans l'épiderme empêche la formation des trichomes tandis que l'expression de *OvoB* (activateur) induit la formation de trichomes ectopiques (Andrews, 2000). *Svb* agit donc comme un activateur de transcription pour la formation des trichomes.



**Figure 25 : Svb est nécessaire et suffisant pour la formation des trichomes.**

Photographies de cuticules produites par des embryons contrôle **A**), mutant *svb* **B**) et surexprimant *Svb* dans les cellules exprimant *Ptc* **C**). Les embryons mutants *svb* génèrent une cuticule nue contenant quelques denticules atrophiés. L'expression ectopique de *svb* dans des cellules normalement lisses est suffisante pour forcer la production de trichomes. Vues ventrales de cuticules de segment abdominal A3 ou A4 d'embryons de stade 16-17.

(D'après Delon, 2003).

De plus, quand les isoformes Ovo, qui sont normalement spécifiques de la lignée germinale, sont exprimées de manière artificielle dans l'épiderme, elles sont capables de reconnaître les mêmes effecteurs que Svb, pour réprimer (OvoA) ou activer (OvoB) leur expression (Chanut, 2006). Les trichomes sont des extensions cytoplasmiques riches en filaments d'actine. L'équipe a montré que Svb déclenche la réorganisation du cytosquelette à la surface apicale des cellules induisant la formation des trichomes (Delon, 2003). L'ensemble de ces données a montré que Svb est nécessaire et suffisant pour induire la formation des trichomes.

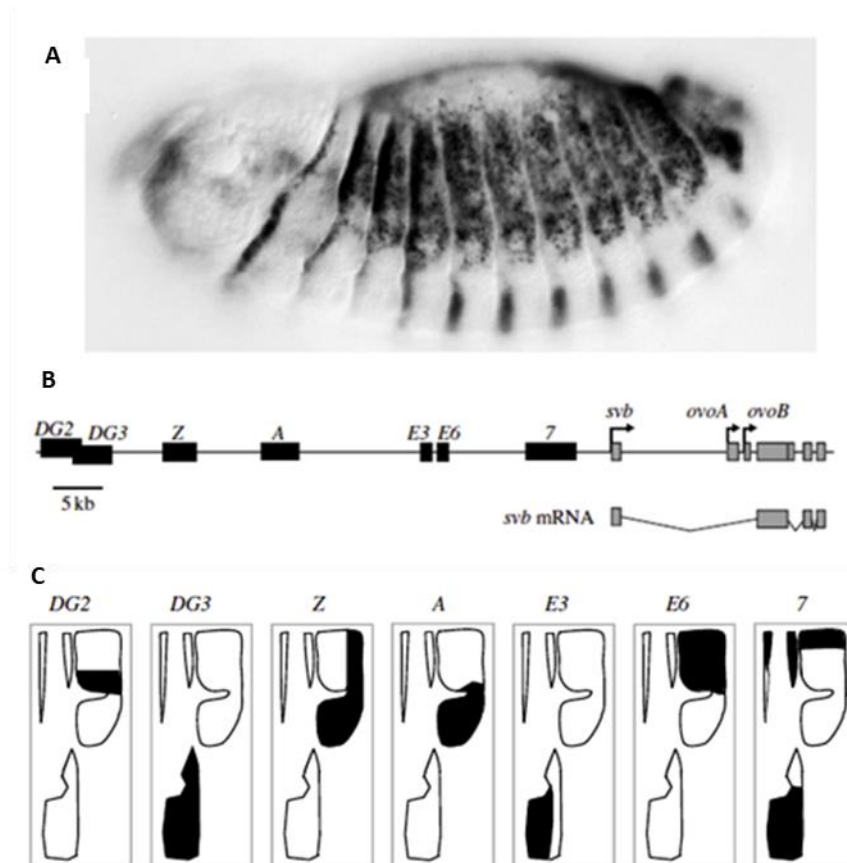
### **3-3 Régulation de l'expression du gène *svb*.**

Les travaux pionniers de François Payre ont montré que *svb* intègre les voies de signalisation Wingless (*wg*) et EGF pour définir les cellules qui vont former des trichomes. Après son rôle dans la segmentation embryonnaire, *wg* un membre fondateur de la voie de signalisation cellulaire Wnt, détermine la production de la cuticule nue. La perte de fonction de *wg* conduit à la production de trichomes par toutes les cellules. Réciproquement, l'expression ectopique du ligand de Wg ou l'activation ectopique de la voie réprime complètement les denticules, conduisant à un phénotype ventral chauve. Ces défauts sont médiés par l'action de Wg qui réprime l'expression de *svb*. D'autre part, l'expression de *svb* dans les cellules du trichome est activée par la voie de l'EGF-r. Le domaine d'expression normal de *svb* coïncide avec la localisation de la signalisation active de l'EGF. De plus, l'inhibition de la voie EGF-r conduit à l'absence de trichomes, résultant d'une diminution de l'expression de *svb*. Pour résumer, les activités réciproques de Wg et EGF-r agissent ensemble pour contrôler finement le domaine d'expression de *svb* dans les cellules épidermiques, déterminant à son tour le patron des trichomes (Payre, 1999).

Le travail collaboratif de notre équipe et celle de David Stern a permis de révéler les régions cis-régulatrices requises pour l'expression de *svb* et a mis en évidence que l'évolution du patron de denticule chez la drosophile est la conséquence de mutations au sein de ces régions. Cinq régions dites enhancer définissent le patron d'expression de *svb* dans l'épiderme embryonnaire, étendues sur 90kb en amont du 1<sup>er</sup> exon (Frankel, 2012).

Un criblage systématique à l'aide de lignées rapportrices couvrant les 50 kb en amont du gène a permis d'identifier trois premières régions. Ces trois régions (*7*, *E* et *A*) sont actives selon des profils distincts et recouvrent le patron d'expression complet de *svb* (Mc Gregor, 2007). Un criblage plus précis, auquel j'ai participé, a permis de restreindre la taille des régions

régulatrices. Nous avons identifié deux régions minimales *E3* et *E6* dans la boîte E (**fig. 26** ; Frankel, 2010). Puis, deux régions supplémentaires, *DG* et *Z*, hautement conservées au cours de l'évolution et couvrant 41 kb en amont de l'enhancer A ont été identifiées. Ces régions ont des profils spatiaux similaires à ceux des régions identifiées précédemment. De plus, la délétion simultanée de ces deux régions n'affecte pas la formation des trichomes et ne provoque pas de létalité (Frankel, 2012). Ces résultats ont permis de définir ses enhancers comme « shadow » car ils assurent la robustesse d'expression de *svb*. L'expression redondante, soutenue par plusieurs enhancers, peut ainsi palier à de possibles mutations délétères (Frankel, 2010 ; Lagha, 2012).



**Figure 26 : L'expression de *svb* est générée par l'activité d'un ensemble d'enhancers.**

- A)** Patron d'expression du gène *svb* dans l'épiderme en fin de stade embryonnaire.  
**B)** Vue schématisée du locus *ovo/svb*. Les boîtes noires indiquent les enhancers de *svb* et les boîtes grises délimitent les exons.  
**C)** Sept enhancers produisent le patron d'expression de *svb* dans les régions dorsales et latérales (en haut) et ventrales (en bas) de l'embryon. Les régions en noir correspondent aux régions d'activité des enhancers. (D'après la revue Frankel, 2013).



A la suite de l'identification des régions régulatrices contrôlant l'expression de *svb*, l'objectif était de rechercher les facteurs de transcription se fixant sur ces enhancers et nécessaires pour leur activité. Alexandre Ahmad Alsawadi, étudiant en thèse au laboratoire, a criblé par une approche RNAi à grande échelle 400 facteurs de transcription exprimés en fin d'embryogenèse. Il a recherché si leur déplétion modifie l'activité des régions 7, E et A. Il a pu identifier dans la région E3 un site de liaison de *pointed*, un effecteur de la voie EGF, ainsi qu'un site de liaison du facteur TCF impliqué dans la voie Wg (Al Hayek, 2020), montrant que cette région intègre les deux voies de signalisation pour la régulation de l'expression de *svb*. En conclusion, *svb* intègre, via plusieurs *enhancers*, les voies de signalisation Wg et EGF afin de déterminer le registre spatial et la robustesse de son patron d'expression dans l'épiderme embryonnaire.

### **3-4 Svb coordonne un réseau génique directement impliqué dans la formation des trichomes.**

Afin de mieux comprendre comment Svb déclenche la réorganisation apicale des cellules épidermiques lors de la formation des trichomes, l'équipe a recherché les gènes cibles de Svb par une approche candidate complétée par une approche globale. L'approche candidate, basée sur le profil d'expression des gènes, ou bien le rôle des gènes dans la formation des trichomes, a permis d'identifier plusieurs dizaines de gènes cibles (Chanut, 2006 ; Fernandez, 2010). Ces gènes codent pour des effecteurs cellulaires qui participent collectivement et directement dans le remodelage de la surface apicale des cellules sous au moins trois grandes fonctions (**fig. 27**). Ils sont impliqués dans la réorganisation du cytosquelette d'actine permettant la déformation de la cellule (Chanut, 2006). De plus, ils codent pour des protéines de la matrice extracellulaire apicale qui soutiennent la forme des trichomes (Fernandes, 2010). Enfin, ils permettent la synthèse de protéines structurantes de la cuticule et des enzymes requises pour la pigmentation de la cuticule (Chanut, 2006 ; Menoret, 2013).

Afin d'obtenir une vision plus étendue de la fonction de Svb, une approche globale combinant une analyse transcriptomique (puces Affymetrix) et des expériences de fixation à l'ADN (ChIP-seq) a été réalisée par Delphine Ménoret, étudiante en thèse (Menoret, 2013). Ce travail a révélé que Svb contrôle l'expression de plus de 150 gènes au cours de la formation des trichomes, dont la fonction de la plupart d'entre eux n'est pas encore connue. Afin de mieux

comprendre comment Svb active ses gènes cibles, l'équipe a recherché les séquences régulatrices de trois gènes cibles en analysant de manière systématique l'activité de leurs régions non codantes dans l'épiderme embryonnaire. Ainsi, dans l'épiderme embryonnaire,

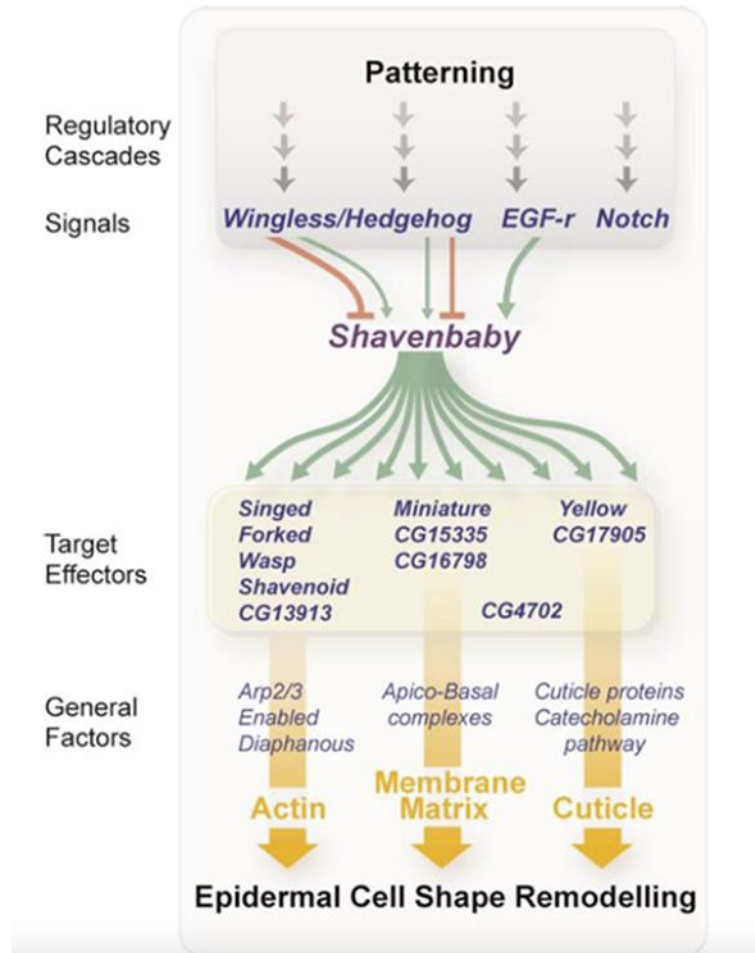


Figure 27 : Modèle de la régulation et la fonction de Svb lors de la formation des trichomes. (D'après Chanut, 2006).

un enhancer a été identifié pour le gène *singed* (*sn1*), deux pour le gène *sha* (*sha1* et *sha3*) et deux pour le gène *miniature* (*Emin* et *EminB*) (Chanut, 2006 ; Menoret, 2013). L'étude s'est appuyée sur l'existence d'un site consensus pour la fixation d'Ovo, CnGTTa nommé OvoQ6, défini par méthode Selex (Lee, 2000). Cependant, ce site est largement représenté dans le génome et n'est pas suffisamment sélectif pour identifier des sites de liaison à Svb. Par une approche statistique, les séquences identifiées comme actives comparées à des séquences non actives dans l'épiderme ont révélé qu'une région activée par Svb contient au moins un site de liaison de 8 bp. Cette séquence appelée *SvbF7*, ACHGTTAK est une extension de la séquence *OvoQ6*. La résolution de ce site a permis d'identifier des régions régulatrices

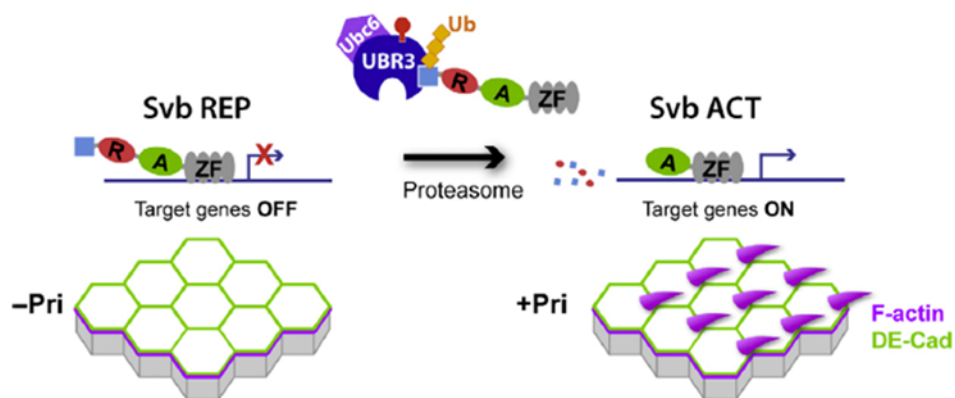
d'autres gènes cibles de Svb. Les régions régulatrices identifiées peuvent contenir un ou plusieurs sites de liaison à Svb. De plus, elles comprennent deux autres motifs, encore non attribués à la fixation d'un facteur de transcription connu. Dans de nombreux cas, la mutation ponctuelle d'un site de fixation à Svb induit une diminution significative de l'activité de la région régulatrice dans l'épiderme (Menoret, 2013).

L'ensemble de ses résultats montre que Svb est le facteur de transcription déterminant pour la formation des trichomes. Par la liaison à une combinaison variable de site de fixation, Svb active un ensemble de gènes cibles qui sont collectivement responsables de la formation des trichomes.

### **3-5 Les smORFs peptides Pri : de nouveaux régulateurs de la morphogénèse épidermique.**

En 2007, deux équipes ont identifié un nouveau régulateur de la différenciation épidermique, l'une par génétique classique (isolement d'un mutant) et l'autre par génétique reverse (Galindo, 2007 ; Kondo, 2007). Ce gène appelé *polished-rice* (*pri*) ou *tarsal-less* (*tal*) a été initialement découvert comme acteur majeur de la segmentation chez le coléoptère *Tribolium castaneum* (nommé mille-pattes) (Savard, 2006). En recherchant des gènes codant de nouveaux longs ARN non codants (lncRNA), l'équipe de Y. Kageyama a révélé le gène *pri* caractérisé par un profil d'expression très dynamique au cours de l'embryogénèse (Inagaki, 2005). *Pri* est exprimé dans de nombreux tissus comme les spiracles, l'intestin et les trachées. Son expression dans l'épiderme présente un profil segmenté recouvrant celui de *svb*. La délétion du gène *pri* provoque une létalité embryonnaire et les embryons présentent des défauts épidermiques similaires à ceux des mutants *svb* puisque les cellules ne forment pas de trichomes (Kondo, 2007). Ce gène est transcrit en un long ARN considéré alors comme non codant (lncRNA, long non coding RNA) puisqu'il contient des cadres ouverts de lecture de moins de 100 aas (Inagaki, 2005 ; Tupy, 2005). Depuis, de nombreuses études ont révélé une nouvelle classe de gènes appelés gènes *smORFS* (small Open Reading Frame), dont fait partie *pri*, codant des peptides et possédant des fonctions biologiques de la levure jusqu'à l'homme. Les deux équipes ont démontré que l'ARN *pri* contient des petits cadres ouverts de lecture, codant des petits peptides de 11 à 32 acides aminés, indispensables pour les fonctions embryonnaires de *pri*.

La ressemblance phénotypique d'absence de trichomes observée en absence de *pri* et *svb*, a mené notre équipe, en collaboration avec celle de Yugi Kageyama, à rechercher le lien moléculaire entre les peptides Pri et Svb. Ces travaux ont démontré que les peptides Pri agissent sur la protéine Svb par modification post-traductionnelle. Le facteur Svb est produit sous forme de protéine de grande taille, agissant comme un répresseur de la transcription. Les peptides Pri induisent une dégradation protéolytique de la partie N-terminale de la protéine Svb, contenant une large partie du domaine répresseur (Kondo, 2010). Svb est ainsi convertie de répresseur transcriptionnel en activateur transcriptionnel, dont la structure est très proche de celle d'OvoB. Jennifer Zanet et Emilie Benrabah ont recherché le mécanisme d'action des peptides Pri. Elles ont montré que les peptides Pri recrutent l'enzyme E3 ubiquitine ligase Ubr3 qui, après avoir ubiquitinylé Svb, le dirige au protéasome pour être partiellement dégradé (fig. 28).



**Figure 28 : Modèle de la maturation de Svb par les peptides Pri et Ubr3.**

Après la liaison des peptides Pri, Ubr3 est capable d'interagir avec le domaine N-terminal de Svb et induire son ubiquitination. Svb est dirigé vers le protéasome pour une dégradation partielle. Svb est sous sa forme courte activatrice de la transcription, dépourvue de son domaine répresseur. (d'après Zanet, 2016).

En conclusion, les peptides Pri induisent, en lien avec le protéasome et Ubr3, la dégradation partielle de Svb et sa conversion de répresseur en activateur de transcription.

Comme indiqué précédemment, le gène *pri* a été identifié à l'origine sous le nom de *mille-pattes* chez *Tribolium castaneum* comme un gène essentiel pour la segmentation pendant l'embryogenèse. Ce processus, piloté par les gènes *gap*, est nécessaire pour générer des subdivisions dans l'embryon précoce. Des études indépendantes réalisées sur des insectes modèles ayant des modes de segmentation différents ont révélé le rôle de *svb* dans la

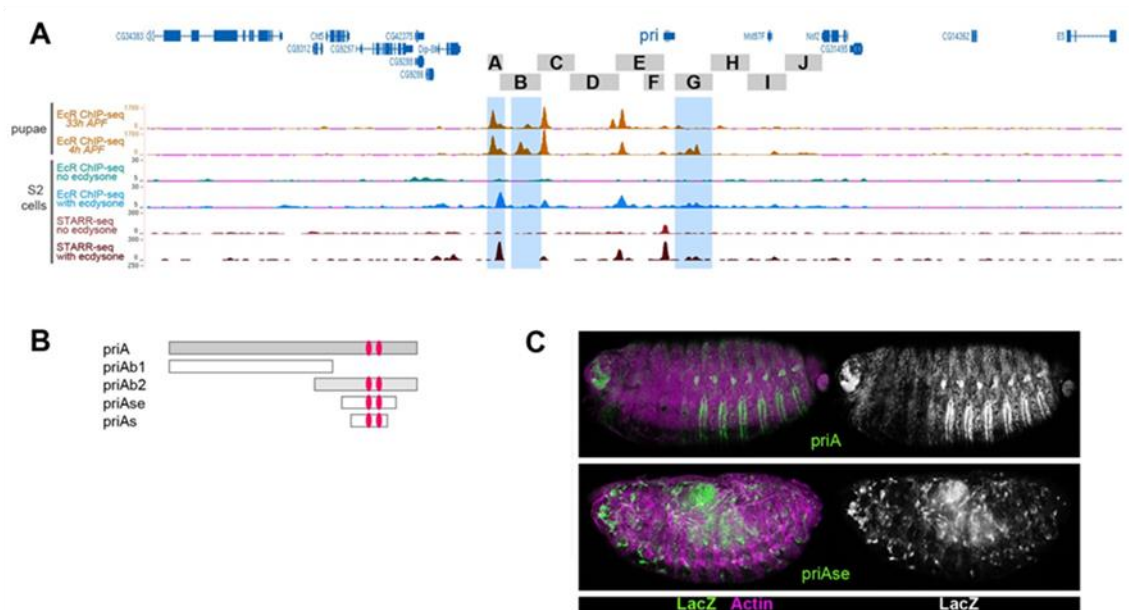
segmentation. Un projet collaboratif avec les équipes ayant réalisées ces observations a conduit à révéler que le complexe moléculaire comprenant les peptides Pri, l'ubiquitine-ligase Ubr3 et le facteur de transcription Shavenbaby (Svb), représente un ancien module de développement requis pour le modelage précoce de l'embryon dans plusieurs ordres d'insectes, à l'exception de la drosophile (Ray, 2019). J'ai contribué à ce projet en construisant une forme Svb activatrice constitutive.

### **3-6 Régulation temporelle de la différenciation épidermique.**

Cette analyse approfondie de la maturation de Svb a dévoilé un processus sophistiqué, en plusieurs étapes pour activer un réseau génique. Une étude entreprise par notre équipe en collaboration avec d'autres équipes internationales, a permis d'apporter une signification biologique au contrôle de l'activité de Svb par les peptides Pri. Cette étude, dirigée dans l'équipe par Hélène Chanut et à laquelle j'ai participé, établit que les peptides Pri assurent un contrôle temporel du processus de différenciation épidermique. En effet, l'expression du gène *pri* est directement sous le contrôle de l'hormone stéroïde ecdysone (Chanut-Delalande, 2014). L'ecdysone est connue pour réguler la temporalité du développement chez les insectes, notamment pour déclencher les grandes transitions développementales comme les mues larvaires et la métamorphose. Elle intègre des signaux externes (lumière, nutriments) (Andersen, 2013 ; Yamanaka, 2013) et internes (croissance des tissus) (Colombani, 2005 ; Delanoue, 2010) pour coordonner de manière systémique le développement de l'organisme. Les travaux montrent qu'en absence de Pri, le programme de formation des trichomes est bloqué, par l'accumulation de la forme répresseur de Svb. En réponse à l'ecdysone, les peptides Pri déterminent le moment auquel Svb est capable d'activer la transcription de ces cibles et induire la formation des trichomes. Ainsi les peptides Pri constituent un relai entre l'information systémique de l'ecdysone et le programme de différenciation des trichomes qui se produit dans les cellules épidermiques. L'ensemble de ce travail a montré que tandis que l'expression de Svb définit le registre spatial des cellules à trichomes, l'expression de *pri* définit quand ce programme de différenciation va s'exécuter.

Afin de mieux comprendre comment l'ecdysone régule l'expression de *pri*, l'équipe a choisi de disséquer de manière systématique les régions cis-régulatrices de *pri*. J'ai initié ce travail et Azza Dib, étudiante en thèse, l'a poursuivi. Nous avons défini que la région fonctionnelle de

pri s'étend sur 50 kb (**fig. 29**). Elle comprend une dizaine de régions régulatrices qui présentent des profils spatio-temporels distincts qui conduisent à l'expression de pri dans les tissus et au cours du développement. Nous avons montré que l'expression de pri dans un tissu donné peut impliquer plusieurs enhancers suggérant des activités redondantes. De plus, certains enhancers peuvent être actifs dans plusieurs tissus au cours du développement. La plupart de ces enhancers ont leur activité qui est directement contrôlée par le récepteur à l'ecdysone EcR. Un élément, priA semble avoir un rôle primordial comme intégrateur du signal ecdysone pour l'expression de pri. Il est actif dans la majorité des tissus exprimant pri et son activité est fortement modulée en présence d'ecdysone. De plus, l'élément priA contient une séquence de 288 pb qui a été identifiée par l'équipe d'Alexander Stark comme répondant de manière très significative à l'ecdysone dans les cellules S2, et dont l'activité a été abolie suite à la mutation de deux sites de liaison EcR (Shlyueva, 2014). Ces deux sites sont conservés au cours



**Figure 29 : Représentation schématique de l'activité des enhancers de pri au cours du développement.**

**A)** Vue de la région génomique englobant le locus *pri* : les régions *pri* testées dans les lignées transgéniques lacZ-reporter sont dessinées comme des boîtes grises. Les différentes pistes représentent des profils d'intensité de signal de CHIP-Seq EcR *in vivo* (4h et 33h APF) en orange, de CHIP-Seq EcR dans des cellules S2 avec ou sans traitement à l'ecdysone (bleu et vert, respectivement), de profilage STARR-seq d'enhancers actifs avec ou sans ecdysone (brun et rouge foncé) dans des cellules S2 (Chanut-Delalande, 2014).

**B)** Schéma de l'enhancer *priA* et de ses sous-régions qui ont été testées dans des lignées de rapporteurs transgéniques lacZ. Les ovales rouges représentent les sites de liaison EcR présentés en C.

**C)** Expression de *priA* et *priAse*, révélée par coloration fluorescente ( $\beta$ -Gal en vert, actine en magenta) ; les panneaux de droite montrent des images en noir et blanc du canal  $\beta$ -Gal pour un meilleur contraste. Toutes les images ont été acquises en utilisant les mêmes montages.

(d'après Dib, 2021).

de l'évolution chez toutes les espèces de drosophiles. Cependant, cet enhancer, que nous avons appelé priAs, n'était pas suffisant pour conduire une expression *in vivo* appropriée dans l'épiderme embryonnaire. Au lieu de cela, nous avons observé une expression ectopique dans des cellules éparses qui représentent probablement des cellules sanguines. Des résultats similaires ont également été observés avec une construction portant une version étendue priAse (421bp). Ces données indiquent que l'activité *in vivo* de priA nécessite des éléments cis-régulateurs supplémentaires de la séquence priA, qui peuvent inclure d'autres sites de liaison EcR. En conclusion, ces données montrent que pri intègre le signal ecdysone à travers un ensemble de régions régulatrices pour façonner son expression spatio-temporelle tout au long du développement (Dib et al., 2021).

## **Chapitre II : Visualisation de l'ADN.**



Comprendre l'organisation des noyaux eucaryotes, la dynamique spatiotemporelle du comportement de la chromatine ..., nécessita le développement d'outils pour visualiser l'ADN. Dans ce chapitre, je vais présenter différentes techniques et approches permettant de visualiser l'ADN, de sa découverte en 1869 avec les premières observations à l'aide de colorants jusqu'à aujourd'hui et les techniques permettant de visualiser des loci génomiques dans des cellules vivantes. Après avoir rapidement mentionné différentes techniques qui ont permis d'observer l'ADN dans des cellules fixées, je développerai les différentes approches actuelles permettant la visualisation de locus génomique unique dans des cellules vivantes.

## 1- Dans les cellules fixées.

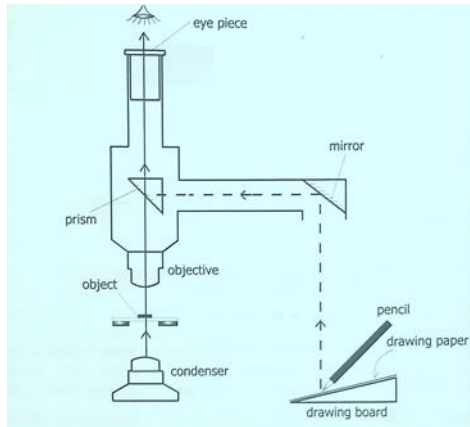
### 1-1 Visualisation de l'ADN en microscopie optique.

A partir de la découverte de l'ADN par Friedrich Miescher en 1869, les premières observations de l'ADN étaient effectuées à l'aide de colorants (**tableau 1**). Les communications scientifiques publiées à l'époque portaient essentiellement sur des descriptions et contenaient peu d'illustrations qui étaient des dessins, souvent obtenus à l'aide de la caméra Lucida. Cet appareil optique effectue une superposition optique de l'objet regardé et de la surface sur

Colorant	Caractéristique(s)	Date	Auteur(s)	Observations	Références
Aniline	Colorant basique ( à base de goudron de rouille)	1882	Flemming	Coloration des chromosomes en division: à donner le nom "chromatine". Dessins avec chambre claire ou "Caméra Lucida".	Fleming, 1882
Solution d'acide chromique	Fixateur	1908	King	Ovogénese chez le crapaud et chromosomes pendant la mitose. Dessins avec chambre claire ou "Caméra Lucida".	King, 1908
Janus green	Colorant basique	1912	Chambers, Kite	Coloration des chromosomes de testicule de punaise, sauterelle.	Kite, 1912
Colorant de Feulgen	Hydrolyse à l'acide chlorhydrique puis coloration avec la fuchsine (ou réactif de Schiff)	1924	Feulgen	Observation des chromosomes: bandes rosées pour les gènes, régions intergéniques plus claires. Encore très utilisé de nos jours, pour la quantification de l'ADN en cytométrie d'imagerie pour évaluer la ploïdie, en histologie.	ref pour imagerie: Blunt, 1932; Subramanian, 1946; Ris, 1949a, b. Très nombreuses publications, par exemple, pour mesurer la quantité d'ADN
Vert de méthyle/pyronine	Colorant basique: - vert de méthyle: colore en vert l'ADN. - pyronine: colore en rouge l'ARN.	1947	Brachet	Test Brachet: permet de localiser ADN et ARN après traitement par ARNase ou ADNase (utilisé en travaux pratiques).	Ris, 1949b

**Tableau 1: Colorants utilisés pour les premières observations de l'ADN en microscopie optique.**

laquelle l'observateur dessine. Il voit simultanément l'image de l'objet observé et la surface dessinée, comme dans une double exposition photographique (**fig 30**). L'une des premières publications de dessins représentant l'ADN fut au crédit de Walther Flemming qui étudiait le



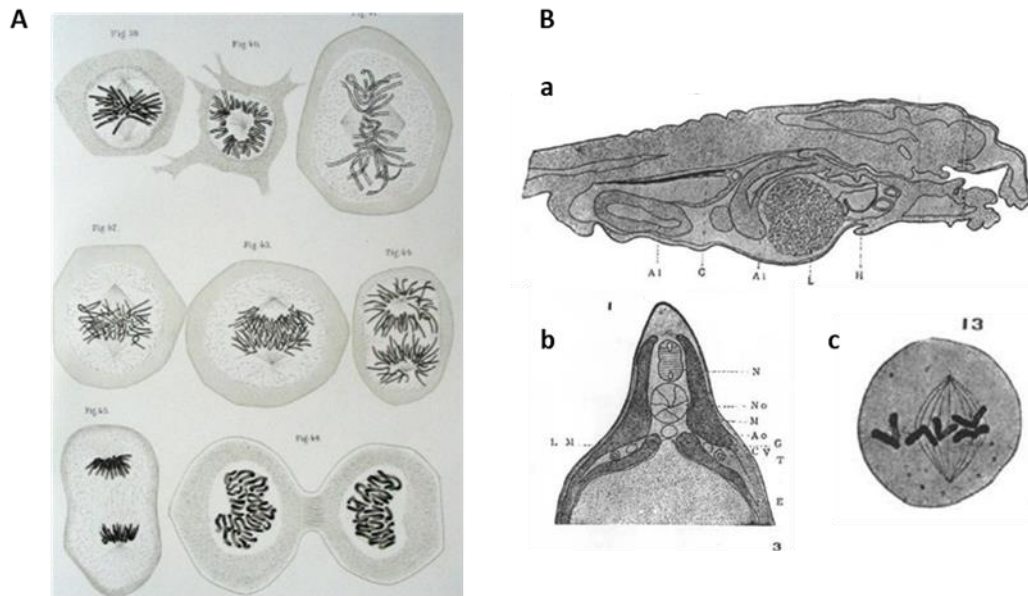
**Figure 30: la camera Lucida.**

Une caméra Lucida est un appareil optique utilisé pour dessiner le diagramme d'objets microscopiques. Elle a été inventée en 1806 par William Hyde Wollaston et Christian Gobrecht. La caméra Lucida effectue une superposition optique de l'échantillon observé et de la surface sur laquelle l'observateur dessine. La personne voit simultanément l'image de l'objet observé et la surface à dessiner, comme dans une double exposition photographique.

(schéma et légende tiré de <https://biolearnspot.blogspot.com/2020/02/camera-lucida.html>).

processus de division cellulaire et effectuait de nombreuses observations à partir de préparations fixées et colorées à l'aide d'un colorant basique à base d'aniline. C'est durant ces expériences qu'il identifie et nomme la chromatine du grec χρώμα (chrôma)=couleur) une substance acide qui capte les colorants basiques. En 1882, il publie un manuel de cytologie qui est illustré de nombreuses planches de dessin (**fig. 31 A** ; Flemming, 1882). En 1908, Helen Dean King rapporte des observations sur l'ovogénèse du crapaud. Dans sa publication figuraient de très nombreux dessins (**fig. 31 B** ; King, 1908).

En 1924, le chimiste allemand Robert Feulgen met au point une réaction colorimétrique permettant de colorer l'ADN, la réaction de Feulgen (Feulgen, 1924). Le principe de la méthode est de dissocier les deux brins d'ADN grâce à une hydrolyse ménagée par l'acide chlorhydrique (HCl) et de colorer ensuite ceux-ci au moyen de la fuchsine qui réagit avec les fonctions réductrices de l'ADN. L'intensité de coloration est proportionnelle à la concentration d'ADN. Elle permet l'observation au microscope photonique des chromosomes, notamment sous forme polyténique. Les bandes colorées rosées qui apparaissent correspondent aux gènes, et les bandes claires aux régions intergéniques. La coloration de Feulgen permet de visualiser l'ADN sur des coupes de tissus et dans les cellules. La coloration de Feulgen a été utilisée tout le long du XX<sup>ème</sup> siècle et a permis d'établir notamment, (i) que le contenu en ADN est le même entre un noyau et son nombre de chromosomes (Vendrely, 1948), (ii) qu'avant la mitose, les cellules doublent leur contenu en ADN (Patau, 1953), (iii) la quantité d'ADN présent dans un seul chromosome (Gaillard, 1968), (iv) la ploïdie dans des cellules



**Figure 31: Illustrations de l'ADN et de coupes cytologiques.**

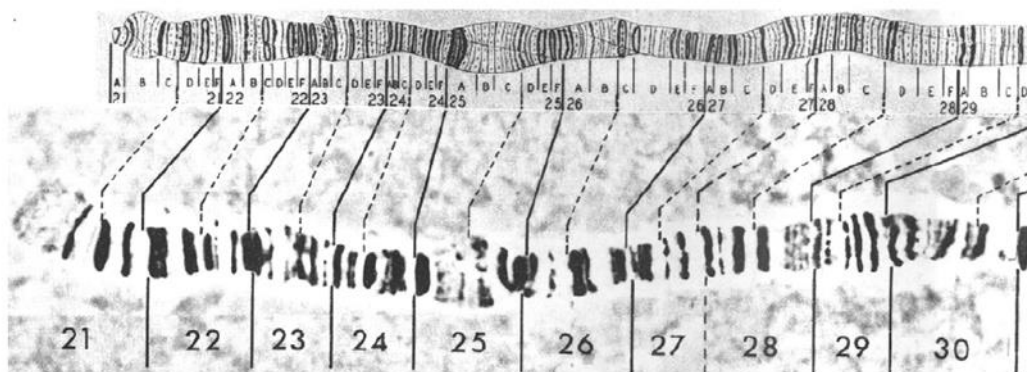
**A)** Illustration de cellules en mitose, une des 100 figures du livre de Flemming *Zellsubstanz, Kern und Zelltheilung*, 1882. Coloration à base d'aniline et dessin à l'aide de la caméra Lucida Flemming, 1882).

**B)** Figures dessinées à l'aide de la caméra Lucida par Helen den King. **a)** coupe longitudinale de la crête germinale chez un têtard âgé de treize jours, **b)** section transversale d'un têtard âgé de six jours, **c)** section longitudinale d'un spindle durant la métaphase, seulement 5 des chromosomes sont montrés (légende et figures tirées de King, 1908).

tumorales (Atkin, 1954). Encore de nos jours, cette coloration est très largement utilisée pour mettre en évidence l'ADN en histologie et dans le cadre de l'enseignement. Les applications actuelles de la réaction de Feulgen concernent principalement la quantification de l'ADN dans les noyaux cellulaires par cytométrie d'image pour l'évaluation de la ploïdie en pathologie tumorale. Par contre au niveau imagerie microscopique, elle est très peu utilisée de nos jours (Chieco, 1999). L'application des principes de Feulgen à la microscopie électronique a récemment permis de développer des procédures spécifiques de coloration de l'ADN pour l'étude de l'organisation structurale de l'ADN *in situ* : la réaction osmium-ammine de type Feulgen, peut être considérée comme une méthode de choix pour la visualisation de l'organisation structurale de structures contenant de l'ADN dans des coupes fines (Derenzini 1995).

Ces marquages de l'ADN avec ces colorants et les observations en microscopie optique qui en découlent ne permettent d'avoir qu'une vision globale de l'ADN. En ce début du XXème siècle, ce qui se pourrait se rapprocher le plus d'une visualisation de l'ADN au niveau de locus génomique serait l'observation des chromosomes polytènes. Ces chromosomes présents dans les cellules de certains organes de larves, comme dans les glandes salivaires de la drosophile,

subissent plusieurs cycles de réplication sans division cellulaire, phénomène d'endoréplication, aboutissant à un nombre élevé de copies de chaque chromosome, ou degré de polyténie, qui peut atteindre jusqu'à 1024. Ces copies de chromatides ne se dissociant pas à la suite des réplications successives restent accolées les unes aux autres et alignées de manière parallèle ce qui donne lieu à des chromosomes fortement agrandis. Au moment de l'interphase, une coloration de Feulgen permet, à l'aide d'un simple microscope photonique, d'observer ces chromosomes polytènes qui présentent une succession stéréotypée de bandes fortement colorées, correspondant à des régions de chromatide condensée (ou hétérochromatine) et de bandes plus claires possédant une chromatine moins condensée (ou euchromatine). Ces observations ont permis à Calvin Bridges, étudiant de Thomas Morgan, d'établir une cartographie précise des chromosomes de la drosophile (**fig 32**). L'ensemble du génome de la drosophile a été divisé en 102 divisions. Le chromosome X comprend les divisions de 1 à 20. Les divisions de 20 à 60 correspondent au chromosome II, 20 à 40 pour le bras gauche, 2L et 40 à 60 pour le bras droit, 2R. Le chromosome III est divisé de 60 à 80 et de 80 à 100, respectivement pour le bras 3L et 3R. Enfin les 2 dernières divisions sont attribuées au chromosome IV. Chaque division est divisée en 6 sections, désignées par les lettres A à F et elles-mêmes en 13 sous-sections (Bridges, 1935, 1938), correspondant à 7956 sections pour l'ensemble du génome de la drosophile ! (Bridges, 1938 ; Lefevre, 1976). Ce qui permet d'avoir une résolution suffisante pour observer des bandes d'environ 20kb (Scalenghe, 1981).



**Figure 32: Extrémité du chromosome 2L de la drosophile.**

La partie supérieure représente un dessin d'après la photographie située en dessous prise par Bridges. Les divisions 20 à 30 correspondent à l'extrémité du bras gauche 2L du chromosome II. Chaque division est divisée en 6 sections, de A à F.

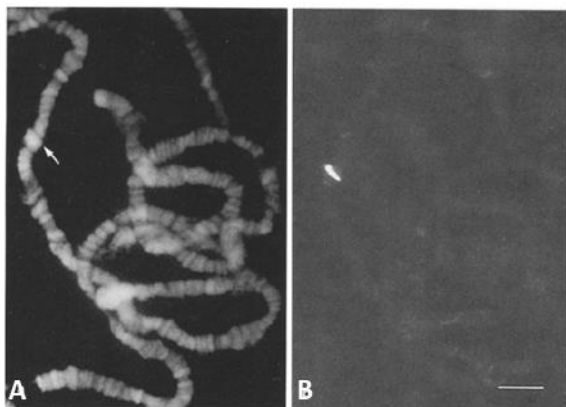
(d'après Bridges, 1938 et Lefevre, 1976).

### 1-2 Visualisation de l'ADN par rayonnement.

Comment ne pas mentionner les techniques de cristallographie par rayon X pour visualiser l'ADN ! Certes, ce ne sont pas les techniques les plus appropriées si l'on s'intéresse à la dynamique de la chromatine ou à l'importance du rôle de l'architecture nucléaire dans la régulation de la transcription des gènes mais elles ont permis la découverte de la double hélice de l'ADN. Et la photo 51, cliché de la diffraction de rayons X, prise par Rosalind Franklin est probablement l'une des photographies scientifiques les plus importantes.

### 1-3 Visualisation de l'ADN par DAPI.

Otto Dann et collaborateurs ont synthétisé pour la première fois en 1971 le 4'-6-Diamidino-2-Phenylindole ou DAPI (Dann, 1971) qui a montré posséder des propriétés de liaison à l'ADN (Williamson, 1974). Russell et ses collaborateurs vont utiliser le DAPI comme un colorant fluorescent hautement spécifique pour l'ADN pour détecter les contaminations de mycoplasme en culture de cellule. En ajoutant du DAPI à des cellules de culture de tissus, ils ont constaté qu'il était rapidement absorbé par l'ADN cellulaire, produisant des noyaux hautement fluorescents et aucune fluorescence cytoplasmique détectable dans les conditions utilisées. Si, par contre, les cellules sont contaminées par des mycoplasmes, des foyers fluorescents discrets caractéristiques sont détectés dans le cytoplasme et à la surface des cellules (Russell, 1975). En 1980, Bauman et ses collaborateurs effectuent une coloration avec le DAPI de chromosomes polytènes en parallèle d'une hybridation *in situ* utilisant un ARN marqué par un fluorochrome (Bauman, 1980). Après l'hybridation d'un ARN 5S de *Drosophila hydei* marqué avec la rhodamine sur des chromosomes polytènes de *Drosophila melanogaster*, une bande est très fluorescente. L'identification de la bande à la position 56F a été facilement réalisée après coloration avec du DAPI, par l'observation alternative de la fluorescence bleue du DAPI et de la fluorescence rouge de la rhodamine en changeant la longueur d'onde d'excitation (**fig. 33** ; Bauman, 1981).



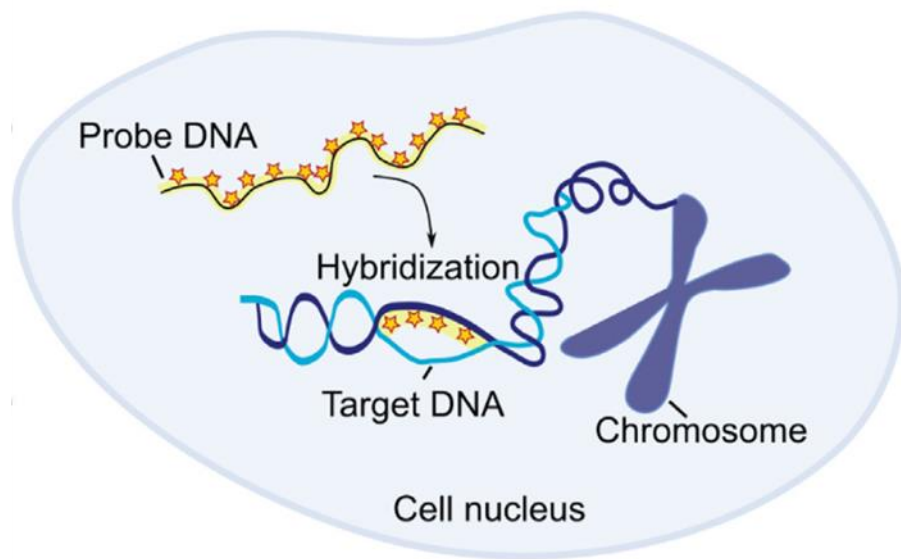
**Figure 33: Marquage des chromosomes polytènes de *Drosophila melanogaster* avec une coloration au DAPI et une hybridation *in situ* avec l'ARN 5 S de *Drosophila hydei* marqué à la rhodamine.**

La coloration au DAPI des chromosomes polytènes de *D. melanogaster* **A**) permet de localiser la bande 56F révélée par hybridation *in situ* avec l'ARN 5S de *D. hydei* marquée à la rhodamine **B**).  
(figure tirée de Bauman, 1981).

#### **1-4 Visualisation de l'ADN par hybridation fluorescente *in situ* (FISH).**

L'hybridation fluorescente *in situ* ou FISH est une technique qui est utilisée pour la détection spatiale et la quantification des acides nucléiques dans leur environnement cellulaire. Elle est devenue une méthode cytogénétique puissante pour l'analyse des cellules et des tissus au niveau du transcriptome et du génome. Dans le domaine du diagnostic, cette technique permet de détecter les cellules malignes contenant des réarrangements chromosomiques (Carter, 1992) ou des aberrations génétiques (Weimer, 2000).

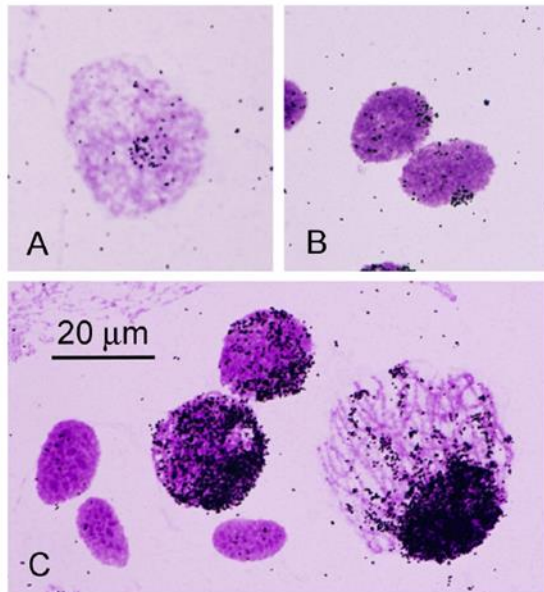
La technique de FISH est une évolution de la méthode originelle, l'hybridation *in situ* ou ISH. La différence entre ISH et FISH se situe au niveau des sondes utilisées : dans la technique FISH, les sondes utilisées sont constituées d'un marqueur fluorescent. L'ISH est une technique cytogénétique permettant la détection, la quantification et la localisation d'acides nucléiques cibles à l'intérieur de cellules ou tissus. La méthode est basée sur l'hybridation de sondes contenant des séquences spécifiques complémentaires (ADN ou ARN) de leur cible à l'intérieur de la cellule (**fig 34** ; Huber, 2018). La sonde d'acide marquée par fluorescence entre dans le noyau de la cellule où il s'hybride à sa cible complémentaire. Les sondes peuvent être ciblées sur l'ADN ainsi que sur l'ARN ce qui permet l'étude des séquences génomiques et des profils d'expression transcriptomiques de cellules individuelles. La réaction peut être visualisée directement à l'aide de sondes marquées par radioactivité ou par fluorescence ou indirectement via la liaison d'antigènes ou par des interactions biotine/streptavidine (Huber, 2018).



**Figure 34: Hybridation *in situ* pour la visualisation de cibles chromosomiques.**

Les sondes ADN (**jaune**) entrent dans les cellules perméabilisées et les noyaux et s'hybrident à leurs séquences ADN cibles complémentaires. Dans les régions où les sondes marquées sont hybridées à leur ADN cible, un signal ISH peut être détecté. (figure et légende tirées de Huber, 2018).

Cette technique d'ISH a été décrite pour la première fois en 1969 par Pardue et Gall (Gall, 1969). Joseph Gall s'intéressait aux chromosomes géants, à ce qu'ils pourraient nous dire sur la structure et la fonction des gènes et notamment les gènes codant pour les ARNs ribosomiques. Avec son étudiante Mary Lou Pardue, ils ont mis beaucoup d'efforts pour développer une méthode permettant de visualiser les gènes sur les chromosomes polytènes. En ce début des années 60, l'utilisation d'anticorps fluorescent n'est pas encore optimale mais Gillespie et Spiegelman ont développé une technique d'hybridation ARN-ADN : ils ont fixé de l'ADN simple brin dénaturé à un filtre de nitrocellulose qui est incubé avec une solution d'ARN marquée au  $P^{32}$  (Gall, 2016). Cette technique s'est avérée extrêmement utile pour l'analyse moléculaire des hybrides ARN-ADN (Gillespie, 1965). Joseph Gall ayant utilisé cette technique à de nombreuses reprises, pensait que c'était possible de l'adapter à des préparations cytologiques. L'idée était, sur un tissu écrasé plat, de dénaturer l'ADN des noyaux sur une lame et de l'hybrider avec de l'ARNr marqué au  $^3H$ . Il a d'abord essayé sur des cellules sanguines d'amphibien mais il ne voyait rien : le problème venait d'une activité spécifique très basse de l'ARN marqué au  $^3H$  et d'une quantité d'ADNr trop faible dans des noyaux diploïdes. Son salut viendra de préparations cytologiques écrasées à partir d'ovocytes de *Xenopus* (fig. 35, Gall, 2016).



**Figure 35: Autoradiographies de noyaux d'ovaire de *Xenopus* après hybridation *in situ* avec de l'ARN ribosomal marqué au  $^3\text{H}$ .**

**A)** Noyau d'oogone (gamétocyste femelle)  
**B)** Deux noyaux leptotènes (première phase de la prophase I de la méiose).

**C)** Trois noyaux folliculaires non marqués et trois noyaux pachytènes (3ème stade de la prophase I) montrant l'augmentation progressive de l'ADNr amplifié à mesure que le noyau s'agrandit. Images récentes en couleur prises à partir de l'une des premières préparations d'hybridation *in situ* réalisées en 1968 par Joseph Gall et Mary Lou Pardue, publiées à l'origine comme Fig. 1 dans Gall, 1969.

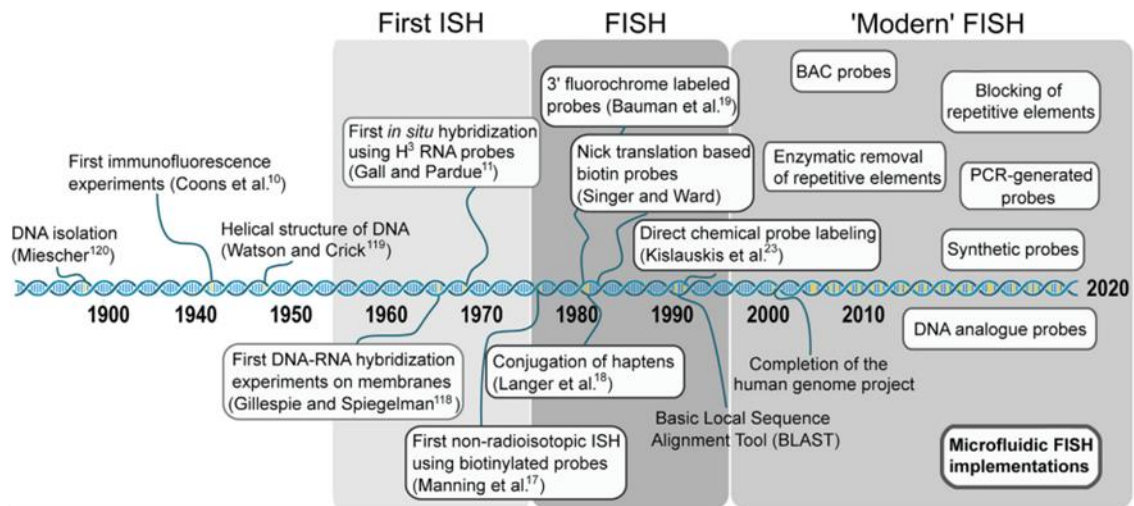
(figure et légende tirées de Gall, 2016).

Parallèlement, John et al (John, 1969) et Buongiorno-Nardelli et Amaldi (Buongiorno-Nardelli, 1970) ont développé indépendamment des techniques d'hybridation *in situ* pour la détection de l'ADNr chez le Xénope et dans des tissus inclus en paraffine. Quelques années plus tard, Gall et Pardue ont visualisé des éléments satellites dans les régions hétérochromatiques des chromosomes de souris démontrant ainsi l'utilisation de l'ISH dans les cellules de mammifères (Pardue, 1970).

Les limites de ces premières expériences ISH étaient la faible sensibilité et la disponibilité limitée de sondes spécifiques à la séquence. Bien que le marquage radio-isotopique soit toujours considéré comme la méthode de détection la plus sensible, il nécessite de longs temps d'exposition (jusqu'à plusieurs semaines pour la détection de  $^3\text{H}$ ), la résolution spatiale est faible et les bruits de fond sont élevés. Ces inconvénients ont conduit au développement d'approches de marquage non isotopique pour les sondes ISH (Huber, 2018).

Au milieu des années 1970, une gamme de sondes non radio-isotopiques a été développée, permettant de nouvelles options de détection des hybrides sonde-cible. La détection directe par fluorescence de cibles chromosomiques a été signalée par Bauman et al. dans les années 1980, qui ont marqué les extrémités 3' des sondes avec un fluorochrome (Bauman, 1980). Les sondes avec des bases marquées à la fluorescéine incorporées par « nick-translation » ont montré des sensibilités de détection de 50-100 kb (Wiegant, 1991) et ont été utilisées pour les premières expériences de FISH multicolores... Dans la **figure 36**, on peut voir la chronologie des développements de l'hybridation *in situ* en fluorescence (Huber, 2018).





**Figure 36: Chronologie des développements de l'hybridation in situ en fluorescence.**

La première d'hybridation in situ a été réalisée par Gall et Pardue en 1969 (Gall, 1969). Les premières versions fluorescentes de la technique FISH sont apparues dans les années 1970, suivies par le marquage direct des sondes vingt ans plus tard. La FISH "moderne" comprend des développements dans la conception et la production des sondes ainsi que dans la réduction du bruit de fond. La combinaison de la microfluidique et de la FISH est apparue au début du 21e siècle. (figure et légende tirées de Huber, 2018).

Les progrès réalisés dans le domaine de la conception des sondes et des stratégies de marquage ont permis à la FISH de devenir une technique très polyvalente permettant de traiter un large éventail de cibles cellulaires avec une spécificité et une sensibilité élevée. La FISH peut donc être utilisée pour dénombrer les chromosomes, détecter les aberrations chromosomiques, étudier des régions chromosomiques spécifiques, ou des niveaux de transcription. Ces aspects ne sont pas seulement intéressants pour les applications de recherche en cytogénétique, mais ils fournissent également la base de diagnostics moléculaires dans une série de domaines médicaux (Huber, 2018).

Cependant, malgré toutes ces qualités et avantages, cette technique de FISH s'effectue dans des cellules fixées et ne permet pas de visualiser l'ADN dans des cellules vivantes.

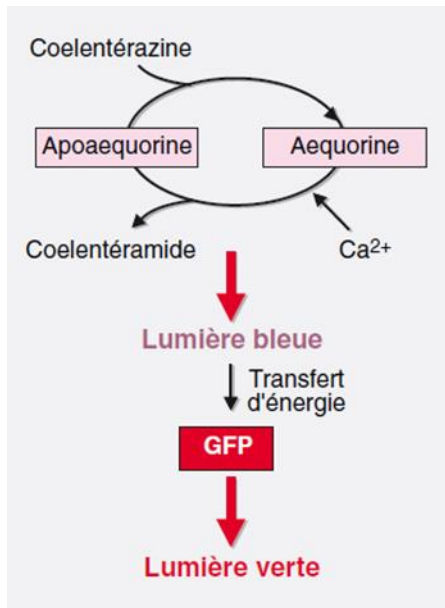
## 2- Visualisation de loci génomiques uniques et d'ARNm en cellules vivantes.

### 2-1 La Green Fluorescent Protein ou GFP.

Comprendre les mécanismes d'un événement qui se produit au sein d'un organisme ou d'une cellule vivante est un enjeu de la biologie cellulaire et moléculaire. L'avènement de la protéine fluorescente verte ou GFP au début des années 90 a constitué une avancée importante permettant la visualisation d'événements cellulaires dans leur environnement. La GFP est devenue en quelques années **l'un des outils les plus puissants de la biologie moléculaire et cellulaire.**

Le gène de la GFP est aujourd'hui utilisé comme gène rapporteur, comme marqueur cellulaire ou encore comme marqueur moléculaire. La fusion de la GFP à des protéines a rendu possible, grâce à la fluorescence émise par la GFP, de localiser des protéines dans les compartiments cellulaires, d'en définir les déplacements lors de stimulations diverses, rendant accessible la visualisation d'événements cellulaires dynamiques. En permettant un marquage facile des protéines chromosomiques et nucléaires et le marquage de régions chromosomiques spécifiques, l'utilisation de la GFP a fourni des possibilités nouvelles pour l'observation directe de la dynamique des chromosomes *in vivo*. Des questions concernant la structure et la dynamique des chromosomes en interphase à la régulation de la transcription, de la réplication, de la recombinaison et la réparation de l'ADN ont été rendues possible à aborder grâce à cette petite protéine fluorescente.

Ce sont les travaux d'Osamu Shimomura qui ont permis la découverte de la GFP. Il travaillait sur la compréhension de l'émission de bioluminescence verte par la méduse *Aequorea victoria*. En 1962, il extrait et purifie la protéine aequorine de la méduse *A. victoria* (Shimomura, 1962). Mais l'activation de l'aequorine *in vitro* ou dans des cellules produit une lumière bleue, alors que la méduse produit une lumière verte ! Cette lumière est le résultat d'une seconde protéine présente chez *A. victoria* qui tire son énergie d'excitation de l'aequorine, la protéine fluorescente verte, GFP (Chalfie, 1994). En 1974, Morise et ces collègues ont purifié, cristallisé et caractérisé la GFP et ils ont démontré, *in vitro*, le transfert d'énergie de l'aequorine à la GFP (Morise, 1974). Dans la **figure 37** est illustré le mécanisme d'action de la GFP dans la méduse *Aequorea victoria* (Georget, 1999).



**Figure 37: Mécanisme d'action de la GFP dans la méduse *Aequorea victoria*.**

En présence d'oxygène, l'aequorine est une protéine dépendante du calcium couplée à la coelentérazine. La fixation de calcium provoque un changement conformationnel de l'aequorine qui induit une oxygénation de la coelentérazine. Cette réaction d'oxydation s'accompagne d'une émission de lumière bleue. Il s'établit alors un transfert d'énergie entre l'aequorine et la GFP. Ce transfert d'énergie conduit à l'émission d'une lumière verte de la GFP.

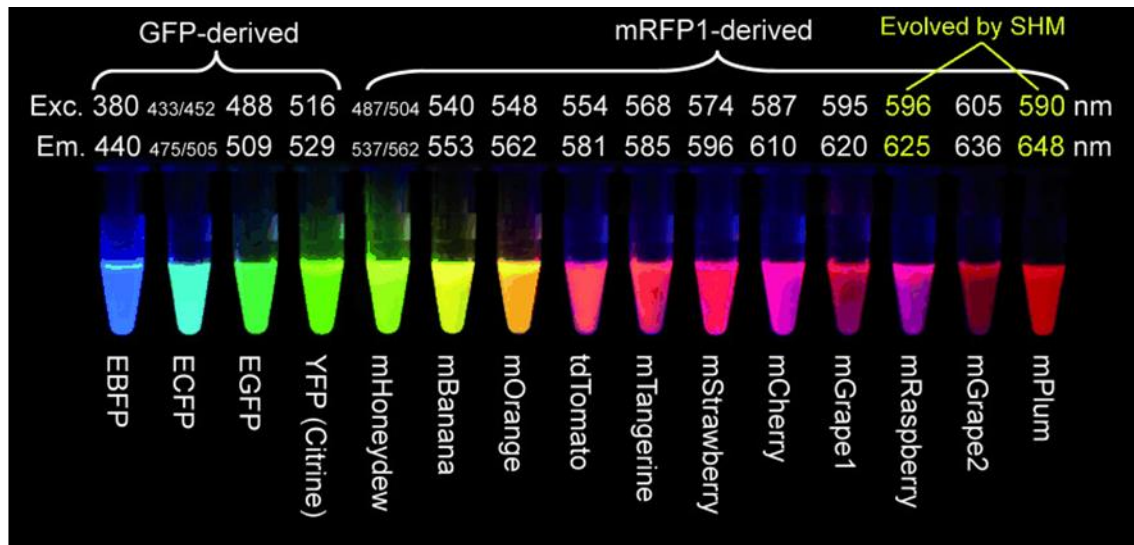
(figure et légende tirées de Georget, 1999).

Le gène codant pour la GFP a été cloné par Prasher en 1992 (Prasher, 1992). Il y a 30 ans !!!! En 1994, Chalfie exprime pour la première fois la GFP chez *E. coli* et chez *Caenorhabditis elegans* (Chalfie, 1994). La même année, Tulle Hazelrigg exprime une protéine de fusion avec la GFP dans des lignées transgéniques de *Drosophila melanogaster*. Elle exprime la protéine Exu fusionnée à la GFP dans les cellules germinales femelles et montre que la protéine de fusion fluoresce fortement dans des cellules vivantes ou fixées pendant l'ovogénèse de la drosophile. Cette protéine de fusion permet de restaurer un allèle nul pour le gène *exu* et, s'exprime et se localise dans un patron spatial et temporel comme le gène sauvage. La haute sensibilité de la GFP fournit d'importants nouveaux détails sur la localisation subcellulaire de la protéine Exu (Wang, 1994).

La protéine GFP présente de nombreux avantages. C'est une molécule de 238 acides aminés. Elle absorbe la lumière bleue à 395nm pour mettre une lumière verte à 509nm. Sa fluorescence est très stable et non invasive. Il suffit de l'éclairer avec la bonne longueur d'onde. La GFP s'accumule dans les cellules en vie en temps réel, ce qui permet d'observer les changements d'expression au cours du temps et elle ne semble interagir avec aucun processus biologique d'intérêt. De nombreux variants spectraux ont été développés notamment par Roger Tsien (**fig. 38** ; Tsien, 2009).

En 2008, les travaux pionniers de Shimomura, Tsien et Chalfie ont été récompensés par un prix Nobel de chimie pour « la découverte et le développement de la protéine fluorescente

verte ». Parmi les nobélisés ne figure pas Douglas Prasher qui a pourtant largement contribué à la découverte de la GFP en clonant son gène ...



**Figure 38: Diversité spectrale des protéines de fluorescence.**

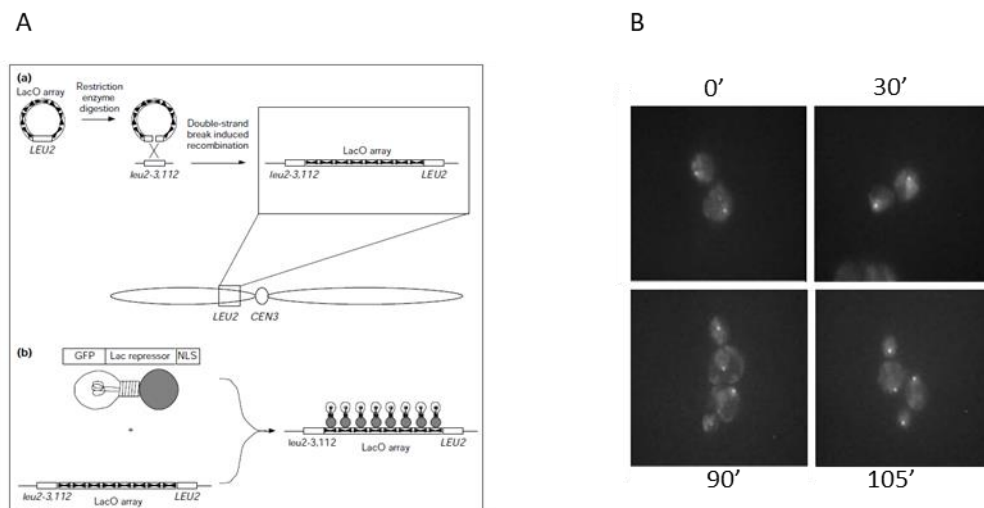
Protéines fluorescentes monomères et dimères en tandem dérivées de la GFP d'*Aequorea* ou de la RFP de *Discosoma*, exprimées dans des bactéries et purifiées. Cette photo est une exposition temporelle de fluorescences excitées à différentes longueurs d'onde et vues à travers différents filtres de coupure. SHM=somatic hypermutation (figure et légende tirées de Tsien, 2009).

## 2-2 Les Systèmes Fluorescents Répresseur/Opérateur : FROS.

L'introduction de nucléotides et de protéines marqués par fluorescence dans les cellules a permis l'observation directe de la dynamique des chromosomes en mitose et en interphase *in vivo*. Cette approche a été considérablement améliorée par les applications de la GFP. Non seulement des protéines nucléaires spécifiques peuvent être marquées avec la GFP mais, en utilisant des répétitions directes de séquences d'opérateurs bactériens, il est possible de marquer des sites chromosomiques spécifiques et de les visualiser en utilisant la GFP fusionnée avec la protéine répresseur appropriée.

La visualisation de l'ADN à des loci génomiques spécifiques nécessite des techniques de marquage qui créent un foyer fluorescent détectable au-dessus du bruit de fond. Jusqu'à la fin des années 90, il n'y avait pas de méthode pour visualiser ces régions dans des cellules vivantes. La première méthode développée utilisait des protéines de liaison à l'ADN fusionnées à des protéines fluorescentes qui reconnaissaient leurs séquences de fixation apparentées. Elle est basée sur le système opérateur (LacO)/répresseur (LacI) provenant de l'opéron lactose de *E. coli*. Cette technique, dénommée FROS pour Système fluorescent

Opérateur/Répresseur, est basée sur l'insertion de nombreuses répétitions de séquences de l'opérateur (LacO) sur lequel le répresseur vient se lier. Profitant de la découverte de la GFP à cette même période, le répresseur est fusionné à une protéine fluorescente. Straight et collaborateur ont inséré 256 copies de l'opérateur LacO à un site précis du génome de la levure qui a été détecté par la liaison du répresseur LacI fusionné à la GFP (fig. 39 ; Straight, 1996 ; Robinett, 1996).



**Figure 39: Première utilisation du système FROS.**

Les chromosomes de la levure peuvent être marqués par intégration dirigée de répétitions de l'opérateur *LacO*.

**A) a)** Les répétitions de l'opérateur *LacO* ont été intégrées par recombinaison homologue près du centromère du chromosome III de *Saccharomyces cerevisiae*. Le plasmide portant les sites *LacO* a été coupé avec *EcoRV* pour cibler l'intégration sur le locus *LEU2* qui se trouve à 22 kb à gauche du centromère. **b)** Schéma de la structure de la fusion répresseur GFP-LacI et de sa liaison au réseau *LacO*.

**B)** Le répresseur GFP-LacI montre la séparation des 2 chromatides sœurs chez la levure au cours du temps.

(figure et légende tirées de Straight, 1996).

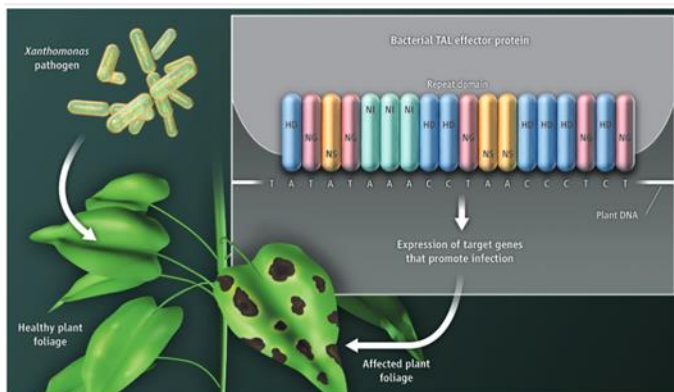
D'autres systèmes FROS ont été développés à partir de couple opérateur/répresseur, (i) à partir du répresseur tétracycline TetR qui se lie sur l'opérateur *tetO*, [TetR/*tetO*] (Lau, 2003 ; Nguen, 2019 ; Tazan, 2020), (ii) à partir du système répresseur/opérateur du phage lambda où le répresseur  $\lambda O$  va se fixer sur les séquences de son opérateur  $\lambda Ci$ , [ $\lambda O$ / $\lambda Ci$ ], (Lassadi, 2015).

La combinaison de différents systèmes FROS permet de marquer différentes régions chromosomiques. Lau et collaborateur ont combiné le système LacI/*lacO* et TetR/*tetO* pour observer l'organisation spatiale et temporelle des chromosomes d'*E. coli* au cours de la réplication. Les régions proches de l'origine de réplication et du terminus de réplication ont

été visualisées simultanément par le système TetR/*tetO* et LacI/*lacO*, respectivement. Cela leur a permis de montrer que la région terminale de la réplication est fréquemment localisée de manière asymétrique, sur le côté du nouveau pôle de la cellule. Cette asymétrie pourrait constituer un mécanisme par lequel la protéine de ségrégation chromosomique FtsK, située au niveau du septum de division, peut agir de manière directionnelle pour s'assurer que la région septale est libre d'ADN avant la fin de la division cellulaire (Lau, 2003).

### 2-3 Autres.

- système TALE (Transcription Activator Like Effector) : les effecteurs TAL ou TALE sont des protéines de liaison à l'ADN séquence spécifique. Elles ont été découvertes chez la bactérie pathogène de plante, *Xanthomonas*. Ces protéines contiennent un nouveau motif de liaison à l'ADN qui jouent un rôle central dans la fonction des effecteurs type activateur de transcription (TALE). Les TALEs sont une arme importante dans la bataille entre l'espèce *Xanthomonas* et leurs plantes hôtes. Ils sont transcrits dans la bactérie et versés dans la plante où ils vont activer la transcription de gènes de la plante leur permettant de se répandre. (Voytas, 2009). Ainsi le TALE, PthXo1 d'un pathogène *Xanthomonas* du riz active l'expression du gène du riz, *Os8N3*, permettant à *Xanthomonas* de coloniser le riz (Yang, 2006). Comme *Os8N3* est également nécessaire au développement normal de la plante, le gène ne peut pas simplement être éliminé pour éviter l'infection. Au contraire, il existe une forte pression sélective pour que la plante accumule les mutations qui empêchent la liaison de PthXo1 et, par conséquent, des changements compensatoires dans la spécificité de l'ADN de l'effecteur TAL. En fin de compte, la reconnaissance de l'ADN détermine l'issue de la guerre entre le pathogène et la plante, et donc les TALEs bénéficieraient d'un mécanisme simple et malléable pour la reconnaissance de l'ADN (Voytas, 2009). Deux études ont simultanément déchiffré le code de reconnaissance des effecteurs TAL (Moscou, 2009 et Boch, 2009). Les TALEs se lient à l'ADN à travers des répétitions en tandem (environ 17-19) de 33 à 35 acides aminés, chacune d'entre elles reconnaissant une seule paire de base spécifique par l'intermédiaire de deux acides aminés adjacents que l'on appelle des di-résidus variables en fonction des répétitions. Les séquences des répétitions sont conservées sauf 2 acides aminés adjacents très variable (à la position 12 et 13) qui sont les candidats pour le déterminant de spécificité (**fig. 40** ; Voytas, 2009 ; Miyanari, 2013). Grâce à ce code simple, les TALEs se sont révélés être une plateforme

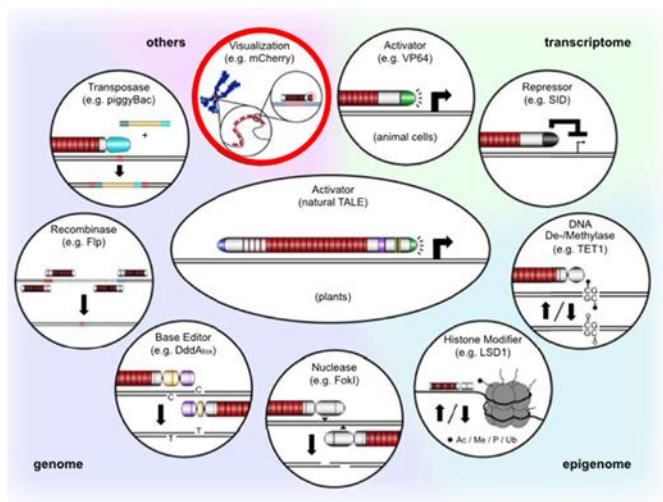


**Figure 40: Mécanisme d'infection de *Xanthomonas*.**

Le domaine de liaison à l'ADN des effecteurs TAL bactériens contient des répétitions en tandem (environ 17; illustrées en bleu, rose, jaune et vert pour AvrBs3) constituées de 34 ou 35 acides aminés. La répétition est hautement conservée, à l'exception des résidus aux positions 12 et 13 (indiqués dans chaque répétition ; H, histidine ; D, acide aspartique ; N, asparagine ; G, glycine ; I, isoleucine ; S, sérine). Le code relie les acides di-aminés dans l'unité de répétition aux nucléotides (T, thymine ; A, adénine ; C, cytosine) dans l'ADN cible. (figure et légende tirées Voytas, 2009).

polyvalente pour l'ingénierie de protéines de liaison à l'ADN dotées d'une fonctionnalité spécifique par la fusion des TALE avec des protéines opérationnelles comme des nucléases, des modulateurs transcriptionnels ou des recombinases (**fig. 41**). Mais les TALEs peuvent être fusionnés avec des protéines fluorescentes pour visualiser le positionnement sub-nucléaire de séquences cibles (**fig. 42**).

La principale limitation de ce système de visualisation à l'aide des TALEs est qu'il est restreint à des séquences répétées naturelles comme les télomères.



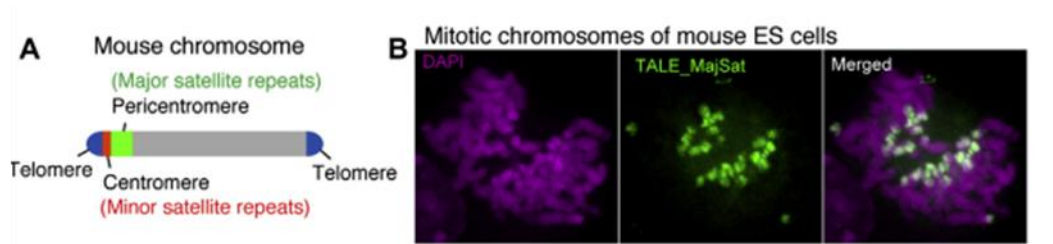
**Figure 41: Différentes applications à partir de l'utilisation des TALEs,.**

En utilisant des domaines fonctionnels d'autres protéines comme partenaires de fusion, divers outils basés sur les TALEs ont été générés. Ces outils peuvent être utilisés pour modifier le transcriptome (activateur et répresseur), l'épigénome (ADN De-/méthylase et modificateur d'histone) ou le génome (nucléase, éditeur de base, recombinase et transposase).

