



HAL
open science

Identification du chaperon d'histones Spt16 acteur essentiel des réarrangements du génome et méthylation des adénines chez *Paramecium tetraurelia*

Amandine Touzeau

► **To cite this version:**

Amandine Touzeau. Identification du chaperon d'histones Spt16 acteur essentiel des réarrangements du génome et méthylation des adénines chez *Paramecium tetraurelia*. Biologie cellulaire. Université Sorbonne Paris Cité, 2018. Français. NNT : 2018USPCC334 . tel-02988266

HAL Id: tel-02988266

<https://theses.hal.science/tel-02988266>

Submitted on 4 Nov 2020

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 de doctorat

De l'Université Sorbonne Paris Cité

Préparée à l'Université Paris Diderot

Ecole doctorale Bio Sorbonne Paris Cité (BioSPC) ED562

Institut Jacques Monod / Equipe Régulation épigénétique de l'organisation du génome

Identification du chaperon d'histones Spt16
acteur essentiel des réarrangements du génome
et méthylation des adénines
chez *Paramecium tetraurelia*

Par Amandine Touzeau

Thèse de doctorat de Biologie cellulaire

Dirigée par Dr Sandra Duharcourt

Présentée et soutenue publiquement à Paris le 25/09/2018

Président du jury : Mme Pascale Lesage / DR / Université Paris Diderot, CNRS

Rapporteurs : M. Éric Meyer / DR / Ecole Normale Supérieure, CNRS

Mme Fabienne Malagnac / Professeure / Université Paris Sud, CNRS

Examineurs : Mme Sophie Polo / DR / Université Paris Diderot, CNRS

M Jean Luc Férat / MC / Université Paris Sud, CNRS

Directrice de thèse : Mme Sandra Duharcourt / DR / Université Paris Diderot, CNRS

Titre : Identification du chaperon d'histones Spt16 acteur essentiel des réarrangements du génome et méthylation des adénines chez *Paramecium tetraurelia*

Résumé :

Chez les eucaryotes, l'organisation de la chromatine est indispensable à la régulation de l'expression des gènes et la stabilité du génome. Les ciliés constituent un excellent modèle d'étude des mécanismes qui permettent de maintenir l'intégrité du génome eucaryote. Notre modèle d'étude, l'eucaryote unicellulaire *Paramecium tetraurelia*, se distingue par l'élimination massive et reproductible de près de 30% de séquences d'ADN germinale lors du développement du macronoyau somatique après les événements sexuels. Ces séquences sont éliminées grâce à un processus en plusieurs étapes impliquant la formation d'hétérochromatine dirigée par de petits ARN suivi d'excision par la transposase domestiquée Piggy Mac (Pgm) et de réparation. Les mécanismes moléculaires permettant de comprendre la reconnaissance spécifique de ces séquences germinales dans le contexte de la chromatine ainsi que la précision de la coupure demeurent mal compris.

Le chaperon d'histone Spt16, associée au partenaire Pob3, fait partie du complexe hétérodimérique FACT (FACilitates Chromatin Transactions). FACT intervient dans de nombreux mécanismes liés au métabolisme de l'ADN tels que la transcription, la réparation, la réplication ou l'accessibilité de la chromatine. Chez *Paramecium tetraurelia*, on a identifié deux protéines homologues de Spt16 et Pob3 exprimées uniquement durant le développement du macronoyau au moment où se produisent les réarrangements. Les protéines Spt16-1 et Pob3-1 fusionnées à la GFP se localisent dans les macronoyaux en développement. Nous avons montré que la protéine Spt16-1 est essentielle à la production d'une descendance sexuelle viable. Le reséquençage du génome après inactivation de *SPT16-1* a montré que Spt16-1 est nécessaire à l'ensemble des réarrangements du génome et conduit à des défauts semblables à ceux obtenus suite à l'inactivation de *PGM*. Spt16-1 agit en aval des voies de dépôt des marques d'hétérochromatine guidées par les petits ARNs mais en amont de l'endonucléase Pgm. Nous avons montré que Spt16-1 était nécessaire à la localisation correcte de Pgm responsable de l'introduction des cassures double brin dans les macronoyaux en développement. Nous proposons un modèle dans lequel Spt16-1 favorise l'interaction de la machinerie d'excision avec la chromatine pour rendre accessible les sites de coupure sur l'ADN à Pgm.

La méthylation des adénines sur l'ADN (6mA), bien connue chez les bactéries pour son rôle dans le système de restriction modification, a été décrite ces deux dernières années chez plusieurs eucaryotes en faible proportion dans le génome. Cependant son rôle chez les eucaryotes est encore peu compris. D'anciennes analyses par chromatographie à l'aide de nucléotides marqués détectaient de l'ordre de 2,5% des adénines méthylées chez *P. tetraurelia* mais sa localisation précise n'avait jamais été montrée. Il avait été proposé que cette modification pourrait signaler avec précision les sites de coupures pendant les réarrangements d'ADN où une partie des séquences germinales présentent des bornes AT. L'abondance des adénines méthylées fait de la paramécie un excellent modèle d'étude de cette modification. En combinant des techniques d'immunofluorescence, d'HPLC-MS et de séquençage, nous avons ainsi décrit la présence et la localisation dans la cellule et au cours du cycle de vie de 6mA chez *P. tetraurelia*. Ces approches ont également permis de montrer l'absence de cytosine méthylée chez la paramécie. Nous avons montré que la méthylation des adénines était retrouvée majoritairement dans le génome somatique et de manière transitoire dans le génome germinale. Elle est établie au cours du développement du macronoyau somatique quand se produisent les réarrangements du génome. Ces résultats préliminaires permettront de poursuivre l'étude de la méthylation afin d'identifier les enzymes responsables et son rôle dans la cellule.

Mots clefs : Chaperon d'histone, Réarrangements programmés du génome, Ciliés, Chromatine, *Paramecium*, méthylation des adénines

Title: Identification of the histone chaperone Spt16, essential component of programmed genome rearrangements and DNA methylation on adenine in *Paramecium tetraurelia*

Abstract:

In eukaryotes, chromatin organization is required for the regulation of gene expression and genome stability. Ciliate provide excellent model to study mechanisms involved in maintain of genome integrity. Our study model, the unicellular eukaryote *Paramecium tetraurelia*, has the particularity to eliminate massively and reproducibly 30% of germinal DNA sequences during the development of the somatic macronucleus after sexual events. Those sequences are eliminated by a multi-step process involving small RNA-directed heterochromatin formation followed by DNA excision by the domesticated transposase Piggy Mac (Pgm) and DNA repair. Molecular mechanisms underlying the specific recognition of those germinal sequences in chromatin context and the precision of the excision, remain elusive.

The histone chaperone Spt16, associated to its partner Pob3, is part of the heterodimeric complex FACT (FACilitates Chromatin Transactions). FACT is implicated in many mechanisms involving DNA metabolism such as transcription, repair, replication or chromatin accessibility. In *P. tetraurelia*, we identified two homologous proteins to Spt16 and Pob3 expressed only during macronucleus development at the time when genome rearrangements occur. Spt16-1 and Pob3-1 fused to GFP are localized in developing macronuclei. We showed that Spt16-1 is required to obtain a viable sexual progeny. Genome re-sequencing after *SPT16-1* inactivation showed that Spt16-1 was required for all DNA elimination events and leads to similar phenotypes and defects to those obtained after *PGM* inactivation. Spt16-1 acts downstream of small RNA-directed heterochromatin formation and upstream of Pgm. We showed that Spt16-1 was required for the correct localization of Pgm responsible for DNA double strand breaks in developing macronuclei. We proposed a model in which Spt16-1 mediates interaction between chromatin and excision machinery that facilitates access to DNA cleavage sites for the Pgm endonuclease.

Adenine DNA methylation, well known in bacteria for its role in restriction modification system, has been described during the last two years in several eukaryotes but in low proportion in the genome. However, its role in eukaryotes remains elusive. Previous analyses by chromatography with radio labelled nucleotides detected around 2,5% of methylated adenine in *P. tetraurelia* but its precise localization and role have never been analyzed. It has been proposed that this modification could mark with precision the excision sites during genome rearrangements when part of the eliminated sequences carry AT boundaries. The abundance of methylated adenine makes of *Paramecium* an excellent model to study this modification. Combining immunofluorescence techniques, HPLC-MS and sequencing, we described the presence and the cellular localization during life cycle of 6mA in *P. tetraurelia*. Those approaches allowed us to show that methylated cytosine are absent in *Paramecium*. We showed that methylation is mainly found in the somatic genome and transiently in germinal genome. It appears during somatic macronucleus development when genome rearrangements occur. Those preliminary results will allow us to pursue the study of adenine methylation by identifying responsible enzymes and its role in the cell.

Keywords: Histone chaperone, Programmed genome rearrangements, ciliate, chromatin, paramecium, adenine methylation

Remerciements

Je tiens d'abord à remercier l'ensemble des membres de mon jury, Mme Fabienne Malagnac, M Éric Meyer, Mme Pascale Lesage, Mme Sophie Polo et M Jean Luc Férat pour avoir accepté d'évaluer mon travail. Je remercie tout particulièrement mes deux rapporteurs pour le temps qu'ils ont donné à la lecture du manuscrit au milieu de l'été.

Je remercie aussi bien sûr énormément Sandra Duharcourt qui m'a encadrée pendant cinq années depuis mon stage de Master 2 lorsque je suis arrivée au laboratoire. Merci à elle pour son implication, sa pédagogie et sa grande culture scientifique, mais aussi sa patience et sa disponibilité. Merci à elle de m'avoir autant transmis scientifiquement en théorie comme en pratique et de m'avoir offert l'opportunité de participer à de nombreux congrès, d'échanger avec de nombreux autres scientifiques et d'encadrer plusieurs étudiants.

Un grand merci à mon comité de thèse, M Raphael Margueron et Mme Fabienne Malagnac pour leur aide et leurs conseils pendant ma thèse. Leurs avis auront été d'une grande aide pour la poursuite de ce projet. Merci également à Slimane Ait-Si-Ali pour son aide précieuse sur le projet Spt16 !

Durant mes cinq années au laboratoire, j'ai eu la chance de travailler avec une équipe attachante, scientifiquement enrichissante mais aussi pleine de bonne humeur et d'amitié. Merci à tous les membres passés ! Merci d'abord à Augustin de Vanssay sans qui le projet Spt16 n'aurait pas vu le jour. Merci aussi à Andrea, Frédéric et Maoussi pour tout ce qu'ils m'ont transmis, mais aussi leur soutien et nos nombreuses discussions qu'elles soient scientifiques ou sur le dernier épisode de Game of Thrones. Merci aussi à tous les membres présents actuellement dans l'équipe : Isadora, Takayuki, Caridad et Adeline avec qui j'ai toujours plaisir à échanger.

Merci également aux autres équipes « Paramécie » avec qui j'ai eu l'occasion de travailler et d'échanger aux cours de ces années de thèses. Merci à Olivier Arnaiz pour son formidable travail sur les données de séquençage. Merci à l'équipe de Mireille Bétermier pour avoir échangé et aidé à produire et analyser certains des résultats Spt16, en particulier Vinciane et Arthur pour leur aide sur les immunofluorescences et les quantifications, mais

aussi Linda, Julien, Marc, etc. Merci aussi à l'équipe d'Éric Meyer pour nos échanges, Sophie, Véronique, Suzanne, et ceux que j'oublie.

Un grand merci à tous ceux de l'IJM qui nous aident à travailler au quotidien au magasin, à l'entretien, à la laverie, aux milieux, à la gestion, etc. Merci également à la plateforme d'imagerie pour nous avoir aidé à la production de nos images.

Merci aussi à tous ceux de l'institut avec qui j'ai échangé au cours de ces années !

Un merci tout particulier à celui qui partage ma vie et qui aura vécu cette thèse avec moi, Eliott, merci d'avoir été là, même dans les périodes les plus difficiles, d'avoir cru en moi et de m'avoir soutenue au quotidien ! Ton aide est inestimable. Ces années auront été parfois dures mais je suis plus qu'heureuse de t'avoir eu à mes côtés.

Merci à toutes celles et ceux en dehors du laboratoire qui m'ont apporté leur aide et leur soutien que ça soit lors de l'écriture du manuscrit ou pendant les années de thèse. Roxane, Galadriel, Cyril, Luna, etc et j'en oublie tellement, merci pour vos petits mots et votre patience. Votre amitié m'est très précieuse.

Merci à ma famille, mes parents en particulier qui m'ont toujours encouragée dans mes projets et mes études et m'ont permis aussi d'en arriver là.

Un merci aussi à ces enseignants qui m'ont fait aimer les sciences et en particulier les SVT pendant ma scolarité et mes années d'université.

Finalement un merci à mes deux petites boules de poils, mes deux petits chats qui m'auront finalement beaucoup aidée pendant la rédaction du manuscrit ...

Table des matières

REMERCIEMENTS	7
PREAMBULE.....	15
INTRODUCTION.....	19
CHAPITRE 1 : CHAPERONS D’HISTONE ET MAINTIEN DE L’INTEGRITE DU GENOME, SPT16 ET LE COMPLEXE FACT, ACTEUR MAJEUR DANS L’ORGANISATION DU GENOME CHEZ LES EUCARYOTES	19
I. DECOUVERTE DU COMPLEXE FACT ET CONSERVATION CHEZ LES EUCARYOTES	19
II. FACT EST UN HETERODIMERE	21
1. LE CHAPERON D’HISTONES SPT16, PIVOT DU COMPLEXE FACT	21
2. SSRP1/POB3, PARTENAIRE DE SPT16	23
3. REGULATION DU COMPLEXE FACT	24
3.1 RECRUTEMENT DE FACT A DES CIBLES SPECIFIQUES.....	24
3.2 DES MODIFICATIONS POST-TRADUCTIONNELLES DES SUBSTRATS DE FACT MODIFIENT SON ACTIVITE	25
3.3 DES MODIFICATIONS POST-TRADUCTIONNELLES DE FACT MODULENT SON ACTIVITE.....	25
3.4 LA QUANTITE DE FACT DANS LA CELLULE EST REGULEE.....	26
III. ROLE DE FACT DANS LA CELLULE, « FACILITES CHROMATIN TRANSCRIPTION TO TRANSACTION ».....	26
1. FACT ASSURE LA DYNAMIQUE DES NUCLEOSOMES : MODELE D’INTERACTION AVEC LES HISTONES	26
1.1 MODELE D’INTERACTION AVEC LE NUCLEOSOME : HYPOTHESE DE L’EVICTION DU DIMERE H2A-H2B.....	28
1.2 MODELE D’INTERACTION AVEC LE NUCLEOSOME : HYPOTHESE DU GLOBAL ACCESS MODEL ET NUCLEOSOME BREATHING	28
2. « FACILITES CHROMATIN TRANSCRIPTION ».....	31
2.1 FACT GARANTIT L’INITIATION CORRECTE DE LA TRANSCRIPTION.....	31
2.2 FACT ASSURE L’ELONGATION DE LA TRANSCRIPTION.....	32
2.3 L’EFFET DE FACT SUR LA TRANSCRIPTION EST-IL GLOBAL OU CIBLE SUR CERTAINS GENES ?	34
3. « FACILITES REPLICATION »	35
4. « FACILITES REPAIR ».....	37
5. FACT PARTICIPE A L’INTEGRITE DES CENTROMERES.....	38
CHAPITRE 2 : METHYLATION DES ADENINES.....	40
I. METHYLATION ET AUTRES MODIFICATIONS DE L’ADN.....	40
1. AU-DELA DES BASES CANONIQUES DE L’ADN	40
2. METHODES DE DETECTION ET D’ANALYSE DE LA METHYLATION DE L’ADN	42
2.1 DETECTER ET QUANTIFIER LA METHYLATION	43
2.2 LOCALISER LA METHYLATION DANS LE GENOME	44
II. METHYLATION DES ADENINES	47
1. 6-METHYLADENINE CHEZ LES PROCARYOTES.....	47
1.1 SYSTEME DE RESTRICTION MODIFICATION ET METHYLTRANSFERASES SOLITAIRES CHEZ LES PROCARYOTES.....	47
1.2 6MA, UNE MARQUE EPIGENETIQUE CHEZ LES BACTERIES.....	48
1.3 ADENINES METHYLTRANSFERASES	49
2. 6-METHYLADENINE CHEZ LES EUCARYOTES.....	51
2.1 LA METHYLATION DES ADENINES EST-ELLE PRESENTE CHEZ LES EUCARYOTES PLURICELLULAIRES ?	51
2.1.1 L’AMELIORATION DES TECHNIQUES DE DETECTION IDENTIFIE 6MA CHEZ DES EUCARYOTES PLURICELLULAIRES .	52
2.1.2 L’ADN DES MAMMIFERES EST-IL METHYLE SUR LES ADENINES ?.....	54
2.2 LA METHYLATION DES ADENINES CHEZ LES EUCARYOTES UNICELLULAIRES.....	55
2.2.1 REPARTITION ET ROLE.....	55

2.2.2 LES CILIES, UN MODELE D'ETUDE DE LA METHYLATION DES ADENINES	56
CHAPITRE 3 : PARAMECIUM TETRAURELIA, MODELE D'ETUDE DE LA REGULATION EPIGENETIQUE DE L'ORGANISATION DU GENOME.....	58
I. CLASSIFICATION, DIMORPHISME NUCLEAIRE ET CYCLE DE VIE.....	58
II. ORGANISATION DES GENOMES.....	60
III. DEVELOPPEMENT DU NOYAU SOMATIQUE	61
1. ENDOREPLICATION.....	61
2. ELIMINATION D'ADN.....	61
2.1 ELIMINATION IMPRECISE ET HETEROGENE DE SEQUENCES.....	62
2.2 ELIMINATION PRECISE D'IES.....	63
3. LA MACHINERIE D'EXCISION ET DE REPARATION.....	66
3.1 PGM, L'ENDONUCLEASE RESPONSABLE DE L'INTRODUCTION DES CASSURES DOUBLE BRINS.....	66
3.2 LE NON-HOMOLOGUOUS-END-JOINING POUR REPARER LES CASSURES.....	67
4. RECONNAITRE LES SEQUENCES A ELIMINER ET CIBLER LA MACHINERIE D'EXCISION.....	69
4.1 CONTROLE EPIGENETIQUE DE L'ELIMINATION.....	69
4.2 ARN NON CODANTS ET GENOME SCANNING MODEL.....	70
4.2.1 PETITS ARN NON CODANTS : SCNARN ET IESARN.....	70
4.2.2 LONGS ARN NON CODANTS MATERNELS.....	71
4.2.3 GENOME SCANNING MODEL.....	72
4.3 L'HISTONE METHYLTRANSFERASE <i>EZL1</i> EST REQUISE POUR LE DEVELOPPEMENT MAC.....	75
4.4 LES MECANISMES D'ELIMINATION DES IES REPOSENT SUR DES VOIES CHEVAUCHANTES.....	76
4.5 DE NOUVEAUX ACTEURS POUR COMPRENDRE COMMENT LA MACHINERIE D'EXCISION EST CIBLEE A L'ENSEMBLE DES SEQUENCES ?	78
OBJECTIFS DE THESE :	79
RESULTATS.....	81
PARTIE I : LE CHAPERON D'HISTONE SPT16 CHEZ <i>PARAMECIUM TETRAURELIA</i> EST UN COMPOSANT ESSENTIEL DES REARRANGEMENTS PROGRAMMES DU GENOME	81
ARTICLE DE RECHERCHE.....	81
RESULTATS SUPPLEMENTAIRES.....	136
I. LE DOMAINE MBD DE SPT16-1 SEMBLE ESSENTIEL A L'ACTIVITE DE SPT16-1 PENDANT LE DEVELOPPEMENT MAC 136	
II. IDENTIFICATION D'HOMOLOGUES DE POB3 CHEZ <i>PARAMECIUM TETRAURELIA</i>	139
1. IDENTIFICATION D'HOMOLOGUES DE LA PROTEINE POB3 CHEZ <i>PARAMECIUM TETRAURELIA</i>	139
2. POB3 N'EST PAS NECESSAIRE AUX EVENEMENTS SEXUELS NI AU DEVELOPPEMENT MAC.....	142
3. POB3 EST LOCALISE DANS LES MACRONOYAUX EN DEVELOPPEMENT	143
PARTIE II : ANALYSE DE LA METHYLATION DES ADENINES CHEZ <i>PARAMECIUM TETRAURELIA</i>	146
I. DETECTION DE BASES MODIFIEES DANS LE GENOME DE <i>PARAMECIUM TETRAURELIA</i>	146
1. DETECTION DE NUCLEOSIDES MODIFIES PAR HPLC-MS DANS LE GENOME.....	146
2. LA METHYLATION DES CYTOSINES EST ABSENTE DU GENOME DE <i>P. TETRAURELIA</i>	148
3. LA METHYLATION DES ADENINES EST PRESENTE DANS L'ADN DE <i>P. TETRAURELIA</i>	149
II. DESCRIPTION DE LA LOCALISATION DE LA METHYLATION DES ADENINES CHEZ <i>PARAMECIUM TETRAURELIA</i>	151
1. LA METHYLATION DES ADENINES EST MAJORITAIREMENT DETECTEE SUR L'ADN.....	151
2. LA METHYLATION DES ADENINES EST ETABLIE PENDANT LE DEVELOPPEMENT DES MACRONOYAUX ET PERDURE PENDANT TOUT LE CYCLE CELLULAIRE.....	152
3. LA METHYLATION DES ADENINES S'ETABLIT EN AVAL DES REARRANGEMENTS D'ADN DANS LES NOUVEAUX MACRONOYAUX.....	153
III. LOCALISATION DE 6MA DANS LE GENOME DE <i>P. TETRAURELIA</i>	155
1. ADAPTATION D'UN PROTOCOLE D'IMMUNOPRECIPITATION DE L'ADN METHYLE SUR 6MA CHEZ <i>P. TETRAURELIA</i> 155	