Ils peuvent également être utilisés pour **visualiser des séquences** d'ADN répétitives comme les centromères ou les télomères.

(Ac = Acétylation, Me = Méthylation, P = Phosphorylation, Ub = signal d'Ubiquitination.

(figure et légende tirées Becker, 2021).



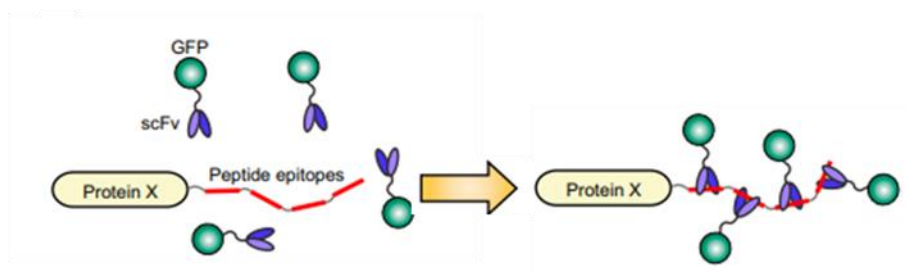
**Figure 42: Marquage spécifique de la séquence de satellite majeur par TALE dans des cellules vivantes.**

**A)** Schéma d'un chromosome de souris. L'emplacement des télomères (bleu), des satellites majeurs péri-centromériques (vert), des satellites mineurs centromériques (rouge) et du bras long du chromosome (gris).

**B)** Images fluorescentes de chromosomes mitotiques de cellules ES exprimant TALE-mClover contre 15 nt de séquence satellite majeur. mClover en vert; DAPI en magenta.

(figure et légende tirées, Miyanari, 2014).

● Sun Tag, CRISPR/Cas9 : le peptide Sun Tag, constitué d'une multimérisation de 4 copies d'épitopes en tandem, est fusionné en C-term d'une protéine. Les épitopes du peptide Sun Tag sont reconnus par des fragments variables simple chaîne (scFv) d'anticorps qui sont fusionnés à la GFP. Cette fixation du fragment scFv sur le peptide Sun Tag permettra de recruter 24 copies de la GFP au niveau d'une protéine cible et de créer un foyer de fluorescence (**fig. 43** ; Tanenbaum, 2014). L'utilisation de ce peptide pour visualiser des loci



**Figure 43: Principe du Sun Tag.**

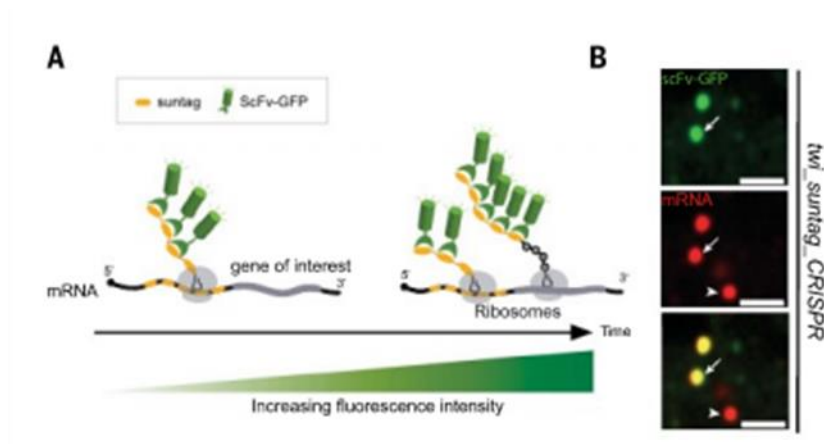
Le peptide Sun Tag en C-term d'une protéine va recruter 24 copie de GFP, suffisant pour créer un foyer de fluorescence.

(figure tirée de Tanenbaum, 2014).

génomiques est restreinte à des protéines qui se lient sur des séquences d'ADN répétées, comme les télomères. De même, l'utilisation de CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) et de la protéine associée CRISPR inactivée ou morte (dCas9) fusionnée à des protéines fluorescentes peut être une alternative intéressante et prometteuse pour visualiser des sites qui sont spécifiquement ciblés par de courts ARN guides complémentaires des séquences génomiques mais elle est essentiellement restreinte à des séquences répétées.



Récemment l'équipe de Mounia Lagha a visualisé la dynamique de traduction dans des embryons vivants (Dufourt, 2021). Ils ont utilisé le principe du peptide Sun Tag. Des répétitions d'épitopes Sun Tag sont ajoutés à la protéine d'intérêt et au cours de la traduction, ils sont détectés avec les fragments scFv couplé à la GFP (**fig. 44**). Cette approche leur a permis de visualiser et de quantifier le timing, la localisation et les cinétiques de traduction d'ARNm unique dans des embryons vivants. En regardant la traduction du facteur de transcription Twist, ils ont identifié une hétérogénéité spatiale dans l'efficacité de la traduction de l'ARNm et révélé l'existence d'usines de traduction, où des ARNm groupés sont co-traduits préférentiellement dans les régions périnucléaires basales (Dufourt, 2021).



**Figure 44: Visualisation de la traduction d'ARNm dans des embryons de *Drosophila melanogaster*.**

**A)** Principe du système Sun Tag.

**B)** Images d'embryons au cycle n.c.14: en vert expression de la scFv-GFP et en rouge une sonde suntag. Deux groupes d'ARNm sont visibles, un qui co-localise avec la scFv-GFP (flèche) et une autre qui ne colocalise pas (tête de flèche).

(figure tirée de Dufourt, 2021).

### 3- Organisation et Ségrégation des génomes bactériens.

Les travaux de ma thèse sont basés sur l'utilisation du système de visualisation de loci génomiques, le système ANCHOR. Ce système a été développé à partir du système de partition des génomes bactériens, le système ParABS. Avant de présenter ce système, je vais tout d'abord discuter de l'organisation des génomes bactériens puis de leurs ségrégations.

### 3-1 Organisation des génomes bactériens.

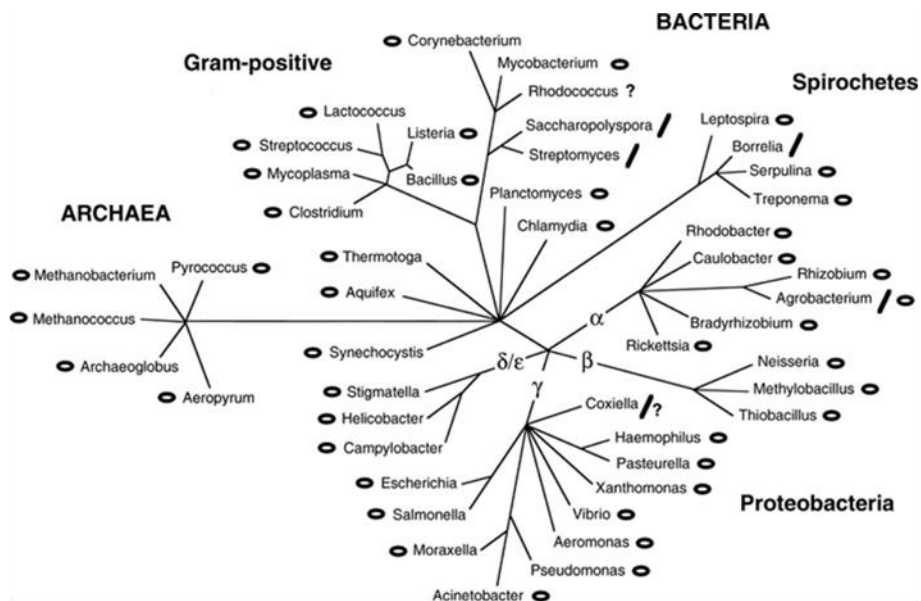
#### 3-1-1 Organisation en chromosomes et plasmides.

Le génome d'une cellule est essentiel à son identité : Il contient l'ensemble des déterminants génétiques définissant les capacités et les limites intrinsèques de chaque espèce donnée. On pensait à l'origine qu'il s'insérait au hasard dans les cellules, sans organisation stéréotypée ou reproductible. Cette hypothèse a été initialement dissipée lorsque des études en microscopie optique de cellules d'*Escherichia coli* colorées avec des colorants spécifiques de l'ADN ont révélé un corps discret d'ADN, situé dans une région irrégulière à l'intérieur des cellules procaryotes dans laquelle se trouve le matériel génétique. Cette région a été nommée nucléoïde (Robinow, 1994). À la différence du noyau des cellules eucaryotes, le nucléoïde des cellules procaryotes n'est pas délimité par une membrane nucléaire.

Au cours du cycle cellulaire, le génome doit être dupliqué par réplication puis correctement ségrégué dans les futures cellules filles afin d'assurer le maintien de l'information génétique au cours des générations. Chez les bactéries, le génome est constitué de deux grands types de molécules d'ADN : **les chromosomes** et les éléments extra-chromosomiques appelés **plasmides**. Ces deux types de molécules, capables de se répliquer de manière autonome dans la cellule, ont été appelées par Jacob et Brenner, réplicons (Jacob, 1963).

Le génome bactérien comprend le plus souvent un unique chromosome, circulaire ou linéaire, complété selon les organismes par un ou plusieurs plasmides (sous forme circulaire et/ou linéaire). Bien que la plupart des chromosomes bactériens soient circulaires, certains sont linéaires, notamment les chromosomes multiples de 1 mégabase (Mb) des espèces *Borrelia* et le chromosome de 8Mb des espèces *Streptomyces* (Badrinarayanan, 2015). Les chromosomes circulaires bactériens se sont sporadiquement linéarisés au cours de l'évolution des procaryotes. Les chromosomes linéaires peuvent avoir été formés par l'intégration de plasmides linéaires (**fig. 45** ; Volff, 2000). Chez *Streptomyces*, l'origine de réplication est localisée dans la région centrale du chromosome, définissant deux bras chromosomiques où des protéines se lient de façon covalente aux extrémités chromosomiques génèrent une topologie fermée du chromosome (Tsai, 2012 ; Yang, 2001). Chez *Borrelia* les extrémités chromosomiques sont organisées en tiges-boucles (Tourand, 2003). Chez *Streptomyces* et *Borrelia*, des mutants circulaires ont pu être générés en laboratoire (Chen et al., 2010 ; Ferdows et al., 1996). Les chromosomes de *Streptomyces* artificiellement circularisés présentent une instabilité génétique égale ou supérieure à la forme linéaire (Chen, 2010 ; Lin,

1997). Une des explications possibles à l'instabilité de la topologie circulaire serait le manque de domaines associés à la terminaison de la réplication, nécessaires à l'intégrité du chromosome



**Figure 45 : Distribution des chromosomes circulaires et linéaires chez les procaryotes.**  
 Les cercles symbolisent les chromosomes circulaires, les barres les chromosomes linéaires.  
 (Figure tirée de Volff, 2000).

Les chromosomes sont généralement des molécules circulaires de grande taille, de l'ordre de la Mégabase (Mb) et porteurs de gènes essentiels au développement de la bactérie (en particulier les gènes ribosomiques). La taille des chromosomes est très variable : allant de 143.795bp pour le plus petit chromosome chez la bactérie endosymbiotique, *Hodgkinia cicadicola*, à 13.033.779bp pour *Sorangium cellulosum* (<https://bionumbers.hms.harvard.edu>).

Les plus petits génomes correspondent à des organismes endosymbiotiques (Moran, 2014), vivant dans un environnement bien défini où les fonctions cellulaires obligatoires sont assurées par l'organisme hôte. À l'inverse, les plus grands génomes bactériens correspondent à des espèces vivant de manière autonome dans des milieux complexes et changeants. En effet dans des conditions stables, des modifications drastiques du génome sont rares (Barrick, 2009). Des bactéries vivant dans de tels milieux doivent être "armées" pour résister à un très grand nombre de stress différents. Cela se traduit par un plus grand nombre de gènes et donc des génomes plus grands (corrélation directe entre la taille des génomes et le nombre de gènes ; Bird, 1995) distinguant deux catégories fonctionnelles : les gènes essentiels,

obligatoires au développement végétatif de la cellule quelles que soient les conditions et les gènes accessoires permettant à la cellule de s'adapter à des conditions données.

La majorité des bactéries ne portent qu'un seul chromosome mais de nombreux génomes bactériens sont constitués de plus d'un chromosome, comme *Vibrio cholerae* ou *Burkholderia cenocepacia* (Egan, 2003 ; Dubarry, 2006). Le plus grand chromosome, appelé chromosome principal, contient la majorité des gènes essentiels (gènes de ménage et gènes codant pour les ARNs ribosomiques) et une origine de réplication similaire à celle des bactéries mono-chromosomiques. Les autres chromosomes, dits chromosomes secondaires, ne portent qu'une partie des gènes essentiels et présentent des caractéristiques hybrides entre chromosomes principaux et plasmides. Leur région origine est relativement variable à l'image de celle observée chez les plasmides et peut être identifiée par leur structure (Egan, 2003), ou par leur homologie de séquence avec les systèmes plasmidiques connus (Dubarry, 2006). De même, les gènes présents sur les chromosomes principaux sont très conservés alors que les gènes présents sur les chromosomes secondaires ont une forte variabilité évolutive comme les plasmides (Cooper, 2010). La frontière entre ces chromosomes secondaires et plasmides est parfois assez floue.

Un plasmide désigne un élément génétique extrachromosomique, physiquement distinct de l'ADN chromosomique qui est capable de réplication autonome. Ce terme plasmide fut introduit par le biologiste moléculaire américain J. Lederberg (Lederberg, 1952). Les plasmides sont des molécules de petites tailles et présents en multi-copie (inférieur à 10kb et plus de 15 copies par chromosome) ou de grandes tailles (supérieur à 30 kb) et en faible nombre de copies. La taille pourrait paraître un critère pour distinguer un chromosome d'un plasmide mais il existe des chromosomes de petites tailles, comme le chromosome 2 de *Butyrivibrio proteoclasticus* de 300kb, et des mégaplasmides de plusieurs Mb comme celui de la bactérie *Ralstonia solanacaerum* GM1100, de 2,096Mb (Salanoubat, 2002). La différence principale entre chromosome et plasmide est la présence de gènes essentiels sur les chromosomes. Bien que les plasmides soient pour la plupart non-essentiels à la survie de la cellule, ils sont souvent porteurs de déterminants génétiques conférant un avantage sélectif à la bactérie hôte (porteurs de gènes de résistance aux antibiotiques, de facteurs de virulence ...). Il existe une corrélation entre la taille des plasmides et la pression écologique que subissent les bactéries qui en sont porteuses (Slater, 2008).

Une autre différence importante entre chromosome et plasmide se situe au niveau de la réplication. L'initiation de la réplication des plasmides peut avoir lieu à tout moment du cycle alors que l'initiation de la réplication des chromosomes secondaires est synchronisée avec le cycle cellulaire. Chez *Vibrio cholerae*, la réplication du chromosome 2 est initiée après celle du chromosome principal de manière à s'achever à peu près en même temps (Rasmussen, 2007 ; Srivastava, 2007).

Néanmoins, les chromosomes secondaires et plasmides peuvent avoir certaines caractéristiques proches comme leur forte variabilité évolutive. L'analyse des vitesses évolutives des gènes présents sur les différents chromosomes des bactéries multi-chromosomiques a permis de constater que les gènes présents sur les chromosomes secondaires étaient en majorité des gènes à évolution rapide, alors que les gènes présents sur les chromosomes principaux sont fortement conservés. Les chromosomes secondaires serviraient donc de réservoir de diversité ou de banc de test de l'évolution (Cooper, 2010). Les caractéristiques plasmidiques des chromosomes secondaires laisseraient penser que ces réplicons seraient d'origine plasmidique. Le transfert de groupe de gènes du chromosome principal vers un plasmide récepteur conduirait ainsi à la formation d'un chromosome secondaire, qui acquerrait ensuite les fonctions et régulations nécessaires à assurer son maintien pérenne dans la descendance. La conservation parmi les espèces proches des groupes de gènes ainsi transférés permet de reconstituer l'histoire de ces remaniements au fil des spéciations (Slater et al., 2009).

### **3-1-2 Organisation du chromosome bactérien.**

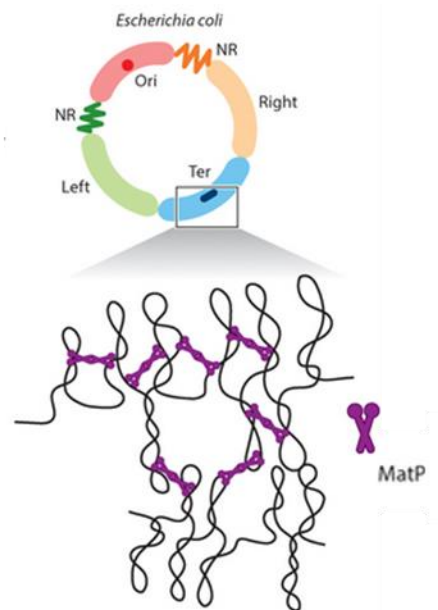
#### **● Organisation en macro-domaines.**

S'il était entièrement étiré, un chromosome bactérien typique mesurerait près d'un millimètre de long, soit environ 1 000 fois la longueur d'une cellule. Les chromosomes doivent être compactés de près de trois ordres de grandeur pour tenir dans les cellules mais elles doivent également organiser leur ADN de manière à ce qu'il soit compatible avec toute une série de processus cellulaires basés sur l'ADN, notamment la réplication, la transcription, la réparation, la recombinaison homologue, le transfert de gènes ... Ce défi est particulièrement aigu chez les bactéries, car la ségrégation des chromosomes se produit en même temps que la réplication de l'ADN, au lieu d'être séparée temporairement comme chez les eucaryotes

(Badrinarayanan, 2015). Cette compaction doit impliquer une organisation structurale des chromosomes.

Des études chez *Salmonella typhimurium* (Garcia-Russell, 2004) puis chez *Escherichia coli* (Valens, 2004) font penser que le mouvement de l'ADN dans le cytoplasme de la cellule est contraint. Des études d'hybridation *in situ* en fluorescence (ou FISH) chez *Escherichia coli* ont démontré que certains loci co-occupent fréquemment le même espace cytoplasmique restreint (Niki, 2000). Ces résultats suggèrent que les chromosomes bactériens sont organisés en domaines de la taille d'un Mb appelés macrodomaines.

Par des techniques de FISH associées à l'utilisation du système fluorescent opérateur/répresseur (FROS) et du système ParB/*parS*, l'agencement spatial des chromosomes a été déduit en suivant les positions subcellulaires des loci individuels (Le, 2014). Chez *Caulobacter crescentus*, 112 loci ont été examinés par FROS dans des cellules contenant un seul chromosome. Les positions spatiales des loci dans la cellule récapitulent la carte génétique, avec l'origine de la réplication (*oriC*) à un pôle de la cellule, l'extrémité de réplication (*Ter*) au pôle opposé, et les bras chromosomiques gauche et droit étant probablement parallèles le long de l'axe de la cellule, dans un modèle appelé configuration *ori-ter* (Viollier, 2004). Une analyse de recombinaisons spécifiques de site entre des sites de recombinaison de l'intégrase  $\lambda$  a montré que l'interaction entre deux sites n'est possible qu'au sein de régions définies du chromosome : des loci au sein d'un même macrodomaine interagissent et recombinent entre eux plus fréquemment que des loci localisés dans différents macrodomaines (Valens, 2004). Ces résultats ont permis de définir quatre macrodomaines chez *E. coli* d'une taille entre 800kb et 1Mb (**fig. 46**) : l'ori autour de l'origine de réplication, le domaine *Ter* autour du terminus de réplication et les macrodomaines Left et Right de chaque côté du *Ter*. Deux autres régions flexibles, ou non-structurées, NR, (qui interagissent avec le reste du chromosome) ont été définies également de chaque côté de l'ori (Valens, 2004). Des analyses par microscopies fluorescentes ont permis de montrer que différents loci situés dans un même macrodomaine sont toujours colocalisés et qu'ils ont un mouvement restreint dans la cellule par rapport à des loci situés dans les régions non-structurées (Espeli, 2008).



**Figure 46 : Organisation en macrodomaine du chromosome d'*Escherichia coli*.**

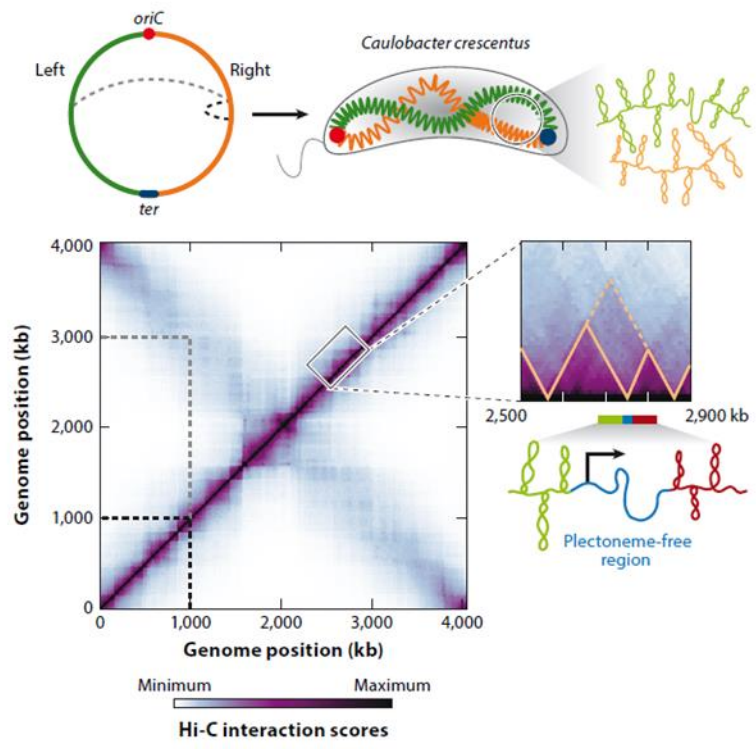
Le chromosome d'*E. coli* est organisé en quatre macrodomaines, Ori, Ter, Left and Right et en deux régions non-structurées (NR). La protéine MatP (en violet) se fixe sur des séquences *matS* spécifiques situées exclusivement dans le domaine Ter et le structure.

(Figure tirée de Badrinarayanan, 2015).

Une avancée dans la compréhension de l'organisation des macrodomaines est venue de la découverte de la protéine MatP (pour Macrodomaine ter Protein) et de ces sites de fixation *matS* (pour macrodomaine ter Sequence). Par une analyse bio-informatique, Mercier et ses collaborateurs ont identifié la séquence *matS*, un palindrome de 13bp, présente à 23 reprises exclusivement dans les 800kb du macrodomaine Ter (Mercier, 2008). La protéine MatP se fixe sous forme de dimère aux sites *matS* et peut former un tétramère avec un dimère MatP lié à un site voisin rapprochant les complexes MatP-*matS* distaux et contribuant à compacter le macrodomaine Ter (Dupaigne, 2012 ; **fig. 46**). Dans les mutants ne possédant pas cette protéine MatP, le macrodomaine Ter est décondensé, interagit avec d'autres macrodomaines, et a une mobilité comparable à celle d'un domaine non-structuré, ce qui indique une désorganisation du domaine Ter. L'inactivation de la protéine MatP entraîne des défauts sévères dans la ségrégation des chromosomes et dans la division cellulaire (Mercier, 2008). Des protéines similaires, mais pas encore identifiées, pourraient structurer les autres macrodomaines d'*E. coli*. Les protéines homologues à MatP semblent être restreintes aux entérobactéries (Badrinarayanan, 2015).

- **Organisation en Domaines d'Interaction Chromosomiques (CIDs).**

Chez *C. crescentus*, les macrodomaines en tant que tels n'ont pas été documentés. Cependant des analyses de capture de conformation du chromosome à haute résolution (Hi-C) ont montré que le chromosome de *C. crescentus* est divisé en 23 régions à forte interaction entre elles ou domaines d'interaction chromosomique (CIDs) de 166kb chacun. Les CIDs sont similaires aux domaines associés topologiquement (TADs) présent chez les eucaryotes. Les loci présents dans un domaine interagissent préférentiellement avec d'autres loci du même domaine (Le, 2013). L'expression active des gènes est nécessaire à la formation des frontières de CID. En effet, les frontières des domaines coïncident souvent avec les gènes les plus fortement exprimés tels que ceux codant pour les protéines ribosomales. Un traitement des cellules à la rifampicine, qui inhibe l'élongation de la transcription, entraîne une perte presque



**Figure 47 : Interactions chromosomiques de *C. crescentus* obtenues par Hi-C.**

Le test de capture de la conformation du chromosome (5C et Hi-C) et la modélisation computationnelle ont révélé l'organisation du chromosome de *Caulobacter crescentus*. La carte Hi-C (Le, 2013) indique la fréquence des interactions ADN-ADN à travers le génome en utilisant l'échelle de couleurs indiquée.

La diagonale la plus proéminente indique les interactions fréquentes au sein d'un bras chromosomique (lignes pointillées noires), et l'autre diagonale, moins proéminente, montre les interactions entre les bras chromosomiques (lignes pointillées grises).

Les triangles orange dans l'encadré (à droite) indiquent les domaines d'interaction chromosomique (CIDs). On pense que les gènes fortement transcrits créent une région moins compacte, sans plectonème (bleu), qui sert à isoler spatialement l'ADN (vert et rouge) dans les domaines adjacents, créant ainsi une frontière de CID.

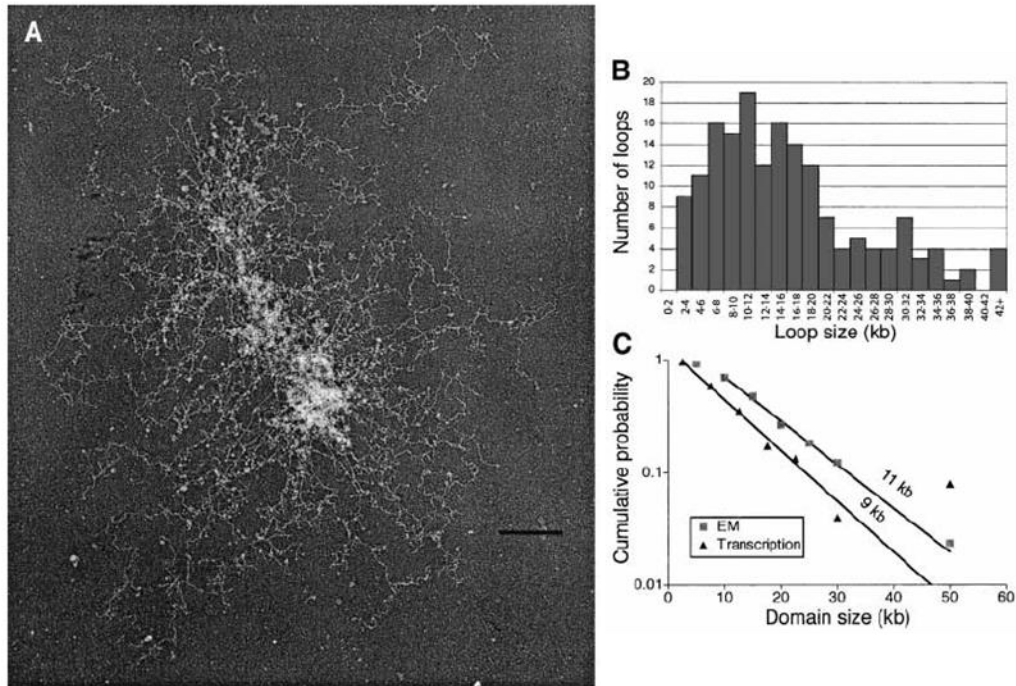
(Légende et figure tirées de Badrinarayanan, 2015).



complète des domaines. De plus, l'insertion d'un gène fortement exprimé à l'intérieur d'un domaine est suffisante pour induire de nouvelles frontières (Le, 2013). Les taux élevés de transcription et le déroulement fréquent de l'ADN créent probablement des régions dans le chromosome exemptes de plectonèmes (accumulation de boucles localisées en un point du segment d'ADN surenroulé conduisant à une structure torsadée) qui se forment lorsque l'ADN super-enroulé s'enroule sur lui-même. Ces régions relâchées sans plectonèmes peuvent former des barrières topologiques bloquant la diffusion de supertours et séparer physiquement des domaines adjacents et diminuant ainsi la probabilité de contact entre les loci des domaines voisins (**fig. 47**). Les CIDs semblent s'établir pendant ou peu après la réplication de l'ADN, et pourraient potentiellement faciliter la ségrégation chromosomique en empêchant les chromosomes nouvellement répliqués de s'enchevêtrer (Le, 2013).

- **Domaines super-enroulés ou boucles de plectonème.**

Des boucles de super-enroulement appelées plectonèmes ou domaines super-enroulés, ont été caractérisées par microscopie électronique chez *Escherichia coli* (Kavenoff, 1976). Ces plectonèmes, enroulés sur eux-mêmes, sont attachés à leur base par des protéines qui aident à isoler topologiquement l'ADN en boucle. Postow et ces collaborateurs ont analysé ces domaines super-enroulés en regardant la réponse transcriptionnelle après une coupure double-brin : seul l'ADN dans un plectonème, topologiquement isolé, serait relâché après une cassure double brin conduisant des changements dans la transcription des gènes sensibles au superenroulement uniquement dans ce domaine. Ils ont estimé à environ 400 plectonèmes dans le chromosome de *E. coli* avec un domaine super-enroulé moyen de 10kb (Postow, 2004). Ils ont confirmé ce chiffre avec celui du nombre de boucles dans les chromosomes de cellules lysées imagés par EM (**fig. 48**). Les limites entre les domaines super-enroulés sont dynamiques et peuvent dépendre à la fois des protéines de liaison à l'ADN et de l'expression des gènes. Les protéines de liaison peuvent relier des loci distants, isolant topologiquement l'ADN formant une boucle et empêchant la propagation de domaines super-enroulés entre domaines adjacents. Comme pour les CIDs, les loci ayant des hauts taux de transcription peuvent être des éléments de frontières qui empêchent la diffusion des plectonèmes bien que le mécanisme sous-jacent ne soit pas très clair (Badrinarayanan, 2015).



**Figure 48 : Mesures des boucles super-enroulées à partir d'images de microscopie électronique à partir de chromosomes isolés d'*Escherichia coli*.**

**A)** Les chromosomes d' *E. coli* ont été délicatement isolés par sédimentation en gradient de saccharose et étalés sur des grilles de microscopie électronique recouvertes de carbone.

**B)** Quantification de la taille des boucles à partir de micrographies électroniques: 169 boucles super-enroulées clairement définies ont été tracées et leurs longueurs ont été comparées à celles d'un plasmide de 7 kb. Les tailles des boucles ont été classées par incréments de 2 kb et représentées sous forme d'histogramme.

**C)** Comparaison de la taille des domaines à partir des expériences de microarray et de microscopie électronique (ME): la probabilité cumulative de la taille des boucles mesurées par la microscopie électronique (carrés gris) et de la taille des domaines provenant des expériences transcriptionnelles de microarray (triangles noirs) est tracée. Les fonctions exponentielles les mieux ajustées indiquent une taille de domaine de 9 kb pour les données de microarray et de 11 kb pour les données de microscopie. Le décalage vertical entre les deux ensembles de données est dû à la rareté des mesures de petites boucles dans l'ensemble de données de microscopie.

(légende et figure tirées de Postow, 2004).

En résumé, les relations entre les différents niveaux de compaction, du chromosome bactérien, macro-domaines/CIDs/plectonèmes n'est pas clairement défini mais une organisation hiérarchique entre les différents niveaux peut être envisagée.

Les macrodomaines de la taille d'une mégabase sont probablement composés de multiples CIDs, chacun d'environ 100-200 kb et contenant de multiples domaines super-enroulés diffusibles, les plectonèmes, d'environ 10 kb de taille.

Les gènes très fortement exprimés semblent jouer un rôle essentiel dans l'établissement des limites des CIDs, qui sont relativement fixes dans une population de cellules.

L'expression d'autres gènes peut former des domaines et des frontières transitoires. La position des gènes à l'intérieur des domaines (à tous les niveaux) peut influencer leur expression, mais la relation précise entre la structure des chromosomes et l'expression des gènes n'est pas connue (Badrinarayanan, 2015).

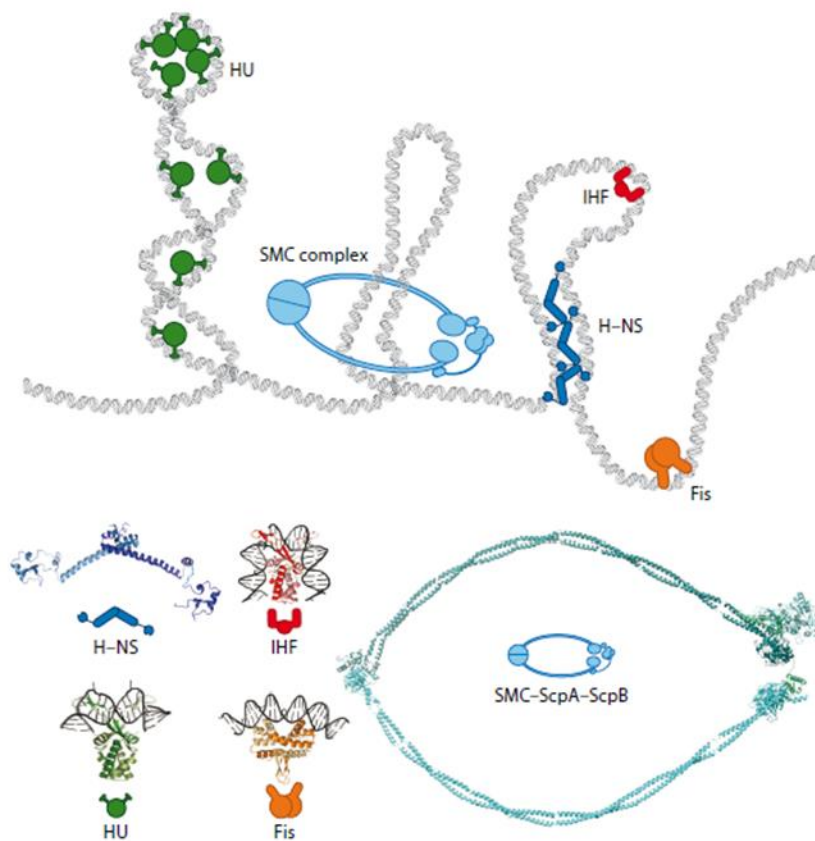
### 3-1-3 Les protéines associées au nucléoïde.

#### ● **Nucleoid Associated Proteins = NAP.**

Les super-enroulements ne peuvent pas expliquer seuls la compaction du chromosome. Une part de cette compaction est assurée par une classe hétérogène de protéines, se liant à l'ADN, appelées NAP pour Nucleoid Associated Proteins (**fig. 49**). Elles sont caractérisées par un petit poids moléculaire et une grande abondance. L'organisation du chromosome bactérien est profondément influencée par ces protéines se liant à l'ADN. En général, les NAPs se lient de manière relativement non spécifique à travers les génomes bactériens et facilitent la compaction en formant, notamment, des ponts entre les loci ou en courbant l'ADN (Dame, 2005 ; Dillon, 2010). Ces protéines ont été identifiées et caractérisées essentiellement chez *E. coli*. Si leur structure est connue, leur rôle est plus difficile à déterminer à cause des redondances fonctionnelles.

- **H-NS : "Histone-like Nucleoid-Structuring protein"** : c'est une petite protéine de 15kDa. Sa partie N-terminale lui permet de former des dimères tandis qu'elle se fixe sur l'ADN par son extrémité C-terminale sur des centaines de sites, de préférence sur des séquences AT-riche ou un ADN incurvé et souvent au niveau de promoteurs sensibles au superenroulement. Elle peut former des ponts entre des régions d'ADN et les rapprocher physiquement et condenser l'ADN surenroulé. La protéine H-NS peut s'oligomériser et s'entendre sur l'ADN incluant des sites de fixation pour l'ARN polymérase ou des activateurs de transcription permettant ainsi à la H-NS de réguler l'expression des gènes (Higgins, 1988 ; Grainger, 2006 ; Kahramanoglou, 2011 ; Singh, 2014).

- **H-U : "Heat-Unstable protein"** : Cette petite protéine de 18kDa est très abondante (près de 30.000 copies par cellule) et est retrouvée dans de nombreuses bactéries. Elle recouvre et enveloppe pratiquement 10% du chromosome (Pietro, 2012) et enroule l'ADN d'une manière similaire aux histones (Azam, 1999). Elle est constituée de deux sous-unités, alpha et bêta, et on peut la trouver sous forme homo ou dimère, en fonction de la phase de croissance d'*E. coli* (Claret, 1997). Des études structurelles suggèrent également que la



**Figure 49: Les protéines associées aux nucléoïdes (NAP) ayant des activités de pontage, d'enveloppement ou de courbure de l'ADN contribuent à l'organisation du chromosome.**

Les fonctions des NAPs sont schématisées en haut, avec les structures cristallines correspondantes en bas.

Les dimères de H-NS (**bleu**) forment un pont avec l'ADN.

La HU abondante (**vert**) introduit une flexion  $\sim 90^\circ$  à l'ADN et peut enrouler l'ADN autour de elle-même, favorisant ainsi les interactions à courte portée de l'ADN.

Lorsque l'IHF (**rouge**) se lie à l'ADN, il induit la formation d'un spectaculaire virage en U de l'ADN, modifiant radicalement la trajectoire du squelette de l'ADN.

Le FIS (**orange**) est une autre NAP ayant une activité de courbure de l'ADN.

Les complexes SMC (**cyan**) forment vraisemblablement une structure en anneau qui peut rassembler et menotter des loci distaux dans la séquence primaire.

(légende et figure tirées de Badrinarayanan, 2015).

protéine H-U peut former une structure octamérique avec l'ADN enroulé autour (Guo, 2007). Des analyses de ChiP-seq indiquent qu'elle n'a pas de spécificité de liaison (Pietro, 2012). En absence de cette protéine, les souches produisent souvent des cellules anucléées suggérant un défaut général de compaction du chromosome ou de ségrégation (Huisman, 1989). Des études de capture de conformation chromosomique (Hi-C) chez des mutants de *Caulobacter* révèlent une diminution du surenroulement avec une diminution significative des interactions à courte distance confortant l'idée que la protéine H-U aide à compacter l'ADN en stabilisant les plectonèmes (Le, 2013). Dans des souches avec une protéine H-U avec une plus grande

affinité pour l'ADN, le niveau global de super-enroulement de l'ADN est augmenté (Kar, 2005). La protéine H-U aurait un effet sur le super-enroulement négatif et un effet global sur la compaction du chromosome.

- **IHF : “integration Host Factor”** : la protéine IHF présente des similitudes de séquence avec la protéine H-U. Elle est composée de deux sous-unités. Elle se fixe à l'ADN de façon plus spécifique, souvent proche de promoteur de gènes, et plie l'ADN d'un angle d'environ de 160° (Rice, 1996). Elle peut donc altérer la forme de l'ADN et faciliter la formation de boucle. Elle est reconnue comme un facteur de régulation de la transcription de nombreux gènes (Swinger, 2004). La protéine IHF a également un impact sur une série d'autres processus basés sur l'ADN, comme l'initiation de la réplication et la recombinaison (Leonard, 2005 ; Mumm, 2006).

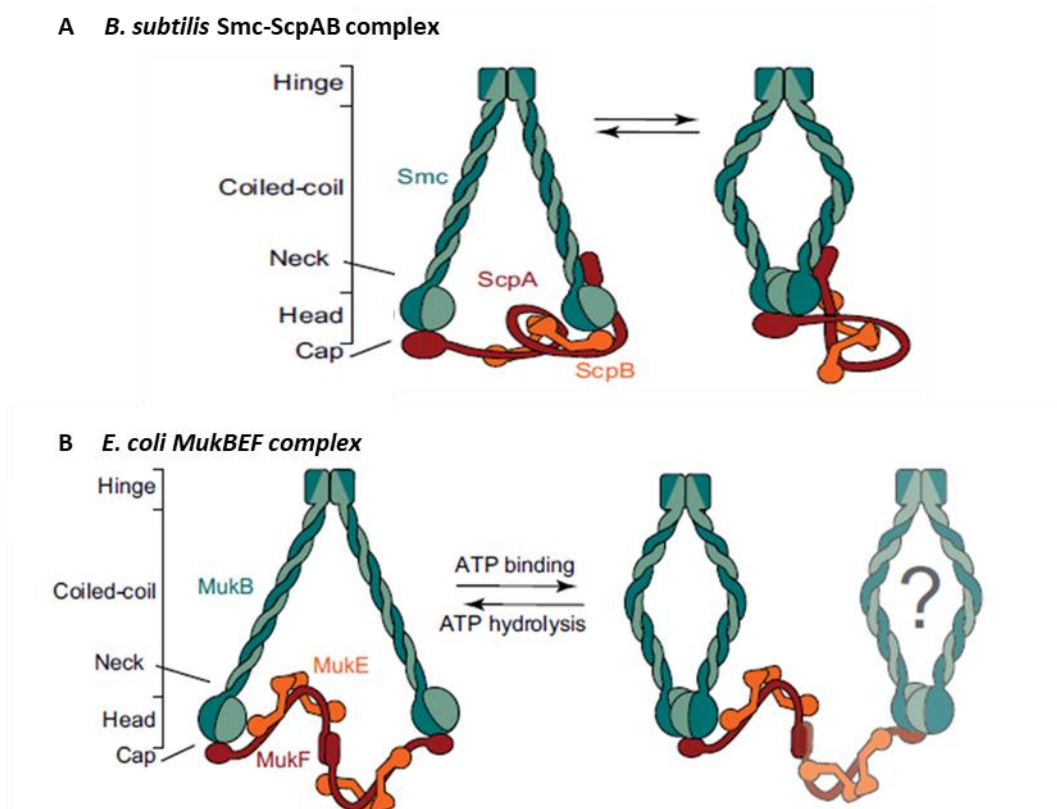
- **FIS : “Factor for Inversion Stimulation”** : c'est une des protéines les plus exprimées chez *E. coli* pendant la phase de croissance (Azam, 1999). Elle est constituée de sous-unités. Sous forme d'homo-dimère, elle fixe à l'ADN au niveau de séquence AT riche au niveau du petit sillon et plie l'ADN de 50 à 90° en formant un complexe nucléoprotéine très stable (Stella, 2010). Elle se fixe à travers le génome, surtout au niveau des promoteurs, impactant transcription, réplication et recombinaison. Elle influe sur la compaction de l'ADN et son organisation. Elle peut maintenir les boucles de plectonèmes superenroulés et influencer sur le niveau global de superenroulement, indirectement, en modulant l'expression de la gyrase (Schneider, 1999).

- **Les complexes SMC, “Structural Maintenance of Chromosome”.**

Le complexe protéique SMC, qui est l'homologue de la condensine eucaryote, est très largement conservé et retrouvé dans de très nombreuses bactéries. Les protéines SMC sont constituées de trois parties distinctes, un domaine ATPase, “Head” formé par les extrémités N et C terminales, une partie centrale, “Coiled-coil” ou super-hélice constituée par deux longs polypeptides antiparallèles enroulés en spirale et un domaine charnière “Hinge” de dimérisation (**fig. 50**). L'homodimérisation via le domaine “hinge” crée une structure en anneau qui peut entourer l'ADN (Nolivos, 2014).

Trois grandes familles de condensines bactériennes ont été identifiées :

- le complexe SMC/ScpA/ScpB : il a été particulièrement étudié chez *B. subtilis* où les protéines SMC s'associent avec les sous-unités ScpA et ScpB qui modulent son activité ATPase et ainsi affectant l'ouverture ou la fermeture de l'anneau homodimérique (**fig. 50 A**).
- le complexe MukB/MukE/MukF chez *E. coli* et d'autres protéobactéries (**fig. 50 B**).
- le complexe MksB/MksE/MksF : ce complexe a une structure secondaire similaire à celui de la protéine MksB mais possède une région, "Coiled-coil" plus courte. Cette troisième famille a été identifiée plus récemment (Petrushenko, 2011).



**Figure 50: Architecture des complexes SMC.**

Les protéines SMC sont composées de trois parties: la partie "Hinge" ou domaine charnière qui permet l'homo-dimérisation, la partie centrale "Coiled-coil" constituée par deux longs peptides anti-parallèles enroulés et la partie "Head", le domaine ATPase, formée par les extrémités N et C terminales de la protéine SMC.

**A)** Chez *B. subtilis*, la protéine SMC s'associe à deux sous-unités, ScpA et ScpB qui modulent l'activité du domaine ATPase permettant l'ouverture ou la fermeture de l'anneau.

**B)** La bactérie *E. coli* possède le complexe MukB/MukE/MukF. En absence d'ATP, deux sous-unités MukF et quatre sous-unités MukE sont liées à MukB. En présence d'ATP, une sous-unité MukF est déplacée, fermant l'anneau. Un dimère de dimère peut être potentiellement formé via la dimérisation de MukF.

(légende et figure tirées de Nolivos, 2014).

La remarquable conservation de l'architecture des différents complexes SMC laisse penser que tous ces complexes partagent un mécanisme biochimique commun pendant leur association avec, et leur action sur, l'ADN chromosomique (Nolivos, 2014).

Des mutations dans le complexe SMC produisent toute une série de défauts chromosomiques, notamment une augmentation de cellules anucléées. Chez *B. subtilis* et *E. coli*, des mutations dans les protéines SMC et MukB entraînent une décondensation visible avec une coloration au DAPI (Tadesse, 2005). Les complexes SMC contribuent vraisemblablement à la ségrégation et à la compaction des chromosomes.

La manière précise dont les protéines SMC affectent la compaction des chromosomes n'est pas encore claire. Par sa structure en forme d'anneau, le complexe SMC pourrait établir un pont entre différents loci du chromosome qui aiderait à compacter l'ADN et pourrait également contraindre les super-enroulements en produisant des boucles d'ADN isolées topologiquement (Badrinarayanan, 2015).

En résumé, les NAP et autres protéines associées aux chromosomes sont clairement des acteurs centraux de l'organisation des chromosomes. Bien que les propriétés biophysiques d'un grand nombre de ces protéines aient été bien étudiées, il reste encore beaucoup à apprendre sur leurs fonctions *in vivo* et sur la manière dont, au niveau global, elles se combinent pour compacter, façonner et organiser le génome et, à leur tour, influencer les fonctions de l'ADN dans les cellules (Badrinarayanan, 2015).

#### **3-1-4 Le super-enroulement négatif.**

Le premier niveau d'organisation du chromosome est le super-enroulement négatif ou surenroulement négatif qui entraîne la formation de domaines topologiques isolés (un super-enroulement négatif correspond à un tour dans le sens contraire de celui de la double hélice d'ADN). Ce super-enroulement semble être le principal mécanisme de compaction du génome bactérien. Alors qu'une seule coupure suffit au relâchement d'un plasmide, un grand nombre de coupures par la DNaseI est nécessaire pour relâcher intégralement le chromosome (Worcel, 1972). En faisant des boucles sur eux-mêmes, les mécanismes de super-enroulements condensent le chromosome sur lui-même mais aussi le séparent d'ADN adjacent (chromatides sœurs, plasmides ...). En maintenant des régions adjacentes proches dans l'espace, les super-enroulements faciliteraient la réparation des cassures double brin en maintenant les

extrémités cassées proches dans l'espace. Différentes études indiqueraient que ces domaines topologiques seraient d'une longueur de 10kb (Postow, 2004). Ils protègent le chromosome d'une entière relaxation qui a été prouvée comme étant létale (Gellert, 1976). Ces super-enroulements négatifs sont régulés par un certain nombre de facteurs dont les enzymes de la famille des topoisomérases qui contrôlent la super-hélicité. Ainsi l'ADN gyrase qui est une enzyme de la famille des ADN topoisomérases de classe II introduit des super-enroulements négatifs, alors que la topoisomérase IV les relâche. Leurs actions antagonistes permettent de réguler les super-enroulements dans le chromosome.

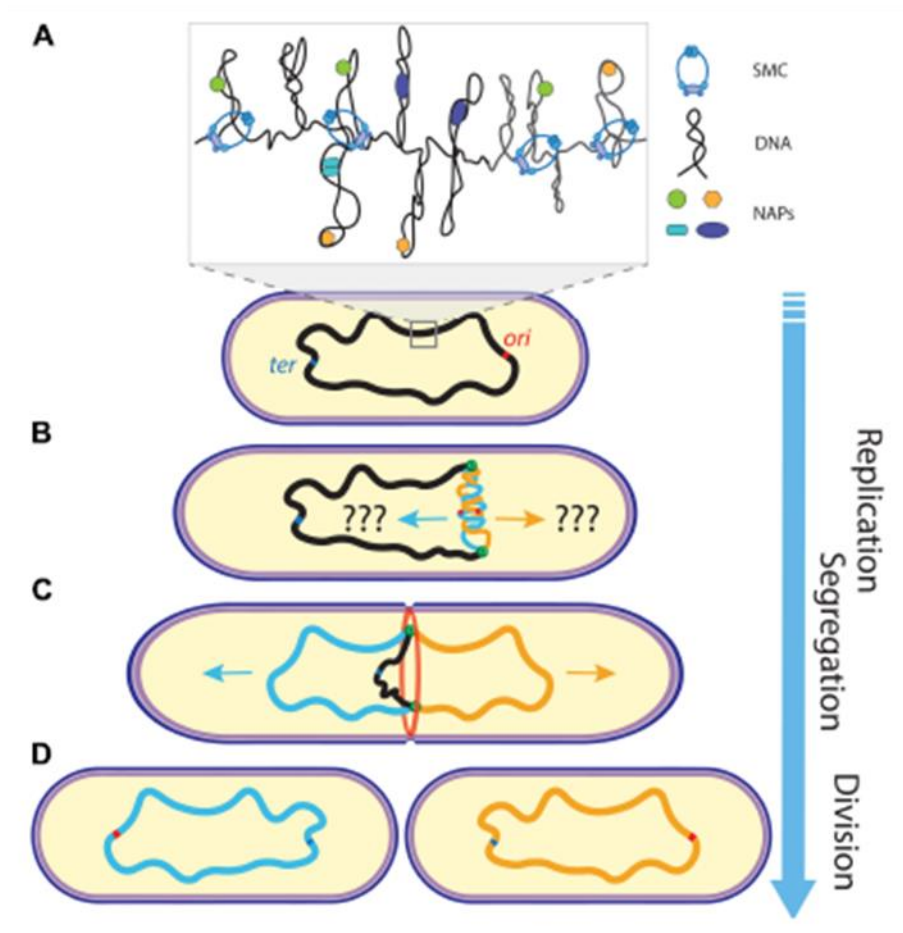
### **3-2 Ségrégation des génomes bactériens.**

Le processus de ségrégation de l'ADN est une étape cruciale du cycle de vie de tous les systèmes vivants. Une ségrégation chromosomique fidèle est essentielle pour garantir que chaque cellule fille hérite d'une copie complète de l'information génétique du parent. La ségrégation des chromosomes n'est pas un processus trivial, surtout chez les bactéries, car l'ADN doit être maintenu dans un état compacté pour tenir dans le volume limité des cellules (Jalal, 2020a). En effet, les génomes des bactéries, d'une taille de 1 à 10 Mb doivent être fortement condensés pour pouvoir tenir dans le volume d'une cellule bactérienne. Comme nous l'avons vu précédemment, les bactéries réalisent cette forte condensation grâce au super-enroulement de l'ADN, aux protéines associées aux nucléoïdes (NAP) qui se lient à l'ADN et à d'autres protéines qui compactent l'ADN comme les complexes SMC (Structural Maintenance of the Chromosome). Contrairement aux cellules eucaryotes, qui ont des phases temporellement distinctes de réplication, condensation des chromatides sœurs et ségrégation, les bactéries ont besoin d'organiser, de compacter et de séparer leurs chromosomes au fur et à mesure que les chromosomes frères sont générés. Ainsi la réplication de l'ADN et la ségrégation des chromosomes se produisent de manière concomitante chez les bactéries, plutôt que d'être séparées dans le temps, comme chez les eucaryotes.

La réplication du chromosome procaryote circulaire commence à un locus d'origine de réplication (ori) dédié et se termine près du terminus de réplication (ter) sur le côté opposé du chromosome alors que la ségrégation se produit en même temps que le processus de réplication (**fig. 51** ; Gogou, 2021). Tout au long des 20 à 200 minutes que dure un cycle bactérien classique, le positionnement séquentiel des régions chromosomiques est



étroitement régulé. De toute évidence, la ségrégation nécessite une coordination temporelle avec la division cellulaire. La constriction du divisome doit être ajournée jusqu'à que la ségrégation soit effectuée, autrement il en résulterait un nucléoïde « guillotiné », des cellules anucléées ou une division cellulaire totalement inhibée (Gogou, 2021).



**Figure 51 : Schéma de la réplication et de la ségrégation des chromosomes chez les bactéries.**

**A)** Un chromosome père (**noir**) avec un site d'origine (**ori**) et de terminaison (**ter**) de réplication: l'encart montre l'organisation compacte du chromosome en domaines super-enroulés organisés par les NAPs et les SMCs.

**B)** Lors de l'initiation de la réplication au niveau de l'ori, les chromosomes frères (**bleu** et **orange**) sont synthétisés par les réplisomes (points **verts**).

**C)** Alors que les réplisomes se déplacent de manière bidirectionnelle le long des bras chromosomiques pour continuer à répliquer l'ADN, les chromosomes frères se séparent en deux moitiés de cellules opposées et la cellule commence à se diviser (à travers l'anneau de divisome indiqué en **rouge**).

**D)** Finalement, deux cellules filles sont formées après la division cellulaire, chacune contenant une copie du chromosome.

La flèche **bleue** qui va de haut en bas indique le temps.

(légende et figure tirées de Gogou, 2021).

### 3-2-1 Ségrégation par diffusion passive.

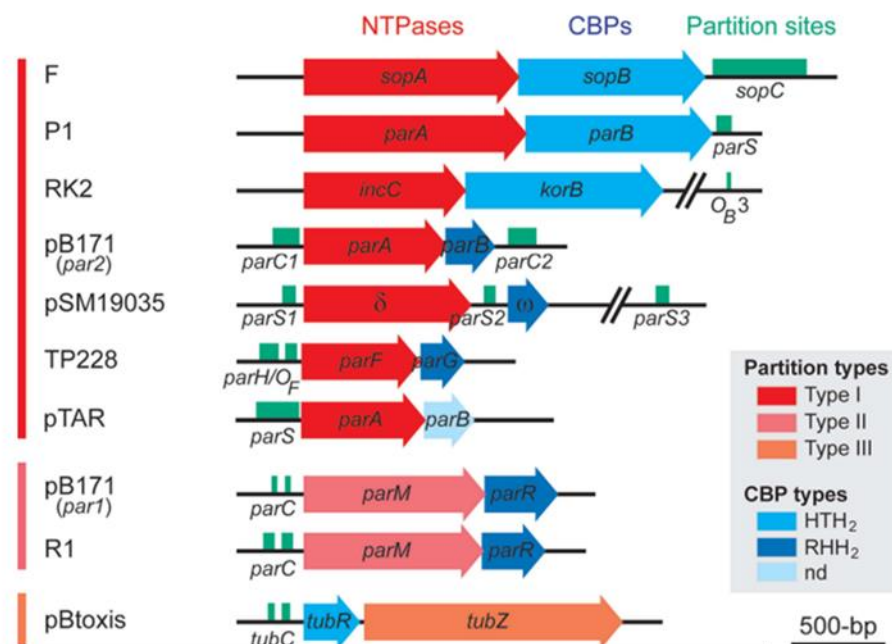
Les processus qui gouvernent la ségrégation des cellules eucaryotes sont basés sur l'utilisation bien caractérisée de fuseaux mitotiques via un mécanisme basé sur la tubuline alors que ceux qui régissent la ségrégation des plasmides et des chromosomes dans les cellules bactériennes sont moins bien compris (Brooks, 2017). Parmi les réplicons bactériens, il faut distinguer entre une diffusion passive pour les plasmides à haut nombre de copies, comme le plasmide ColE1, présents entre 20 et 30 copies par chromosome et une ségrégation active pour la plupart des chromosomes et des plasmides à bas nombres de copies, comme le plasmide F ou P1 d'*E. coli*, présent entre 1 à 5 copies par chromosome selon les conditions de croissance.

La diffusion passive des plasmides dans le cytoplasme engendre une répartition aléatoire lors de la division cellulaire. Le plasmide ColE1 qui est de l'ordre d'une dizaine de copies par chromosome peut se maintenir de façon passive dans les bactéries au cours des générations. La probabilité d'apparition d'une cellule sans plasmide, par division cellulaire, suit la formule suivante :  $P_0=2^{1-n}$ , où  $n$  est le nombre de copies du plasmide avant la division cellulaire. La perte d'un plasmide présent en 10 exemplaires est prédite toutes les 512 générations, ce qui est peu fréquent et négligeable. Par contre, un plasmide présent à deux copies par cellule sera perdu toutes les deux générations. Les plasmides à bas nombre de copies tels que le plasmide F, ne peuvent pas se maintenir dans les cellules par diffusion passive.

### 3-2-2 Ségrégation active.

Les plasmides à faible nombre de copies et la plupart des chromosomes codent pour des systèmes de partition dédiés (*Par*) afin de séparer activement l'ADN vers les cellules filles avant la division cellulaire (Bouet, 2019). Ces systèmes de partition sont présents au niveau d'un locus *par* qui est composé d'un opéron constitué de deux gènes qui codent pour une protéine motrice NTPase, et pour une protéine adaptatrice ou protéine de liaison au centromère (CBP) qui se fixe spécifiquement sur une séquence centromérique ou site de partition, située au niveau de ce locus, en amont ou en aval de l'opéron. Dans la **figure 52** est représentée l'organisation génétique générale de ce locus qui peut différer selon le type de plasmide (Bouet, 2019). Les protéines constituant ce système de partition sont génériquement appelées *ParA* pour la protéine motrice NTPase, *ParB* pour la protéine adaptatrice (ou CBP) qui se fixe sur la séquence centromérique, *parS*. Cependant, il existe une

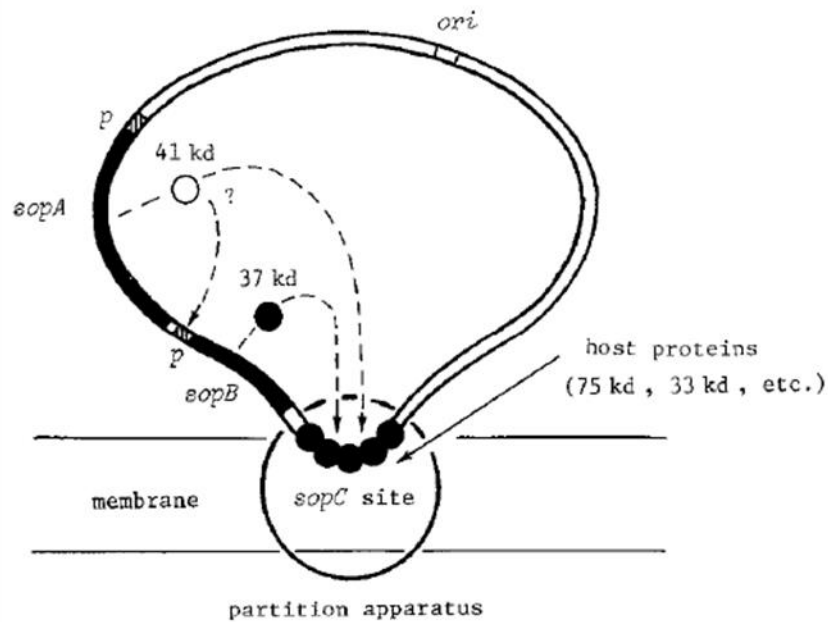
grande hétérogénéité de nomenclature entre les différents systèmes de partition étudiés (Bouet, 2019).



**Figure 52 : Organisation génétique des loci de partition plasmidiques.**

Les gènes codant pour les NTPases et les CBPs (Centromere Binding Protein) sont représentés par des flèches **rouges** et **bleues**, respectivement, et les sites du centromère sont représentés par des cases **vertes**. Les loci de partition de type I (*parABS*), de type II (*parMRC*) et de type III (*tubRZC*) se distinguent par leurs signatures NTPase, Walker-A (**rouge foncé**), actine-like (**rouge clair**) et tubuline-like (**rouge orange**), respectivement. Les CBPs abritent un motif de liaison à l'ADN HTH2 (**bleu clair**) ou RHH2 (**bleu foncé**). Pour le type III, l'ordre de la NTPase et CBP est inverse de celui des types I et II. Tous les plasmides schématisés se trouvent chez les *Enterobacteriaceae* sauf pTAR, pSM19035, et pBtoxis. Les noms historiques des gènes et centromères sont indiqués à l'intérieur des flèches ou sous les cases, respectivement. (légende et figure adaptées de Bouet, 2019).

Le premier locus de partition a été mis en évidence par des expériences de délétions et de complémentations sur la stabilité du plasmide F de chez *E. coli* : il a été appelé *sopABC* (pour stability of plasmid) avec le gène *sopA* et *sopB* correspondant respectivement aux protéines ParA et ParB ; *sopC* correspondant à la région centromérique *parS* (**fig. 53** ; Ogura, 1983). Ce même type de locus, appelé *parABS*, a été retrouvé sur le plasmide P1 de *E. coli* puis décrit sur les autres plasmides à faible nombre de copies (Austin, 1983a, 1983b ; Gerdes, 2010 ; Salje, 2010b). Un locus *par* fonctionnellement équivalent a été trouvé quelques années plus tard pour la ségrégation du chromosome de *Bacillus subtilis* où ParA et ParB sont appelées respectivement Soj et SpoJ car ces gènes avaient déjà été définis pour leurs rôles dans la sporulation de *Bacillus subtilis* (Ireton, 1994)



**Figure 53 : Premier modèle hypothétique pour l'appareil de partition du plasmide F.**

Régions pleines: les gènes *sopA* et *sopB*; régions hachurées: les promoteurs (p) des gènes; ori: l'origine de réplication du plasmide.

(légende et figure tirées de Ogura, 1983).

Le site *parS* contient des séquences d'ADN spécifiques qui agissent comme un centromère sur lesquelles la protéine adaptatrice ParB, sous forme de dimère se fixe pour former un complexe de partition. La protéine motrice ParA est une NTPase qui lie et hydrolyse l'ATP en ADP pour fournir l'énergie afin de déplacer le complexe ParB-*parS* et diriger sa ségrégation. Trois types de systèmes de partition ont été identifiés et classés en fonction de la signature de leurs NTPases (Brooks, 2017 ; Debaugny, 2018 ; Bouet, 2019 ; Jalal, 2020a) :

- type I, ParABS : avec une ATPase à motif « Walker A box », appelée ParA.
- type II, ParMRC : avec une ATPase de type actine, appelée ParM.
- type III, ParTubZRC : avec une GTPase de type tubuline, appelée TubZ.

Un quatrième système de partition a été trouvé plus récemment qui utilise une simple protéine avec une structure en super hélice, non-NTPase (Simpson, 2003).

Bien que les loci qui codent pour les systèmes Par présentent une remarquable similitude dans leur organisation génétique, des différences fondamentales dans la séquence et la structure de leurs NTPases ont conduit à des mécanismes divergents de partition des plasmides (Brooks, 2017). Chaque réplicon code pour son propre système ParABS et sa propre position intracellulaire dépend des interactions des trois composants ParABS. Le mécanisme dirigé par

ParA qui assure la localisation correcte et la ségrégation dirigée des réplicons dépend du positionnement du complexe de partition ParBS à l'intérieur du nucléoïde (Le Gall, 2016).

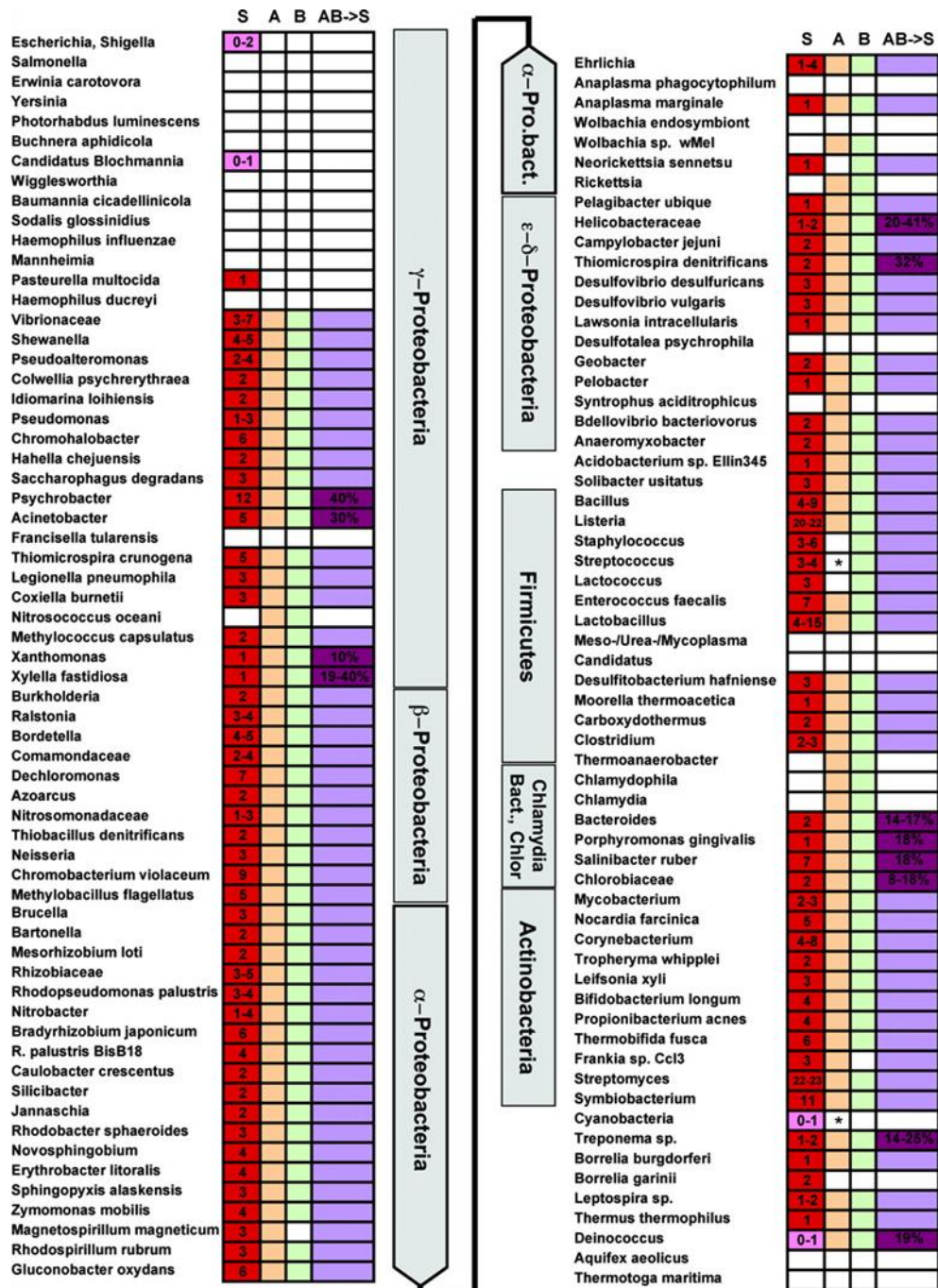
La ségrégation des chromosomes bactériens peut être divisée en trois étapes : (i) la séparation et relocalisation rapide des origines nouvellement répliquées, (ii) la ségrégation progressive des corps des chromosomes, et (iii) la séparation des terminus de réplication avant la fermeture du septum de division (Jalal, 2020a).

### 3-2-2-1 Ségrégation des origines de réplication.

L'un des premiers modèles de ségrégation des chromosomes bactériens a été proposé par François Jacob et ses collègues (Jacob, 1963) qui ont suggéré que les origines nouvellement répliquées pouvaient être ancrées à la membrane cellulaire et la croissance et l'allongement des cellules par leur centre contribueraient passivement à leur ségrégation. Il est maintenant établi que l'élongation des bactéries ne se limite pas à la croissance de la zone médiane de la cellule mais se produit sur des sites répartis tout le long de la membrane (Typas, 2011). De plus, des études sur le suivi des origines ont montré que la vitesse de déplacement des origines est plus rapide que la vitesse d'élongation de la cellule (Viollier, 2004 ; Wang, 2010).

Le locus *par* a tout d'abord été découvert pour les plasmides à faible nombre de copies et a été montré être essentiel pour leur maintien dans les cellules (Austin, 1983a, b ; Abeles, 1985 ; Mori, 1986). Le premier système de partition équivalent pour la ségrégation des chromosomes a été trouvé chez *B. subtilis* (Ireton, 1994 ; Lin, 1998).

A partir d'une approche bio-informatique pour rechercher des sites *parS* putatifs dans tous les réplicons séquencés, incluant chromosomes et éléments extra-chromosomiques (400 génomes procaryotes séquencés), Livny et collaborateurs (**fig. 54**) ont trouvé que 69% des souches contiennent des sites chromosomiques putatifs *parS*. Les espèces possédant ces sites sont présentes dans toutes les branches des procaryotes. Dans l'immense majorité des cas, les sites *parS* ont été identifiés dans la région proximale des origines de réplication du chromosome et à proximité du locus *parAB*. De plus aucun site *parS* caractéristique d'un chromosome principal n'a été identifié sur un chromosome secondaire. Ainsi les bactéries arborant plusieurs réplicons apparaissent posséder distincts locus *par* spécifique par réplicon (**fig. 54** ; Livny, 2007).



**Figure 54: Phylogénie du système ParABS.**

Les phylums et classes phylogénétiques sont indiqués en gris. Les chiffres indiqués dans la première colonne (S) indiquent le nombre de sites parS contenus dans les souches de chaque branche. Une case rouge indique que toutes les souches de la branche ont un site parS. Une case rose indique que seulement certaines souches de la branche ont un site parS. Une case vide indique qu'aucune souche de la branche ne contient de site parS. Dans les colonnes A et B, les cases sont colorées si toutes les souches d'une branche ont une séquence codant un homologue de ParA et de ParB respectivement. Les astérisques indiquent qu'une seule souche de la branche a une séquence codant un homologue de ParA ou de ParB. Dans la dernière colonne (AB ->S), les cases colorées correspondent aux branches dans lesquelles pour toutes les souches, les 3 composants du système ParABS sont contenus dans une région correspondante à 10% de la taille du chromosome. Les branches pour lesquelles les cases sont plus foncées indiquent que la distance est de plus de 10% pour au moins une souche. La distance moyenne est indiquée par les pourcentages. (légende et figure tirées de Livny, 2007).

### 3-2-2-1a Les systèmes de partition de type II et III.

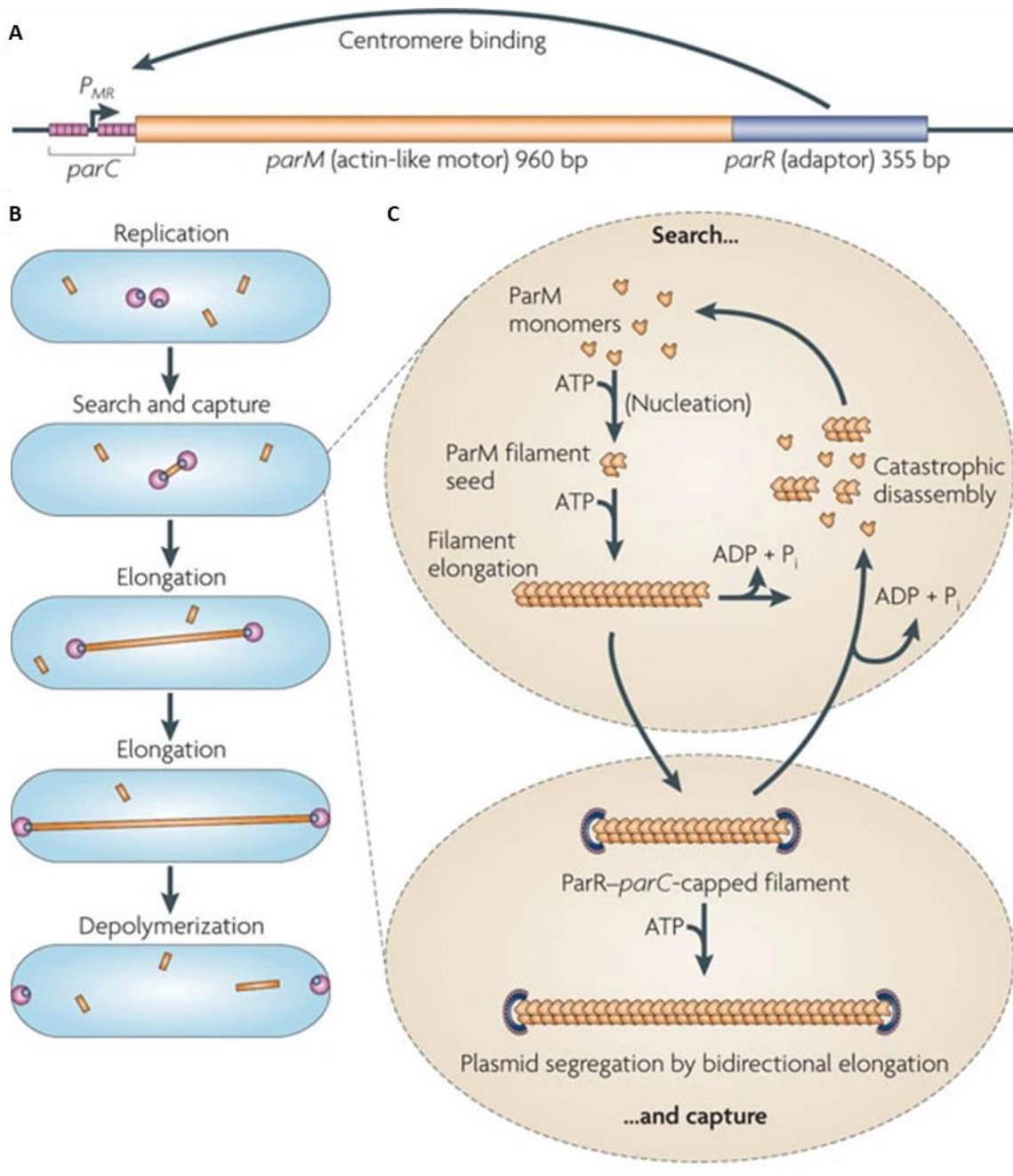
#### ● Les systèmes de partition de type II.

Le système de partition de plasmide, ParMRC, de type II est probablement le système le mieux caractérisé d'un point de vue moléculaire. Le locus *parMRC* a été isolé à partir du plasmide R1 à bas nombre de copies de chez *E. coli*. Ce locus *parMRC* est peu répandu, présent seulement sur quelques plasmides et n'a pas pu être identifié sur des chromosomes bactériens (Gerdes, 1985 ; Salje, 2010a).