2. L'IMMUNOPRECIPITATION PAR L'ANTICORPS 6MA N'A PAS PERMIS D'ENRICHER EN REGIONS METHYLEES	157
---	-----

DISCUSSION..... 160

CHAPITRE 1 : LE CHAPERON D'HISTONE SPT16 CHEZ *PARAMECIUM TETRAURELIA* EST UN COMPOSANT ESSENTIEL DES REARRANGEMENTS PROGRAMMES DU GENOME 160

I. DEMONSTRER L'ACTIVITE CHAPERON D'HISTONE DE SPT16-1 160

1. A LA RECHERCHE DES PARTENAIRES DE SPT16-1 161

2. LE DOMAINE MBD POURRAIT ETRE REQUIS POUR L'ACTIVITE DE SPT16-1 DANS LES REARRANGEMENTS DU GENOME..... 162

3. DEMONSTRER LES INTERACTIONS ENTRE LE DOMAINE MBD DE SPT16-1 ET LES HISTONES..... 163

II. EXISTENCE D'UN COMPLEXE FACT CHEZ *P. TETRAURELIA* 164

1. A LA RECHERCHE D'UN PARTENAIRE DE SPT16-1..... 164

1.1 *P. TETRAURELIA* DISPOSE D'UN HOMOLOGUE DE *POB3*..... 164

1.2 *POB3-1* A UNE LOCALISATION SIMILAIRE A *SPT16-1*..... 165

1.3 DIMERISATION DE *POB3-1* ET *SPT16-1* 165

2. IMPLICATION DE *POB3-1* DANS LE DEVELOPPEMENT MAC..... 166

3. MODELE D'ACTION DES DIFFERENTS *POB3* ET *SPT16* CHEZ *P. TETRAURELIA*..... 167

III. COMPRENDRE COMMENT SPT16-1 PERMET LES REARRANGEMENTS DU GENOME..... 168

1. L'ACTION DE SPT16-1 DANS LE DEVELOPPEMENT MACRONUCLEAIRE N'EST PAS REGULE PAR LES SCNARN OU LES MARQUES D'HETEROCHROMATINE..... 169

1.1 *SPT16-1* N'EST PAS RECRUTE PAR L'INTERMEDIAIRE DES SCNARNS OU DES MARQUES D'HETEROCHROMATINE 170

1.2 *SPT16-1* PERMET DE PRECISER COMMENT SONT RECONNUES ET ELIMINEES LES SEQUENCES NE DEPENDANT PAS DE *Dcl2/3* OU *Ez11* 170

2. *SPT16-1* POURRAIT FACILITER L'ACCES A L'ADN DE LA MACHINERIE D'EXSION 171

3. *SPT16-1* POURRAIT-IL FACILITER LA REPARATION CHEZ *P. TETRAURELIA* ? 174

4. *SPT16-1* ET SON EFFET SUR LA TRANSCRIPTION CHEZ *P. TETRAURELIA*..... 175

CHAPITRE 2 : METHYLATION DES ADENINES 177

I. IDENTIFICATION DE MODIFICATIONS DANS L'ADN DE *P. TETRAURELIA*..... 177

1. 5mC, 5hmC, 4mC ET 8oxoG NE SONT PAS DETECTEES CHEZ *P. TETRAURELIA* 177

2. DETECTION DE 6MA DANS L'ADN DE *P. TETRAURELIA*..... 179

II. LA METHYLATION DES ADENINES EST ETABLIE PENDANT LE DEVELOPPEMENT MACRONUCLEAIRE 180

1. METHYLATION *DE NOVO* ET DE MAINTENANCE CHEZ *P. TETRAURELIA*..... 180

2. A LA RECHERCHE DE METHYLTRANSFERASES DES ADENINES 181

3. LA METHYLATION DES ADENINES DANS LES MIC..... 184

4. METHYLATION DES ADENINES ET REARRANGEMENTS DU GENOME..... 185

III. LOCALISER DANS LE GENOME DE *P. TETRAURELIA* LA METHYLATION DES ADENINES 185

1. ADAPTATION DU PROTOCOLE DE MEDIP..... 185

2. LOCALISER 6MA PAR SMRT SEQUENCING..... 187

BIBLIOGRAPHIE 188

ANNEXES 1..... 204

ANNEXE 2..... 205

ANNEXE 3..... 206

Table des illustrations

FIGURE 1 : REPRESENTATION SCHEMATIQUE DU COMPLEXE FACT ET DE SES DOMAINES CHEZ LA LEVURE	21
FIGURE 2 : MODELE D'INTERACTION DE FACT AVEC LE NUCLEOSOME	27
FIGURE 3 : MODELE D'INTERACTION DES EXTREMITES C-TERMINALES DE SPT16 ET POB3 AVEC LE NUCLEOSOME	30
FIGURE 4 : ROLE DE FACT DANS L'INHIBITION DE LA TRANSCRIPTION CRYPTIQUE.	34
FIGURE 5 : MODELE D'ACTION DU COMPLEXE FACT PENDANT LA REPLICATION.	37
FIGURE 6 : METHYLATION SUR LES BASES DE L'ADN	42
FIGURE 7 : PRESENCE DE METHYLATION DES CYTOSINES OU DES ADENINES DANS LE MONDE VIVANT	46
FIGURE 8 : ROLE DE LA METHYLATION DES ADENINES CHEZ LES BACTERIES	49
FIGURE 9 : CYCLE DE VIE DE PARAMECIUM TETRAURELIA	60
FIGURE 10 : ELIMINATION D'ADN ET REPARATION DES CASSURES DOUBLE BRIN PENDANT LE DEVELOPPEMENT DU NOUVEAU MAC	65
FIGURE 11 : MODELE DE GENOME SCANNING	72
FIGURE 12 : MODELE D'ELIMINATION D'ADN ET VOIES D'EXCISION DES IES	76
FIGURE 13 : ANALYSE FONCTIONNELLE DU DOMAINE MBD DE SPT16-1	138
FIGURE 14 : IDENTIFICATION D'HOMOLOGUES DE POB3 CHEZ PARAMECIUM TETRAURELIA	141
FIGURE 15 : LOCALISATION DE LA PROTEINE POB3-1 A L'AIDE D'UNE FUSION PROTEIQUE A LA GFP	145
FIGURE 16 : QUANTIFICATION PAR HPLC-MS DES NIVEAUX DE METHYLATIONS DES CYTOSINES ET DES ADENINES DE L'ADN GENOMIQUE TOTAL DE P. TETRAURELIA ET P. SEXAURELIA	148
FIGURE 17 : DETECTION DE 6MA DANS LE GENOME DE PARAMECIUM TETRAURELIA AVEC DES ANTICORPS ANTI-6MA.	151
FIGURE 18 : LOCALISATION DE 6MA DANS LE GENOME DE P. TETRAURELIA PAR MEDIP	154
FIGURE 19 : MODELES D'ACTION DE SPT16-1 ET DE COMPOSITION DE COMPLEXES FACT CHEZ P. TETRAURELIA	168
FIGURE 20 : MODELE D'ACTION DE SPT16-1 DANS LES REARRANGEMENTS DU GENOME	173
FIGURE 21 : ANALYSE PHYLOGENETIQUE DES ADENINES METHYLTRANSFERASES PUTATIVES A DOMAINE MTA70 DE P. TETRAURELIA.	184

Préambule

L'ADN des cellules eucaryotes est organisé dans le noyau sous forme de chromatine : l'ADN y est arrangé autour de protéines histones et non histones dont l'unité est le nucléosome (Hammond et al., 2017). Chaque nucléosome est constitué d'environ 147 paires de bases s'enroulant autour d'un octamère d'histones. Le nucléosome est composé de deux paires de chaque histone, un tétramère d'H3 et H4 et deux dimères d'H2A et H2B. Cette organisation dite canonique peut être modifiée par l'incorporation de variants d'histone et de modifications post-traductionnelles des histones, ainsi que des modifications chimiques sur les bases de l'ADN (Hammond et al., 2017; Iyer et al., 2011). Ce niveau de complexité supplémentaire permet de réguler la conformation de la chromatine, l'expression des gènes et l'architecture générale du noyau. Le maintien d'une organisation dynamique de la chromatine et des modifications associées est nécessaire à l'ensemble des mécanismes liés au métabolisme de l'ADN et plus largement au maintien de l'intégrité des génomes eucaryotes. Un degré de compaction élevé de la chromatine empêche l'accès à l'ADN des différentes machineries de réplication, transcription, etc. Tandis qu'un niveau de compaction plus faible permet l'accès à l'ADN et favorise ces processus cités avant (Corpet and Almouzni, 2009). Pour comprendre comment est maintenue l'intégrité des génomes eucaryotes il est donc important de comprendre comment est régulée l'architecture chromatinienne.

Pour permettre à la fois le maintien de cette organisation de la chromatine ainsi que sa flexibilité, une classe particulière de protéines nommées les chaperons d'histones prennent en charge les histones libres ou dans le nucléosome. Bien que ces protéines partagent des fonctions communes, elles regroupent une grande diversité de domaines fonctionnelles avec peu ou pas de similarité de séquence et des propriétés structurales différentes (Hammond et al., 2017; Laskey et al., 1978). L'un des points communs de ces protéines reste cependant la présence de domaines riches en résidus acides permettant des interactions fortes avec les histones et rentrant en compétition avec l'ADN. On peut définir les chaperons d'histones comme des protéines prenant en charge *in vivo* les histones non incorporées au nucléosome et promouvant l'assemblage du nucléosome *in vitro* à partir d'histones isolées et d'ADN. Si

cette définition permet de regrouper cette catégorie de protéines, elle ne reflète pas la grande diversité de fonctions auxquelles elles sont associées telles que la réplication, la transcription, la réparation, l'intégrité des centromères, l'incorporation de variants d'histones, etc (Adam et al., 2015; Hammond et al., 2017; Ransom et al., 2010; Venkatesh and Workman, 2015). L'ensemble de ces processus nécessite de perturber l'organisation de la chromatine tout en s'assurant de la restaurer par la suite, en prenant en compte le recyclage des histones, l'incorporation de nouvelles histones mais aussi le maintien de l'information épigénétique. Les chaperons fonctionnent en réseau et à l'aide de partenaires co-chaperons au sein de complexes qui ont l'avantage de permettre l'interaction la plus complète avec les histones (Hammond et al., 2017). Les analyses aux rayons X des structures cristallisées de chaperons d'histone ont identifié les domaines et modes d'interaction avec les histones et les partenaires des complexes de co-chaperons. L'un de ces complexes, FACT (FAcilitates Chromatin transaction) composé de deux co-chaperons Spt16 et Pob3/Ssrp1 est retrouvé chez de nombreux eucaryotes (Formosa, 2012; Jeronimo and Robert, 2016). Ses interactions fortes et multiples avec toutes les histones lui assurent un rôle dans quasiment tous les processus impliquant la chromatine telle que la réplication, la transcription ou encore la réparation, en faisant un chaperon particulièrement important pour assurer l'intégrité des génomes eucaryotes.

Les modifications chimiques portées par les bases de l'ADN apportent un autre niveau de complexité assurant le maintien de l'intégrité des génomes et influençant la structure de la chromatine et sa dynamique (Iyer et al., 2011). Il s'agit principalement de méthylations. La modification la plus abondante et la plus fréquente en particulier chez les eucaryotes concerne la méthylation des cytosines (Breiling and Lyko, 2015). Elle est connue pour être impliquée dans de nombreux processus, telle que la régulation de l'expression des gènes, l'empreinte génétique, l'inactivation de chromosomes, la déposition de modifications post-traductionnelles des histones, etc. Les adénosines méthylées, la deuxième modification la plus fréquente dans le monde vivant, sont surtout retrouvées dans les génomes procaryotes associées au système de restriction modification, une forme de système immunitaire primitif (O'Brown and Greer, 2016; Ratel et al., 2006). Des travaux récents ainsi que l'amélioration des techniques d'analyse des modifications de l'ADN ont mis en évidence la présence d'adénosines méthylées dans des proportions relativement faibles chez certains eucaryotes

sans identifier de rôle précis pour le moment (Fu et al., 2015; Greer et al., 2015; Koziol et al., 2016; Liang et al., 2018a, 2018b; Mondo et al., 2017; Wang et al., 2017; Wu et al., 2016; Zhang et al., 2015). Des travaux plus anciens avaient déjà détecté la méthylation des adénosines chez des ciliés dans des proportions similaires à celles retrouvées pour la méthylation des cytosines chez les autres eucaryotes, sans y associer de rôle (Cummings et al., 1974; Gorovsky et al., 1973; Harrison et al., 1986).

Notre organisme modèle, le cilié *Paramecium tetraurelia*, nous permet d'étudier les mécanismes épigénétiques qui contrôlent l'intégrité des génomes eucaryotes. Cet unicellulaire a séparé ses fonctions somatiques et germinales dans deux types de noyaux différents dans un même cytoplasme (Betermier and Duharcourt, 2014). Le génome somatique est produit à partir du génome germinale par l'élimination massive et reproductible de nombreuses séquences répétées et de séquences uniques dérivées de transposon (Arnaiz et al., 2012). La reconnaissance et le ciblage de ces séquences multiples par la machinerie d'excision fait intervenir des mécanismes épigénétiques impliquant notamment des petits ARN ou des modifications de la chromatine. Ces acteurs identifiés des réarrangements du génome ne permettent cependant pas d'expliquer l'ensemble des séquences éliminées, ni la précision de la coupure et l'accès de la machinerie d'excision à l'ADN.

Pour permettre de mieux comprendre les mécanismes qui gouvernent l'élimination d'ADN lors du développement du noyau somatique chez *Paramecium tetraurelia*, mon travail de thèse a consisté à :

- 1) Etudier le rôle d'un homologue du chaperon d'histones Spt16 dans les réarrangements du génome. Nous avons mis en évidence que cette protéine est nécessaire à l'ensemble des évènements d'élimination.
- 2) Etudier la méthylation des adénosines chez *Paramecium tetraurelia*, une modification abondamment présente qui pourrait guider la précision de l'élimination d'ADN mais dont le rôle exact reste encore incompris.

Dans l'introduction qui va suivre, je vais dans un premier temps décrire le fonctionnement du chaperon d'histones Spt16 chez les eucaryotes et du complexe FACT auquel il appartient. Dans un second temps, je présenterais les caractéristiques de la

méthylation des adénosines chez les procaryotes et chez les eucaryotes chez qui elle a été décrite. Enfin je décrirais notre organisme modèle, *Paramecium tetraurelia*, ce que nous connaissons des mécanismes gouvernant l'élimination d'ADN et ce qui en fait un excellent modèle d'étude du maintien de l'intégrité des génomes eucaryotes.

Introduction

Chapitre 1 : Chaperons d'histone et maintien de l'intégrité du génome, Spt16 et le complexe FACT, acteur majeur dans l'organisation du génome chez les eucaryotes

L'organisation dynamique de la chromatine assure le maintien de la stabilité des génomes eucaryotes. La chromatine doit à la fois permettre l'accès à l'ADN pour les différentes machineries de réplication, transcription, et autres facteurs interagissant avec l'ADN mais aussi conserver son intégrité structurale et les modifications qui lui sont apportées. Pour assurer ces fonctions, les eucaryotes ont notamment recours aux chaperons d'histones, des protéines interagissant avec les histones libres ou dans le nucléosome et capables de modifier la structure de la chromatine. Les chaperons d'histones lient à la fois les histones libres pour favoriser l'assemblage des nucléosomes et sont aussi capables de les désassembler pour favoriser l'accès à l'ADN. Parmi ces chaperons, le complexe FACT, très conservé chez les eucaryotes est connu pour son interaction avec la chromatine et son implication dans de nombreux processus cellulaires en faisant un acteur majeur dans l'organisation du génome.

I. Découverte du complexe FACT et conservation chez les eucaryotes

Le complexe FACT pour FACilitates Chromatin Transcription est composé de deux protéines, les chaperons d'histone Spt16 et Ssrp1 ou Pob3 et est particulièrement conservé chez les eucaryotes (Figure 1). FACT a été identifié ainsi que ses composants pour son rôle dans la transcription lui donnant son nom chez *Saccharomyces cerevisiae* (Orphanides et al., 1998, 1999).

La protéine Spt16 a été mise en évidence chez *S. cerevisiae* grâce à deux cribles génétiques pour identifier des mutants de la transcription et du cycle cellulaire. Dans le premier crible, les mutants de gènes *SPT* pour SuPpressor of Ty, inhibent les mutations dues aux insertions des éléments transposables Ty et delta en rétablissant la transcription des gènes adjacents (Clark-Adams et al., 1988; Malone et al., 1991). La mutation Spt16, est un suppresseur de l'insertion des éléments delta et par Northern Blot on montre qu'elle altère la transcription des gènes dans lesquels ces éléments sont insérés (Malone et al., 1991). Dans le deuxième crible, des cellules de *S. cerevisiae* dont la taille a augmenté sont isolées pour identifier des mutations thermosensibles du cycle cellulaire (Prendergast et al., 1990). La mutation thermosensible *CDC68* entraîne le blocage en phase G1 des cellules et empêche la progression dans le cycle cellulaire (Prendergast et al., 1990). Par Northern Blot, on montre que le défaut de progression dans le cycle cellulaire est dû à une diminution de la transcription des cyclines *CLN1*, 2 et 3 (Rowley et al., 1991). Par complémentation fonctionnelle et des analyses d'interaction génétique, il est montré que les mutations *CDC68* et *SPT16* correspondent au même gène *SPT16* (Rowley et al., 1991).

La protéine Pob3 est identifiée pour la première fois également chez *S. cerevisiae* à l'aide d'un crible permettant l'identification de nouveaux facteurs de la réplication (Wittmeyer and Formosa, 1997). Une chromatographie d'affinité est réalisée à partir de la sous-unité catalytique Pol1 de la DNA polymérase alpha. Les protéines Pob3 mais aussi Spt16 sont identifiées par ce crible et interagissent avec la Pol1 (Wittmeyer and Formosa, 1997). Des expériences d'immunoprécipitation de Spt16 la caractérisent en complexe exclusif avec Pob3 au sein du complexe FACT chez *S. cerevisiae* initialement nommé CP (Brewster et al., 1998).

Si les premières études du complexe FACT et de ses composants se font chez *S. cerevisiae*, ils sont présents chez de nombreux autres eucaryotes avec quelques variations. Des homologues de la protéine Spt16 sont retrouvés chez *H. sapiens*, *Xenopus laevis*, *A. thaliana*, etc. (Are et al., 2016; Formosa, 2012; Okuhara et al., 1999; Orphanides et al., 1998). La protéine Pob3 est l'homologue des protéines Ssrp1 qui composent avec Spt16 le complexe FACT dans la majorité des autres eucaryotes. Pob3 ne dispose pas du domaine HMG Box à l'extrémité C-terminale de Ssrp1 et est uniquement retrouvée chez les levures *S. cerevisiae* et *S. pombe* (Wittmeyer and Formosa, 1997). Dans ces espèces, la fonction du

domaine HMG box de Ssrp1 est assurée par une troisième protéine s'associant plus faiblement au complexe FACT, Nhp6 (Brewster et al., 2001).

II. FACT est un hétérodimère

1. Le chaperon d'histones Spt16, pivot du complexe FACT

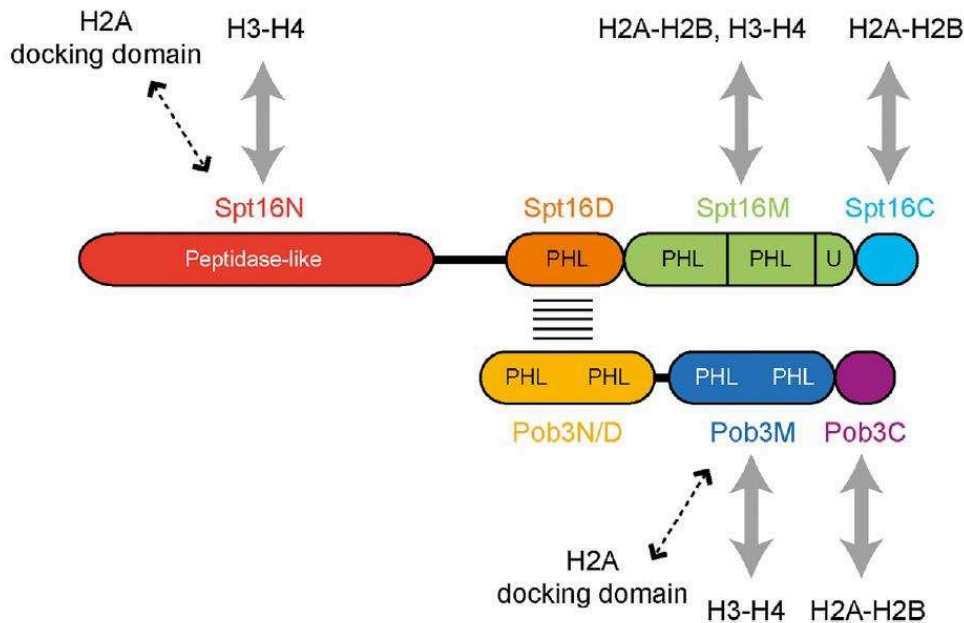


Figure 1 : Représentation schématique du complexe FACT et de ses domaines chez la levure

Le complexe FACT est composé des sous-unités Spt16 et Pob3/Ssrp1. Sont représentés les domaines Spt16N de type peptidase, Spt16D pour la dimérisation, Spt16M et Spt16C en C-terminal, les domaines Pob3N/D pour la dimérisation, Pob3M et Pob3C en C-terminal. Sont aussi montrés les interactions physiques (flèches pleines) ou génétiques (flèches en pointillés) entre les histones et le complexe. D'après (Jeronimo and Robert, 2016).

Spt16 est une grande protéine, d'environ 120kDa (Figure 1) indispensable à la croissance et à la survie des cellules (Formosa, 2012). Elle est constituée de quatre domaines :

- 1) En N-terminal, on retrouve un domaine en lobe ou à la conformation dite « pita-bread ». Sa séquence et sa structure, résolue par diffraction de rayons X sur le domaine N cristallisé, sont homologues aux familles d'aminopeptidases, des enzymes qui catalysent le clivage des acides aminés en N-terminal (Stuwe et al., 2008). Mais aucune activité peptidase ni aucun des résidus nécessaires à l'activité catalytique des aminopeptidases n'est retrouvé dans le domaine N de Spt16 (Stuwe et al., 2008; VanDemark et al., 2008). Le domaine N n'est pas

indispensable à l'activité de Spt16 mais son absence devient létale si elle est associée à un allèle mutant de Pob3. Le domaine pourrait avoir des fonctions redondantes avec Pob3 (O'Donnell et al., 2004). Des interactions du domaine N-terminal avec les histones H3-H4 ont été rapportées chez *S. pombe*, *H. sapiens* et chez les plantes par analyses d'interactions génétiques ou par résolution aux rayons X de la structure du domaine N cristallisé (Are et al., 2016; Stuwe et al., 2008; VanDemark et al., 2008). On détecte également des interactions génétiques entre le domaine N-terminal et le « docking domain » d'H2A, un domaine en C-terminal de l'histone H2A stabilisant l'interface entre H2A et H3-H4 et nécessaire pour l'organisation du nucléosome (Stuwe et al., 2008; VanDemark et al., 2008).

- 2) A la suite du domaine N, se trouve le domaine D qui assure la dimérisation de Spt16 avec Pob3 au sein du complexe FACT (Formosa, 2012). On y retrouve un domaine dit d'homologie à la pleckstrine (PHL), domaine retrouvé dans une grande variété de protéines eucaryotes comme des phospholipases, protéines transmembranaires, kinases ou encore protéines de signalisation cellulaire et capable de lier une grande variété de ligands dans la cellule. Ce domaine n'interagit pas avec les histones mais montre une interaction avec la sous-unité Pol1 de l'ADN polymérase alpha de *S. cerevisiae* (Hondele et al., 2013).
- 3) Le domaine M succède au domaine D dans Spt16. Il est composé de deux domaines PHL et d'un motif dit U-turn (Hondele et al., 2013; Jeronimo and Robert, 2016; Kemble et al., 2013). Le motif U-turn est un coude dans la protéine formant une poche hydrophobe permettant l'interaction avec les histones (Hondele et al., 2013). L'analyse aux rayons X de la structure cristallisée du domaine M de Spt16 de *S. cerevisiae* montre une interaction entre les histones H3-H4 et les domaines PHL (Kemble et al., 2013). L'analyse aux rayons X de la structure cristallisée du domaine M de Spt16 de *Chaetomium thermophilum* en complexe avec les histones H2A-H2B montre une interaction avec ces histones grâce au motif U-turn (Hondele et al., 2013). Le motif U-turn forme une poche hydrophobe complémentaire du motif hydrophobe de l'extrémité N-terminale de H2B dont les acides aminés nécessaires ont été identifiés par l'analyse de mutants (Hondele et al., 2013). Le domaine M est requis pour la fonction correcte de

Spt16 (Belotserkovskaya et al., 2003; Formosa, 2012; Schlesinger and Formosa, 2000).

- 4) La protéine Spt16 finit par un domaine en C-terminal riche en acides aminés acides et non structuré, une caractéristique commune aux chaperons d'histones (Formosa, 2012; Hoffmann and Neumann, 2015; Jeronimo and Robert, 2016; Kemble et al., 2015). La délétion du domaine C dans une construction protéique fusionnant Spt16 d'*H. sapiens* à la protéine FLAG empêche la formation de nucléosomes, la transcription et l'interaction avec les histones (Belotserkovskaya et al., 2003; Kemble et al., 2015; Schlesinger and Formosa, 2000). L'analyse aux rayons X de la structure cristallisée du domaine C et par titration calorimétrique isotherme (ITC) montre qu'il interagit fortement avec les histones H2A-H2B, grâce à des résidus aromatiques et acides composant le MBD pour Minimum Binding Domain, et essentiellement via H2B bien que des contacts soient détectés avec H2A (Kemble et al., 2015). La mutation de ces acides aminés chez Spt16 de *S. cerevisiae* empêche l'interaction avec H2A-H2B.

2. Ssrp1/Pob3, partenaire de Spt16

Pob3 ou Ssrp1 suivant les espèces est la deuxième sous-unité du complexe FACT (Figure 1). Chez la majorité des eucaryotes on la nomme Ssrp1 et comme Spt16 elle est essentielle à la viabilité des cellules. Pob3 est son équivalent chez certaines levures et champignons et son inactivation ne produit pas nécessairement de létalité comme chez *S. pombe* (Formosa, 2008). La différence entre les deux protéines provient aussi de l'absence chez Pob3 d'un motif HMGB présent chez Ssrp1 dont la fonction est remplacée par la protéine Nhp6 qui lie faiblement le complexe FACT et interagit avec l'ADN. Nhp6 n'est pas essentielle suggérant que la liaison à l'ADN du complexe FACT n'est pas une fonction essentielle (Formosa, 2012; Ruone et al., 2003). Pob3/Ssrp1 se compose en plus du domaine HMGB de trois domaines à l'organisation et à la structure similaires à Spt16 :

- 1) Le premier domaine en N-terminal de Pob3/SSrp1, N/D, est constitué de deux motifs PHL et assure la dimérisation avec Spt16 (Formosa, 2012).
- 2) Le domaine suivant M est homologue du domaine M de Spt16 et est aussi constitué de deux motifs PHL. Grâce à des analyses de co-immunoprécipitation,

on montre que le domaine M interagit avec la protéine RPA qui lie l'ADN simple brin pendant la réplication (VanDemark et al., 2006). L'analyse par rayons X de la structure cristallisée du domaine M de *S. cerevisiae* identifie une interaction avec les histones H3-H4 (Zunder et al., 2012). Des interactions génétiques du domaine M de Pob3 de *S. cerevisiae* avec le docking domain d'H2A ont été rapportées comme pour le domaine N de Spt16 (VanDemark et al., 2008). Des mutations dans le domaine M produisent des défauts de réplication et de transcription.

- 3) Comme Spt16, Pob3/Ssrp1 se termine en C-terminal par un domaine C riche en résidus acides. Par ITC et analyses de mutants chez *S. cerevisiae*, il a été montré que le domaine C de Pob3 est capable d'interagir avec les histones H2A-H2B via des résidus acides et aromatiques similaires à ceux de Spt16. Il est supposé que les domaines C de Spt16 et Pob3 lient chacun un dimère H2A-H2B dans le nucléosome (Hoffmann and Neumann, 2015; Kemble et al., 2015).
- 4) Ssrp1 à la différence de Pob3 contient en plus en C-terminal un motif HMGB (High Mobility Group Box) composé de trois hélices qui lie l'ADN. Ce motif absent de Pob3 est alors remplacé par la protéine Nhp6 de la famille HMGB qui interagit faiblement avec le complexe FACT. Nhp6 peut aussi se lier aux nucléosomes (Rhoades et al., 2004) et est requis pour une interaction correcte de FACT avec le nucléosome en liant l'ADN (McCullough et al., 2018).

3. Régulation du complexe FACT

Compte tenu de son abondance et des rôles multiples que le complexe endosse, sa quantité, son interaction avec ses substrats ou son activité sont régulées par un certain nombre de processus que je vais détailler.

3.1 Recrutement de FACT à des cibles spécifiques

Le complexe FACT peut être recruté à des cibles spécifiques (promoteurs, gènes, origines de réplication, etc) permettant de réguler son activité. La protéine Swi6 est un cofacteur de la transcription chez *S. cerevisiae* appartenant au complexe SBF régulant l'expression de cyclines pendant la phase G1 du cycle cellulaire. Elle lie directement FACT pour le recruter sur des promoteurs de gènes qu'elle régule et en favoriser la transcription. La protéine Whi5

est une protéine réprimant la transcription chez *S. cerevisiae* en phase G1. Elle inhibe le recrutement de FACT sur les mêmes promoteurs que ceux régulés par Swi6 (Takahata et al., 2009a). HP1c est une isoforme de la protéine à chromodomaines HP1 qui lie l'euchromatine chez *D. melanogaster*. Elle recrute FACT sur les gènes actifs et le lie aux formes actives de la polymérase II pour assurer leur transcription (Kwon et al., 2010). Le recrutement ou non de FACT sur des gènes particuliers permet d'en réguler la transcription et donc le niveau d'expression.

3.2 Des modifications post-traductionnelles des substrats de FACT modifient son activité

Des modifications post-traductionnelles des substrats du complexe FACT, les histones, peuvent influencer le recrutement du complexe FACT et son activité. Chez *S. cerevisiae*, l'acétylation des lysines des histones H4 régule la fonction de FACT dans la réplication et la transcription. Les phénotypes associés à des défauts dans Pob3 ou Spt16 sont supprimés par la mutation de déacétylases ou acétylases (Formosa et al., 2001, 2002; VanDemark et al., 2006). La méthylation des lysines de l'histone H3 est aussi capable de modifier l'activité de FACT. Les phénotypes causés par des mutations de Spt16 ou Pob3 chez *S. cerevisiae* sont supprimés par la délétion de la protéine Set2 méthylant H3K36. Ces phénotypes sont par contre aggravés par la délétion de la protéine Set1 méthylant H3K4 (Biswas et al., 2006). Il a été montré *in vitro* chez *S. cerevisiae* que l'ubiquitination de la lysine 120 de l'histone H2B favorise le réassemblage des nucléosomes pendant la transcription par l'ARN polymérase II (Fleming et al., 2008; Pavri et al., 2006). Chez *H. sapiens*, la phosphorylation du variant d'histone H2AX par la protéine kinase DNA-PK facilite son échange dans le nucléosome par FACT pendant la réparation (Heo et al., 2008).

3.3 Des modifications post-traductionnelles de FACT modulent son activité

Des modifications post-traductionnelles du complexe FACT peuvent réguler son activité. Chez *H. sapiens*, la protéine PARP1, Poly (ADP-ribose) polymérase 1, poly-ADP-ribosyle Spt16 réprimant l'échange du variant d'histone H2AX par le complexe FACT (Heo et al., 2008). Chez *S. cerevisiae*, Rtt101, sous-unité d'une E3 ubiquitine ligase, ubiquitine Spt16 permettant l'association de FACT avec l'hélicase MCM du réplisome aux origines de réplication (Han et al., 2010). Chez *D. melanogaster*, on a montré *in vitro* que Ssrp1 est phosphorylée

empêchant l'interaction du domaine HMGB avec l'ADN et l'ouverture de la chromatine par FACT (Tsunaka et al., 2009).

3.4 La quantité de FACT dans la cellule est régulée

La régulation du complexe FACT peut aussi se faire en contrôlant son abondance dans la cellule. Les estimations les plus récentes par Western Blot donnent 42000 dimères Spt16-Pob3 par cellule haploïde chez la levure pour 75000 nucléosomes, ce qui équivaut à un complexe FACT pour deux nucléosomes (McCullough et al., 2015). Chez *H. sapiens*, le ratio est d'un complexe FACT pour 300 nucléosomes (Formosa, 2008). Son abondance dans la cellule rend le complexe FACT disponible rapidement pour un grand nombre de processus. Pour autant, le ciblage de FACT au protéasome après ubiquitination de Spt16 par l'ubiquitine ligase San1 chez *S. cerevisiae*, assure un niveau optimal de FACT pour permettre la transcription (Sen et al., 2016).

III. Rôle de FACT dans la cellule, « FACilitates Chromatin Transcription to Transaction »

L'acronyme FACT signifie initialement FACilitates Chromatin Transcription et les premières études de la sous-unité Spt16 la lient à la transcription. Désormais l'acronyme de FACT ne saurait se limiter à transcription étant donné les nombreux autres rôles qui ont été identifiés. Les rôles détaillés dans la suite du paragraphe ont permis de passer de FACilitates transcription à transaction.

1. FACT assure la dynamique des nucléosomes : modèle d'interaction avec les histones

A la différence des enzymes remodelant la chromatine et dépendantes de l'ATP, le complexe FACT n'hydrolyse pas l'ATP et ne déplace pas les histones par rapport à l'ADN. FACT augmente l'accessibilité à l'ADN en réorganisant les nucléosomes grâce à ses contacts multiples avec les différentes histones. La description de l'organisation modulaire de Spt16 et Pob3/Ssrp1 a mis en évidence que leurs différents domaines interagissaient à de multiples reprises avec les deux dimères H2A-H2B et H3-H4. Les mécanismes expliquant comment

FACT réorganise le nucléosome sont encore mal compris mais des résultats suggèrent un modèle de respiration des nucléosomes permettant à FACT d'assurer des rôles dans la cellule qui peuvent sembler antagonistes (Formosa, 2012; Hondele et al., 2013; Jeronimo and Robert, 2016) (Figure 2).

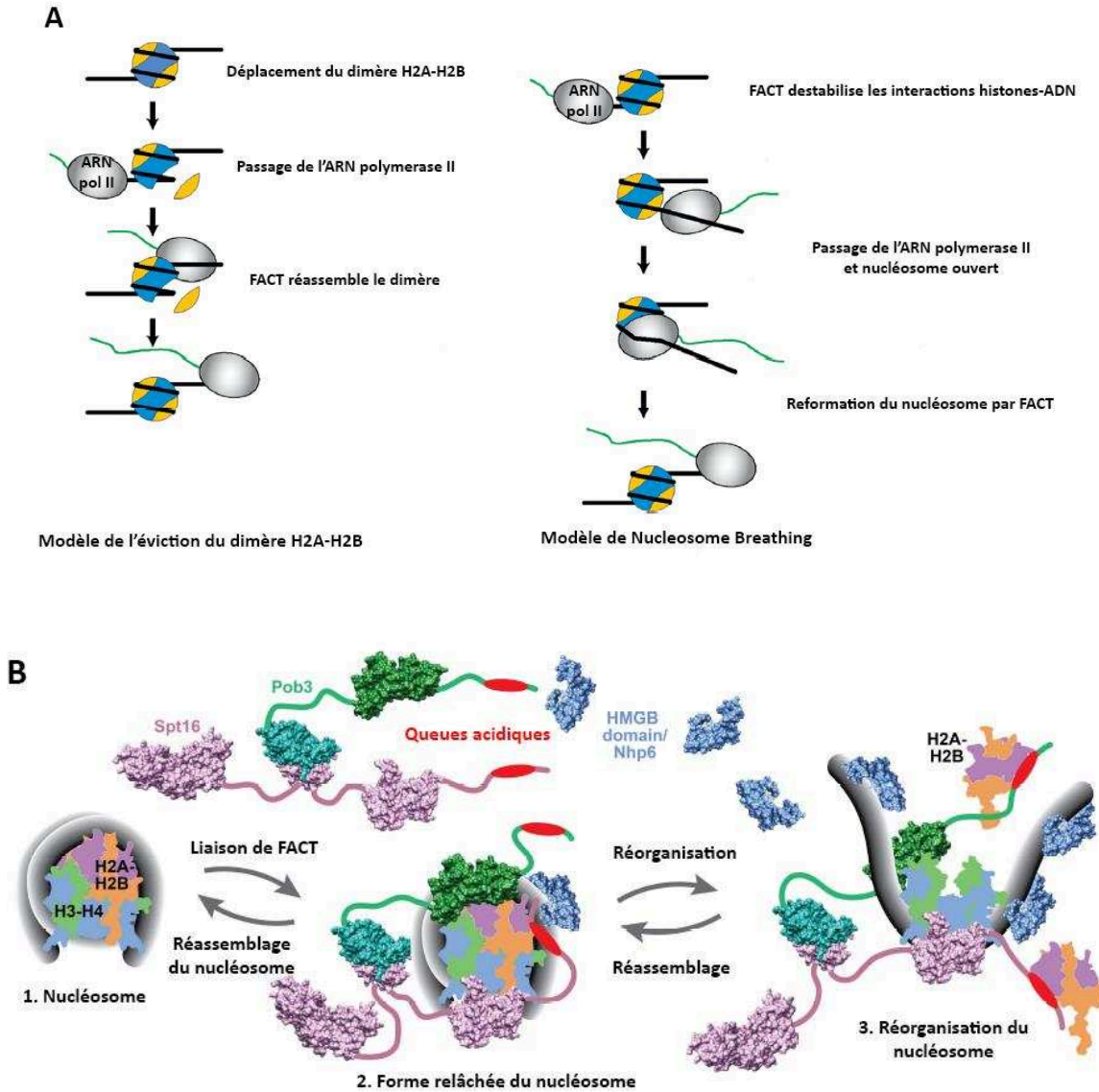


Figure 2 : Modèle d'interaction de FACT avec le nucléosome

A. Sont présentées les deux hypothèses expliquant l'interaction entre FACT et le nucléosome. A gauche dans le modèle d'éviction du dimère, l'interaction avec FACT provoque la perte du dimère H2A-H2B permettant le passage de l'ARN polymérase II puis nécessite le réassemblage du nucléosome en incorporant à nouveau le dimère H2A-H2B. A droite, le modèle de nucléosome breathing propose que l'interaction de FACT avec le nucléosome déstabilise les interactions ADN-Histones favorisant le relâchement du nucléosome et laissant le passage libre à l'ARN polymérase II. D'après (Jeronimo et al., 2016). **B.** Modèle détaillé du Nucleosome Breathing. Le complexe FACT prend en charge l'intégrité du nucléosome grâce notamment aux extrémités acides des protéines Spt16 et Pob3 qui interagissent avec le dimère H2A-H2B. Le nucléosome s'équilibre entre une forme plus fermée ou une forme plus ouverte permettant le passage des machineries, l'accès à l'ADN garanti par le domaine HMGB ou Nhp6, l'éviction et le remplacement d'histones et de variants d'histones. D'après (McCullough et al., 2018).

1.1 Modèle d'interaction avec le nucléosome : Hypothèse de l'éviction du dimère H2A-H2B

Le premier modèle proposé pour expliquer l'interaction et l'action du complexe FACT avec le nucléosome reposait sur l'hypothèse d'un déplacement du dimère H2A-H2B pour créer un hexasome. Plusieurs évidences suggéraient ce mécanisme qui s'accordait avec les rôles jusque-là identifiés pour FACT où la déstabilisation du nucléosome à la suite de la perte du dimère H2A-H2B faciliterait le passage de l'ARN polymérase II. Dans une première étape, FACT lierait le nucléosome grâce aux domaines C liant H2A-H2B. Dans une deuxième étape, FACT déplace le dimère créant un hexasome à partir de l'octamère d'histones et permettant le passage des machineries de réplication ou transcription par exemple. Enfin dans une troisième étape le dimère revient au nucléosome (Winkler et al., 2011) (Figure 2A). Plusieurs arguments supportent ce modèle. L'incubation prolongée de FACT d'*H. sapiens* avec des nucléosomes immobilisés provoque la perte de dimères H2A-H2B et le passage de l'ARN polymérase II entraîne une perte sensible de dimères H2A-H2B sans perte complète du nucléosome (Belotserkovskaya et al., 2003). L'immobilisation du dimère H2A-H2B au nucléosome chez *H. sapiens* inhibe l'action de FACT pendant la transcription (Orphanides et al., 1999). Il est également montré que FACT promeut l'éviction de nucléosomes nécessaire à l'assemblage des machineries de transcription en 5' des gènes qui doivent être induits rapidement comme des gènes du cycle cellulaire chez *S. cerevisiae* (Biswas et al., 2006; Ransom et al., 2009; Takahata et al., 2009a, 2009b). Enfin l'inactivation de FACT chez *S. cerevisiae* empêche la reformation des nucléosomes après le passage de la machinerie de transcription supposant la perte d'histones (Formosa et al., 2002).

1.2 Modèle d'interaction avec le nucléosome : Hypothèse du Global access model et nucleosome breathing

Pour autant, plusieurs observations sont en contradiction avec le modèle d'éviction et suggèrent un modèle dit Global Access ou Nucleosome Breathing. Au lieu d'une éviction du dimère H2A-H2B, ce modèle propose un relâchement du nucléosome, offrant plus d'accessibilité à l'ADN, grâce à la prise en charge du dimère H2A-H2B par le complexe FACT (Figure 2A et 2B). La structure dynamique s'équilibre entre un état fermé avec un nucléosome canonique et un état plus ouvert où le dimère H2A-H2B s'éloigne de l'interface

créée par H3-H4. Plusieurs arguments sont en faveur de ce modèle d'interaction de FACT avec le nucléosome. FACT altère les interactions histones/ADN sans supprimer les interactions entre histones au sein du nucléosome (Formosa, 2012; Hondele and Ladurner, 2013; Jeronimo and Robert, 2016). Des mutations altérant l'interface entre les dimères H2A-H2B et les tétramères H3-H4 favorisent la réorganisation du nucléosome et le rendent instable. Ces mutations suppriment génétiquement les défauts dans la transcription, associés à la perte du complexe FACT chez *S. cerevisiae*. Ces résultats montrent l'importance du maintien d'une conformation dynamique du nucléosome pour l'action de FACT (McCullough et al., 2011). Les modèles d'interaction de FACT avec les histones, décrits par analyse des structures aux rayons X des protéines Spt16 et Pob3 cristallisées de *C. thermophilum*, décrivent des domaines M et C interagissant avec les dimères H2A-H2B et tétramères H3-H4, suggérant une prise en charge globale du nucléosome (Hondele and Ladurner, 2013; Hondele et al., 2013). Chez *S. cerevisiae*, l'induction de gènes dépendant de FACT comme le promoteur Gal10 s'accompagne d'une plus grande accessibilité de l'ADN aux nucléases et enzymes de restriction sans pour autant s'accompagner d'une diminution de dimères H2A-H2B (Xin et al., 2009). Dans cette étude, les auteurs ajoutent que même si la perte du dimère H2A-H2B peut être induite par FACT, elle n'est pas forcément liée au niveau d'accessibilité de l'ADN. L'augmentation de l'accessibilité de l'ADN par le complexe FACT n'est pas due à la perte du dimère H2A-H2B. *In vitro*, lorsque des mutations sont introduites dans les histones H3 et H4 pour déstabiliser l'interface entre le dimère H2A-H2B et le tétramère H3-H4 dans un nucléosome reconstitué, FACT d'*H. sapiens* est toujours capable de transcrire le fragment d'ADN portant ce nucléosome. Lorsqu'ensuite, les histones sont réticulées au nucléosome par le DMS, sans que cela n'affecte la mobilité du nucléosome sur l'ADN, FACT d'*H. sapiens* est toujours capable de lever la pause de l'ARN polymérase II pendant la transcription. Ces résultats indiquent que la rupture de l'interface entre le dimère H2A-H2B et le tétramère H3-H4 n'est pas requise pour le fonctionnement de FACT pendant la transcription (Hsieh et al., 2013). Les études les plus récentes de la structure des domaines M et C de Pob3 et Spt16 de *S. cerevisiae* vont également dans le sens de ce modèle. Par titration calorimétrique isotherme, on a montré que Pob3 et Spt16 lient simultanément les dimères H2A-H2B par leurs extrémités C riches en résidus acides. L'analyse aux rayons X de la structure cristallisée du domaine C de Spt16 en interaction avec le dimère H2A-H2B et des résidus requis pour l'interaction suggère que les domaines C de Spt16 et Pob3 agissent

comme une pince, rentrant en compétition avec l'ADN. Les auteurs proposent un modèle dans lequel FACT maintient le nucléosome dans une conformation plus ouverte sans en altérer complètement la structure (Kemble et al., 2015) (Figure 2B et 3).