Le locus opéron *parMRC* (**fig. 55 A**) code pour une protéine motrice ATPase de type actine, ParM, et une protéine adaptatrice, ParR. Il contient la région centromérique du plasmide, *parC*, composée de deux blocs de 5 séries de séquences répétées de 11bp chacune, interrompus par une région de 39bp contenant le promoteur. ParR est une petite protéine adaptatrice de 13kDa, avec un motif ruban-hélice-hélice (RHH<sub>2</sub>) qui se lie spécifiquement sur le site de partition, *parC* (Dam, 1994 ; Jensen, 1994 ; 1998). La protéine ATPase, ParM, est structurellement similaire à l'actine eucaryote, formant des faisceaux de filaments double brins (Van den Ent, 2002). Les protéines ParM forment de courts filaments en présence d'ATP qui présentent une forte instabilité dynamique, résultant de l'hydrolyse de l'ATP (Møller-Jensen, 2003 ; **fig. 55 C**). L'utilisation d'un analogue non hydrolysable de l'ATP permet de stabiliser ces polymères en formant une protection contre la dépolymérisation (Garner, 2004 ; Popp, 2007).

Le mécanisme moléculaire fondamental de la ségrégation de l'ADN par le système ParMRC est basé sur des faisceaux de filaments ParM de type actine qui, liés aux plasmides par ParR via les séquences *parC* poussent les plasmides vers les pôles opposés de la cellule par un mécanisme de polymérisation d'insertion (**fig. 55 B, C**). La polymérisation des protéines ParM fournit la force pour ségréger les plasmides.

ParM subit un changement dynamique entre des périodes de croissance et de réduction, suggérant un mécanisme de « recherche et capture » où les filaments ParM peuvent explorer continuellement le volume de la cellule pour se lier au complexe ParR-*parC* (Garnier, 2007). Cette instabilité dynamique de ParM résultant de l'hydrolyse de l'ATP est cruciale pour la partition des plasmides. Les filaments ParM sont dynamiquement instables à moins qu'ils ne soient coiffés par le complexe nucléoprotéique ParR-*parC* lié au plasmide (Garnier, 2007). Ce n'est que lorsqu'ils sont coiffés par un complexe ParR-*parC* aux deux extrémités que les



**Figure 55: Mécanisme du système de ségrégation de plasmide ParMRC codé par le plasmide R1.**

**A)** Organisation génétique de l'opéron *parMRC*: le centromère, *parC*, est composé de blocs de 5 courtes séquences répétées qui sont interrompues par une région de 39bp contenant le promoteur pour les protéines ParM et ParR. ParR, une protéine adaptatrice, se lie à *parC* et réprime la transcription.

**B)** Ségrégation des plasmides par les filaments d'actine dynamique, ParM. Les plasmides sont en **rose**, ParR en **bleu** et ParM en **orange**.

**C)** Recherche et capture: le mécanisme moléculaire de ségrégation du plasmide par ParM. ParM forme de courts filaments en présence d'ATP, qui soit subissent un désassemblage « catastrophique » suite à l'hydrolyse de l'ATP en ADP, soit sont coiffés à leur extrémité par la fixation du complexe ParR-*parC*, et effectuent une élongation stable bidirectionnelle.

(légende et figure adaptées de Salje, 2010).



filaments se stabilisent ; leur élongation bidirectionnelle ultérieure garantit que les plasmides liés aux filaments se déplacent vers les pôles opposés.

La protéine ParM est cinétiquement apolaire et présente une élongation équivalente à chaque extrémité contrairement à l'actine ou la tubuline qui ont des propriétés cinétiques distinctes à chaque extrémité. Cela est probablement nécessaire pour permettre aux filaments ParM de déplacer les plasmides dans les deux directions, aux pôles opposés de la cellule (Garner, 2004 ; Popp, 2007).

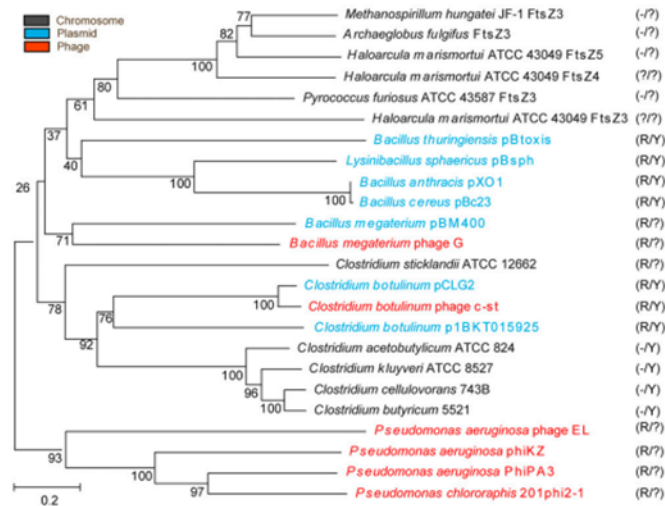
Ce modèle est conforté par une reconstitution fonctionnelle du système *in vitro* (Garner, 2007). Cette reconstitution, qui fait encore défaut pour les autres types de systèmes de ségrégation, ne nécessite aucun acteur extérieur au système ParMRC. Ceci indique que le système est autonome et n'interagit probablement pas avec un éventuel facteur cellulaire pour assurer sa fonction de ségrégation.

- **Les systèmes de partition de type III.**

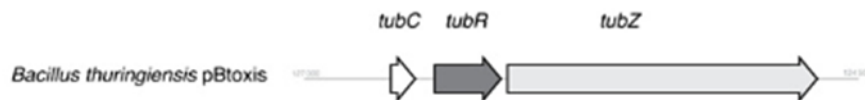
La découverte de la protéine motrice NTPase TubZ de type tubuline a permis d'identifier un troisième type de système de partition des génomes bactériens. Ce système de partition de type III consiste en une protéine motrice GTPase de type tubuline, appelé TubZ, une protéine adaptatrice, TubR et un site centromérique, *tubC*. Ce système de partition TubZRC a été retrouvé dans de nombreux grands plasmides porteurs des facteurs de virulence du genre *Bacillus*, notamment le plasmide pBtoxis de *Bacillus thuringiensis* (Tang, 2006) et le plasmide pXO1 de *Bacillus anthracis* (Akhtar, 2009) et *Bacillus cereus* (Hoshino, 2012). Il a été également trouvé dans des bactériophages des espèces *Clostridium* et *Pseudomonas* (Sakaguchi, 2005 ; Kraemer, 2012). Contrairement aux homologues de la famille FtsZ codés par le chromosome, et nécessaires à la division cellulaire bactérienne, les protéines de type tubuline transmises par les plasmides et les phages ont des fonctions différentes (**fig. 56 A** ; Fink, 2017).

Le système le mieux étudié est celui trouvé sur le plasmide pBtoxis de *B. thuringiensis*. Ce plasmide contient un opéron qui code pour deux protéines (**fig. 56 B**) : ORF156 pour la protéine TubZ et ORF157 pour la protéine TubR (Berry, 2002 ; Larsen, 2007 ; Tang, 2007). La protéine TubR est une petite protéine de liaison à l'ADN de 11,6 kDa qui possède un motif hélice-tour-hélice (HTH<sub>2</sub>) trouvé dans un grand nombre de protéines de liaison à l'ADN, procaryotes ou eucaryotes. Elle ne présente aucune homologie de séquence avec une protéine connue. Elle se lie sur une séquence centromérique, *tubC*, qui est localisée

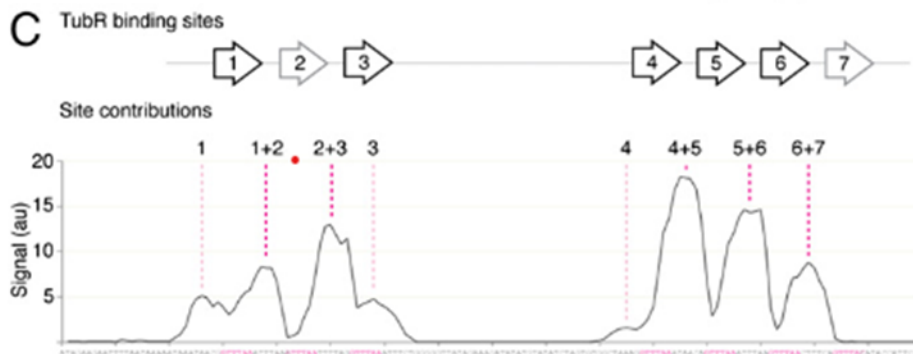
A



B



C ;



**Figure 56: Distribution et organisation génomique des systèmes TubZ au sein des plasmides et des bactériophages.**

**A)** Arbre phylogénétique des systèmes TubZ: les systèmes codés par les plasmides sont représentés en **bleu**, ceux transmis par les phages en **rouge** et les gènes chromosomiques apparentés en **gris**. A droite, (R) et (Y) indiquent la présence ou absence de TubR et TubY, respectivement, près de *tubZ*.

Les points d'interrogation indiquent la présence de protéines putatives qui partagent avec TubR seulement un motif HTH et une petite taille ou qui ne comprennent seulement des longs motifs de super-hélice comme TubY (d'après Oliva, 2012).

**B)** Représentation du locus *tubZRC* du plasmide pBtoxis de *Bacillus thuringiensis* (d'après Aylett, 2012).

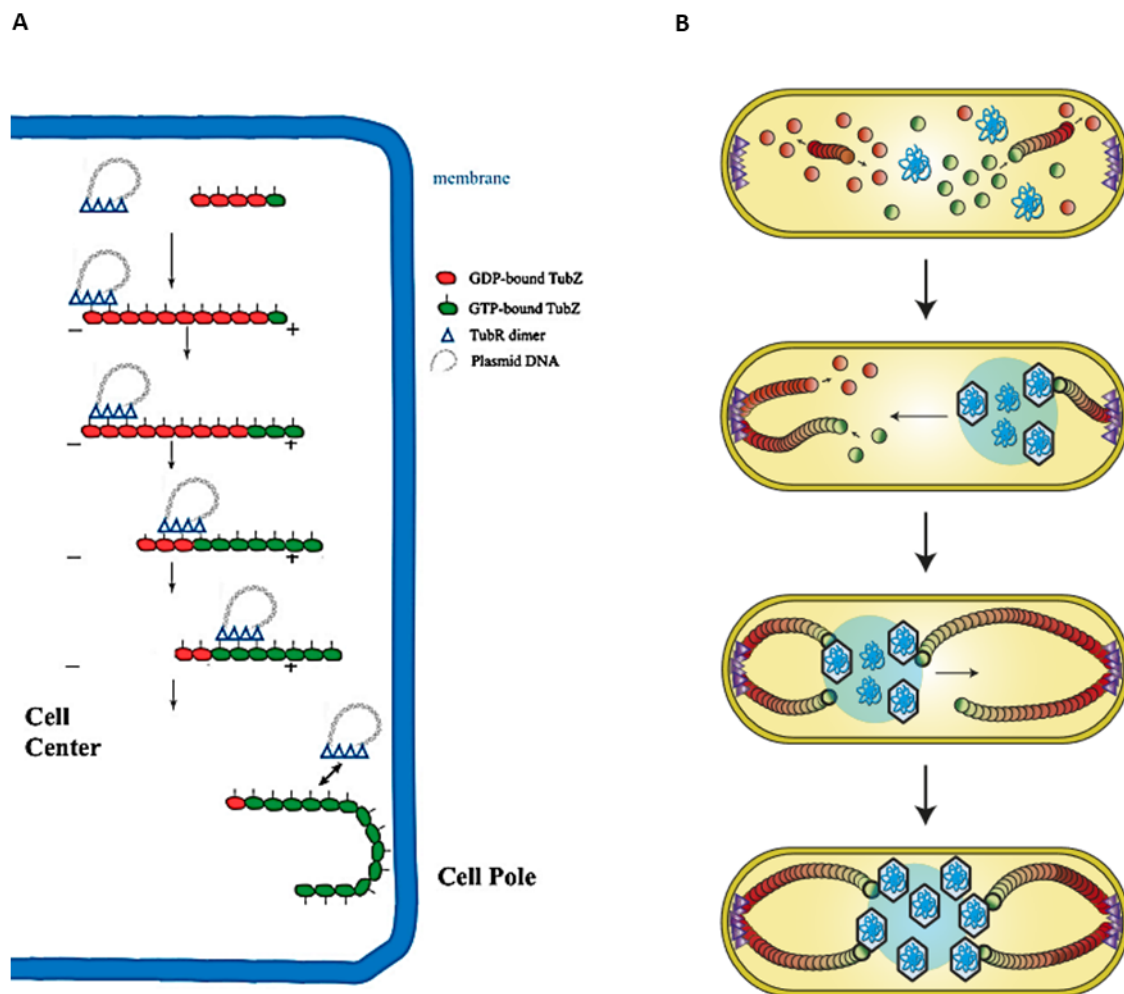
**C)** Tracé du signal enregistré pour chaque point de la biopuce dans un balayage de 1 pb sur la région *tubC* de *B. thuringiensis* (de la position 126688 à 126496). Les sites de liaison attribués au TubR de *B. thuringiensis* sont indiqués ci-dessus, et les séquences correspondantes sont colorées en **magenta**. Le(s) site(s) résultant de chaque pic ont été annotés au-dessus du graphique.

(d'après Aylett, 2012).

directement en aval des gènes *tubZ* et *tubR* sur le plasmide. Cette séquence est composée de 7 répétitions de 12bp qui sont réparties en deux blocs de 3 et 4, séparés par 45bp (Tang, 2007 ; Aylett, 2012 ; **fig. 56 B, C**). La protéine TubZ est une protéine de 54,4 kDa qui peut s'assembler en filaments de manière GTP-dépendante. La protéine fait partie de la famille des GTPases de type tubuline qui peut s'assembler *in vitro* en présence de GTP et également *in vivo* (Larsen, 2007) où l'extrémité positive d'un filament s'allonge, tandis que l'extrémité négative se rétracte, ce qui entraîne une polymérisation dans un mouvement dit de « tapis roulant ». La présence du complexe TubR-*tubC* accentue et stabilise la polymérisation (Aylett, 2012 ; Oliva, 2012). L'utilisation de l'analogue non hydrolysable du GTP, le GTP $\gamma$ S, stabilise ces polymères en formant une protection contre la dépolymérisation comme dans les systèmes de type II. Le complexe de partition prévient donc la dépolymérisation des filaments de tubuline en empêchant l'hydrolyse du GTP.

Dans la **figure 57 A** est représenté le mécanisme de partition de type III dit du « tapis roulant » chez *Bacillus thuringiensis* : le complexe TubR-pBtoxis se lie à l'extrémité C-terminale des protéines TubZ via TubR. L'élimination des sous-unités TubZ au niveau de l'extrémité négative du polymère et l'ajout de sous-unités à l'extrémité positive allongent le filament et permettent au complexe TubR-pBtoxis de se déplacer comme sur un tapis roulant vers l'un des pôles de la cellule. Arrivé au niveau d'un de ces pôles, le complexe TubR-pBtoxis se détache du filament qui repart en direction de l'autre pôle pour récupérer le deuxième complexe TubR-pBtoxis et le transporter vers l'autre pôle de la cellule (Ni, 2010).

Certains bactériophages contiennent des homologues de la protéine TubZ, appelé PhuZ, et le système TubZ serait requis pour la ségrégation de bactériophages circulaires comme pour le cycle du phage 201 $\phi$ 2-1 chez *Pseudomonas chlororaphis* (**fig. 57 B**). Dans les cellules infectées par le phage 201 $\phi$ 2-1, les filaments PhuZ semblent être ancrés aux pôles des cellules et s'allonger en direction des génomes phagiques afin de les maintenir au centre de la cellule pour être répliqués et encapsidés efficacement (Kraemer, 2012 ; Erb, 2014).



**Figure 57: Modèle du système de partition de type III, dit du « tapis roulant ».**

**A)** Chez *Bacillus thuringiensis*: la protéine adaptatrice TubR, sous forme de dimère, se lie à un plasmide pBtoxis répliqué au niveau de la région centromérique *tubC*, et entre en contact avec la région C-terminale de la protéine TubZ (indiquée par des lignes pointant à partir des "cercles" TubZ) dans un polymère TubZ en mouvement de type « tapis roulant ». Les sous-unités TubZ sont éliminées au niveau de l'extrémité négative (-) du filament et ajoutées à l'extrémité positive (+). TubR est tirée le long du polymère en croissance par son interaction TubR-TubZ jusqu'à ce que le plasmide atteigne le pôle cellulaire où il se détache lorsqu'il entre en contact avec la membrane au niveau du pôle de la cellule. TubZ part dans la direction de l'autre pôle et peut récupérer l'autre complexe TubR-pBtoxis et le délivrer de la même manière au pôle cellulaire opposé.  
(d'après Ni, 2010).

**B)** Cycle du phage 201φ2-1 chez *Pseudomonas chlororaphis*: après que le phage ait infecté une cellule, le chromosome hôte est dégradé et de courts filaments de PhuZ (TubZ de phage) apparaissent et s'étendent des pôles de la cellule jusqu'au nucléoïde du phage. Les sous-unités liées au GTP (vert) s'ajoutent à une extrémité pour permettre de positionner l'ADN au centre de la cellule et d'être efficacement répliqués et/ou empaquetés dans les capsides. Une fois l'assemblage des phages terminé, la cellule se lyse, expulsant les phages matures.  
(d'après Erb, 2014).

### 3-2-2-1b Les systèmes de partition de type I : le système ParABS.

Les systèmes de partition de type I, ParABS, sont les plus répandus pour les plasmides à faible nombre de copies et le seul type connu pour être présent sur les chromosomes bactériens (Salje, 2010b).

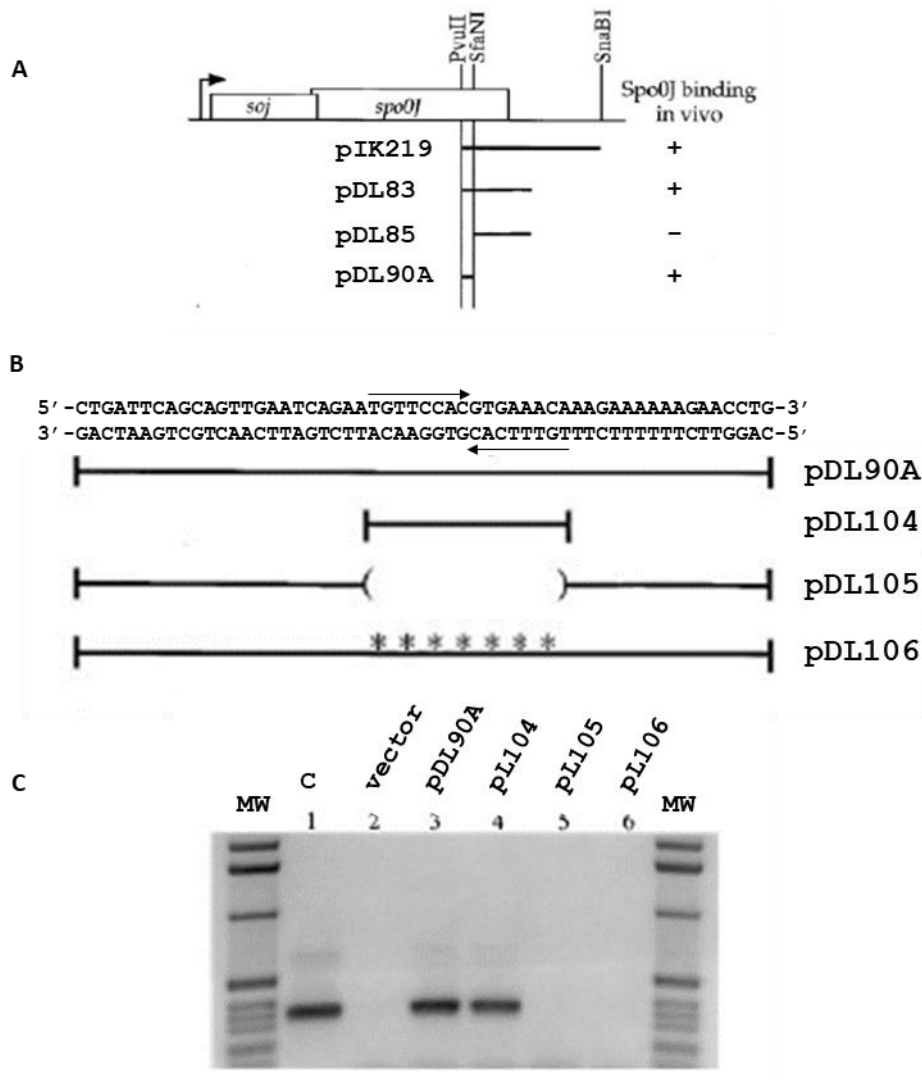
Ils contiennent, comme pour les systèmes de type II et III, trois éléments essentiels : une séquence nucléotidique de type centromérique, *parS*, une protéine adaptatrice de liaison à l'ADN, ParB qui se fixe sur *parS*, et une protéine motrice ATPase de type Walker A (aussi appelée P-loop), ParA qui est impliquée dans le positionnement du réplicon en déplaçant le complexe ParB-*parS* grâce à l'énergie fournie par l'hydrolyse de l'ATP.

Les séquences centromériques *parS*, appelées également centromères bactériens, sont des séquences d'ADN agissant en cis qui, avec les protéines ParA et ParB sont impliquées dans la ségrégation des plasmides et chromosomes. La liaison spécifique de ParB sur *parS* permet l'assemblage d'un large complexe nucléoprotéique, appelé « complexe de partition ». Dans les grandes familles de systèmes de partition, on observe une grande diversité au niveau des protéines ParB et du site *parS*, ainsi qu'un niveau de l'organisation du complexe de partition (Hayes, 2006 ; Bouet, 2019).

- **La séquence centromérique, *parS*.**

Le premier site d'ADN impliqué dans la partition chromosomique a été identifié chez *Bacillus subtilis* en 1998 par Lin et Grossman (Lin, 1998). Ce site a été identifié *in vivo*, par immunoprécipitation, comme le site de liaison pour la protéine de partition chromosomique Spo0J, un membre de la famille des protéines de partition ParB. La protéine Spo0J est une protéine de liaison à l'ADN site-spécifique qui reconnaît une séquence de 16bp trouvée dans le gène *spo0J*. Ce site, appelé *parS*, contient une répétition inversée de 8bp au niveau du gène *spo0J* (fig. 58). Huit sites ont été trouvés permettant la fixation de la protéine Spo0J *in vivo*, situés dans les 20% du chromosome situé autour de l'origine de réplication. La présence de ce site dans un plasmide instable de *B. subtilis* le stabilise d'une manière Spo0J dépendante et confère une fonction de partition pour un réplicon hétérologue (Lin, 1998). A partir de la séquence palindromique consensus de 16bp identifiée chez *B. subtilis*, « 5'-TGTTNACGTGAAACA-3' » et des génomes disponibles dans les bases de données de l'époque, Lin et Grossman ont

identifié 10 autres génomes contenant ces séquences *parS*. Par une approche expérimentale, différentes équipes ont trouvé 7 autres espèces bactériennes contenant des sites *parS*.



**Figure 58: Identification *in vivo* du site de fixation de Spo0J (*parS*) localisé dans le gène *spo0J*.**

Les plasmides contenant différents inserts de la région *spo0J* sont testés *in vivo* pour la liaison de Spo0J par immuno-précipitation.

**A)** L'opéron *soj-spo0J* est dessiné schématiquement, et les inserts contenus dans les différents plasmides sont indiqués dessous. Le présence (+) ou l'absence (-) des quatre différents plasmides dans l'immuno-précipitation est indiquée. Le site de liaison *Spo0J* est contenu dans le fragment de 55bp PvuII-SfaNI dans le plasmide pDL90A.

**B)** La séquence de 16bp contenant une répétition inversée imparfaite de 8bp est indiquée par des flèches en dessus et en dessous de la séquence de 55bp insérée dans le plasmide pDL90A. Les inserts contenus dans les plasmides pDL104, pDL105 et pDL106 sont schématisés. pDL106 contient 7 changements par rapport au sauvage.

**C)** La séquence de 16bp est le site de liaison de Spo0J. Réactions PCR avec des amorces spécifiques du plasmide et l'ADN indiqué. Ligne 1: contrôle avec le vecteur purifié. Lignes 2-6: ADN immuno-précipité à partir des souches contenant le plasmide indiqué.

(légende et figure tirées de Lin, 1998).

En 2007, Livny, Yamaichi et Waldor, utilisant une matrice consensus *parS* à partir de *Bacillus subtilis* et *Streptomyces coelicolor*, ont effectué une approche bio-informatique pour rechercher des sites *parS* sur 400 séquences chromosomiques disponibles dans les bases de données. 69% des espèces bactériennes possèdent des séquences *parS* spécifiques, recouvrant toutes les branches de l'arbre évolutif procaryote (Livny, 2007). Ces données confirment les approches expérimentales, où les sites *parS* retrouvés dans 17 génomes sont très similaires à la séquence consensus de *B. subtilis*. Ces séquences *parS* sont le plus souvent proche de l'origine de réplication, 92% localisées dans les 15% du chromosome autour de l'origine de réplication (Livny, 2007). Le nombre de site *parS* est très variable que ce soit entre les différents genres bactériens qu'entre différentes espèces d'un même genre (**fig. 54**) : d'un seul site pour *Thermus thermophilus* ou *Pasteurella multocida*, à 20-22 pour *Listeria* ou 22-23 pour *Streptomyces* (Livny, 2007). En moyenne on retrouve 1 à 8 sites. Dans le cas de bactéries multichromosomiques, les chromosomes principaux et secondaires ne possèdent pas le même site *parS* afin d'éviter les problèmes d'incompatibilité entre réplicons portant le même système de ségrégation.

Replicon	ATPase (avg residues)	DNA-binding protein (avg residues)	Minimal centromere repeat
Universal chromosome	ParA (250–270) <sup>b</sup>	ParB (280–330) <sup>b</sup>	tGTTNCACGTGAAACa <sup>f</sup> ↔↔
J2315 c1	ParA (259)	ParB (305)	TGTTTCACGTGAAACa ↔↔
J2315 c2	ParA (220)	ParB (353)	gTTTATGCGCATAAAc ↔↔
J2315 c3	ParA (231)	ParB (343)	gTTGTACGTGACAAc ↔↔
J2315 pBC	ParA (223)	ParB (290)	qITGGCTCGAGCCAAg ↔↔
<i>Ralstonia</i> c2	ParA (222)	ParB (334)	cTTTCAGCGCTGAAAg ↔↔
Plasmid RK2	IncC (364–259) <sup>d</sup>	KorB (358)	TTTAGCSGCTAAA ↔↔
<i>Vibrio</i> c2	ParA (407)	ParB (323)	NTTTACANTGTAAAN ↔↔
Phage P1 (Ia) Plasmid F (Ia)	ParA (398) SopA (391)	ParB (333) SopB (323)	(A) ATTTTCAM (B) TCGCCA <sup>f</sup> TGGGACCACGGTCCCA ↔↔
<i>Rhizobiales</i> MP	RepA (400–435) <sup>b</sup>	RepB (300–370) <sup>b</sup>	GTTNNCNGCNGNNAAC <sup>f</sup> ↔↔
Plasmid TP228 (Ib)	ParF (206)	ParG (76)	ACTCWWWW
Plasmid pTAR (Ib)	ParA (222)	ParB (94)	AYCCGRT
Plasmid pB171 (Ib)	ParA (214)	ParB (91)	TTATKA

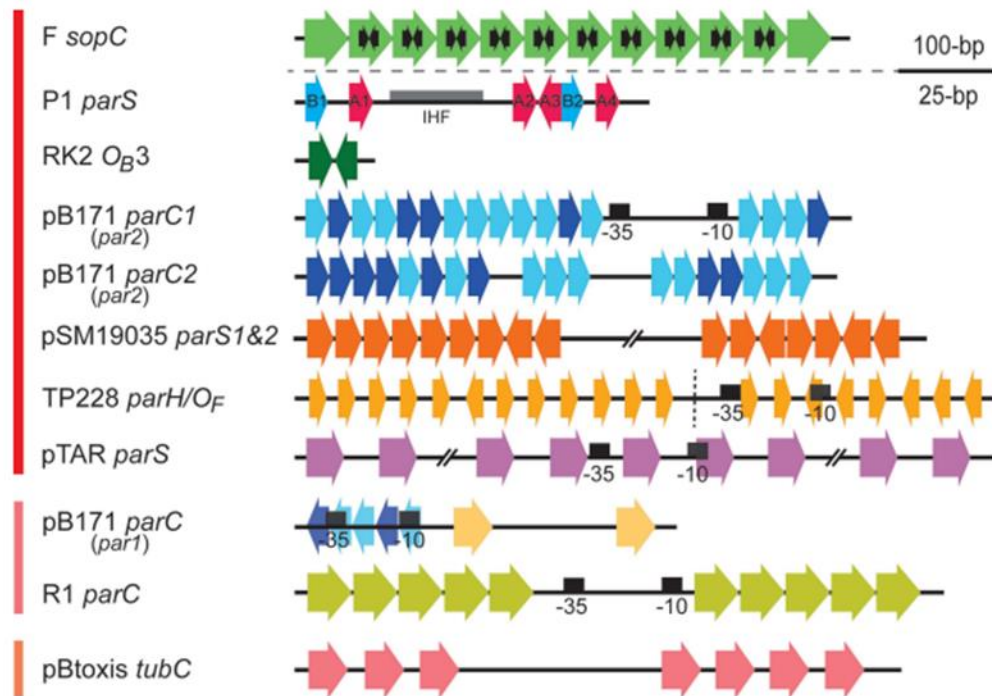
**Tableau 2: Caractéristiques des loci *par* des réplicons de *B. cenocepacia* J2315 et autres systèmes.**

c1, c2 et c3 se réfèrent à des chromosomes; MP à un mégaplasme.

Les flèches convergentes indiquent une symétrie. Les lettres en minuscules indiquent des bases qui varient dans quelques répétitions naturelles détectées.

(figure tirée de Passot, 2012).

La séquence palindromique consensus de 16bp identifiée chez *B. subtilis*, « 5'-TGTTNCACGTGAAACA-3' » rassemble un certain nombre de critères : (i) au moins 14bp (ii) un îlot central du palindrome composé de 2 nucléotides CG (iii) une région en 5' riche en T (Passot, 2012).



**Figure 59: L'organisation et la composition des centromères sont très diverses.**

Les répétitions directes et inversées sont représentées par des flèches orientées de même couleur, indiquant les motifs conservés. Pour  $parS_F$ , les flèches inversées noires représentent les sites de liaison ParB à répétition inversée de 16 pb au sein des répétitions directes de 43 pb (flèches vertes). Le centromère de P1 est composé de deux boîtes B de 6 pb (bleu) et de quatre boîtes A de 7 pb (rouge) présents des deux côtés d'un site de liaison du facteur hôte d'intégration (IHF) (rectangle gris).

Le centromère de RK2 ( $O_B3$ ) est constitué d'une répétition inversée de 13 pb. Pour pB171, les régions  $parC1$  et  $parC2$  de  $par2$  sont composées de 17 (2 groupes) et 18 (3 groupes) répétitions, respectivement, d'un motif de 6 pb. Le site  $parS$  de  $par1$  ne comprend que deux motifs identiques de 10 pb (flèches orange) en orientation directe séparés par une répétition directe de 31 pb ; les flèches bleues chevauchant les séquences promotrices -35 et -10 correspondent au début de  $parC1$  à partir du locus  $par2$  impliqué dans la régulation croisée entre les deux systèmes Par de pB171. pSM19035 porte trois loci  $parS$  composés de répétitions contiguës de 7 pb en orientation directe ou inverse (seuls  $parS1$  et  $parS2$  sont représentés). Pour  $parSTP228$ , les deux régions centromériques,  $parH$  (à gauche) et  $OF$  (à droite), délimitées par une ligne pointillée verticale, sont composées de 12 et 8 répétitions dégénérées (4 pb) séparées par des espaceurs riches en AT (4 pb), respectivement. Le centromère du pTAR contient 13 répétitions de 9 pb, chacune séparée par 8 pb, englobant les boîtes des promoteurs -35 et -10. Le centromère du plasmide R1 comprend deux réseaux, espacés de 39 pb, composés de cinq répétitions directes de 11 pb. Le centromère de pBtoxis,  $tubC$ , comprend deux réseaux de trois et quatre motifs de 12 pb, séparés par 54 pb. (légende et figure tirées de Bouet, 2019).

Contrairement aux unités  $parS$  chromosomiques qui sont des palindromes de 16 pb de séquence remarquablement uniforme ("universelle") dispersés dans la région entourant l'origine de réplication, les séquences centromériques plasmidiques, localisées proches des



gènes *parAB*, sont constituées de séquences répétées. Leur nombre, leur séquence ainsi que leur organisation peut varier d'un plasmide à un autre (**tableau 2 ; fig. 52 ; fig. 59** ; Lin, 1983 ; Hayes, 2006 ; Passot, 2012). Les centromères des plasmides sont généralement constitués de répétitions directes d'ADN, mais peuvent inclure des répétitions inversées et/ou des répétitions directes à symétrie interne. Les répétitions peuvent être étendues, par exemple, *sopC* du plasmide F consiste en 12 répétitions en tandem d'un motif de 43 pb (Mori, 1989), mais seulement 10 sont fonctionnelles (Pillet, 2011) et *parS* de pTAR d'*Agrobacterium tumefaciens* contient 13 répétitions heptamériques qui sont séparées par un tour d'hélice à chaque fois (Gallie, 1987 ; Kalnin, 2000). En revanche, le site de partition  $O_{B3}$  du plasmide RK2 est une séquence centrale palindromique plus simple de 13 pb, bien que des sites apparentés dispersés dans ce plasmide puissent également contribuer à la ségrégation (Williams, 1998). La liaison des facteurs de partition aux centromères induit probablement des changements topologiques dans la structure de l'ADN, en particulier dans les sites qui présentent des motifs de répétition étendus. Par conséquent, la structure d'ordre supérieur du ségrosome est cruciale pour son bon fonctionnement.

- **Les protéines adaptatrices ou les protéines se fixant sur les séquences *parS* : les protéines ParB.**

Chez les bactéries, la ségrégation ou partition des plasmides à faible nombre de copies et des chromosomes dépend de l'activité des protéines de liaison à l'ADN site-spécifiques qui reconnaissent une ou plusieurs copies d'un site d'ADN semblable à un centromère, les séquences *parS*. Ces "protéines de liaison au centromère" (ou CBP pour centromere binding protein) ou protéines adaptatrices ou ParB travaillent généralement de concert avec une ATPase, ce qui entraîne un mouvement et un positionnement dynamiques des plasmides ou des domaines chromosomiques au cours du cycle cellulaire (Wang, 2013 ; Baxter, 2014 ; Bouet, 2014).

Les protéines de liaison au centromère appartiennent à l'une des deux classes structurales suivantes : les protéines contenant un domaine de liaison à l'ADN de type hélice-tour-hélice (HTH<sub>2</sub>) ou ruban-hélice-hélice (RHH<sub>2</sub>). Chez les bactéries, toutes les protéines des systèmes de partition des chromosomes bactériens et de nombreux plasmides possèdent un domaine HTH. Elles partagent des propriétés similaires *in vivo* et *in vitro*, notamment une organisation similaire des domaines et des propriétés de liaison à l'ADN. Ils forment tous de grands

complexes de partition *in vivo* qui peuvent être visualisés comme des foyers en utilisant des approches de fluorescence (Funnell, 2016).

Ces protéines ParB partagent une organisation en domaine similaire bien que la conservation de la séquence primaire ne soit pas très forte parmi les membres de cette famille. Les protéines peuvent être divisées en trois domaines (**fig. 60**) :

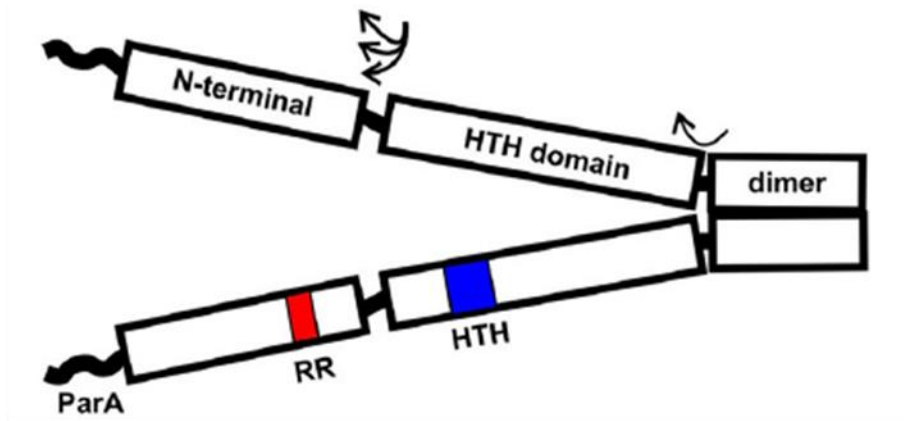
- **un domaine N-terminal (NTD)** qui contient un motif riche en arginine très conservé (GERRxRA) qui permet l'interaction protéine-protéine et protéine-ligand. Ce domaine est nécessaire pour l'oligomérisation des protéines et l'interaction avec ParA et stimule son hydrolyse de l'ATP (Surtees, 1999 ; Yamaichi, 2000 ; Funnell, 2016 ; Jalal, 2020a). Récemment une activité CTPase a été recensée au niveau de ce domaine NTD. ParB lie et hydrolyse la cytidine triphosphate (CTP) qui se fixe au niveau du patch d'arginine (Osorio-Valeriano, 2019 ; Soh, 2019).

- **un domaine central de liaison à l'ADN (DBD)** qui contient un motif type hélice-tour-hélice (HTH) permettant la fixation spécifique des protéines ParB sur les séquences *parS* mais aussi des liaisons avec l'ADN non-spécifique (Sanchez, 2013 ; Funnell, 2016 ; Jalal, 2020a).

- **un domaine C-terminal (CTD)** qui est le domaine le moins bien conservé parmi les homologues de ParB. Ce domaine contient un motif en « glissière de leucine » ou leucine zipper qui est un motif de dimérisation et permet l'homodimérisation des protéines ParB (Surtees, 1999 ; Funnell, 2016 ; Jalal, 2020a). Le CTD de la protéine ParB de *B. subtilis* a aussi un patch d'acides aminés riches en lysine qui lui fournit des activités additionnelles de liaison à l'ADN non spécifique et de condensation de l'ADN (Fisher, 2017).

Des liens flexibles relient les domaines, et une flexibilité dans l'organisation, l'orientation et le repliement des domaines a été observée dans des expériences biochimiques et des structures cristallines.

Actuellement, la structure globale d'une protéine ParB n'est pas disponible. La flexibilité de ParB, due aux liens des d'acides aminés qui relient les domaines consécutifs, a entravé les efforts visant à cristalliser et à résoudre la structure d'une protéine complète. Néanmoins, des idées sur la structure-fonction ont été obtenues à partir d'études de cristallographie aux rayons X et de RMN utilisant un seul domaine ou des variantes tronquées de domaines de ParB provenant de diverses espèces bactériennes (Leonard, 2004 ; Schumacher, 2010 ; Jalal, 2020a).



**Figure 60: Schéma de la structure de la protéine ParB, représentée sous forme d'un dimère.**

Les rectangles noirs représentent les régions de la protéine pour lesquelles il existe des informations structurales : le domaine N-terminal, le domaine de liaison à l'ADN HTH et le domaine dimère C-terminal. Les trois régions sont reliées par des séquences de liaison flexibles (flèches). La longueur de la liaison entre les domaines est représentée ici comme courte, comme dans les structures publiées de HpSpo0J et P1 ParB (Schumacher, 2005 ; Chen, 2015), mais peut être plus longue dans d'autres ParBs. La ligne ondulée représente la région qui interagit avec ParA dans beaucoup de ParBs, mais pas tous. La position du motif HTH (bleu) et du motif patch arginine conservé (RR, rouge) est indiquée dans un monomère. (légende et figure tirées de Funnell, 2016).

Les comparaisons structurales ont suggéré que ParB, en particulier son NTD, peut adopter de multiples conformations alternatives qui pourraient faciliter l'assemblage d'un complexe nucléoprotéique d'ordre supérieur.

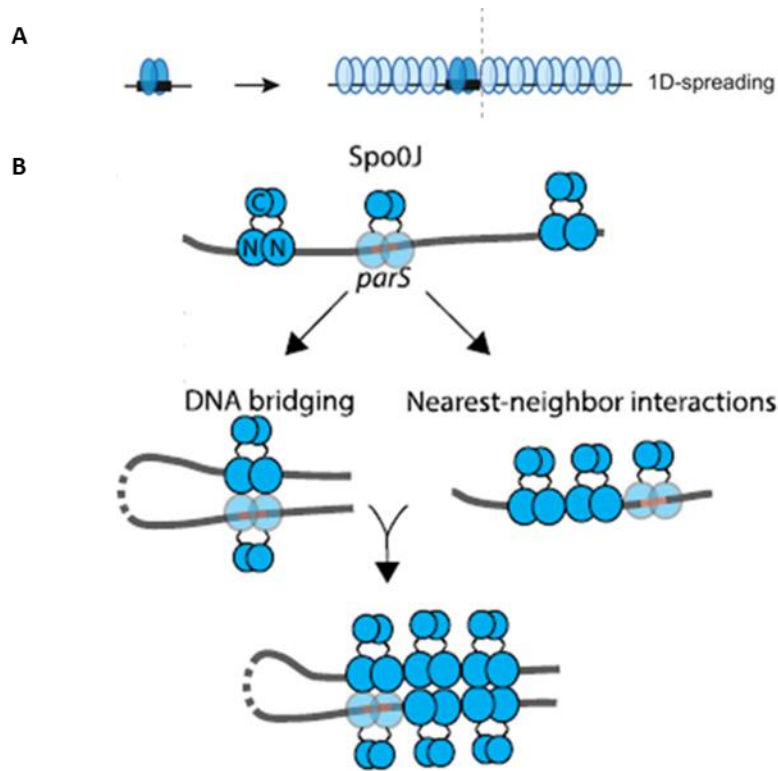
L'assemblage du complexe de partition commence par la reconnaissance de *parS* par un dimère de ParB, suivi par le chargement de plusieurs dimères de ParB pour former un très grand complexe protéine-ADN (Baxter, 2014). Ces complexes d'ordre supérieur sont nécessaires à la fois comme substrats et comme activateurs des mécanismes de partition. Les séquences *parS* servent de point de nucléation pour l'assemblage de complexe de partition ParB-*parS* de haut poids moléculaire. Mesurée par des approches de ChIP-seq (Sanchez, 2015), la fixation de ParB *in vivo* s'étend, s'étale au-delà des séquences *parS* sur l'ADN environnant, sur plus de 10kb autour du site *parS* (ChIP +PCR, Rodionov, 1999 ; Murray, 2006). Ce phénomène d'étalement ou « spreading » fait référence à la liaison de ParB aux régions d'ADN adjacentes du centromère d'une manière non-spécifique. Ce phénomène d'étalement a été initialement remarqué par la répression de gènes situés au voisinage des sites *parS* (Lynch, 1995 ; Rodionov, 1999). Cette capacité de réduire au silence les gènes environnants, spécialement quand ParB est surexprimée est une conséquence mais n'est pas requis pour la partition (Rodionov, 2004). Cependant ce « spreading » est nécessaire car des mutants ParB qui ne « s'étaient » pas, ont des défauts de partition (Debaugny, 2018). En particulier les

patches d'arginine dans ParB sont nécessaires à cet étalement (Rodionov, 1999 ; Graham, 2014 ; Debaugny, 2018). Ce « spreading » peut être bloqué par des complexes nucléoprotéiques tel que les complexes d'initiation de la réplication dans le cas des plasmides P1 et F (Rodionov, 1999 ; Sanchez 2015) ou par des complexes répresseur-opérateur pour les chromosomes bactériens (Murray, 2006).

Cet étalement n'est pas une caractéristique habituelle pour les protéines de liaison à l'ADN site-spécifique mais est commun aux protéines ParB à motif HTH et reflète comment ParB s'assemble en des complexes d'ordre supérieur (Rodionov, 1999 ; Murray, 2006 ; Sanchez, 2015).

Plusieurs modèles ont été proposés concernant l'assemblage d'un complexe nucléoprotéique ParB-ADN d'ordre supérieur impliqué dans la partition des génomes.

Le premier modèle et le plus simple, « étalement en une dimension » ou « **1D-spreading** », décrit l'étalement comme une association latérale protéine-protéine le long de l'ADN sous la forme d'un filament sur une dimension (**fig. 61 A**). Les dimères ParB se lient spécifiquement et fortement sur les séquences *parS* puis se propagent le long de l'ADN adjacent, dans les deux directions, par une polymérisation latérale de ParB due à des interactions protéine-protéine, formant un filament à partir de *parS* et pouvant couvrir jusqu'à 10 à 20 kb (Rodionov, 1999 ; Murray, 2006). Cependant des études ultérieures ont montré que la concentration intracellulaire de ParB est trop faible pour former un filament en une dimension s'étalant sur une telle distance. Par une technique d'empreinte de pied ou « foot-printing », il a été montré qu'une protéine ParB (en l'occurrence Spo0J, l'expérience a été faite chez *B. subtilis*) occupe 30bp, ce qui implique qu'il faudrait plusieurs centaines de molécules Spo0J par site *parS* (Murray, 2006). Par une approche d'immunoblots quantitatifs, Graham et collaborateurs ont montré qu'en moyenne chaque cellule contient 440 dimères Spo0J pour 24 sites *parS* soit 20 dimères Spo0J pour chaque site *parS* (Graham, 2014). Avec ce modèle de « 1D-spreading », cela permettrait de couvrir seulement 500bp par site *parS*, beaucoup moins que les 10-20kb observés (Graham, 2014). Ceci est substantiellement plus bas que les 10-20kb d'ADN lié par ParB, observé par CHIP-chip (Murray, 2006). Cela est aussi vrai pour les ParB plasmidiques (Sanchez, 2015).



**Figure 61: Modèles pour l'étalement de ParB et l'assemblage des complexes de partition (I).**

**A) Modèle du «1D-spreading»:** le dimère ParB (ovale **bleu foncé**) se lie spécifiquement à la séquence centromérique *parS* (rectangle **noir**). Les dimères de ParB se propagent par des interactions entre voisins les plus proches le long de la molécule d'ADN (ligne noire) et forment des filaments qui s'éloignent de *parS* dans les deux directions (figure tirée de Sanchez, 2015).

**B) Modèle du « spreading and bridging »** (étalement et pontage): les dimères Spo0J (équivalent de ParB chez *B. subtilis*) se lient fortement aux sites *parS* et plus faiblement à l'ADN non-spécifique. Le domaine N-terminal de liaison à l'ADN (N) et le domaine C-terminal de dimérisation (C) sont montrés dans des cercles. Les dimères ParB éloignés peuvent interagir ensemble formant des ponts d'ADN (DNA bridging) et piégeant des boucles d'ADN. Des dimères situés à côté peuvent également coopérer (nearest-neighbor interactions) pour former des complexes d'ordre supérieur plus grand.

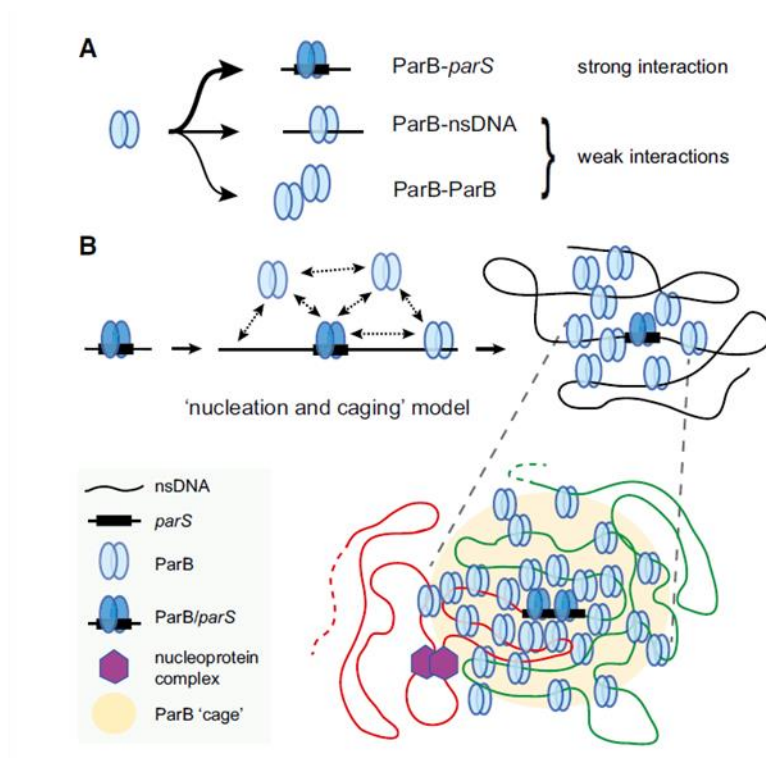
(légende et figure tirées de Graham, 2014).

Le deuxième modèle « étalement et pontage » ou « **spreading and bridging**, » suggéré par Graham et collaborateurs combinerait un étalement de ParB mais sur des distances plus limitées avec des liaisons intra et intermoléculaires formant des boucles et des ponts permettant de construire de grands complexes nucléoprotéiques (**fig. 61 B** ; Graham, 2014). Ils ont réalisé des études de microscopies à molécule unique où l'ADN du bactériophage lambda est fixé par une de ces extrémités à une lame de microscope et étiré par un flux de tampon. Suite à l'introduction de ParB de *B. subtilis* et à sa fixation, l'ADN est compacté à partir de son extrémité libre. Avec un marquage fluorescent de ParB, ils ont montré que ParB peut se fixer sur l'ADN non spécifiquement et qu'il n'y a pas d'enrichissement détecté au niveau de *parS*. De plus cette compaction ne commence seulement qu'après une fixation

suffisante de ParB à l'ADN. L'ADN étiré par le flux a été condensé par ParB d'une manière qui est la plus cohérente avec le pontage de ParB à travers les boucles dans la même ou à travers différentes molécules d'ADN. De plus des mutations dans le patch d'arginine altèrent la formation de boucles de l'ADN *in vitro*. Ainsi l'activité de pontage de l'ADN permet à un nombre limité de molécules ParB de rapprocher des régions de l'ADN distantes de plusieurs kilobases dans un espace tridimensionnel pour former un complexe nucléoprotéique compact (Graham, 2014). La modélisation informatique suggère que ni les interactions 1D ni les interactions 3D ne pouvaient à elles seules générer des foyers de ParB mais qu'une combinaison du « spreading » associé à la formation de ponts en 3D sont nécessaires pour recréer le complexe nucléoprotéique condensé ParB-DNA observé *in vivo* (Broedersz, 2014). La principale mise en garde de ce modèle est que ces expériences n'ont pas pu démontrer une quelconque spécificité de séquence pour *parS* contrairement à ce qui est observé *in vivo*. Ce dernier modèle s'appuie sur des études analysant la liaison de ParB à l'ADN étiré par le flux, dans lesquelles aucun effet de la présence de *parS* n'a été détecté. En effet, ParB de *B. subtilis* peut faire des boucles pour condenser l'ADN *in vitro* indépendamment de la présence de *parS*. Ceci contredit les données *in vivo* montrant que l'absence de sites *parS* empêche la formation de clusters. En effet *parS* est absolument nécessaire pour le regroupement des molécules ParB (Erdmann, 1999). De plus, le patch riche en lysine présent dans le CTD de la protéine ParB de *B. subtilis* n'est pas hautement conservé ; la protéine ParB de *Caulobacter crescentus* ne possède pas ces résidus lysine équivalents et ne peut ni faire des boucles ni condenser l'ADN *in vitro* (Jalal, 2019).

A la même époque, un autre modèle permettant de mieux expliquer le confinement de ParB observé *in vivo* dépendant de *parS* a été énoncé par l'équipe de Jean-Yves Bouet (fig. 62 ; Sanchez, 2015 ; Debaugny, 2018). Son modèle « nucléation et encagement » ou « **Nucleation and caging** », propose qu'un réseau de liaison stochastique de ParB explique le regroupement des molécules de ParB autour de *parS*. Le site *parS* agit comme un centre de nucléation de ParB, tandis que des interactions protéine-protéine et protéine-ADN faibles mais synergiques confinent spatialement ParB dans un volume à l'intérieur des cellules. Pour développer ce modèle, une approche combinant microscopie de haute résolution, modèle mathématiques, études quantitatives à l'échelle du génome et biochimie a été entreprise tout d'abord en utilisant le système de partition de type I du plasmide F d'*Escherichia coli* (Sanchez, 2015) puis

le système de partition des chromosomes ParABS du chromosome principal de *Vibrio cholerae* (Debaugny, 2018).



**Figure 62: Modèles pour l'étalement de ParB et l'assemblage des complexes de partition (II): «Nucleation and caging» (Nucléation et engagement).**

**A)** Interactions de la protéine ParB avec ces partenaires: le dimère ParB a une forte affinité pour le site d'ADN *parS* (flèche épaisse) et une faible affinité (flèche fine) pour l'ADN non spécifique (nsDNA) et pour lui-même.

**B)** Un dimère ParB se lie fortement au site *parS* à haute affinité (à gauche). Les interactions transitoires de ParB avec lui-même et avec ParB-nsDNA fournissent un réseau d'interactions plus faibles qui nucléent la formation d'une zone ParB hautement confinée (milieu). En empêchant la diffusion rapide de ParB en dehors du complexe ParB/*parS*, ces interactions indépendantes mais synergiques regroupent activement la plupart des ParB autour de *parS* (à droite). L'ADN situé à proximité de *parS* pénétrerait de préférence dans cette région à haute densité de ParB. Cela entraîne la liaison stochastique de ParB sur les séquences d'ADN proches du centromère qui dépendent des boucles naturelles de l'ADN. L'assemblage du complexe de partition du plasmide F basé sur le modèle « nucleation and caging » est schématisé entre les lignes pointillées, le cercle **jaune** représentant la « cage ParB ». Le centromère est symbolisé par un rectangle **noir** sur lequel les dimères ParB<sub>F</sub> se fixent avec une haute affinité. Des dimères additionnels ParB sont attirés à proximité de ParB/*parS* par le réseau d'interactions ParB. Les lignes **vertes** et **rouges** représentent l'ADN adjacent à *parS*, respectivement à droite et à gauche. L'assemblage de complexes nucléoprotéiques, tels que les complexes RepE/itéron (hexagones **violet**) sur le côté gauche de *parS*<sub>F</sub> peut réduire la proximité spatiale des régions d'ADN environnantes du complexe de partition. (légende et figure tirées de Sanchez, 2015).

Ils ont regardé la localisation cellulaire de molécules ParB<sub>F</sub> seules dans des cellules vivantes avec un microscope PALM (Photo-Activated Localization Microscopy) puis leurs distributions spatiales à l'aide d'un microscope SMLM (Single Molecule Localization Microscopy) qui permet la détection et la localisation de molécules individuelles à l'échelle du nanomètre (Sanchez, 2015). L'analyse des données d'imagerie ont montré que la liaison de ParB<sub>F</sub> à *parS*

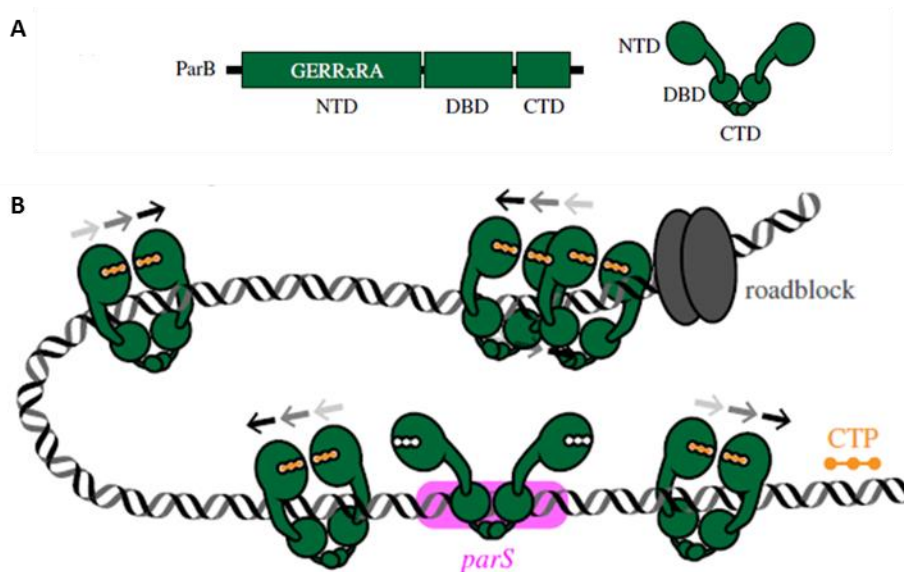
entraîne une concentration locale très élevée de la protéine *in vivo*, où plus de 90 % de ParB<sub>F</sub> dans la cellule sont confinées en cluster au niveau de *parS*. Cette distribution de ParB<sub>F</sub> autour de *parS* est similaire sur chromosome (Debaugny, 2018). Des données de CHIP-seq indiquent que la fixation de ParB<sub>F</sub> est très forte au niveau de *parS* puis diminue fortement. Ainsi ParB<sub>F</sub> présente une forte affinité pour le site *parS* et des affinités faibles pour l'ADN non spécifique ou pour lui-même (**fig. 62 A**). La nucléation de ParB<sub>F</sub> par *parS* crée et maintient une concentration localisée très élevée de ParB dans une "cage" par de nombreuses interactions faibles mais dynamiques avec lui-même (interactions dimère-dimère via les domaines N-terminaux), ainsi qu'avec l'ADN non spécifique autour de *parS*. Ces interactions peuvent se produire de manière stochastique à une fréquence très élevée, en particulier au niveau du point de concentration maximale de ParB *in vivo*. Des expériences de redistribution de fluorescence après photoblanchiment (FRAP) ont montré que les molécules ParB<sub>F</sub> s'échangent très rapidement entre les différents clusters mettant en lumière la nature dynamique des cages ParB-ADN. La modélisation informatique de ParB autour de *parS* ne va pas dans le sens d'une propagation latérale 1D seule ou d'une combinaison de propagation 1D et de pontage 3D, excluant les deux modèles précédents. Ce modèle « **nucleation and caging** » est cohérent avec les données *in vivo* et il a été appliqué au système de partition des chromosomes de *Vibrio cholerae* indiquant que ce mécanisme stochastique d'auto-assemblage est largement conservé des plasmides aux chromosomes (**fig. 62 B** ; Sanchez, 2015 ; Funnell, 2016 ; Debaugny, 2018 ; Jalal, 2020a).

Récemment un nouveau co-facteur de ParB a été découvert : le dimère ParB lie la cytidine triphosphate (CTP) et l'hydrolyse en cytidine diphosphate (CDP) plus un phosphate inorganique (Osorio-Valeriano, 2019 ; Soh, 2019 ; Jalal, 2020b). Une structure co-cristalline montre la liaison de CTP avec le patch d'arginine dans le NTD de la protéine ParB de *B. subtilis*. CTP (ou CDP) est pris en sandwich entre deux NTD, ainsi promouvant une nouvelle interface de dimérisation au niveau du NTD qui n'avait pas été observée auparavant (Osorio-Valeriano, 2019 ; Soh, 2019).

Un nouveau modèle est suggéré : « le glissement latéral de la pince ParB-CTP sur l'ADN » ou « **sliding on DNA** ». Des expériences de liaisons covalentes ou « cross-linking » et d'imagerie de molécule unique ont montré que lorsque le dimère ParB lie la CTP au niveau de son NTD,



une nouvelle dimérisation au niveau du NTD s'effectue et cela crée une sorte de pince qui attrape l'ADN à l'intérieur de sa cavité centrale. La liaison de CTP au dimère ParB fixé sur son site *parS* induirait un changement de conformation du domaine de liaison à l'ADN (DBD) qui rend incompatible la liaison avec *parS*. Ainsi CTP libérerait ParB de son site *parS* permettant un glissement de ParB de son site *parS* à haute affinité vers de l'ADN voisin à faible affinité. CTP servirait probablement de faire passer ParB d'un mode de nucléation à un mode coulissant (fig. 63 ; Osorio-Valeriano, 2019 ; Soh, 2019 ; Jalal, 2020a). Plusieurs arguments viennent étayer ce modèle de pince ParB-CTP, (i) des protéines de liaisons à forte affinité pour l'ADN peuvent bloquer l'étalement du couple ParB-CTP de *B. subtilis* et *C. crescentus* sur l'ADN *in vitro* (fig. 63), (ii) ParB de *C. crescentus* ne s'accumule que sur l'ADN dont les deux extrémités sont bloquées pour éviter un trop grand étalement. L'hydrolyse de CTP par ParB pourrait servir à recycler ParB entre les modes de nucléation et de translocation (Soh, 2019, Jalal, 2020b). Ce modèle « **sliding on DNA** » ne serait pas incompatible avec le modèle « **nucleation and caging** ».



**Figure 63: Modèles pour l'étalement de ParB et l'assemblage des complexes de partition (III):**  
«Sliding on DNA» (glissement sur l'ADN):

**A)** Les protéines chromosomiques ParB sont composées d'un domaine N-terminal (NTD), un domaine central de liaison à l'ADN (DBD) et un domaine C-terminal (CTD). Le NTD héberge un motif riche en arginine conservé (GERRxRA) qui permet les interactions ParB-ParB et ParB-cytidine triphosphate (CTP). Représentation d'un dimère ParB: homodimérisation au niveau de son domaine C-terminal (CTD).

**B)** ParB se propage en glissant sur l'ADN. ParB passe d'une pince ouverte à une pince fermée lors de la liaison au CTP (orange). ParB et *parS* sont colorés en vert et magenta, respectivement. Les flèches au-dessus des complexes ParB-CTP indiquent leur glissement progressif sur l'ADN. Une protéine de liaison à forte affinité pour l'ADN (en gris) peut bloquer de manière unidirectionnelle le glissement de ParB sur l'ADN.

(légende et figure tirées de Jalal, 2020).

- **Une protéine motrice ATPase de type Walker A : la protéine ParA.**

L'ATPase ParA est le troisième élément du système de partition ParABS, de type I, qui permet la ségrégation des plasmides à faible nombre de copies et de la plupart des chromosomes bactériens. C'est l'élément moteur du système en fournissant l'énergie pour permettre la ségrégation des réplicons, en fixant et hydrolysant l'ATP.

ParA est une ATPase qui fait partie de la grande superfamille des GTPases à boucle P (P-loop) qui comprend deux classes, TRAFAC et SIMIBI. ParA fait partie de cette dernière qui fonctionne sous forme de dimères (Leipe, 2002). La caractéristique principale de cette ATPase est qu'elle contient un motif « atypique Walker A » : [KGGxxGKT], contenant deux lysines conservées (Motallebi-Veshareh, 1990 ; Lutkenhaus, 2012). La seconde lysine est commune à tous les motifs Walker A et elle est impliquée dans la liaison et l'hydrolyse de l'ATP tandis que la première est appelée « la signature lysine » pour la classe ParA des ATPases. Elle participe à la dimérisation de ParA en se liant au phosphate de l'ATP lié à une autre sous-unité. Cette lysine est essentielle pour l'hydrolyse de l'ATP et elle est l'équivalent fonctionnel du doigt d'arginine trouvé dans les protéines GAP qui activent les activités GTPases des protéines Rho et Ras (Lutkenhaus, 2012).

Les premières études sur la ségrégation des plasmides et des chromosomes ont proposé un mécanisme de poussée/traction de l'ADN par un filament linéaire ou hélicoïdal de ParA, semblable à l'appareil du fuseau mitotique chez les eucaryotes. Selon ce modèle, ParA liée à l'ATP se polymériserait en une structure filamenteuse le long de la cellule, avec l'extrémité du filament capturant le complexe nucléoprotéique ParB-ADN. ParB se lie à ParA et stimule son activité ATPase pour hydrolyser l'ATP, dépolymérisant ainsi le filament de ParA et entraînant simultanément le complexe ParB-ADN (donc le plasmide/chromosome) le long du filament rétracté vers le pôle cellulaire opposé (Fogel, 2006 ; Ptacin, 2010). De telles observations ont conduit à rechercher un système de traction basé sur un filament actif, à travers les pôles. Malgré la capacité de polymérisation de ParA *in vitro* (Leonard, 2005), aucun filament de ce type n'a été trouvé *in vivo* (Le Gall, 2016). Au lieu de ça, ParA a été montrée se lier à l'ADN chromosomal de façon non spécifique ce qui apparaît être une étape nécessaire dans son interaction avec ParB (Bouet, 2007 ; Castaing, 2008). De plus ParB lie et hydrolyse la CTP, *parS* dépendante qui à son tour est essentiel pour l'interaction ParB-ParA.

Les récents progrès réalisés dans les expériences de reconstitution *in vitro* (Hwang, 2013 ; Vecchiarelli, 2013, 2014), parallèlement aux approches de microscopie à super-résolution (Lim, 2014 ; Le Gall, 2016), ont fourni de nouvelles informations sur le fonctionnement interne de la partition médiée par ParA qui remettent en question les modèles établis basés sur les filaments. Collectivement, ces expériences montrent que, contrairement à la mitose eucaryote, la machinerie de partition de type ParA conduit le mouvement dirigé de l'ADN sans la formation de structures filamenteuses.

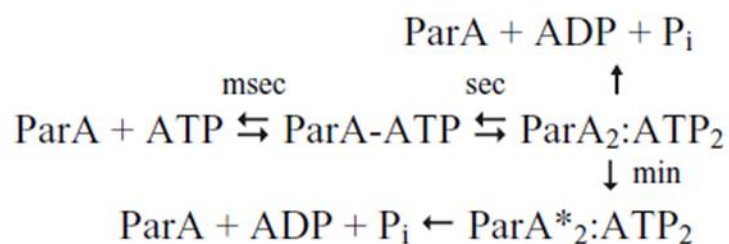
Des études récentes ont convergé vers un nouveau modèle dans lequel la partition de type ParA est déterminée par une motilité dirigée par un mouvement « brownien à cliquet ».

En quoi consiste ce concept de « cliquet Brownien » ou « Brownian ratchet » : considérons une particule diffusant dans une dimension divisée en intervalle, où chaque frontière entre deux intervalles est comme un cliquet. Une particule peut passer librement à travers une frontière venant de la gauche, mais elle ne peut pas revenir en arrière une fois passée cette frontière. A cause de l'irréversibilité imposée par le cliquet, la particule diffusible subit un mouvement dirigé vers la droite même si la particule, intrinsèquement diffuse dans toutes les directions. Ce mouvement dirigé nécessite certainement un apport d'énergie ; mais plutôt que de fournir directement l'énergie pour le mouvement, l'apport d'énergie peut servir à établir les cliquets et ainsi rectifier le mouvement brownien pour produire un mouvement biaisé (Hu, 2017). Il est important de noter que ce concept de « cliquet Brownien » ne va pas à l'encontre des lois de la thermodynamique mais fournit un mécanisme alternatif pour générer un mouvement dirigé communément trouvé dans, ou proposé pour, plusieurs protéines motrices (Hu, 2017). Deux conditions sont requises pour ce « cliquetage Brownian » : un environnement asymétrique qui nécessite de l'énergie et impose la direction du mouvement, et une diffusion aléatoire de l'objet qui "surfe" sur cette asymétrie, perpétuant le mouvement dirigé (Hu, 2017).

Trois modèles « de cliquet Brownien » médiés par parA sont proposés et tous partagent la même biochimie du système ParA ; mais ils diffèrent dans les aspects mécaniques (Hu, 2017) : le modèle « de « Diffusion-ratchet », le modèle de « DNA-relay » ou « relais de l'ADN » et le modèle de « Hitch-Hiking » ou « Auto-stop ».

- **le modèle « diffusion-ratchet »** : ce modèle a été proposé par l'équipe de Mizuuchi à la suite de travaux notamment de Ling Hwang et Anthony Vecchiarelli (Vecchiarelli, 2010, 2013, 2014 ;

Hwang, 2013). Mizuuchi et ses collègues ont tout d'abord regardé comment l'ATP promeut et régule l'activité de liaison à l'ADN de l'ATPase ParA. Par des analyses en microscopie TIRFM (Total Internal Reflection Fluorescence Micoscopy) ils ont regardé l'interaction entre l'ADN et une version fluorescente de ParA. Ils ont montré que ParA se lie à l'ADN non-spécifique (ADNns) en présence de ATP sous forme dimérique. Cette liaison à l'ADN ne nécessite pas une hydrolyse de l'ATP mais la fixation de l'ATP par ParA induit une lente transition conformationnelle en plusieurs étapes de ParA qui permet à ParA de se fixer à l'ADNns. La cinétique fournit un délai temporel pour permettre un lent cycle entre les formes liées et non liées de ParA. La première transition (2-50s) correspond la fixation de l'ATP qui induit un changement de conformation. La deuxième transition, plus longue (10min), permettrait à ParA-ATP de passer dans un état stable ce qui induirait un délai temporel pour ParA avant de pouvoir se refixer sur l'ADN (Vecchiarelli, 2010). Un mutant ParA qui fixe l'ATP mais qui ne peut pas subir de changement conformationnel, ne peut pas se fixer à l'ADN et est déficient pour la partition, *in vivo*.

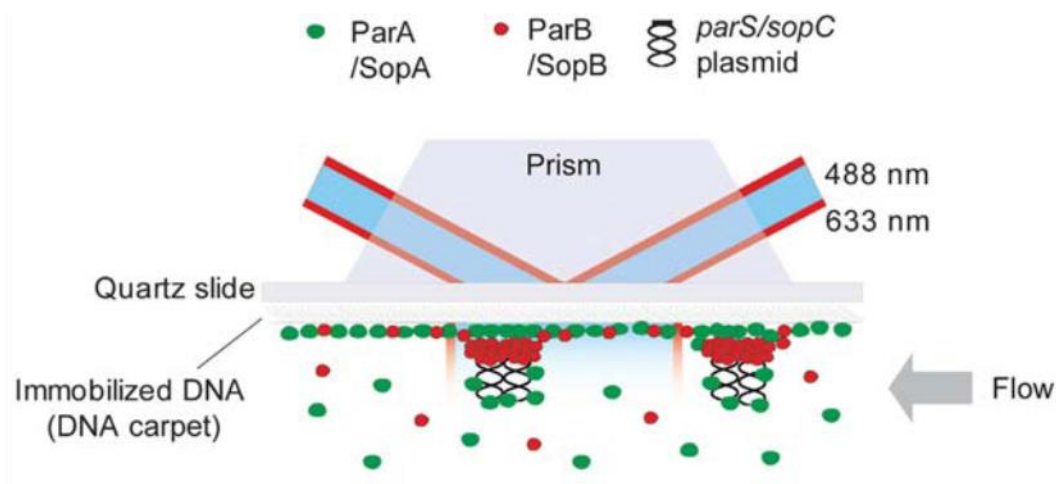


Mécanisme cinétique qui illustre cette transition

ParA\*<sub>2</sub>:ATP<sub>2</sub> = intermédiaire stable de ParA lié à l'ATP sous-forme dimère, à la suite d'un lent changement de conformation, ATP spécifique, et qui précède l'hydrolyse.

Ainsi ce délai temporel, combiné avec la stimulation de l'activité ATPase de ParA par ParB liée au plasmide ou ADN, génère une distribution inégale de ParA associée au nucléoïde et fournit une force motrice pour la ségrégation. Le nucléoïde servirait de matrice sur laquelle les plasmides se déplaceraient sous l'action de ParA et ParB (Vecchiarelli, 2010).

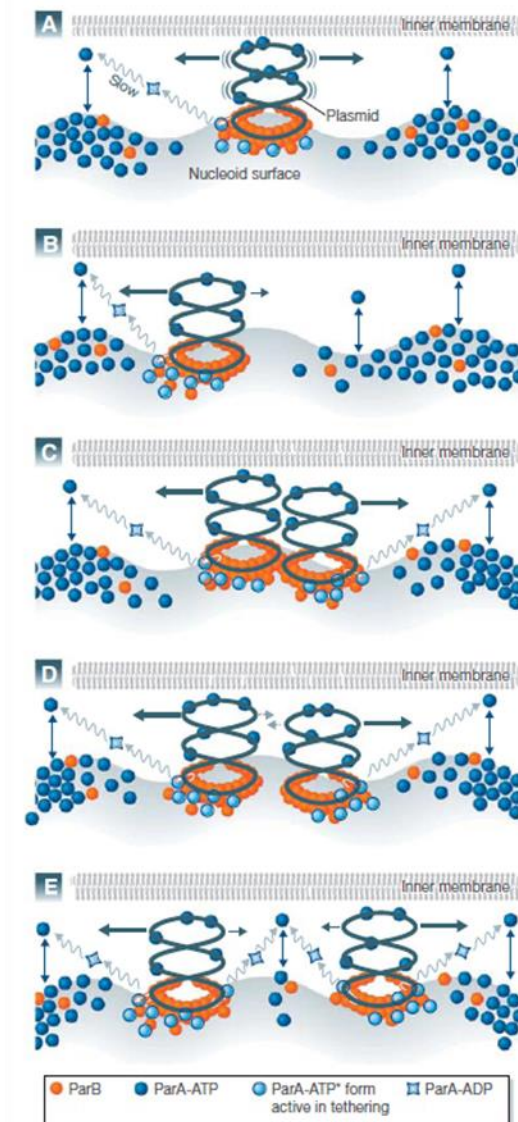
Pour étudier les interactions entre ParA et les autres intervenants du système de partition, Hwang et ces collègues ont reconstitué, *in vitro*, un système de partition du plasmide P1 médié par ParA à l'intérieur d'une cellule à écoulement recouverte d'ADN, qui agit comme un nucléoïde artificiel (**fig. 64**). La dynamique du système a été visualisée à l'aide de la TIRFM qui permet d'illuminer seulement les molécules qui sont associées au tapis d'ADN (Hwang, 2013). Ainsi ParB, liée aux séquences *parS* du plasmide, interagit avec ParA-ATP fixée non



**Figure 64: Schéma de la cellule d'écoulement utilisée pour reconstituer le système de partition du plasmide P1, *in vitro*.**

La surface intérieure de la cellule d'écoulement a été recouverte d'ADNs soniqué pour imiter la surface du nucléoïde (tapis d'ADN). Des composants du système Par marqués par fluorescence ont été introduits dans la cellule à flux et la dynamique du système a été visualisée à l'aide de la TIRFM à prisme bicolore. (légende et figure tirées de Hwang, 2013).

spécifiquement au tapis d'ADN. ParB stimule l'activité ATPase de ParA qui hydrolyse l'ATP, via un motif « doigt d'arginine » (ou arginine finger ; Ah-Seng, 2009) ce qui entraîne la dissociation du dimère ParA en monomères individuels qui ne se lient plus au tapis d'ADN, ce qui conduit au détachement du plasmide. La stimulation de l'activité ATPase crée un gradient local de ParA-ATP avec moins de ParA-ATP liée au tapis d'ADN près du complexe ParB-plasmide. Ce mouvement brownien accélère l'élimination de ParA du tapis d'ADN environnant, créant une zone de déplétion de ParA. Les séries de transitions dépendant de l'ATP retarde la réactivation de ParA pour se fixer à l'ADN. Ce délai temporel, observé par Vecchiarelli en 2010, aurait pour conséquence la diffusion de ParA et défavoriserait la fixation de ParA-ATP au même endroit. Le mouvement initial du complexe ParB-ADN dans une direction choisie renforce le mouvement continu dans la même direction, ce qui entraîne un mouvement directionnel à longue distance de l'ADN. Les travaux de Mizuuchi et de ces collègues soutiennent un modèle de diffusion-ratchet, où ParB fixée au plasmide repousse et redistribue le gradient de ParA sur le nucléoïde, qui à son tour permet de mobiliser le plasmide (**fig. 65** ; Vecchiarelli, 2010, 2013, 2014 ; Hwang, 2013).