La liaison aux histones H2A-H2B est une activité essentielle du complexe FACT et nécessaire aux nombreux rôles qu'il assume dans la cellule. Le modèle du Global Access ou Nucleosome Breathing réconcilie les différents résultats obtenus jusqu'à présent et n'exclue pas non plus l'éviction du nucléosome (Figure 2B et 3) (Xin et al., 2009). Il est en accord avec la structure adoptée par les deux protéines ainsi que les nombreuses interactions entre les différentes parties du nucléosome et les domaines de FACT. Il permet de comprendre pourquoi FACT a des rôles parfois opposés stabilisant ou déstabilisant le nucléosome. L'essentiel des études de structure de FACT avec les histones a été réalisé via des domaines isolés. Il reste encore à bien comprendre comment se fait l'interaction globale du complexe avec le nucléosome.

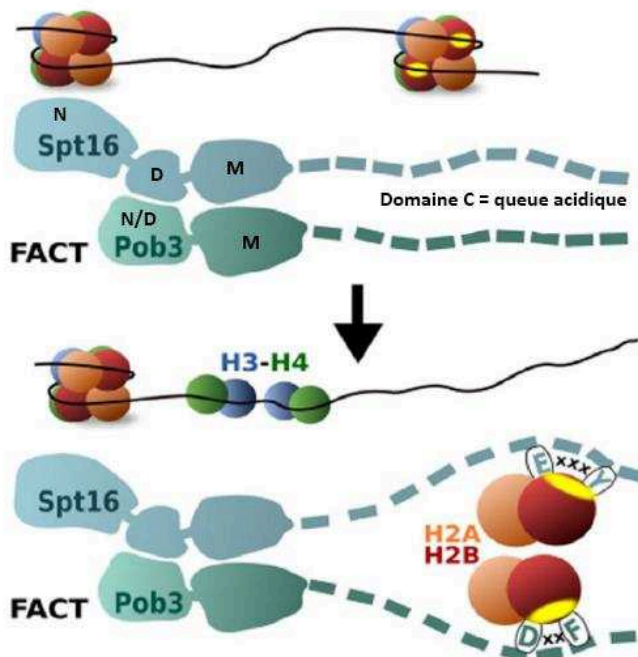


Figure 3 : Modèle d'interaction des extrémités C-terminales de Spt16 et Pob3 avec le nucléosome

Sont représentées les protéines Pob3 et Spt16 dans le complexe FACT interagissant avec le nucléosome. Les extrémités acides en C-terminal lient les dimères H2A-H2B sous forme de pince pour relâcher le nucléosome. Sont également représentés les acides aminés conservés nécessaires au Minimum Binding Domain pour l'interaction avec les histones. D'après (Kemble et al., 2015).

2. « **Facilitates Chromatin Transcription** »

C'est son rôle dans la transcription, initialement identifié, qui a valu à FACT son acronyme. Ainsi Spt16 fut d'abord associé à la transcription de gènes du cycle cellulaire et à des défauts dans la transcription suite à des insertions d'éléments transposables (Clark-Adams et al., 1988; Malone et al., 1991; Prendergast et al., 1990; Winston and Carlson, 1992). Mais c'est aussi pour son association avec l'ARN polymérase II que le complexe est identifié chez *H. sapiens* (Orphanides et al., 1998) et pour la régulation de l'expression de gènes chez *S. cerevisiae* (Brewster et al., 1998). FACT est requis pour la transcription d'un grand nombre de gènes et a des effets à la fois négatifs et positifs sur l'expression des gènes. Il fait le lien entre les étapes d'initiation et d'élongation et accompagne l'ARN polymérase II pour assurer une transcription fidèle.

2.1 *FACT garantit l'initiation correcte de la transcription*

Si FACT est essentiellement retrouvé sur les corps des gènes pendant l'élongation, il est aussi associé à l'initiation de la transcription (Mayer et al., 2010). La réorganisation de la chromatine par FACT peut à la fois faciliter l'accès de la machinerie de transcription aux sites d'initiation mais également empêcher que la polymérase démarre la transcription ailleurs qu'aux sites d'initiation prévus en maintenant une chromatine répressive (transcription cryptique) (Mason and Struhl, 2003) (Figure 4).

Il a ainsi été montré dans des mutants de SPT16 chez *S. cerevisiae* par des expériences de CHIP que le facteur général de la transcription TBP était moins présent sur les promoteurs de certains gènes. *In vitro*, on constate que FACT promeut la liaison de TBP sur des nucléosomes contenant un motif TATA. La sensibilité à la DNase I de cette séquence d'ADN est augmentée lorsqu'elle est incubée uniquement avec FACT mais diminue si on y ajoute en plus TBP qui vient se lier au motif. Ces résultats montrent que FACT désassemble la chromatine pour permettre aux facteurs de la transcription d'interagir avec les promoteurs (Biswas et al., 2005). Dans une autre étude, les auteurs se sont intéressés à la région HBR dans la queue de l'histone H2B. Par CHIP, ils observent que la délétion de cette région très conservée de H2B réduit le désassemblage de la chromatine dans les promoteurs de certains gènes et réduit la transcription. Par des expériences de pulldown et de mobilité sur gel, ils

observent que FACT lie cette région HBR et permet donc le désassemblage des nucléosomes aux promoteurs pour initier la transcription (Zheng et al., 2014).

FACT intervient également pour garantir le choix correct des sites d'initiation de la transcription et éviter les phénomènes de transcription cryptique. H2A.Z est un variant d'histone très conservé et retrouvé dans les promoteurs, les enhancers ou les centromères. Il est incorporé dans les promoteurs par la protéine SWR-C, une enzyme remodelant la chromatine et dépendante de l'ATP chez *S. cerevisiae*. Par des expériences de ChIP on observe que H2A.Z est par contre déplété des corps des gènes. Par ChIP on note que dans des mutants de Spt16, les corps de gènes sont déplétés en nucléosomes mais sont envahis par SWR-C attirée par les régions déplétées en nucléosomes. H2A.Z se retrouve donc dans les corps de gènes plutôt que dans les promoteurs. Par des expériences de gène rapporteur et par Northern Blot les auteurs montrent enfin que la mauvaise localisation de H2A.Z promeut la transcription cryptique en dehors des promoteurs corrects. On en conclue que FACT inhibe la perte de nucléosomes pendant la transcription et l'invasion par le variant H2A.Z empêchant ainsi l'utilisation de sites d'initiation aberrants de la transcription et des phénomènes de transcription cryptique (Jeronimo et al., 2015) (Figure 4).

2.2 FACT assure l'élongation de la transcription

Des expériences d'immunoprécipitation de la chromatine (ChIP) et d'immunofluorescence avec des anticorps contre FACT et l'ARN polymérase II chez *D. melanogaster* montrent que FACT et l'ARN polymérase II sont retrouvées conjointement sur les gènes exprimés (Saunders et al., 2003). Chez *S. cerevisiae* et *H. sapiens*, des expériences de ChIP-seq localisent FACT dans les corps des gènes transcrits par l'ARN polymérase II et en plus grande quantité pour des gènes plus exprimés (Feng et al., 2016a; Li et al., 2007; Mayer et al., 2010; True et al., 2016a). FACT promeut l'élongation lors de la transcription de plusieurs manières : en permettant à l'ARN polymérase II de repartir lorsqu'elle fait des pauses, mais aussi par le désassemblage des nucléosomes en amont de l'ARN polymérase II et enfin en réassemblant les nucléosomes en aval du passage de l'ARN polymérase II.

L'ARN polymérase II effectue deux pauses lors de la transcription, à 15 et 45 nucléotides en aval du site d'initiation de la transcription. *In vitro*, l'étude de la traversée de l'ARN polymérase II de *S. cerevisiae* et de *H. sapiens* sur des séquences avec des nucléosomes

reconstitués montre que la pause de la polymérase est réduite par la diminution des contacts entre le dimère H2A-H2B et l'ADN et entre le tétramère H3-H4 et l'ADN (Hsieh et al., 2010, 2013; Ujvári et al., 2008). FACT en prenant en charge le nucléosome via ses interactions avec le dimère H2A-H2B et en réduisant les contacts avec l'ADN, permet de dérouler l'ADN du nucléosome et à la polymérase II de poursuivre la transcription malgré ses pauses (Hsieh et al., 2013).

Le passage de l'ARN polymérase II nécessite de déstabiliser les nucléosomes c'est ce que fait le complexe FACT en interagissant avec le nucléosome et en maintenant la chromatine dans un état dynamique. Des expériences d'incubation de FACT d'*H. sapiens* avec des nucléosomes ont montré sur gel SDS PAGE une perte du dimère H2A-H2B qui s'intensifie lorsque l'on augmente la quantité de FACT dans la réaction. Les auteurs ont réalisé ensuite des expériences de transcription *in vitro* de séquences portant des mono-nucléosomes à des concentrations ioniques faibles défavorisant la déstabilisation de la chromatine et créant une barrière pour l'ARN polymérase II. L'ajout de FACT a permis, de manière similaire à une augmentation de concentration ionique, la transcription des séquences et le passage de l'ARN polymérase II. Les auteurs en ont conclu que FACT promouvait le désassemblage de la chromatine pour faciliter le passage de l'ARN polymérase II (Belotserkovskaya et al., 2003) (Figure 2A). *In vivo*, des expériences d'immunoprécipitation de la chromatine chez *S. cerevisiae* ont montré qu'au cours du temps après induction de la transcription, les histones H2B et H3 étaient moins présentes sur les corps de certains gènes et ce seulement lorsque FACT était bien présent dans la cellule (Xin et al., 2009).

Le complexe FACT contribue au réassemblage de la chromatine après le passage de l'ARN polymérase II. Il a été montré qu'*in vitro* l'incubation de FACT d'*H. sapiens* avec des histones permet leur dépôt sur de l'ADN (Belotserkovskaya et al., 2003) et qu'*in vivo* FACT est nécessaire à l'assemblage de nucléosomes (Xin et al., 2009). Chez *S. cerevisiae*, des expériences d'immunoprécipitation de la chromatine ont montré une perte des histones H2B et H3 sur les corps des gènes dans des mutants du complexe FACT (Jamai et al., 2009). La perte des histones est empêchée lorsque la transcription est inhibée par la thiotuline. Ces résultats montrent que FACT assure le réassemblage de la chromatine pendant la transcription. D'autres études montrent une augmentation d'histones solubles chez des mutants SPT16 qui est empêchée par l'inhibition de l'ARN polymérase II par l'alpha-

amanitine (Morillo-Huesca et al., 2010). FACT permet de restaurer des nucléosomes canoniques après le passage de la ARN polymérase en redéposant notamment les histones H3/H4 évincées par la polymérase II. Par analyse des profils de génomes traités par la Mnase, on observe que la déplétion de Spt16 entraîne des modifications dans la position des nucléosomes essentiellement dans le corps des gènes. Il s'avère que l'inhibition de la transcription par la thiotuline restaure le positionnement correct des nucléosomes malgré l'absence de Spt16. Le complexe FACT maintient une organisation stable des nucléosomes sur la chromatine compatible avec l'élongation de la transcription (Feng et al., 2016b). La quantité de FACT et sa stabilité doivent cependant être finement régulées pendant la transcription pour optimiser le désassemblage et le réassemblage des nucléosomes. Des auteurs ont en effet identifié une ubiquitine ligase San 1 qui cible Spt16 au protéasome et régule la quantité de FACT dans la cellule. Dans des cellules de *S. cerevisiae* délétées de San 1, par ChIP on observe que la quantité de Spt16 dans les corps des gènes augmente alors que dans le même temps on constate une diminution de la quantité d'ARN polymérase II et une diminution du dépôt de dimères H2A-H2B et de tétramères H3-H4. La multiplicité des rôles de FACT impose une régulation fine pendant la transcription (Sen et al., 2016).

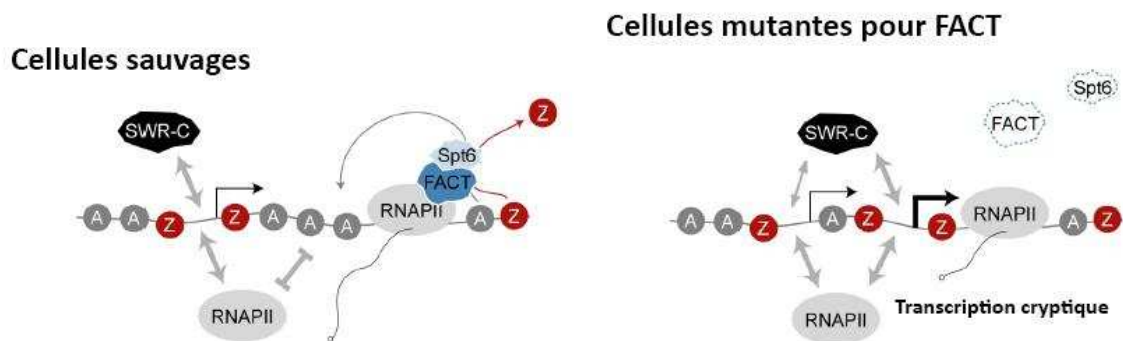


Figure 4 : Rôle de FACT dans l'inhibition de la transcription cryptique.

Dans des cellules sauvages pour FACT, le complexe accompagne l'ARN polymérase II pendant l'élongation, déroulant l'ADN et s'assurant que le variant d'histone H2AZ restreint aux promoteurs n'est pas déposé par SWR-C. Dans le cas d'un mutant du complexe FACT, le variant d'histone H2AZ envahit la région transcrite déposé par SWR-C et initie la transcription à d'autres endroits que le promoteur. D'après (Jeronimo et al., 2015).

2.3 L'effet de FACT sur la transcription est-il global ou ciblé sur certains gènes ?

Malgré le rôle essentiel du complexe FACT dans la transcription et la régulation de nombreux gènes, les changements observés dans l'expression des gènes dans des transcriptomes par micro-array ou RNAseq de mutants *SPT16* ou *POB3* restent généralement

modérés (Li et al., 2007; Pfab et al., 2018). On observe à la fois qu'un faible nombre de gènes voit son expression modifiée par l'absence de FACT mais les changements d'expression restent également minimales. Pour expliquer ces observations, on imagine que dans certains cas l'étude par déplétion plutôt que par délétion des gènes Spt16 et Pob3 sous-estime le nombre de gènes réellement touchés. Les composants de FACT étant essentiels pour les eucaryotes, il n'est pas forcément possible d'étudier les phénotypes d'organismes délétés pour les protéines de FACT. Les analyses du complexe sur la transcription par la polymérase II sont généralement réalisées par des expériences *in vitro* dans lesquels ne sont pas forcément présents l'ensemble des facteurs présents *in cellulo*. Il est alors possible que l'on surestime le rôle *in vivo* de FACT dans la transcription (Jeronimo and Robert, 2016). De plus FACT n'est pas le seul facteur promouvant la transcription. D'autres facteurs comme Spt5 ou Spt6 participent à l'élongation de la transcription par exemple (Mayer et al., 2010). On imagine qu'ils ont alors des fonctions redondantes. Il est encore difficile de comprendre si le rôle de Spt16 sur la transcription est un rôle global ou ciblé sur des gènes spécifiquement. L'effet global sur la transcription n'est pas incompatible avec le recrutement sur des gènes particuliers pour augmenter leur efficacité de transcription et réguler leur niveau d'expression. Malgré tout chez *S. cerevisiae*, des résultats récents montrent que l'absence de Spt16 dans la cellule réduit l'occupation de l'ARN polymérase II sur l'ensemble du génome et suggère un rôle global sur la transcription du complexe FACT en combinaison d'autres facteurs de transcription (Pathak et al., 2018).

3. « Facilitates replication »

FACT joue également un rôle important dans la réplication chez les eucaryotes. Comme pour la transcription, la chromatine est une barrière au passage de la machinerie de réplication et sa structure doit être relâchée pour faciliter le passage de l'ADN polymérase ce qui nécessite le désassemblage des nucléosomes. La chromatine doit aussi être doublée ce qui nécessite l'assemblage de nouveaux nucléosomes.

De nombreuses études ont montré l'implication de FACT dans la réplication, d'abord par l'étude des phénotypes de mutants de *SPT16* et *POB3* mais aussi par l'étude de leurs partenaires. L'absence de FACT dans plusieurs organismes eucaryotes provoque des défauts de réplication et de progression dans le cycle cellulaire (Abe et al., 2011; Okuhara et al.,

1999; Tan et al., 2006). Les mutants de FACT sont plus sensibles au stress réplicatif induit par l'hydroxyurée, une drogue qui bloque les fourches de réplication (Hoffmann and Neumann, 2015; O'Donnell et al., 2004; Schlesinger and Formosa, 2000). On a montré que les composants du complexe FACT interagissent physiquement avec la machinerie de réplication. Pob3 et Spt16 ont été initialement détectés par immunoprécipitation chez *S. cerevisiae* pour leur interaction avec la protéine Pol1 sous-unité de la DNA polymérase alpha (Wittmeyer and Formosa, 1997). Des expériences de purification d'affinité en tandem montrent que les deux sous-unités interagissent avec la DNA polymérase alpha via leurs domaines de dimérisation (Spt16D et Pob3N) (Hondele et al., 2013). Chez *S. cerevisiae*, FACT est co-purifiée par immunoprécipitation avec le complexe d'élongation de la réplication (Gambus et al., 2006) et avec la protéine RPA qui lie les brins d'ADN simple brins et la primase pendant la réplication (VanDemark et al., 2006). Enfin des expériences de CHIP et d'immunoprécipitation chez *H. sapiens* montrent que FACT interagit avec l'hélicase MCM nécessaire pour dérouler l'ADN lors de la réplication. FACT facilite l'activité de MCM *in vitro* et les deux protéines sont retrouvées par CHIP aux origines de réplication (Han et al., 2010; Tan et al., 2006) (Figure 5).

Si les mécanismes d'assemblage/désassemblage de la chromatine pendant la réplication par FACT sont encore peu connus, plusieurs constatations suggèrent que FACT promeut le désassemblage des nucléosomes et leur réassemblage sur l'ADN nouvellement répliqué. *In vitro*, FACT d'*H. sapiens* incubé avec des nucléosomes immobilisés favorise leur déstabilisation et la perte de dimères H2A-H2B et de tétramères H3-H4. Il favorise aussi *in vitro* le dépôt de nucléosome sur l'ADN (Belotserkovskaya et al., 2003; Yang et al., 2016). Le modèle d'interaction de FACT avec les nucléosomes par Nucleosome breathing est compatible avec l'idée que FACT faciliterait à la fois l'assemblage et le désassemblage de la chromatine pendant la réplication (Hondele and Ladurner, 2013; Xin et al., 2009). Le fait que FACT soit retrouvé aux origines de réplication et en lien avec la machinerie de réplication et qu'il facilite l'activité de l'hélicase MCM suggère qu'il facilite le passage de la machinerie et contrôle le désassemblage des nucléosomes (Han et al., 2010; Tan et al., 2006). Une autre étude montre des interactions génétiques et physiques par copurification par TAP-TAG entre FACT et des composants connus d'assemblage des nucléosomes pendant la réplication, les protéines Rtt106 et CAF-1. Ces interactions se produisent en particulier pendant la phase S.

Les auteurs montrent aussi que des mutants de ces facteurs et de Spt16 dans le domaine M interagissant avec H3-H4 sont particulièrement sensibles au stress réplcatif induit par l'hydroxyurée. Ces résultats suggèrent que FACT s'associe aux facteurs d'assemblage des nucléosomes à la réplication et permet le dépôt des nucléosomes sur l'ADN répliqué (Yang et al., 2016) (Figure 5). La progression de la fourche de réplication étant ralentie par l'absence de FACT et compte tenu de son interaction avec la machinerie de réplication, il est vraisemblable que FACT soit aussi nécessaire au désassemblage des nucléosomes (Abe et al., 2011).

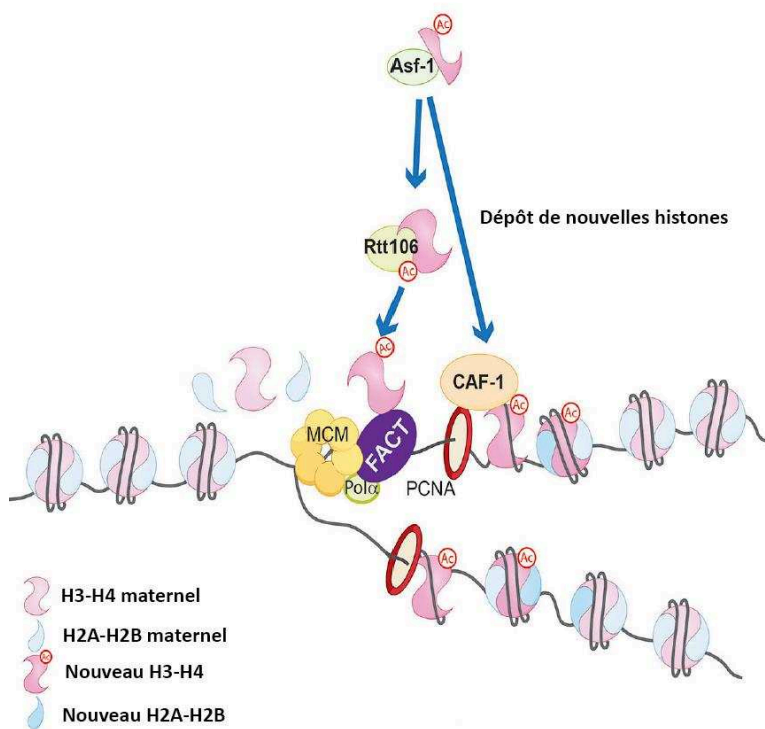


Figure 5 : Modèle d'action du complexe FACT pendant la réplication.

FACT interagit avec la machinerie de réplication et l'hélicase MCM pour faciliter le déroulement de l'ADN et la progression de la machinerie de réplication tout en désassemblant la chromatine. FACT collabore avec les chaperons d'histone CAF-1 et Rtt106 pour réassembler les nucléosomes. D'après (Yang et al., 2016).

4. « Facilitates repair »

Les rôles de FACT ne s'arrêtent pas à la transcription et à la réplication. Lors de la réparation suite à des dommages de l'ADN, il faut ainsi laisser passer la machinerie de réparation, la cibler et reconstituer par la suite la chromatine de manière analogue à la transcription et à la réplication. Lors de dommages par les UV chez *H. sapiens*, la protéine

p53 est activée par phosphorylation par la caséine kinase II pour ensuite inhiber le cycle cellulaire ou induire l'apoptose de la cellule. Par des expériences de co-immunoprécipitation, une étude a montré que la caséine kinase II interagit avec les protéines Spt16 et Ssrp1. L'absence de Spt16 et Ssrp1 empêche la caséine kinase II d'interagir spécifiquement avec P53 en cas de dommages à l'ADN (Keller and Lu, 2002). Dans ce cas, il ne s'agit pas d'une interaction de FACT avec le nucléosome mais d'une interaction avec d'autres protéines. Dans une autre étude, on observe que l'incorporation du variant d'histones H2A.X et sa phosphorylation par la kinase DNA-PK après une cassure double brin de l'ADN facilite la reconnaissance des sites de dommages par la machinerie de réparation. *In vitro*, FACT catalyse l'échange entre H2A et H2A.X dont la phosphorylation par la DNA-PK déstabilise la structure du nucléosome et facilite l'échange (Heo et al., 2008). Dans une autre étude on montre que Spt16 ciblé aux sites de dommages aux UV, facilite l'échange des histones H2A-H2B en suivant par microscopie à fluorescence des histones fusionnées à la GFP. Spt16 promeut le redémarrage de la transcription aux sites des dommages (Dinant et al., 2013). FACT recrute RNF20, l'E3 ubiquitine ligase responsable de l'ubiquitination de H2B lors de cassures double brins et facilite le relâchement du nucléosome pour l'accès de RNF20 et de la machinerie de réparation. La modification d'H2B permet le recrutement des protéines utiles à la réparation par recombinaison homologue (Oliveira et al., 2014). Une étude plus récente propose un rôle pour Ssrp1 indépendamment de Spt16 dans la réparation des cassures simple brins. Ssrp1 est recruté aux dommages et facilite la décondensation de la chromatine et l'échange de H2B (Gao et al., 2017).

5. FACT participe à l'intégrité des centromères

Les centromères chez les eucaryotes sont des portions de chromatine spécialisées flanquées d'hétérochromatine, incorporant un variant de l'histone H3, Cenp-A et requises pour l'assemblage du kinétochore. L'assemblage du kinétochore est requis à chaque division cellulaire pour permettre la ségrégation correcte des chromosomes. De nombreuses autres protéines interagissent avec le kinétochore et la chromatine pour assurer la formation et le fonctionnement correct du centromère et des résultats suggèrent que FACT en fait partie. L'absence de Pob3 chez *S. pombe* n'affecte pas la viabilité des cellules mais il est requis pour la maintenance de l'hétérochromatine péricentromérique et la ségrégation des

chromosomes indépendamment des voies du RNAi (Lejeune et al., 2007). Les sous-unités de FACT sont également capables d'interagir avec Cenp-A, le variant d'histone H3 conférant l'identité du centromère (Lejeune et al., 2007). L'assemblage du kinétochore au centromère est permis par l'incorporation de deux variants d'histones qui ne sont pas conservés au cours des divisions cellulaires, CENP-T et CENP-W. L'utilisation d'inhibiteurs de la réplication (Hydroxyurée) et de l'ARN polymérase II ont montré que le dépôt de CENP-T/-W ne dépend ni de la synthèse d'ADN ni de la transcription. L'analyse en spectrométrie de masse des protéines interagissant avec CENP-W après immunoprécipitation d'une version GFP-CENP-W, révèle la présence de FACT. *In vitro*, Spt16 interagit avec les variants CENP-T/-W et l'étude montre *in vivo* par immunofluorescence que FACT promeuvent l'échange d'histones CENP-T/-W et leur déposition au centromère (Prendergast et al., 2016).

En conclusion, Le complexe FACT joue des rôles multiples dans la cellule. Son abondance lui permet d'interagir fréquemment avec les nucléosomes et de réguler l'accès à l'ADN des nombreuses machineries de transcription, réplication, réparation etc. Le modèle d'interaction de FACT avec le nucléosome, dit Nucleosome Breathing permet d'expliquer les fonctions parfois contradictoires du complexe visant à la fois à assurer l'accès à l'ADN mais aussi à conserver l'intégrité de la chromatine. Il est cependant encore difficile de comprendre comment le complexe dans son ensemble interagit avec le nucléosome. Beaucoup des études réalisées se concentrent sur des domaines spécifiques de Spt16 ou Pob3 et essentiellement chez *H. sapiens* ou *S. cerevisiae* bien que le complexe FACT soit bien conservé chez les eucaryotes. Enfin, il est encore difficile de comprendre le rôle de FACT à l'échelle du génome et les analyses de transcriptome en l'absence des protéines Spt16, Pob3 et Ssrp1 peinent à montrer un effet global. Des études récentes identifiant de nouveaux rôles pour FACT suggèrent que le complexe intervient dans la cellule au-delà de ses fonctions dans la transcription ou la réplication.

Chapitre 2 : Méthylation des adénines

Le maintien de l'intégrité du génome ne relève pas uniquement de la régulation de la chromatine mais repose également sur le niveau de régulation supplémentaire apporté par les modifications des bases de l'ADN. Il s'agit le plus fréquemment de méthylation des adénines et des cytosines impliquées dans une très grande variété de processus. La méthylation des adénines essentiellement étudiées chez les procaryotes bénéficie de l'amélioration des techniques d'analyses qui permettent de l'étudier chez les eucaryotes.

I. Méthylation et autres modifications de l'ADN

1. Au-delà des bases canoniques de l'ADN

Des modifications des quatre bases canoniques de l'ADN sont retrouvées dans les trois grands domaines du vivant (eucaryotes, bactéries et archées). Ces modifications jouent un rôle dans une grande variété de processus régulant l'expression des gènes, protégeant contre les invasions du génome par des éléments transposables, recrutant des protéines interagissant et/ou modifiant l'ADN ou la chromatine, discriminant des séquences d'ADN de provenance différente (ADN nouvellement répliqué, ADN d'agent pathogène), signalant des dommages de l'ADN, etc. Sans altérer la séquence d'ADN et à un niveau de résolution supérieur à celui du nucléosome, elles sont une source importante d'information supplémentaire pour le génome.

Les modifications les plus fréquemment retrouvées sur l'ADN sont les ajouts de groupements méthyle sur des carbones C, azotes N ou oxygènes O des différentes bases de l'ADN (Iyer et al., 2011). Le numéro accompagnant la base modifiée indique la position de l'atome dans le cycle aromatique de la base portant le groupement méthyl. Les méthylations peuvent être produites suite à des lésions sur l'ADN et ne sont alors pas déposées par des enzymes : (N1-méthyladenine 1mA, N3-méthyladenine 3mA, N7-méthyladenine 7mA, N3-méthylcytosine 3mC, N2-méthylguanane 2mG, O6-méthylguanane 6mG, N7_méthylguanane 7mG, N3-méthylthymine 3mT, O4-méthylthymine 4mT). Ces méthylations sont induites par des composés endogènes ou de l'environnement comme des drogues, ou des produits du

métabolisme. Ces modifications doivent être réparées car elles provoquent des mutations et ont des effets toxiques pour la cellule (Drabløs et al., 2004). Les méthylations peuvent être directement catalysées par des méthyltransférases : N6-méthyladenine 6mA, N4-méthylcytosine 4mC, et C5-méthylcytosine 5mC (Figure 6). Ces trois dernières méthylations sont les plus largement observées dans le monde vivant, elles ont de plus la capacité de s'insérer dans la double hélice d'ADN sans en modifier la structure ni l'appariement des bases tout en étant facilement accessibles pour des protéines interagissant avec elles (Jurkowska and Jeltsch, 2016). D'autres modifications de l'ADN ont été décrites : comme des déaminations des cytosines par l'enzyme AID qui produit un uracile à la place de la cytosine, des oxydations comme la C5-hydroxyméthylcytosine (5hmC) issu de 5mC, ou bien la base J issue de la thymine et retrouvée chez certains parasites (Breiling and Lyko, 2015; Grosjean, 2013; Iyer et al., 2011).

5mC est majoritairement étudiée chez les organismes eucaryotes et en particulier chez les mammifères en proportions importantes chez qui elle régule l'expression des gènes, le développement, l'expression des éléments transposables ou encore inactive le chromosome X (Breiling and Lyko, 2015; Edwards et al., 2017). On la retrouve alors au sein de motifs CpG. 5mC est aussi largement étudiée chez les champignons ou chez les plantes chez qui elle est déposée sur des motifs plus variés (Law and Jacobsen, 2010). Elle n'est par contre pas détectée chez des organismes comme *S. cerevisiae*, *S. Pombe* et *C. elegans* et les proportions infimes chez *D. melanogaster* ne permettent pas d'affirmer qu'elle est présente chez cette dernière espèce (Capuano et al., 2014; Proffitt et al., 1984). Elle est bien entendue retrouvée chez un grand nombre de bactéries où elle participe au système de restriction modification qui sera détaillé un peu plus loin, permettant de protéger l'hôte contre l'invasion d'un pathogène. 5mC régule aussi l'expression des gènes ou encore la virulence chez *E. coli* ou *H. pylori* (Chen et al., 2016; Kahramanoglou et al., 2012; Kumar et al., 2012; Militello et al., 2014; Sánchez-Romero et al., 2015). L'absence des méthyltransférases catalysant le dépôt de 5mC chez les procaryotes produit des défauts d'expression de gènes liés à la réponse au stress ou encore à la virulence chez des pathogènes (Sánchez-Romero et al., 2015).

La méthylation de l'ADN 4mC est beaucoup moins décrite que 5mC car étudiée essentiellement chez quelques bactéries thermophiles (Janulaitis et al., 1983; Morgan et al., 2016). Elle est associée au système de restriction modification et participe chez *Helicobacter*

pylori à la régulation de gènes de virulence (Kumar et al., 2018). Chez les eucaryotes, 4mC ne semble pas détectée hormis par séquençage par SMRT sequencing détaillée juste après chez les espèces dont on cherchait à détecter la méthylation des adénines (Ye et al., 2017). Cependant aucune autre méthode indépendante n'a confirmé la présence de cette modification chez les eucaryotes. Il est possible qu'il s'agisse d'un artefact lié au SMRT sequencing.

La répartition et les proportions de méthylation des cytosines et adénines varient grandement au sein des espèces, elles peuvent être observées simultanément chez certaines espèces quand chez d'autres on ne détecte que l'une de ces modifications voire aucune jusqu'à présent (Figure 7). La présence de méthyltransférases et déméthylases spécifiques d'une modification ou d'une autre est aussi l'indice de leur présence mais leur activité enzymatique n'a pas forcément été démontrée. Nous nous concentrerons sur la méthylation des adénines, les enzymes la catalysant dans les génomes procaryotes et eucaryotes et ses rôles chez les différentes espèces lorsqu'il a été trouvé.

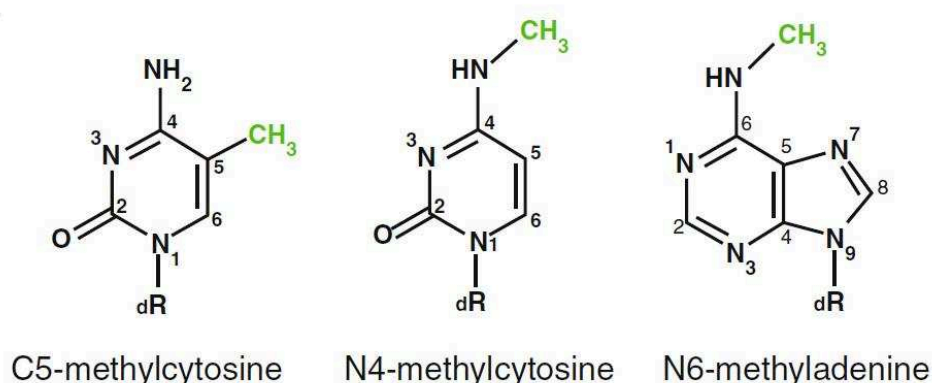


Figure 6 : Méthylation sur les bases de l'ADN

Sont représentées les formules chimiques des trois principales méthylations retrouvées dans le monde vivant. D'après (Breiling and Lyko, 2015).

2. Méthodes de détection et d'analyse de la méthylation de l'ADN

C'est au milieu du XX^e siècle que sont découverts les premiers nucléosides ne correspondant pas aux bases canoniques déjà connues, A, T, G et C (Dunn and Smith, 1958; Hotchkiss, 1948). Les méthodes de détection et d'analyse des nucléosides modifiés ont grandement avancé depuis les chromatographies sur papier des années 50. Elles bénéficient

aujourd'hui des techniques de séquençage les plus récentes permettant de localiser précisément sur le génome les modifications.

2.1 Détecter et quantifier la méthylation

La chromatographie liquide à haute performance couplée à la spectrométrie de masse permet de détecter sans ambiguïté les modifications de l'ADN et de les quantifier si nécessaire dans l'échantillon étudié (Sulewska et al., 2007). Elle nécessite cependant de disposer de molécules standards pour chaque modification recherchée dont on connaît la masse et le temps de rétention spécifique. Les échantillons d'ADN préalablement digérés en nucléosides sont séparés par chromatographie liquide par affinité avec une colonne apolaire. Les nucléosides séparés sont ensuite analysés en spectrométrie de masse qui assigne à chaque molécule une masse et un temps de rétention. Par comparaison avec les signaux produits par les nucléosides standards on peut ainsi assigner aux molécules détectées le nucléoside modifié correspondant. Il est ainsi théoriquement possible de détecter n'importe quel nucléoside modifié. Il peut malgré tout arriver que des nucléosides modifiés présentent une masse identique comme dans le cas de méthylation sur des positions différentes d'une même base. Dans ce cas, il n'est pas toujours possible de séparer en chromatographie les deux molécules et il faudra démontrer par une autre méthode que c'est bien la modification portée par un atome spécifique que l'on souhaite quantifier (Capuano et al., 2014; Huang et al., 2015).

Il est possible de détecter des nucléosides modifiés à l'aide d'anticorps spécifiques. L'anticorps peut être utilisé en Dot-Blot pour détecter la présence ou l'absence d'une modification dans des échantillons d'ADN. Par immunofluorescence, l'anticorps spécifique permet de localiser dans la cellule et de suivre la dynamique au cours du temps de la modification de l'ADN. Ces techniques ont l'avantage de donner un résultat rapide mais ont l'inconvénient de ne pas être quantitatives. Elle est aussi limitée par l'existence d'un anticorps spécifique pour chaque modification. Certains anticorps peuvent aussi confondre la détection de modifications similaires comme l'anticorps commercial anti-6mA qui reconnaît également 1mA (Greer et al., 2015). Il est nécessaire de confirmer par une méthode indépendante la présence de la modification cherchée. Enfin comme pour la

chromatographie liquide, la détection par un anticorps peut être biaisée par une contamination bactérienne qu'il est impossible de discriminer.

Certaines enzymes de restriction ont la propriété de digérer l'ADN en fonction du statut de méthylation d'un nucléotide et permettent donc de détecter la présence des modifications sur des sites identifiés au préalable (Fu et al., 2015; Sulewska et al., 2007). Si l'on prend l'exemple de la méthylation des adénines, l'enzyme DpnI digère des sites 5'-G6mATC-3' méthylés quand l'enzyme DpnII digère des sites 5'-GATC-3' non méthylés. Cette méthode ne permet pas de quantifier la modification recherchée comme les précédentes évoquées. Elle nécessite de connaître au préalable les sites du génome portant la modification recherchée et dépend de la présence des sites digérés par les enzymes dans leur génome.

2.2 Localiser la méthylation dans le génome

L'amélioration importante des techniques de séquençage à haut débit a permis de localiser avec une précision toujours plus fine la méthylation dans le génome. A l'aide d'anticorps spécifiques de la modification étudiée, on peut immunoprécipiter les fragments d'ADN portant cette modification (Hsu et al., 2014). L'immunoprécipitation ou MedIP est particulièrement utilisée pour la méthylation des cytosines et adénines. Ce n'est pas une méthode quantitative et elle n'est pas non plus exhaustive. Elle a de plus l'inconvénient de n'offrir qu'une résolution de l'ordre de 200pb environ en fonction de la taille des fragments d'ADN immunoprécipités. Elle peut permettre l'identification de motifs mais elle reste trop peu précise. Elle est cependant intéressante pour établir des profils de méthylation du génome dans des conditions différentes.

Différentes méthodes de séquençage sont accessibles pour quantifier et localiser plus précisément que par immunoprécipitation les modifications de l'ADN, jusqu'à une résolution au nucléotide près. Par exemple, la conversion des cytosines méthylées en uridine par traitement au sodium bisulfite permet d'analyser la présence de 5mC par séquençage haut débit (Patterson et al., 2011). On peut améliorer la technique en enrichissant en séquences méthylées par une immunoprécipitation préalable ou à l'aide d'enzymes de restriction sensibles ou non à la méthylation (Plongthongkum et al., 2014). Cette méthode n'est pas disponible pour la méthylation des adénines.

Ce sont les techniques de séquençage par molécule unique qui ont permis de localiser dans le génome avec une plus grande précision des modifications comme la méthylation des adénines. Le SMRT (Single Molecule Real Time) sequencing analyse la cinétique d'incorporation des bases par la DNA polymérase et permet de détecter avec une résolution au nucléotide près théoriquement un grand nombre de modifications (6mA, 4mC, 5mC, 5hmC, 8oxoG, etc) (Beaulaurier et al., 2015; Fang et al., 2012; Flusberg et al., 2010). L'ADN polymérase incorpore des nucléotides marqués avec un fluorophore ce qui permet de suivre en temps réel la cinétique d'incorporation de chaque nucléotide complémentaire au brin d'ADN servant de modèle (Eid et al., 2009). Le principal avantage de cette méthode est de permettre la détection de la fluorescence en même temps que l'on détermine la séquence au contraire des autres méthodes de séquençage. Elle ne nécessite pas d'amplification préalable. L'incorporation d'un nucléotide est détectée par un pic de fluorescence dont la couleur identifie le nucléotide. La durée entre deux pics de fluorescence est aussi variable en fonction du nucléotide incorporé. L'incorporation de bases en face de bases méthylées sur la matrice modifient la cinétique d'incorporation des nucléotides. On peut ainsi déterminer les nucléotides modifiés précisément, identifier des motifs et quantifier la présence de la modification dans le génome. Elle a cependant l'inconvénient de ne pas être aussi efficace pour toutes les modifications (Flusberg et al., 2010). Elle l'est particulièrement pour 6mA mais beaucoup moins pour 5mC dont la couverture du génome doit être beaucoup plus grande pour détecter un signal (x25 vs x250). Au préalable une conversion par l'enzyme Tet pour l'oxyder en 5-carboxylcytosine (5-caC) réduit la couverture requise pour détecter 5mC. Cependant la conversion par l'enzyme Tet peut aussi dans certains contextes cibler 4mC et conduire à de faux positifs. La technique n'est pas non plus facilement utilisable pour des génomes de grande taille car elle requiert une couverture importante des molécules séquencées et peut vite devenir peu abordable (Eid et al., 2009; Flusberg et al., 2010). Elle est cependant très adaptée à l'étude de petits génomes comme les génomes bactériens ou de petits eucaryotes. La technologie nanopore est également capable de détecter la présence de modifications de l'ADN qui émettent des signaux différents des nucléosides non modifiés. Elle est surtout utilisable pour la détection de 5mC sans nécessiter de conversion préalable des cytosines méthylées, mais elle est aussi capable de détecter la méthylation des adénines avec cependant moins d'efficacité que 5mC (McIntyre et al., 2017; Simpson et al., 2017). Le séquençage nanopore mesure les changements de courant induits par le passage

d'une molécule d'ADN à travers le nanopore lui-même traversé par un courant ionique. La présence d'une méthylation modifie le courant produit. On peut ainsi déterminer au nucléotide près la position d'une méthylation. La technique est moins couteuse que le SMRT sequencing et nécessite moins d'ADN. Cependant les modifications autres que 5mC et 6mA ne peuvent être détectées et le séquençage nanopore ne parvient pas à détecter l'ensemble des méthylations (90% des 5mC et 80% des 6mA).