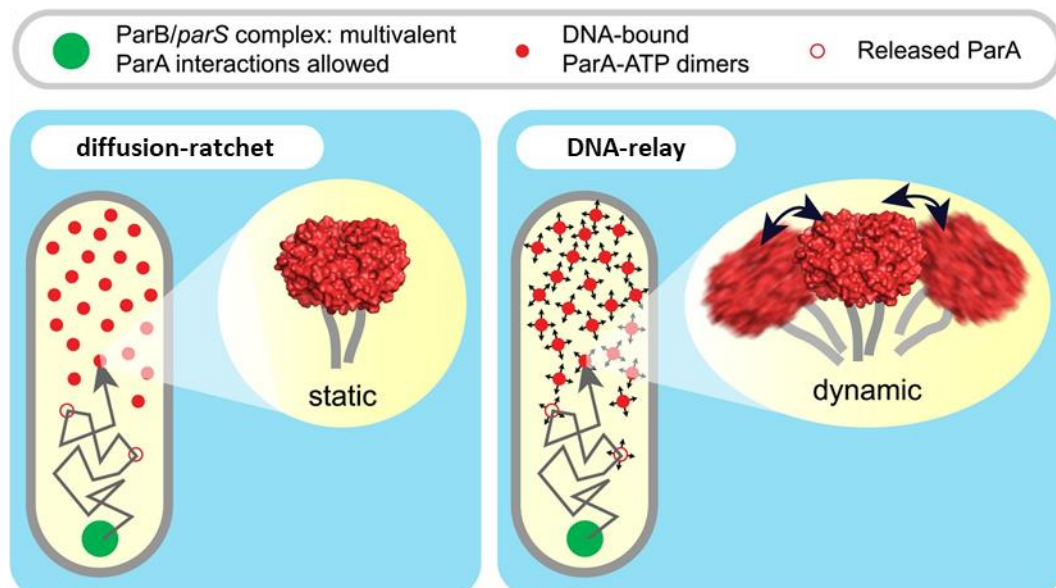
**Figure 65: Schéma du modèle de « diffusion-ratched » .**

ParA-ATP subit un lent changement de conformation après la liaison avec l'ATP (cercles **bleus pleins**) et s'associe ensuite de manière non spécifique au nucléoïde et au plasmide. L'interaction des dimères ParB (cercles **orange**) liés aux sites parS avec de multiples molécules adjacentes de ParA-ATP\* (cercles **bleu pâle**: forme de ParA-ATP attachée) attache les complexes de partition plasmidique au tapis nucléoïde. L'hydrolyse stimulée par ParB des molécules ParA-ATP attachées libère Par-ADP (**étoiles bleues pâles**) et en résulte une perte progressive des points d'ancrage. Cela conduit à la formation d'une zone de déplétion en ParA-ATP sur le nucléoïde. Le plasmide se déplace jusqu'à la limite de la zone de déplétion pour continuer à s'associer à ParA-ATP (**A et B**).

Après la réplication et la décaténation, les molécules filles sont libres de diffuser indépendamment et de s'associer indépendamment avec le gradient ParA-ATP aux bords distaux de la déplétion (**C à E**). L'état d'équilibre est atteint lorsque deux sœurs sont aux quarts de la position du nucléoïde. La lente reformation de la forme ParA-ADP à ParA-ATP qui peut de nouveau s'associer au nucléoïde fournit l'instabilité dynamique qui sous-tend la diffusion en cliquet.

(légende et figure tirées de Sherratt, 2013).

- le modèle « DNA-relay » : combinant des approches biochimiques, d'imageries quantitatives, et de modélisations mathématiques chez *Caulobacter crescentus*, Lim et ses collaborateurs (Lim, 2014) ont présenté une évolution du modèle « diffusion-ratchet » où ils ont introduit une notion « d'élasticité et de mouvement dynamique de l'ADN » qui aide ou dirige la translocation du chromosome sur de courtes distances. C'est le modèle « DNA-relay » (fig. 66 ; Lim, 2014). Sur la base de la modélisation informatique, il a été avancé que la diffusion à courte distance d'un complexe ParB-ADN le long du gradient de ParA-ATP pourrait ne pas être suffisante pour expliquer une ségrégation unidirectionnelle robuste du chromosome vers le nouveau pôle cellulaire et que le modèle de diffusion à cliquet devrait être étendu pour incorporer une composante d'élasticité de l'ADN (Lim, 2014 ; Jalal, 2020).



**Figure 66: Schéma du modèle de « DNA-relay » .**

Le modèle « DNA-relay » entraîne une translocation robuste du complexe ParB/*parS*.

A gauche, représentation graphique du modèle " diffusion-ratchet " : le complexe ParB/*parS* diffuse (disque **vert**) et interagit avec les dimères ParA-ATP immobilisés dans l'ADN (disques **rouges remplis**), stimule l'activité ATPase, ce qui entraîne la dissociation de ParA de l'ADN (disques **rouges ouverts**).

A droite, , représentation graphique du modèle " DNA-relay " : identique au modèle " diffusion-ratchet " sauf les dimères ParA-ATP, liés à ParB, fluctuent de façon aléatoire en fonction du mouvement de l'ADN associé. Lorsque le complexe ParB/*parS* est associé à un ou plusieurs dimères ParA-ATP liés à l'ADN, il subit la force élastique qui régit la dynamique des loci d'ADN associés aux dimères ParA.

(légende et figure tirées de Lim, 2014).

Le suivi d'un locus chromosomique en « time-lapse » montre que sa localisation est dynamique, qu'il bouge autour de son point d'équilibre au cours du temps, impliquant une élasticité de l'ADN. Dans ce modèle « DNA-relay », les dimères ParA-ATP liés à un locus d'ADN

fluctuent autour du point d'équilibre de ce locus. Lorsque le système de partition s'attache à un dimère de ParA-ATP qui est étiré par rapport à son point d'équilibre, cela entraîne une force élastique qui déplace le complexe vers le point d'équilibre du locus sur lequel est fixé ParA jusqu'à ce que l'hydrolyse de l'ATP le détache. Dans ce modèle, les complexes ParA-ATP liés à l'ADN peuvent exploiter la dynamique élastique du chromosome pour relayer le complexe de partition sur une longue distance d'un emplacement de l'ADN à un autre (Lim, 2014 ; Jalal, 2020).

- **le modèle « hitch-hiking »** : les modèles précédents proposaient que les composants du complexe de partition ségrégent à travers l'espace entre la surface du nucléoïde et la membrane. Ces différents modèles proposent des schémas de localisation distincts des composants Par à l'intérieur ou autour du nucléoïde. Pour détecter directement la localisation 3D des complexes de partition du plasmide F ou du chromosome de *B. subtilis* à l'intérieur du volume du nucléoïde, Le Gall et ses collaborateurs (Le Gall, 2016) ont combiné un marquage à haute densité du chromosome entier avec de la microscopie à illumination structurée (3D-SIM). La localisation des complexes de partition est détectée par un marquage fluorescent de la protéine ParB. Les complexes de partition sont confinés à l'intérieur du nucléoïde bactérien que ce soit pour les systèmes plasmidiques ou chromosomiques (Le Gall, 2016), consistant avec les premières observations à partir de ParB de *B. subtilis* (Marbouty, 2015). Les complexes de partition plasmidiques et chromosomiques sont positionnés le long de l'axe longitudinal. Les complexes de ségrégation du plasmide F, ParBS<sub>F</sub> sont localisés dans le quart du nucléoïde, tandis que pour le chromosome ParBS<sub>Bsu</sub>, ils sont positionnés près des bords du nucléoïde. Utilisant la microscopie 3D-SIM, la protéine ParA est également distribuée à l'intérieur du nucléoïde, en patch dans un gradient asymétrique le long de l'axe longitudinal avec une plus grande densité à un pôle de la cellule (Le Gall, 2016). En utilisant une combinaison de mutants pour ParA qui affectent la liaison à l'ADN ou l'activité d'hydrolyse de l'ATP, il s'est avéré que la localisation volumétrique des complexes de partition à l'intérieur du nucléoïde nécessite la fixation de ParA à l'ADNs aussi bien que la stimulation de l'activité ATPase de ParA par ParB. ParA détermine la localisation 3D des complexes de partition à l'intérieur du nucléoïde (Le Gall, 2016). En 2015, une nouvelle caractéristique structurale du nucléoïde bactérien de *B. subtilis* a été décrite, les HDRs pour High Density chromosomal Regions : ces HDRs représentent des régions du nucléoïde qui contiennent une quantité



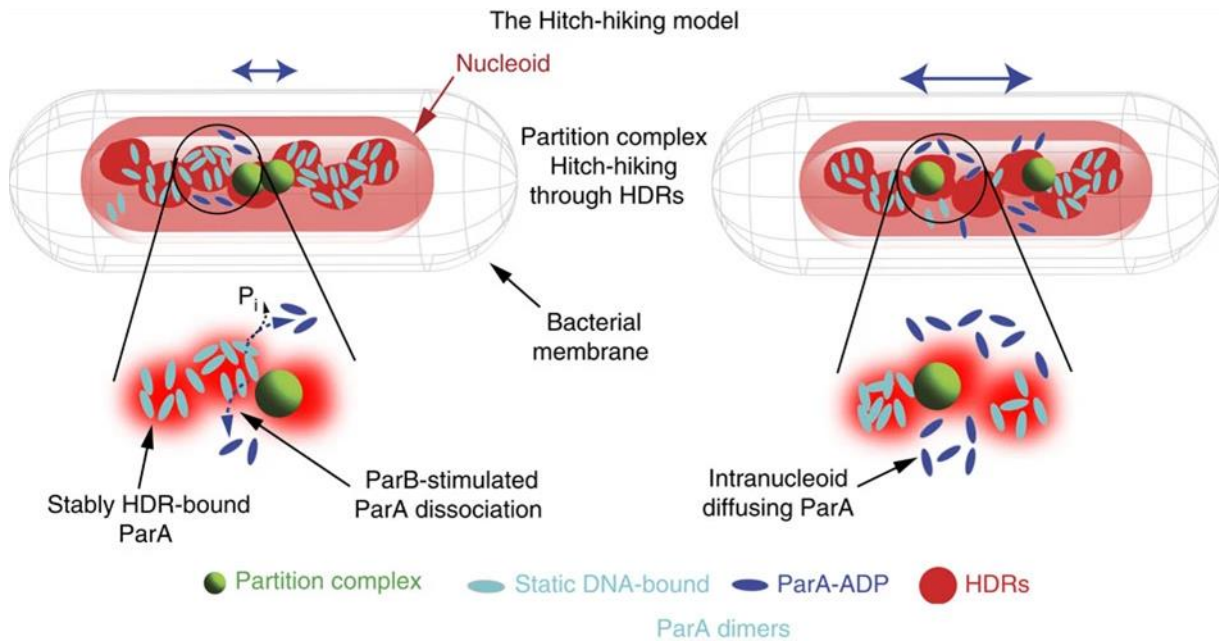
d'ADN plus dense et où l'ADN est plus fortement compacté probablement en raison de l'association stochastique et dynamique des domaines d'interaction chromosomique. Ces HDRs ont également été caractérisées pour le chromosome de *E. coli* (Marbouty, 2015). Des analyses utilisant la 3D-SIM deux couleurs, ont montré que ParB et les HDRs ont un degré de colocalisation ( $63 \pm 36\%$ ) plus grand que le hasard (20%) suggérant que les complexes de partition peuvent interagir avec les HDRs pendant la ségrégation et persistent à travers le cycle cellulaire. De même, ParA et les HDRs présentent un large degré de corrélation de position ( $82 \pm 24\%$ ) même plus élevé que celui avec le complexe de partition (Marbouty, 2015 ; Le Gall, 2016). Le mouvement dynamique d'une molécule unique de ParA a été suivi dans des cellules vivantes par spt-PALM (single-particle tracking PALM) qui a révélé deux états : une population de ParA qui explore de larges régions du nucléoïde et une population de ParA statique, localisée au niveau des patchs. Le nombre de patchs statiques ParA est similaire au nombre de patchs ParA colocalisants avec les HDRs, observés par 3D-SIM. Cette localisation de ParA aux HDRs est importante pour la localisation volumétrique et une ségrégation correcte des complexes de partition (Le Gall, 2016).

Pour résumer les résultats des travaux de l'équipe de Marcello Nollmann : tous les composants de la machinerie de partition sont localisés à l'intérieur du nucléoïde pendant la ségrégation. La liaison à l'ADN et l'activité ATPase de la protéine de transport ParA sont essentielles pour cette localisation. Ces résultats appellent à une réévaluation de plusieurs aspects des modèles précédemment proposés.

A partir de ces données ils présentent un nouveau mécanisme : le modèle « hitch-hiking » (**fig. 67**) :

- (i) ParA-ATP s'assemble en petits patchs dans les régions à forte densité d'ADN (HDRs),
- (ii) les complexes de partition ParB-*parS* stimulent fortement l'activité ATPase de ParA fixée sur l'ADNs conduisant à sa dissociation de l'ADN (voir modèle « diffusion-ratchet » ; Vecchiarelli, 2010 ; 2014). ParB, présent dans les complexes de partition, interagit avec ParA liée à l'ADNs dans les HDRs entraînant leur dissociation progressive et la libération de ParA-ADP qui diffuse rapidement dans la cellule. ParA peut de nouveau fixer l'ATP, se dimériser et se lier à l'ADNs (**fig. 67**, schéma de droite).
- (iii) une fois le complexe de partition ParB-*parS* a épuisé ParA dans son voisinage, il peut diffuser vers la HDR le plus proche enrichie en ParA, se déplaçant ainsi le long de la distribution hétérogène de ParA, d'HDRs en HDRs, permettant la ségrégation des réplicons de la bactérie.

Le mouvement des complexes de partition est composé d'une composante de diffusion brownienne et d'un biais directionnel, les constituants essentiels d'un mouvement de brownien-ratchet.



**Figure 67: Schéma du modèle de « Hitch-hiking » pour la ségrégation du chromosome bactérien.**

Le contour de la cellule est indiquée par un filet gris, le nucléoïde par un cylindre rouge, le complexe de partition ParBS par une sphère verte. Le dimère ParA est soit dans un état stable lié à l'ADN, représenté en cyan, soit libre lié à l'ADP, Par-ADP représenté en bleu. Les régions du nucléoïde avec un contenu en ADN plus dense, ou HDR, sont représenté par des cercles rouges diffus.

Le dimère ParA-ATP (cyan) est fixé sur l'ADNns au niveau des HDRs. Le complexe de partition ParB-parS (vert) va activer l'activité ATPase de ParA par la liaison de ParB avec ParA ce qui va entraîner la dissociation de ParA de l'ADNns et la diffusion de ParA-ADP (bleu) dans la cellule. Le délai temporel pendant lequel ParA ne peut pas fixer de nouveau de l'ATP va entraîner une zone de déplétion en ParA dans l'HDR et le complexe de partition ParB-parS qui n'est plus retenu au nucléoïde va diffuser vers l'HDR la plus proche enrichi en ParA et de nouveau interagir avec ParA-ATP et se fixer au nucléoïde. Ainsi le complexe de partition va progresser le long de l'axe longitudinal de la bactérie en « sautant » d'HDR en HDR.

Après duplication des origines de réplication, chaque complexe va migrer à un pôle de la cellule sans possibilité de revenir en arrière en raison de l'effet « Brownien de diffusion à cliquet » . (figure tirée de Le Gall, 2016).

### 3-2-2-2 Ségrégation de l'ensemble du chromosome.

A l'inverse de la ségrégation active, rapide et dirigée des origines de réplication, médiée par un système spécifique, la ségrégation du « corps » du chromosome est un processus lent et progressif qui suit la réplication (Nielsen, 2006a, 2006b) et ne nécessite probablement pas de mécanisme dédié. La ségrégation des origines de réplication via le système ParABS pourrait

entraîner le reste du chromosome de chaque côté de la cellule. Les deux copies du chromosome seraient ainsi ségréguées de façon concomitante à la réplication.

Deux études suggèrent que la ségrégation pourrait résulter en grande partie de forces entropiques, arguant que, si le chromosome est un polymère auto-évitant, la maximisation de l'entropie va intrinsèquement séparer les deux chromosomes (Jun, 2006, 2010). Cependant, l'entropie seule ne peut pas expliquer la vitesse de ségrégation des chromosomes en masse chez *E. coli*.

Des études menées chez *E. coli* ont montré que les loci frères répliqués restent ensemble pendant environ 7 à 10 minutes avant d'être rapidement séparés (Joshi, 2011). De plus, les chromosomes frères forment probablement des précaténanes, des structures dans lesquelles l'ADN des chromosomes frères s'enchevêtre topologiquement (Joshi, 2011, 2013 ; Wang, 2008a). Ainsi, les régions récemment répliquées ne peuvent pas se séparer tant qu'elles ne sont pas démêlées via la topoisomérase IV. Une fois libéré des contraintes topologiques et des attaches à base de protéines, l'ADN dupliqué se déplace dans les deux sens, produisant probablement un état plus détendu dans les chromosomes frères naissants.

Bien que les forces purement physiques jouent un rôle majeur, voire dominant, dans la ségrégation des chromosomes, les NAP et d'autres protéines associées aux chromosomes y contribuent probablement aussi (Junier, 2014). Par exemple, les NAP qui compactent l'ADN, comme les protéines HU et IHF, facilitent probablement la ségrégation des chromosomes récemment dupliqués vers les côtés opposés des cellules (Hong, 2011 ; Swiercz, 2013). Des mutants des protéines HU et IHF présentent des défauts de ségrégation avec une augmentation de cellules anucléées (Huisman, 1989, Kaidow, 1995). De même des mutants de la gyrase, de la topoisomérase I ou IV, enzymes impliquées dans le super-enroulement négatif, peuvent entraîner ces mêmes défauts de ségrégation. La Topoisomérase IV, qui résout les précaténanes entre les chromosomes frères, pourrait être particulièrement critique. Lorsque son activité est perturbée, les cellules peuvent achever la réplication des chromosomes, mais ils restent souvent colocalisés (Wang, 2008a).

Un autre acteur clé de la ségrégation chromosomique est le complexe SMC/ScpA/ScpB, ou le complexe apparenté MukB/MukE/MukF que l'on retrouve chez *E. coli* et d'autres  $\gamma$ -protéobactéries (Britton, 1998 ; Danilova, 2007 ; Jensen, 1999 ; Niki, 1991). Chez *E. coli*, l'absence de MukB entraîne la perte de l'orientation gauche-ori-droite des chromosomes et conduit à une augmentation de la formation de cellules anucléées, ce qui indique que MukBEF

peut favoriser une ségrégation chromosomique correcte. Bien qu'il ne soit pas très clair comment MukBEF intervient au niveau de ce phénomène, des études ont montré qu'il stimule directement la topo IV, indiquant que MukBEF contribue aux démêlages des chromosomes frères (Hayama, 2010 ; Li, 2010 ; Nicolas, 2014).

Chez *B. subtilis* les protéines SMC pourraient favoriser la ségrégation des chromosomes, car les cellules dépourvues de SMC présentent une série de défauts de partition des chromosomes (Britton, 1998 ; Gruber, 2009 ; Sullivan, 2009), notamment au niveau de la ségrégation des origines (Gruber, 2014 ; Wang, 2014a, 2014b). Comme chez *E. coli*, la SMC est recrutée dans les régions du chromosome proches de l'origine par une interaction directe avec ParB. Cependant, contrairement à la protéine MukB d'*E. coli*, SMC ne semble pas fonctionner en favorisant l'activité de la Topo IV (Wang, 2014b) mais serait plutôt impliquer pour condenser l'ADN chromosomique, aidant les chromosomes frères à éviter de s'enchevêtrer et augmentant l'efficacité globale de la ségrégation chromosomique.

### **3-2-2-3 Terminaison de la ségrégation.**

Même si la ségrégation des origines de réplication entraîne l'ensemble du chromosome, la ségrégation finale des extrémités des chromosomes nécessite un mécanisme dédié, pour éviter notamment que deux copies répliquées restent entremêlées en raison d'un surenroulement généré par la réplication ou que des chromosomes dimériques se forment à la suite de recombinaisons entre chromosomes frères (Adams, 1992 ; Peter, 1998 ; Steiner, 1998).

L'ADN translocase FtsK est un composant majeur de l'appareil de ségrégation. Elle se lie et stimule la topoisomérase IV pour décaténiser les chromosomes (Espeli, 2003).

De plus FtsK active directement la recombinase à tyrosine XerCD qui résout les problèmes de chromosomes dimères (Grainge, 2007 ; Steiner, 1999). FtsK reconnaît de courts motifs de 8bp appelés KOPS (FtsK Orienting Polar Sequence) qui sont surreprésentés dans le chromosome et dont l'orientation est orientée et guident FtsK vers des sites *dif* (deletion induced filamentation) où FtsK active la recombinase à tyrosine XerCD (Bigot, 2005 ; Löwe, 2008 ; Sivanathan, 2006). XerCD ouvre l'ADN au niveau d'un site *dif*. La recombinaison site-spécifique entre deux loci est activée par une interaction entre FtsK et XerD en activant son activité catalytique (Bigot, 2005).

L'activité de la translocase FtsK n'est pas essentielle chez *E. coli*, vraisemblablement parce que les chromosomes dimères ne sont produits que dans ~15% des cellules par cycle de réplication, mais en absence des protéines MukBEF, *ftsK* est essentiel à la vitalité. Comme ces protéines MukBEF sont une composante majeure de la ségrégation des chromosomes, Ftsk devrait contribuer à la ségrégation des chromosomes en favorisant la décaténation et la résolution des chromosomes (Badrinarayanan, 2015).

Chez la plupart des bactéries, la cytokinèse est activement retardée jusqu'à ce que les chromosomes frères soient complètement séparés vers les côtés opposés de la cellule, empêchant la guillotine de l'ADN. Chez *E. coli*, le mécanisme responsable, appelé occlusion du nucléoïde, implique une protéine appelée SlmA qui se lie à des sites spécifiques de l'ADN qui sont enrichis dans la région terminale-proximale du chromosome. SlmA se lie également et bloque la polymérisation de FtsZ (Bernhardt, 2005). Par conséquent, SlmA bloque la division cellulaire jusqu'à ce que les régions terminales-proximales du chromosome aient été séparées du milieu de la cellule (Bernhardt, 2005).

Chez *B. subtilis*, un mécanisme similaire se produit, impliquant la protéine non apparentée Noc, qui se lie à des séquences d'ADN à travers le chromosome. Noc ne cible pas spécifiquement FtsZ ou d'autres protéines de division cellulaire ; au lieu de cela, elle semble former de grands complexes nucléoprotéiques qui bloquent physiquement l'appareil de division (Adams, 2015 ; Wu, 2004,2009).

## **4- Le système ANCHOR.**

### **4-1 Principe et applications du système de visualisation de locus génomique en temps réel : le système ANCHOR.**

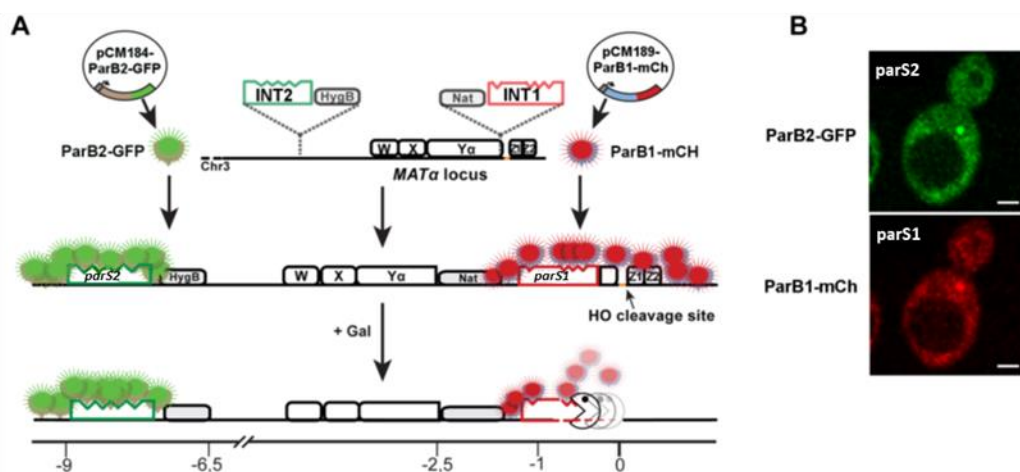
A partir du système de partition des génomes bactériens de type I, le système ParABS, l'équipe de Kerstin Bystricky a développé un nouveau système de marquage non intrusif de locus génomique en temps réel, le système ANCHOR. Ce système est basé sur les propriétés de fixation de la protéine adaptatrice ParB sur les séquences centromériques, *parS*. Les protéines ParB (appelée OR dans le système ANCHOR) se fixent sous forme de dimères sur des séquences palindromiques de 16bp, contenues dans les séquences *parS*, constituant une unité ANCH, entre 300 et 1000 bp. Elles recrutent, par des interactions protéine-protéine, d'autres

dimères ParB qui se lient de manière non spécifique à l'ADN adjacent pour former un complexe nucléoprotéine. Bien que le mécanisme d'assemblage du complexe nucléoprotéique ParB-ADN ne soit pas complètement explicité (voir ci-dessus § 3-2-2-1b : le système ParABS), il faut retenir que la protéine ParB (ou OR) se fixe fortement sur des séquences spécifiques *parS* (ANCH), qu'elle s'étend avec des interactions faibles que ce soit ParB-ADNs ou ParB-ParB créant un environnement autour des séquences *parS* permettant de recruter suffisamment de protéines ParB pour créer un foyer de fluorescence (Audibert, 2020). Le principe du système ANCHOR est basé sur l'oligomérisation de protéines de liaison à l'ADN sur des séquences spécifiques plutôt que sur une répétition de séquences d'ADN comme sont définis les autres systèmes de visualisation de loci génomiques (FROS, TALEN, CRISPR/Cas9 ...). Les composants du système ANCHOR développé par l'équipe de Kerstin Bystricky sont dérivés du système de partition des réplicons de l'espèce multi-chromosomales *Burkholderia cenocepacia*, pour les chromosomes c2 et c3 (**tableau 2** ; Saad, 2014).

Au début des années 2000, le système ParB-*parS* avait déjà été utilisé pour visualiser des régions spécifiques d'ADN chromosomique dans des bactéries vivantes (Li, 2002a, b ; Nielsen, 2006a, b ; Youngren, 2014). Il était basé sur le système de partition du plasmide P1 d'*Escherichia coli* ou sur celui du plasmide de virulence pMT1 de *Yersinia Pestis*. Ils ont été utilisés, notamment, dans des études pour déterminer la dynamique de ségrégation des chromosomes chez *E. coli* (Li, 2002a, b ; Nielsen, 2006a, b ; Youngren, 2014). Mais le facteur bactérien « Integration Host Factor » (IHF) est nécessaire que la protéine ParB forme un complexe effectif sur sa séquence *parS* spécifique. De plus une protéine ParB délétée de ses 30 premiers résidus devait être utilisée pour éviter qu'elle interagisse avec la protéine ParA. Sous ces conditions, le mécanisme qui dirige le mouvement des plasmides est absent. L'utilisation était restreinte à *E. coli* qui ne possède pas de système de partition chromosomique (Wang, 2008b ; Germier, 2018). L'originalité du système ANCHOR est qu'il repose sur le système de partition ParABS qui ne nécessite pas de présence de facteurs bactériens et peut être utilisé dans tous les organismes vivants.

Ce système a tout d'abord été validé chez la levure par l'équipe Kerstin Bystricky qui travaille sur la réparation des cassures double brin de l'ADN ou DSB (**fig. 68** ; Saad, 2014). Le mécanisme de réparation le plus précis est basé sur la recombinaison homologue (HR) dans laquelle les

simples brins générés à côté de la cassure recherchent une réplique intacte qui est copiée dans le site cassé. Les changements dans la dynamique des chromosomes au cours des premières étapes qui créent les brins uniques n'ont pas été analysés en raison du manque d'outils permettant l'analyse de ce processus dans des cellules vivantes individuelles. Les systèmes fluorescents opérateur/répresseur (FROS) utilisables pour marquer les loci chez les eucaryotes ne conviennent pas pour cette étude. En effet la nature répétitive et la grande taille des séquences de fixation de l'opérateur peuvent interférer et/ou perturber les processus d'ADN à courte distance et provoquer des événements de recombinaison préjudiciables. En outre, les répresseurs LacI et TetR étroitement liés peuvent créer des sites fragiles et constituer une barrière (Possoz, 2006 ; Dubarry, 2011 ; Jacomme, 2011 ; Sofueva, 2011).



**Figure 68: Système de visualisation de loci ParB-*parS* chez la levure.**

**A)** Les sites *parS1* et *parS2* sont insérés respectivement à 76bp et 3,4kb en amont du site de coupure HO dans le locus *MAT* sur le chromosome III de la levure. ParB1-mCherry et ParB2-GFP produites après l'ajout de doxycycline se lient d'abord à leurs sites spécifiques puis, par interaction protéine-protéine et liaison non spécifique à l'ADN, aux séquences flanquantes, créant ainsi des foyers visibles. L'ajout de galactose induit l'HO, et la DSB qu'elle crée déclenche la résection qui n'est pas entravée par la protéine ParB1 faiblement liée. La disparition de la tache fluorescente rouge signale le passage de la ou des nucléases.

**B)** Images représentatives de la distribution de ParB dans des cellules marquées à *parS1* et *parS2*. Barre = 2 µm. (légende et figure tirées de Saad, 2014).

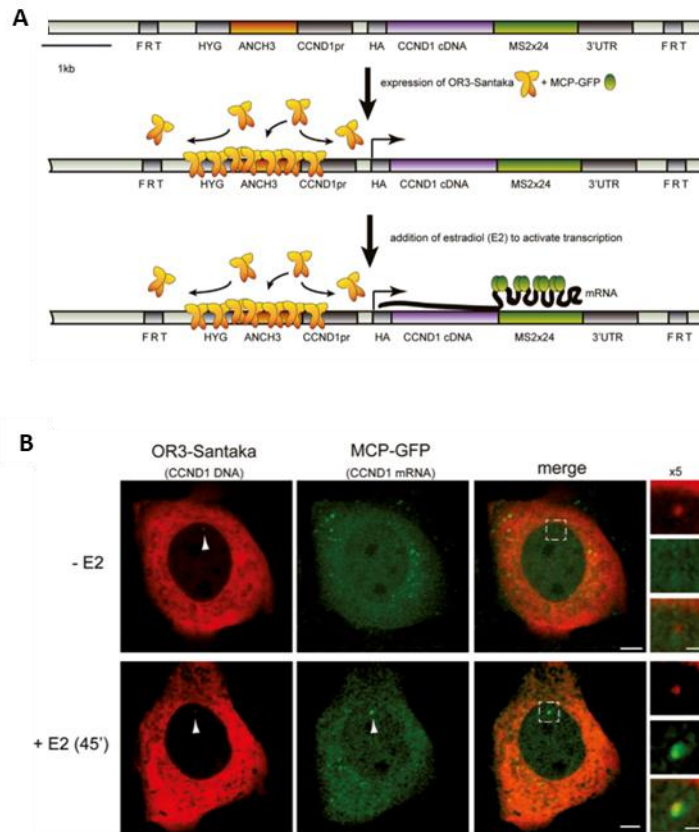
L'équipe de Kerstin Bystricky a développé une méthode originale en utilisant l'oligomérisation des protéines du système ANCHOR plutôt que la multiplicité des opérateurs pour former des foyers visibles. Les sites *parS1* et *parS2* sont insérés respectivement à 76bp et 3,4kb en amont du site de coupure HO du locus *MAT* sur le chromosome III de *Saccharomyces cerevisiae* (fig.

**68**). En induisant une coupure au niveau de ce site HO, une DSB est créée. La réparation de cette DSB déclenche la résection dont l'action des nucléases est visualisée par la disparition du foyer de fluorescence générée par la fixation de la protéine ParB1-mCherry sur ses séquences de fixation *parS1* situées à proximité du site de coupure HO (à 76bp exactement). La présence de la protéine ParB1 n'entrave pas l'action des nucléases. Le foyer de fluorescence dû à la fixation de ParB2 sur les séquences *parS2* est toujours visible permettant de continuer de localiser le locus *MAT*. Ainsi l'utilisation de ce système ANCHOR a permis à l'équipe de Kerstin Bystricky de fournir la première analyse en temps réel de la résection dans des cellules vivantes uniques. Le suivi de la dynamique d'un locus chromatinien à côté d'une DSB a révélé un confinement transitoire de la région chromatinienne endommagée au cours des toutes premières étapes de la résection, ce qui est cohérent avec la nécessité de maintenir les extrémités de l'ADN en contact (Saad, 2014).

L'insertion de séquences *parS* n'altère pas la dynamique de la chromatine, son statut transcriptionnel ou sa sensibilité au dommage à l'ADN. La taille réduite de la séquence ANCH insérée (<1kb) contenant un nombre réduit de sites de liaison facilite l'intégration et la stabilité des sites de liaison. Le système de visualisation ANCHOR permet des mesures cinétiques directes sur des cellules uniques, évitant ainsi toute interférence avec le processus étudié. Ces caractéristiques rendent le système largement applicable à l'étude du positionnement et de la dynamique des chromosomes à petite échelle dans des contextes autres que le processus de réparation étudié ici, tels que l'expression des gènes, la réplication et la recombinaison (Saad, 2014).

Le système ANCHOR a été ensuite adapté pour étudier le comportement de la chromatine pendant l'activation de la transcription d'un unique gène dans des cellules tumorales mammaires humaines, MCF-7 (**fig. 69**). Pour cette étude, Germier et collaborateurs (Germier, 2017) ont utilisé un troisième système ANCHOR qui a été développé et optimisé par la société NeoVirTech SA. Une construction comprenant la séquence centromérique ANCH3 en amont du gène cycline D1 (*CCND1*) sous contrôle de son promoteur endogène est insérée à l'aide de sites FRT dans des cellules MCF-7 FRT (**fig. 69 A**). Le système ANCHOR est couplé avec le système MS2/MCP qui permet de voir les transcrits primaires, en insérant 24 sites MS2 dans la partie 3'UTR du gène *CCND1*. L'expression du gène *CCND1* est inductible par l'ajout d'estradiol. L'expression transitoire des protéines OR3-Sankara (ou ParB) et MCP-eGFP





**Figure 69: Visualisation en temps réel du locus du gène Cyclin D1 dans des cellules humaines.**

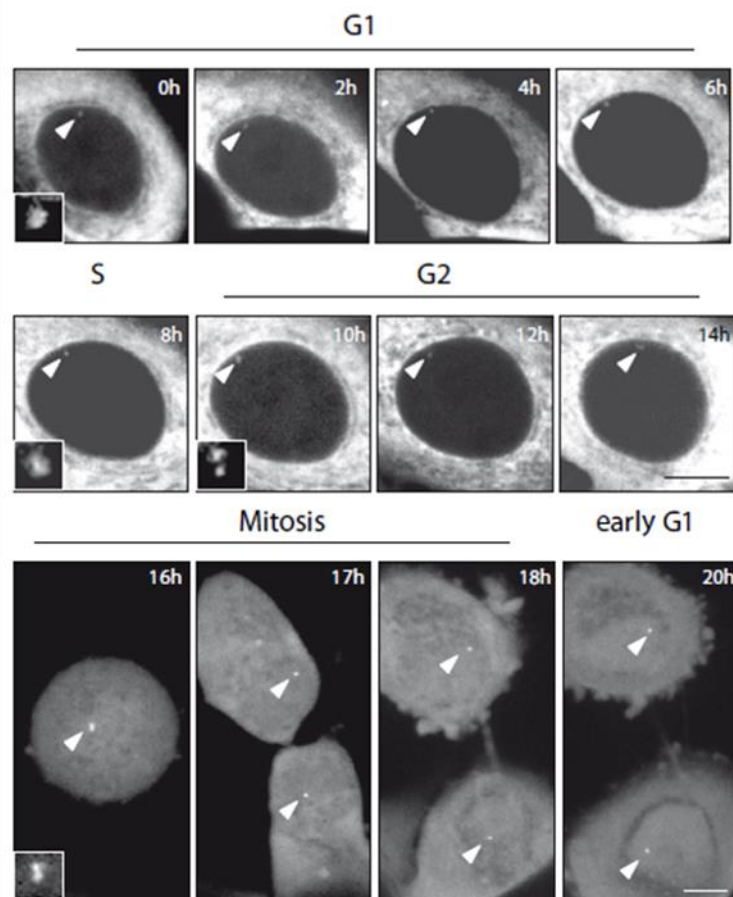
**A)** Représentation schématique d'une construction (ANCH3-CCND1-MS2), insérée de façon stable, comprenant le gène de la *cycline D1* (*CCND1*) sous son promoteur endogène, adjacent à une séquence ANCH3 unique, 24 répétitions MS2 dans le 3'UTR et un gène de sélection de l'hygromycine (HYG). La construction est flanquée de sites FRT pour l'intégration dans les cellules MCF-7 FRT. La transfection transitoire avec des protéines OR3 et MCP, fusionnées avec des protéines fluorescentes entraîne leur accumulation au niveau des séquences ANCH3 et MS2, respectivement (après stimulation par l'estradiol, E2).

**B)** Images représentatives de cellules ANCH3-CCND1-MS2 transfectées de manière transitoire et exprimant OR3-Santaka et MCP-eGFP. Le locus du gène *CCND1* (spot rouge) se co-localise avec l'ARNm transcrit (point vert) localisé par MCP-EGFP (spot vert) qui s'associe aux boucles de la tige MS2 45 minutes après l'ajout de 100 nM d'estradiol (E2). La même cellule est représentée avant et après l'ajout de E2. Barres d'échelle, 5 et 2 mm (pour les images recadrées). (légende et figure tirées de Germier, 2017).

permet respectivement de localiser le locus *CCND1* et de visualiser la synthèse des transcrits primaires du gène *CCND1* (**fig. 69 B**). La protéine Santaka est une protéine fluorescente synthétique non aequorine de la société ATUM. Le suivi du locus marqué par la protéine OR3-Santaka dans des cellules vivantes a permis de démontrer que l'initiation de la transcription restreint le mouvement de la chromatine (Germier, 2017, 2018). De plus, l'intensité des foyers de fluorescence de OR3-Santaka ne varie pas durant l'activation de la transcription. Les événements permettant la synthèse d'ARNm d'un locus marqué n'interfèrent pas avec l'accumulation de la protéine OR3 qui, réciproquement, n'altère pas les mécanismes

transcriptionnels. Ainsi cette étude a également permis de démontrer que le système ANCHOR combiné avec le système MS2 est idéalement adapté pour visualiser simultanément l'ADN et l'ARNm au niveau d'un unique gène dans des cellules humaines vivantes avec une haute résolution spatio-temporelle (Germier, 2017).

Des cellules MCF-7 contenant le transgène CCND1 marqué par le système ANCHOR3 ont été suivies pendant 24h à travers une division cellulaire (**fig. 70** ; Germier, 2018). Pendant l'interphase, le foyer fluorescent ne change pas significativement sa position spatiale dans le noyau. Au bout de 10 heures, deux spots juxtaposés sont visibles, suggérant que le chromosome portant le site ANCH3 est dupliqué et que la cellule est en phase G2. Lorsque les cellules rentrent en mitose, l'enveloppe nucléaire se désintègre et la protéine OR3-Sankara

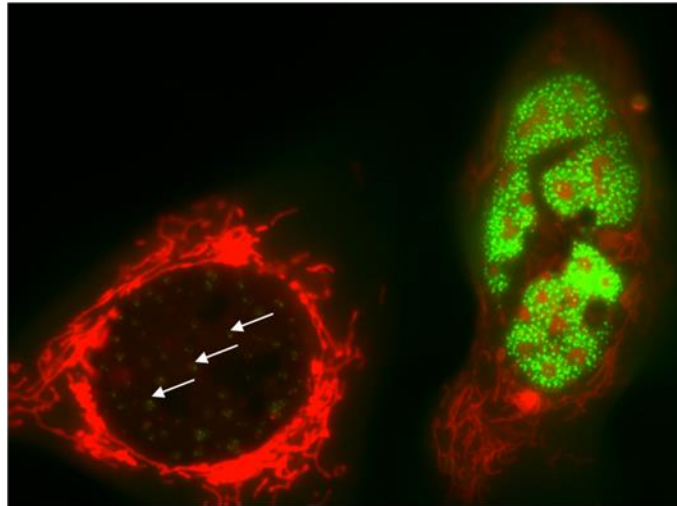


**Figure 70: Le système ANCHOR à travers la division cellulaire.**

Les cellules MCF7 contenant le transgène CCND1 marqué par ANCHOR ont été suivies pendant une division cellulaire (24 h). La cellule est en phase G1 (0h→6 h) avec une tache visible correspondant à la protéine OR3-santaka liée à la séquence ANCH. La cellule en phase S (8h). La cellule en phase G2 (10 h→14 h) avec la présence de 2 foyers distincts. La mitose s'est produite entre 16 h et 18 h, donnant lieu à deux cellules filles contenant chacune un loci OR3-santaka (20 h ; G1 précoce). Barres d'échelle = 5 µm. (légende et figure tirées de Germier, 2018).

occupe tout l'espace de la cellule. A la fin de la cytokinèse, le locus marqué par le système ANCHOR est visible dans les deux cellules filles (**fig. 70** ; Germier, 2018). Plusieurs observations se dégagent de ce suivi à travers une division cellulaire, (i) la chromatine des chromosomes mitotiques est accessible à suffisamment de protéines OR3-Santaka de 70kDa pour produire un foyer fluorescent, corroborant les données actuelles selon lesquelles les chromosomes forment des structures de type fibreuse tout au long de la mitose (Ou, 2017) (ii) les sites génomiques marqués par le système ANCHOR peuvent être imagés sur une longue période, (iii) le locus est visible au cours des différentes phases du cycle de la division cellulaire.

Le système ANCHOR a également été utilisé pour marquer des génomes petits et compacts, y compris des virus. Les cinétiques d'infection et de réplication de l'ADN du virus de l'herpès marqué par le système ANCHOR peuvent être visualisées en temps réel (Mariamé, 2018) et utilisées comme outil pour cribler des candidats antiviraux ([www.neovirtech.com](http://www.neovirtech.com)). La biologie du virus est difficile à explorer en raison de son extrême sophistication et de l'absence d'un modèle animal pertinent. Bernard Mariamé et ces collaborateurs, en collaboration avec la société NeoVirTech ont appliqué le système ANCHOR à un cytomégalovirus humain, le virus herpès, (HCMV) permettant de suivre en temps réel l'infection et la réplication du HCMV dans des cellules humaines vivantes. Ils ont créé un virus recombinant, ANCHOR-HCMV, contenant une séquence ANCH et le gène codant pour la protéine OR-GFP. L'infection de cellules permissives avec le virus recombinant ANCHOR-HCMV permet de visualiser le cycle viral presque complet jusqu'à la fragmentation et la mort des cellules (**fig. 71**). L'analyse quantitative de la cinétique d'infection et de la réplication de l'ADN viral a révélé un comportement du HCMV spécifique au type cellulaire et une sensibilité aux inhibiteurs. Le système ANCHOR est un outil efficace pour l'étude des virus à ADN complexes. Il permet de suivre et de quantifier la réplication du HCMV dans les cellules vivantes. Il est applicable à pratiquement tous les virus à ADN ou présentant une phase d'ADN à double brin. Ce système ANCHOR appliqué au virus est un outil puissant pour permettre de cribler de grandes bibliothèques de produits chimiques. Pour preuve de principe, ils l'ont appliqué à un antiviral connu, le ganciclovir. Associé à la microscopie automatisée, les concentrations inhibitrices à 50 et 90% (IC<sub>50</sub> et IC<sub>90</sub>) de cet antiviral ont été déterminées rapidement avec un investissement minimal en temps. Ainsi le système ANCHOR est un système très prometteur pour le

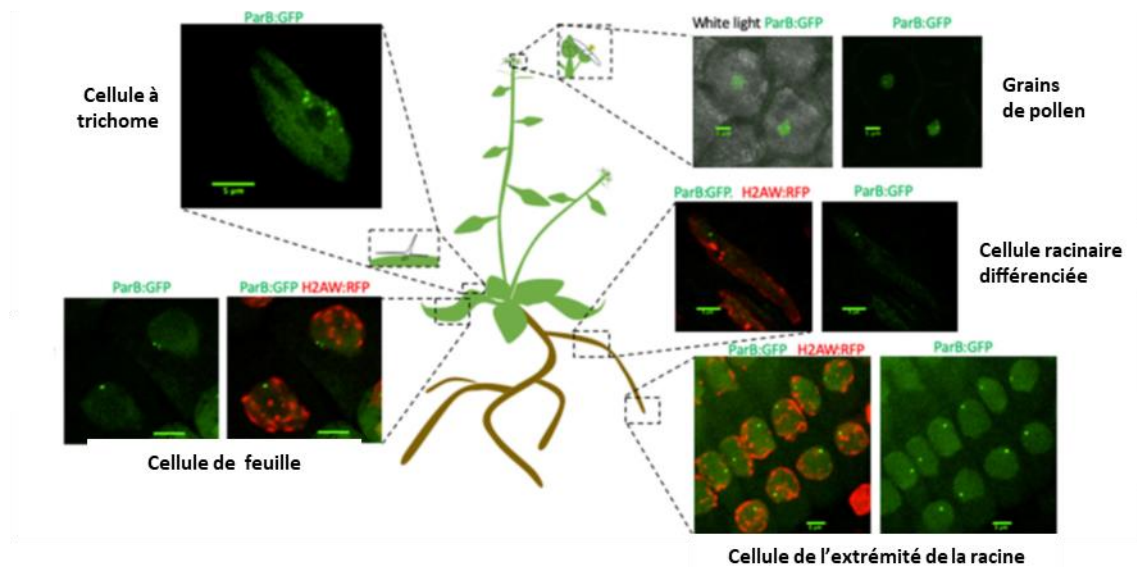


**Figure 71: Cellules infectées par un virus marqué par le système ANCHOR.**

Image de 2 cellules infectées par un virus ayant intégré le système ANCHOR, à 2 stades différents de l'infection. Chaque spot vert représente une charge virale (flèche blanche). (image tirée du site de la société NeoVirTech: [www.neovirtech.com](http://www.neovirtech.com)).

développement d'applications biotechnologiques innovantes (Mariamé, 2018). La société de biotechnologie NeoVirTech a été créée à la suite de cette découverte permettant de visualiser en temps réel dans les cellules vivantes l'infection et la réplication d'un virus ou de n'importe quel vecteur viral. Ainsi NeoVirTech développe des virus et vecteurs viraux auto-fluorescents pour l'imagerie et la découverte de nouveaux médicaments candidats antiviraux dans le domaine de la santé humaine, de la santé animale et des Biodéfenses.

En 2021, l'équipe de Frédéric Pontvianne a adapté l'utilisation du système ANCHOR pour visualiser un locus unique dans des tissus de plantes fixées ou en temps réel (**fig. 72** ; Meschichi, 2021a). Ils s'intéressent aux mouvements des chromosomes et des fibres de chromatine durant la différenciation cellulaire, sujet chez les plantes qui n'est pas bien étudié. Utilisant le système de visualisation de locus ANCHOR mis en place par Frédéric Pontvianne et ses collaborateurs chez les plantes, ils regardent les mouvements de la chromatine durant la différenciation cellulaire. Ils mesurent les mouvements des foci *parS*-ParB-GFP dans les cellules méristématiques et dans les cellules différenciées provenant de l'épiderme de racine par l'utilisation de microscopie confocale en temps réel par et en quantifiant les mouvements par analyse MSD (Mean Square Displacement, Meschichi, 2021b). Ils montrent que le mouvement de la chromatine est plus élevé dans les cellules méristématiques que dans les cellules différenciées, confortant l'idée que la chromatine dans les cellules indifférenciées aurait une conformation plus dynamique (Meschichi, 2021a).



**Figure 72:** Le système ANCHOR permet de suivre un locus à copie unique dans une cellule vivante et dans différents tissus chez *Arabidopsis thaliana*.

Représentation schématique d'une plante, *Arabidopsis thaliana*, illustrant les différents tissus dans lesquels le signal *parS*-ParB-GFP a été détecté par imagerie en temps réel. ParB-GFP, en vert et H2AW-RFP en rouge. Barres d'échelle= 5µm; (légende et image tirées de Meschichi, 2021a).

Depuis la création des deux premiers systèmes ANCHOR, d'autres systèmes ont été développés et optimisés, notamment par la société NeoVirTech, avec des spécificités ParB/*parS* distinctes et provenant de bactéries exotiques. Ces systèmes ont été mis en place chez la levure, dans des cellules de mammifères en culture, dans des virus à ADN, très récemment chez les plantes et j'ai mis en place ce système chez la drosophile (voir partie Résultat, chapitre III).

#### **4-2 Le système ANCHOR et les autres systèmes permettant de visualiser des loci génomiques en temps réel.**

La visualisation de l'ADN au niveau de loci génomiques spécifiques en temps réel nécessite des techniques de marquage qui créent un foyer fluorescent détectable au-dessus du bruit de fond (**tableau 3**). Idéalement, un système de marquage ne devrait perturber ni le locus ni l'ADN voisin, persister pendant toute la durée de l'expérience et ne nécessiter aucune modification ou une modification minimale du génome. Les systèmes qui ont été développés pour l'imagerie de cellules vivantes peuvent être groupés en deux catégories :

Système	Origine	Application, avantages, inconvénients	Éditeur de génome	Références
<b>FROS</b> (Fluorescent Repressor Operateur System)	<i>E. coli</i>	Levure; difficile chez métazoaires (sites fragiles); rupture potentiel de la chromatine, interférence avec transcription	nécessaire: insertions de 5 à 10kb de séquences répétées	Straight, 1996; Marshall, 1997; Lassadi, 2015
<b>TALE</b> (Transcription Activator like Effector proteins)	Activateur de la transcription chez <i>Xanthomonas oryzae</i>	N'importe quelle type cellulaire; limitée aux séquences répétées	non	Miyazari, 2013, 2014; Ma, 2013
<b>CRISPR/Cas9</b>	<i>Streptococcus pyrogenes</i> (Sp), <i>Neisseria meningitidis</i> (Nm) et <i>Streptococcus thermophilus</i> (St1).	N'importe quelle type cellulaire; limitée aux séquences répétées ou sites multiples (plusieurs kb); déroulement local de l'ADN/formation de triple hélice.	non	Chen, 2013; Ma, 2015
<b>Suntag</b>	Anticorps	Potentiellement tout type de cellule ; très spécifique, manque de polyvalence ; complexe final très grand (jusqu'à 1400kDa) ; forte probabilité d'instabilité, d'agrégation et d'interférence avec la fonction de la protéine étudiée	nécessaire: insertions de sites de liaison protéiques	Tanenbaum, 2014
<b>ANCHOR</b>	Systèmes de partition des chromosomes de <i>Burkholderia</i>	Tout type de cellule ; polyvalent ; interférence négligeable avec les processus liés à l'ADN (transcription, réplication)	nécessaire: insertions de séquences uniques de 0,4 à 1kb	Saad, 2014; Germier, 2017, 2018; Mariamé, 2018; Meschichi, 2021

**Tableau 3: Outils permettant le marquage de l'ADN dans des cellules eucaryotes vivantes.**  
(légende et tableau tirés de Bystricky, 2015).

- Les systèmes qui visualisent des séquences répétitives endogènes qui sont la cible des protéines TALE ou CRISPR/Cas9 inactives, fusionnées à des protéines fluorescentes. Ces systèmes basés sur la fixation des protéines TALE et Cas9 présentent l'avantage de ne pas nécessiter l'insertion de séquences dans le génome qui pourraient perturber les processus dépendants de l'ADN comme la transcription ou la réplication. L'avantage d'utilisation de ces 2 systèmes est malheureusement aussi leur inconvénient majeur puisqu'ils sont restreints aux séquences répétées naturelles et l'utilisation de ces systèmes a été limitée au marquage de séquences naturelles répétées comme les télomères (Chen, 2013 ; Ren, 2017).
- Les systèmes qui nécessitent l'insertion de manière aléatoire ou spécifique à un site de cassettes de séquences étrangères sur lesquelles se fixent des protéines de liaison à l'ADN fusionnées à des protéines fluorescentes : les systèmes FROS et ANCHOR. Le système FROS qui a été le premier système de marquage de loci génomiques développé (Straight, 1996) est constitué de centaines de répétitions de séquences d'opérateurs bactériens et de protéines répressives (LacI, TetR, C<sup>lambd</sup>) fusionnées à des protéines fluorescentes qui se lient à chacune d'elles. Le système FROS s'est avéré un outil précieux pour déterminer la position et la

dynamique de loci d'ADN chez la levure, et dans une moindre mesure chez la drosophile et les cellules de mammifères (Straight, 1996 ; Marshall, 1997 ; Lassadi, 2015). Malgré des applications réussies, bien que très utile pour déterminer la position nucléaire d'éléments chromosomiques, la grande multiplicité des sites nécessaires pour un signal et la forte liaison des éléments répresseurs sont souvent problématiques. Ils peuvent interférer avec la structure et la fonction normale de la chromatine, (i) en perturbant la régulation de la transcription (Dubarry, 2011), (ii) en bloquant ou perturbant la fourche de réplication (Possoz, 2006 ; Payne, 2006 ; Sofueva, 2011), (iii) en affectant l'organisation locale de la chromatine et réduire au silence des gènes rapporteurs (Watanabe, 2005), (iv) en favorisant des cassures double brin (Sofueva, 2011), (v) en générant des sites fragiles (Jacome, 2011). Le système ANCHOR, quant à lui, est basé sur l'amplification d'un signal à partir d'un petit site de liaison, *parS*, de moins d'1kb, par l'oligomérisation de la protéine de liaison spécifique, ParB. La séquence *parS* a l'avantage d'être courte, non répétitive et non perturbatrice de la structure de la chromatine. Il peut être utilisé pour marquer n'importe quel locus. Comme les complexes ANCHOR sont générés par des interactions ParB-ParB ou ParB-ADN de force modérée, ils peuvent être rapidement dispersés et reformés, et sont donc moins intrusifs que le marquage par les systèmes FROS. Nous avons pu voir, comme décrit ci-dessus (§ 4-1), que le système ANCHOR n'interfère pas avec les différents processus liés à l'ADN comme la transcription ou la réplication. L'insertion stable du système ANCHOR dans le génome permet de visualiser des événements à long terme ouvrant la voie au suivi de sites génomiques pendant le développement et la différenciation des tissus et organes. Le fait de disposer d'un insert stable tel qu'une unité ANCHOR dans un matériel fragile, comme les cellules souches, permet également de contourner la nécessité d'une transfection ou d'une injection avant les expériences et limite le risque d'endommagement des cellules. De plus, le système ANCHOR peut être combiné avec d'autres systèmes de marquage, comme par exemple le système de marquage des transcrits primaires (Germier, 2017). Les foyers de fluorescence résistent à la fixation et peuvent être combinés avec des signaux d'immunofluorescence (Bystricky, 2015 ; Germier, 2018).

Le système ANCHOR a été utilisé pour visualiser des loci génomiques dans des cellules vivantes chez la levure, dans des cellules de mammifères en culture, dans des virus à ADN, et chez les plantes.

**Le sujet de ma thèse s'articulera sur l'utilisation de ce système de visualisation de locus génomique en temps réel, le système ANCHOR. La première partie portera sur la mise en place du système ANCHOR chez la drosophile. Dans la deuxième partie de ma thèse, j'ai mis en évidence et caractérisé, grâce à ce système ANCHOR, des structures cytoplasmiques pseudo-nucléaires dans des cellules S2 de drosophile.**



## **Résultats**

**Chapitre III : Visualisation de loci génomiques en temps réel dans des tissus d'un organisme entier, la drosophile.**

Le développement d'outils pour la visualisation de loci génomiques spécifiques en temps réel est nécessaire pour comprendre l'organisation des noyaux eucaryotes, la dynamique spatiotemporelle du comportement de la chromatine, ou bien pour étudier l'importance du rôle de l'architecture nucléaire dans la régulation de la transcription des gènes. Ainsi la visualisation ou l'analyse en temps réel de l'ADN à un locus génomique spécifique, dans des cellules eucaryotes vivantes nécessite des techniques de marquage qui permettent d'assembler suffisamment de molécules fluorescentes pour produire un signal stable et persistant le temps de l'expérience. La visualisation de loci génomiques endogènes spécifiques dans des cellules vivantes demeure difficile. Ceci est largement dû à l'insuffisance des techniques pour parvenir à marquer des loci génomiques site-spécifiques de manière à être peu intrusifs afin de préserver les caractéristiques natives et les fonctions de la chromatine (Bystricky, 2015). Les technologies existantes pour visualiser des loci génomiques sont basées soit sur l'insertion, faite au hasard ou de manière site spécifique, de séquences étrangères sur lesquelles des protéines de liaisons à l'ADN fusionnées à des molécules fluorescentes se fixent, soit en ciblant la fixation de protéines fluorescentes fusionnées aux protéines TALE ou Cas9 sur des séquences endogènes, souvent répétitives telles que les télomères (Germier, 2018). L'exemple le plus classique de ces approches est le système fluorescent opérateur/ répresseur ou FROS qui est composé de centaines de répétitions d'opérateurs bactériens (comme LacO ou TetO) sur lesquels viennent se fixer des protéines répresseurs spécifiques (LacI ou TetR) fusionnées à une protéine fluorescente. Bien que très utile pour déterminer la position nucléaire d'éléments chromosomiques, la grande multiplication des sites nécessaires pour obtenir un signal a pour conséquence d'insérer des séquences de grandes tailles (supérieur à 10kb). Cette grande insertion combinée à la liaison étroite du répresseur peut interférer avec la structure normale de la chromatine et sa fonction : problème d'instabilité, perturbation de la régulation de la transcription, création de sites fragiles, d'obstacles, et particulièrement dans les processus dépendants du déroulement de l'ADN tel que la réplication ou la transcription (Payne, 2006 ; Dubarry, 2011 ; Jacome, 2011 ; Sofueva, 2011).

L'équipe de Kerstin Bystricky de l'unité de biologie Moléculaire, Cellulaire et du Développement (MCD) de Toulouse a développé un nouveau système de visualisation de loci génomiques en temps réel, le système ANCHOR, basé sur le système ParABS de partition des

chromosomes bactériens. Pour rappel, ce système ANCHOR est un système en deux parties, composé de séquences nucléiques courtes, moins de 1kb, *parS* (ou ANCH) et des protéines ParB (ou OR). Ces séquences *parS* insérées dans le génome de l'hôte sont spécifiquement reconnues par les protéines ParB qui, une fois fixées, s'oligomérisent et s'étendent à partir des séquences *parS* sur l'ADN environnant. Comme les protéines ParB peuvent être fusionnées à un marqueur fluorescent, leur concentration importante sur les séquences *parS* crée un foyer fluorescent, facilement détectable en microscopie à fluorescence. Ce système a été validé chez la levure (Saad, 2014), les cellules de mammifères (Germier, 2017) ou encore chez les virus (Mariamé, 2018).

La première partie de ma thèse a consisté à évaluer si le système ANCHOR peut fonctionner dans un organisme entier, la drosophile et s'il interfère avec le développement de la drosophile ou avec la transcription de gènes voisins.

Les réponses apportées aux questions suivantes permettront de valider ce système chez la drosophile

#### **- Ce système est-il toxique ?**

L'expression des protéines bactériennes ParB chez la drosophile et leur fixation sur leurs séquences vont-elles interférer avec le développement et la viabilité de la drosophile ?

#### **- Ce système est-il fonctionnel ?**

Que sera la distribution intracellulaire de ces protéines en présence ou pas de leur site de fixation ? Est-ce que les protéines ParB pourront se fixer sur leurs séquences, intégrées dans le génome de la drosophile ? La présence d'une copie de séquences *parS* sera-t-elle suffisante pour obtenir une concentration assez importante de fluorescence permettant de visualiser le foyer ?

#### **- Ce système est-il spécifique ?**

Est-ce que les différentes protéines ParB sont spécifiques de leurs séquences de fixation *parS* ? Est-ce qu'il est possible d'utiliser différents systèmes ParB/*parS* afin de visualiser simultanément différents loci ?

## - Ce système modifie-t-il l'activité de régions régulatrices ?

Est-ce que la fixation des protéines ParB sur leurs séquences *parS* a une influence sur l'activité d'une région régulatrice située à proximité ? Et inversement, l'activité d'une région régulatrice située à proximité d'une séquence *parS* influe-t-elle sur la fixation des protéines ParB sur leurs séquences *parS* et la création d'un spot de fluorescence ?

### 1- Les différents systèmes ANCHOR.

Au cours de ma thèse, j'ai testé chez la drosophile quatre systèmes ANCHOR différents afin de déterminer le ou les couples ParB/*parS* qui fonctionnent le mieux pour visualiser un locus génomique chez la drosophile et avoir plusieurs combinaisons possibles pour pouvoir localiser différents locus simultanément.

Chaque système ANCHOR est constitué d'un couple ParB/*parS* différent, chacun spécifique pour la partition d'un réplicon bactérien. Au cours de ce travail, les couples ParB/*parS* ou ANCHOR impliqués dans la partition des réplicons de *Burkholderia cenocepacia* J2315, pour les chromosomes c2 et c3 (Dubarry, 2006 ; Passot, 2012) sont nommés respectivement ParB1/*parS1* (ANCHOR1) et ParB2/*parS2* (ANCHOR2) et celui pour la partition du plasmide F d'*Escherichia coli* (Sanchez, 2015) ParB4/*parS4* (ANCHOR4). Les plasmides correspondant aux systèmes ANCHOR1 et 2 m'ont été fournis par Kerstin Bystricky, ceux du système ANCHOR4 par Jean-Yves Bouet de l'unité LMGM de Toulouse. Le couple ParB3/*parS3* ou ANCHOR3 sont la propriété de la société NeoVirTech, dont la nature et les séquences sont confidentielles.

J'ai cloné les séquences *parS* en amont du gène rapporteur lacZ codant pour la protéine  $\beta$ Galactosidase, dont l'expression est contrôlée ou non par différentes régions régulatrices (enhancer ou CRM) de gènes, dans un vecteur permettant leur intégration dans le génome de la drosophile en utilisant le système d'intégration par recombinaison Attb/Attp (Bischof, 2007). Ces séquences *parS* sont intégrées dans le génome de la drosophile à différentes positions cytologiques (voir annexe 2, listes des lignées transgéniques). Dans le **tableau 4** sont indiquées les caractéristiques des différentes séquences *parS* clonées. Les séquences *parS1* sont constituées de trois modules contenant chacun un site de fixation pour les protéines ParB1. Les séquences *parS2* comprennent deux modules avec chacun un site de fixation pour

Séquence de fixation <i>parS</i>	Nombre de site de fixation par module	Nombre de modules par constructions	Longueur de chaque module (bp)
<i>parS1</i>	1	3	420bp
<i>parS2</i>	1	2	383bp
<i>parS3</i>	10	1	981bp
<i>parS4</i>	7	1	505bp

**Tableau 4: Caractéristiques des séquences *parS* utilisées pour chaque système.**

Les constructions plasmidiques contenant les séquences *parS* 1, 2, 3 et 4 ont respectivement 3, 2, 10 et 7 sites de fixation pour leur protéine ParB spécifique.

les protéines ParB2. Quant aux séquences *parS3* et *parS4*, elles possèdent un seul module avec, respectivement dix et sept sites de fixation pour les protéines PaB3 et ParB4. Les séquences *parS1*, *parS2* et *parS4* correspondent aux séquences endogènes de chaque réplicon. La séquence *parS3* a été optimisée par la société NeoVirTech. Donc les séquences permettant la fixation des protéines ParB contiennent de 3, 2, 10 et 7 sites de fixation pour respectivement les protéines ParB1, ParB2, ParB3 et ParB4. Les séquences permettant la fixation des protéines ParB sont comprises entre 505bp et 1260bp : 1260, 766, 981 et 505bp pour respectivement *parS1*, *parS2*, *parS3* et *parS4*. La longueur des séquences *parS* nécessaires pour la fixation des protéines ParB est très inférieure à celle utilisée dans les systèmes de visualisation de loci de type FROS, où les séquences de fixation des protéines correspondantes peuvent dépasser les 10kb (Straight, 1996). De plus, les séquences *parS* contiennent peu de séquences répétées contrairement à celles des systèmes FROS où la multiplication des séquences répétées peut être supérieure à 250 (Straight, 1996), ce qui entraîne des difficultés dans la construction des plasmides les contenant. La construction, la manipulation des plasmides contenant les séquences *parS* ne pose aucun problème, de même que leur intégration dans le génome de la drosophile.

Les séquences codant pour la synthèse des protéines ParB chez la drosophile sont clonées dans un vecteur permettant l'expression conditionnelle de ces protéines à l'aide du système Gal4/UAS (Brand, 1993) et sont intégrées dans le génome de la drosophile à différentes positions génomiques en utilisant le système d'intégration par recombinaison Attb/Attp. Ces protéines sont fusionnées à différentes protéines fluorescentes correspondant à différentes couleurs : soit en vert avec la protéine GFP ou mVenus (un variant monomérique de la protéine YFP), soit en rouge avec la protéine mCherry.

Un large éventail de possibilités d'utilisation est disponible pouvant permettre notamment la visualisation de différents loci.

En **annexe 1**, se trouve la liste des différents plasmides que j'ai construits au cours de ma thèse. A partir de ces différents plasmides, toute une série de lignées transgéniques ont été générées, dont la liste se trouve en **annexe 2**. La partie « Matériels et Méthodes » se trouvent en **annexe 3** où est présentée notamment les différentes méthodologies de clonage.

## **2- Le système ANCHOR est-il toxique pour le développement de la drosophile ?**

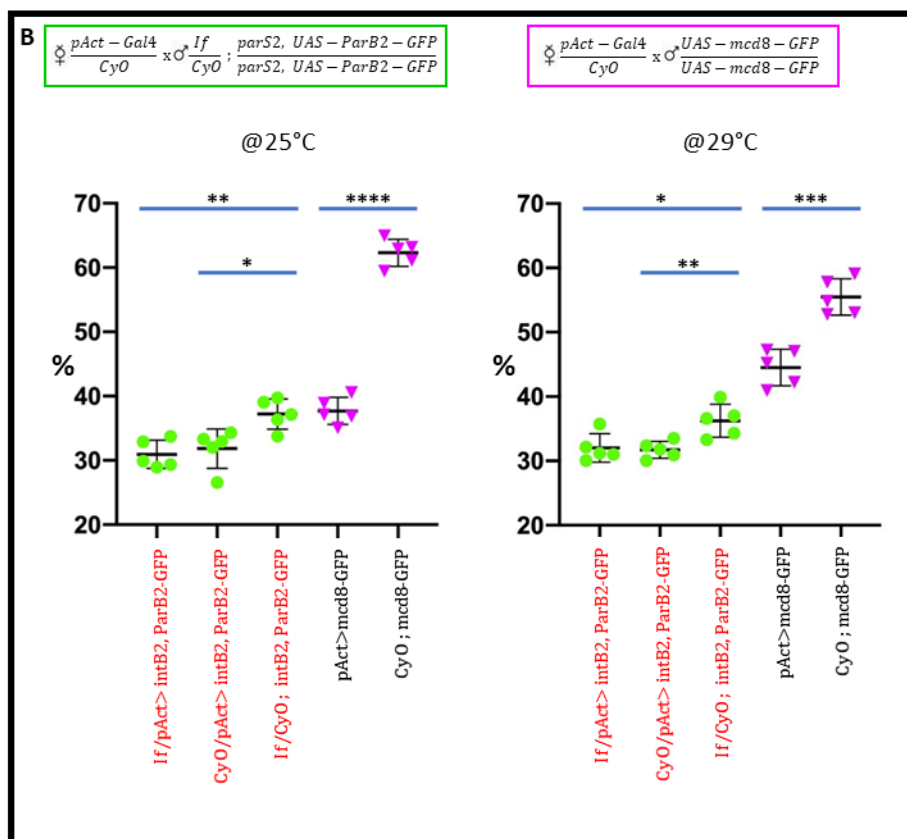
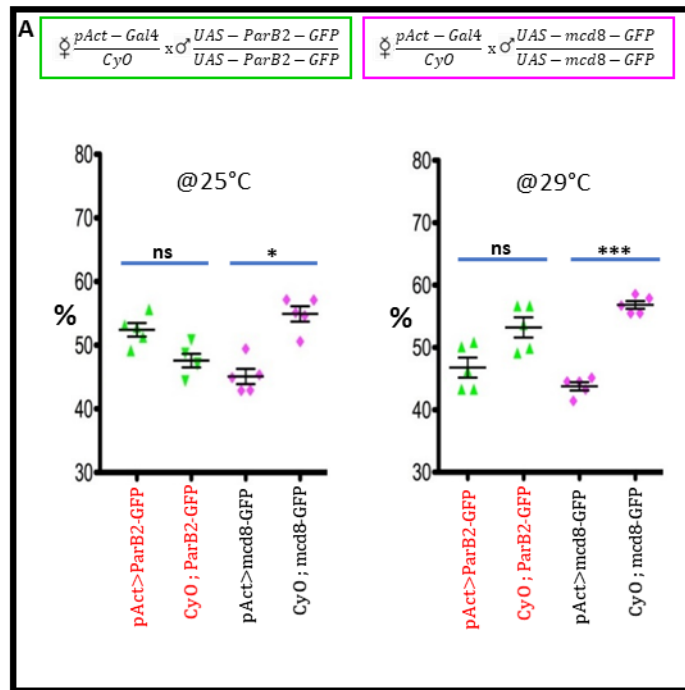
Le système ANCHOR est dérivé des systèmes de partition des chromosomes bactériens. Les lignées transgéniques obtenues à partir des différentes constructions plasmidiques vont exprimer les protéines ParB. Quelles sont les conséquences de l'expression de ces protéines d'origine bactérienne sur le développement et la viabilité de la drosophile ?

Avant d'utiliser ce système ANCHOR, il est nécessaire de tester sa toxicité.

Afin de répondre à cette question, j'ai effectué tout d'abord des tests de toxicité de l'expression de la protéine ParB fusionnée à la protéine fluorescente GFP chez la drosophile, en présence ou non de ces séquences de fixation. Le promoteur utilisé pour piloter le gène Gal4 permettant l'expression de la protéine ParB est le promoteur de l'actine, qui est un promoteur fort permettant l'expression de la protéine Gal4 chez la drosophile de manière ubiquitaire, aussi bien au cours du développement que chez l'adulte. Ainsi la protéine ParB2 sera exprimée dans l'ensemble des tissus de la drosophile et au cours de son développement et de sa vie d'adulte. Pour cette étude, j'ai utilisé le couple *parS2*/ParB2.

Dans la **figure 73 A**, sont présentés les résultats du test de toxicité pour l'expression de la protéine ParB2 chez la drosophile. En ordonnée est représenté le pourcentage de chaque génotype, indiqué en abscisse, correspondant à un croisement.

J'ai croisé les lignées [pAct-Gal4/CyO] et [UAS-ParB2-GFP/UAS-ParB2-GFP] à 25°C et à 29°C. Dans la descendance, on obtient la lignée [pAct-Gal4 ; ParB2-GFP] qui permet l'expression de la protéine ParB2-GFP et la lignée [CyO ; ParB2-GFP] chez laquelle la protéine ParB2-GFP n'est pas exprimée. Comme témoin, j'ai utilisé la lignée [UAS-mcd8-GFP/UAS-mCD8-GFP]. mCD8 est un marqueur de lymphocyte de souris qui fusionné à la protéine GFP permet de l'adresser à



**Figure 73: Test de toxicité du système ANCHOR chez la drosophile**

**A)** Test de toxicité de la protéine ParB2-GFP: l'expression de la protéine ParB2 n'entraîne pas de létalité chez la drosophile.

**B)** Test de toxicité de la fixation de la protéine ParB2-GFP sur ces séquences *parS2*: la fixation de la protéine ParB2-GFP sur ces séquences *parS2* pourrait entraîner une très faible létalité chez la drosophile.

(cinq croisements par génotype, répliqués quatre fois, [10 femelles/7 mâles] par tubes, à 25°C et 29°C).

En ordonnée, les pourcentages de chaque génotype, indiqué en abscisse, issus des croisements notés au dessus.



la membrane plasmidique. Cette protéine de fusion est exprimée de manière ubiquitaire, comme pour la protéine ParB2-GFP, grâce au pilote pAct-Gal4.

A 25°C et à 29°C, on constate qu'il n'y a pas de différence significative entre les deux lignées [pAct-Gal4 ; ParB2-GFP] et [CyO ; ParB2-GFP], ce qui signifie que l'expression de la protéine ParB2-GFP dans ces conditions n'est pas toxique pour la drosophile. Par contre, dans la cadre de la lignée témoin, on constate une différence significative entre [pAct-Gal4/UAS-mcd8-GFP] où la protéine mcd8-GFP est fortement exprimée au niveau de la membrane plasmique et [CyO/UAS-mcd8-GFP] sans expression de la protéine mcd8-GFP. Une forte expression de la protéine mcd8-GFP à la surface des cellules semblerait générer une certaine létalité chez la drosophile.

Les résultats de toxicité de la fixation de la protéine ParB2-GFP sur ces séquences *parS2* sont exposés dans la **figure 73 B**. La lignée pilote [pAct-Gal4/CyO] est croisée avec la lignée [If/CyO ; *parS2*, UAS-ParB2GFP/ *parS2*, UAS-ParB2GFP/]. A l'issue de ce croisement, on obtient soit les lignées [pAct-Gal4/If ; *parS2*, UAS-ParB2-GFP] et [pAct-Gal4/CyO ; *parS2*, UAS-ParB2-GFP] qui permettent l'expression de la protéine ParB2-GFP et sa fixation sur ses séquences spécifiques *parS2*, soit la lignée [CyO/If ; *parS2*, UAS-ParB2-GFP] chez laquelle la protéine ParB2-GFP n'est pas produite. Dans ce croisement, le génotype [CyO/CyO ; *parS2*, UAS-ParB2-GFP] ne fournit pas d'adulte car il est létal. Que ce soit à 25°C ou à 29°C, on constate une légère différence significative entre les lignées exprimant la protéine ParB2-GFP et celles qui ne l'expriment pas. Ce qui pourrait signifier que la fixation de la protéine ParB2-GFP sur ses séquences *parS2* engendrerait une certaine létalité. Ceci est à minorer si on compare avec la lignée témoin dans laquelle on observe une différence significative entre la lignée produisant la protéine mCD8-GFP et celle ne la produisant pas. La présence importante de la GFP au niveau de la membrane plasmique pourrait expliquer cette létalité. Il est à noter que si l'on effectue ces tests statistiques à partir du nombre de mouches pour chaque lignée et non pas le pourcentage que représente chaque lignée, on n'obtient pas de différence significative entre les lignées chez lesquelles la protéine ParB-2GFP se fixe sur ses séquences et celles où la protéine ParB2-GFP n'est pas exprimée.

Au vu de ces résultats, on peut considérer que le système ANCHOR n'est pas toxique pour la drosophile.