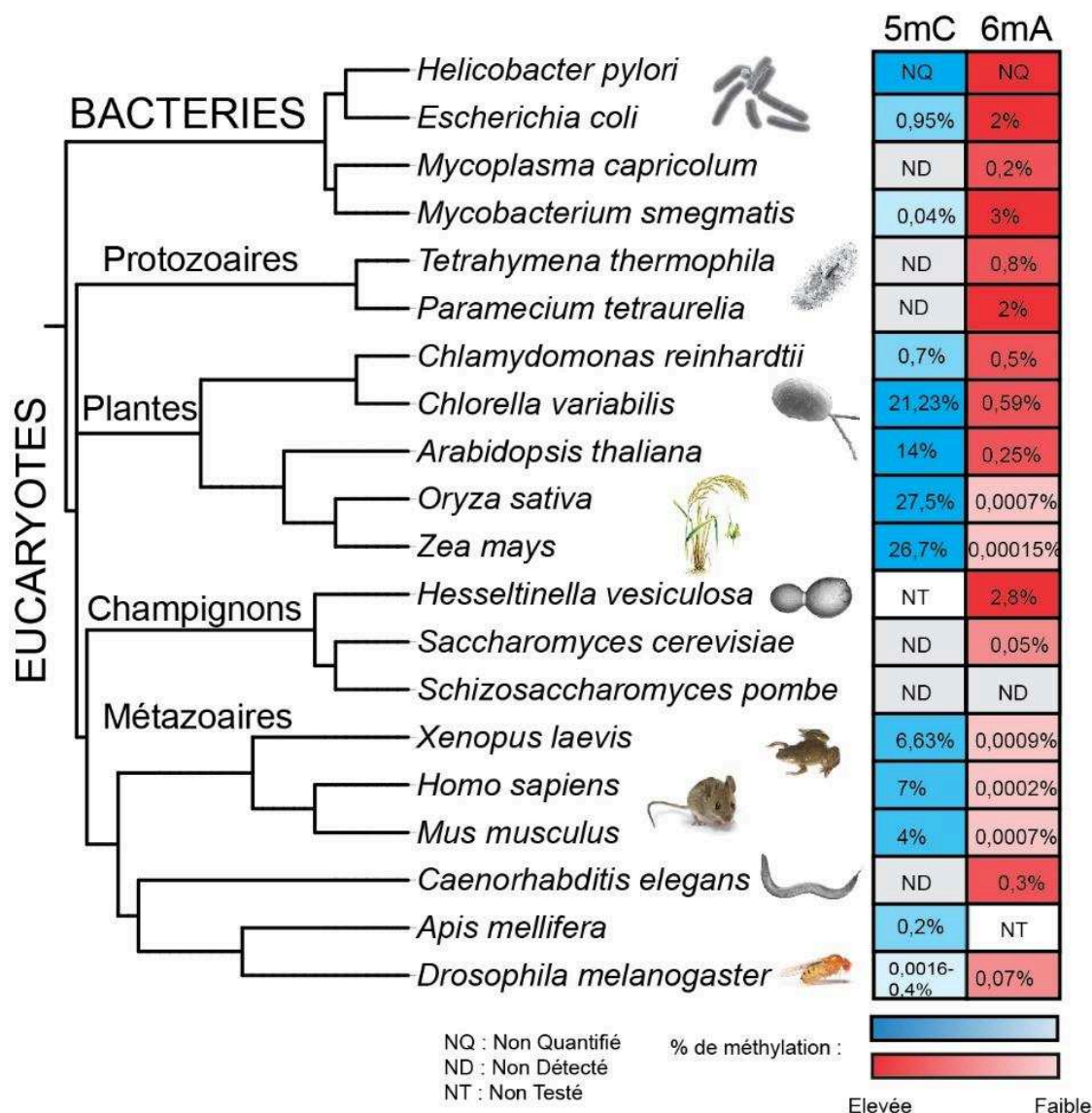


Figure 7 : Présence de méthylation des cytosines ou des adénines dans le monde vivant

Sont représentées plusieurs espèces pour lesquelles ont été recherchées la présence de 5mC ou 6mA dans le génome. Ces espèces sont représentatives des différents grands domaines du monde vivant. Pour chaque espèce les pourcentages s'ont calculés à partir d'expériences d'HPLC-MS ou par séquençage à haut débit comme détaillé dans les paragraphes suivants. Plus les méthylations sont abondantes plus les couleurs sont saturées. Pour la bactérie *Helicobacter pilori*, les pourcentages ne sont pas calculés mais le séquençage par SMRT sequencing atteste de grosses quantités de 5mC et 6mA dans son génome. NQ signifie que la méthylation n'a pas été quantifiée. ND signifie que malgré les analyses la méthylation n'a pas été détectée. NT signifie que la méthylation n'a pas été recherchée. Références : (Liang et al., 2018a, 2018b; Mondo et al., 2017; O'Brown and Greer, 2016; Wang et al., 2017)

II. Méthylation des adénines

La méthylation des adénines (6mA) est la deuxième modification la plus abondamment retrouvée dans le monde vivant après la méthylation des cytosines. Elle a été détectée pour la première fois en 1955 chez *E. coli* (Dunn and Smith, 1958). Elle a été très étudiée chez les procaryotes ainsi que les enzymes catalysant son dépôt sur l'ADN. Sa présence chez les eucaryotes semblait anecdotique. Elle est peu étudiée et on ne connaît que peu d'adénines méthyltransférases eucaryotes. Mais sa détection récente chez de nouvelles espèces grâce aux techniques de séquençage haut débit et les rôles associés suggèrent qu'il s'agirait d'une nouvelle marque épigénétique.

1. 6-méthyladénine chez les procaryotes

1.1 Système de restriction modification et méthyltransférases solitaires chez les procaryotes

C'est chez les bactéries que la méthylation des adénines est identifiée pour la première fois (Dunn and Smith, 1958). Chez *E. coli*, on estime par SMRT sequencing que 6mA est retrouvée dans des proportions similaires à 5mC, environ 2% (Fang et al., 2012) (Figure 7). L'analyse de la méthylation par SMRT sequencing facilite la détection de 6mA chez les bactéries et détermine son abondance dans le génome, les sites de méthylation et le brin méthylé (Beaulaurier et al., 2015). 6mA est ainsi retrouvé dans un grand nombre de bactéries, et dans toutes les branches phylogénétiques (Ratel et al., 2006). Les abondances de 6mA dans les génomes procaryotes varient grandement, de 0,2% chez *Mycoplasma capricolum* à 3% chez *Mycobacteria smegmatis* déterminées par HPLC (O'Brown and Greer, 2016) (Figure 7). Des adénines méthyltransférases sont retrouvées chez des virus comme des bactériophages ou des virus d'archées (Ratel et al., 2006). Par exemple, des expériences de complémentation des mutants Dam- d'*E. coli* incapables de méthyler les adénines, ont montré que la méthyltransferase M du virus d'archée ϕ Ch1 restaure la méthylation chez *E. coli* sur l'adénine. La présence de 6mA a été attestée en parallèle chez ce virus par HPLC (Baranyi et al., 2000).

Chez les procaryotes, la méthylation des adénines est impliquée dans le système de restriction modification comme les autres modifications des bases (5mC, 4mC), protégeant le génome de l'hôte contre les invasions d'ADN étranger (Casadesús and Low, 2006). Chaque système est composé d'une enzyme de restriction associée à une méthyltransferase de l'ADN. Dans la majorité des cas, l'ADN bactérien méthylé est protégé de la coupure par sa propre enzyme de restriction qui est inhibé par la méthylation (Loenen et al., 2014; Wion and Casadesús, 2006). Ce mécanisme est originellement décrit comme un système immunitaire primitif protégeant l'ADN de l'hôte contre l'invasion par des phages ou autres ADN invasifs. Des résultats suggèrent aussi qu'il permettrait de réguler l'expression de gènes dont les niveaux d'expressions sont modifiés par la délétion des systèmes de restriction modification (Sánchez-Romero et al., 2015). Rien que chez *E. coli*, on trouve 9 adénines méthyltransferases. Parmi elles, existent des méthyltransferases solitaires sans enzymes de restriction associées capables de méthyliser les adénines. Dam chez les gammaprotéobactéries et archées méthyle le motif GATC. CcRM chez les alphaprotéobactéries méthyle le motif GANTC avec une préférence pour l'ADN hémi-méthylé (Ratel et al., 2006). L'absence de Dam et CcRM provoquent de nombreux phénotypes, les souches d'*E. coli* dam⁻ sont viables mais Dam est un gène essentiel chez *Vibrio cholerae* par exemple (Wion and Casadesús, 2006). Les souches déficientes pour Dam sont plus sensibles aux dommages de l'ADN, plus promptes à la recombinaison et aux mutations. De la même manière, CcRM est essentielle chez certaines bactéries. Dam hautement processive assure la maintenance de la méthylation au cours de la réplication après le passage de la fourche de réplication qui laisse les sites hémi-méthylés (Wion and Casadesús, 2006).

1.2 6mA, une marque épigénétique chez les bactéries

Grâce aux méthyltransferases solitaires, la méthylation des adénines n'est pas restreinte au système de restriction modification chez les bactéries (Figure 8). Elle joue ainsi un rôle important dans la réplication et la réparation par mismatch repair. Le nouveau brin d'ADN répliqué est transitoirement non-méthylé. Les sites GATC hémi-méthylés sont reconnus par la protéine SeqA inhibant la ré-initiation de la réplication aux origines de réplication. Les sites GATC complètement méthylés aux origines de réplication sont reconnus par DNA A qui

séparent les brins et initie la réplication (Lu et al., 1994; Ratel et al., 2006). 6mA permet aussi de discriminer le brin parental du brin naissant lors de la réparation de l'ADN par Mismatch repair (Ratel et al., 2006). 6mA est impliquée dans le contrôle de la transcription. Certains sites de méthylation sont positionnés dans des promoteurs ou des régions régulatrices et modulent les interactions protéines-ADN comme avec la RNA polymérase bactérienne ou des facteurs de transcription. CcRM et Dam régulent l'expression de nombreux gènes de virulence et la méthylation des adénines contrôle la pathogénicité chez *E. coli*, *H. pylori*, *V. cholerae*, etc. Elles contrôlent l'expression de gènes requis pour la conjugaison ou le cycle cellulaire ou l'expression de transposases régulant la transposition (Casadesús and Low, 2006; O'Brown and Greer, 2016; Sánchez-Romero et al., 2015; Wion and Casadesús, 2006).

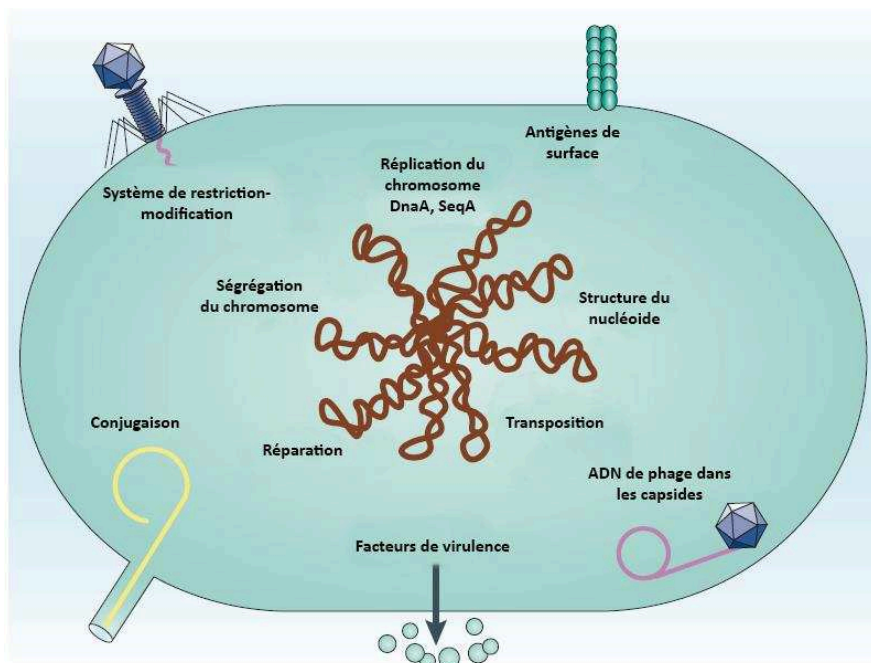


Figure 8 : Rôle de la méthylation des adénines chez les bactéries

La méthylation des adénines chez les bactéries a de nombreux rôles représentés ici et détaillés dans les paragraphes précédents. Elle permet notamment d'assurer le système de restriction modification, la conjugaison, la réplication, la réparation, la ségrégation du chromosome, la transposition, la structure du nucléole, l'expression d'antigènes de surface ou de facteurs de virulence ou encore l'encapsidation d'ADN de phages. D'après (Wion and Casadesús, 2006).

1.3 Adénines méthyltransférases

La majorité des méthyltransférases de l'ADN utilisent le coenzyme AdoMet ou SAM (S-adenosyl méthionine) comme donneur de méthyl et ont une structure et une organisation communes. Les méthyltransférases de 6mA et 4mC méthylant sur le même atome, un azote,

partagent une organisation similaire différente de celles de 5mC méthyltransférases. Des expériences de méthylation *in vitro* ont d'ailleurs montré que les méthyltransférases bactériennes pour 6mA étaient aussi capables de méthyler 4mC et inversement (Jeltsch et al., 1999). L'étude aux rayons X de la structure de l'adénine méthyltransferase M.TaqI cristallisée de *Thermus aquaticus*, a montré que le domaine catalytique est composé de 6 feuillets beta en parallèle et d'un septième feuillet inséré de manière antiparallèle entre le cinquième et le sixième feuillet (Jurkowska and Jeltsch, 2016; Labahn et al., 1994; Malone et al., 1995). Le domaine est divisé en deux sous-domaines reliés par deux hélices alpha : un contenant le site de liaison à AdoMet et l'autre pour la liaison à l'adénine à méthyler (Jurkowska and Jeltsch, 2016). L'azote 6 N6 de l'adénine se positionne en liaison avec les résidus (DNS)PP(YFW). Ce domaine catalytique est retrouvé aussi bien dans des adénines méthyltransférases de l'ADN que de l'ARN et chez les bactéries comme chez leurs homologues eucaryotes.

Des analyses phylogénétiques des adénines méthyltransférases de l'ADN et de l'ARN retrouvés chez les bactéries et leurs homologues eucaryotes ont mis en évidence que l'ensemble des adénines méthyltransférases proviennent probablement des adénines méthyltransférases d'ARN ribosomiaux universellement conservées (Jurkowska and Jeltsch, 2016). La plupart des méthyltransférases procaryotes sont associées au système de restriction modification et ont été acquises par des transferts latéraux entre des espèces éloignées. Les méthyltransférases solitaires Dam chez les gamma protéobactéries et CcrM chez les alpha protéobactéries dérivent d'opérons de système de restriction modification (Jurkowska and Jeltsch, 2016; Sánchez-Romero et al., 2015). Elles ont été acquises à partir de méthyltransférases du système de restriction modification et ont perdu l'enzyme de restriction correspondante pour réguler des fonctions cellulaires comme la réplication, la réparation, la régulation de l'expression de gènes chez les bactéries.

On retrouve par homologie avec les enzymes procaryotes chez les eucaryotes de nombreuses adénines méthyltransférases de l'ADN et de l'ARN dont on n'a pas forcément les évidences d'une activité catalytique ni si cette activité cible les ADN ou les ARN. Les analyses phylogénétiques suggèrent que ces enzymes ont été acquises à 13 reprises par les eucaryotes et leur virus à partir des procaryotes et de leur virus (Iyer et al., 2015). Une grande partie des adénine méthyltransférases putatives eucaryotes appartiennent à la

famille MT-A70 du nom de la sous-unité MT-A70/METTTL3 de 70kDa de l'ARN méthyltransferase d'*H. sapiens* ajoutant 6mA sur les pré-ARNm. Elles ont évolué à partir des systèmes de restriction modification bactériens M.MbolIA et M.MunI (Iyer et al., 2015). Elles contiennent notamment les ARN méthyltransférases d'*H. sapiens* METTL3 et 14 ou Ime4 de *S. cerevisiae*. Plusieurs des méthyltransférases putatives eucaryotes de cette famille contiennent en plus du domaine catalytique MT-A70, des domaines Zinc-finger ou PHD supposés interagir avec la chromatine. Les eucaryotes ne possédant pas de système de restriction modification, les méthyltransférases ont vraisemblablement évoluées pour remplir de nouvelles fonctions chez les eucaryotes comme la méthylation des adénines sur les ARN. La seule ADN méthyltransferase eucaryote appartenant à la famille MT-A70 dont l'activité a été démontrée *in vivo* est Damt-1 de *C. elegans* (Greer et al., 2015). L'absence de Damt-1 chez *C. elegans* entraîne une diminution de 6mA attestée par Dot-Blot à l'aide d'un anticorps spécifique. Le nombre important de méthyltransférases putatives, leur diversité et la détection récente de 6mA chez un certain nombre d'eucaryotes, que je détaillerai par la suite, laissent imaginer que la méthylation des adénines pourrait être plus largement répandue chez les eucaryotes.

2. 6-méthyladénine chez les eucaryotes

2.1 La méthylation des adénines est-elle présente chez les eucaryotes pluricellulaires ?

Même si la méthylation des adénines a été découverte peu de temps après celle des cytosines, on a longtemps considéré qu'elle était restreinte aux procaryotes et présente de manière anecdotique chez quelques eucaryotes unicellulaires. L'amélioration récente des techniques de séquençage a permis de détecter à des niveaux plus faibles la méthylation des adénines dans des eucaryotes plus variés. Pour autant les faibles niveaux détectés sont parfois proches des limites de détection des techniques utilisées et les analyses fonctionnelles peinent encore à associer un rôle à la présence de 6mA chez les eucaryotes (Figure 7).

2.1.1 L'amélioration des techniques de détection identifie 6mA chez des eucaryotes pluricellulaires

A partir des études menées jusqu'à maintenant, *C. elegans*, eucaryote pluricellulaire, ne semble pas présenter de méthylation des cytosines dans son génome (5mC) (Simpson et al., 2017). Ainsi, si certaines analyses par HPLC-MS semblent détecter un peu de 5mC, les résultats ne sont pas reproduits et aucun homologue des méthyltransférases de 5mC Dnmt1 ou 3 n'est retrouvé (Capuano et al., 2014) (Figure 7). L'analyse de différents génomes compilés issus de plusieurs stades du cycle de vie de *C. elegans*, par SMRT sequencing, a permis de quantifier et positionner 6mA : 0,3% des adénines réparties sur le génome sans enrichissement particulier dans un type de séquences sur deux motifs AGAA et GAGG (Greer et al., 2015). Par immunofluorescence avec un anticorps spécifique de 6mA, la méthylation est détectée dans les noyaux de manière ubiquitaire à l'exception du sperme chez des nématodes sauvages. 6mA n'est par contre pas détectée en Dot-Blot avec le même anticorps dans l'ADN de vers sauvages, elle n'est visible qu'après 5 à 15 générations de nématodes mutants pour la déméthylase d'H3K4me2, spr-5. Ces résultats suggèrent que la détection par Dot-Blot serait moins sensible que par Immunofluorescence. Par UHPLC-MS, les auteurs confirment la présence de 6mA bien que les proportions retrouvées soient plus variables : de 0,01% à 0,4% de 6mA/A dans des vers sauvages. L'abondance de 6mA retrouvée par SMRT sequencing ne reflète pas les variations qui doivent sans doute dépendre du stade de vie de *C. elegans* et amène à relativiser le pourcentage de 0,3% retrouvé. Les auteurs identifient par homologie une méthyltransferase putative Damt-1 appartenant à la famille MT-A70 et dont le site catalytique contient les acides aminés DPPW caractéristiques des adénine méthyltransférases (Iyer et al., 2015). Ils identifient aussi une déméthylase putative NMAD-1 de la famille des protéines ALKB. Par détection de 6mA par Dot-Blot ou UHPLC-MS, les déplétions de ces protéines produisent respectivement une diminution et une augmentation de la quantité de 6mA dans le génome. Etant donné que les auteurs détectent plus facilement 6mA et une augmentation de la méthylation dans un contexte mutant pour la déméthylase spr-5, ils proposent un lien entre la méthylation des adénines et la méthylation H3K4me2. Observant de plus que le phénotype sur la méthylation s'intensifie avec les générations du mutant spr-5 en même temps que le phénotype de stérilité lié à la mutation spr-5, les auteurs suggèrent que les modifications d'histone H3K4me2 et 6mA participent à

la transmission de l'information épigénétique de manière transgénérationnelle. Malgré tout la distribution très uniforme de la méthylation dans le génome de *C. elegans*, sa faible abondance dans des nématodes sauvages et la nécessité d'être dans un contexte mutant pour observer une quantité facilement détectable de 6mA, amènent à relativiser sur l'importance de la méthylation chez *C. elegans* (Greer et al., 2015).

Chez *Drosophila melanogaster*, l'existence de la méthylation de l'ADN est particulièrement controversée (Figure 7). La plupart des analyses échouent à détecter 5mC ou à des niveaux très bas en HPLC-MS proche des limites de détection et avec une grande variabilité (Capuano et al., 2014). La drosophile ne dispose que d'un homologue de la méthyltransferase Dnmt2 qui ne semble pas avoir d'activité catalytique. L'analyse récente par UHPLC-MS/MS du génome de *D. melanogaster*, échouant à détecter la méthylation des cytosines, détecte un relativement faible pourcentage d'adénines méthylées 0,07% (Zhang et al., 2015). Ce pourcentage varie grandement durant l'embryogénèse. Une déméthylase Dmad, homologue de la protéine Tet responsable de la déméthylation de 5mC chez les mammifères (Shi et al., 2017), est exprimée au moment où le niveau de 6mA est le plus bas. Dmad est essentiel pour le développement embryonnaire. *In vitro*, les auteurs mesurent le niveau de déméthylation 6mA sur de l'ADN double brin de thymus de veau par UHPLC-MS après incubation avec un extrait nucléaire d'embryon de *D. melanogaster* contenant ou non Dmad. Ils détectent une diminution de la déméthylation de 6mA lorsque Dmad est déplété par un anticorps ou des ARN double brin des embryons ou en utilisant des extraits nucléaires de *D. melanogaster* mutantes pour *DMAD*. Par immunoprécipitation de l'ADN méthylé dans les ovaires de drosophile avec un anticorps spécifique de mouches sauvages ou mutantes pour *DMAD*, un peu moins de la moitié des adénines méthylées est enrichie sur les transposons par rapport au reste du génome. Cette fraction augmente dans un mutant Dmad dans lequel on détecte en plus une augmentation globale de la méthylation. Cette augmentation de 6mA s'accompagne de l'augmentation de l'expression des transposons mesurée par RT-qPCR amenant les auteurs à suggérer que la déméthylation des transposons par Dmad soit corrélée à leur expression (Zhang et al., 2015). Les auteurs utilisent comme contrôle de l'immunoprécipitation, l'ADN immunoprécipité par un anticorps IgG et observent un enrichissement de 6mA qui s'accroît encore plus dans le contexte mutant pour *DMAD*. Etant donnée la faible quantité de 6mA détectée à l'origine, il aurait été intéressant

de réaliser comme contrôle l'immunoprécipitation du même ADN amplifié par PCR et donc dépourvu de modifications de l'ADN avec le même anticorps anti-6mA. Cela aurait permis de s'assurer de la spécificité de l'anticorps anti-6mA. Enfin il est à noter que comme pour *C. elegans*, les auteurs doivent se placer dans un contexte mutant pour observer une quantité plus facilement mesurable de 6mA amenant à relativiser son intérêt pour *D. melanogaster* dans un contexte physiologique.

Le séquençage par SMRT sequencing a également détecté la présence d'adénines méthylées chez *Arabidopsis thaliana* (Liang et al., 2018a) (Figure 7). La proportion augmente au cours du développement et varie suivant les organes passant de 0.05% à 0.25% d'adénines méthylées. Elles sont distribuées le long du génome, enrichies dans les régions péri-centromériques et uniquement en intergénique. 6mA est positivement corrélée avec l'expression des gènes et le passage d'un état végétatif à reproductif de la plante suggérant qu'elle régule l'expression des gènes liés au développement. La présence de 6mA est également confirmée par un anticorps spécifique et par HPLC-MS. Cependant aucune méthyltransferase n'a pu être identifiée pour le moment.

Chez d'autres eucaryotes pluricellulaires comme le Xénope, le moustique ou d'autres plantes, les quantités de 6mA détectées sont très faibles et proches des limites de détection. Les quelques études recherchant des adénines méthylées concluent à des taux relativement faibles, 0,0009% chez le Xénope (Koziol et al., 2016), 0,00015 à 0,0007% chez *Oryza sativa* et *Zea mays* (Huang et al., 2015). Dans le blé, des proportions plus importantes sont retrouvées dans l'ADN mitochondrial. On observe également une petite quantité de 6mA chez le moustique *Aedes albopictus* (Adams et al., 1979) (Figure 7).

2.1.2 L'ADN des mammifères est-il méthylé sur les adénines ?

Seules quelques études semblent détecter de la méthylation des adénines chez les mammifères mais dans des proportions extrêmement faibles en comparaison des autres organismes (Figure 7). Récemment, par HPLC-MS, l'abondance en 6mA a été mesurée dans différents types cellulaires chez *H. sapiens* et le rat (Huang et al., 2015). Chez l'humain, on détecte environ 0,0002% d'adénines méthylées et chez le rat de 0,00006 à 0,0003%. Des taux similaires sont retrouvés dans des cellules embryonnaires de souris par SMRT-ChIP sequencing en enrichissant en séquences contenant le variant H2AX. Les auteurs observant

un enrichissement relatif de 6mA sur les jeunes transposons L1, proposent que 6mA réprime transcriptionnellement les éléments L1 (Wu et al., 2016). Ces études sont cependant à prendre avec précaution, aucune confirmation par une autre méthode de la présence de 6mA n'a été faite et les taux détectés restent très faibles, proches des limites de détection des techniques utilisés. Il est aussi difficile d'exclure avec des taux si faibles la présence d'un symbiote ou d'une contamination. Il semble plus vraisemblable que la méthylation des adénines soit absente des mammifères (O'Brown and Greer, 2016).

2.2 La méthylation des adénines chez les eucaryotes unicellulaires

2.2.1 Répartition et rôle

Chez les eucaryotes unicellulaires, la présence de 6mA est attestée depuis plus longtemps. 6mA y est retrouvée dans des proportions plus importantes comme chez les algues vertes *Chlorella* et *Chlamydomonas reinhardtii*, les ciliés *Paramecium tetraurelia*, *Tetrahymena thermophila* et *Oxytricha*, le parasite de la malaria *Plasmodium falciparum* ou certains champignons (Figure 7).

La présence de 6mA et 5mC dans le génome de *C. reinhardtii* était connue depuis les années 70 (Hattman et al., 1978), avoisinant les 0,3 à 0,5% d'adénines méthylées. 1/3 des sites méthylés sont des motifs CATG et GATC. 6mA est enrichi autour du TSS de gènes transcriptionnellement actifs et localisée principalement dans les linkers d'ADN entre les nucléosomes. Elle occupe des positions clairement différentes de celles de 5mC présent dans les corps des gènes. Il est proposé que 6mA participe au positionnement des nucléosomes autour des TSS et facilite l'initiation de la transcription quand 5mC à l'opposé facilite l'élongation ou inhibe l'initiation de la transcription (Fu et al., 2015).

L'analyse des génomes de plusieurs champignons par SMRT sequencing et HPLC-MS révèle des abondances et des répartitions très variables suivant les espèces (Mondo et al., 2017). Les espèces du sous-règne *Dikarya* étudiées ont des niveaux faibles de 6mA (0,048% à 0,21%) qui sont difficilement retrouvés par HPLC-MS ou 6mA Immunoprécipitation. Il est possible que dans des abondances aussi faibles de 6mA, soit le SMRT sequencing détecte trop de faux positifs soit la résolution et la sensibilité de l'IP et de l'HPLC-MS sont insuffisantes. Ce constat amène à relativiser aussi les taux encore plus faibles de méthylation

retrouvés dans d'autres eucaryotes comme la *Drosophile* ou chez les mammifères. Il amène aussi à se poser la question des contrôles à réaliser d'abord dans les immunoprécipitations mais aussi dans les expériences de SMRT sequencing. Il serait très utile de séquencer dans les deux cas le même ADN que celui utilisé pour l'expérience mais amplifié par PCR pour effacer les méthylations de l'ADN soit par SMRT sequencing soit après immunoprécipitation par l'anticorps anti 6mA. Dans le cas de l'immunoprécipitation comme je l'indiquais plus haut pour l'étude chez *D. melanogaster*, cela permettrait de vérifier la spécificité de l'anticorps et la réalité des pics d'enrichissement de 6mA détectés. Dans le cas du SMRT sequencing, ce contrôle distinguerait la présence réelle de 6mA des faux positifs. Parmi les autres sous-règnes de champignons rassemblés sous l'appellation « early diverging fungi », 6mA est plus abondant jusqu'à 2,8% chez *Hesseltinella vesiculosa*, dans un contexte ApT, avec le motif TAT en particulier et méthylés symétriquement (Mondo et al., 2017). Les sites sont rassemblés dans des clusters positionnés sur les promoteurs. La méthylation est associée avec des gènes actifs et positionnée dans les zones d'éviction des nucléosomes supposant d'une part qu'elle facilite la transcription et d'autre part pourrait réguler le positionnement des nucléosomes sur les promoteurs pour faciliter l'accès de la machinerie de transcription (Mondo et al., 2017). Très récemment, par HPLC-MS/MS et SMRT sequencing, 6mA a aussi été détectée chez *S. cerevisiae* à hauteur de 0.05% d'adénines méthylées. Elle semble dynamique pendant le cycle cellulaire et enrichie dans les corps de gènes et promoteurs avec une concentration au niveau du TSS (Liang et al., 2018b).

2.2.2 Les ciliés, un modèle d'étude de la méthylation des adénines

6mA est identifié simultanément dans les génomes des ciliés *Paramecium* et *Tetrahymena* dans les années 70 (Cummings et al., 1974; Gorovsky et al., 1973). On détecte 0,8% d'adénines méthylées pour *Tetrahymena* et environ 2% pour la Paramécie. Comme tous les ciliés ils ont la particularité de présenter deux types de noyaux germinaux et somatiques au sein d'un même cytoplasme. Le noyau germinal est transcriptionnellement inactif et diploïde, tandis que le noyau somatique est responsable de l'expression des gènes et hautement polyplôïde (Betermier and Duharcourt, 2014). Le dimorphisme nucléaire de la paramécie et sa physiologie seront détaillés dans le chapitre 3 de cette introduction. Dans les expériences citées, le pourcentage d'adénines méthylées est quantifié par chromatographie sur de l'ADN extrait après fractionnement cellulaire pour séparer les deux

types de noyaux et les mitochondries après incorporation de nucléosides radiomarqués 6mA. Elle est majoritairement retrouvée dans le noyau somatique et absente de l'ADN mitochondrial. Le noyau germinal de *Paramecium* contient aussi une proportion importante de 6mA mais il est possible qu'il ait été contaminé par des bactéries ou le noyau somatique hautement polyploïde (Cummings et al., 1974). On ne peut exclure également une contamination bactérienne lors de l'extraction des noyaux somatique, les ciliés étant cultivés dans un milieu enrichi en bactéries elles-mêmes méthylées. Seuls quelques travaux ultérieurs suggèrent que 6mA pourrait être impliquée dans les réarrangements du génome programmés au cours du développement chez *Paramecium* (Coyne et al., 2012; Nakayama and Endoh, 2003) (voir Chapitre 3). Chez *Tetrahymena*, à l'aide d'enzymes de restriction, plusieurs sites méthylés sur la séquence GATC positionnés dans les linkers d'ADN entre les nucléosomes ont été identifiés (Capowski et al., 1989; Harrison and Karrer, 1989; Harrison et al., 1986; Karrer and VanNuland, 2002). En immunofluorescence et en Dot-Blot à l'aide d'un anticorps spécifique pour 6mA, le signal de la méthylation est détecté uniquement dans le noyau somatique végétatif et s'établit progressivement pendant le développement du nouveau noyau somatique. Il reste absent du noyau germinal. Plus récemment, l'analyse du génome de *Tetrahymena* par SMRT sequencing a confirmé l'ensemble de ces observations bien que le pourcentage d'adénines méthylées soit un peu plus faible (0,6%) que ce que les chromatographies montraient. 6mA est préférentiellement localisé sur des sites 5'-AT-3' dans les linkers entre les nucléosomes. Elle est dans les corps de gènes transcrits par la RNA polymérase II et en aval des sites d'initiation de la transcription (Wang et al., 2017). Aucune enzyme n'a pour le moment été identifiée ni de rôle précis associé. Il a malgré tout été observé que le niveau de méthylation des gènes n'était pas corrélé à leur niveau d'expression mesuré par RNAseq suggérant que 6mA ne soit pas lié à la régulation de l'expression des gènes chez *Tetrahymena*. Les ciliés et en particulier *P. tetraurelia* avec 2% d'adénines méthylées sont particulièrement riches en méthylation des adénines parmi les eucaryotes. Ils sont un modèle de choix pour l'étude de cette modification et des rôles associés.

Chapitre 3 : *Paramecium tetraurelia*, modèle d'étude de la régulation épigénétique de l'organisation du génome

I. Classification, dimorphisme nucléaire et cycle de vie

Paramecium tetraurelia est un unicellulaire cilié appartenant au groupe des alvéolés parmi les eucaryotes et au groupe d'espèces apparentées *Paramecium aurelia* (Derelle and Lang, 2012). Les ciliés de formes très diverses sont caractérisés par la présence de cils sensoriels et motiles sur toute la cellule lui permettant de se déplacer en nageant, se nourrir et sentir les signaux de l'environnement. Si certains ciliés sont des parasites ou des symbiontes, *P. tetraurelia* est un organisme libre retrouvé dans les eaux douces sur tous les continents (Beisson et al., 2010; Betermier and Duharcourt, 2014).

Comme tous les ciliés, *P. tetraurelia* a la particularité de présenter deux types de noyaux dans un même cytoplasme lui permettant de séparer en deux compartiments distincts les fonctions somatiques et germinales. Les deux micronoyaux ou MICs sont diploïdes et assurent les fonctions germinales. Le macronoyau ou MAC est hautement polyploïde, $800n$, et assure les fonctions somatiques. Les MICs sont transcriptionnellement inactifs et ne sont pas requis pour la croissance végétative où ils se divisent par mitose. Ils assurent la transmission de l'information génétique d'une génération à la suivante par méiose (Brygoo et al., 1980; Tam and Ng, 1986). Le MAC est responsable de l'expression des gènes et détermine le phénotype de la cellule. Il se divise pendant la croissance végétative, se séparant en deux moitiés sans formation apparente de fuseau mitotique ni condensation des chromosomes. Le MAC est perdu suite aux événements sexuels et un nouveau MAC est formé à partir d'une copie mitotique du noyau zygotique issu des produits de méiose des MICs (Tucker et al., 1980).

Le cycle de vie de la paramécie se divise en deux phases, d'une part la croissance végétative et d'autre part les événements sexuels (Figure 9). La croissance végétative consiste en des divisions clonales et asexuées lorsque la nourriture est disponible et à des

températures entre 14°C et 30°C. Les évènements sexuels sont induits par la restriction alimentaire et peuvent aboutir à deux modes de reproduction, soit par autogamie impliquant une seule cellule, soit par conjugaison impliquant deux partenaires de type sexuel opposé (Beisson et al., 2010). L'autogamie est un processus d'auto-fertilisation se produisant chez des cellules d'âge élevé (correspondant à un nombre de divisions clonales élevé) (Sonneborn, 1974). Les deux MICs se divisent par méiose produisant 8 gamètes haploïdes. 7 de ces noyaux haploïdes sont détruits, un seul est conservé et se divise par mitose pour donner deux noyaux haploïdes identiques. Ces deux noyaux fusionnent pour former le noyau zygotique diploïde entièrement homozygote (caryogamie). Le noyau zygotique se divise deux fois pour donner quatre noyaux diploïdes : deux noyaux formeront les nouveaux MICs, et deux noyaux formeront les nouveaux MACs. Le destin de chaque noyau sous forme de MIC ou de MAC est déterminé par le positionnement de chaque noyau dans la cellule après la deuxième division zygotique. Les noyaux au pôle postérieur de la cellule deviennent des MACs et ceux au pôle antérieur deviennent des MICs (Grandchamp and Beisson, 1981). En parallèle, un gradient intracellulaire d'ions Ca^{2+} et K^{+} entre les deux pôles participerait à la détermination des MICs et des MACs (Betermier and Duharcourt, 2014). A l'issue de la première division dite caryonidale après les évènements sexuels, les MICs se divisent par mitose et chaque MAC ségrège dans une cellule fille (caryonide). La conjugaison fait intervenir deux partenaires de types sexuels opposés, mating type O et E et chez des cellules relativement jeunes (nombre de divisions clonales peu élevé) (Sonneborn, 1974). Dans le cas de la conjugaison, après les évènements de méiose, les partenaires échangent réciproquement par un pont cytoplasmique, un des deux noyaux haploïdes. Dans chaque cellule, le noyau haploïde résident et celui issu de l'autre cellule fusionnent pour donner un noyau zygotique. Puis chaque cellule poursuit indépendamment le développement de ses nouveaux MICs et MACs (Beisson et al., 2010). Tout le long de ce processus, le MAC maternel est progressivement détruit en une trentaine de fragments qui persisteront dans le cytoplasme et seront dilués et perdus au fur et à mesure des divisions cellulaires (Berger, 1973). Tant que les nouveaux MACs ne sont pas entièrement fonctionnels, le MAC maternel fragmenté continue d'assurer l'expression des gènes.

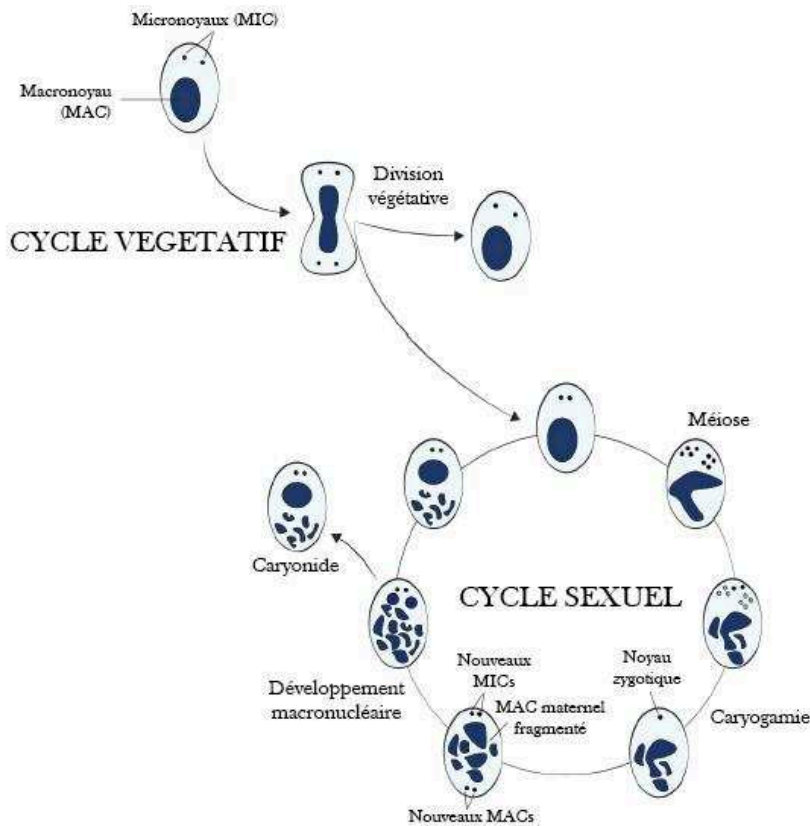


Figure 9 : Cycle de vie de *Paramecium tetraurelia*

Pendant le cycle végétatif la cellule se divise par fission clonale. Le cycle sexuel est initié par la restriction alimentaire et se poursuit soit par autogamie soit par conjugaison nécessitant deux partenaires de signe sexuel opposé. N'est représentée ici que l'autogamie mais le cycle est similaire et détaillé plus haut. Les deux MIC germinaux se divisent par méiose produisant 8 gamètes haploïdes dont un seul est conservé et se divise pour donner deux noyaux haploïdes identiques qui fusionnent lors de la caryogamie pour donner le noyau zygotique. Le noyau zygotique se divise deux fois par mitose pour donner deux nouveaux MICs et deux nouveaux MACs. Après la première division, les deux MICs se divisent par mitose et un MAC s'égère dans chaque cellule fille. Pendant ce temps, le MAC maternel est dégradé en une trentaine de fragments persistant dans le cytoplasme.

II. Organisation des génomes

Les noyaux MIC et MAC ne diffèrent pas seulement dans la cellule par leur fonction biologique mais aussi par le contenu et la structure du génome et des chromosomes bien qu'ils soient originaires du même noyau zygotique. Le MIC germinale contient une version complète du génome dont la complexité est de 100Mb environ tandis que le MAC somatique correspond à une version réarrangée du génome micronucléaire dont ont été éliminés 25% de séquences pendant la différenciation des nouveaux MACs lors des événements sexuels (Arnaiz et al., 2012; Prescott, 1994).

Le génome macronucléaire d'une complexité de 72Mb est très riche en AT (72%) (Arnaiz et al., 2012; Aury et al., 2006). Il est composé de 188 scaffolds dont 150 au moins sont

terminés par des répétitions télomériques en faisant des chromosomes. Ils vont d'une taille de 45kpb à 1Mpb. Le génome de la paramécie est un des plus riches et denses en gènes parmi les génomes eucaryotes avec 40000 gènes annotés et une fraction codante représentant 78% du génome (Aury et al., 2006; Zagulski et al., 2004). Les régions intergéniques sont courtes (350pb en moyenne) et avec très peu de séquences répétées. Les gènes contiennent 90000 très petits introns de 20 à 34 nt. La richesse en gènes est la conséquence de 3 duplications globales du génome successives dont les deux plus récentes précèdent la diversification des 15 espèces apparentées du groupe *aurelia* (Aury et al., 2006; Johri et al., 2017). 40 à 50% des gènes dupliqués à l'issue de la duplication la plus récente sont encore retenus dans les génomes des espèces *Paramecium aurelia* (McGrath et al., 2014). Ces paralogues sont nommés par la suite onhologues. Le nombre exact de chromosomes du MIC n'est pas connu mais estimé entre 30 et 60 paires visibles en métaphase (Betermier and Duharcourt, 2014).

III. Développement du noyau somatique

Pendant le développement du MAC somatique, le génome subit de nombreux cycles d'endoréplication et une élimination massive et reproductible d'ADN dans le génome MAC.