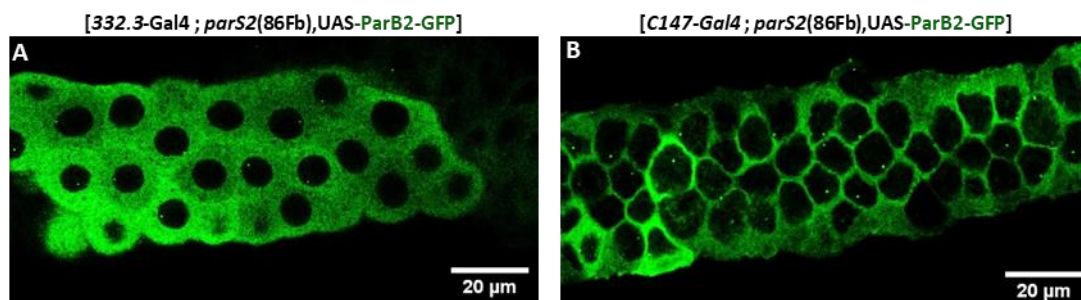
### 3- Le système ANCHOR est-il fonctionnel ?

#### 3-1 Sélection des pilotes permettant l'expression de la protéine ParB dans les glandes salivaires.

Afin de valider l'utilisation du système ANCHOR au sein d'un organisme, j'ai tout d'abord entrepris cette étude dans les glandes salivaires de drosophile parce qu'elles possèdent de grosses cellules avec des chromosomes polytènes qui, suite à plusieurs cycles de réplication sans division cellulaire, ont un nombre élevé de copies de chromosome, ou degré de polyténie, qui peut atteindre 1024. Chaque chromosome amplifié reste accolé et parallèle aboutissant à des chromosomes géants. Pour cette étude, je sélectionne des larves au stade L3 pour lesquelles j'ai défini un temps de développement correspondant à des glandes salivaires avec les cellules les plus grosses mais précédant le phénomène d'histolyse.

Pour exprimer les protéines ParB dans les glandes salivaires, j'ai testé différents pilotes provenant du Bloomington Drosophila Stock Center aux Etats-Unis : *Sgs3-Gal4*, *332.3-Gal4*, et *C147-Gal*.

J'ai tout de suite éliminé le pilote *Sgs3-Gal4*, car lorsque ce pilote dirige l'expression de la protéine ParB fusionnée à la GFP, je vois peu d'expression de la protéine GFP dans les glandes salivaires de larves, ce qui n'est pas le cas avec les deux autres pilotes. Les deux autres pilotes, *332.3-Gal4* et *C147-Gal4*, permettent d'avoir une bonne expression de la protéine ParB-GFP dans les glandes salivaires (**fig. 74**).



**Figure 74: Test de différents pilotes pour l'expression de la protéine ParB2-GFP dans les glandes salivaires.**

Expression de la protéine ParB2-GFP avec le pilote *332.3-Gal4* **A**) et *C147-Gal4* **B**).  
(**anti-GFP** ; x63).

J'ai choisi le pilote *C147-Gal4* car l'expression de la protéine ParB-GFP est plus forte comparée à celle avec le pilote *332.3-Gal4*. Pour toutes les expériences qui seront effectuées dans les glandes salivaires de larve, les protéines ParB sont exprimées avec ce pilote.

### **3-2 Un site *parS* est suffisant pour générer un foyer fluorescent dans le noyau.**

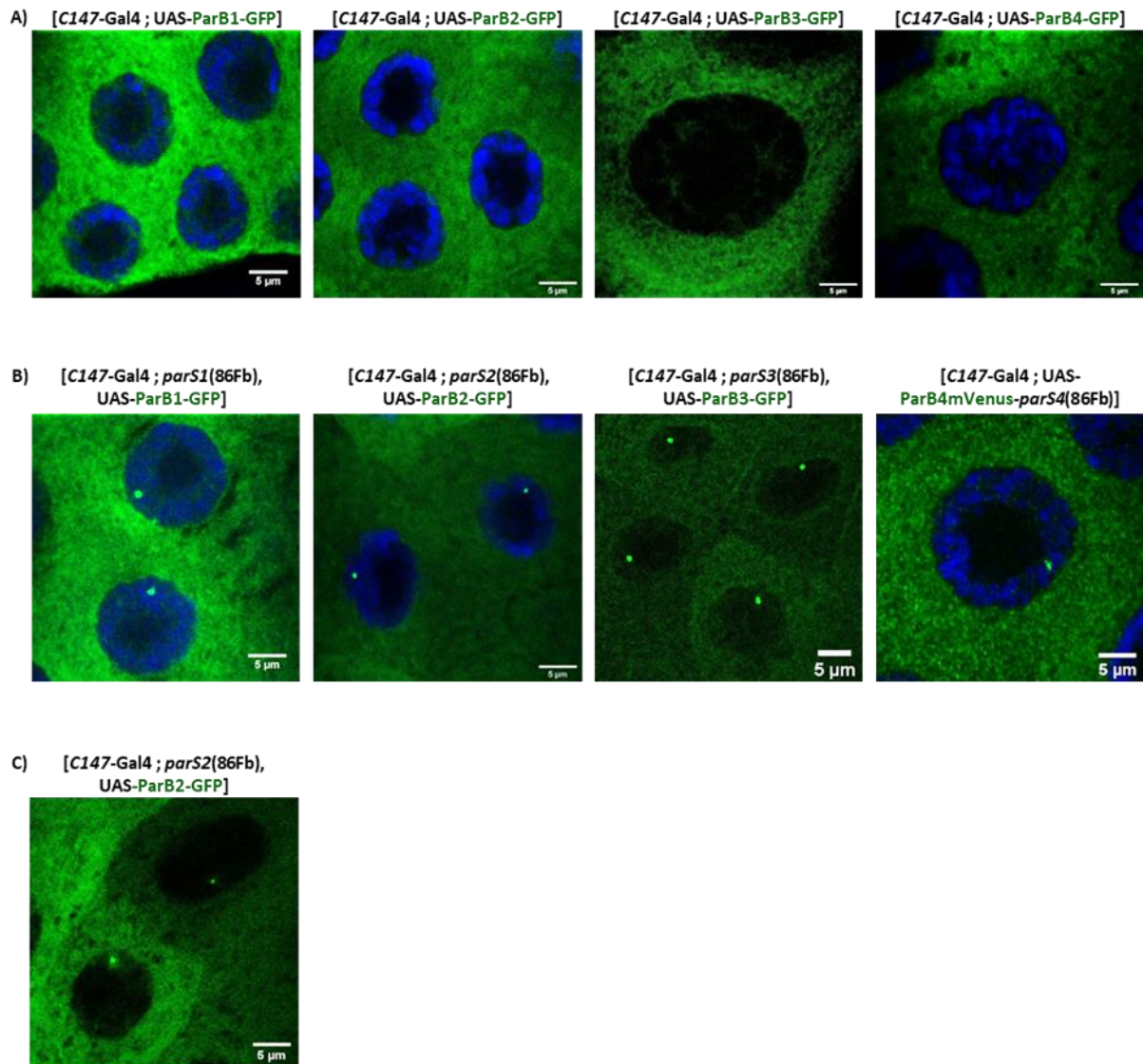
J'ai regardé dans les glandes salivaires de drosophile l'expression des protéines ParB pour les quatre systèmes ANCHOR, afin de déterminer comment se localise la protéine ParB dans la cellule, en absence puis en présence des séquences *parS*.

Dans les glandes salivaires de troisième stade larvaire, en absence de leur séquence de fixation *parS*, l'expression des protéines ParB fusionnées à la protéine fluorescente GFP et dirigées par le pilote *C147-Gal4*, montre une localisation diffuse des protéines ParB-GFP qui s'accumulent massivement dans le cytoplasme, quel que soit le système ANCHOR auxquelles elles appartiennent (**fig. 75 A**). Cette localisation cytoplasmique est identique à celle que l'on peut observer quand la protéine ParB est exprimée dans des cellules de mammifères en absence de ses séquences de fixation *parS* (Fok, 2021).

En présence des séquences de fixation *parS*, insérées dans le génome de la drosophile sur le troisième chromosome en position cytologique 86Fb, la protéine ParB-GFP est toujours présente dans le cytoplasme mais un foyer de fluorescence est observé dans le noyau (**fig. 75 B**). Les séquences *parS*, qui ont une taille pour la plupart inférieure à 1kb, permettent de stabiliser la protéine ParB et d'accumuler suffisamment de fluorescence pour créer un foyer facilement détectable en microscopie. Ce spot fluorescent, correspondant à la fixation de la protéine ParB-GFP sur ses séquences *parS*, permet de localiser le locus génomique 86Fb dans le noyau. En présence des séquences *parS*, la protéine ParB est visible dans le noyau uniquement au niveau du spot de fluorescence. Cette modification de la distribution intracellulaire de la protéine ParB en présence des séquences *parS* est identique pour les quatre systèmes ANCHOR utilisés (**fig. 75 B**). Ce système fonctionne aussi avec les autres protéines fluorescentes fusionnées aux protéines ParB

La robustesse du système ANCHOR permet de visualiser un spot en temps réel et *in vivo* sur tissu vivant (**fig. 75 C**). Le pilote *C147-Gal4* dirige l'expression de la protéine ParB2-GFP dans

les glandes salivaires, qui va se fixer sur ses séquences *parS* et permettre la visualisation d'un spot fluorescent dans des cellules vivantes.



**Figure 75: La fixation de la protéine ParB sur sa séquence *parS* permet de visualiser un locus génomique.**

**A)** Quand les protéines ParB sont exprimées dans les glandes salivaires de drosophile au troisième stade larvaire, elles présentent une expression diffuse, localisée dans le cytoplasme.

(anti-GFP + *topro*; x63 zoom 5).

**B)** La présence des séquences de fixation, *parS*, insérées dans le génome en position cytologique 86Fb, modifie la distribution des protéines ParB et permet de visualiser un spot de fluorescence localisé dans le noyau au niveau du locus génomique 86Fb.

(anti-GFP + *topro*; x63 zoom 5).

**C)** Le système ANCHOR permet de visualiser un locus génétique, en 86Fb, dans des cellules vivantes.

(GFP ; x63 zoom 5).

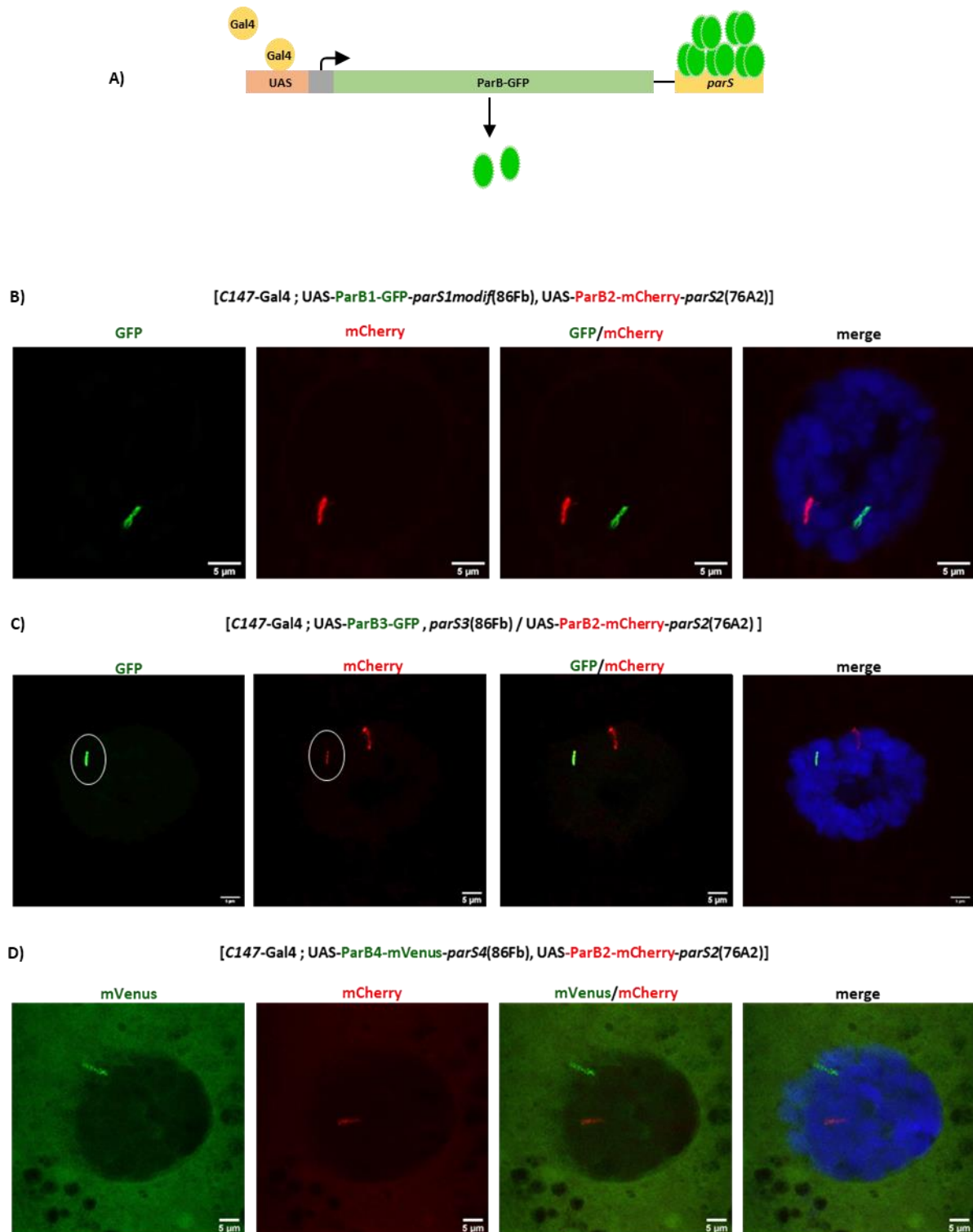
### 3-3 Spécificité de fixation de la protéine ParB2.

La spécificité de fixation des diverses protéines aux séquences de liaison *parS* est importante à connaître si on veut utiliser plusieurs systèmes ANCHOR en parallèle pour visualiser différents loci.

Chez la drosophile, un locus génomique peut être visualisé avec quatre systèmes ANCHOR différents. Cependant le couple ParB2/*parS2* (ANCHOR2) est particulièrement efficace avec un excellent rapport (foyer fluorescent)/(bruit de fond), que ce soit dans les glandes salivaires de drosophile ou dans d'autres tissus comme l'épiderme (voir §4). J'ai essentiellement utilisé ce système ANCHOR2 par la suite. Il est étonnant que ce système ANCHOR2 fonctionne si bien par rapport aux trois autres systèmes car les constructions plasmidiques utilisées pour obtenir les lignées transgéniques ayant intégrées les séquences *parS2* dans le génome de la drosophile contiennent 2 sites de fixation pour la protéine ParB2 (2 modules avec chacun 1 site de fixation) alors que le système ANCHOR1 possède 3 sites de fixation et les systèmes ANCHOR3 et 4, respectivement 10 et 7 sites de liaison (**tableau 4**).

J'ai observé la réactivité croisée du système ANCHOR2 par rapport aux trois autres systèmes. Pour faciliter les croisements génétiques, j'ai construit des plasmides qui contiennent les deux éléments du système ANCHOR, les séquences *parS* et les séquences permettant la synthèse des protéines ParB (**fig. 76 A**). J'ai utilisé des lignées transgéniques qui ont intégrées dans leur génome ces constructions plasmidiques, contenant des séquences *parS* et permettant l'expression des protéines ParB fusionnées à des protéines fluorescentes. L'intégration au même endroit du génome des séquences exprimant la protéine ParB et les séquences de fixation *parS* n'empêche ni l'expression des protéines ParB et ni leur fixation sur leur séquence *parS* permettant de créer un foyer de fluorescence.

Pour tester la réactivité croisée des différents systèmes ANCHOR, j'ai croisé la lignée contenant le système ANCHOR2 [*C147-Gal4* ; *UAS-ParB2-mCherry-parS2*] avec les différentes lignées contenant les systèmes ANCHOR 1, 3 ou 4. La protéine ParB2 est fusionnée à la protéine fluorescente mCherry tandis que les protéines ParB1, 3 et 4 sont fusionnées à la protéine GFP (**fig. 76**). Les séquences *parS2* sont intégrées en position cytologique 76A2, sur le bras gauche du chromosome III, tandis que les séquences *parS1*, 3 ou 4 sont en position 86Fb sur le bras droit du chromosome III.



**Figure 76: Spécificité de fixation de la protéine ParB2.**

A) Schéma des constructions plasmidiques permettant l'expression des protéines ParB et contenant les séquences parS correspondantes.

Test de fixation de la protéine ParB2 sur les séquences parS1 B), parS3 C) et parS4 D) et des protéines ParB1 B), ParB3 C) et ParB4 D) sur la séquence parS2.

La protéine parB2 ne se fixe pas sur les séquences parS1 et parS4, mais une fixation peut être observée sur les séquences parS3 (cercle blanc) C). Les protéines ParB1, ParB3 et ParB4 ne se fixent pas sur les séquences parS2.

(GFP + mCherry + topro ; 63zoom5).

Pour chaque expérience on peut voir deux spots de fluorescence (**fig. 76**). Chaque protéine se localise sur des sites distincts dans le noyau, en 76A2 pour la protéine parB2 et en 86Fb pour les autres protéines ParB. Il n'y a pas de réaction croisée entre le système ANCHOR2, le système ANCHOR1 (**fig. 76 A**) et le système ANCHOR4 (**fig. 76 C**). La protéine ParB2 ne se fixe pas sur les séquences *parS1* et *parS4* et inversement les protéines ParB1 et ParB4 ne se fixent pas sur les séquences *parS2*. Mais la protéine ParB2 peut se fixer sur les séquences *parS3* (**fig. 76 B**). Au niveau du spot de fluorescence verte, indiquant la fixation de la protéine ParB3-GFP sur sa séquence *parS3*, on peut voir un faible signal de fluorescence rouge indiquant la fixation de la protéine ParB2 sur les séquences *parS3*.

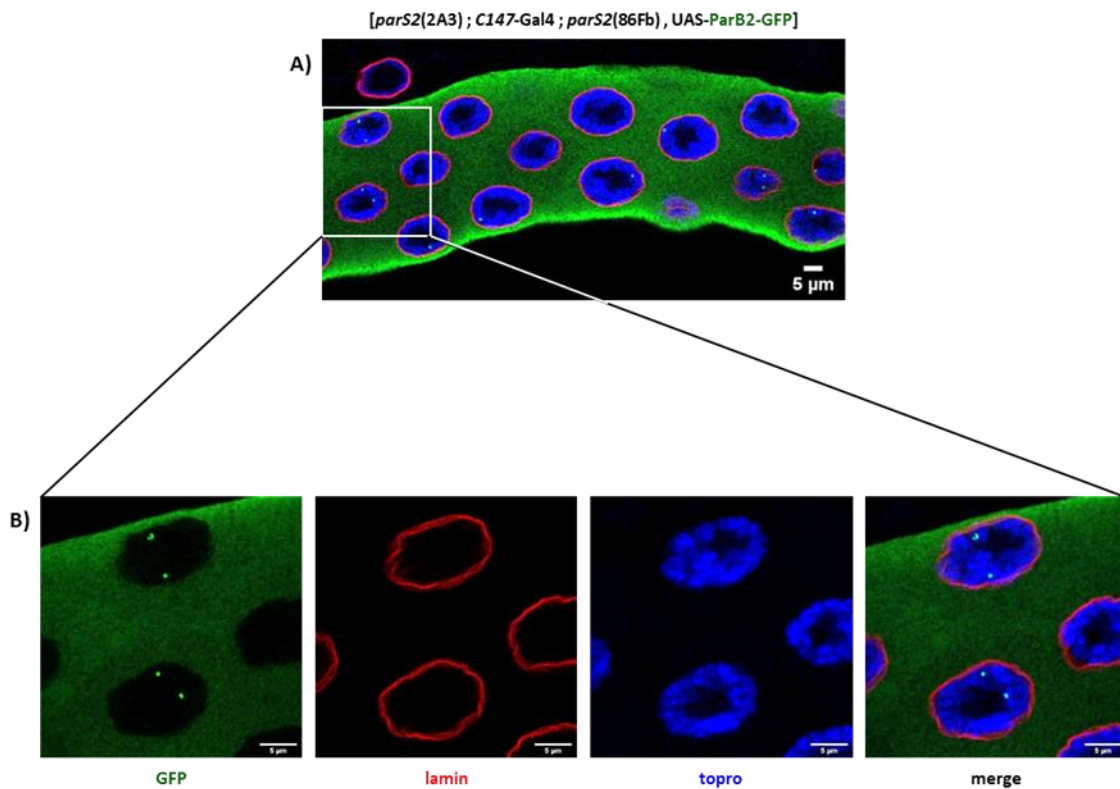
Pour conclure, il est possible de visualiser simultanément deux loci ayant des sites de fixation *parS* différents et il n'y a pas de réaction croisée entre le système ANCHOR2 et les systèmes ANCHOR1 et 3.

Pour visualiser deux loci, il convient d'utiliser les systèmes ANCHOR1 et 2 ou les systèmes ANCHOR2 et 4.

### **3-4 Visualisation par la protéine ParB2 de deux loci génomiques différents.**

La **figure 77** illustre la localisation de deux loci génomiques différents par le système ANCHOR2. Les séquences de fixation *parS2* sont intégrées dans le génome de la drosophile en deux positions cytologiques différentes : sur le chromosome X à la position 2A3 et sur le chromosome III en 86Fb. La fixation de la protéine ParB2 fusionnée à la protéine fluorescente GFP sur ses séquences *parS2* permet de visualiser deux foyers de fluorescence distincts et de localiser les positions cytologiques 2A3 et 86FB. On peut remarquer l'excellent rapport entre le spot de fluorescence et le bruit de fond dans le noyau. Un marquage avec un anticorps contre la lamine A, une protéine de la membrane nucléaire, permet de remarquer que la protéine ParB2 est, en dehors du spot de fluorescence dans le noyau, localisée que dans le cytoplasme.

L'image dans la **figure 77 A** est prise sur un seul plan focal. On peut observer que les deux foyers de fluorescence ne sont pas visibles dans tous les noyaux. On peut voir un seul spot dans certains noyaux et même aucun spot dans d'autres. Si on parcourt l'ensemble du noyau sur le plan focal en Z, on visualisera toujours deux spots par noyaux.



**Figure 77: Visualisation de deux loci génomiques différents.**

La protéine ParB2 permet de localiser deux loci chromosomiques distincts en se fixant sur la séquence *parS2* insérée à deux positions cytologiques différentes, sur le chromosome X en 2A3 et sur le chromosome III en 86Fb. En **A)** x63. En **B)** x63zoom5. (anti-GFP + anti-lamin + topro).

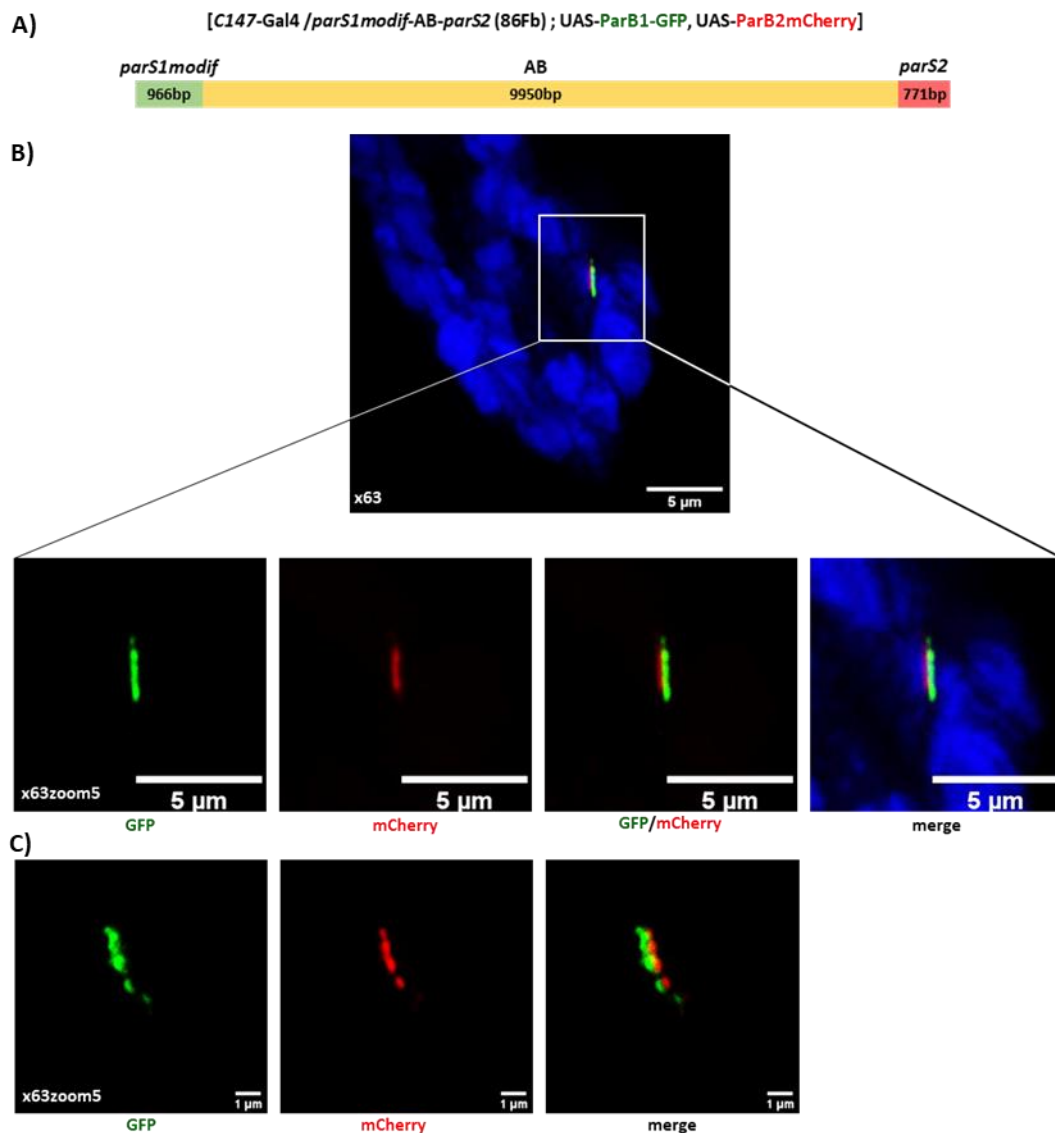
### 3-5 Visualisation de deux loci séparés par dix kilobases.

Les régions régulatrices jouent un rôle clef dans le contrôle de l'expression des gènes. Dans un génome compact comme celui de la drosophile, 20 à 30% des régions régulatrices peuvent interagir avec leurs promoteurs sur des distances dépassant les 20 kb (Ghavi-Helm, 2014). Cependant une région régulatrice se situe généralement entre 4 et 10kb par rapport à ses gènes cibles (Furlong, 2018). Si on veut comprendre comment les régions régulatrices communiquent avec leurs promoteurs, il est nécessaire de développer des outils en imagerie en temps réel, peu intrusifs pour suivre les positions spatiales de régions régulatrices et de leur promoteur cible. Le système ANCHOR peut permettre d'apporter des réponses à ces questions.

**Le système ANCHOR est-il suffisamment sensible pour pouvoir visualiser  
deux loci séparés par 10kb ?**



Afin de répondre à cette question, j'ai construit un plasmide qui contient deux séquences *parS* différentes, *parS1* et *parS2*, séparées par 10kb (fig. 78 A). Dans la figure 78 B, on peut voir deux foyers de fluorescence distincts résultant de la liaison des protéines ParB1 et ParB2 sur leurs séquences de fixation respectives séparées par 10kb. Ainsi seulement 10kb entre deux séquences *parS* sont suffisants pour accumuler assez de fluorescence pour visualiser deux spots sans pour autant se chevaucher (fig. 78 B). Ces deux foyers de fluorescences séparés par 10kb peuvent être également observés dans des cellules vivantes (fig. 78 C).



**Figure 78: Visualisation de deux loci séparés par dix kilobases.**

**A)** Construction plasmidique utilisée pour estimer à quelle distance deux séquences *parS* doivent se situer pour distinguer deux spots de fluorescences distincts.

**B)** Les séquences de fixation pour les protéines ParB1 et ParB2 sont séparées par 10kb, ce qui est suffisant pour visualiser deux spots distincts.

(GFP + mCherry + topopro).

**C)** Deux foyers de fluorescences séparés par 10kb dans des cellules vivantes.

(GFP + mCherry + topopro).

#### **4- Le système ANCHOR dans l'épiderme embryonnaire et dans d'autres tissus de la drosophile.**

Dans le paragraphe précédent, nous avons vu la mise en place et l'efficacité du système ANCHOR dans les glandes salivaires de drosophile de troisième stade larvaire pour visualiser un ou plusieurs loci génomiques en temps réel. Mais le génome de ces cellules est particulier. Les cellules de drosophile de troisième stade larvaire sont des cellules de grande taille et leurs chromosomes sont polytènes. Les séquences *parS* intégrées dans le génome sont présentes jusqu'à 1024 copies.

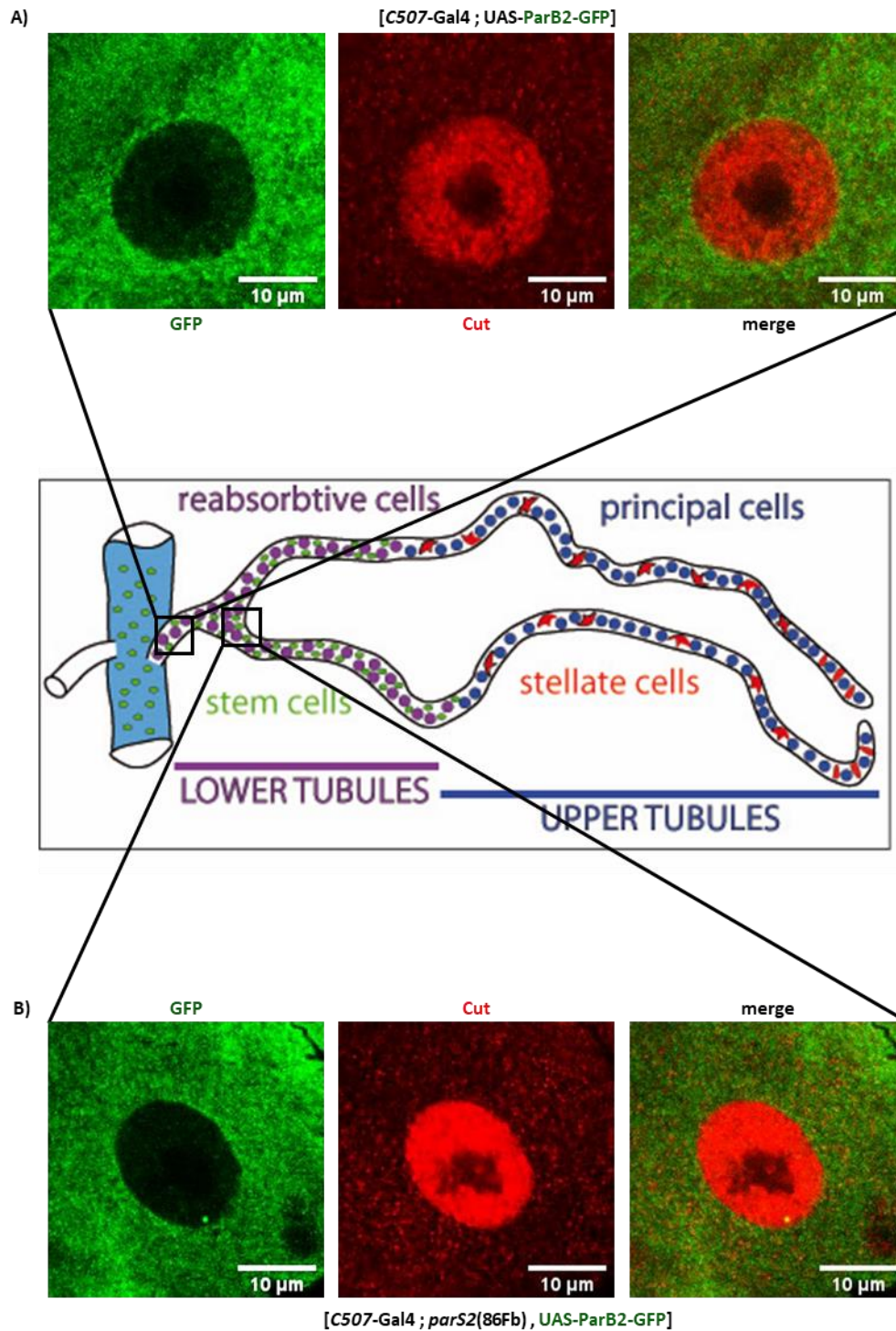
Mais comment se comporte le système ANCHOR dans d'autres systèmes où les cellules n'ont pas une polyténie si élevée comme dans les disques imaginaux ou les cellules épidermiques embryonnaires qui sont diploïdes ? et chez l'adulte ? Est-il possible de voir un foyer de fluorescence lorsque les séquences *parS* sont présentes sur un seul chromosome ?

##### **4-1 Chez l'adulte, dans des cellules polyploïdes.**

Je vais tout d'abord regarder, avec l'aide de Cédric Polesello, comment se comporte le système ANCHOR chez l'adulte, dans les cellules réabsorbantes du tube de Malpighi. Les tubules de Malpighi assurent la fonction excrétrice du système rénal chez la drosophile. Un schéma représentant les tubules de Malpighi est représenté en **figure 79**. Le pilote *C507-Gal4* exprime les protéines dans les tubules inférieurs et notamment dans les cellules réabsorbantes qui, comme leur nom l'indique, sont capables de réabsorber l'eau ainsi que les ions  $K^+$  et  $Cl^-$  tout en sécrétant du  $Ca^{2+}$  (Bohère, 2017). Ces cellules peuvent être marquées par un anticorps dirigé contre la protéine Cut, qui est exprimée principalement dans les noyaux polyploïdes des cellules réabsorbantes.

En absence des séquences *parS2*, la protéine ParB2-GFP, exprimée dans les cellules réabsorbantes des tubules de Malpighi par le pilote *C507-Gal4*, est strictement localisée dans le cytoplasme (**fig. 79 A**).

En présence des séquences *parS2*, on peut observer un spot fluorescent dans le noyau des cellules réabsorbantes résultant de la fixation des protéines ParB2-GFP sur ses séquences *parS2* (**fig. 79 B**).



**Figure 79: Visualisation d'un locus génomique dans des cellules des tubules de Malpighi chez la drosophile adulte.**

*C507-Gal4* dirige l'expression dans le segment inférieur des tubes de Malpighi et notamment dans les cellules de réabsorption. En absence des séquences *parS*, la protéine ParB est exprimée dans le cytoplasme. En présence des séquences *parS*, la protéine ParB se fixe dessus et crée un foyer de fluorescence dans le noyau.

(GFP + anti-cut; x63zoom5).

C. Polesello

## 4-2 Dans des cellules diploïdes.

### 4-2-1 Dans les disques imaginaux.

Le système ANCHOR permet de visualiser des loci génomiques en temps réel dans des cellules polyplœïdes comme les glandes salivaires ou les cellules réabsorbantes du tube de Malpighi.

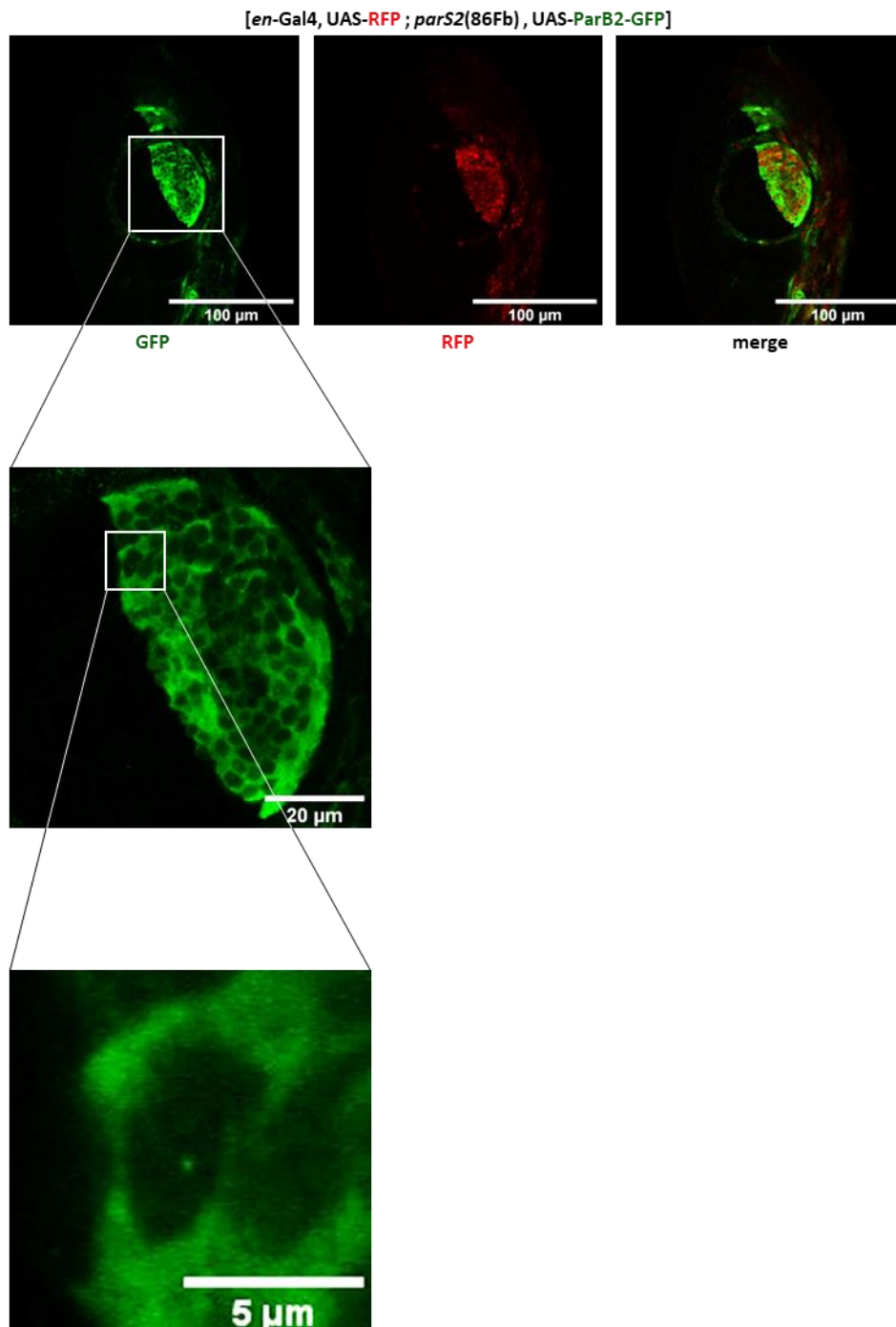
Qu'en est-il dans des tissus où les cellules sont diploïdes ?

Avec l'aide de Jennifer Zanet, j'ai regardé le système ANCHOR dans les disques imaginaux qui sont des structures à partir desquelles les différentes parties de l'insecte adulte, comme les pattes, les ailes, les pièces buccales, sont formées durant le stade pupal. Les cellules de ces disques sont diploïdes. Des disques imaginaux de patte sont disséqués à partir de la lignée suivante :

[*en-Gal4*, UAS-RFP ; *parS2*(86Fb), UAS-ParB2-GFP].

L'expression des protéines ParB2-GFP est dirigée par le pilote *en-Gal4* : la protéine Gal4 est sous le contrôle de la région régulatrice du gène *engrailed* (*en*) qui s'exprime dans la partie postérieure des disques imaginaux. La protéine fluorescente RFP, sous contrôle de séquence UAS, va permettre de visualiser le domaine d'expression du pilote *en-Gal4*.

Dans la **figure 80**, on peut voir l'expression de la protéine ParB2-GFP dans la partie postérieure du disque de patte et en zoomant sur quelques cellules exprimant cette protéine, on peut observer un foyer de fluorescence consécutif à la fixation des protéines ParB2-GFP sur ses séquences *parS2* localisées en 86Fb.



**Figure 80: Visualisation d'un locus génomique dans un disque imaginal de patte.**

Le pilote *en-Gal4* permet d'exprimer les protéines dans la partie postérieure du disque imaginal de patte qui est visualisé par la marquage anti-RFP. L'expression de la protéine ParB2-GFP par le pilote *en-Gal4* permet de visualiser un foyer fluorescent localisé en 86Fb.

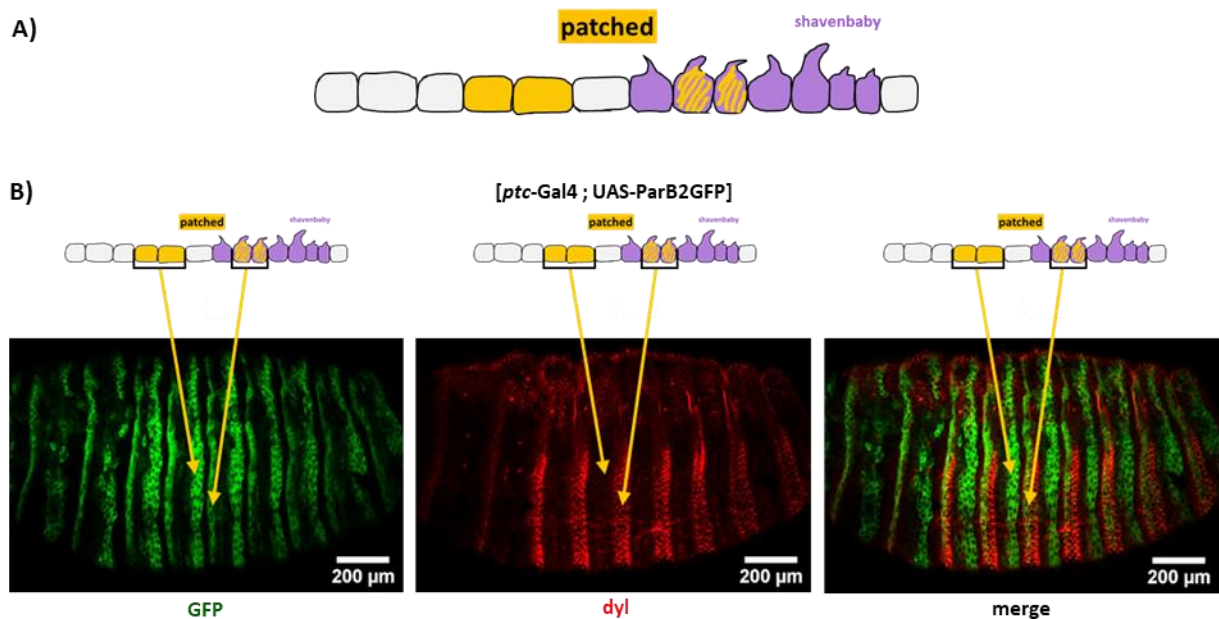
(anti-GFP + anti-RFP ; x63 et encart x63zoom5).

J. Zanet

#### 4-2-2 Dans l'épiderme embryonnaire de la drosophile.

Les cellules embryonnaires épidermiques sont de petites cellules diploïdes. Pour diriger l'expression des protéines ParB dans l'épiderme embryonnaire, j'ai utilisé le pilote *patched-Gal4* (*ptc-Gal4*).

Dans la **figure 81 A**, sont schématisées douze rangées de cellules embryonnaires ventrales de l'épiderme de la drosophile. Six rangées de cellules présentent, en fin de stade embryonnaire, dans leur domaine apical une excroissance cytoplasmique, appelée denticule ou trichrome dont la formation est sous le contrôle du facteur de transcription shavenbaby (*svb*). Son domaine d'expression, dans la partie ventrale de l'épiderme, est représenté en violet. Les six autres rangées de cellules ne forment pas de denticule en raison de l'absence de l'expression du gène *svb*. Ces rangées présentent une cuticule nue. Le gène *patched* s'exprime dans deux rangées de cellules où la cuticule est nue et dans deux rangées dans le domaine d'expression du gène *svb* (**fig. 81 B**). J'ai choisi ce pilote pour son activité dans le territoire où le gène *svb* est actif et dans celui où il ne s'exprime pas.



**Figure 81: Domaine d'expression de *ptc-Gal4*.**

**A)** Domaine d'expression du pilote *ptc-Gal4* au niveau de l'épiderme ventral de l'embryon de la drosophile.  
**B)** Le pilote *ptc-Gal4* permet de diriger l'expression de la protéine ParB2-GFP dans deux rangées de cellules de cuticule nue et dans 2 rangées de cellules dans le domaine d'expression de *Svb* qui permet la formation de denticules (vue latérale; antérieur à gauche, postérieur à droite). Le gène *dyl* est une cible directe du facteur de transcription *Svb* et le domaine d'expression du gène *dyl* phénocopie celui du gène *svb*. (anti-GFP + anti-dyl; x20).

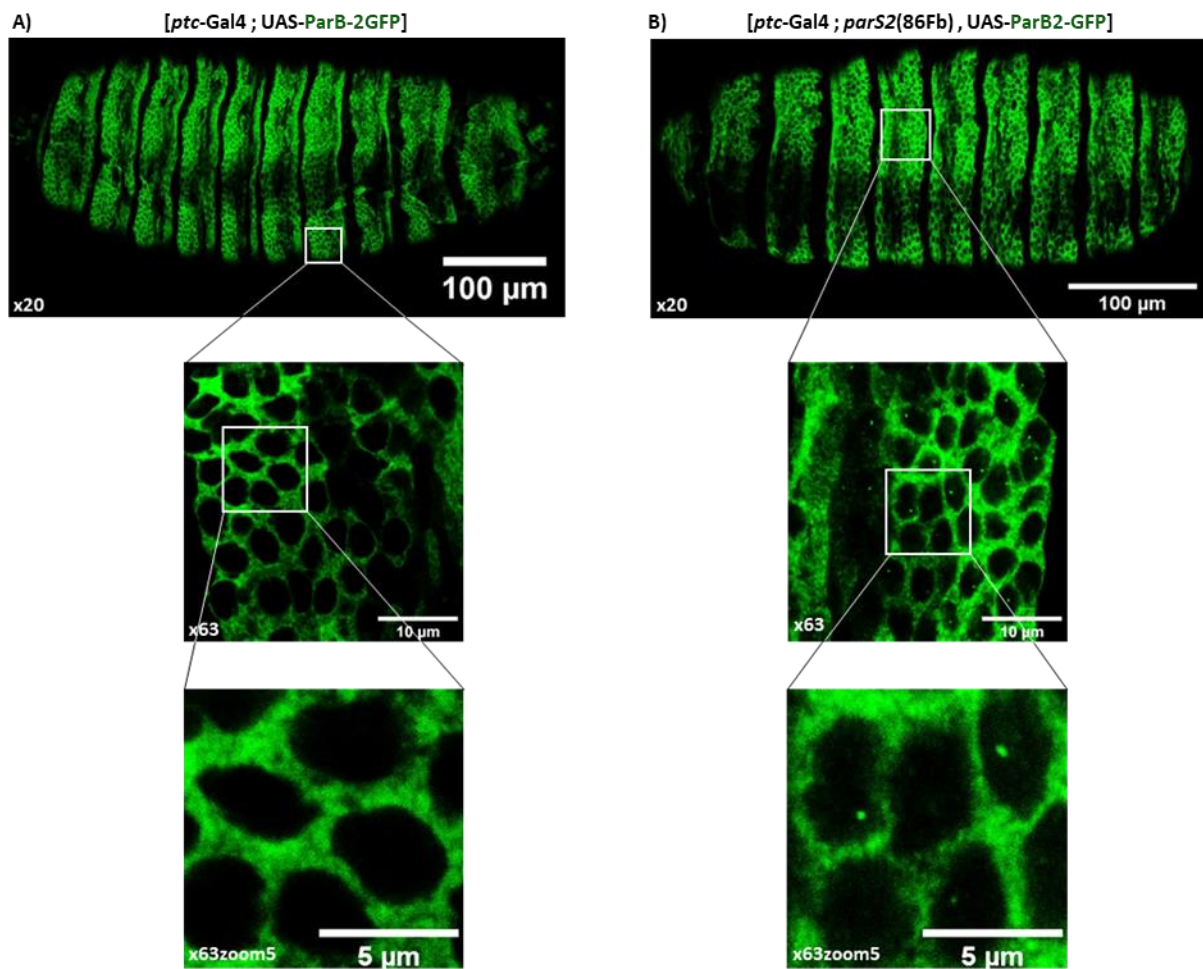
Quand la protéine ParB2-GFP est exprimée grâce au pilote *ptc*-Gal4, on peut observer son expression dans deux rangées de cellules. Pour savoir quelles sont les cellules qui présenteront une cuticule nue et celles qui formeront des denticules, j'ai effectué un marquage contre la protéine Dusky-like (Dyl), dont le gène est une cible directe du facteur de transcription *svb*. Le domaine d'expression de la protéine Dyl phénocopie celui du gène *svb* (**fig. 81 B**).

On peut voir une colocalisation de la protéine PaB2-GFP avec la protéine Dyl indiquant la présence de la protéine ParB2-GFP dans deux rangées de cellules épidermiques qui formeront des denticules. De même, on peut voir un marquage de la protéine ParB2-GFP dans des cellules qui ne présentent pas de marquage contre la protéine Dyl soulignant l'expression de la protéine ParB2-GFP dans des cellules qui auront une cuticule nue (**fig. 81 B**).

Il est à noter que le domaine d'expression de ce pilote *ptc*-Gal4 peut s'étendre au-delà de ces quatre rangées de cellules, notamment lorsqu'il est utilisé à 25°C. L'activité du système Gal4/UAS varie en fonction de la température. Plus la température sera élevée et plus l'expression des gènes sera importante et étendue à d'autres rangées de cellules.

J'ai induit l'expression de la protéine ParB2 fusionnée à la GFP avec le pilote *ptc*-Gal4 dans les cellules de l'épiderme en fin d'embryogénèse (stade 12 à 16). En absence de ses séquences spécifiques de fixation *parS2*, la protéine ParB2-GFP est localisée massivement dans le cytoplasme (**fig. 82 A**), comme dans les cellules des glandes salivaires (§3-2) ou dans les cellules de mammifères (Fok, 2021).

En présence des séquences *parS2* intégrées au génome de la drosophile au niveau de la position cytologique 86Fb, on observe, comme dans les cellules des glandes salivaires, une modification de la distribution intracellulaire des protéines ParB2-GFP et l'apparition d'un spot de fluorescence net qui est visible à l'intérieur du noyau, correspondant à la fixation des protéines ParB2 sur leur séquence de liaison *parS2* (**fig. 82 B**). Les images dans la **figure 82** sont prises avec un seul plan focal. On n'observe pas de spots dans toutes les cellules, car ils sont localisés sur différents plans focaux. Ces images ont été obtenues en condition hétérozygote pour les séquences *parS2* : elles ne sont présentes que sur un seul chromosome. Le génome de la drosophile ne contient donc que deux sites de fixation *parS2* qui sont suffisants pour recruter suffisamment de protéine ParB2 pour créer un foyer de fluorescence.



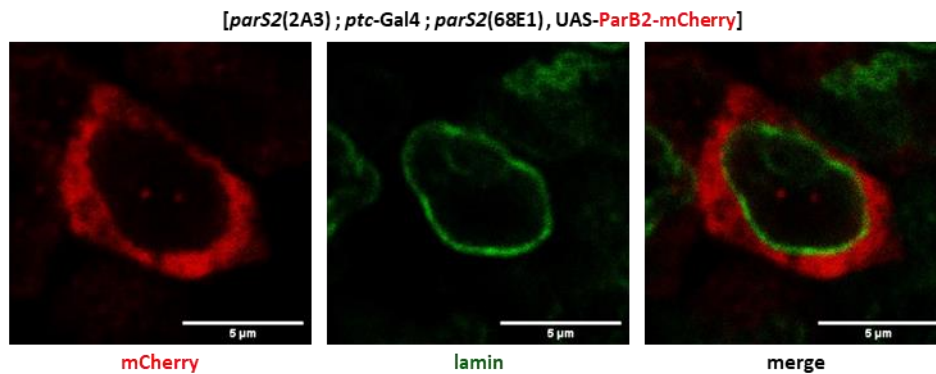
**Figure 82: Visualisation d'un spot de fluorescence dans le noyau de cellules épidermiques.**

En absence des séquences de fixation *parS2* **A**), la protéine ParB2 est exclusivement localisée dans le cytoplasme. En présence des séquences *parS2* **B**), un spot de fluorescence est détecté à l'intérieur du noyau.

(vue ventrale, antérieure à gauche, postérieure à droite; anti-GFP).

Dans la **figure 83**, on peut visualiser deux foyers de fluorescence dans le noyau correspondant à la fixation des protéines ParB2 fusionnées à la protéine fluorescente mCherry sur leurs séquences *parS2* insérées dans le génome de la drosophile en deux positions cytologiques différentes : sur le chromosome X en position 2A3 et sur le chromosome III en position 68E1. Pour chaque séquence *parS*, nous sommes en situation hétérozygote avec une seule copie par chromosome.





**Figure 83: Visualisation de deux loci génomiques différents.**

Le système ParB2/*parS2* permet de visualiser 2 loci génétiques différents grâce à la fixation des protéines ParB2 sur leurs séquences *parS2* localisées en 2 positions cytotogiques différentes (en 2A3 et en 68E1). (anti-mCherry + anti-lamin; x63zoom5).

Le système ANCHOR permet de visualiser des loci génomiques en temps réel dans un organisme pluricellulaire, la drosophile, et tout au long de la vie de l'insecte : de l'embryon à l'adulte en passant par la larve et aussi bien dans des cellules possédant des chromosomes polytènes que dans de petites cellules diploïdes.

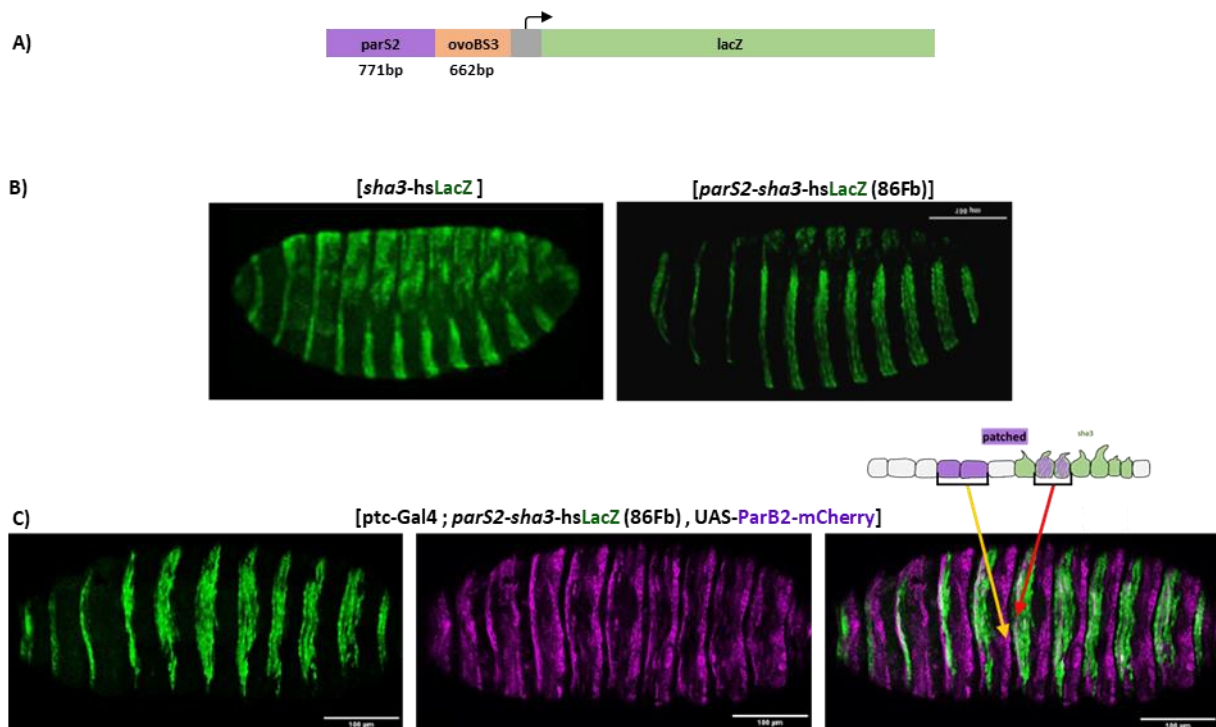
#### **4-3 La fixation des protéines ParB sur leurs séquences *parS* modifie-t-elle l'activité de régions régulatrices ?**

Les systèmes existant pour visualiser des loci génomiques de type FROS sont basés sur l'insertion dans le génome de séquences de grande taille pouvant être supérieures à 10Kb et contenant un nombre important de répétition de séquences (jusqu'à 256). Ces insertions sont intrusives et peuvent modifier l'environnement chromatinien notamment au niveau des régions de régulation (Delker, 2022).

Qu'en est-il du système ANCHOR dont les séquences *parS* sont de petites tailles (inférieures à 1kb) et contiennent peu de séquences répétées ? L'intégration des séquences *parS* localisées au voisinage d'une séquence de régulation a-t-elle une influence sur l'activité de cette région ? Et inversement, l'activité d'une région régulatrice a-t-elle une influence sur la visualisation de locus génomique par le système ANCHOR ?

Afin de répondre à ces questions, j'ai utilisé comme séquence de régulation, la région cis-régulatrice nommée *sha3* du gène *shavenoïd* (*sha*), une cible directe du gène *svb* dont le domaine d'expression récapitule celui de *svb*.

La construction plasmidique que j'ai utilisée pour répondre à cette question est représentée dans la **figure 84 A** et elle est intégrée dans le génome de la drosophile en position 86Fb. Les séquences *parS2* sont situées en 5' de la séquence de régulation *sha3* qui dirige l'expression du gène rapporteur *lacZ*. Le domaine d'expression du gène *lacZ* obtenu avec cette construction, que l'on peut voir dans l'image de droite dans la **figure 84 B**, est similaire à celui qui a été obtenu à partir des séquences *sha3* clonées dans le vecteur pCasperhs43lacZ (image de gauche **fig. 84 B** ; Menoret, 2013). L'introduction des séquences *parS2* en amont de celles de *sha3* ne modifie pas le domaine d'expression de *sha3* (**fig. 84 B**).



**Figure 84: Le domaine d'expression de la région régulatrice *sha3* n'est pas modifié par la présence des séquences *parS2* et/ou des protéines ParB2.**

**A)** Schéma de la construction plasmidique introduite dans le génome de la drosophile à la position 86Fb.

**B)** A gauche, domaine d'expression de la région régulatrice *sha3* du gène *shavenoïd* cloné dans pCasSpeRhs43lacZ (x20; vue latérale; Menoret, 2013). A droite domaine d'expression de la région régulatrice *sha3* à partir de la construction schématisée en **A**) (x20; vue ventrale).

(anti-βGal).

(antérieur à gauche, postérieur à droite).

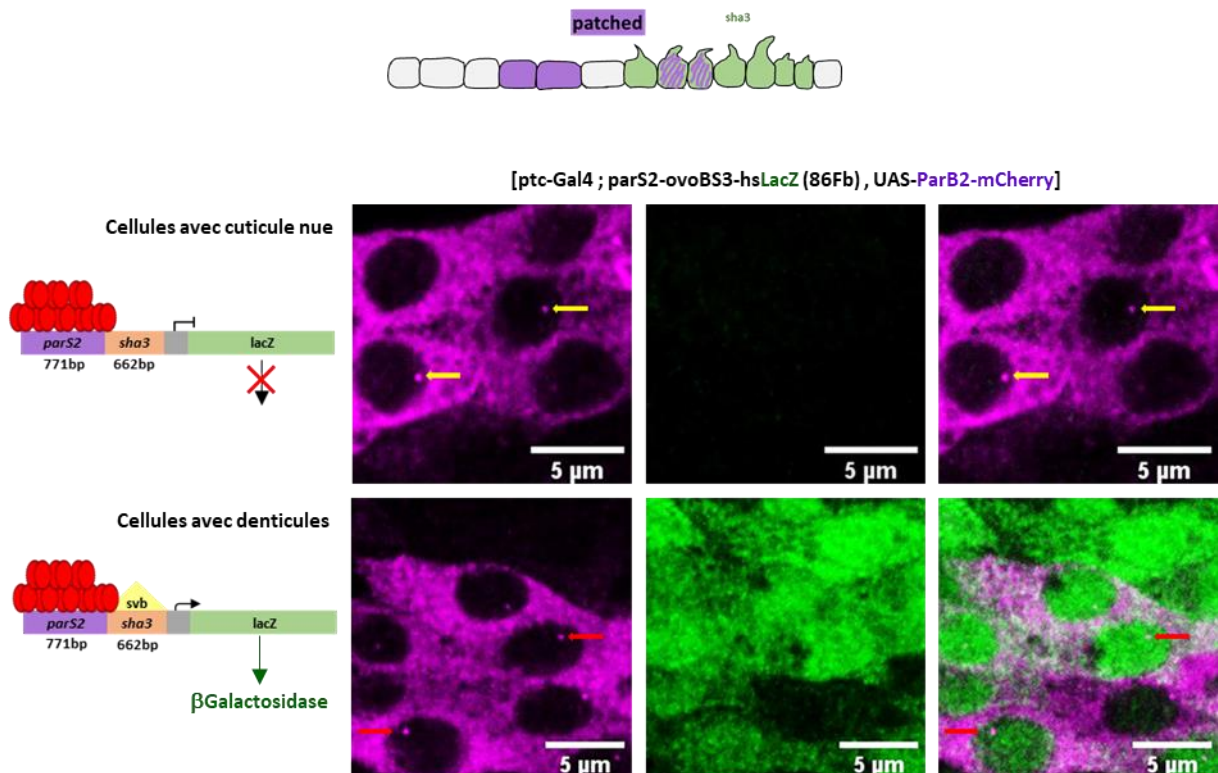
**C)** La présence du système de visualisation ANCHOR ne modifie pas le domaine d'expression de la région régulatrice *sha3* du gène *shavenoïd* (x20; vue ventrale; antérieur à gauche, postérieur à droite).

(anti-βGal + anti-mCherry ).

Le pilote *ptc*-Gal4 permet l'expression des protéines ParB2-mCherry dans deux rangées de cellules qui auront une cuticule nue car le gène *svb* ne s'exprime pas (flèche jaune **fig. 84 C**) et dans deux rangées de cellules qui formeront des denticules car le gène *svb* est actif (flèche

rouge, **fig. 84 C**). L'expression des protéines ParB2 ne modifie pas le domaine d'expression de la région régulatrice *sha3* qui est similaire à celui que l'on peut observer en absence des protéines ParB2 (**fig. 84 B** photo de droite) ou à celui en absence des séquences *parS2* (**fig. 84 B** photo de gauche).

Si on effectue un agrandissement au niveau de ces régions, dans les cellules où la cuticule sera nue, on peut observer dans le noyau un foyer de fluorescence correspondant à la fixation des protéines ParB2 sur leurs séquences *parS2* (flèche jaune, **fig. 85**). Dans le territoire où les cellules formeront une denticule, on peut voir un marquage contre la protéine  $\beta$ Galactosidase qui témoigne de l'activité de la région régulatrice *sha3*. Dans le noyau de ces mêmes cellules, on peut visualiser un foyer de fluorescence (flèche rouge, **fig. 85**).



**Figure 85: La visualisation de locus génomique n'empêche pas l'activité d'une région régulatrice.**

La fixation des protéines ParB2 sur leur site n'a pas d'influence sur l'activité d'une région régulatrice située à proximité. Réciproquement, l'activité d'une région régulatrice située à proximité d'un site *parS2* n'a pas d'influence sur la fixation des protéines ParB2.

(anti- $\beta$ Gal + anti-mCherry; x63zoom5).

La relation entre le système ANCHOR et l'activité d'une région régulatrice peut être résumée par la **figure 86** :

- La fixation des protéines ParB sur leurs séquences *parS* n'empêche pas l'activité de la région de régulation située à proximité.
- L'activité d'une région régulatrice n'empêche pas la fixation des protéines ParB2 sur leurs séquences *parS2* situées à proximité et la création d'un foyer de fluorescence.

L'insertion de séquences *parS* est peu intrusive et la visualisation d'un locus génomique avec le système ANCHOR ne perturbe pas l'activité d'une séquence régulatrice située à proximité.

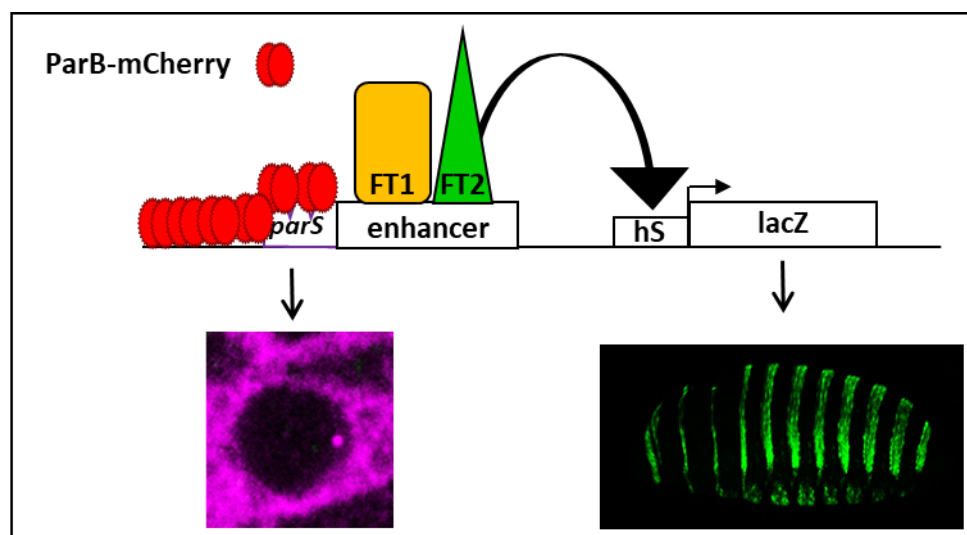


Figure 86: L'activité d'une région régulatrice n'est pas perturbée par la création d'un foyer de fluorescence et la fixation des protéines ParB sur leurs séquences *parS* n'est pas entravée par l'activité d'une région de régulation située à proximité.

## 5- Discussions et perspectives.

Dans cette première partie de ma thèse, nous avons vu la mise en place et l'efficacité du système ANCHOR pour visualiser des loci génomiques en temps réel dans un organisme multicellulaire, la drosophile.

J'ai montré que l'expression des protéines ParB chez la drosophile n'entraîne pas de létalité. Mais la fixation de la protéine ParB sur les séquences *parS* semble indiquer une légère toxicité pour la drosophile. Cette toxicité, toute relative, n'est pas préjudiciable dans l'utilisation de ce système chez la drosophile et n'a pas affecté les expériences que j'ai pu faire avec le système ANCHOR. Au vu des résultats, on peut considérer que le système ANCHOR n'est pas toxique pour la drosophile.

Ce système ANCHOR peut être utilisé pour visualiser des loci génomiques à tous les stades du développement de la drosophile : de l'embryon à l'adulte en passant par les stades larvaires. Chaque fois que j'ai testé ce système ANCHOR dans un tissu, un foyer de fluorescence a pu être observé, que ce soit dans les chromosomes polytènes où chaque locus est représenté jusqu'à plus de mille copies, ou dans des cellules polypoïdes comme dans les cellules principales des tubes de Malpighi, ou encore dans les petites cellules épidermiques diploïdes.

Nous disposons de quatre systèmes ANCHOR différents provenant de différentes origines bactériennes : *Burkholderia cenocepacia*, *Ralstonia pickettii* ou *Escherichia coli*. Ainsi différentes combinaisons peuvent être utilisées simultanément car les différentes protéines ParB se fixent sur des séquences *parS* qui leurs sont spécifiques, même pour des protéines provenant de la même bactérie, comme c'est le cas pour ParB1/*parS1* et ParB2/*parS2*.

Parmi les quatre systèmes ANCHOR dont je dispose, le couple ParB2/*parS2* provenant du complexe de partition du chromosome 3 de *Burkholderia cenocepacia* est celui qui s'est révélé le plus performant. Il peut être utilisé simultanément avec deux autres couples : le couple ParB1/*parS1* provenant du complexe de partition du chromosome 2 de *Burkholderia cenocepacia* et le couple ParB4/*parS4* d'*Escherichia coli* pour la partition du plasmide F. Or les séquences *parS2* ne contiennent que deux sites de fixation pour les protéines parB2 comparées aux autres séquences *parS*.

Pour les quatre systèmes ANCHOR, l'expression des protéines ParB fusionnées à une protéine fluorescente, en absence de leur séquence de fixation *parS*, est exclusivement cytoplasmique, quel que soit le tissu où la protéine est produite. En présence des séquences *parS*, la distribution intracellulaire de la protéine ParB est modifiée et un foyer de fluorescence est observé dans le noyau, uniquement au niveau du locus où la séquence *parS* se trouve. C'est cet excellent rapport fluorescence/bruit de fond qui permet de visualiser ce foyer de fluorescence.

En comparant avec les autres stratégies existantes pour visualiser un locus génomique, comme le système *LacO/LacI*, un des avantages notables du système ANCHOR est l'utilisation de séquences de petite taille (moins de 800bp pour *parS2*) et sans élément répété, comparé aux multiples répétitions du site *LacO*. En effet, au-delà de la facilité pour construire les éléments plasmidiques, l'intégration de ces séquences *parS* dans le génome de la drosophile est peu intrusive sur l'activité chromatinien environnant. Ainsi la présence de séquence *parS* au voisinage de région régulatrice n'a pas d'influence sur leur activité et inversement l'activité de région régulatrice n'empêche pas la création d'un foyer de fluorescence. Ces résultats sont confortés par ceux obtenus par Thomas Germier dans l'équipe de Kerstin Bystricky. Dans les cellules de mammifères qu'il utilise, il a montré que le système ANCHOR n'interfère pas avec les processus de réplication ou de transcription (Germier, 2018).

Les outils du système ANCHOR que j'ai développés pour leur utilisation dans la drosophile ont été transmis à deux équipes, celle de Thomas Gregor et celle Sarah Bray. Les résultats de leurs travaux respectifs ont permis de confirmer et de renforcer le système ANCHOR comme un système performant pour visualiser des loci génomiques en temps réel dans des cellules vivantes.

Thomas Gregor s'intéresse à la communication entre les régions régulatrices et leurs promoteurs cibles et étudie plus spécialement comment la topologie de la chromatine participe activement à l'activation des gènes. Il combine des techniques d'édition du génome et des techniques d'imageries en temps réel dans des cellules vivantes pour visualiser simultanément l'interaction entre séquences régulatrices et promoteur et transcription. Pour répondre à ces questions, il a mis à profit le développement d'outils de visualisation de locus génomique en temps réel, le système ANCHOR développé par l'équipe de Kerstin Bystricky et que j'ai mis en place chez la drosophile, qu'il a combiné avec le système MS2/MCP permettant

de suivre la synthèse des transcrits primaires. Dans son article publié en 2018 (Chen, 2018), il a caractérisé, à l'échelle de la cellule, l'interaction dynamique entre la topologie d'une région régulatrice, son promoteur et l'activité transcriptionnelle. Développant un système d'imagerie à trois couleurs, il a visualisé simultanément la localisation d'une région régulatrice endogène, la localisation du promoteur d'un gène rapporteur où il a inséré une séquence *parS* et son activité transcriptionnelle dans des embryons vivants. Il a observé des foyers de fluorescence ParB dans tous les noyaux (diploïdes) de l'embryon en cours de développement, permettant de suivre la position et le mouvement du locus où se trouve le gène rapporteur. En dehors des autres conclusions de cet article, il a renforcé l'utilisation du système ANCHOR pour visualiser des loci génomiques en temps réel dans des cellules vivantes.

Sarah Bray s'intéresse à la voie de signalisation Notch et dans son article publié en 2018, en collaboration avec l'équipe de Kerstin Bystricky et la nôtre (Gomez-Lamarca, 2018), elle regarde la réactivité de la voie de signalisation Notch au niveau transcriptionnel à travers une analyse *in vivo* en temps réel du facteur de liaison à l'ADN CSL ou Suppressor of Hairless (Su(H)). Elle met en évidence que l'activation de la voie de signalisation Notch *in vivo* provoque des changements dans la dynamique nucléaire de CSL, par une analyse *in vivo* des mécanismes sous-jacents de la réponse transcriptionnelle à la voie de signalisation Notch. Ainsi à l'aide du système ANCHOR, elle a pu visualiser en temps réel le locus du gène Enhancer of split, cible du gène Su(H), en y insérant par CRISPR/Cas9 la séquence *parS*.

**Les résultats que j'ai obtenus au cours de ma thèse sur la mise en place du système ANCHOR chez la drosophile, combinés aux travaux de ces deux équipes utilisant les outils que j'ai développés, permettent de valider chez la drosophile, ce système de visualisation de locus génomique en temps réel.**

En février 2022, Richard Mann a publié un article dans lequel il développe des outils pour marquer des loci génomiques dans des disques imaginaux avec trois systèmes de protéines se liant à l'ADN (Delker, 2022) : LacI/*LacO*, ParB1/*parS1* et ParB2/*parS2*. Il a optimisé le système LacI/*LacO* en diminuant le nombre de site nécessaire pour permettre d'observer un foyer de

fluorescence suffisamment fort pour sortir du bruit de fond. Il a développé ses propres couples ParB1/*parS1* et ParB2/*parS2*. Par rapport au système ANCHOR développé par l'équipe de Kerstin Bystricky, il y a certaines différences au niveau de la structure des protéines utilisées ou dans la manière de les utiliser. Au niveau des trois protéines, LacI et les deux protéines ParB, il a introduit une séquence de localisation au noyau (NLS) en position C-terminale des protéines ainsi qu'un connecteur entre ces trois protéines et la protéine fluorescente.

L'expression de ces protéines est sous le contrôle d'un promoteur inductible par la chaleur (heat-shock-sensitive hsp70 promoteur ou hsp70). Il induit la synthèse des protéines par un choc thermique de quinze minutes puis il dissèque les disques imaginaires au bout de 4 heures. Il compare ces trois systèmes et il considère que son système ParB2/*parS2* est le plus performant. Dans son étude, il n'a effectué que des marquages sur tissus fixés et il n'a pas testé ses systèmes en cellules vivantes. Un des problèmes dans l'utilisation du système LacI/*LacO* est l'insertion de large séquences *LacO* dans le génome. Richard Mann a regardé si son système LacI/*LacO* optimisé pouvait perturber la transcription. Contrairement à ses systèmes ParB1/*parS1* ou ParB2/*parS2*, la fixation de LacI perturbe la transcription (Delker, 2022).

Pour observer un foyer fluorescent au niveau d'un locus génomique, il faut que le signal du focus soit supérieur à celui du bruit de fond. Dans le système ANCHOR, ce rapport signal/bruit de fond est excellent notamment parce que les protéines ParB, en absence de leur séquence *parS*, sont localisées dans le cytoplasme. Les protéines ParB ne restent dans le noyau que lorsqu'elles trouvent leurs séquences de fixation, les séquences *parS*, sur lesquelles elles vont se stabiliser, s'agglomérer, s'étendre et accumuler suffisamment de fluorescence pour constituer un foyer visible en microscopie.

Dans les systèmes développés par Richard Mann, pour visualiser en temps réel un locus génomique, pour suivre la dynamique des changements au niveau de la chromatine, est-ce que quinze minutes d'induction de la synthèse des protéines ParB seront suffisantes pour faire perdurer le signal ? si l'induction devait durer plus longtemps est-ce que l'accumulation des protéines ParB dans le noyau permettrait toujours d'observer un foyer de fluorescence ?



Au-delà de la visualisation de loci génomiques, le système ANCHOR pourrait permettre d'identifier des facteurs trans-régulateurs se fixant sur des régions régulatrices de l'activité de gène, en se servant des propriétés de fixation spécifiques de la protéine ParB sur ses séquences *parS*, de faible taille (inférieur à 1kb) et sa capacité à s'agglomérer, pour développer une approche permettant la purification et l'identification de ces facteurs trans-régulateurs.

J'aimerais développer une stratégie de « ChiP-reverse » utilisant les propriétés du système ANCHOR pour permettre l'identification biochimique de facteurs de transcription et des protéines associées à une région de régulation. La mise au point de cette approche s'effectuera en culture de cellules S2. Des lignées stables de cellules S2 seront établies ayant intégré dans leur génome les séquences permettant l'expression des protéines ParB fusionnées à une molécule fluorescente, ainsi que les séquences de fixation *parS* situées en amont d'une région de régulation d'un gène. Pour mettre au point cette technique, j'utiliserai une région régulatrice du gène *polished rice* ou *pri*, qui est contrôlée par la fixation du récepteur à l'ecdysone. Pour activer cette région régulatrice, il suffit d'ajouter de l'ecdysone dans le milieu de culture. Cette région servira de témoin positif. Pour purifier les protéines se fixant sur cette région de régulation, j'utiliserai le principe de l'immunoprécipitation de la chromatine ou CHIP où un anticorps dirigé contre la GFP ou la mCherry fusionnée à la protéine ParB permettra de titrer le complexe [ParB/*parS*/Protéines/région-régulatrice]. L'identification des protéines liées à cette région s'effectuera par spectrométrie de masse. La présence du récepteur à l'ecdysone dans nos résultats me permettra de valider cette approche.

La faisabilité démontrée en culture de cellule, j'appliquerai cette approche *in vivo* chez la drosophile en intégrant par CRISPR les séquences *parS* près d'un élément de régulation.

**Chapitre IV : Caractérisation de structures  
cytoplasmiques pseudo-nucléaires dans  
les cellules S2.**

## 1- Développement du système ANCHOR en cellules S2.

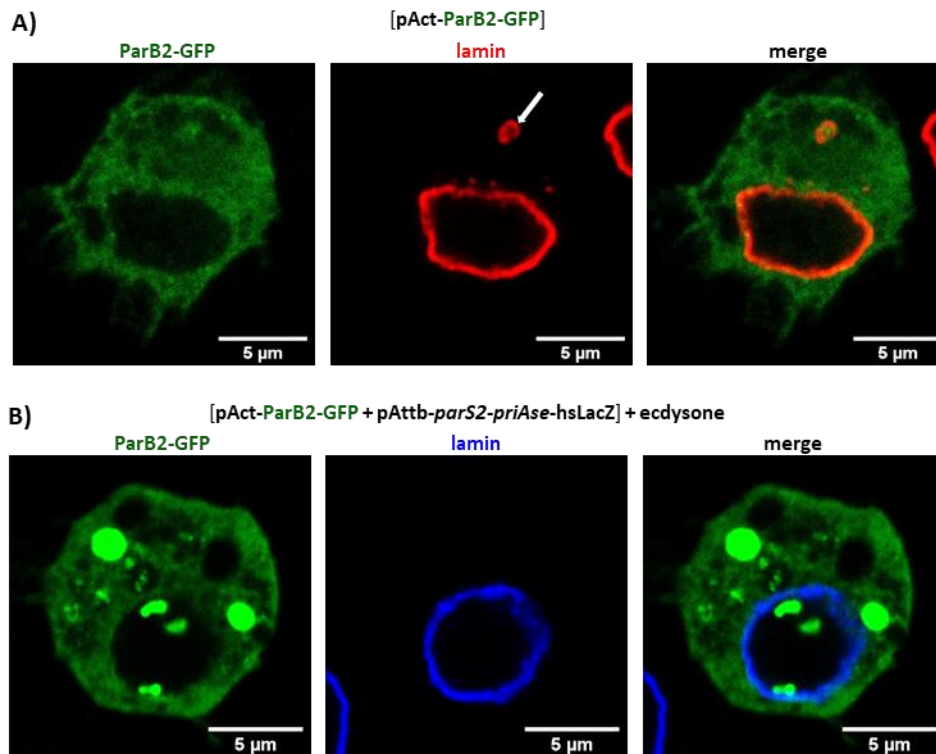
### 1-1 Distribution intracellulaire des protéines ParB2 en présence des séquences *parS*.

Nous avons vu dans le chapitre précédent la mise en place du système ANCHOR dans un organisme entier multicellulaire, la drosophile et avec quelle efficacité il est possible de visualiser un locus génomique que ce soit dans de petites cellules diploïdes comme les cellules embryonnaires de l'épiderme ou dans des grosses cellules polyploïdes comme les cellules des glandes salivaires avec ses chromosomes polytènes possédant plus de mille copies pour chaque locus.

J'ai utilisé les cellules en culture S2 pour lesquelles nous possédons un grand nombre d'outils, facilement manipulables, afin de tester plus en détails comment le système se comporte vis-à-vis de la régulation transcriptionnelle.

J'ai transfecté de manière transitoire dans des cellules S2 un plasmide permettant l'expression de la protéine ParB2 fusionnée à la GFP. En absence des séquences *parS*, la protéine ParB2 est principalement localisée dans le cytoplasme. La lamine A est utilisée pour visualiser la membrane nucléaire (**fig. 87 A**).

En présence d'un plasmide contenant des séquences de fixation *parS2*, on peut visualiser des foyers de fluorescence localisés dans le noyau **et/ou** dans le cytoplasme correspondant à la fixation des protéines ParB sur leurs séquences *parS* (**fig. 87 B**). Je reviendrai ultérieurement sur les structures identifiées dans le cytoplasme (Partie 2 de ce chapitre). Le nombre, la taille et la localisation des spots de fluorescence varient d'une cellule à l'autre. Le comportement des protéines ParB est similaire que ce soit dans un organisme multicellulaire comme la drosophile ou en culture de cellules : en absence des séquences *parS*, les protéines ParB ont une expression diffuse dans le cytoplasme. Intégrées dans le génome de la drosophile ou sur des plasmides transfectés de manière transitoire dans des cellules en culture (c'est à dire non intégrés dans le génome des cellules), la présence des séquences *parS* modifie la distribution intracellulaire des protéines ParB qui vont aller se fixer dans le noyau et/ou dans le cytoplasme sur ces séquences et former des foyers de fluorescence.



**Figure 87: Un site *parS* est suffisant pour modifier la distribution intracellulaire des protéines ParB.**

**A)** En transfection transitoire dans des cellules S2 de drosophile, la protéine ParB a une expression diffuse, exclusivement dans le cytoplasme.

(GFP + anti-lamin; x63 zoom5).

**B)** La présence de plasmides contenant la séquence de fixation *parS2* modifie la distribution de la protéine ParB et la fixation de ParB2 sur sa séquence permet de visualiser un spot de fluorescence localisé dans le cytoplasme et/ou dans le noyau.

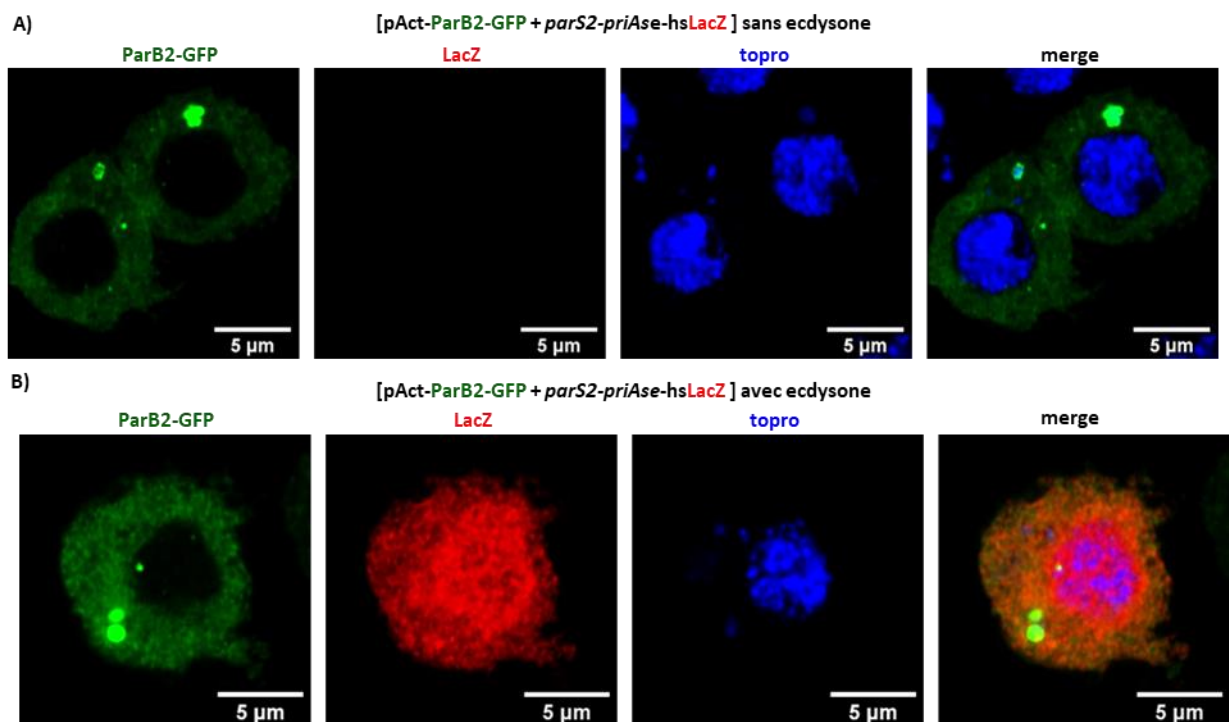
(GFP + anti-lamin; x63 zoom5).

## 1-2 Effet du système ANCHOR sur l'activité transcriptionnelle.

J'ai voulu tester si le système ANCHOR peut modifier l'activité transcriptionnelle. Pour cela, j'ai utilisé le système MS2/MCP, fourni par Mounia Lagha de l'Institut de Génétique Moléculaire de Montpellier, qui permet de détecter les transcrits primaires. Je vais utiliser la séquence régulatrice du gène *pri*, *priAse* qui est très sensible à l'ecdysone (Shlyeuva, 2014) et des travaux dans l'équipe ont montré que le récepteur de l'ecdysone se fixe directement sur la région régulatrice *priAse* (Dib, 2021). Le gène *pri* n'est pas exprimé dans les cellules S2. Je vais tout d'abord regarder comme se comporte cette séquence régulatrice dans des cellules S2 en présence ou en absence d'ecdysone.

J'ai co-transfecté un plasmide contenant cette séquence *priAse* contrôlant l'expression du gène rapporteur *lacZ* et des séquences *parS2* avec un plasmide exprimant les protéines parB2. En présence d'ecdysone (**fig. 88 B**), le gène *lacZ* est exprimé et on peut détecter un fort signal de la protéine  $\beta$ galactosidase. En absence d'ecdysone (**fig. 88 A**), on ne détecte pas d'expression de la protéine  $\beta$ galactosidase.

Ainsi, bien que le gène *pri* ne s'exprime pas dans les cellules S2, sa région de régulation *priAse* permet de réguler l'expression d'un gène rapporteur en présence d'ecdysone.



**Figure 88: La séquence régulatrice *priAse* du gène *pri* permet l'expression d'un gène rapporteur dans les cellules S2 en présence d'ecdysone.**

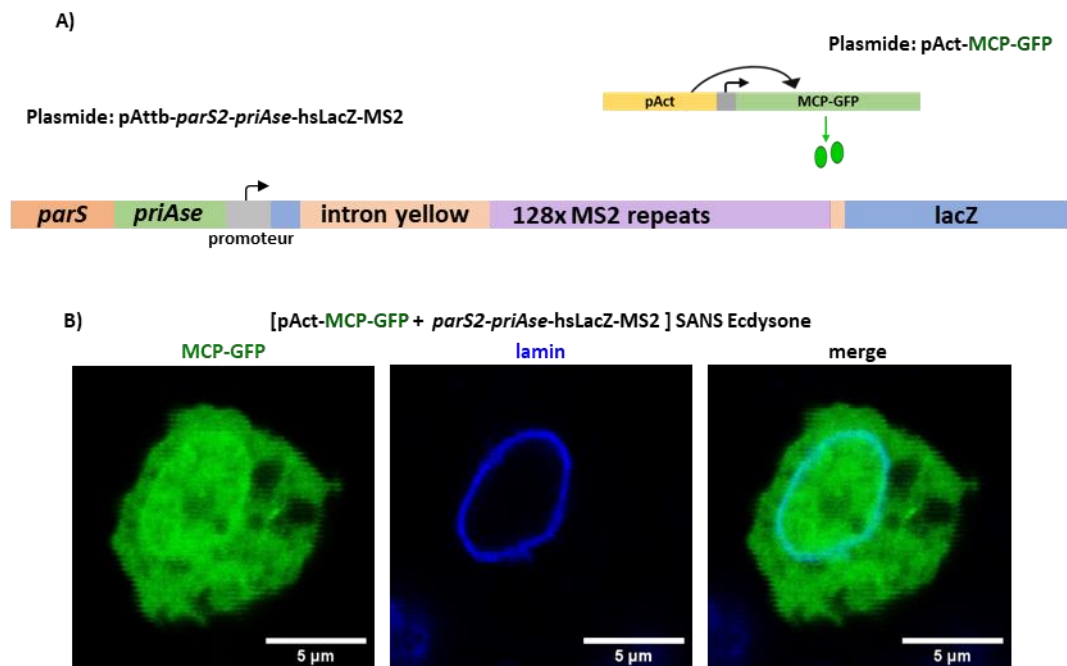
En absence d'ecdysone, le gène rapporteur *LacZ* n'est pas exprimé **A**). En présence d'ecdysone **B**), un fort signal de la protéine  $\beta$ galactosidase est détecté.

( anti-GFP + anti- $\beta$ Gal + topro; x63 zoom 5).

Je vais utiliser cette séquence régulatrice car elle présente l'avantage que pour activer l'expression d'un gène rapporteur, il suffit de rajouter dans le milieu de culture de l'ecdysone. Il n'est pas nécessaire de co-transfecter un plasmide exprimant un facteur de transcription permettant l'expression d'un gène rapporteur.

J'ai introduit une séquence intronique ainsi que les séquences *MS2* au niveau du gène rapporteur *lacZ*, sous le contrôle de la séquence régulatrice *priAse* (**Fig. 89 A**). J'ai cotransfecté cette construction avec le plasmide [pAct-MCP-GFP] qui permet d'exprimer la protéine MCP fusionnée à la protéine fluorescente GFP.

Dans les conditions « contrôle » (sans Ecdysone), donc sans activité transcriptionnelle, la protéine MCP se localise uniformément dans la cellule (**fig. 89 B**).



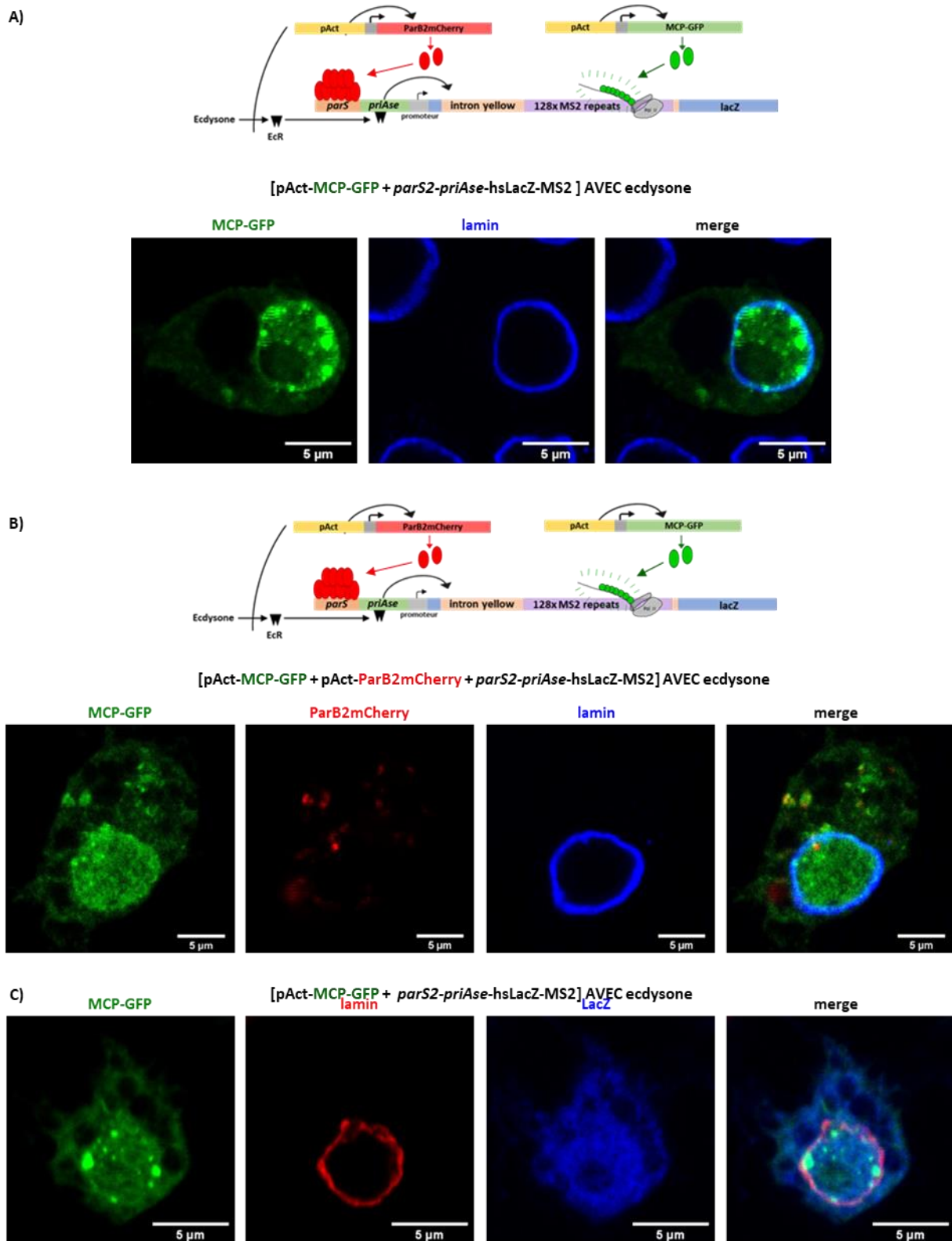
**Figure 89: Utilisation du système MS2/MCP pour analyser l'effet du système ANCHOR sur l'activité transcriptionnelle.**

**A)** Introduction de séquences introniques du gène *yellow* et des séquences *MS2* au niveau du gène rapporteur *lacZ* (schéma adapté de Ferraro, 2016).

**B)** En absence d'ecdysone, l'expression de la protéine MCP en cellule S2 est diffuse à l'ensemble de la cellule.

(GFP + anti-lamin; x63zoom5).

En présence d'ecdysone dans le milieu, le récepteur de l'ecdysone EcR est activé et se fixe sur des séquences contenues dans la région *priAse* induisant l'expression du gène rapporteur *LacZ* qui contient des séquences *MS2*. La protéine MCP se fixant sur les séquences répétées *MS2* néo-transcrites, s'accumule notamment dans le noyau (**fig. 90 A**). Ainsi, la présence d'ecdysone permet au travers de la région régulatrice *priAse* de promouvoir l'activité transcriptionnelle et la synthèse des transcrits primaires du gène rapporteur *lacZ-MS2* qui sont détectés par la protéine MCP (**fig. 90 A**).



**Figure 90: Effet du système ANCHOR sur l'activité transcriptionnelle.**

A) La protéine MCP-GFP se fixe sur les séquences MS2 contenues dans les transcrits primaires du gène *LacZ* en présence d'ecdysone.

(GFP + anti-lamin; x63zomm5).

B) La protéine ParB co-localise avec la protéine MCP au niveau de foci, indiquant que la protéine parB2 ne perturbe pas l'activité transcriptionnelle.

(GFP + mCherry + anti-lamin; x63zomm5).

C) Les transcrits primaires *LacZ* contenant les séquences MS2 sont traduits correctement et sont détectés par un marquage avec l'anticorps anti-βGalactosidase.

(GFP + anti-lamin + anti-βGal; x63zomm5).

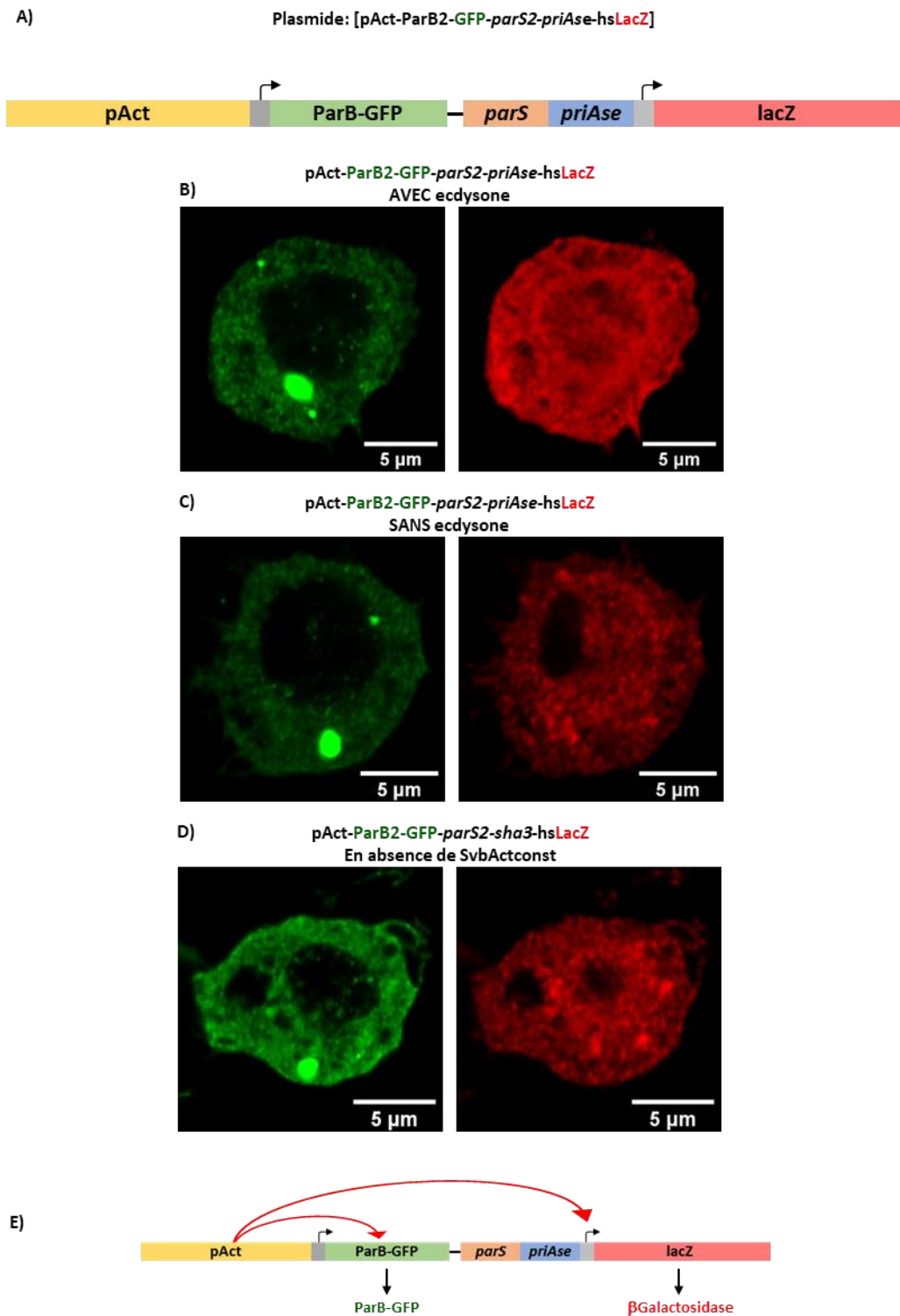
Pour déterminer si une activité transcriptionnelle peut avoir lieu en présence du système ANCHOR, j'ai ajouté à ces conditions expérimentales, un plasmide permettant la synthèse de la protéine ParB2 fusionnée à la protéine fluorescente mCherry. Le plasmide [pAct-ParB2-mCherry], sous le contrôle du promoteur actine, va exprimer la protéine ParB2-mCherry qui va se fixer sur ses séquences *parS2* et qui permettra de localiser le plasmide produisant les transcrits primaires contenant les séquences *MS2*. La protéine ParB2-mCherry co-localise avec la protéine MCP sous forme de foci et ne perturbe pas l'activité transcriptionnelle révélée par la présence de la protéine MCP (**fig. 90 B**). Ces expériences suggèrent que l'activité transcriptionnelle n'est pas affectée par l'activité du système ANCHOR. J'ai également testé que la traduction des transcrits générés s'effectue correctement. Pour cela j'ai suivi la présence de la protéine  $\beta$ -Galactosidase dont l'expression est sous le contrôle de la région régulatrice *priAse*. En absence d'Ecdysone, sans activité transcriptionnelle la protéine  $\beta$ -Galactosidase n'est pas produite. En présence d'Ecdysone, les transcrits produits permettent la synthèse de la protéine (**fig. 90 C**).

**L'ensemble de ces résultats permet de conclure que le système ANCHOR est compatible avec une activité transcriptionnelle des éléments situés à proximité des séquences *parS*.**

### **1.3 Effet du système ANCHOR sur les séquences insulatrices.**

Afin de limiter le nombre de plasmides transfectés dans les cellules, j'ai construit un plasmide comprenant les deux composants du système ANCHOR : la protéine ParB-GFP sous le contrôle du promoteur fort actin5C, la séquence *parS* ainsi que la séquence codant pour la protéine  $\beta$ -Galactosidase, sous le contrôle de la séquence régulatrice *priAse* (**fig. 91 A**). En présence d'ecdysone, le gène rapporteur *LacZ* est exprimé et un fort signal de la protéine  $\beta$ -Galactosidase est observé (**fig. 91 B**). Cependant, en absence d'ecdysone, j'ai pu détecter une expression anormale du gène rapporteur *LacZ*, qui devrait être nulle (**fig. 91C**). J'ai effectué la même expérience mais en utilisant la région régulatrice *sha3* à la place de la séquence *priAse*. La région *sha3* en absence de la forme activatrice du gène *svb* ne permet pas d'activer la





**Figure 91: Dans un plasmide contenant deux cadres de lecture, le premier agit en trans sur le second.**

A) Schéma du plasmide.

B) En présence d'ecdysone, le gène rapporteur est exprimé.

C) En absence d'ecdysone, l'expression du gène *lacZ* est dû à l'action du promoteur actine. Le promoteur actine agit sur le deuxième cadre de lecture ouvert pour permettre l'expression du gène *LacZ*.

D) Le même résultat est obtenu avec la région régulatrice *sha3*.

(anti-GFP + anti-βGal; x63 zoom5).

E) Schéma récapitulant les résultats de l'expérience.

transcription d'un gène rapporteur. J'obtiens le même résultat : expression du gène rapporteur en absence du facteur activateur (**fig. 91 D**).

Cette expression « parasite » du gène *LacZ* doit être liée à la présence du promoteur actin5C : le gène *lacZ* subit l'influence de l'activité transcriptionnelle associée à la synthèse de la protéine ParB2 (schéma **fig. 91 E**). Un mal pour un bien ! J'ai détourné l'utilisation de cette construction pour analyser si la présence de séquences insulatrices de part et d'autre des séquences contenant [*parS2-priAse-hsLacZ*] pourraient les isoler de l'activité du promoteur actin5C et indiquer si des plasmides non intégrés au génome bactérien pourraient avoir une certaine organisation.

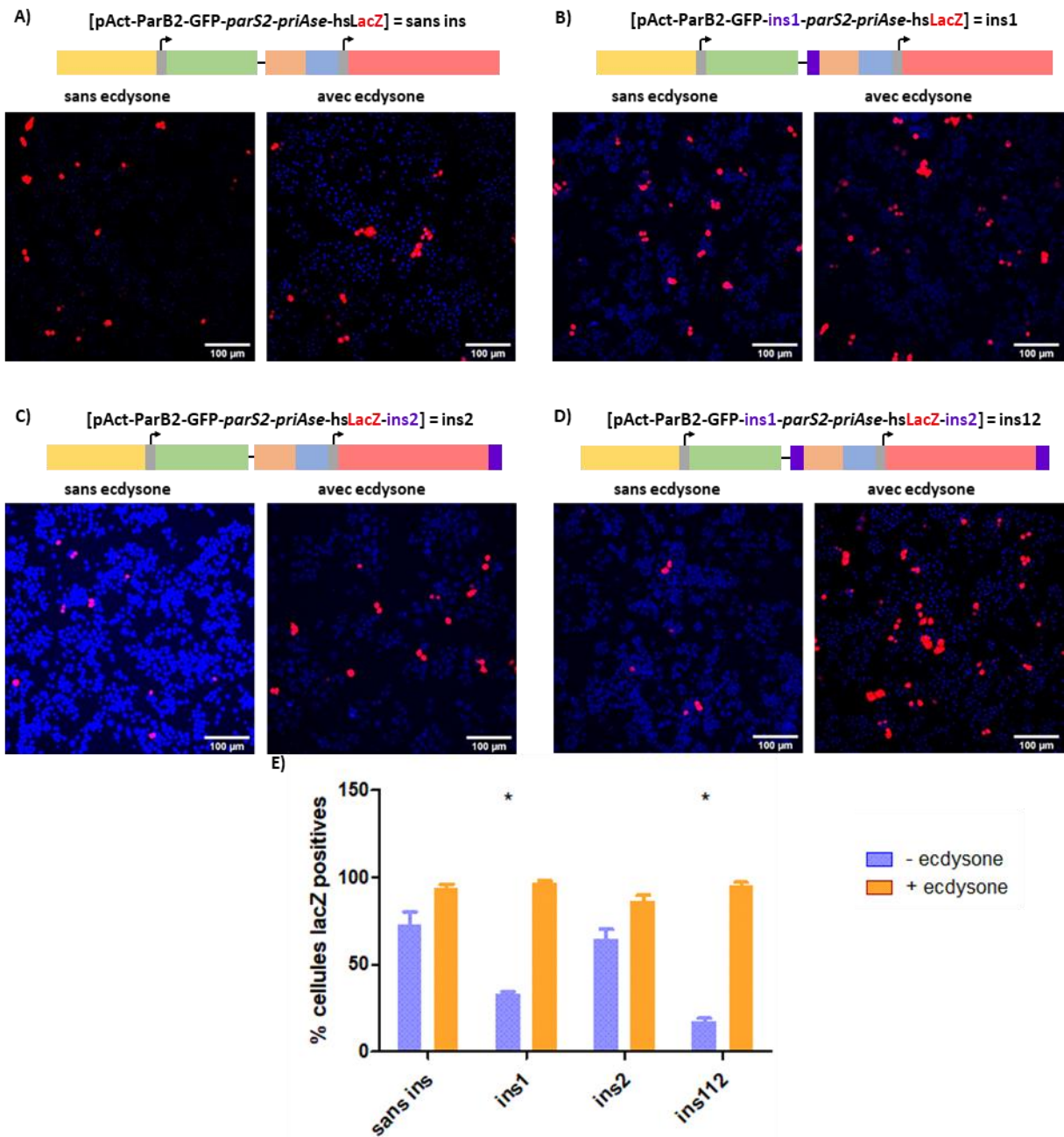
J'ai introduit des séquences insulatrices *gypsy* en 5' des séquences [*parS2-priAse-hsLacZ*] (**fig. 92 B = ins1**) et/ou en 3' de ces séquences (**fig. 92 C = ins2** et **fig. 92 D = ins12**). J'ai transfecté ces constructions ainsi que celle sans séquence insulatrice, comme contrôle (**fig. 92 A = sans ins**). J'ai mesuré l'activité transcriptionnelle de *priAse* en comptant le nombre de cellules qui expriment la protéine  $\beta$ galactosidase en présence ou en absence d'ecdysone (**fig. 92 E**).

En absence de séquence insulatrice (**A = sans ins**) ou avec une séquence insulatrice insérée en 3' (**C = ins2**), il n'y a pas de différence significative dans le nombre de cellules exprimant le gène rapporteur *LacZ*. En absence d'ecdysone, l'expression de la protéine  $\beta$ galactosidase est due à l'activité du promoteur actin5C. La présence de séquences insulatrices en 3' ne bloque pas l'influence transcriptionnelle du promoteur actin5C sur l'expression du gène *lacZ* (**fig. 92 E**).

En revanche, en présence de séquences insulatrices positionnées en 5' (**B = ins1**) ou en 5' et en 3' (**D = ins12**), on note une différence significative dans le nombre de cellules exprimant le gène rapporteur *LacZ* entre absence et présence d'ecdysone ; le nombre de cellules exprimant le gène *LacZ* est beaucoup plus faible en absence d'ecdysone qu'en présence d'ecdysone. Cette différence est d'autant plus prononcée que les séquences [*parS2-priAse-hsLacZ*] sont flanquées des séquences insulatrices en 5' et 3' (**fig. 92 E**).

Ces résultats montrent que l'introduction des séquences insulatrices en 5' ou en 5' et en 3' du deuxième cadre de lecture ouverte protège l'expression du gène *LacZ* des activités transcriptionnelles du promoteur actin5C. Ainsi la présence des séquences *gypsy* indique

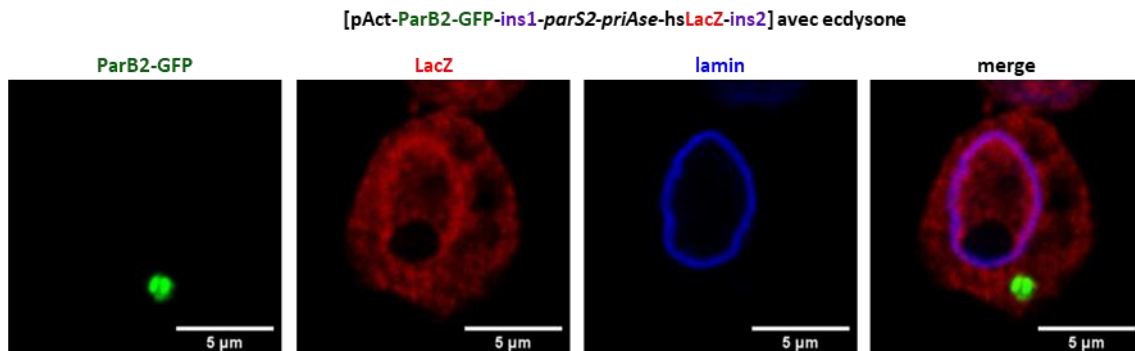
qu'un plasmide peut acquérir une certaine organisation lui permettant d'isoler un cadre de lecture ouverte de l'influence d'activités transcriptionnelles situées en dehors de ce cadre.



**Figure 92: Les séquences *gypsy* peuvent isoler les séquences [*parS*-*priAse*-hsLacZ] de l'activité du promoteur actin5C.**

En absence de séquence isolatrice, le promoteur actin5C induit l'expression du gène *LacZ* **A)**. L'introduction d'isolateur en 5' **B)** ou de part et d'autre de la deuxième ORF **D)** permet isoler cette région de l'effet du promoteur actin5C. La présence d'isolateur en 3' a peu d'effet **C)**. **E)** Quantification du nombre de cellules exprimant le gène *LacZ* en fonction de la présence ou non d'isolateurs. (anti-βGal + topro; x20).

L'ensemble de ces expériences a été réalisé en présence de la protéine ParB2-GFP et de la séquence *parS2* localisée en amont de la séquence régulatrice *priAse* et entre les insulateurs. Dans la **figure 93**, on peut voir un foyer de fluorescence vert correspondant à la fixation de la protéine ParB2 sur ses séquences *parS2*.



**Figure 93: Le système ANCHOR ne perturbe pas l'activité des séquences insulatrices.**

La fixation de la protéines ParB2 sur ses séquences *parS* ne perturbe pas l'activité des séquences insulatrices *gypsy*.

( GFP + anti-LacZ + anti-lamin; x20).

De ces expériences, nous pouvons conclure, (i) la présence des séquences insulatrices *gypsy* n'empêche pas la fixation de la protéine ParB2 sur ses séquences *parS*, (ii) le système ANCHOR ne perturbe pas l'activité des séquences insulatrices, (iii) la présence des séquences *gypsy* souligne que les plasmides peuvent adopter une conformation structurale leur permettant d'isoler une région promotrice de l'influence d'autres promoteurs.

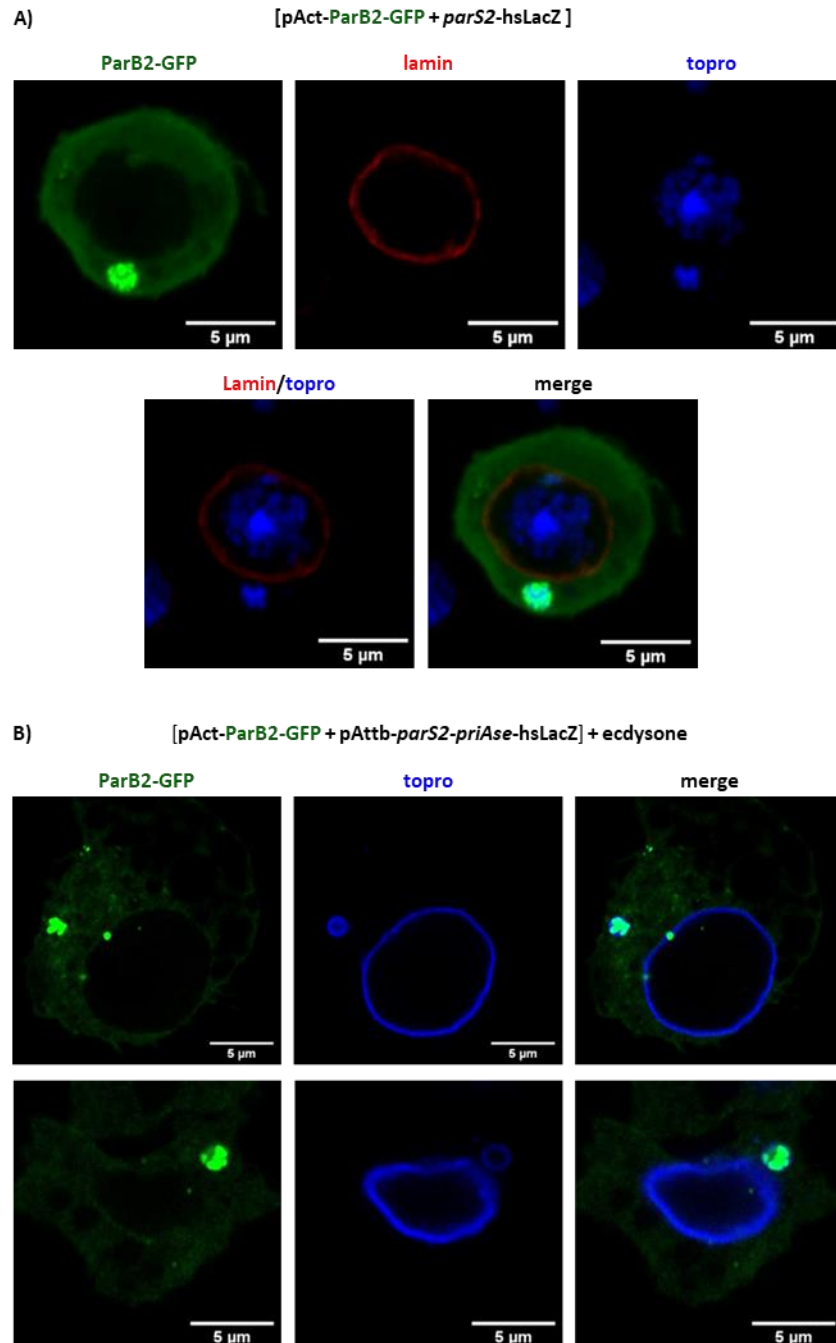
## 2- Structures cytoplasmiques pseudo-nucléaires en cellules S2.

Comme je l'ai mentionné précédemment, lors de l'utilisation du système ANCHOR dans les cellules S2 en transfection transitoire, j'ai observé la présence de foyers de fluorescence dans le noyau. J'ai également observé des foyers de fluorescence dans le cytoplasme qui ont attiré mon attention (**Fig. 87**). Dans cette partie, je me suis attaché à caractériser ces structures.

### 2.1 Identification des structures pseudo-nucléaires cytoplasmiques.

En effectuant un marquage avec la molécule TO-PRO3, un intercalant fluorescent de l'ADN double brin, sur des cellules S2 en culture co-transfectées avec un plasmide exprimant les

protéines ParB2 et un plasmide contenant les séquences *parS2* (fig. 94 A), j'ai pu remarquer que les foyers fluorescents contiennent de l'ADN, correspondant probablement à des agrégats de plasmides contenant des séquences *parS*, localisées dans le cytoplasme.



**Figure 94: Les agrégats plasmidiques cytoplasmiques auraient une structuration pseudo-nucléaire**

**A)** Les agrégats plasmidiques cytoplasmiques peuvent être visualisés par un marquage au TO-PRO3, colorant de l'ADN (GFP + topro; x63 zoom 5).

**B)** Les agrégats plasmidiques cytoplasmiques visualisés par la fixation de la protéine parB2 sur sa séquence *parS2* peuvent être marqués à l'aide d'un anti-corps dirigé contre la lamine A.

(GFP + anti-lamin; x63 zoom 5).

De plus, un marquage avec un anticorps dirigé contre la lamine A (**fig. 94 B**) a révélé que ces agrégats cytoplasmiques plasmidiques pouvaient être entourés de lamine, composant de l'enveloppe nucléaire. **Ainsi la présence de lamine autour de ces agrégats plasmidiques cytoplasmiques pourrait indiquer une structuration de type pseudo-nucléaire.** Il est à noter que tous les foyers de fluorescence situés dans le cytoplasme ne sont pas tous entourés de lamine mais ce phénomène n'est pas rare.

## **2.2 Le système ANCHOR est-il responsable de la formation des structures pseudo-nucléaires?**

L'utilisation du système ANCHOR dans des cellules en culture co-transfectées avec un plasmide exprimant les protéines ParB et un plasmide contenant des séquences *parS* a permis de mettre en évidence ces agrégats plasmidiques situés dans le noyau et le cytoplasme. Et ceux présents dans le cytoplasme peuvent être entourés de lamine pouvant faire penser à des structures pseudo-nucléaires. Ces résultats sembleraient indiquer que les plasmides transfectés dans des cellules en culture resteraient localisés et regroupés au sein de la cellule. Connaissant la fonction et le mécanisme d'action des protéines ParB, on peut se demander si ces agrégats plasmidiques ne pourraient pas être une conséquence de l'utilisation du système ANCHOR.

En effet, les protéines ParB se fixent sur leurs séquences *parS* sous forme de dimères, s'agglomèrent et s'étendent autour de leur point de fixation, accumulant suffisamment de fluorescence pour former un foyer de fluorescence visible en microscopie. Les protéines ParB possèdent un domaine central pour se lier à leurs séquences spécifiques *parS* mais aussi à de l'ADN non spécifique (Sanchez, 2103).

L'équipe de Jean-Yves Bouet a proposé le modèle, « Nucleation and caging », pour expliquer le fonctionnement du système de partition ParABS (Sanchez, 2015). Je l'ai déjà présenté dans mon introduction bibliographique mais il est important de reprendre quelques points qui pourraient permettre de comprendre la formation de ces agrégats plasmidiques.

Les protéines dimères ParB ont une forte affinité pour leurs séquences *parS* mais une faible affinité pour l'ADN non spécifique et entre elles. Les interactions des protéines ParB avec elles-

mêmes et avec l'ADN non spécifique formeraient un réseau de faible affinité qui permettrait la formation d'une zone confinée, ou cage, contenant une grande quantité de protéine ParB.

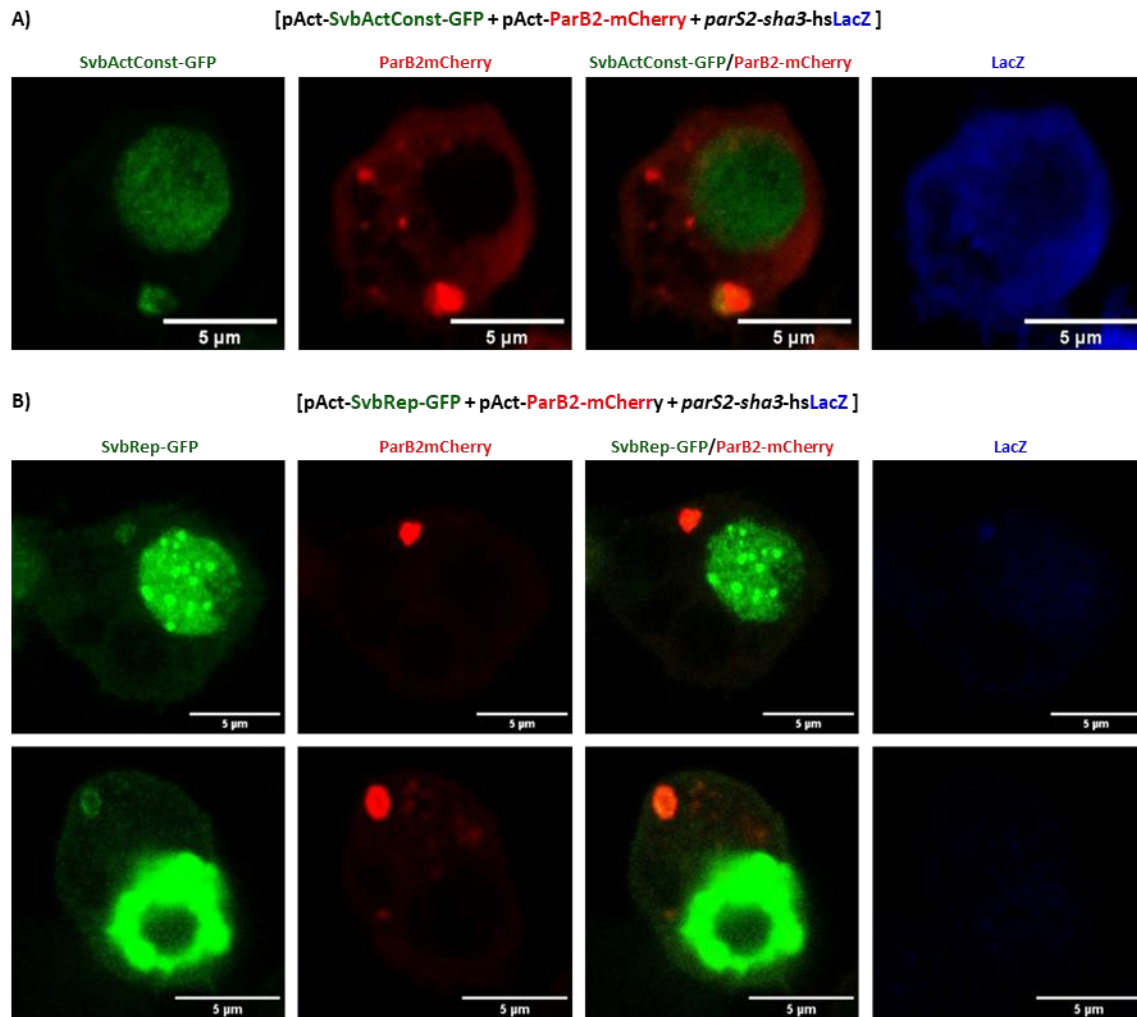
### **Le comportement des protéines Par pourrait-il s'appliquer aux plasmides pour expliquer la formation de ces agrégats cytoplasmiques ?**

#### **La formation de ces agrégats cytoplasmiques plasmidiques est-elle ParB-dépendante ?**

Pour répondre à cette question, l'expérience à réaliser est de regarder si sans les plasmides exprimant les protéines ParB, ces structures pseudo-nucléaires cytoplasmiques peuvent se produire. En absence des protéines ParB pour visualiser ces structures, la fixation d'un facteur de transcription sur ces séquences de régulation cible pourrait permettre de localiser ces structures pseudo-nucléaires cytoplasmiques.

J'ai utilisé le facteur de transcription Shavenbaby (Svb) et une de ses séquences cibles préalablement identifiées, *sha3*, élément de régulation du gène *shavenoid* (Menoret, 2013).

La co-transfection d'un plasmide contenant les séquences *parS* et la séquence régulatrice *sha3*, dirigeant l'expression du gène rapporteur *LacZ*, avec un plasmide exprimant la forme activatrice de Svb, *SvbActConst*, fusionnée avec la protéine fluorescente GFP, permet la synthèse de la protéine  $\beta$ -Galactosidase (**fig. 95 A**). Nous pouvons également voir une co-localisation des protéines ParB avec Svb. Svb se localise bien dans ces structures cytoplasmiques pseudo-nucléaires, identifiées par la présence de la protéine ParB. Il est à noter que la forme répressive de Svb, *SvbRep*, se localise également au niveau de ces structures cytoplasmiques et il n'y a pas expression de la protéine  $\beta$ -Galactosidase (**fig. 95 B**). De plus, comme décrit dans la littérature, la protéine Svb activatrice a une expression diffuse dans le noyau, alors que sa forme répressive est localisée dans le noyau sous forme de dot (Kondo, 2010). En conclusion, un facteur de transcription peut s'accumuler dans les structures cytoplasmiques pseudo-nucléaires et permettre leur localisation. Donc **SvbActConst et SvbRep se fixent sur la séquence sha3 au niveau des structures cytoplasmiques pseudo-nucléaires.**



**Figure 95: Expression du facteur Shavenbaby en cellules S2.**

La forme activatrice SvbActConst **A**) et répressive SvbRep **B**) de Svb se co-localisent avec la protéine ParB au niveau des structures cytoplasmiques.

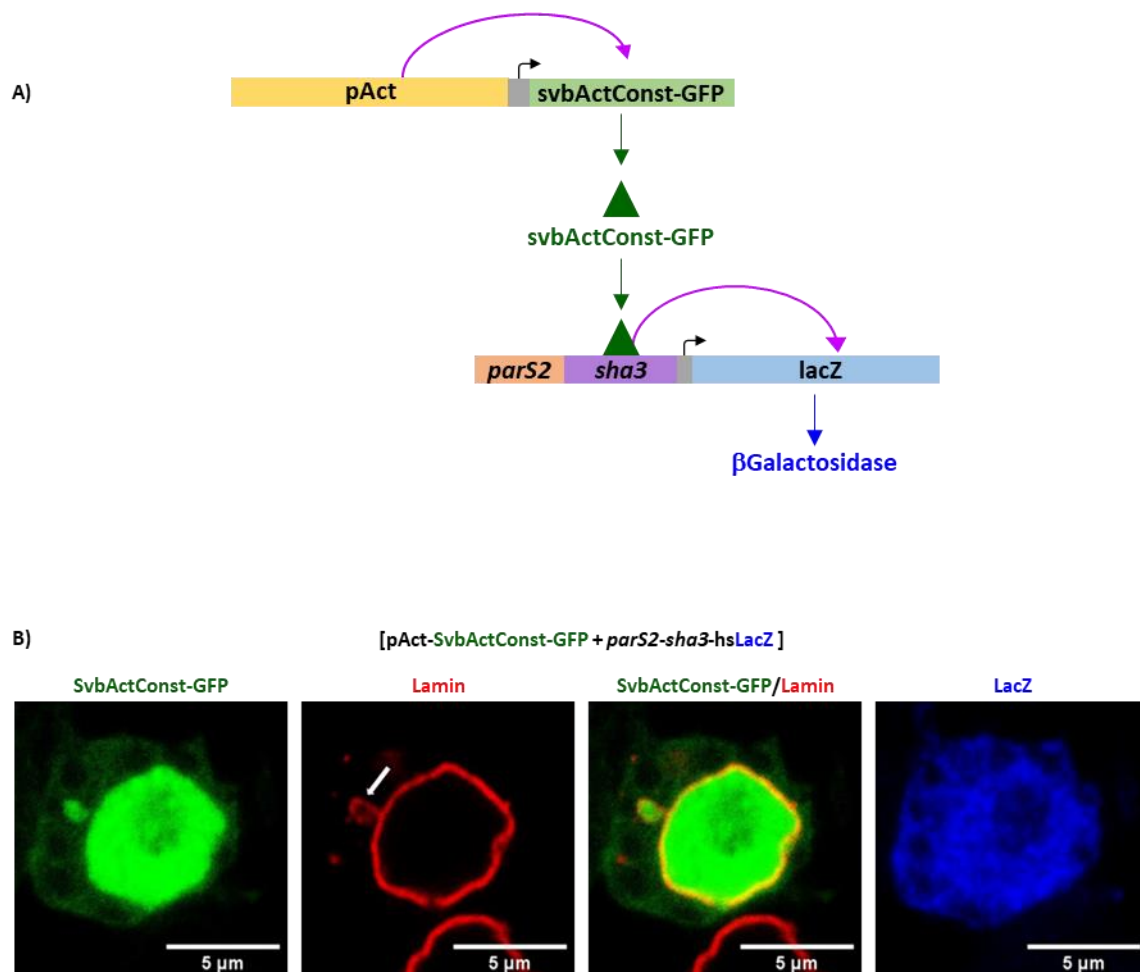
SvbActConst permet la synthèse de la protéine  $\beta$ -Galactosidase.

En transfection transitoire, SvbActConst et SvbRep se comportent différemment dans le noyau. SvbActConst a une expression diffuse dans tout le noyau **A**) tandis que SvbRep se concentre sous forme de foci **B**).

(GFP + mCherry + anti- $\beta$ Gal;  $\times 63$  zoom 5).

En absence des protéines ParB et afin de faciliter l'identification des structures pseudo nucléaires cytoplasmiques, j'ai réalisé un marquage contre la lamine. La forme activatrice de Svb fusionnée à la GFP s'accumule dans le cytoplasme au niveau de structures entourées de lamine. De plus l'expression de la protéine  $\beta$ -Galactosidase certifie la présence du plasmide [*parS-sha3-hsLacZ*] (**fig. 96 A** schéma du principe de l'expérience et **B**).





**Figure 96: La formation des structures plasmidiques cytoplasmiques pseudo-nucléaires n'est pas dépendante de la présence de la protéine ParB2.**

**A)** Principe de l'expérience.

**B)** L'activité de la séquence régulatrice *sha3*, cible du facteur de transcription Svb, visualisée par le marquage avec l'anticorps anti- $\beta$ Gal indique la présence des plasmides [pAct-svbActConst-GFP] et [*parS2-sha3-hsLacZ*] dans la cellule.

La co-localisation d'un marquage avec un anti-corps contre la protéine lamine A et du facteur de transcription Svb fusionné à la protéine fluorescente GFP indique la présence de structures cytoplasmiques pseudo-nucléaires (flèche blanche) malgré l'absence des protéines ParB2.

(GFP + anti-lamin + anti- $\beta$ Gal; x63 zoom 5).

**On peut donc conclure que la formation des structures plasmidiques cytoplasmiques pseudo-nucléaires n'est pas dépendante de la présence des protéines ParB.**

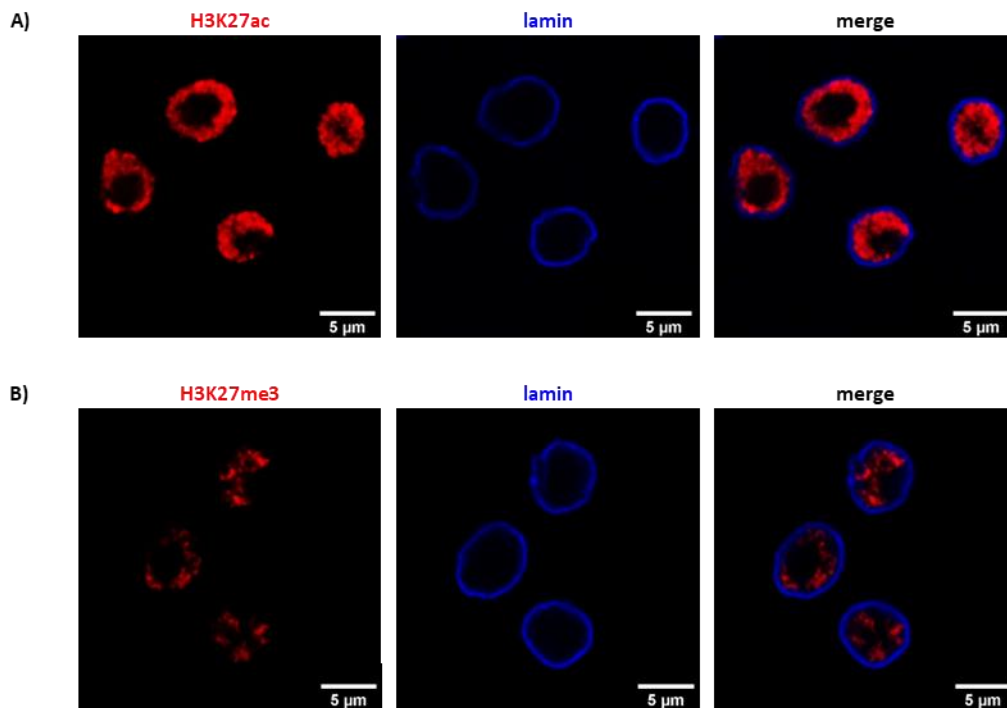
### 2.3 Caractérisation des structures pseudo-nucléaires.

Pour caractériser ces structures cytoplasmiques pseudo-nucléaires, j'ai tout d'abord recherché si l'ADN qu'elles contiennent adopte une certaine organisation de type chromatinien.

Un des principaux constituants protéiques des chromosomes sont les histones qui sont étroitement associées à l'ADN pour former les nucléosomes. Les histones du nucléosome peuvent être modifiées post-traductionnellement sur leur chaîne latérale (acétylation méthylation, phosphorylation...). Ces modifications sont souvent appelées marques d'histones. L'enrichissement de ces modifications au niveau d'éléments génétiques permettent de dresser un état de la chromatine. En 2017, Delandre et Marshall ont classé la chromatine en cinq états différents allant de permissive (2 états), région de la chromatine avec une activité transcriptionnelle importante, à répressive (3 états) avec comme son nom l'indique une absence d'activité transcriptionnelle. Ces différents états indiquent le niveau d'accessibilité de la chromatine en fonction des protéines qui les composent et des différentes marques d'histones associées (Delandre, 2017).

Pour caractériser l'état de ces structures cytoplasmiques pseudo-nucléaires, j'ai choisi deux marques d'histones, chacune représentante d'un état de la chromatine : la marque d'histone H3K27ac pour un état permissif et la marque d'histone H3K27me3 pour un état répressif.

J'ai tout d'abord analysé la localisation de ces deux marques d'histones dans des cellules S2 en utilisant un anticorps dirigé contre chacune de ces marques et un anticorps ciblant la lamine A permettant de localiser le noyau. Dans les cellules S2, l'expression des marques d'histones H3K27ac (**fig. 97 A**) et H3K27me3 (**fig. 97 B**) associées à la chromatine se trouve logiquement dans le noyau.



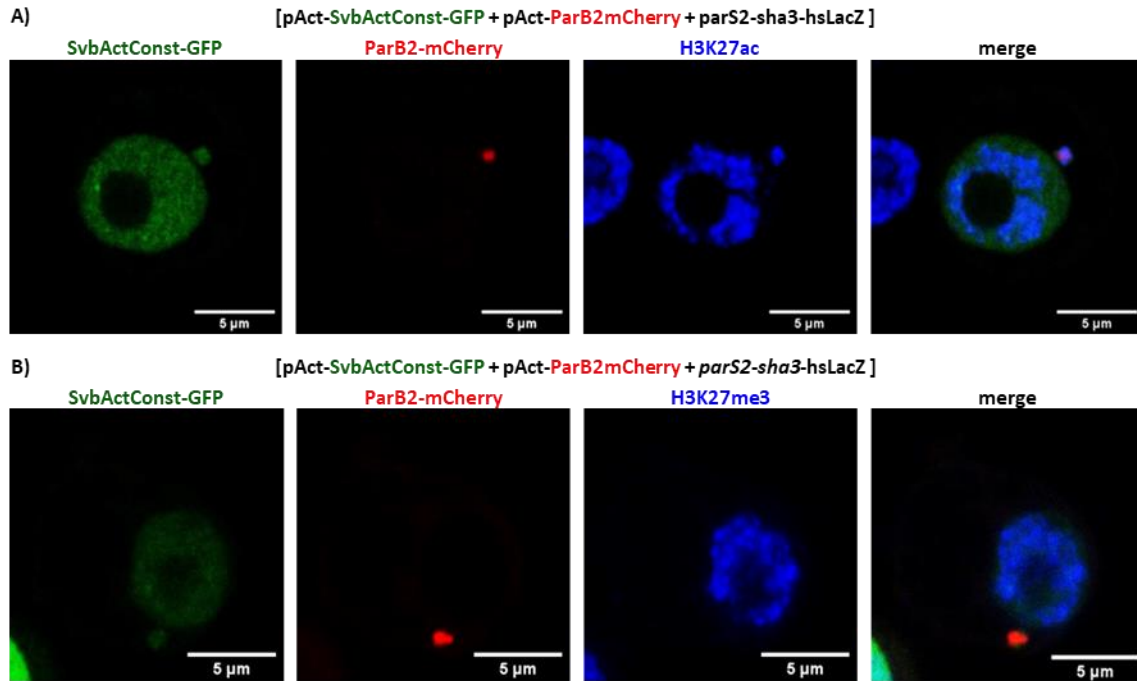
**Figure 96: Localisation des marques d'histones H3K27ac et H3K27me3 en cellules S2.**

Dans les cellules S2, les protéines H3K23ac **A)** ou H3K27me3 **B)** sont exprimées de façon exclusive dans le noyau.

**(A) anti-H3K27ac** ou **B) anti-H3K27me3 + lamin**; x63 zoom).

Pour caractériser l'état chromatinien des structures pseudo-nucléaires, j'ai analysé par immunocytochimie la localisation de ces deux marques d'histones en présence des différentes formes du facteur de transcription Svb et du système ANCHOR qui révèle les structures pseudo-nucléaires.

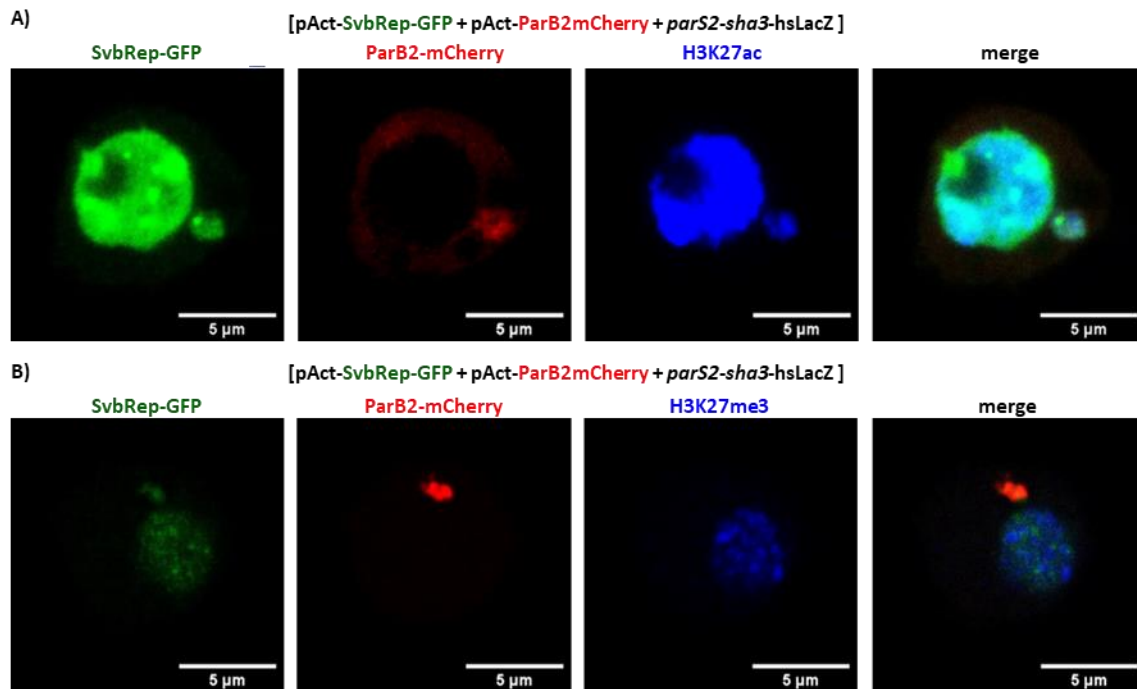
En présence de la forme activatrice de Svb, on peut remarquer dans la **figure 98 A**, une co-localisation dans le cytoplasme au niveau des structures pseudo-nucléaires entre la protéine ParB2, le facteur SvbActConst et la marque d'histone H3K27ac. Par contre, la marque d'histone H3K27me3 présente une localisation strictement nucléaire et ne se localise pas dans les structures cytoplasmiques pseudo-nucléaires (**fig. 98 B**). J'ai effectué les mêmes expériences mais cette fois en présence de la forme répressive du facteur de transcription Svb, SvbRep. On observe les mêmes localisations en ce qui concerne la marque d'histone H3K27ac (**fig. 99 A**) ou H3K27me3 (**fig. 99 B**).



**Figure 98: Localisation des marques d'histones H3K27ac A) et H3K27me3 B) au niveau des structures pseudo-nucléaires en présence de SvbActConst.**

La marque d'histone H3K27ac co-localise avec le facteur SvbActConst et la protéine ParB2 au niveau des structures cytoplasmiques pseudo-nucléaires **A)**, tandis que la marque d'histone H3K27me3 a une localisation strictement nucléaire **B)**.

( **GFP + mCherry + anti-H3K27ac A)** ou **H3K27me3 B)**; x63 zoom 5).



**Figure 99: Localisation des marques d'histones H3K27ac A) et H3K27me3 B) au niveau de structures pseudo-nucléaires en présence de SvbRep.**

La marque d'histone H3K27ac co-localise avec le facteur SvbRep et la protéine ParB2 au niveau des structures cytoplasmiques pseudo-nucléaires **A)**, tandis que H3K27me3 a une localisation strictement nucléaire **B)**.

( **GFP + mCherry + anti-H3K27ac A)** ou **H3K27me3 B)**; x63 zoom 5).

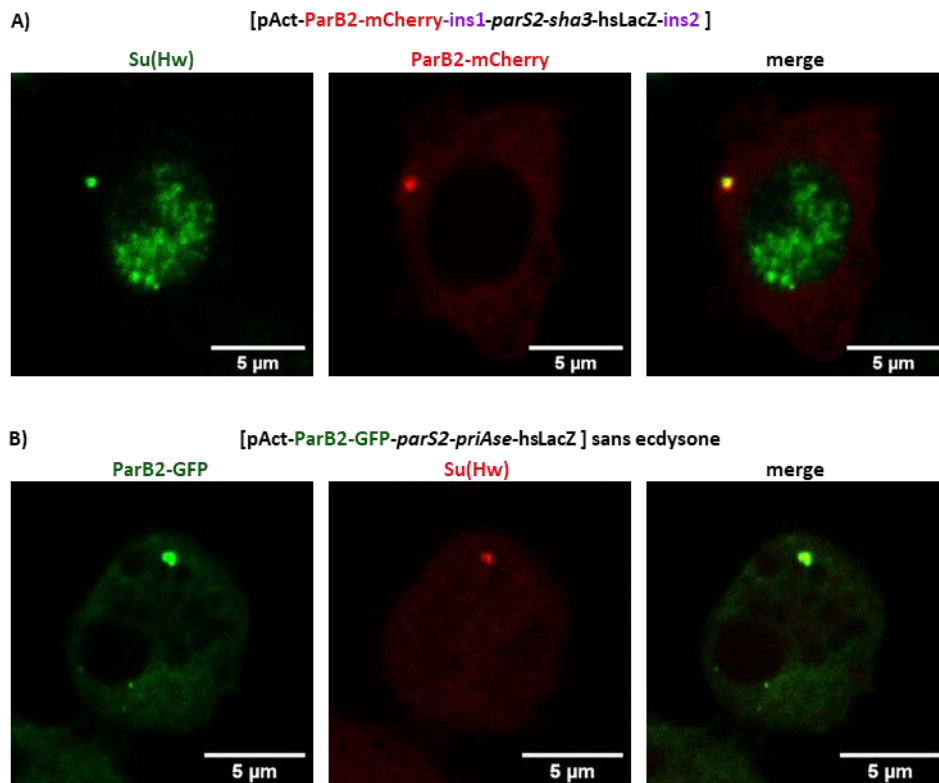
Que Svb soit sous forme activatrice ou répressive, la marque d'histone H3K27ac est présente au niveau des structures cytoplasmiques pseudo-nucléaires, tandis que la marque H3K27me3 est toujours nucléaire. Ces agrégats plasmidiques peuvent être également localisés dans le noyau et je ne les ai pas observés associés avec la marque d'histone H3K27me3, quelle que soit la forme de Svb utilisée (données non montrées).

A partir de ces observations, il semblerait que ces structures cytoplasmiques pseudo-nucléaires acquièrent une certaine organisation qui se rapprocherait d'une organisation de la chromatine de type permissive permettant l'accessibilité des facteurs de transcription à leur séquence cible. Ces résultats sont en accord avec ceux obtenus par Alexandra Mancheno qui, lors de sa thèse dans l'équipe et par des approches génomiques, montre que les formes SvbAct et Rep se fixent sur les mêmes séquences régulatrices dont *sha3*.

**On peut donc conclure que l'ADN plasmidique contenu dans les structures cytoplasmiques pseudo-nucléaires, acquiert une organisation chromatinienne de type permissif insensible à l'activité transcriptionnelle de Svb.**

Tirant profit des constructions contenant des insulateurs utilisées précédemment, j'ai testé si leur présence permet de relocaliser des protéines se fixant sur les insulateurs gypsy dans les structures pseudo-nucléaires. La protéine Suppressor of Hairy wing (Su(Hw)) fait partie du complexe chromatinien insulateur gypsy et se fixe spécifiquement sur une région de cet élément. A l'aide d'un anticorps dirigé contre Su(Hw), fourni par le Professeur Pr. Pamela Geyer de l'Université de l'Iowa, j'ai examiné si cette protéine se localise au niveau de ces structures cytoplasmiques pseudo-nucléaires.

En absence de séquences insulatrices, le comportement de la protéine Su(Hw) semble ambigu car elle est détectée ou non au niveau de ces structures cytoplasmiques pseudo-nucléaires (**fig. 100 B** et images non montrées). En revanche, la protéine su(Hw) se localise au niveau des structures cytoplasmiques pseudo-nucléaires en présence des séquences insulatrices (**fig. 100 A**).



**Figure 100: Localisation de la protéine Su(Hw) au niveau des structures pseudo-nucléaires**

**A)** En présence des séquences insulatrices, la protéine Su(Hw) se co-localise au niveau des structures cytoplasmiques pseudo-nucléaires,

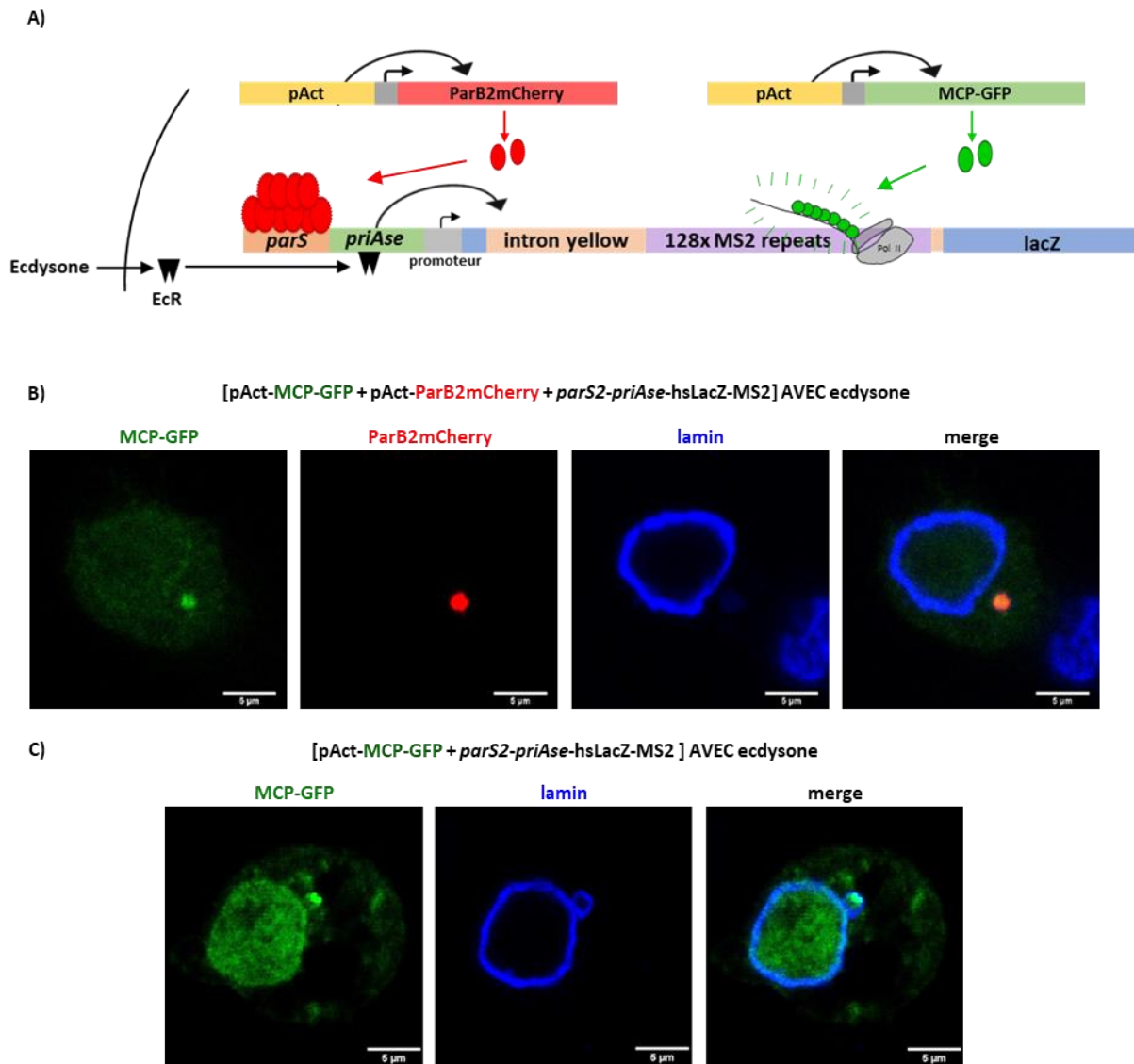
**B)** En absence des séquences insulatrices, la protéine Su(Hw) peut être présente au niveau de ces structures.

( anti-Su(Hw) + mCherry; x63 zoom 5).

Ces résultats suggèrent que la protéine Su(Hw) se fixe sur les séquences insulatrices au sein des structures pseudo-nucléaires, confirmant qu'elles acquièrent une organisation de type chromatinien.

#### 2.4 Peut-il y avoir une activité transcriptionnelle au niveau de ces structures cytoplasmiques pseudo-nucléaires ?

Les résultats précédents suggèrent qu'une activité transcriptionnelle pourrait se produire au niveau de ces structures plasmidiques cytoplasmiques pseudo-nucléaires. Pour répondre à cette interrogation, j'ai à nouveau utilisé le système MS2/MCP, qui permet de détecter les transcrits primaires.



**Figure 101: Activité transcriptionnelle dans ces structures cytoplasmiques pseudo-nucléaires.**

L'expérience est schématisée en A).

B) En présence d'ecdysone, l'expression de la protéine MCP est détectée au niveau de structures dans le cytoplasme et co-localise avec la protéine ParB2.

(GFP + mCherry + anti-lamin; x63zomm5).

C) En absence de la protéine ParB2, une activité transcriptionnelle est retrouvée dans le cytoplasme au niveau de structures pseudo-nucléaires, visualisées par un marquage avec un anticorps dirigé contre la lamine A.

(GFP + anti-lamin ; x63zoom5).

Pour regarder si une activité transcriptionnelle peut avoir lieu au niveau de ces structures plasmidiques cytoplasmiques pseudo-nucléaires, j'ai co-transfecté un plasmide permettant d'exprimer avec le promoteur constitutif de l'actine, la protéine MCP fusionnée à la GFP, un plasmide permettant la synthèse de la protéine ParB2 fusionnée à la protéine fluorescente mCherry et un plasmide contenant les séquences *parS2* à proximité de la région régulatrice

*priAse* qui sous contrôle de l'ecdysone exprime le gène rapporteur LacZ contenant des séquences MS2 (**fig. 101 A**).

Comme attendu, la protéine ParB2 se localise au sein de ces structures cytoplasmiques pseudo-nucléaires contenant l'ADN du plasmide [*parS2-priAse-hsLacZ-MS2*] (**fig. 101 B**). De manière plus surprenante, la protéine MCP se localise également dans le cytoplasme au niveau des structures cytoplasmiques pseudo-nucléaires, mettant ainsi en évidence, la synthèse des transcrits primaires dans ces structures. D'après ces observations, on peut donc proposer qu'**une activité transcriptionnelle aurait lieu au niveau de ces structures plasmidiques cytoplasmiques pseudo-nucléaires.**

En absence de la protéine ParB2, on peut observer que la protéine MCP se localise tout de même dans ces structures cytoplasmiques entourées par la lamine A (**fig. 101 C**). Cela confirme, si nécessaire, que la formation de ces structures est indépendante de la présence des protéines ParB2.

**Pour conclure, l'ensemble des résultats acquis sur ces structures plasmidiques cytoplasmiques pseudo-nucléaires soutient que ces structures acquièrent un certain degré d'organisation de type chromatine permissive leur permettant d'avoir une activité transcriptionnelle.**

Le système ANCHOR fonctionne aussi en transfection transitoire et la puissance de ce système a permis d'identifier des structures pseudo-nucléaires contenant de l'ADN plasmidique. L'ensemble des outils disponibles dans ces cellules associé au système ANCHOR devrait permettre dans le futur d'appréhender plus finement la dynamique de la chromatine et la régulation de l'expression génique.



### **3- Discussions et perspectives.**

De par sa simplicité de manipulation et les nombreux outils disponibles, les cellules S2 constituent un modèle de choix pour étudier les mécanismes de la dynamique chromatinienne. C'est pour cela que j'ai développé le système ANCHOR dans ce système cellulaire, tout d'abord en transfection transitoire, mais avec pour objectif à plus long terme de générer des lignées stables, contenant des séquences *parS* intégrées dans le génome et permettant ainsi de travailler dans des meilleures conditions expérimentales.

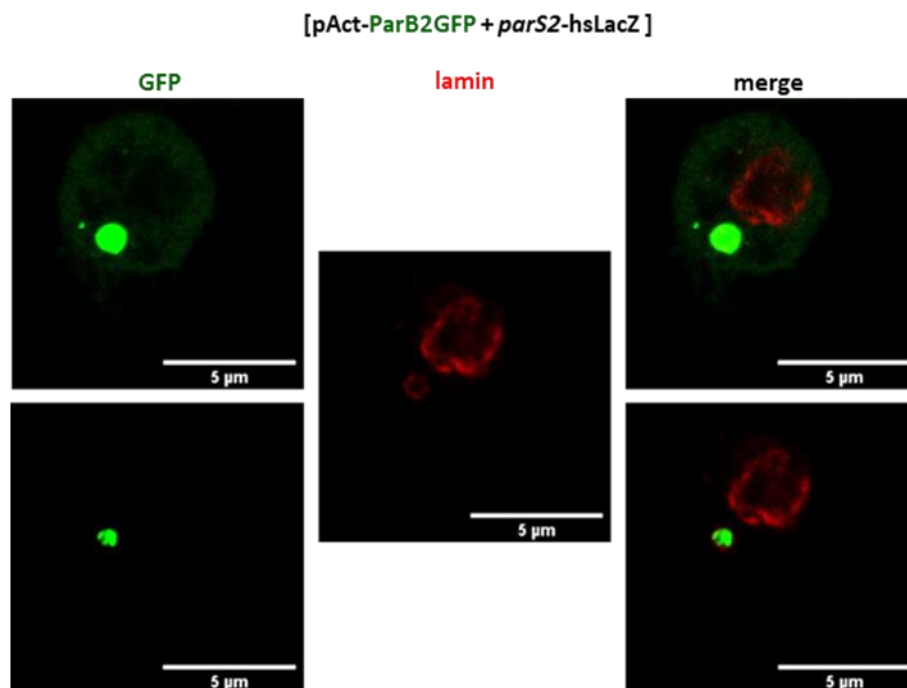
Au cours de ce développement, j'ai pu mettre en évidence que le système ANCHOR fonctionne pleinement dans ce système cellulaire et qu'il permet de visualiser l'ADN plasmidique, non intégré au génome des cellules hôtes. Ces expériences ont mis en évidence que les plasmides peuvent se regrouper sous formes d'agrégats et cela indépendamment du système ANCHOR (**fig. 94 et 96**). Ces agrégats sont visibles aussi bien dans le cytoplasme que dans le noyau des cellules transfectées et posent la question des mécanismes d'import des plasmides au sein des noyaux.

#### **3-1 Mécanismes d'import des plasmides dans le noyau.**

Les plasmides transfectés qui se retrouvent dans le cytoplasme des cellules sont rapidement exposés aux nucléases présentes dans le cytosol. La fixation de nombreuses protéines sur l'ADN plasmidique, assure à la fois sa compaction et sa protection. Afin que cet ADN puisse être transcrit, il doit être intégré au noyau qui est isolé du reste de la cellule par la membrane nucléaire. Pour atteindre son but, l'ADN plasmidique peut profiter de la destruction de l'enveloppe nucléaire qui se produit lors de la division cellulaire. Pour les cellules interphasiques, les choses sont donc plus difficiles et s'il semble possible qu'une petite fraction des plasmides puissent passer par simple diffusion à travers les pores nucléaires, ce phénomène reste anecdotique. Pour envahir les noyaux interphasiques, les plasmides profitent des propriétés d'import des protéines qui leur sont associées comme certaines histones et plus particulièrement des facteurs de transcription. En interphase, la plus grande fraction des plasmides atteignant le noyau le fait grâce à un transport actif, en utilisant les pores nucléaires (Bai, 2018). Il est aussi intéressant de noter que le transport des plasmides dans le cytoplasme, vers les pores nucléaires, nécessite également un transport actif faisant

intervenir le cytosquelette cellulaire et notamment les microtubules (Mesika, 2005 ; Salman, 2005 ; Vaughan, 2006). L'import nucléaire est donc facilité par la fixation de facteurs de transcription sur l'ADN plasmidique sur des séquences qualifiées de séquence de ciblage. Un exemple typique de ces séquences de ciblage est l'enhancer SV40.

En effet, dans des études de micro-injection et d'hybridation *in situ*, réalisées pour étudier le mouvement de l'ADN plasmidique, il a été découvert que les plasmides contenant l'intégralité du génome du SV40 (5243 pb) ou seulement 50 pb correspondant à l'enhancer SV40 étaient en fait capables d'entrer dans les noyaux des cellules interphasiques. Les mêmes plasmides, dépourvus de l'enhancer SV40 restent dans le cytoplasme (Dean, 1997, 1999). Dans mes expériences, j'ai pu observer des choses similaires. Lors de la co-transfection d'un plasmide codant pour la protéine ParB2 et d'un plasmide [*parS2*-hslacZ] qui ne contient que les séquences *parS*, j'ai pu observer des foyers fluorescents **uniquement dans le cytoplasme** (fig. 102). Pour le système ANCHOR, cette observation a été également vue dans des cellules de mammifères (Fok, 2021). L'ajout d'une séquence régulatrice, *priAse* au sein de ce plasmide,



**Figure 102: Les structures pseudo-nucléaires constituées de plasmides ne contenant que des séquences *parS* sont localisées uniquement dans le cytoplasme.**

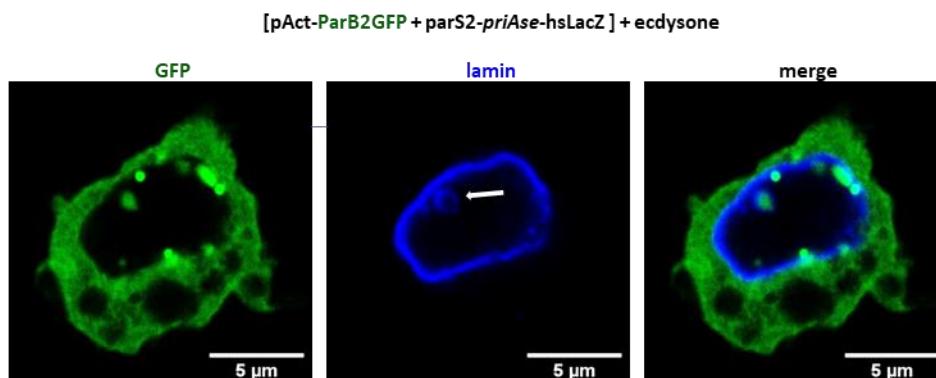
Les deux images avec l'expression de la GFP ont été réalisées sur la même cellule mais avec une intensité du laser différente.

(GFP + anti-lamin; x63zoom5).

en présence d'ecdysone régulant l'activité du gène rapporteur *lacZ*, permet la translocation du plasmide [*parS2-priAse-hslacZ*] dans le noyau (**fig 87**). Mes résultats confirment donc que la présence de séquences transcriptionnellement actives, permettant la fixation de facteur de transcription, facilite le transport des plasmides du cytoplasme vers le noyau où ils seront transcrits. Dans mon cas, il serait intéressant de tester la localisation du plasmide [*parS2-priAse-hslacZ*], en absence d'ecdysone et donc sans activité transcriptionnelle.

### 3-2 Formation des structures pseudo-nucléaires.

Nous venons de voir que la présence de séquences spécifiques facilite l'import des plasmides dans le noyau et qu'en leur absence, ces derniers restent localisés dans le cytoplasme. Mes travaux mettent en évidence que les plasmides au sein du cytosol peuvent former des agrégats qui s'organisent en structures pseudo-nucléaires identifiables par la présence de lamine et d'histones (**fig. 94**). Il m'est également arrivé de retrouver ces agrégats plasmidiques entourés de lamines dans le noyau (**fig. 103**, flèche blanche).



**Figure 103: A l'intérieur du noyau, des structures pseudo-nucléaires entourées de lamine peuvent être observées (flèche blanche).**  
(GFP + anti-lamin ; x63zoom5).

On peut se demander pourquoi des plasmides transfectés restent groupés et acquièrent une structure pseudo-nucléaire qui peut être localisée dans le cytoplasme ? Est-ce que la technique de transfection utilisée, générant elle-même des précipités favorise la formation des agrégats et induit en erreur la cellule faisant penser à un noyau ou micro-noyau ? On peut également se demander si ces structures persistent lors de la division cellulaire. Enfin, il serait

intéressant répéter ces expériences de transfection transitoire utilisant le système ANCHOR dans d'autres types de cellules en culture provenant de mammifère comme les cellules NIH3T3, pour regarder si ces structures cytoplasmiques pseudo-nucléaires sont également observables.

### 3-3 Le système ANCHOR, Shavenbaby et accessibilité chromatinienne.

Dans notre équipe, les travaux de thèse d'Alexandra Mancheno-Ferris ont porté sur la caractérisation, par des approches génomiques, des mécanismes moléculaires qui contrôlent la fixation sur l'ADN du facteur de transcription Shavenbaby. Alexandra a montré que la forme activatrice de Svb, SvbAct, et la forme répressive, SvbRep, se fixent sur les mêmes régions cis-régulatrices pour contrôler de manière antagoniste l'expression d'environ 200 gènes cibles directes dans les cellules S2. Alexandra a également mis en évidence que l'environnement chromatinien dans lequel Svb évolue contraint sa fixation et par conséquent modifie le répertoire de ses gènes cibles (Mancheno-Ferris, 2021). Ainsi Svb, régule différentes cohortes de gènes cibles en fonction du type cellulaire dans lequel il se trouve, les cellules S2, les cellules épidermiques embryonnaires ou encore les cellules souches intestinales.

Étant donné que Svb ne s'exprime pas de manière endogène dans les cellules S2, ces dernières offrent une opportunité de tester l'accessibilité de ce facteur de transcription sur ses séquences cibles. En utilisant des lignées cellulaires stables de cellules S2 permettant l'expression de SvbAct ou de SvbRep, Alexandra a montré que ces deux protéines ne se fixent pas sur les séquences régulatrices contrôlant l'expression du gène *shavenoid* car elles se trouvent dans un environnement chromatinien répressif.

Dans la **figure 104** est représenté le locus du gène *shavenoid* ainsi que les différentes régions régulatrices contenant des motifs de fixation de svb OvoQ6 (prédits par *cisTargetX*). Ces régions ont été testées en essais transgéniques pour appréhender leur activité au sein de l'épiderme embryonnaire. Si la région régulatrice *sha3*, récapitule l'expression épidermique de *shavenoid*, la région *sha2* s'est révélée négative, ne permettant pas une expression spécifique. Les analyses de ChIPSeq indiquent que Svb ne se fixe pas sur cette région que ce soit dans des cellules épidermiques embryonnaires ou en cellules S2.

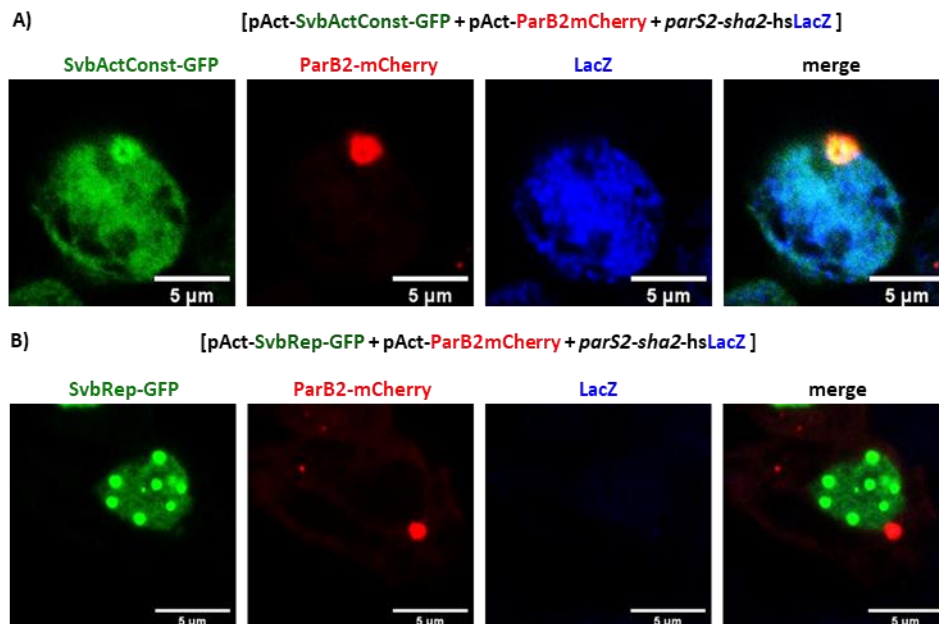


**Figure 104: Locus du gène *shavenoïd***

La figure représente le gène *shavenoïd* ainsi que ces régions régulatrices prédites par *cisTargetX*. Les différentes régions, *sha1*, *2*, *3*, *7k* et *int* ont été testées par des essais transgéniques en regardant l'expression du gène rapporteur *lacZ*. En rose, est indiquée les régions qui ne dirigent pas une expression spécifique dans l'épiderme embryonnaire. Les régions en bleu affichent une activité activatrice reproduisant le domaine endogène du gène *shavenoïd* (Menoret, 2013).

En utilisant le système ANCHOR, et bénéficiant de la présence des structures pseudo-nucléaires, j'ai voulu utiliser la région *sha2* comme une région « témoin » et regarder comment elle se comporte en transfection transitoire dans des cellules S2 pour voir :

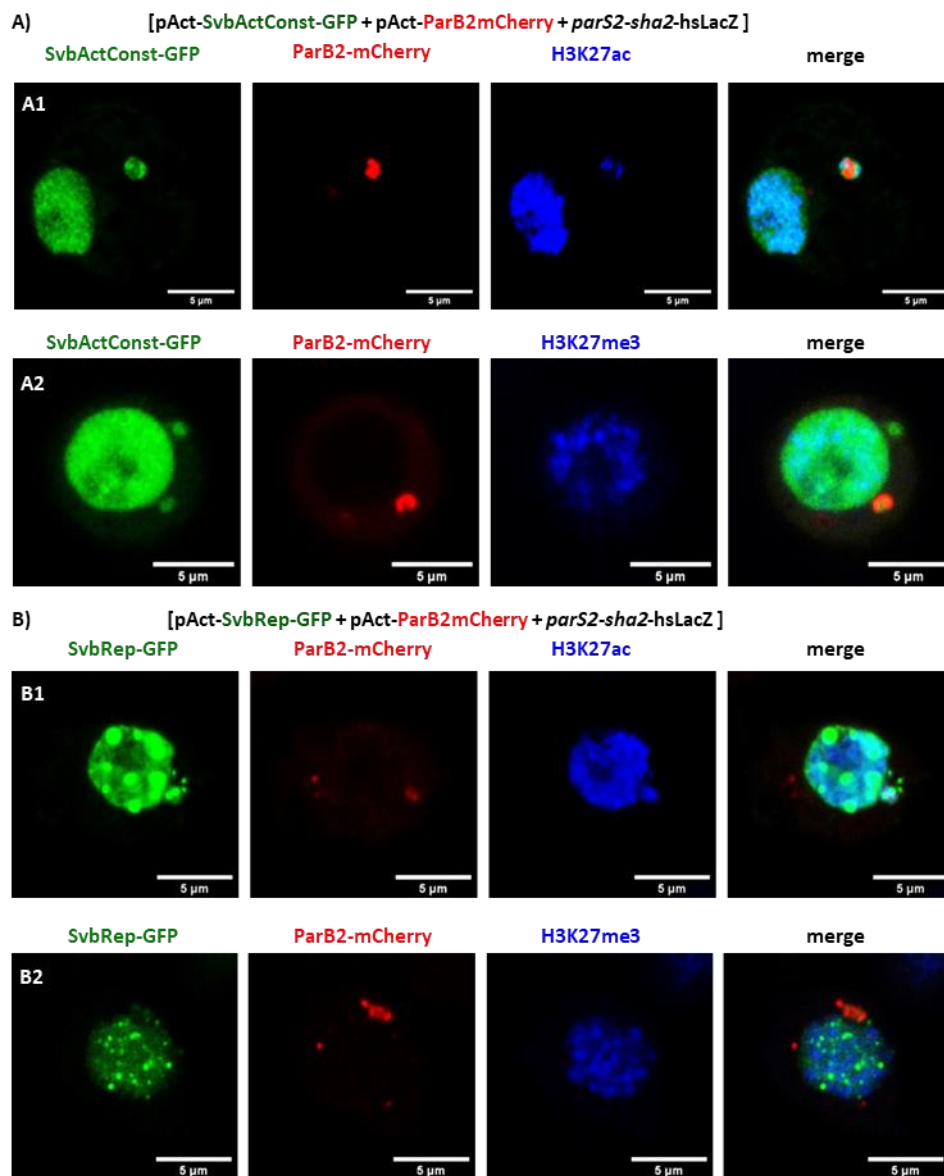
- si SvbAct et SvbRep peuvent se fixer dessus.
- si en présence de SvbAct, *sha2* est capable d'activer l'expression d'un gène rapporteur.
- si en présence de la région *sha2*, on peut observer ces structures cytoplasmiques pseudo-nucléaires et quelles seraient leur état chromatinien.



**Figure 105: Activité de la région régulatrice *sha2* en cellules S2.**

SvbActCont A) tout comme tout comme SvbRep B) peut se fixer sur le site *sha2*. En présence de SvbActConst, la région régulatrice *sha2* permet l'expression du gène rapporteur *LacZ*. ( GFP + mCherry + anti-βGal ; x63 zoom 5)

Dans la **figure 105** est représentée l'activité transcriptionnelle de la région *sha2* en transfection transitoire dans des cellules S2, en présence de SvbActConst (**fig. 105 A**) ou SvbRep (**fig. 105 B**). La présence des structures pseudo-nucléaires cytoplasmiques permet de voir que SvbActConst et Svb Rep s'accumulent sur les plasmides contenant la région *sha2*.



**Figure 106: En présence de la région régulatrice *sha2*, les structures cytoplasmiques pseudo-nucléaires présentent une structure de type chromatine accessible.**

Comme pour la région régulatrice *sha3*, co-localisation de la protéine H3K27ac avec SvbActConst (**A1**) et SvbRep (**B1**) au niveau des structures cytoplasmiques pseudo-nucléaires.

La protéine H3K27me3 a une localisation exclusivement nucléaire que ce soit avec SvbActConst (**A2**) et SvbRep (**B2**) et n'est jamais localisée au niveau de ces structures.

( GFP + Cherry + anti-H3K27ac ou anti-h3K27me3 ; x63 zoom5)

Mais, seulement la protéine SvbActConst, est capable d'induire l'expression du gène rapporteur *lacZ* dépendant de la région régulatrice *sha2*. Cette expérience révèle qu'une

région régulatrice, inaccessible dans le génome des cellules peut présenter la fixation de facteur de transcription lorsqu'elle est contenue dans un plasmide libre de contraintes chromatiniennes. L'accessibilité de l'ADN plasmidique est d'ailleurs révélée par la présence de la marque d'histone H3K27ac, identifiant les enhanceurs actifs au sein des structures pseudo-nucléaires cytoplasmiques (**fig. 106**).

Ainsi en transfection transitoire, dans des cellules S2, le site OvoQ6 contenu dans la région régulatrice de *sha2* est accessible au sein de l'ADN plasmidique et permet la fixation de Svb. Dans ce cas, la région *sha2* tout comme la région *sha3* présente une activité favorisant l'expression du gène situé en aval de ces séquences.

Ces résultats obtenus avec la région *sha2* soulignent l'importance de l'accessibilité de la chromatine aux facteurs de transcription et complètent ceux obtenus par Alexandra durant sa thèse. La régulation de l'accessibilité à la chromatine est fondamentale dans l'expression des génomes au cours du développement ou de la vie d'une cellule pour permettre notamment de maintenir les fonctions de chaque type cellulaire.

# **Annexes**



## Annexe 1 : Liste des plasmides construits

pAttb-UAST-ParB1-GFP  
pAttb-UAST-ParB1-mCherry  
pAttb-UAST-ParB1-GFP-*parS1*modif  
pAttb-UAST-ParB2-GFP  
pAttb-UAST-ParB2-mCherry  
pAttb-UAST-ParB2-GFP-*parS2*  
pAttb-UAST-ParB2-mCherry-*parS2*  
pAttb-UAST-ParB3-GFP  
pAttb-UAST-ParB4-mVenus  
pAttb-UAST-ParB4-GFP  
pAttb-UAST-ParB4-mVenus-*parS4*  
pAttb-*parS1*-hsLacZ  
pAttb-*parS1*modif-hsLacZ  
pAttb-*parS2*-hsLacZ  
pAttb-*parS2-sha3*-hsLacZ  
pAttb-*parS2-sha2*-hsLacZ  
pAttb-*parS2-priAse*-hsLacZ  
pAttb-*parS3*-hsLacZ  
pAttb-*parS4*-hsLacZ  
pAttb-*parS2-sha3*-hsLacZ-intron-yellow-MS2  
pAttb-*parS2*-hsLacZ-intron-yellow-MS2  
pAttb-*parS2-priAse*-hsLacZ-intron-yellow-MS2  
pAct-MCP-GFP  
pAttb-*parS1*modif-AB-*parS2*-hsLacZ  
pAct-ParB1-GFP  
pAct-ParB1-mCherry  
pAct-ParB2-GFP  
pAct-ParB2-mCherry  
pAct-ParB1-GFP-*parS1*modif  
pAct-ParB1-mCherry-*parS1*modif  
pAct-ParB2-GFP-*parS2*  
pAct-ParB2-GFP-*parS2-priAse*  
pAct-ParB2-GFP-*parS2-priAse*-hsLacZ  
pAct-ParB2-mCherry-*parS2*  
pAct-ParB2-mCherry-*parS2-priAse*  
pAct-ParB2-mCherry-*parS2-priAse*-hsLacZ  
pAct-ParB2-GFP-*parS2-sha3*  
pAct-ParB2-GFP-*parS2-sha3*-hsLacZ  
pAct-ParB2-mCherry-*parS2-sha3*  
pAct-ParB2-mCherry-*parS2-sha3*-hsLacZ  
pAct-ParB2-GFP-**ins1**-*parS2-priAse*-hsLacZ  
pAct-ParB2-GFP-*parS2-priAse*-hsLacZ-**ins2**  
pAct-ParB2-GFP-**ins1**-*parS2-priAse*-hsLacZ-**ins2**

## Annexe 2 : Liste des lignées transgéniques construites

10967-1.2M | yw-; + ; UAST-ParB1-GFP / TM6 (86Fb)  
 10967-2.2M | yw-; + ; UAST-ParB1-mCherry / TM6 (86Fb)  
 10967-3.1M | yw-; + ; UAST-ParB2-GFP / TM6 (86Fb)  
 10967-4.2M | yw-; + ; UAST-ParB2-mCherry / TM6b (86Fb)  
 10967-5.1M | yw-; + ; *parS1*-hsLacZ / TM6 (86Fb)  
 10967-6.1M | yw-; + ; *parS2*-hsLacZ / TM3 (86Fb)  
 10967-9.1M | yw ; + ; *parS2-sha3*-hsLacZ / TM6 (86Fb)  
 11872-1.1M | yw ; + ; UAST-ParB1-GFP / TM6 (76A2)  
 11872-1.2M | yw ; + ; UAST-ParB1-GFP Hz (76A2)  
 11872-2.1M | yw ; + ; UAST-ParB2-GFP Hz (76A2)  
 11872-2.2M | yw ; + ; UAST-ParB2-GFP / TM6 (76A2)  
 11872-3.1M | yw ; + ; *parS1*-hsLacZ Hz (76A2)  
 11872-3.2M | yw ; + ; *parS1*-hsLacZ / TM3 (76A2)  
 11872-4.1M | yw ; + ; *parS2*-hsLacZ / TM6 (76A2)  
 11872-4.2M | yw ; + ; *parS2*-hsLacZ Hz (76A2)  
 11872-5.2M | yw ; + ; *parS1*-hsLacZ / TM3 (68E1)  
 11872-5.4M | yw ; + ; *parS1*-hsLacZ Hz (68E1)  
 11872-6.3M | yw ; + ; *parS2*-hsLacZ / TM6 (68E1)  
 22844-1-4M | yw ; + ; *parS1*modif-hsLacZ / TM3+6 (86Fb)  
 yw ; + ; *parS3*-hsLacZ / TM3 (86Fb) L1  
 yw ; + ; *parS3*-hsLacZ / TM3 (86Fb) L2  
 yw ; + ; *parS3-sha3*-hsLacZ / TM3 (86Fb) L1  
 yw ; + ; *parS3-sha3*-hsLacZ / TM3 (86Fb) L2  
 18475-1.2M | yw , *parS2*-hsLacZ / FM7i (2A3) ; + ; +  
 18475-2.1M | yw ; *parS2*-hsLacZ / CyO (43A1) ; +  
 14598-3.1M | yw ; *parS2-sha3*-hsLacZ / TM3 (76A2) ; +  
 yw ; *ptc-Gal4*/(CwL) ; UAST-ParB3-GFP / TM3 (65B2)  
 yw ; + ; UAST-ParB3-mCherry L1 Hz (89E11)  
 yw ; If/CwL ; UAST-ParB3-GFP (65B2) , *parS3* (86Fb) / TM3  
 yw ; + ; UAST-ParB3-GFP (65B2) , *parS3* (86Fb) / TM3 (R9.3)  
 25811-9-1M | yw ; + ; UAST-ParB4-mVenus / TM3 (86Fb)  
 25811-10-2M | yw ; + ; *parS4*-hsLacZ / TM3 (86Fb)  
 yw ; If/CwL ; UAST-ParB4-mVenus-*parS4* / TM3 (86Fb) D  
 yw ; If/CwL ; UAST-ParB4-GFP / TM3 (86Fb) A  
 27224-1-1M | yw ; + ; UAST-ParB1-GFP-*parS1*modif / TM3 (86Fb)  
 27224-3-2M | yw ; + ; UAST-ParB2-GFP-*parS2* / TM3 (86Fb)  
 27224-4-1M | yw ; UAST-ParB2-GFP-*parS2* / TM3 (43A1) ; +  
 27224-5-2M | yw ; + ; UAST-ParB2-mCherry-*parS2* / TM3 (76A2)  
 27224-6-3M | yw ; UAST-ParB2-mCherry-*parS2* Hz (43A1)  
 27224-2-1M | yw ; If / CwL ; UAST-ParB1-GFP-*parS1*modif / TM3 (76A2)  
 27224-3-1M | yw ; If / CwL ; UAST-ParB2-GFP-*parS2* / TM3 (86Fb)  
 27224-5-1M | yw ; If / CwL ; UAST-ParB2-mCherry-*parS2* / TM3 (76A2)  
 27224-2-1M | yw ; *ptc-Gal4* / (CwL) ; UAST-ParB1-GFP-*parS1*modif / TM3 (76A2)  
 27224-3-1M | yw ; *ptc-Gal4* / (CwL) ; UAST-ParB2-GFP-*parS2* / TM3 (86Fb)  
 28522-5-1M | yw ; *parS1*modif-AB-*parS2*-hsLacZ / CyO ; + (1) chrII  
 28522-5-1M | yw ; *parS1*modif-AB-*parS2*-hsLacZ (Hz) ; TM3/TM6b (1) chrII  
 28522-5-5F | yw ; + ; *parS1*modif-AB-*parS2*-hsLacZ / TM3 ; + (5) chrIII  
 yw ; If/CwL ; UAST-ParB1-GFP (76A2) ; UAST-ParB2mCherry (86Fb) [R7,1] / TM3  
 yw ; C147-Gal4 / CwL ; UAST-ParB1-GFP (76A2) ; UAST-ParB2mCherry (86Fb) [R7-1] / TM3  
 yw ; *ptc-Gal4* / CwL ; UAST-ParB1-GFP (76A2) ; UAST-ParB2mCherry (86Fb) [R7-1] / TM3  
 31051-1-1M | yw ; *parS2-sha3*-hsLacZ-MS2 (43A1) / CyO ; +  
 yw ; *parS2-sha3*-hsLacZ-MS2 (43A1) / CyO ; UAST-ParB2-mCherry (76A2)/TM3 (1)  
 yw ; + ; pL402-UAS-MCP-GFP (65B2)/TM3b (from Mounia Lagha)  
 yw ; If/CwL ; UAS-MCP-GFP (65B2) / TM3 (from Mounia Lagha)  
 yw ; *ptc-Gal4* / CwL ; UAS-MCP-GFP (65B2) / TM3  
 31051-1-1M | yw ; *parS2-sha3*-hsLacZ-MS2 (43A1) / CwL ; TM3/TM6b  
 yw ; UASp-Svb-ActConst-GFP (3A) / CwL ; + / (TM3) chrII

## Annexes 3 : Matériels et Méthodes

### Constructions plasmidiques et lignées transgéniques

En **annexe 1** se trouve la liste des différentes constructions plasmidiques qui sont citées dans ce mémoire de thèse. Les différentes amorces utilisées pour obtenir ces constructions sont listées dans le **tableau 5**. Toutes les constructions ont été vérifiées par séquençage.

Les séquences pour les systèmes **ANCHOR 1** (pGFP-c2-ParB1-GFP ou mCherry / pUC-*parS1*) et **ANCHOR2** (pGFP-c2-ParB2-GFP / pUC-*parS2*) m'ont été fournies par l'équipe de Kerstin Bystricky de l'unité de biologie MCD à Toulouse.

Les séquences ParB1-GFP, ParB1-mCherry et ParB2-GFP ont été obtenues par digestion des plasmides pGFP-c2-ParB1-GFP, pGFP-c2-ParB1-mCherry et pGFP-c2-ParB2-GFP, respectivement avec les enzymes de restriction NheI/XbaI. Ces séquences ont été clonées, par ligation avec la T4 DNA ligase d'Invitrogen, suivant le protocole du fabricant, dans le vecteur pAttb-UAST préalablement coupé par l'enzyme de restriction XbaI et déphosphorylé par l'action de l'alkaline phosphatase calf intestinal (CIP) de chez New England Biolabs (NEB) suivant le protocole du fabricant pour obtenir les constructions **pAttb-UAST-ParB1-GFP**, **pAttb-UAST-ParB1-mCherry** et **pAttb-UAST-ParB2-GFP**.

La construction **pAttb-ParB2-mCherry** a été obtenue par ligation de la séquence mCherry provenant de la digestion du plasmide pAttb-UAST-ParB1-mCherry par NdeI/XbaI dans le plasmide pAttb-ParB2-GFP coupé par les enzymes NdeI/XbaI avec élimination de la séquence codant pour la GFP.

Les séquences *parS1* et *parS2* ont été obtenues par digestion avec les enzymes de restriction HindIII/EcoRI à partir des plasmides pUC-*parS1* et pUC-*parS2*, respectivement. Elles ont été clonées par ligation avec la T4 DNA ligase d'Invitrogen, suivant le protocole du fabricant, dans le vecteur pAttb-hsLacZ digéré par les enzymes de restriction HindIII/EcoRI pour obtenir les constructions **pAttb-*parS1*-hsLacZ** et **pAttb-*parS2*-hsLacZ**.

Les séquences du système **ANCHOR3** (ParB3/*parS3*) sont la propriété de la société NeoVirTech SA et sont confidentielles. Les séquences ParB3 et *parS3* ont été clonées, respectivement dans

les vecteurs pAttb-UAST et pAttb-hsLacZ pour obtenir les constructions **pAttb-UAST-ParB3-GFP** et **pAttb-*parS3*-hsLacZ**.

Les séquences du système **ANCHOR4** (ParB4/*parS4*), contenues dans le vecteur pJYB234-ParB4-mVenus-*parS4*, m'ont été fournies par l'équipe de Jean-Yves Bouet du LMGM à Toulouse. Les séquences ParB4, ParB4-mVenus, ParB4-mVenus-*parS4* et *parS4* ont été obtenues par PCR avec la polymérase Phusion de chez ThermoFisher, suivant le protocole du fabricant, à partir du vecteur pJYB234-ParB4-mVenus-*parS4*, en utilisant respectivement les couples d'amorces [parB4-fw/parB4-rev], [parB4-mVenus-fw/parB4-mVenus-rev], [parB4-mVenus-fw/*parS4*-rev] et [*parS4*-fw/*parS4*-rev]. Ces séquences amplifiées ont été insérées dans divers vecteurs par recombinaison homologue en utilisant le système InFusion de chez Takara, suivant le protocole du fabricant.

La séquence ParB4 a été clonée dans le vecteur pAttb-UAST-GFP digéré par BglII/NdeI pour obtenir la construction **pAttb-UAST-ParB4-GFP**.

La séquence ParB4-mVenus a été clonée dans le vecteur pAttb-UAST digéré par EcoRI/KpnI pour obtenir la construction **pAttb-UAST-ParB4-mVenus**.

La séquence ParB4-mVenus-*parS4* a été clonée dans le vecteur pAttb-UAST digéré par EcoRI/NotI pour obtenir la construction **pAttb-UAST-ParB4-mVenus-*parS4***.

La séquence *parS4* a été clonée dans le vecteur pAttb-hsLacZ digéré par HindIII/NotI pour obtenir la construction **pAttb-*parS4*-hsLacZ**.

La construction **pAttb-*parS1modif*-hsLacZ** a été obtenue par insertion du fragment *parS1modif* synthétisé par la société Integrated DNA Technologies (IDT) dans le vecteur pAttb-hsLacZ digéré par HindIII/EcoRI, en utilisant le système InFusion de la société Takara.

Le fragment *parS1modif*, amplifié par PCR à partir de pAttb-*parS1modif*-hsLacZ avec le couple d'amorces [*newparS2*-fw/*newparS2*-rev] est introduit dans le plasmide pAttb-ParB1-GFP au niveau du site XbaI avec le système InFusion de chez Takara pour obtenir la construction **pAttb-UAST-ParB1-GFP-*parS1modif***.

Le fragment *parS2*, amplifié par PCR à partir de pAttb-*parS2*-hsLacZ avec le couple d'amorces [*newparS2*-fw/*newparS2*-rev] est introduit dans le plasmide pAttb-ParB2-GFP et pAttb-ParB2-

mCherry au niveau du site XbaI avec le système InFusion de chez Takara pour obtenir les constructions **pAttb-UAST-ParB2-GFP-*parS2*** et **pAttb-UAST-ParB2-mCherry-*parS2***.

La construction **pAttb-*parS2-sha3-hsLacZ*** a été obtenue par ligation de la séquence *sha3* provenant de la digestion du plasmide pCaSpeR-*sha3* (Menoret, 2013) par NdeI/XbaI dans le plasmide pAttb-ParB2-GFP coupé par les enzymes NdeI/XbaI avec élimination de la séquence codant pour la GFP.

La construction **pAttb-*parS2-sha2-hsLacZ*** a été réalisée par insertion du fragment PCR *sha2* obtenu à partir du plasmide pCaSpeR-*sha123* avec le couple d'amorces [*sha2-fw/sha2-rev*], dans le plasmide pAttb-*parS2-hsLacZ* au niveau du site XbaI avec le système InFusion de chez Takara.

La construction **pAttb-*parS2-priAse-hsLacZ*** a été réalisée par insertion du fragment PCR *parS2* obtenu à partir du plasmide pAttb-*parS2-hsLacZ* avec le couple d'amorces [*parS2-priAse-fw/parS2-priAse-rev*], dans le plasmide pAttb-*priAse-hsLacZ* (Dib, 2021) au niveau du site HindIII avec le système InFusion de chez Takara.

La construction **pAttb-*parS2-sha3-hsLacZ-intron-yellow-MS2*** a été obtenue par l'assemblage par recombinaison de quatre fragments avec le système InFusion de Takara, selon les instructions du fabricant. Les quatre fragments sont :

- le plasmide pAttb-*parS2-sha3-hsLacZ* linéarisé par digestion avec l'enzyme Bsu36I dont le site est situé dans le gène *lacZ*.
- la partie 5' de l'intron du gène *yellow* obtenue par PCR à partir du plasmide pL19 fourni par l'équipe de Mounia Lagha à Montpellier, avec le couple d'amorces [*yellow-1<sup>er</sup>-intron-fw/yellow-1<sup>er</sup>-intron-rev*], contenant en 5' le site de restriction Bsu36I et en 3' le site de restriction BamHI.
- le fragment MS2, contenant 128 répétitions de la séquence MS2, provenant de la digestion par les enzymes de restriction BamHI et NheI du plasmide pL19.
- la partie 3' de l'intron du gène *yellow*, synthétisé par la société Integrated DNA Technologies (IDT), contenant en 5' le site de restriction NheI et en 3' le site de restriction Bsu36I.

La construction **pAttb-*parS2*-*hsLacZ*-intron-yellow-MS2** a été obtenue par la ligation, avec la T4 DNA ligase d'Invitrogen, suivant le protocole du fabricant, du fragment *parS2* HindIII/NotI provenant du plasmide pAttb-*parS2*-*hsLacZ* avec le plasmide pAttb-*parS2*-*sha3*-*hsLacZ*-intron-yellow-MS2 digéré par HindIII/NotI (élimination du fragment *parS2*-*sha3*).

La construction **pAttb-*parS2*-*priAse*-*hsLacZ*-intron-yellow-MS2** a été réalisée par insertion du fragment PCR *priAse* obtenu à partir du plasmide pAttb-*priAse*-*hsLacZ* avec le couple d'amorces [*priAse*-NotI-fw/*priAse*-NotI-rev], dans le plasmide pAttb-*parS2*-*hsLacZ*-intron-yellow-MS2 au niveau du site NotI avec le système InFusion de chez Takara.

La construction **pAct-MCP-GFP** a été obtenue par ligation de la séquence MCP-GFP provenant de la digestion KpnI/XbaI du plasmide pL402-MCP-GFP, fourni par l'équipe de Mounia Lagha à Montpellier, dans le plasmide pAct5C coupé par les enzymes KpnI/XbaI.

La construction **pCaSpeR-*parS1*modif-AB-*parS2*-*hsLacZ*** a été obtenue par l'assemblage par recombinaison de quatre fragments avec le système InFusion de Takara, selon les instructions du fabricant. Les quatre fragments sont :

- le fragment NotI-AB-XhoI et le fragment XhoI-pCaSpeR-*hslacZ*-NotI provenant de la digestion du plasmide pCaSpeR-AB-*hsLacZ* par XhoI/NotI.
- le fragment *parS1*modif(AB) contenant les sites XhoI en 5' et en 3', obtenu par PCR à partir du plasmide pAttb-*parS1*modif-*hsLacZ* avec le couple d'amorces [*parS1*modif-AB-fw/*parS1*modif-AB-rev].
- le fragment *parS2*(AB) contenant les sites NotI en 5' et en 3', obtenu par PCR à partir du plasmide pAttb-*parS2*-*hsLacZ* avec le couple d'amorces [*parS2*-AB-fw/*parS2*-AB-rev].

Les fragments ParB1-GFP, ParB1-mCherry, ParB2-GFP et ParB2-mCherry, issus respectivement des plasmides pAttb-ParB1-GFP, pAttb-ParB1-mCherry, pAttb-ParB2-GFP et pAttb-ParB2-mCherry par digestion KpnI/XbaI ont été clonés dans le vecteur pAct5C au niveau des sites KpnI et XbaI pour obtenir les constructions plasmidiques **pAct-ParB1-GFP**, **pAct-ParB1-mCherry**, **pActParB2-GFP** et **pAct-ParB2-mCherry**.

Le fragment *parS1*modif obtenu par PCR à partir du plasmide pAttb-*parS1*modif-*hsLacZ* avec le couple d'amorces [*newparS2*-fw/*newparS2*-rev] a été cloné dans les vecteurs pAct-ParB1-GFP et pAct-ParB1-mCherry au niveau du site XbaI avec le système InFusion de chez Takara

pour obtenir les constructions **pAct-ParB1-GFP-*parS1modif*** et **pAct-ParB1-mCherry-*parS1modif***.

Le fragment *parS2* obtenu par PCR à partir du plasmide pAttb-*parS2*-hsLacZ avec le couple d'amorces [*newparS2-fw/newparS2-rev*] a été cloné dans les vecteurs pAct-ParB2-GFP et pAct-ParB2-mCherry au niveau du site XbaI avec le système InFusion de chez Takara pour obtenir les constructions **pAct-ParB2-GFP-*parS2*** et **pAct-ParB2-mCherry-*parS2***.

Le fragment *parS2-sha3* obtenu par PCR à partir du plasmide pAttb-*parS2-sha3*-hsLacZ avec le couple d'amorces [*newparS2-fw/newparS2-rev*] a été cloné dans les vecteurs pAct-ParB2-GFP et pAct-ParB2-mCherry au niveau du site XbaI avec le système InFusion de chez Takara pour obtenir les constructions **pAct-ParB2-GFP-*parS2-sha3*** et **pAct-ParB2-mCherry-*parS2-sha3***.

Le fragment *parS2-priAse* obtenu par PCR à partir du plasmide pAttb-*parS2-priAse*-hsLacZ avec le couple d'amorces [*act-parS2-priAse-fw/act-parS2-priAse-rev*] a été cloné dans les vecteurs pAct-ParB2-GFP et pAct-ParB2-mCherry au niveau du site XbaI avec le système InFusion de chez Takara pour obtenir les constructions **pAct-ParB2-GFP-*parS2-priAse*** et **pAct-ParB2-mCherry-*parS2-priAse***.

Le fragment *priAse*-hsLacZ obtenu par PCR à partir du plasmide pAttb-*priAse*-hsLacZ (Dib, 2021) avec le couple d'amorces [*priAse-hslacZ-fw/priAse-hsLacZ-rev*] a été cloné dans les vecteurs pAct-ParB2-GFP-*parS2* et pAct-ParB2-mCherry-*parS2* au niveau du site PspXI avec le système InFusion de chez Takara pour obtenir les constructions **pAct-ParB2-GFP-*parS2-priAse*-hsLacZ** et **pAct-ParB2-mCherry-*parS2-priAse*-hsLacZ**.

Le fragment hsLacZ obtenu par PCR à partir du plasmide pAttb-*priAse*-hsLacZ (Dib, 2021) avec le couple d'amorces [*hslacZ-fw/priAse-hsLacZ-rev*] a été cloné dans les vecteurs pAct-ParB2-GFP-*parS2-sha3* et pAct-ParB2-mCherry-*parS2-sha3* au niveau du site XbaI avec le système InFusion de chez Takara pour obtenir les constructions **pAct-ParB2-GFP-*parS2-sha3*-hsLacZ** et **pAct-ParB2-mCherry-*parS2-sha3*-hsLacZ**.

Le fragment *ins1* obtenu par PCR à partir du plasmide pAttb-hsLacZ-*ins* avec le couple d'amorces [*ins1-HindIII-fw/ins1-HindIII-rev*] est cloné dans le vecteur pAttb-ParB2-GFP-*parS2-priAse*-hsLacZ au niveau du site HindIII avec le système InFusion de chez Takara pour obtenir la construction **pAct-ParB2-GFP-*ins1-parS2-priAse*-hsLacZ**.

Le fragment *ins2* obtenu par PCR à partir du plasmide pAttb-hsLacZ-*ins* avec le couple d'amorces [*ins2*-PspOMI-fw/*ins2*-PspOMI-rev] est cloné dans le vecteur pAttb-ParB2-GFP-*parS2*-*priAse*-hsLacZ au niveau du site PspOMI avec le système InFusion de chez Takara pour obtenir la construction **pAct-ParB2-GFP-*parS2*-*priAse*-hsLacZ-*ins2***.

La construction **pAct-ParB2-GFP-*ins1*-*parS2*-*priAse*-hsLacZ-*ins2*** est réalisée par l'assemblage par recombinaison de quatre fragments avec le système InFusion de Takara : les fragments *ins1* et *ins2* obtenue par PCR et le plasmide pAct-ParB2-GFP-*parS2*-*priAse*-hsLacZ digéré par HindIII/PspOMI.

Les lignées transgéniques (liste en **annexe2**) ont été générées soit par la plateforme de transgénése de l'unité soit par la société BestGene Inc., en utilisant le système PhiC31. Les protéines ParB sont exprimées dans l'épiderme embryonnaire de drosophile avec le pilote *ptc-Gal4*, dans les glandes salivaires avec les pilotes, *sgs3-Gal4* ou *332.3-Gal4* ou *C147-Gal4*, dans les disques imaginaux de patte avec le pilote *en-Gal4* et dans les cellules réabsorbantes des tubules de Malpighi de drosophile adulte avec le pilote *C507-Gal4*.

Amorces	Séquences
<b>parB4-fw</b> <b>parB4-rev</b>	ATTCGTTAACAGATCTATGAAGCGTGCGCCTGTTATTCC CCCTTGCTCACCATATGGGGTGCTGGCTTTTCAAGTTCC
<b>parB4-mVenus-fw</b> <b>parB4-mVenus-rev</b>	AGGGAATTGGGAATTCATATGAAGCGTGCGCCTGTTATTCC AAAGATCCTCTAGAGGTACCTTACTTGTACAGCTCGTCCATGCCG
<b>parS4-fw</b> <b>parS4-rev</b>	AGGGAATTGGGAATTCATATGAAGCGTGCGCCTGTTATTCC GGATCCGCTAGCGGCCGCTTGACAGGCATTGATGGAATC
<b>newparS2-fw</b> <b>newparS2-rev</b>	AAGGGGATCCTCTAGAGCTTGGATTTCACTGGAAGTAGG ACAAAGATCCTCTAGATTATACTCCGGCGCTCCTCGAG
<b>sha2-fw</b> <b>sha2-rev</b>	GAAGTTCAAGGAATTCCACCTGCCGTGGAACTTTTCGAAAG GGATCCGCTAGCGGCCGATCCGAACACTGAAGACTGTAGACTG
<b>parS2-priAse-fw</b> <b>parS2-priAse-rev</b>	TAGCGGATCCAAGCTTGCTTGGATTTCACTGGAAGTAGG TTAGTGGTGTAAGCTTTTATACTCCGGCGCTCCTCGAG
<b>yellow-1<sup>er</sup>-intron-fw</b> <b>yellow-1<sup>er</sup>-intron-rev</b>	GTGCGATCTTCTGAGGCCCGTGATGGTAAGTGAAGTTAAATATG ATGTACTAGTGATCCCCATGGGCTTTTTTAAATCAAATTTATTTC
<b>priAse-NotI-fw</b>	CGGTACCCCGCGGCCGCTTACACCACTAAAAAGTCCAAAAAGAAGG



<i>priAse-NotI-rev</i>	GGATCCGCTAGCGGCCGGAATTCAGCATGAATATTCGGAGAC
<i>parS1modif-AB-fw</i>	GATCAGATCCGCGCCGCGCTTGGATTTCACTGGAAGTAGG
<i>parS1modif-AB-rev</i>	TAGTTCTAGAGCGGCCGCTTGTGAATTGAATTGTCGCTC
<i>parS2-AB-fw</i>	TACCGTCGACCTCGAGGCTTGGATTTCACTGGAAGTAG
<i>parS2-AB-rev</i>	TCCGGCGCTCCTCGAGCACTTTGCTTGTGAATTGAATTGTCG
<i>act-parS2-priAse-fw</i>	AAGGGGATCCTCTAGAGAATACATTTAATTTAGAAAATGCTTGG
<i>act-parS2-priAse-rev</i>	GCGCGCCGGCTCTAGAAGCATGAATATTCGGAGACAATG
<i>priAse-hsLacZ-fw</i>	TAGCGGATCCCTCGAGGACACCACTAAAAAGTCCAAAAAGAAGG
<i>priAse-hsLacZ-rev</i>	TCCGGCGCTCCTCGAGGTTATTTTTGACACCAGACCAACTG
<i>hs-LacZ-fw</i>	TAGCGGATCCCTCGAGGAGCGCCGGAGTATAAATAGAGG
<i>Ins1-HindIII-fw</i>	TAGCGGATCCAAGCTTGGGCTGCATCACGTAATAAG
<i>Ins1-HindIII-rev</i>	GCAGGCATGCAAGCTTGAGACAATTGATCGGCTAAATGG
<i>Ins2-PspOMI-fw</i>	ATAATCTAGAGGGCCCTCACGTAATAAGTGTGCGTTG
<i>Ins2-PspOMI-rev</i>	TACCTTCAAGGGCCCAATTGATCGGCTAAATGGTATGGC

**Tableau 5 : Liste des amorces utilisées pour obtenir les différentes constructions plasmidiques citées dans ce mémoire de thèse.**

### **Immunocoloration sur embryons, larves.**

Fixation et dévitellinisation des embryons : les embryons, âgés de 14 à 18h de développement sont déchorionnées pendant 2min dans la javel 37%/H<sub>2</sub>O, volume à volume, puis transférés dans des tubes corex contenant 5ml d'heptane. Ils sont fixés avec l'ajout de 5ml d'une solution de paraformaldéhyde 4% dans du PBS1X pendant 20min sous agitation à 22°C. Après avoir éliminé la phase inférieure (formaldéhyde), les embryons sont dévitellinisés avec l'addition de 5ml de méthanol puis par une forte agitation pendant une quarantaine de seconde. Les embryons dévitellinisés tombent au fond du tube. Après avoir éliminé les deux phases, ils sont rincés trois fois avec 5ml de méthanol. Ils peuvent être stockés dans du méthanol à -20°C.

Immunocoloration sur embryons : les embryons sont réhydratés progressivement et perméabilisés avec du PBT (PBS 1X, Triton X-100 0,1%), puis trois lavages de 5 min sont effectués dans du PBT avant d'être incubés pendant 1heure sous agitation à 22°C dans du PAT (PBS 1X, BSA 1%, Triton X-100 1%) afin de saturer les sites non spécifiques. Les embryons sont

incubés avec l'anticorps primaire dans du PAT pendant toute la nuit, sous agitation, à 4°C. La concentration des différents anticorps primaires utilisés est : anti-GFP (mouse) 1/800<sup>e</sup> ; anti-mCherry (rat) 1/800<sup>e</sup> ; anti-Cut (mouse) 1/200<sup>e</sup> ; anti-dyl (rabbit) 1/600<sup>e</sup> ; anti-lamine A (mouse) 1/200<sup>e</sup> ; anti-βgalactosidase (rabbit) 1/1000<sup>e</sup>. Après trois rinçages de 10 min dans du PAT puis deux de 10min dans du PBT, les embryons sont incubés avec les anticorps secondaires dans du PBT pendant deux heures à 22°C. Les anticorps secondaires, Alexa-Goat-Anti-Rabbit 488/555/647 (GAR488 ou GAR555 ou GAR647), ou Alexa-Goat-Anti-Mouse 488/555/647 (GAM488 ou GAM555 ou GAM647) ou Alexa-Goat-Anti-Rat 555 (GARat555) sont utilisés au 1/800<sup>e</sup> sauf pour les anticorps GAR647 et GAM647 qui sont utilisés au 1/200<sup>e</sup>. Les noyaux peuvent être colorés avec du TO-PRO et visualisés dans du rouge lointain. Après trois rinçages de 10 min avec du PBT, les embryons sont montés entre lame et lamelle dans du milieu de montage Vectashield.

Les observations et acquisitions sont effectuées sur un microscope confocal Leica SPE.

Immunocoloration sur glandes salivaires : 20 à 30 mouches sont mises dans des pondoirs où les femelles vont pondre leurs œufs, pendant huit heures à 25°C, dans des boîtes de Petri de 45mm de diamètre contenant de l'agar et du rouge ponceau sur lequel de la levure a été déposée. Les pondoirs sont laissés vieillir pendant 72heures à 25°C. Les larves possédant le bon génotype sont triées sous loupe à fluorescence. Les glandes salivaires sont disséquées dans du PBS1X froid puis fixées pendant 20 minutes à 22°C dans une solution de paraformaldéhyde 4% dans du PBS1X. Après un rinçage rapide dans du PBS1X, trois lavages de 5 min dans du PBS1X sont effectués suivi par un lavage rapide avec du PAT avant d'incuber les glandes salivaires pendant 1heure sous agitation à 22°C dans du PAT afin de saturer les sites non spécifiques. Puis l'immunocoloration des glandes salivaires s'effectue comme pour les embryons. Pour les protéines fusionnées à des protéines fluorescentes comme la GFP ou la mCherry, il n'est obligé d'utiliser un anticorps dirigé contre ces protéines. La fixation n'altère pas leur pouvoir de fluorescence. Comme pour les embryons, les noyaux peuvent être colorés avec du TO-PRO et visualisés dans du rouge lointain. Après trois rinçages de 10 min avec du PBT, les glandes salivaires sont montées entre lame et lamelle dans du milieu de montage Vectashield.

Pour effectuer une observation dans des cellules vivantes, les glandes salivaires sont montées directement, sans fixation, entre lame et lamelle dans du Vectashield.

Les observations et acquisitions sont effectuées sur un microscope confocal Leica SPE.

### **Immunocoloration sur cellule de drosophile S2 :**

Les cellules de drosophile S2 sont cultivées dans du milieu de culture Schneider supplémenté avec 10% de sérum de veau foetal et 1% de pénicilline/streptomycine.

Transfection transitoire en cellules S2 : les expériences de transfection transitoire dans les cellules S2 sont réalisées dans des boites de 24 puits. Mettre une lamelle de 14mm de diamètre dans chaque puit, puis déposer 200.000 cellules dans 500µl de milieu par puit. Le mélange des différents plasmides à transférer (50ng par plasmide) est effectué dans de l'Opti-MEN™ (Gibco) dans un volume de 21µl puis 0.6µl de réactif de transfection FuGENE<sup>HD</sup> (Promega) est rajouté en dernier. Après 20min à 22°C, le mélange à transférer est déposé dans chaque puit puis mélangé en faisant des 8 et incubé pendant 48h à 25°C. S'il est nécessaire d'induire l'expression par l'ecdysone, au bout de 24h, 5nmol de 20-HydroxyEcdysone (soit 10µM final) est ajouté dans chaque puit à partir d'une solution de 5mg/ml préparé dans l'éthanol.

Immunocoloration sur cellules S2 : après 48h, retirer le milieu de culture des cellules avec une pipette automatique et rajouter 250µl de solution de paraformaldéhyde 4% dans du PBS1X pour fixer les cellules. Au bout de 10min, retirer la solution de paraformaldéhyde à la pipette et effectuer deux rinçages avec 500µl de PBS1X. Après un rinçage rapide avec 500µl de PBT 0,1% (PBS 1X, Triton X-100 0,1%), saturer les sites non spécifiques avec 500µl de PAT 0,1% (PBS1X, 50mg de BSA, Triton X-100 0,1%) pendant 1h à 22°C. Prélever la lamelle et la déposer sur un parafilm où l'élimination des diverses solutions (anticorps ou rinçages) s'effectuera à l'aide d'une pompe aspirante. 50µl de solutions contenant les anticorps primaires dans du PAT 0,1% sont déposées sur chaque lamelle et incubé pendant 2h à l'abri de la lumière. Après 4 rinçages de 10min avec 75µl de PBT 0,1%, les cellules sont incubées avec 50µl de solutions contenant les anticorps secondaires dans du PAT 0,1%.

La concentration des différents anticorps primaires utilisés est : anti-GFP (mouse) 1/500<sup>e</sup> ; anti-mCherry (rat) 1/500<sup>e</sup> ; anti-lamine A (mouse) 1/250<sup>e</sup> ; anti- $\beta$ galactosidase (rabbit) 1/2000<sup>e</sup> ; anti-H3K27ac (rabbit) 1/1000<sup>e</sup> ; anti-H3K27me3 (rabbit) 1/500<sup>e</sup> ; anti-Su(Hw) (chèvre) 1/200<sup>e</sup> fourni par la Professeure Pamela Geyer (Université de l'Iowa). Les anticorps secondaires sont utilisés avec la même concentration que pour les embryons ou les glandes salivaires. Pour les protéines fusionnées aux protéines fluorescentes GFP ou mCherry, la plupart des images ont été obtenues sans utiliser d'anticorps dirigé contre elles. Les noyaux peuvent être colorés avec du TO-PRO et visualisés dans du rouge lointain. Les lamelles sont déposées sur une lame avec les cellules côté lame, dans 3 $\mu$ l de Vectashield.

Les observations et acquisitions sont effectuées sur un microscope confocal Leica SPE.

## Annexe 4 : Liste des publications obtenues pendant la thèse

- Dib, A., Zanet, J., Mancheno-Ferris, A., Gallois, M., Markus, D., Valenti, P., Marques-Prieto, S., Plaza, S., Kageyama, Y., Chanut-Delalande, H., and Payre, F. (2021). Pri smORF Peptides Are Wide Mediators of Ecdysone Signaling, Contributing to Shape Spatiotemporal Responses. *Front Genet.* 12:714152.
- Montigny, A., Tavormina, P., Duboe, C., San Clémente, H., Aguilar, M., Valenti, P., Laressergues, D., Combier, J.P., and Plaza, S. (2021). *Drosophila* primary microRNA-8 encodes a microRNA-encoded peptide acting in parallel of miR-8. *Genome Biol.* 22(1):118.
- Ray, S., Rosenberg, M.I., Chanut-Delalande, H., Decaras, A., Schwertner, B., Toubiana, W., Auman, T., Schnellhammer, I., Teuscher, M., Valenti, P., Khila, A., Klingler, M., and Payre, F. (2019). The mlpt/Ubr3/Svb module comprises an ancient developmental switch for embryonic patterning. *Elife.* 8:e39748.
- Bohère, J., Mancheno-Ferris, A., Al Hayek, S., Zanet, J., Valenti, P., Akino, K., Yamabe, Y., Inagaki, S., Chanut-Delalande, H., Plaza, S., Kageyama, Y., Osman, D., Polesello, C., and Payre, F. (2018). Shavenbaby and Yorkie mediate Hippo signaling to protect adult stem cells from apoptosis. *Nat Commun.* 9(1):5123.
- Gomez-Lamarca, M.J., Falo-Sanjuan, J., Stojnic, R., Abdul Rehman, S., Muresan, L., Jones, M.L., Pillidge, Z., Cerda-Moya, G., Yuan, Z., Baloul, S., Valenti, P., Bystricky, K., Payre, F., O'Holleran, K., Kovall, R., and Bray, S.J. (2018) Activation of the Notch Signaling Pathway In Vivo Elicits Changes in CSL Nuclear Dynamics. *Dev Cell.* 44(5):611-623.e7.
- Salis, P., Payre, F., Valenti, P., Bazellieres, E., Le Bivic, A., and Mottola, G. (2017). Crumbs, Moesin and Yurt regulate junctional stability and dynamics for a proper morphogenesis of the *Drosophila* pupal wing epithelium. *Sci Rep.* 7(1):16778.

## **Références bibliographiques**

**Abeles, A.L., Friedman, S.A., and Austin, S.J.** (1985). Partition of unit-copy miniplasmids to daughter cells. III. The DNA sequence and functional organization of the P1 partition region. *J. Mol. Biol.* 185(2): 261-272.

**Actualités, Institut Pasteur** (2021). Découverte de l'ARN messenger, en 1961. <https://www.pasteur.fr/fr/journal-recherche/actualites/decouverte-arn-messenger-1961>

**Adams, D., Shekhtman, E., Zechiedrich, E., Schmid, M., and Cozzarelli, N.** (1992). The role of topoisomerase-IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. *Cell.* 71(2):277–288.

**Adams, M.D., ..., and Venter, J.G.** (2000). The genome sequence of *Drosophila melanogaster*. *Science.* 287(5461): 2185-2195.

**Adams, D.W., Wu, L.J., and Errington, J.** (2015). Nucleoid occlusion protein Noc recruits DNA to the bacterial cell membrane. *EMBO J.* 34(4):491–501.

**Aerts, S.** (2012). Chapter five - Computational Strategies for the Genome-Wide Identification of cis-Regulatory Elements and Transcriptional Targets. In: Plaza S, Payre F (eds) *Current Topics in Developmental Biology*. Academic Press, pp 121–145

**Ah-Seng, Y., Lopez, F., Pasta, F., Lane, D., and Bouet, J.Y.** (2009). Dual role of DNA in regulating ATP hydrolysis by the SopA partition protein. *J Biol Chem.* 284(44): 30067-30075.

**Akhtar, P., Anand, S.P., Watkins, S.C., and Khan, S.A.** (2009). The tubulin-like RepX protein encoded by the pXO1 plasmid forms polymers in vivo in *Bacillus anthracis*. *J Bacteriol.*191:2493–2500.

**Al Hayek, S., Alsawadi, A., Kambris, Z., Boquete, J.P., Bohere, J., Immarigeon, C., Ronsin, B., Plaza, S., Lemaitre, B., Payre, F., and Osman, D.** (2021). Steroid-dependent switch of OvoL/Shavenbaby controls self-renewal versus differentiation of intestinal stem cells. *EMBO J.* 40(4): e104347.

**Andrews, J., Garcia-Estefania, D., Delon, I., Lu, J., Mevel-Ninio, M., Spierer, A., Payre, F., Pauli, D., and Oliver, B.** (2000). OVO transcription factors function antagonistically in the *Drosophila* female germline. *Development.* 127(4): 881-892.

- Anonymous.** (1970). Central dogma reversed. *Nature*. 226(5252):1198-1199.
- Apgar, J., Holley, R.W., and Merrill, S.H.** (1962). Purification of the alanine-, valine-, histidine-, and tyrosine-acceptor ribonucleic acids from yeast. *J. Biol. Chem.* 237(3): 796-802.
- Arber, W., and Linn, S.** (1969). DNA modification and restriction. *Annu. Rev. Biochem.* 38: 476-500.
- Astbury, W.T., and Bell, F.O.** (1938a). X-ray study of thymonucleic acid. *Nature*. 141(3573): 747-748.
- Astbury, W.T., and Bell, F.O.** (1938b). Some recent developments in the X-ray study of proteins and related structures. *Cold Spring Harb Symp. Quant. Biol.* 6: 109-121
- Aulner, N., Monod, C., Mandicourt, G., Jullien, D., Cuvier, O., Sall, A., Janssen, S., Laemmli, U.K., and Käs, E.** (2002). The AT-hook protein D1 is essential for *Drosophila melanogaster* development and is implicated in position-effect variegation. *Mol Cell Biol* 22(4) :1218–1232.
- Austin, S., and Abeles, A.** (1983a). Partition of unit-copy miniplasmids to daughter cells. I. P1 and F miniplasmids contain discrete, interchangeable sequences sufficient to promote equipartition. *J. Mol. Biol.* 169(2): 353-372.
- Austin, S., and Abeles, A.** (1983b). Partition of unit-copy miniplasmids to daughter cells. II. The partition region of miniplasmid P1 encodes an essential protein and a centromere-like site at which it acts. *J. Mol. Biol.* 169(2): 373-387.
- Avery, O.T., MacLeod, C.M., and McCarty, M.** (1944). Studies on the chemical nature of the substance inducing transformation of Pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *J. Exp. Med.* 79(2): 137-158
- Aylett, C.H., and Lowe, J.** (2012). Superstructure of the centromeric complex of TubZRC plasmid partitioning systems. *Proc Natl Acad Sci.* 109(41): 16522–16527.
- Azam, T.A., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A.** (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.* 181(20): 6361–70.



- Badrinarayanan, A., Le, T.B., and Laub, M.T.** (2015). Bacterial chromosome organization and segregation. *Annu. Rev. Dev. Biol.* 31: 171-199.
- Baehrecke, E.H., (1996).** Ecdysone signaling cascade and regulation of *Drosophila* metamorphosis. *Archives of Insect Bioch. and Physio.* 33: 231-244
- Baltimore, D.** (1970). RNA-dependent DNA polymerase in virions of RNA tumor viruses. *Nature.* 226(5252): 1209-1211.
- Bai, H., Lester, G.M.S., Petishnok, L.C., and Dean, D.A.** (2017). Cytoplasmic transport and nuclear import of plasmid DNA. *Biosci Rep.* 37(6): BSR20160616.
- Baltzer, F.** (1964). Theodor Boveri. *Sciences.* 144(3260): 809-815.
- Bannister, A.J., and Kouzarides, T.** (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21(3) : 381–395.
- Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K, Schneider, D., Lenski, R.E, and Kim, J.F.** (2009). Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature.* 461(7268):1243-1247.
- Bauman, J., Wiegant, J., Borst, P., and van Duijn, (1980).** A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. *P. Exp Cell Res.* 128(2):485-490.
- Bauman, J.G., Wiegant, J., Van Duijn, P., Lubsen, N.H., Sondermeijer, P.J., Hennig, W., and Kubli, E.** (1981). Rapid and high resolution detection of *in situ* hybridisation to polytene chromosomes using fluorochrome-labeled RNA. *Chromosoma.* 84(1):1-18.
- Baxter, J.C, and Funnell, B.E.** (2014). Plasmid partition mechanisms. *Microbiol. Spectr.*2(6): PLAS-0023-2014.
- Becker, S., and Boch, J.** (2021). TALE and TALEN genome editing technologies. *Gene and Genome Editing.* (2021) 100007.
- Benos, P.V., ..., Valenti, P., ..., and Glover. D.M.** (2000). From sequence to chromosome: the tip of the X chromosome of *D. melanogaster*. *Science.* 287(5461): 2220-2222.

**Berche, P.** (2016). Histoire de la biologie moléculaire. Revue de Biologie Médicale. 333 – novembre 2016: 41-51.

**Berg, P., and Ofengand, E.J.** (1958). An enzymatic mechanism for linking amino acids to RNA. Proc. Natl. Acad. Sci. 44(2): 78-86.

**Bernhardt, T.G., and de Boer, P.A.J.** (2005). SlmA, a nucleoid-associated, FtsZ binding protein required for blocking septal ring assembly over chromosomes in *E. coli*. Mol. Cell 18(5):555–564.

**Berry, C., O’neil, S., Ben-Dov, E., Jones, A.F., Murphy, L., Quail, M.A., Holden, M.T., Harris, D., Zaritsky, A., and Parkhill, J.** (2002). Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis subsp. israelensis*. Appl Environ Microbiol 68:5082–5095.

**Bigot, S., Saleh, O.A., Lesterlin, C., Pages, C., El Karoui, M., Dennis, C., Grigoriev, M., Allemand, J.F., Barre, F.X., and Cornet, F.** (2005). KOPS: DNA motifs that control *E. coli* chromosome segregation by orienting the FtsK translocase. EMBO J. 24(21): 3770–3780.

**Bird, A.P.** (1995). Gene number, noise reduction and biological complexity. Trends Genet. 11(3): 94-100.

**Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K.** (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc. Natl. Acad. Sci. 104(9) : 3312-3317.

**Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U.** (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science. 326(5959): 1509-1512.

**Bohere, J. (2017).** Etude fonctionnelle de Shavenbaby dans l’homéostasie du système rénal chez la drosophile. Génétique animale. Université Paul Sabatier - Toulouse III, 2017. Français. NNT :2017TOU30158.

**Boivin, A., and Vendrely, R.** (1947). Sur le rôle possible des deux acides nucléiques dans la cellule vivante. Experientia. 3(1): 32-34.

- Boopathi, R., Dimitrov, S., Hamiche, A., Petosa, C., and Bednar, J.** (2020). Cryo-electron microscopy of the chromatin fiber. *Curr Opin Struct Biol.* 64 : 97–103.
- Bouet, J.Y., Nordström, K., and Lane, D.** (2007). Plasmid partition and incompatibility—the focus shifts. *Mol Microbiol.* 65(6): 1405–1414.
- Bouet, J.Y., Stouf, M., Lebailly, E., and Cornet, F.** (2014) Mechanisms for chromosome segregation. *Curr Opin Microbiol.* 22: 60-65.
- Bouet, J.Y., and Funnell, B.E.** (2019). Plasmid Localization and Partition in *Enterobacteriaceae*. *EcoSal Plus.* Jun;8(2).
- Brachet, J., and Chantrenne, H.** (1956). The function of the nucleus in the synthesis of cytoplasmic proteins. *Cold Spring Harb Symp Quant Biol.* 21: 329-237.
- Brand, A.H, and Perrimon, N.** (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development.* 118(2):401-415.
- Breathnach, R., Mandel, J.L., and Chambon, P.** (1977). Ovalbumin gene is split in chicken DNA. *Nature.* 270(5635): 314-319.
- Brenner, S.** (1957). On the impossibility of all overlapping triplet codes in information transfer from nucleic acid to proteins. *Proc. Natl. Acad. Sci.* 43(8): 687-694.
- Brenner, S., Jacob, F., and Meselson, M.** (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature.* 190(4776): 576-581.
- Brenner, S, Stretton, A.O.W., and Kaplan, S.** (1965). Genetic Code: The ‘Nonsense’ Triplets for Chain Termination and their Suppression. *Nature.* 206(4988): 994-998.
- Brenner, S, Barnett, L., Katz, E.R., and Crick, F.H.C.** 1967. UGA: a third nonsense triplet in the genetic code. *Nature.* 213(5075): 449-450.
- Bridges, C.B.,** (1935). Salivary chromosome maps: with a key to the banding of the chromosomes of *Drosophila melanogaster*. *J. Hered.* 26: 60-64.
- Bridges, C.B.,** (1938). A revised map of the salivary gland X-chromosome of *Drosophila melanogaster*. *J. Hered.* 29: 11-13.

- Britton, R.A., Lin, D.C., and Grossman, A.D.** (1998). Characterization of a prokaryotic SMC protein involved in chromosome partitioning. *Genes Dev.* 12(9):1254–1259.
- Broedersz, C.P., Wang, X., Meir, Y., Loparo, J.J., Rudner, D.Z., and Wingreen, N.S.** (2014). Condensation and localization of the partitioning protein ParB on the bacterial chromosome. *Proc. Natl. Acad. Sci.* 111(24): 8809–8814.
- Brooks, A.C, and Hwang, L.C.** 2017). Reconstitutions of plasmid partition systems and their mechanisms. *Plasmid.* 91 : 37-41.
- Bruch, J.F., Eloit, M., Tabareau-Delalande, F., El Gani, M., Kervarrec, T., Pettereau, A., et Jousset, S.** (2019). Étude historique de la visualisation de l'acide désoxyribonucléique. *Revue Francophone des Laboratoires.* 2019 (511): 71-80.
- Buchler, N.E., Gerland, U., and Hwa, T.** (2003). On schemes of combinatorial transcription logic. *Proc Natl Acad Sci.* 100(9): 5136–5141.
- Bulger, M., and Groudine, M.** (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell.* 144(3) :327–339.
- Buongiorno-Nardelli, M., and Amaldi, F.** (1970). Autoradiographic detection of molecular hybrids between RNA and DNA in tissue sections. *Nature.* 225(5236):946-948.
- Bystricky, K.** (2015). Chromosome dynamics and folding in eukaryotes: Insights from live cell microscopy. *FEBS lett.* 589(2015):3014-3022.
- Carbonell, A., Mazo, A., Serras, F., and Corominas, M.** (2013). Ash2 acts as an ecdysone receptor coactivator by stabilizing the histone methyltransferase Trr. *Mol Biol Cell.* 24(3) :361–372.
- Carter, N.P., Ferguson-Smith, M.A., Perryman, M.T., Telenius, H., Pelmear, A.H, Leversha, M.A., Glancy, M.T., Wood, S.L., Cook, K., and Dyson, H.M.** (1992). Reverse chromosome painting: a method for the rapid analysis of aberrant chromosomes in clinical cytogenetics. *J Med Genet.* 29(5): 299-307.
- Castaing, J.P., Bouet, J.Y., and Lane, D.** (2008). F plasmid partition depends on interaction of SopA with non-specific DNA. *Mol Microbiol.* 70(4): 1000–1011.

**Cerqueira, A.V., and Lemos, B.** (2019). Ribosomal DNA and the Nucleolus as keystones in nuclear architecture, organization, and function. *Trends Genet.* 35(10) : 710–723.

263(5148):802-5.

**Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C.** (1994). Green fluorescent protein as a marker for gene expression. *Science.* 263(5148):802-805.

**Chang, B., Chen, Y., Zhao, Y., and Bruick, R.K.** (2007). JMJD6 is a histone arginine demethylase. *Science.* 318(5849) : 444–447.

**Chanut-Delalande, H., Fernandes, I., Roch, F., Payre, F., and Plaza, S.** (2006). Shavenbaby couples patterning to epidermal cell shape control. *PLoS biology.* 4(9): e290.

**Chanut-Delalande, H., Hashimoto, Y., Pelissier-Monier, A., Spokony, R., Dib, A., Kondo, T., Bohere, J., Niimi, K., Latapie, Y., Inagaki, S., Dubois, L., Valenti, P., Polesello, C., Kobayashi, S., Moussian, B., White, K.P., Plaza, S., Kageyama, Y., and Payre, F.** (2014). Pri peptides are mediators of ecdysone for the temporal control of development. *Nat Cell Biol.* 16(11): 1035-1044.

**Chapeville, F., Lipmann, F., Von Ehrenstein, G., Weisblum, B., Ray, W.J., and Benzer, S.** (1962). On the role of soluble ribonucleic acid in coding for amino acids. *Proc. Natl. Acad. Sci.* 48(6): 1086-1092.

**Chargaff, E., Zamenhof, S., and Green, C.** (1950a). Composition of human desoxyribose nucleic acid. *Nature.* 165(4202): 756-757.

**Chargaff, E.** (1950b). Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia.* 6(6): 201-209.

**Chargaff, E., Lipshitz, R., Green, C., and Hodes, H.E.** (1951). The composition of the deoxyribonucleic acid of salmon sperm. *J. Biol. Chem.* 192(1): 223-230.

**Chargaff, E., Lipshitz, R., and Green, C.** (1952). Composition of the desoxyribose nucleic acids of four genera of sea-urchin. *J. Biol. Chem.* 195(1): 155-160.

**Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B.** (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155(7): 1479–1491.

- Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., and Gregor, T.** (2018). Dynamic interplay between enhancer-promoter topology and gene activity. *Nature Genet.* 50(9): 1296-1303.
- Chen, W., He, F., Zhang, X., Chen, Z., Wen, Y., and Li, J.** (2010). Chromosomal instability in *Streptomyces avermitilis*: major deletion in the central region and stable circularized chromosome. *BMC Microbiol.* 10: 198.
- Chieco, P., and Derenzini, M.** (1999). The Feulgen reaction 75 years on. *Histochem Cell Biol.* 111(5):345-58.
- Claret, L., and Rouviere-Yaniv, J.** (1997). Variation in HU composition during growth of *Escherichia coli*: the heterodimer is required for long term survival. *J. Mol. Biol.* 273(1):93–104.
- Claude, A.** (1943). The constitution of protoplasm. *Science.* 97(2525): 451-456.
- Cobb, M.** (2015). Who discovered messenger RNA? *Curr. Biol.* 25(13): R526-53.
- Cohen, S.S.** (1949). Growth requirements of bacterial viruses. *Bacteriol Rev.* 13(1): 1-24
- Cooper, V.S., Vohr, S.H., Wrocklage, S.C., and Hatcher, P.J.** (2010). Why genes evolve faster on secondary chromosomes in bacteria. *PLoS Comput Biol* 6: e1000732.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A., and Jaenisch, R.** (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci.* 107(50) : 21931–21936.
- Crick, F.H.C.** (1955). On degenerate templates and the adaptor hypothesis: a note for the RNA Tie Club. Unpublished note. Wellcome Library for the History and Understanding of Medicine. Francis Harry Compton Crick Papers. 11646.  
<http://resource.nlm.nih.gov/101584582X73>
- Crick, F.H.C.** (1956). Ideas on protein synthesis.  
<https://wellcomecollection.org/works/xmscu3g4>
- Crick, F.H.C.** (1966). The genetic code: III. *Sci. Am.* 215(4):55-60.

- Crick, F.H.C.** (1970). Central Dogma of molecular biology. *Nature*. 227(5258): 561-563.
- Crick, F.H.C., Griffith, J.S., and Orgel, L.E.** (1957). Codes without commas. *Proc. Natl. Acad. Sci.* 43(5): 416-421.
- Crick, F.H.C., and Brenner, S.** (1959). Some footnotes on protein synthesis: a note for the RNA TIE club. *Profiles in Medicine*. <https://profiles.nlm.nih.gov/ps/access/SCBBFV.pdf>.
- Crick, F.H.C, Leslie Barnett, Brenner, S, and Watts-Tobin, R.J.** (1961). General nature of the genetic code for proteins. *Nature*. 192(4809): 1227-1232.
- Cubeñas-Potts, C., Rowley, M.J., Lyu, X., Li, G., Lei, E.P., and Corces, G.** (2017). Different enhancer classes in *Drosophila* bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.* 45(4) : 1714–1730.
- Dahm, R.** (2005). Friedrich Miescher and the discovery of DNA. *Dev. Biol.* 15; 278 (2): 274-288.
- Dahm, R.** (2008). Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* 135(2): 387-399.
- Dam, M., and Gerdes, K.** (1994) Partitioning of plasmid R1. Ten direct repeats flanking the *parA* promoter constitute a centromere-like partition site *parC*, that expresses incompatibility. *J Mol Biol* 236: 1289–1298.
- Dame, R.T.** (2005). The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol Microbiol.* 4(56): 858–870.
- Danilova, O., Reyes-Lamothe, R., Pinskaya, M., Sherratt, D., and Possoz, C.** (2007). MukB colocalizes with the *oriC* region and is required for organization of the two *Escherichia coli* chromosome arms into separate cell halves. *Mol. Microbiol.* 65(6):1485–1492.
- Dann, O., Bergen, G., Demant, E., and Volz. G.** (1971). Trypanocide Diamidine des 2-Phenyl-benzofurans, 2-Phenyl-indens und 2-Phenyl-indol. *Ann. Chem* . 749: 68-89.
- Danna, K., and Nathans, D.** (1971). Specific cleavage of Simian Virus 40 DNA by restriction endonuclease of *Hemophilus Influenzae*. *Proc. Natl. Acad. Sci.* 68(12): 2913-2917.
- Debaugny, R.E, Sanchez, A., Rech, J., Labourdette, D., Dorignac, J., Geniet, F., Palmeri, J., Parmeggiani, A., Boudsocq, F., Anton Leberre, V., Walter, J.C., and Bouet, J.Y.** (2018). A

conserved mechanism drives partition complex assembly on bacterial chromosomes and plasmids. *Mol. Syst. Biol.* 14(11): e8516.

**Dean, D.A.** (1997). Import of plasmid DNA into the nucleus is sequence specific. *Exp. Cell Res.* 230(2): 293–302.

**Dean, D.A., Dean, B.S., Muller, S. and Smith, L.C.** (1999) Sequence requirements for plasmid nuclear entry. *Exp. Cell Res.* 253(2): 713–722.

**Delandre, C., and Marshall, O.J.** (2017). United colours of chromatin? Developmental genome organisation in flies. *Biochem. Soc. Trans.* 47(2): 691-700.

**Delker, R.K, Munce, R.H, Hu, M., and Mann, R.S.** (2022). Fluorescent labeling of genomic loci in *Drosophila* imaginal discs with heterologous DNA-binding proteins. *Cell Rep. Methods.* 2(3): 100175.

**Delon, I., Chanut-Delalande, H., and Payre, F.** (2003). The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in *Drosophila*. *Mech Dev.* 120(7): 747-758.

**Derenzini, M.** (1995). Ultrastructural cytochemistry. In: Morel G (ed) Visualization of nucleic acids. CRC Press, Boca Raton, pp 69–93.

**Dib, A., Zanet, J., Mancheno-Ferris, A., Gallois, M., Markus, D., Valenti, P., Marques-Prieto, S., Plaza, S., Kageyama, Y., Chanut-Delalande, H., and Payre, F.** (2021). Pri smORF peptides are wide mediators of ecdysone signaling, contributing to shape spatiotemporal responses. *Front. Genet.* 12: 714152

**Dillon, S.C., and Dorman, C.J.** (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* 8(3): 185–95.

**Dingwall, C., and Laskey, R.** (1992). The nuclear membrane. *Science.* 258(5084) : 942–947.

**Dubarry, M., Liodice, I., Chen, C.L., Thermes, C., and Taddei, A.** (2011). Tight protein-DNA interactions favor gene silencing. *Genes Dev.* 25: 1365–1370.

**Dubarry, N., Pasta, F., and Lane, D.** (2006). ParABS systems of the four replicons of *Burkholderia cenocepacia*: new chromosome centromeres confer partition specificity. *J. Bacteriol.* 188(4) : 1489-1496.



- Dufourt, J., Bellec, M., Trullo, A., Dejean, M., De Rossi, S., Favard, C., and Lagha, M. (2021).** Imaging translation dynamics in live embryos reveals spatial heterogeneities. *Science*. 372(6544):840-844.
- Dupaigne, P., Tonthat, N.K., Espéli, O., Whitfill, T., Boccard, F., and Schumacher, M.A. (2012).** Molecular basis for a protein-mediated DNA-bridging mechanism that functions in condensation of the *E. coli* chromosome. *Mol. Cell* 48(4): 560–71.
- Egan, E.S., and Waldor, M.K. (2003).** Distinct replication requirements for the two *Vibrio cholerae* chromosomes. *Cell*. 114(4): 521–530.
- Erb, M.L., Kraemer, J.A., Coker, J.K., Chaikerasitak, V., Nonejuie, P., Agard, D.A., and Pogliano, J. (2014).** A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells. *Elife*: 3: e03197.
- Erdmann, N., Petroff, T., and Funnell, B.E. (1999)** Intracellular localization of P1 ParB protein depends on ParA and parS. *Proc. Natl Acad. Sci.* 96(26): 14 905–14 910.
- Espeli, O., Levine, C., Hassing, H., and Mariani, K.J. (2003).** Temporal regulation of topoisomerase IV activity in *E. coli*. *Mol. Cell*. 11(1):189–201.
- Espeli, O., Mercier, R., and Boccard, F. (2008).** DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Mol. Microbiol.* 68(6): 1418–27.
- Ferdows, M.S., Serwer, P., Griess, G.A., Norris, S.J, and Barbour, A.G. (1996).** Conversion of a linear to a circular plasmid in the relapsing fever agent *Borrelia hermsii*. *J. Bacteriol.* 178(3): 793-800.
- Fernandes, I., Chanut-Delalande, H., Ferrer, P., Latapie, Y., Waltzer, L., Affolter, M., Payre, F., and Plaza, S. (2010).** Zona pellucida domain proteins remodel the apical compartment for localized cell shape changes. *Dev Cell*. 18(1): 64-76.
- Ferraro, T., Esposito, E., Mancini, L., Ng, S., Lucas, T., Coppey, M., Dostatni, N., Walczak, A.M., Levine, M., and Lagha, M. (2016)** Transcriptional Memory in the *Drosophila* Embryo. *Curr Biol*. 26(2):212-218.

- Feulgen, R., and Rossenbeck, H.** (1924). Mikroskopisch-chemischer Nachweis einer Nukleinsäure von Typus der Thymonukleinsäure und die darauf beruhende selektive Färbung von Zellkernen in mikroskopischen Präparaten. *Hoppe-Seylers Z Physiol Chem.* 135:203–248.
- Fink, G., and Aylett, C.H.S.** (2017). Tubulin-Like Proteins in Prokaryotic DNA Positioning. *Subcell Biochem.* 84:323-356.
- Filion, G.J., van Bommel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., and van Steensel, B.** (2010). Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell.* 143(2) : 212–224.
- Fisher, G.L., Pastrana, C.L., Higman, V.A., Koh, A., Taylor, J.A., Butterer, A., Craggs, T., Sobott, F., Murray, H., Crump, M.P., Moreno-Herrero, F., and Dillingham, M.S.** (2017). The structural basis for dynamic DNA binding and bridging interactions which condense the bacterial centromere. *Elife.* 6: e28086.
- Flemming, W.** (1882). *Zellsubstanz, Kern und Zelltheilung.* Leipzig, Maison d'édition F.C.W Vogel.
- Fogel, M.A., and Waldor, M.K.** (2006). A dynamic, mitotic-like mechanism for bacterial chromosome segregation. *Genes Dev.* 20(23): 3269–3282.
- Fok, E.T., Fanucchi, S., Bystricky, K., and Mhlanga, M.M.** 2021). Visualization of chromatin dynamics by live cell microscopy using CRISPR/Cas9 gene editing and ANCHOR labeling. *Methods Mol. Biol.* 2157: 197-212.
- Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L.** (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* 466(7305): 490-493.
- Frankel, N., Wang, S., and Stern, D.L.** (2012). Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proc Natl Acad Sci.* 109(51): 20975-20979.
- Franklin, R.E., and Gosling, R.G.** (1953). Molecular configuration in sodium thymonucleate. *171(4356):* 740-741.

- Fraser, R.D.B.** (1953). The structure of deoxyribose nucleic acid. <http://scarc.library.oregonstate.edu/coll/pauling/dna/notes/fraser-structure-01.html>
- Funnell, B.E.** (2016). ParB Partition Proteins: Complex Formation and Spreading at Bacterial and Plasmid Centromeres. *Front Mol Biosci.* 3:44.
- Furberg, S.** (1952). On the structure of nucleic acids. *Acta Chem. Scand.* 6: 634-640
- Furlong, E.E.M., and Levine, M.** (2018). Developmental enhancers and chromosome topology. *Science.* 361(6409) : 1341–1345.
- Gaillard, J.L.J., Van Duijn, P., and Schaberg, A.** (1968). Photometric determination of DNA in human chromosomes *Exp Cell Res* 53: 417-431.
- Galindo, M.I., Pueyo, J.I., Fouix, S., Bishop, S.A., and Couso, J.P.** (2007). Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS biology* 5(5): e106.
- Gall, J.G., and Pardue, M.L.** (1969). Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci.* 63(2):378-383.
- Gall, J.G.** (2016). The origin of in situ hybridization - A personal history. *Methods.* 98:4-9.
- Gallant, P., Shio, Y., Cheng, P.F., Parkhurst, S.M., and Eisenman, R.N.** (1996). Myc and Max homologs in *Drosophila*. *Science.* 274(5292) : 1523–1527.
- Gallie, D.R. and Kado, C.I.** (1987). *Agrobacterium tumefaciens* pTAR *parA* promoter region involved in autoregulation, incompatibility and plasmid partitioning. *J. Mol. Biol.* 193(3): 465–478.
- Garcia-Russell, N., Harmon, T.G., Le, T.Q., Amaladas, N.H., Mathewson, R.D., and Segall, A.M.** (2014). Unequal access of chromosomal regions to each other in *Salmonella*: Probing chromosome structure with phage 1 integrase-mediated long-range rearrangements. *Mol Microbiol.* 52: 329–344.Ce
- Garner, E.C., Campbell, C.S., and Mullins, R.D.** (2004) Dynamic instability in a DNA-segregating prokaryotic actin homolog. *Science* 306: 1021–1025.

**Garner, E.C., Campbell, C.S., Weibel, D.B. and Mullins, R.D.** (2007). Reconstitution of DNA segregation driven by assembly of a prokaryotic actin homolog. *Science* 315(5816): 1270-1274.

**Gaszner, M., and Felsenfeld, G.** (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet.* 7(9) : 703–713.

**Gellert, M., O’Dea, M.H, Itoh, T., and Tomizawa, J.** (1976). Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase. *Proc. Natl. Acad. Sci. USA.* 73: 4474–4478.

**Genome News Network.** (a)

[http://www.genomenetwork.org/resources/timeline/1888\\_Boveri.php](http://www.genomenetwork.org/resources/timeline/1888_Boveri.php)

**Genome News Network.** (b)

[http://www.genomenetwork.org/resources/timeline/1902\\_Boveri\\_Sutton.php](http://www.genomenetwork.org/resources/timeline/1902_Boveri_Sutton.php)

**Genuth, N.R., and Barna, M.** (2018). The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *71(3):* 364-374.

**Georget, V., Nicolas, J.C., and Sultan, C.** (1999). *La green fluorescent protein: application à la dynamique intracellulaire des récepteurs stéroïdiens.* *Médecine/sciences.* 15 : 45-55.

**Gerdes, K., Larsen, J.E., and Molin, S.** (1985). Stable inheritance of plasmid R1 requires two different loci. *J Bacteriol* 161: 292.

**Gerdes, K., Howard, M., and Szardenings, F.** (2010) Pushing and pulling in prokaryotic DNA segregation. *Cell.* 141(6): 927-42.

**Germier, T., Kocanova, S., Walther, N., Bancaud, A., Shaban, H.A., Sellou, H., Politi, A.Z., Ellenberg, J., Gallardo, F., and Bystricky, K.** (2017). Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophys. J.* 113(7):1383-1394.

**Germier, T., Audibert, S., Kocanova, S., Lane, D., and Bystricky, K.** (2018). Real-time imaging of specific genomic loci in eukaryotic cells using the ANCHOR DNA labelling system. *Methods.* 142: 16-23.

**Geyer, P.K., and Corces, V.G.** (1992). DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. *Genes Dev.* 6(10) : 1865–1873.

**Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E. (2014).** Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*. 512(7512):96-100.

**Giegé, R. (2006).** The early history of tRNA recognition by aminoacyl-tRNA synthetases. *J Biosci*. 31(4): 477-88.

**Gillespie, D., and Spiegelman, S.J. (1965).** A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane. *Mol Biol*. 12(3):829-42.

**Godson, G.N., Barrell, B.G., Staden, R., and Fiddes, J.C. (1978).** Nucleotide sequence of bacteriophage G4 DNA. *Nature*. 276(5685): 236-247.

**Gogou, C., Japaridze, A., and Dekker, C. (2021).** Mechanisms for chromosome segregation in bacteria. *Front. Microb*. 12: 685687.

**Gomez-Lamarca, M.J., Falo-Sanjuan, J., Stojnic, R., Abdul Rehman, S., Muresan, L., Jones, M.L., Pillidge, Z., Cerda-Moya, G., Yuan, Z., Baloul, S., Valenti, P., Bystricky, K., Payre, F., O'Holleran, K., Kovall, R., and Bray, S.J. (2018).** Activation of the Notch signaling pathway *in vivo* elicits changes in CSL nuclear dynamics. *Develop. Cell*. 44(5): 611-623.

**Graham, T.G.W., Wang, X., Song, D., Etson, C.M., van Oijen, A.M., Rudner, D.Z., and Loparo, J.J. (2014).** ParB spreading requires DNA bridging. *GenesDev*. 28: 1228–1238.

**Grainge, I., Bregu, M., Vazquez, M., Sivanathan, V., Ip, S.C., and Sherratt, D.J. (2007).** Unlinking chromosome catenanes *in vivo* by site-specific recombination. *EMBO J*. 26(19):4228–4238.

**Grainger, D.C., Hurd, D., Goldberg, M.D., and Busby, S.J.W. (2006).** Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res*. 34(16): 4642–52.

**Griffith, F. (1928).** The significance of Pneumococcal types. *Journal of Hygiene*. 27(2): 113-159.

**Gros, F., Hiatt, H., Gilbert, W., Kurland, C.G., Risebrough, R.W., and Watson, J.D. (1961).** Unstable ribonucleic acid revealed by pulse labelling of *Escherichia Coli*. *Nature*. 190(4776): 581-585.

- Gruber, S., and Errington, J.** (2009). Recruitment of condensin to replication origin regions by ParB/SpoOJ promotes chromosome segregation in *B. subtilis*. *Cell*. 137(4):685–696.
- Gruber, S., Veening, J.-W., Bach, J., Blettinger, M., Bramkamp, M., and Errington, J.** (2014). Interlinked sister chromosomes arise in the absence of condensin during fast replication in *B. subtilis*. *Curr. Biol*. 24(3):293–298.
- Grunberg-Manago, M., Ortiz, P.J., and Ochoa, S.** (1955). Enzymatic synthesis of nucleic acidlike polynucleotides. *Science*. 122(3176): 907-910.
- Guillebault, D., and Cotterill, S.** (2007). The *Drosophila* Df31 protein interacts with histone H3 tails and promotes chromatin bridging in vitro. *J Mol Biol*. 373(4) : 903–912.
- Guo, F., and Adhya, S.** (2007). Spiral structure of *Escherichia coli* H<sub>u</sub>α provides foundation for DNA supercoiling. *Proc. Natl Acad. Sci*. 104(11): 4309–4314.
- Gyurkovics, H., Gausz, J., Kummer, J., and Karch, F.** (1990). A new homeotic mutation in the *Drosophila* bithorax complex removes a boundary separating two domains of regulation. *EMBO J*. 9(8) : 2579–2585.
- Hall, K.** (2011). William Astbury and the biological significance of nucleic acids, 1938-1951. *Stud. Hist. Philo. Biol. Biomed. Sci*. 42(2011): 119-128.
- Hayes, F., and Barillà, D.** (2006). The bacterial segrosome: a dynamic nucleoprotein machine for DNA trafficking and segregation. *Nat Rev Microbiol*. 2006 4(2): 133-43.
- Hayama R, Mariani KJ.** (2010). Physical and functional interaction between the condensin MukB and the decatenase topoisomerase IV in *Escherichia coli*. *Proc. Natl Acad. Sci*. 107(44): 18826–18831.
- Hergeth, S.P., and Schneider, R.** (2015). The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep*. 16(11): 1439–1453.
- Hershey, A.D, and Chase, M.** (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol*. 36(1): 39-56.
- Higgins, C.F., Dorman, C.J., Stirling, D.A., Waddell, L., Booth, I.R., May, G., and Brenner, E.** (1988). A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. *Cell* 52(4): 569–84.

**Ho, T.C.S., Chan, A.H.Y., and Ganesan, A.** (2020). Thirty Years of HDAC Inhibitors: 2020 Insight and Hindsight. *J Med Chem.* 63(21) : 12460–12484.

**Hoagland, M.B., Zamecnik, P.C., and Stephenson, M.L.** (1957). Intermediate reactions in protein biosynthesis. *Biochim. Biophys. Acta.* 24(1): 215-216.

**Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht L.I., and Zamecnik, P.C.** (1958). A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* 231(1): 241-57.

**Hoagland, M.B.** (1959). Nucleic acids and proteins. *Sci. Am.* 201: 55-61.

**Holley, R.W., Apgar, J., Merrill, S.H, and Zubkoff, P.L.** (1961). Nucleotide and oligonucleotide compositions of the alanine-, valine-, and tyrosine-acceptor “soluble” ribonucleic acids of yeast. *J. Am. Chem. Soc.* 83(23): 4861-4862.

**Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., and Zamir, A.** (1965a). Structure of a ribonucleic acid. *Science.* 147(3664): 1462-1465.

**Holley, R.W., Everett, G.A., Madison, J.T., and Zamir, A.** (1965b). Nucleotide sequences in the yeast alanine transfer ribonucleic acid. *J. Biol. Chem.* 240(5): 2122-2128

**Holley, R.W.** (1965c). Structure of an alanine transfer ribonucleic acid. *JAMA.* 194(8): 868-871.

**Holley, R.W.** (1966). The nucleotide sequence of a nucleic acid. *Sci. Am.* 214(2): 30-39.

**Hong, S.H., and McAdams, H.H.** (2011). Compaction and transport properties of newly replicated *Caulobacter crescentus* DNA. *Mol. Microbiol.* 82(6):1349–1358.

**Horn P.J., and Peterson, C.L. (2002).** Molecular biology. Chromatin higher order folding--wrapping up transcription. *Science.* 297(5588):1824-1827.

**Hoshino, S., and Hayashi, I.** (2012). Filament formation of the FtsZ/tubulin-like protein TubZ from the *Bacillus cereus* pXO1 plasmid. *J Biol Chem.* 287: 32103–32112.

**Hu, L., Vecchiarelli, A.G., Mizuuchi, K., Neuman, K.C., and Liu, J.** (2017). Brownian ratchet mechanisms of ParA-mediated partitioning. *Plasmid.* 92: 12-16.

**Huber, D., Voith von Voithenberg, L., and Kaigala, G.V.** (2018). Fluorescence in situ hybridization (FISH): History, limitations and what to expect from micro-scale FISH? *Micro Neno Eng.* 15-24.

- Huisman, O., Faelen, M., Girard, D., Jaffé, A., Toussaint, A., and Rouvière-Yaniv, J.** (1989). Multiple defects in *Escherichia coli* mutants lacking HU protein. *J. Bacteriol.* 171(7):3704–3712.
- Hwang, L.C., Vecchiarelli, A.G., Han, Y.W., Mizuuchi, M., Harada, Y., Funnell, B.E., and Mizuuchi, K.** (2013). ParA-mediated plasmid partition driven by protein pattern self-organization. *EMBO J.* 32(9): 1238–1249.
- Inagaki, S., Numata, K., Kondo, T., Tomita, M., Yasuda, K., Kanai, A., and Kageyama, Y.** (2005). Identification and expression analysis of putative mRNA-like non-coding RNA in *Drosophila*. *Genes Cells.* 10(12): 1163-1173.
- Ireton, K., Gunther, N.W., and Grossman, A.D.** (1994). *spo0J* is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*. *J. Bacteriol.* 176(17): 5320-5329.
- Jackson, D.A., Symons, R.H., and Berg, P.** (1972). Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing Lambda Phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci.* 69(10): 2904-2909.
- Jacob, F., and Monod, F.** (1961a). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3:318-3356.
- Jacob, F., and Monod, F.** (1961b). On the regulation of gene activity. *Cold Spring Harb Symp Quant Biol.* 26: 193-211.
- Jacob, F., and Brenner, S.** (1963). [On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon]. *Comptes Rendus Hebd Séances Académie Sci.* 256 : 298–300.
- Jacob, F., Brenner, S., and Cuzin, F.** (1963). On the regulation of DNA replication in bacteria. *Cold Spring Harb. Symp. Quant. Biol.* 28: 329-348.
- Jacome, A., and Fernandez-Capetillo, O.** (2011). Lac operator repeats generate a traceable fragile site in mammalian cells. *EMBO Rep* 12(10): 1032–1038.
- Jalal, A.S.B., Pastrana, C.L., Tran, N.T., Stevenson, C.E., Lawson, D.M., Moreno-Herrero, F., and Le, T.B.K.** (2019). Structural and biochemical analyses of *Caulobacter crescentus* ParB reveal the role of its N-terminal domain in chromosome segregation. *bioRxiv*, 816959.



**Jalal, A.S.B., and Le, T.B.K.** (2020a). Bacterial chromosome segregation by the ParABS system. *Open Biol.* 10: 200097.

**Jalal, A.S.B, Tran, N.T., and Le, T.B.K.** (2020b). ParB spreading on DNA requires cytidine triphosphate in vitro. *eLife* 9: e53515.

**Jensen, R.B., Dam, M., and Gerdes, K.** (1994) Partitioning of plasmid R1. The parA operon is autoregulated by ParR and its transcription is highly stimulated by a downstream activating element. *J Mol Biol* 236: 1299–1309.

**Jensen, R.B., Lurz, R., and Gerdes, K.** (1998) Mechanism of DNA segregation in prokaryotes: replicon pairing by parC of plasmid R1. *Proc Natl Acad Sci.* 95(15) : 8550–8555.

**Jensen, R.B., and Shapiro, L.** (1999). The *Caulobacter crescentus smc* gene is required for cell cycle progression and chromosome segregation. *Proc Natl Acad Sci.* 96(19):10661–10666.

**Jiang, S., and Mortazavi, A.** (2018). Integrating ChIP-seq with other functional genomics data. *Brief Funct Genomics.* 17(2) : 104–115.

RNA-DNA hybrids at the cytological level.

**John, H.A., Birnstiel, M.L., and Jones, K.W.** (1969). RNA-DNA hybrids at the cytological level. *Nature.* 223(5206):582-587.

**Joshi, M.C., Bourniquel, A., Fisher, J., Ho, B.T, Magnan, D., Kleckner, N., and Bates, D.** (2011). *Escherichia coli* sister chromosome separation includes an abrupt global transition with concomitant release of late-splitting intersister snaps. *Proc Natl Acad Sci.* 108(7):2765–2770.

**Joshi, M.C., Magnan, D., Montminy, T.P., Lies, M., Stepankiw, N., and Bates, D.** (2013). Regulation of sister chromosome cohesion by the replication fork tracking protein SeqA. *PLOS Genet.* 9(8):e1003673.

**Jun, S., and Mulder, B.** (2006). Entropy-driven spatial organization of highly confined polymers: lessons for the bacterial chromosome. *Proc Natl Acad Sci.* 103(33):12388–12393.

**Jun, S., and Wright, A.** (2010). Entropy as the driver of chromosome segregation. *Nat. Rev. Microbiol.* 8(8) :600–607.

- Junier, I., Boccard, F., and Espéli, O.** (2014). Polymer modeling of the *E. coli* genome reveals the involvement of locus positioning and macrodomain structuring for the control of chromosome conformation and segregation. *Nucleic Acids Res.* 42(3) :1461–1473.
- Kahn, A.** (2003). L'hélice de la vie. *M/S : médecine et sciences.* 19(4): 491-495.
- Kahramanoglou, C., Seshasayee, A.S.N, Prieto, A.I., Ibberson, D., Schmidt, S., Zimmerman, J., Benes, V., Fraser, G.M., and Luscombe. N.M.** (2011). Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res.* 39(6) :2073–91.
- Kaidow, A., Wachi, M., Nakamura, J., Magae, J., and Nagai, K.** (1995). A nucleate cell production by *Escherichia coli*  $\Delta$ hns mutant lacking a histone-like protein, H-NS. *J. Bacteriol.* 177(12):3589–3592.
- Kalnín, K., Stegalkina, S., and Yarmolinsky, M.** (2000). pTAR-encoded proteins in plasmid partitioning. *J. Bacteriol.* 182(7): 1889–1894.
- Kar, S., Edgar, R., and Adhya, S.** (2005). Nucleoid remodeling by an altered HU protein: reorganization of the transcription program. *Proc. Natl. Acad. Sci.* 102(45): 16397–402.
- Kassis, J.A., Kennison, J.A., and Tamkun, J.W.** (2017). Polycomb and Trithorax Group Genes in *Drosophila*. *Genetics.* 206(4) : 1699–1725.
- Kavenoff, R., and Ryder, O.A.** (1976). Electron microscopy of membrane-associated folded chromosomes of *Escherichia coli*. *Chromosoma.* 55(1): 13–25.
- Kislauskis, E.H., Li, Z., Singer, R.H., and Taneja, K.L.** (1993). Isoform-specific 3'-untranslated sequences sort alpha-cardiac and beta-cytoplasmic actin messenger RNAs to different cytoplasmic compartments. *J Cell Biol.* Oct;123(1):165-72.
- King, H.D.** (1908). The oogenesis of *Bufo Lentiginosus*. *J. Morphology.* 19(2): 369-438.
- Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S., and Kageyama, Y.** (2007). Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat Cell Biol.* 9(6): 660-665.

**Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., Kobayashi, S., Payre, F., and Kageyama, Y.** (2010). Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Sciences*. 329(5989): 336-339.

**Kornberg, R.D., and Lorch, Y.** (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*. 98(3) : 285–294.

**Kraemer, J.A., Erb, M.L., Waddling, C.A., Montabana, E.A., Zehr, E.A., Wang, H., Nguyen, K., Pham, D.S., Agard, D.A., and Pogliano, J.** (2012). A phage tubulin assembles dynamic filaments by an atypical mechanism to center viral DNA within the host cell. *Cell* 149(7): 1488–1499.

**Lagha, M., Bothma, J.P., and Levine, M.** (2012). Mechanisms of transcriptional precision in animal development. *Trends in genetics*. 28(8): 409-416.

**Lamborg, M.R., and Zamecnik, P.C.** (1960). Amino acid incorporation into protein by extracts of *E. coli*. *Bioch. Biophys. Acta*. 42: 206-211

**Langer-Safer, P.R, Levine, M., and Ward, D.C.** (1982). Immunological method for mapping genes on *Drosophila* polytene chromosomes. *Proc Natl Acad Sci*. 79(14):4381-4385.

**Larsen, R.A., Cusumano, C., Fujioka, A., Lim-Fong, G., Patterson, P., and Pogliano, J.** (2007). Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in *Bacillus thuringiensis*. *Genes Dev*. 21:1340–1352

**Lassadi, I., Kamgoué, A., Goiffon, I., Tanguy-le-Gac, N., and Bystricky, K.** (2015) Differential chromosome conformations as hallmarks of cellular identity revealed by mathematical polymer modeling. *PLoS Comput. Biol*. 11(6): e1004306.

**Lau, Y.F., Filipe, S.R., Søballe, B., Økstad, O.A., Barre, F.X., and Sherratt, D.J.** (2003). Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Mol Microbiol*. 49(3):731-43.

**Le, T.B., Imakaev, M.V., Mirny, L.A., and Laub, M.T.** (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. *Science*. 342(6159): 731–34.

**Le, T.B., and Laub, M.T.** (2014). New approaches to understanding the spatial organization of bacterial genomes. *Curr. Opin. Microbiol*. 22 :15–21.

- Lederberg, J.** (1952) Cell genetics and hereditary symbiosis. *Physiol Rev* 32: 403–430.
- Lee, S., and Garfinkel, M.D.** (2000). Characterization of *Drosophila* OVO protein DNA binding specificity using random DNA oligomer selection suggests zinc finger degeneration. *Nucleic Acids Res.* 28(3): 826-834.
- Lefevre, G.,** (1976). A photographic representation and interpretation of the polytene chromosomes of *Drosophila melanogaster* salivary glands. In *The genetics and biology of Drosophila* (Eds Ashburner M., Novitski E.), a, 31-66. Academic Press, London, New-York and San Francisco.
- Le Gall, A., Cattoni, D.I., Guilhas, B., Mathieu-Demaziere, C., Oudjedi, L., Fiche, J.-B., Rech, J., Abrahamsson, S., Murray, H., Bouet, J.-Y., and Nollmann, M.** (2016). Bacterial partition complexes segregate within the volume of the nucleoid. *Nat. Commun.* 7: 12107.
- Leipe, D.D., Wolf, Y.I., Koonin, E.V., and Aravind, L.** (2002) Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* 317(1): 41–72.
- Leonard, A.C., Grimwade, J.E.** (2005). Building a bacterial orisome: emergence of new regulatory features for replication origin unwinding. *Mol. Microbiol.* 55(4) : 978–85.
- Levene, P.A.** (1919). The structure of yeast nucleic acid. IV. Ammonia Hydrolysis. *J. Biol. Chem.* 40: 415-424
- Li, Y., and Austin, S.** (2002a). The P1 plasmid is segregated to daughter cells by a 'capture and ejection' mechanism coordinated with *Escherichia coli* cell division. *Mol Microbiol.* 46(1): 63-74.
- Li, Y., Sergueev, K., and Austin, S.** (2002b). The segregation of the *Escherichia coli* origin and terminus of replication. *Mol Microbiol.* 46(4): 985-96.
- Li, Y., Stewart, N.K., Berger, A.J., Vos, S., Schoeffler, A.J., Berger, J.M., Chait, B.Y., and Oakley, M.G.** (2010). *Escherichia coli* condensin MukB stimulates topoisomerase IV activity by a direct physical interaction. *Proc Natl Acad Sci.* 107(44):18832–18837.
- Lim, H., Surovtsev, I., Beltran, B., Huang, F., Bewersdorf, J., and Jacobs-Wagner, C.** (2014). Evidence for a DNA-relay mechanism in ParABS-mediated chromosome segregation. *eLIFE* 3, e02758.

- Lin, D.C., and Grossman, A.D.** (1998). Identification and characterization of a bacterial chromosome partitioning site. *Cell*. 92(5): 675-685.
- Lin, Y.S., and Chen, C.W.** (1997). Instability of artificially circularized chromosomes of *Streptomyces lividans*. *Mol. Microbiol.* 26(4): 709-719.
- Littlefield, J.W., Keller, E.B., Gross, J., and Zamecnik, P.C.** (1955). Studies on cytoplasmic ribonucleoprotein particles from the liver of the rat. *J. Biol. Chem.* 217(1): 111-23.
- Livny, J., Yamaichi, Y., Waldor, M.K.** (2007). Distribution of centromere-like parS sites in bacteria: insights from comparative genomics. *J. Bacteriol.* 189(23): 8693-8703.
- Lorton, B.M., and Shechter, D.** (2019). Cellular Consequences of Arginine Methylation. *Cell Mol Life Sci.* 76(15) : 2933–2956.
- Löwe, J., Ellonen, A., Allen, M.D., Atkinson, C., Sherratt, D.J., and Grainge, I.** (2008). Molecular mechanism of sequence-directed DNA loading and translocation by FtsK. *Mol. Cell* 31(4): 498–509.
- Luger, K., Mäder, A.W., Richmond, R.K, Sargent, D.F., and Richmond, T.J.** (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 389(6648): 251-260.
- Luger, K., and Hansen, J.C.** (2005). Nucleosome and chromatin fiber dynamics. *Curr Opin Struct Biol.* 15(2) : 188–196.
- Lutkenhaus, J.** (2012). The ParA/MinD family puts things in their place. *Trends Microbiol* 20(9): 411–418.
- Lynch, A.S., and Wang, J.C.** (1995). SopB protein-mediated silencing of genes linked to the *sopC* locus of *Escherichia coli* F plasmid. *Proc Natl Acad Sci.* 92(6): 1896–1900.
- Ma, H., Reyes-Gutierrez, P., and Pederson, T.** (2013). Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proc. Natl. Acad. Sci.* 110 (52): 21048-21053.
- Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S.A., Zhang, S., and Pederson, T.** (2015). Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci.*, 112(10): 3002-3007.

**Mancheno-Ferris, A.** (2021). Caractérisation du facteur de transcription Shavenbaby par approches génomiques chez la drosophile. Université Paul Sabatier - Toulouse III, 2017.

**Manning, J.E., Hershey, N.D., Broker, T.R., Pellegrini, M., Mitchell, H.K., and Davidson, N.** (1975). A new method of in situ hybridization. *Chromosoma*. 24;53(2):107-17.

**Marbouty, M., Le Gall, A., Cattoni, D.I., Cournac, A., Koh, A., Fiche, J.B., Mozziconacci, J., Murray, H., Koszul, R., and Nollmann, M.** (2015). Condensin- and replication-mediated bacterial chromosome folding and origin condensation revealed by Hi-C and super-resolution imaging. *Mol Cell*. 59(4): 588-602.

**Mariamé, B., Kappler-Gratias, S., Kappler, M., Balor, S., Gallardo, F., and Bystricky, K.** (2018). Real-Time visualization and quantification of human cytomegalovirus replication in living cells using the ANCHOR DNA labeling technology. *J. Virol*. 92(18): e00571-18.

**Marmorstein, R., and Zhou, M.-M.** (2014). Writers and Readers of Histone Acetylation: Structure, Mechanism, and Inhibition. *Cold Spring Harb Perspect Biol*. 6(7) : a018762.

**Marshall, W., Straight, A., Marko, J., Swedlow, J., Dernburg, A., Belmont, A., Murray, A.W., Agard, D.A., and Sedat, J.W.** (1997) Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr. Biol*. 7(12): 930–939.

**Matthaei, J.H., and Nirenberg, M.W.** (1961). Characteristics and stabilization of DNAase-sensitive protein synthesis in *E. coli* extracts. *Proc. Natl. Acad. Sci*. 47(10): 1580-1588.

**Maxam, A.M., and Gilbert, W.** (1977). A new method for sequencing DNA. *Proc. Natl. Acad. Sci*. 74(2): 560-564.

**McCarty, M.** (1946a). Purification and properties of desoxyribonuclease isolated from beef pancreas. *J. Gen. Physiol*. 29(3): 123-139.

**McCarty, M., and Avery, O.T.** (1946b). Studies on the chemical nature of the substance inducing transformation of Pneumococcal types: II. Effect of desoxyribonuclease on the biological activity of the transforming substance. *J. Exp. Med*. 83(2): 89-96.

**McGregor, A.P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D.G., Payre, F., and Stern, D.L.** (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature*. 448(7153) : 587-590.

**Menoret, D., Santolini, M., Fernandes, I., Spokony, R., Zanet, J., Gonzalez, I., Latapie, Y., Ferrer, P., Rouault, H., White, K.P., Besse, P., Hakim, V., Aerts, S., Payre, F., and Plaza, S.** (2013). Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization. *Genome Biol.* 14(8): R86.

**Mercier, R., Petit, M.-A., Schbath, S., Robin, S., El Karoui, M., Boccard, F., and Espéli, O.** (2008). The MatP/matS site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. *Cell.* 135(3): 475–85.

**Meschichi, A., Ingouff, M., Picart, C., Mirouze, M., Desset, S., Gallardo, F., Bystricky, K., Picault, N., Rosa, S., and Pontvianne, F.** (2021a). ANCHOR: A Technical Approach to Monitor Single-Copy Locus Localization in Planta. *Front Plant Sci.* 12: 677849.

**Meschichi, A., and Rosa, S.** (2021). Visualizing and measuring single locus dynamics in *Arabidopsis thaliana*. *Methods Mol. Biol.* 2200: 213–224.

**Meselson, M. and Stahl, F.** (1958). The replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci.* 44(7): 671-682.

**Mesika, A., Kiss, V., Brumfeld, V., Ghosh, G. and Reich, Z.** (2005) Enhanced intracellular mobility and nuclear accumulation of DNA plasmids. associated with a karyophilic protein. *Hum. Gene Ther.* 16: 200–208.

**Mével-Ninio, M., Terracol, R., and Kafatos, F.C.** (1991). The ovo gene of *Drosophila* encodes a zinc finger protein required for female germ line development. *EMBO J.* 10(8): 2259-2266.

**Mével-Ninio, M., Terracol, R., Salles, C., Vincent, A., and Payre, F.** (1995). ovo, a *Drosophila* gene required for ovarian development, is specifically expressed in the germline and shares most of its coding sequences with shavenbaby, a gene involved in embryo patterning. *Mech Dev.* 49(1-2): 83-95.

**Møller-Jensen, J., Borch, J., Dam, M., Jensen, R.B., Roepstorff, P., and Gerdes, K.** (2003) Bacterial Mitosis: ParM of Plasmid R1 Moves Plasmid DNA by an Actin-like Insertional Polymerization Mechanism. *Mol Cell* 12: 1477–1487.

**Moran, N.A., and Bennett, G.M.** (2014). The tiniest tiny genomes. *Annu. Rev. Microbiol.* 68: 195-215.

- Morgan, T.H.** (1910). Sex limited inheritance. *Science*. 22; 32(812): 120-122.
- Morgan, T.H., Sturtevant, A.H., Muller, H.J., and Bridges, C.B.** (1915). The mechanism of Mendelian Heredity. Henry Holt Company. New York.
- Mori, H., Kondo, A., Ohshima, A., Ogura, T., Hiraga, S.** (1986). Structure and function of the F plasmid genes essential for partitioning. *J. Mol. Biol.* 192(1): 1-15.
- Mori, H., Mori, Y., Ichinose, C., Niki, H., Ogura, T., Kato, A., and Hiraga, S.** (1989). Purification and characterization of SopA and SopB proteins essential for F plasmid partitioning. *J Biol Chem.* 264(26): 15535–15541.
- Morise, H., Shimomura, O., Johnson, F.H., and Winant, J.** (1974). Intermolecular energy transfer in the bioluminescent system of *Aequorea*. *Biochemistry*. 13(12):2656-2662.
- Moscou, M.J., and Bogdanove, A.J.** (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*. 326(5959):1501.
- Motallebi-Veshareh, M., Rouch, D.A., and Thomas, C.M.** (1990). A family of ATPases involved in active partitioning of diverse bacterial plasmids. *Mol Microbiol* 4(9): 1455–1463.
- Mumm, J.P., Landy, A., and Gelles, J.** (2006). Viewing single lambda site-specific recombination events from start to finish. *EMBO J.* 25(19): 4586–4595.
- Murray, H., Ferreira, H., and Errington, J.** (2006). The bacterial chromosome segregation protein Spo0J spreads along DNA from *parS* nucleation sites. *Mol. Microbiol.* 61(5): 1352–1361.
- Miyazari, Y., Ziegler-Birling, C., and Torres-Padilla, M.-E.** (2013) Live visualization of chromatin dynamics with fluorescent TALEs. *Nat. Struct. Mol. Biol.* 20(11): 1321–1324.
- Miyazari, Y.,** (2014). TAL effector-mediated genome visualization (TGV). *Methods*. 69(2): 198-204.
- Nguyen, T.S., Gladyshev, E.** (2020). Developing a *tetO*/TetR system in *Neurospora crassa*. *Fungal Genet Biol.* 136:103316.



- Ni, L., Xu, W., Kumaraswami, M., and Schumacher, M.A.** (2010). Plasmid protein TubR uses a distinct mode of HTH-DNA binding and recruits the prokaryotic tubulin homolog TubZ to effect DNA partition. *Proc Natl Acad Sci.* 107(26): 11763–117.
- Nicolas, E., Upton, A.L., Uphoff, S., Henry, O., Badrinarayanan, A., Sherratt, D.** (2014). The SMC complex MukBEF recruits topoisomerase IV to the origin of replication region in live *Escherichia coli*. *Mbio* 5(1): e01001–13.
- Niki, H., Jaffé, A., Imamura, R., Ogura, T., and Hiraga, S.** (1991). The new gene *mukB* codes for a 177 kd protein with coiled-coil domains involved in chromosome partitioning of *E. coli*. *EMBO J.* 10(1): 183–193.
- Niki, H., Yamaichi, Y., and Hiraga, S.** (2000). Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev* 14: 212–223.
- Nielsen, H.J., Li, Y., Youngren, B., Hansen, F.G., and Austin, S.** (2006a). Progressive segregation of the *Escherichia coli* chromosome. *Mol. Microbiol.* 61(2): 383–393.
- Nielsen, H.J., Ottesen, J.R., Youngren, B., Austin, S.J., and Hansen, F.G.** (2006b). The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. *Mol. Microbiol.* 62(2): 331–338.
- Nirenberg, M.W., and Matthaei, J.H.** (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci.* 47(10): 1588-1602.
- Nirenberg, M.W.** (1963). The genetic code: II. *Scie. Am.* 208: 80-94.
- Nirenberg, M.W., and Leder, P.** (1964). RNA codewords and protein synthesis. *Science.* 145(3639): 1399-1407.
- Nolivos, S., and Sherratt, D.** (2014). The bacterial chromosome: architecture and action of bacterial SMC and SMC-like complexes. *FEMS Microbiol. Rev.* 38(3): 380–92.
- Ogura, T., and Hiraga, S.** (1983) Partition mechanism of F plasmid: two plasmid gene-encoded products and a cis-acting region are involved in partition. *Cell.* 332(2):351-60.
- Olins, D.E., and Olins, A.L.** (2003). Chromatin history: our view from the bridge. *Nat. Rev. Mol. Cell. Biol.* Oct; 4(10): 809-814.

- Oliva, M.A., Martin-Galiano, A.J., Sakaguchi, Y., and Andreu, J.M.** (2012). Tubulin homolog TubZ in a phage-encoded partition system. *Proc Natl Acad Sci.* 109(20): 7711–7716.
- Osorio-Valeriano, M., Altegoer, F., Steinchen, W., Urban, S., Liu, Y., Bange, G., and Thanbichler, M.** (2019) ParB-type DNA segregation proteins are CTP-dependent molecular switches. *Cell* 179(7): 1512–1524.
- Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., O’Shea, C.C.** (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science.* [unevirisu357\(6349\): eaag0025](https://doi.org/10.1126/science.1257583).
- Özdemir, I., and Gambetta, M.C.** (2019). The Role of Insulation in Patterning Gene Expression. *Genes.* 10(10) : 767.
- Pardue, M.L., and Gall, J.G.** (1970). Chromosomal localization of mouse satellite DNA. *Science.* 168(3937):1356-1358.
- Passot, F.M., Calderon, V., Fichant, G., Lane, D., and Pasta, F.** (2012). Centromere binding and evolution of chromosomal partition systems in the *Burkholderiales*. *J. Bacteriol.* 194(13): 426-436.
- Pauling, L., and Corey, R.B.** (1951). Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proc. Natl. Acad. Sci.* 37(5): 235-239.
- Pauling, L., and Corey, R.B.** (1953a). A proposed structure for the nucleic acids. *Proc. Natl. Acad. Sci.* 39(2): 84-97.
- Pauling, L., and Corey, R.B.** (1953b). Structure of the nucleic acids. *Nature.* 171(4347): 346.
- Payne, B.T, van Knippenberg, I.C., Bell, H., Filipe, S.R., Sherratt, D.J., and McGlynn, P.** (2006). Replication fork blockage by transcription factor-DNA complexes in *Escherichia coli*. *Nucleic Acids Res.* 34(18): 5194-5202.
- Payre, F., Vincent, A., and Carreno, S.** (1999). *ovo/svb* integrates Wingless and DER pathways to control epidermis differentiation. *Nature.* 400(6741): 271-275.
- Peifer, M., and Bender, W.** (1986). The anterobithorax and bithorax mutations of the bithorax complex. *EMBO J.* 5(9) : 2293–2303.

- Peter, B.J., Ullsperger, C., Hiasa, H., Mariani, K.J., and Cozzarelli, N.R.** (1998). The structure of supercoiled intermediates in DNA replication. *Cell*. 94(6):819–827.
- Petrushenko, Z.M., She, W., and Rybenkov, V.V.** (2011). A new family of bacterial condensins. *Mol. Microbiol.* 81, 881–896.
- Pillet, F., Sanchez, A., Lane, D., Anton Leberre, V., and Bouet, J.Y.** (2011). Centromere binding specificity in assembly of the F plasmid partition complex. *Nucleic Acids Res.* 39(17): 7477–7486.
- Piunti, A., and Shilatifard, A.** (2021). The roles of Polycomb repressive complexes in mammalian development and cancer. *Nat Rev Mol Cell Biol.* 22(5) : 326–345.
- Popp D., Yamamoto, A., Iwasa, M., Narita, A., Maeda, K., and Maéda Y.** (2007). Concerning the dynamic instability of actin homolog ParM. *Biochem. Biophys. Res. Commun.* 353(1): 109–114.
- Possoz, C., Filipe, S.R., Grainge, I., and Sherratt, D.J.** (2006). Tracking of controlled *Escherichia coli* replication fork stalling and restart at repressor-bound DNA in vivo. *EMBO J.* 25(11): 2596–2604.
- Postow, L., Hardy, C.D., Arsuaga, J., and Cozzarelli, N.R.** (2004). Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev.* 18(14): 1766–1779
- Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G., and Cormier, M.J.** (1992). Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene.* 111(2):229–233.
- Prieto, A.I., Kahramanoglou, C., Ali, R.M., Fraser, G.M., Seshasayee, A.S.N., and Luscombe, N.M.** (2012). Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12. *Nucleic Acids Res.* 40(8):3524–3537.
- Ptacin, J.L., Lee, S.F., Garner, E.C., Toro, E., Eckart, M., Comolli, L.R., Moerner, W.E., and Shapiro, L.** (2010). A spindle-like apparatus guides bacterial chromosome segregation. *Nat. Cell Biol.* 12(8): 791–798.
- Quénet, D., McNally, J.G., and Dalal, Y.** (2012). Through thick and thin: the conundrum of chromatin fibre folding in vivo. *EMBO Rep.* 13(11) : 943–944.

**Rasmussen, T., Jensen, R. B. and Skovgaard, O.** (2007). The two chromosomes of *Vibrio cholerae* are initiated at different time points in the cell cycle. *EMBO J.* 26: 3124-3131.

**Ray, S., Rosenberg, M.I., Chanut-Delalande, H., Decaras, A., Schwertner, B., Toubiana, W., Auman, T., Schnellhammer, I., Teuscher, M., Valenti, P., Khila, A., Klingler, M., Payre, F.** (2019). The mlpt/Ubr3/Svb module comprises an ancient developmental switch for embryonic patterning. *Elife.* 8: e39748.

**Ren, R., Deng, L., Xue, Y., Suzuki, K., Zhang, W., Yu, Y., Wu, J., Sun, L., Gong, X., Luan, H., Yang, F., Ju, Z., Ren, X., Wang, S., Tang, H., Geng, L., Zhang, W., Li, J., Qiao, J., Xu, T., Qu, J., and Liu, G.H.** (2017). Visualization of aging-associated chromatin alterations with an engineered TALE system. *Cell Res.* 27(4): 483-504.

**Rice, P.A., Yang, S., Mizuuchi, K., and Nash, H.A.** (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell.* 87(7) :1295–1306.

**Roberston, C.W.,** (1936). The metamorphosis of *Drosophila melanogaster*, including an accurately timed account of the principal morphological changes. *J. Morphology.* 59(2): 351-399.

**Robinett, C.C., Straight, A., Li, G., Wilhelm, C., Sudlow, G., Murray, A., and Belmont, A.S.** (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell Biol.* 135(6 Pt 2):1685-1700.

**Robinow, C., and Kellenberger, E.** (1994). The bacterial nucleoid revisited. *Microbiol. Rev.* 58(2): 211-232.

**Rodionov, O., Lobočka, M., and Yarmolinsky, M.** (1999). Silencing of genes flanking the P1 plasmid centromere. *Science.* 283(5401): 546–549.

**Rodionov, O., and Yarmolinsky, M.** (2004). Plasmid partitioning and the spreading of P1 partition protein ParB. *Mol.Microbiol.* 52(4): 1215–1223.

**Rosa, S., and Shaw, P.** (2013). Insights into chromatin structure and dynamics in plants. *Biology.* 2(4) : 1378–1410.

**Roseman, R.R., Pirrotta, V., and Geyer, P.K.** (1993). The su(Hw) protein insulates expression of the *Drosophila melanogaster white* gene from chromosomal position-effects. *EMBO J.* 12(2): 435–442.

**Russell, W.C., Newman, C., and Williamson, D.H.** (1975). A simple cytochemical technique for demonstration of DNA in cells infected with mycoplasmas and viruses. *Nature.* 253(5491):461-462.

**Saad, H., Gallardo, F., Dalvai, M., Tanguy-le-Gac, N., Lane, D., and Bystricky, K.** (2014). DNA dynamics during early double-strand break processing revealed by non-intrusive imaging of living cells. *PLoS Genet.* 10(3): e1004187.

**Sakaguchi, Y., Hayashi, T., Kurokawa, K., Nakayama, K., Oshima, K., Fujinaga, Y., Ohnishi, M., Ohtsubo, E., Hattori, M., and Oguma, K.** (2005). The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. *Proc Natl Acad Sci.* 102(48): 17472–17477.

**Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., et al.** (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature.* 415(6871): 497–502.

**Salje, J., Gayathri, P., and Löwe, J.** (2010a). The ParMRC system: molecular mechanisms of plasmid segregation by actin-like filaments. *Nat Rev Micro.* 8: 683-692.

**Salje, J.** (2010b). Plasmid segregation: how to survive as an extra piece of DNA. *Crit. Rev. Biochem. Mol. Biol.* 45(4): 296–317.

**Salman, H., Abu-Arish, A., Oliel, S., Loyter, A., Klafter, J., Granek, R. and Elbaum, M.** (2005). Nuclear localization signal peptides induce molecular delivery along microtubules. *Biophys. J.* 89(3): 2134–2145.

**Sanchez, A., Rech, J., Gasc, C., and Bouet, J.Y.** (2013). Insight into centromere-binding properties of ParB proteins: a secondary binding motif is essential for bacterial genome maintenance. *Nucleic Acid. Res.* 39(17): 7477-7486.

- Sanchez, A., Cattoni, D.I., Walter, J.C., Rech, J., Parmeggiani, A., Nollmann, M., and Bouet, J.Y.** (2015). Stochastic Self-Assembly of ParB Proteins Builds the Bacterial DNA Segregation Apparatus. *Cell Syst.* 1(2): 163-173.
- Sanger, F., Nicklen, S., and Coulson, A.R.** (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* 74(12): 5463-5467.
- Sanger, F., Coulson, A.R., Friedmann, T., Air, G.M., Barrell, B.G., Brown, N.L., Fiddes, J.C., Hutchison 3rd, C.A., Slocombe, P.M., and Smith, M.** (1978). The nucleotide sequence of bacteriophage *phiX174*. *J. Mol. Biol.* 125(2): 225-246.
- Santolini, M., Mora, T., and Hakim, V.** (2014). A general pairwise interaction model provides an accurate description of in vivo transcription factor binding sites. *PLoS One.* 9(6) : e99015.
- Savard, J., Marques-Souza, H., Aranda, M., and Tautz, D.** (2006). A segmentation gene in *tribolium* produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell.* 126(3): 559-569.
- Scalenghe, F., Turco, E., Edström, J.E., Pirrotta, V., and Melli, M.,** (1981). Microdissection and cloning of DNA from specific region of *Drosophila melanogaster* polytene chromosomes. *Chromosoma.* 82(2): 205-216.
- Schneider, R., Travers, A., Kutateladze, T., and Muskhelishvili, G.** (1999). A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol. Microbiol.* 34(5): 953–964.
- Schotta, G., Ebert, A., and Reuter, G.** (2003). SU(VAR)3-9 is a conserved key function in heterochromatic gene silencing. *Genetica.* 117(2-3) : 149–158.
- Schumacher, M.A., Piro, K.M., and Xu, W.** (2010) Insight into F plasmid DNA segregation revealed by structures of SopB and SopB–DNA complexes. *Nucleic Acids Res* 38(13): 4514–4526.
- Sherratt, D.** (2013). Plasmid partition: sisters drifting apart. *EMBO J.* 32(9): 1208-1210.
- Signer, R., Caspersson, T., and Hammarsten, E.** (1938). Molecular shape and size of thymonucleic acid. *Nature.* 141(3559): 122

- Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y** (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*. 119(7): 941–953.
- Shlyueva, D., Stelzer, C., Gerlach, D., Yáñez-Cuna, J.O., Rath, M., Boryń, Ł.M., Arnold, C.D., and Stark, A.** (2014). Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. *Mol Cell*. 54(1):180-192.
- Shimomura, O., Jonson, F.H., and Saiga, Y.** (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J Cell Comp Physiol*. 59:223-239.
- Simpson, A.E., Skurray, R.A., and Firth N.** (2003). A single gene on the staphylococcal multiresistance plasmid pSK1 encodes a novel partitioning system. *J. Bacteriol*. 185(7): 2143-2152.
- Singer, R.H., and Ward, D.C.** (1982). Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog. *Proc Natl Acad Sci*. 79(23):7331-7335.
- Singh, S.S., Singh, N., Bonocora, R.P., Fitzgerald, D.M., Wade, J.T., and Grainger, D.C.** (2014). Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev*. 28(3): 214–219.
- Sivanathan, V., Allen, M.D., de Bekker, C., Baker, R., Arciszewska, L.K., Freund, S.M., Bycroft, M., Löwe, J., and Sherratt, D.J.** (2006). The FtsK gamma domain directs oriented DNA translocation by interacting with KOPS. *Nat. Struct. Mol. Biol*. 13(11):965–972.
- Slater, F.R., Bailey, M.J., Tett, A.J., and Turner, S.L.** (2008). Progress towards understanding the fate of plasmids in bacterial communities. *FEMS Microbiol Ecol*. 66: 3–13.
- Slater, S. C., Goldman, B. S., Goodner, B., Setubal, J. C., Farrand, S. K., Nester, E. W., Burr, T. J., Banta, L., Dickerman, A. W., et al.** (2009). Genome sequences of three *Agrobacterium Biovars* help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol*. 191: 2501-2511.

**Smith, H.O., and Wilcox, K.W.** (1970). A restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties. *J. Mol. Biol.* 51: 379-391.

**Sofueva, S., Osman, F., Lorenz, A., Steinacher, R., Castagnetti, S., Ledesma, J., and Whitby, M.C.** (2011). Ultrafine anaphase bridges, broken DNA and illegitimate recombination induced by a replication fork barrier. *Nucleic Acids Res* 39(15): 6568–6584.

**Soh, Y.M., Davidson, I.F., Zamuner, S., Basquin, J., Bock, F.P., Taschner, M., Veening, J.W., De Los Rios, P., Peters, J.M., and Gruber, S.** (2019), Self-organization of *parS* centromeres by the ParB CTP hydrolase. *Science*. 366(6469): 1129–1133.

**Söll, D, Ohtsuka, E., Jones, D.S., Lohrmann, R., Hayatsu, H., Nishimura, S., and Khorana, H.G.** (1965). Studies on polynucleotides, XLIX. Stimulation of the binding of aminoacyl-sRNA's to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids. *Proc Natl Acad Sci.* 54(5): 1379-1385.

**Springer, M.** (1987). La découverte des enzymes qui tranchent l'ADN: Ou de l'imprévisible utilité du phénomène de restriction-modification chez les bactéries. *Esprit.* 128(7): 17-25.

**Srivastava, P., and Chatteraj, D.K.** (2007). Selective chromosome amplification in *Vibrio cholerae*. *Mol. Microbiol.* 66: 1016–1028

**Steiner, W.W., and Kuempel, P.L.** (1998). Sister chromatid exchange frequencies in *Escherichia coli* analyzed by recombination at the *dif* resolvase site. *J. Bacteriol.* 180(23):6269–6275.

**Steiner, W., Liu, G., Donachie, W.D., and Kuempel, P.** (1999). The cytoplasmic domain of FtsK protein is required for resolution of chromosome dimers. *Mol. Microbiol.* 31(2):579–583.

**Stella, S., Cascio, D., and Johnson, R.C.** (2010). The shape of the DNA minor groove directs binding by the DNA bending protein Fis. *Genes Dev.* 24(8):814–826.

**Stern, D.L., and Frankel, N.** (2013). The structure and evolution of cis-regulatory regions: the shavenbaby story. *Philos Trans R Soc Lond B Biol Sci.* 368(1632) : 20130028.

**Straight, A.F., Belmont, A.S., Robinett, C.C., and Murray, A.W.** (1996). GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Curr. Biol.* 6(12): 1599-1608.



**Sturtevant, A.H.** (1913). The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.* 14: 43-59.

**Sullivan, N.L., Marquis, K.A., and Rudner, D.Z.** (2009). Recruitment of SMC by ParB-*parS* organizes the origin region and promotes efficient chromosome segregation. *Cell.* 137(4):697–707.

**Surtees, J.A., and Funnell, B.E.** (1999). P1 ParB domain structure includes two independent multimerization domains. *J Bacteriol.* 181(19):5898-908.

**Sutton, W.S.** (1902). On the morphology of the chromosome group in *Brachystola magna*. *Biological Bulletin.* 4: 24-39.

**Sutton, W.S.** (1903). The chromosomes in heredity. *Biological Bulletin.* 4: 24-39.

**Swiercz, J.P., Nanji, T., Gloyd, M., Guarné, A., and Elliot, M.A.** (2013). A novel nucleoid-associated protein specific to the actinobacteria. *Nucleic Acids Res.* 41(7):4171–4184.

**Swinger, K.K., and Rice, P.A. (2004).** IHF and HU: flexible architects of bent DNA. *Curr. Opin. Struct. Biol.* 14: 28–35.

**Szymanski, M., and Barciszewski, J.** (2017). The path to the genetic code. *Biochim Biophys Acta Gen Subj.* 1861(11 Pt A):2674-2679.

**Tadesse, S., Mascarenhas, J.K., Kösters, B., Hasilik, A., and Graumann, P.L. (2005).** Genetic interaction of the SMC complex with topoisomerase IV in *Bacillus subtilis*. *Microbiology* 151(11): 3729–3737.

**Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D.** (2014) A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell.* 159(3): 635–646.

**Tang, M., Bideshi, D.K., Park, H.W., and Federici, B.A.** (2007.) Iteron-binding ORF157 and FtsZ-like ORF156 proteins encoded by pBtoxis play a role in its replication in *Bacillus thuringiensis subsp. israelensis*. *J Bacteriol.* 189:8053–8058.

**Tasan, I., Su, C.J., Enghiad, B., Zhang, M., Mishra, S., and Zhao, H.** (2020). Two-Color Imaging of Nonrepetitive Endogenous Loci in Human Cells. *ACS Synth Biol.* 9(9):2502-2514.

- Teich, M.** (1975). Essay Review: A Single Path to the Double Helix? : The Path to the Double Helix. *History of Sciences*. 13(4): 264-283.
- Temin, H. M., and Mizutami, S.** (1970). RNA-dependent DNA polymerase in virions of Rous Sarcome virus. *Nature*. 226(5252):1211-1213.
- Thach, R.E., Dewey, K.F., Brown, J.C., and Doty, P.** (1966). Formylmethionine Codon AUG as an Initiator of Polypeptide Synthesis. *Science*. 153(3734): 416-418.
- Thoma, F., Koller, T., and Klug, A.** (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol*. 83(2 Pt 1) : 403–427.
- Tissieres A., Schlessinger D., and Gros F.** (1960). Amino acid incorporation into proteins by *Escherichia coli* ribosomes. *Proc. Natl. Acad. Sci*. 46(11): 1450-1463.
- Tourand, Y., Kobryn, K., and Chaconas, G.** (2003). Sequence-specific recognition but position-dependent cleavage of two distinct telomeres by the *Borrelia burgdorferi* telomere resolvase, ResT. *Mol. Microbiol*. 48(4): 901-911.
- Tsai, H.-H., Shu, H.-W., Yang, C.-C, and Chen, C.W.** (2012). Translesion-synthesis DNA polymerases participate in replication of the telomeres in *Streptomyces*. *Nucleic Acids Res*. 40: 1118-1130.
- Tsien, R.Y.** (2009). Constructing and exploiting the fluorescent protein paintbox (Nobel Lecture). *Angew Chem Int Ed Engl*. 48(31):5612-5626.
- Tsugita, A., and Fraenkel-Conrat, H.** (1960). The amino-acid composition and C-terminal sequence of a chemically evoked mutant of TMV. *Proc. Natl. Acad. Sci*. 46(5): 636-642.
- Tupy, J.L., Bailey, A.M., Dailey, G., Evans-Holm, M., Siebel, C.W., Misra, S., Celniker, S.E., and Rubin, G.M.** (2005). Identification of putative noncoding polyadenylated transcripts in *Drosophila melanogaster*. *Proc Natl Acad Sci*. 102(15): 5495-5500
- Typas, A., Banzhaf, M., Gross, C.A., and Vollmer W.** (2011). From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat. Rev. Microbiol*. 10(2): 123-136.

**Udvardy, A., Maine, E., and Schedl, P.** (1985). The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J Mol Biol.* 185(2) : 341–358.

**Valens, M., Penaud, S., Rossignol, M., Cornet, F., and Boccard, F.** (2004). Macrodomain organization of the *Escherichia coli* chromosome. *EMBO J.* 23: 4330–4341.

**Van den Ent, F., Møller-Jensen, J., Amos, L.A., Gerdes, K., and Löwe, J.** (2002). F-actin-like filaments formed by plasmid segregation protein ParM. *EMBO J.* 21:6935–6943.

**van Steensel, B., and Belmont, A.S.** (2017). Lamina-associated domains: links with chromosome architecture, heterochromatin and gene repression. *Cell.* 169(5) : 780–791.

**Vaughan, E.E. and Dean, D.A.** (2006). Intracellular trafficking of plasmids during transfection is mediated by microtubules. *Mol. Ther.* 13(2): 422–428

**Vecchiarelli, A.G., Han, Y.W., Tan, X., Mizuuchi, M., Ghirlando, R., Biertümpfel, C., Funnell, B.E., and Mizuuchi, K.,** (2010). ATP control of dynamic P1 ParA–DNA interactions: a key role for the nucleoid in plasmid partition. *Mol. Microbiol.* 78(1): 78–91.

**Vecchiarelli, A.G., Hwang, L.C., and Mizuuchi, K.** (2013). Cell-free study of F plasmid partition provides evidence for cargo transport by a diffusion-ratchet mechanism. *Proc. Natl. Acad. Sci.* 110(15): E1390–E1397.

**Vecchiarelli, A.G., Neuman, K.C., and Mizuuchi, K.** (2014). A propagating ATPase gradient drives transport of surface-confined cellular cargo. *Proc. Natl. Acad. Sci.* 111(13): 4880–4885.

**Vendrely, R., and Vendrely, C.** (1948). The content of deoxyconucleic acid in the cell nucleus through organs, individuals and animal species. *Experientia.* 4(11):434-436.

**Victor, J.M.** (2012). Le structure de l'ADN en double hélice. *Bibnum, Sciences de la Vie.* <https://journals.openedition.org/bibnum/pdf/503>.

**Viollier, P.H., Thanbichler, M., McGrath, P.T., West, L., Meewan, M., McAdams, H.H., and Shapiro, L.** (2004) Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. *Proc. Natl. Acad. Sci.* 101(25): 9257-9262.

- Volff, J.N., and Altenbuchner, J.** (2000). A new beginning with news ends: linearization of circular chromosomes during bacterial evolution. *FEMS Microb. Letters*. 145: 2209-2220.
- Volkin, E., and Astrachan, L.** (1956). Intracellular distribution of labeled ribonucleic acid after phage infection of *Escherichia coli*. *Virology*; 2(4): 433-437.
- Volkin, E., Astrachan, L., and Countryman, L.** (1958). Metabolism of RNA phosphorus in *Escherichia coli* infected with bacteriophage T7. *Virology*. 6(2): 545-555.
- Volkin, E.** (2001). The discovery of mRNA. *Mutat. Res.* 488(2): 87-91.
- Voytas, D.F., and Joung, J.** (2009). Plant science. DNA binding made easy. *Science*. 326(5959):1491-1492.
- Wang, S., and Hazelrigg, T.** (1994). Implications for bcd mRNA localization from spatial distribution of exu protein in *Drosophila oogenesis*. *Nature*. 369(6479):400-403.
- Wang, X., Reyes-Lamothe, R., and Sherratt, D.J.** (2008a). Modulation of *Escherichia coli* sister chromosome cohesion by topoisomerase IV. *Genes Dev.* 22(17):2426–2433.
- Wang, X., Reyes-Lamothe, R., and Sherratt, D.J.** (2008b). Visualizing genetic loci and molecular machines in living bacteria. *Biochem Soc Trans.* 36(Pt 4) :749-53.
- Wang, X., and Sherratt, D.J.** (2010). Independent segregation of the two arms of the *Escherichia coli ori* region requires neither RNA synthesis nor MreB dynamics. *J. Bacteriol.* 192(23): 6143-6153.
- Wang, X., Montero Llopis, P., and Rudner, D.Z.** (2013). Organization and segregation of bacterial chromosomes. *Nat Rev Genet.* 14(3): 191-203.
- Wang, X., Montero Llopis, P., and Rudner, D.Z.** (2014a). *Bacillus subtilis* chromosome organization oscillates between two distinct patterns. *Proc. Natl. Acad. Sci.* 111(35):12877–12882.
- Wang X, Tang OW, Riley EP, and Rudner DZ.** (2014b). The SMC condensin complex is required for origin segregation in *Bacillus subtilis*. *Curr. Biol.* 24(3):287–292.

- Watanabe, K., Pecinka, A., Meister, A., Schubert, I., and Lam, E.** (2005). DNA hypomethylation reduces homologous pairing of inserted tandem repeat arrays in somatic nuclei of *Arabidopsis thaliana*. *Plant J.* 44(4): 531-540.
- Watson, J.D.** (1963). Involvement of RNA in the synthesis of proteins. *Science.* 140(3562): 17-26.
- Watson, J.D., and Crick, F.H.C.** (1953a). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature.* 171(4356): 737-738.
- Watson, J.D., and Crick, F.H.C.** (1953b). Genetical implications of the structure of deoxyribonucleic acid. *Nature.* 171(4361): 964-967.
- Watson, J.D., and Crick, F.H.C.** (1953c). The structure of DNA. *Cold Spring Harb. Symp.* 18: 123-131.
- Weimer, J., Kiechle, M., and Arnold, N.** (2000). FISH-microdissection (FISH-MD) analysis of complex chromosome rearrangements. *Cytogenet Cell Genet.* 88(1-2):114-118.
- Wilkins, M.H.F., Stokes, A.R., and Wilson, H.R.** (1953). Molecular structure of deoxyribose nucleic acid. *Nature.* 171(4356): 738-740.
- Williams, D.R., Macartney, D.P., and Thomas, C.M.** (1998). The partitioning activity of the RK2 central control region requires only *incC*, *korB* and *KorB*-binding site  $O_{B3}$  but other *KorB*-binding sites form destabilizing complexes in the absence of  $O_{B3}$ . *Microbiology* 144(12): 3369–3378.
- Wittmann, H.G.** (1960). Comparison of the tryptic peptides of chemically induced and spontaneous mutants of Tobacco Mosaic Virus. *Virology.* 12(4): 609-612.
- Woodcock, C.L.F., Safer, J.P., and Stanchfield, J.E.** (1976). Structural repeating units in chromatin. I. Evidence for their general occurrence. *Exp Cell Res.* 97 : 101–110.
- Worcel, A., and Burgi, E.** (1972). On the structure of the folded chromosome of *Escherichia coli*. *J. Mol. Biol.* 71: 127–147.
- Wu, L.J., and Errington, J.** (2004). Coordination of cell division and chromosome segregation by a nucleoid occlusion protein in *Bacillus subtilis*. *Cell.* 117(7): 915–25.

**Wu, L.J., Ishikawa, S., Kawai, Y., Oshima, T., Ogasawara, N., and Errington, J. (2009).** Noc protein binds to specific DNA sequences to coordinate cell division with chromosome segregation. *EMBO J.* 28(13): 1940–52.

**Yamaichi, Y., and Niki, H. (2000).** Active segregation by the *Bacillus subtilis* partitioning system in *Escherichia coli*. *Proc. Natl. Acad. Sci.* 97(26) : 14656–14661.

**Yang, B., Sugio, A., and White, F.F. (2006).** Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *Proc Natl Acad Sci.* 103(27): 10503-10508.

**Yang, M.C., and Losick, R. 2001.** Cytological evidence for association of the ends of the linear chromosome in *Streptomyces coelicolor*. *J. bacterial.* 183(17): 5180-5186.

**Youngren, B., Nielsen, H.J., Jun, S., and Austin, S. (2014).** The multifork *Escherichia coli* chromosome is a self-duplicating and self-segregating thermodynamic ring polymer. *Genes Dev.* 28(1): 71-84.

**Zanet, J., Benrabah, E., Li, T., Pelissier-Monier, A., Chanut-Delalande, H., Ronsin, B., Bellen, H.J., Payre, F., and Plaza, S. (2015).** Pri sORF peptides induce selective proteasome-mediated protein processing. *Science.* 349(6254): 1356-1358.

**Zanet, J., Chanut-Delalande, H., Plaza, S., and Payre, F. (2016).** Small Peptides as Newcomers in the Control of *Drosophila* Development. *Curr Top Dev Biol.* 117: 199-219.