1. Endoréplication

Le génome somatique passe de l'état diploïde $2n$ à l'état polyploïde $800n$ lors du développement du nouveau MAC (Figure 10A). La ploïdie finale est atteinte après la première division cellulaire à la fin des événements sexuels (Berger, 1973). Peu de choses sont connues sur l'endoréplication mais on sait que 4 à 5 cycles de réplication se produisent en même temps que les événements d'éliminations d'ADN dans le nouveau MAC (Betermier and Duharcourt, 2014).

2. Élimination d'ADN

Les réarrangements massifs et hautement reproductibles d'ADN qui se produisent dans le nouveau MAC conduisent à la perte d'environ 25% de séquences spécifiques du génome MIC. Jusqu'à présent il était compliqué d'accéder au génome micronucléaire du fait de la

petite taille des micronoyaux, sa faible ploïdie par rapport à celle du MAC et de l'absence d'assemblage MIC. C'est d'abord la comparaison entre des extractions d'ADN des génomes MIC et MAC et l'étude de la librairie de phages du génome MIC qui a permis d'identifier des séquences spécifiques du génome germinale et absentes du génome somatique (Preer et al., 1992; Steele et al., 1994). Le séquençage du génome MAC dont les réarrangements d'ADN sont bloqués par l'inactivation de l'endonucléase responsable de l'introduction de cassures double brins aux bornes des séquences à éliminer a montré que 25% des séquences de ce génome MAC non réarrangé ne s'alignaient pas sur le génome MAC de référence (Arnaiz et al., 2012). L'amélioration récente des techniques de purification du MIC par cytométrie en flux et les premiers assemblages du génome MIC séquencé ont confirmé que l'ensemble de ces séquences sont contenues dans le génome MIC et éliminées du génome MAC (Guérin et al., 2017). Cependant, cet assemblage est encore très fragmenté et ne permet pas d'identifier plus que quelques familles d'éléments transposables parmi les nombreuses séquences répétées ni d'établir la structure des chromosomes MIC que l'on estime entre 30 et 60 paires visibles en métaphase (Betermier and Duharcourt, 2014; Guérin et al., 2017).

2.1 *Elimination imprécise et hétérogène de séquences*

90% des séquences spécifiques du MIC éliminées pendant le développement du MAC sont constituées de longues portions d'ADN de plusieurs kb contenant des séquences répétées comme des mini-satellites ou des transposons (Figure 10A). L'élimination imprécise de ces fragments d'ADN est couplée à la fragmentation des chromosomes MAC. Elle s'accompagne de l'ajout de télomères de longueurs variables *de novo* aux extrémités des nouveaux chromosomes MAC plus courts que les chromosomes MIC. Les sites d'ajout des télomères peuvent varier le long de la séquence. Les extrémités peuvent parfois être jointes ensemble formant une délétion intrachromosomique au lieu d'un ajout de télomères (Arnaiz et al., 2012; Le Mouël et al., 2003).

Un certain nombre d'éléments répétés ont été identifiés et le séquençage récent du génome micronucléaire offre de nouvelles informations sur son contenu en séquences répétées. Trois principales familles de transposons de la superfamille Tc1/Mariner avaient été initialement décrites après le séquençage du génome macronucléaire dans lequel le processus d'élimination avait avorté (Arnaiz et al., 2012). Les transposons des familles

nommées Thon et Sardine contiennent de longs TIRs (Terminal Inverted Repeats) palindromiques et deux phases de lecture pour une transposase DDE de la famille IS630/Tc1 mariner et pour une tyrosine recombinase. La famille des transposons Anchois contient, elle, des TIRs plus courts sans palindromes, le même type de transposase DDE mais pas de tyrosine recombinase. Le séquençage du génome MIC purifié a permis de détecter de nouvelles familles d'éléments transposables. 13 éléments identifiés appartiennent à la superfamille IS360/Tc1 mariner des transposons d'ADN de classe II. Mais la majorité des éléments transposables appartiennent à la classe I de transposons non-LTR, LINE, contenant tous une phase de lecture avec des domaines pour une reverse transcriptase et pour une endonucléase apurinique/apyrimidique. Quelques éléments SINE moins fréquents sont aussi détectés (Guérin et al., 2017).

2.2 Elimination précise d'IES

Des 27 Mb d'ADN éliminés pendant la différenciation du nouveau MAC, 14% environ correspondent à un autre type de séquences plus courtes et excisées très précisément, les Internal Eliminated Sequences (IES) (Arnaiz et al., 2012) (Figure 10A). Un petit nombre d'IES sont d'abord découvertes en comparant des extractions d'ADN du génome MIC et du génome MAC et grâce à la librairie de phages du génome MIC (Preer et al., 1992; Steele et al., 1994). Le séquençage du génome MAC dont les réarrangements d'ADN sont bloqués et celui du MIC ont permis d'identifier l'ensemble des 45000 IES (Arnaiz et al., 2012; Guérin et al., 2017). Les IES sont de courtes séquences, en copie unique, non codantes réparties tout le long du génome. Elles peuvent être retrouvées en intergénique mais 47% des gènes contiennent au moins une IES. Leur excision est donc cruciale chez la paramécie pour restaurer l'intégrité des phases ouvertes de lecture et permettre l'expression correcte de ces gènes. Toutes les IES sont excisées au nucléotide près à leurs bornes TA.

Les IESs sont toutes bornées par deux dinucléotides TA en 5' et 3' et après excision ne reste qu'une seule de ces bornes (Arnaiz et al., 2012; Betermier and Duharcourt, 2014). Une séquence consensus faiblement conservée de 8pb est retrouvée dans les IES aux extrémités : (5'-TAYAGYNR-3'). Cette séquence rappelle les répétitions inversées des extrémités des éléments transposables Tc1/Mariner. On propose que les IES proviennent d'une invasion plus ancienne par ces éléments transposables (Klobutcher and Herrick, 1997). Le TA est

essentiel pour l'excision et des mutations dans la séquence consensus peuvent empêcher l'excision correcte (Gratias et al., 2008; Matsuda et al., 2004; Mayer and Forney, 1999). Le modèle IBAF est proposé après l'identification des premières IES : Invade, Bloom, Abdicate, Fade (Klobutcher and Herrick, 1997). Il propose que les IES dérivent de transposons ADN qui ont envahi le génome mais se sont détériorés avec le temps par l'accumulation de mutations, délétions et substitutions. La pression de sélection conserve le faible consensus aux extrémités des IES pour permettre une excision précise et efficace. Il s'avère en plus que les éléments TC1/Mariner s'intègrent préférentiellement dans des dinucléotides TA qui sont dupliqués suite à l'intégration confirmant l'hypothèse d'une invasion par ces éléments (Plasterk et al., 1999). Les IES sont de tailles réduites : 93% d'entre elles ont une taille inférieure à 150pb et 35% entre 26 et 35pb (Arnaiz et al., 2012) (Figure 10B). En comparant la taille d'IES contenues dans des gènes, on a montré que les IES acquises avant la première duplication globale du génome sont les plus courtes (Arnaiz et al., 2012; Aury et al., 2006). Ces constatations vont dans le sens du modèle, confirmant le déclin des IES au cours du temps. Certains transposons comme Anchois sont excisés de manière très précise pendant le développement renforçant l'hypothèse du modèle et le lien entre IES et transposons (Arnaiz et al., 2012).

Les IES ont des tailles relativement courtes, la grande majorité sont inférieures à 150pb. La distribution de leur taille présente une périodicité de 10,2pb correspondant à un tour d'hélice de la molécule d'ADN. Cette périodicité pourrait refléter des contraintes du mécanisme d'excision (Figure 10B) nécessitant que les deux extrémités d'une IES soient présentes sur une même face de la double hélice d'ADN. Si une seule des bornes TA est mutée, il n'y a pas de cassures doubles brins à aucune des deux bornes y compris la sauvage suggérant que les deux extrémités dépendent l'une de l'autre pour l'excision (Gratias et al., 2008). Suite à l'excision de longues IES, des cercles d'IES correspondant à la fermeture covalente des extrémités des IES sont détectés (Bétermier et al., 2000). On imagine que deux machineries d'élimination s'associent à chaque extrémité des IES sur la même face et interagissent en formant une boucle d'ADN. Les IES les plus courtes, un pic correspondant à 26-32pb, ne nécessiteraient pas de boucle d'ADN pour que les deux machineries interagissent. Le pic suivant, 32-42pb, est par contre sous-représenté. Cette taille interdite s'expliquerait par le fait que les IES soient à la fois trop longues pour que les machineries

d'excision interagissent sans boucle et trop courtes pour qu'une boucle d'ADN se forme. Etant éliminées moins efficacement elles seraient contre sélectionnées (Arnaiz et al., 2012).

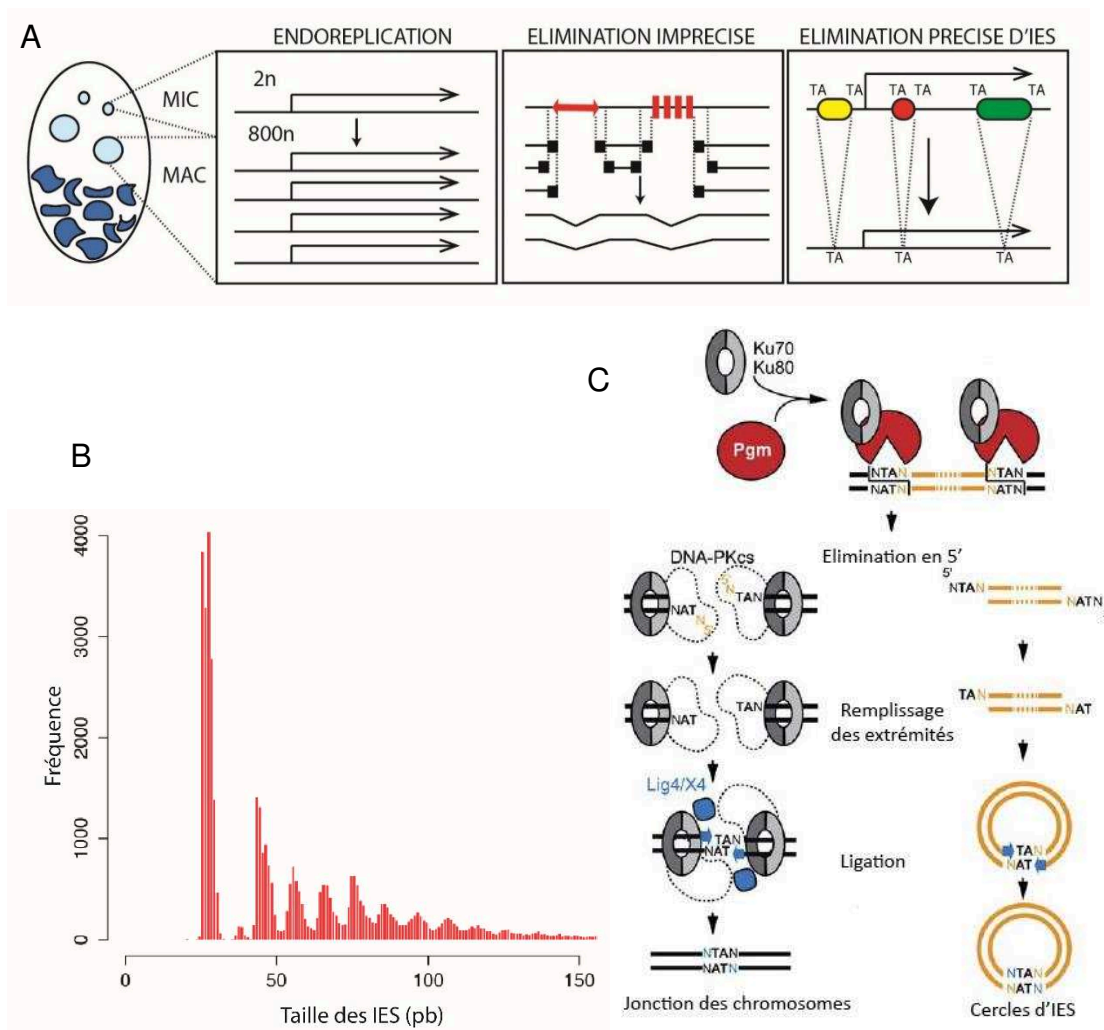


Figure 10 : Elimination d'ADN et réparation des cassures double brin pendant le développement du nouveau MAC

A. Réarrangements programmés pendant le développement du nouveau MAC. Un chromosome MIC est représenté en haut et les chromosomes MAC résultants en bas. A gauche, des cycles d'endoréplication permettent de passer d'une ploïdie de 2n à une ploïdie de 800n. Au centre, dans le chromosome MIC des séquences répétées et éléments transposables sont représentés en rouge et éliminés imprécisément permettant la fragmentation des chromosomes et l'ajout de télomères en noir. A droite, le chromosome MIC contient une unité de transcription symbolisée par la flèche interrompue par des IES en couleur bornées par les TA. Ces IES sont éliminées précisément pour restaurer les phases ouvertes de lecture ne laissant qu'un seul TA.

B. Distribution de taille des IES : fréquence des IES inférieures à 150 pb en fonction de la taille. 93% des IES ont une taille inférieure à 150 pb et ont une périodicité de 10 pb environ correspondant à un tour d'hélice de la molécule d'ADN. Le premier pic contient 35% des IES d'une taille comprise entre 26 et 28pb. Les pics suivants sont centrés sur 45-46, 55-56, 65-66pb. Cette distribution périodique est observée quelle que soit la localisation des IES. Adapté de (Arnaiz et al., 2012)

C. Clivage et réparation des cassures double brins par le Non.Homologous.End.Joining. Les deux brins d'ADN sont représentés avec l'IES en orange et les séquences somatiques en noir. Pgm en rouge, interagit avec l'hétérodimère Ku en gris pour venir cliver les bornes TA laissant une cassure double brins et 4 bases libres à chaque borne. Dans le système classique de NHEJ, l'hétérodimère Ku se lie à chaque cassure et recrute DNA-Pkcs. Les extrémités se lient grâce au TA et le nucléotide en 5' est éliminé. La ligase IV et Xccr4 remplissent les extrémités 3' avant la dernière étape de ligation. Les IES excisées se circularisent par un mécanisme similaire si elles sont suffisamment longues et sont rapidement dégradées. Adapté de (Betermier and Duhaucourt, 2014)

3. La machinerie d'excision et de réparation

3.1 Pgm, l'endonucléase responsable de l'introduction des cassures double brins

L'hypothèse IBAF expliquant l'origine et l'évolution des IES propose également que leur excision soit assurée par une transposase domestiquée intégrée dans le génome de la paramécie (Klobutcher and Herrick, 1997). Elle pourrait ainsi éliminer l'ensemble des IES et permettre progressivement leur disparition du génome. Si les IES ont pour origine les transposons de la famille Tc1/Mariner, on imagine que l'endonucléase responsable de leur excision soit une transposase issue de cette famille. L'analyse du clivage des IES dément cette hypothèse. Le clivage est initié par des cassures double brins produisant des extrémités libres de 4 nucléotides en 5' centrées à chaque borne TA (Gratias and Bétermier, 2003). Il se trouve que les transposases Tc1/Mariner produisent des extrémités libres de 2 nucléotides en 3' et laissent les deux TA sur le site donneur (Plasterk et al., 1999). Les transposons de classe II de la famille PiggyBac ne génèrent pas d'empreinte sur le site donneur et la géométrie de la coupure correspond à celle observée pour le clivage des IES. *In vitro*, la transposase PiggyBac réalise des coupures double brins aux extrémités des transposons laissant des extrémités libres 5' de 4 nucléotides TTAA (Mitra et al., 2008).

Un homologue de la transposase PiggyBac est retrouvé dans le génome de la paramécie et exprimé uniquement pendant le développement du nouveau MAC (Baudry et al., 2009) (Figure 12). L'inactivation du gène pendant le développement provoque la létalité de la progéniture sexuelle et la rétention de l'ensemble des séquences spécifiques du MIC (IES et éléments répétés). Il est nommé PiggyMac (Pgm). Présent en une seule copie, il n'est plus un transposon actif mais une transposase domestiquée. Pgm comprend un domaine catalytique caractéristique avec une triade d'acides aminés aspartiques et un domaine riche en cystéines également retrouvé chez la transposase domestiquée Tbp2 de *Tetrahymena thermophila* (Dubois et al., 2017). Tbp2 est requise pour l'élimination d'ADN et colocalise avec les marques de l'hétérochromatine et la protéine à chromodomaine Pdd1p (Cheng et al., 2010). Pgm dispose d'une longue extension en C-terminal qui semble adopter une conformation Coiled-coil absente des autres transposases PiggyBac. Des expériences de complémentation fonctionnelle avec des versions mutées ou délétées dans les différents domaines de Pgm ont montré que le domaine catalytique avec sa triade d'acides aminés

aspartiques et le domaine CR sont nécessaires à l'activité de Pgm. Le domaine C-terminal Coiled-Coil est lui requis pour la multimérisation de Pgm formant le complexe d'excision (Dubois et al., 2017).

Les séquençages des génomes de cellules inactivées *PGM* et du génome MIC ont confirmé que Pgm était requis pour l'élimination de l'ensemble des 45000 IES et des éléments répétés éliminés imprécisément (Arnaiz et al., 2012; Guérin et al., 2017). La complexité du génome inactivé *PGM* est très proche du génome MIC, environ 100Mb. Pour autant les deux génomes ne sont pas complètement équivalents. 3Mb d'ADN spécifique du MIC, enrichis en mini satellites et déplétés en éléments transposables, ne sont que très faiblement couverts dans le génome inactivé *PGM*. Il est possible que Pgm ne soit pas requis pour l'élimination de cette fraction du génome micronucléaire. On émet aussi l'hypothèse que l'efficacité de l'ARN interférence (ARNi) diminuant avec le temps, l'inactivation de *PGM* soit moins efficace dans à la fin de l'autogamie. La quantité de Pgm redevient suffisante dans la cellule à la fin de son inactivation par ARN interférence permettant ainsi l'excision de ces 3 Mb de séquences en fin de développement MAC. On note ainsi que l'inactivation de facteurs du développement agissant en amont de Pgm n'empêche pas l'élimination de ces 3Mb d'ADN MIC spécifique ce qui est cohérent avec cette deuxième hypothèse (Guérin et al., 2017).

3.2 Le Non-Homologous-End-Joining pour réparer les cassures

La réparation des cassures double brins (CDB) hautement toxiques pour la cellule doit faire suite au processus d'excision. Elle doit être fidèle pour les IES afin de maintenir des phases de lecture correctes mais aussi préserver l'organisation linéaire des chromosomes. La réparation chez *P. tetraurelia* fait intervenir un mécanisme de Non-Homologous End Joining Pathway dont les composants classiques ont été progressivement identifiés (Betermier and Duharcourt, 2014; Lieber, 2010). Alors que cette voie de réparation des cassures double brins est généralement considérée comme sujette aux erreurs contrairement à la recombinaison homologue, elle est chez *P. tetraurelia* particulièrement précise et sans erreurs. La conformation des extrémités 5' sortantes et leur maintien par l'appariement des TA centraux permettent la précision de la réparation (Figure 10C) (Betermier and Duharcourt, 2014).

Les premiers acteurs identifiés sont le complexe Ligase IV/Xrcc4 par ARN interférence (Kapusta et al., 2011) (Figure 10C). Leur inactivation pendant le développement en plus de provoquer la létalité de la progéniture sexuelle, entraîne l'accumulation de cassures double brins et inhibe la réparation et la circularisation des longues IES. Il est montré par quantification au iodure de propidium pendant le développement que la quantité d'ADN dans les nouveaux MACs de cellules inactivées Ligase IV est diminuée par rapport à des cellules sauvages indiquant un défaut de réplication. Ce phénotype est sauvé par l'inactivation de *Pgm* montrant que le défaut de réplication reflète un défaut de réparation des cassures double brins accumulées dans les nouveaux MACs. A partir d'un échantillon d'IES, on montre que le complexe est requis pour la ligation des extrémités et l'ajout de nucléotide manquant en 3' pour fermer la cassure, mais n'est pas nécessaire à l'élimination du nucléotide terminal en 5' de la cassure.

Pour assurer une réparation efficace et précise des très nombreuses cassures double brins déposées par Pgm dans les MACs en développement, les processus d'excision et de réparation via la machinerie du NHEJ sont couplés chez *P. tetraurelia* (Figure 10C). Parmi les acteurs du NHEJ, l'hétérodimère Ku70/80 se lie aux extrémités des cassures double brins et facilite leur appariement et la réparation (Betermier and Duharcourt, 2014; Lieber, 2010). Chez la paramécie, il est exprimé spécifiquement pendant le développement et est requis pour l'introduction de CDB aux extrémités des IES (Marmignon et al., 2014). L'absence de Ku80c provoque la rétention des IES testées, on ne détecte pas de CDB ni de produits d'excision que sont les cercles d'IES. Contrairement à la Ligase IV, Ku70a/80c ne sont pas nécessaires à l'endoréplication. Des versions de Ku70a et Ku80c fusionnées à HA et de Pgm fusionnée à MBP sont produites dans un système hétérologue de cellules d'insectes. Des expériences de co-immunoprécipitation de MBP-Pgm et HA-Ku70a, HA-Ku80c montrent que les protéines interagissent en complexe dans des extraits cellulaires solubles indépendamment de la présence d'ADN.

Dans le NHEJ, Ku recrute ensuite la DNA-Pkcs, une protéine kinase qui stabilise les extrémités reliées, facilite l'élimination du nucléotide en 5' de la CDB (Figure 10C). Un homologue de cette protéine est retrouvé chez la paramécie et sa déplétion entraîne la persistance tardivement dans le développement d'extrémités libres de chromosomes

(Betermier and Duhaucourt, 2014). La machinerie du NHEJ est couplée étroitement à la machinerie d'excision et permet ainsi une réparation efficace et précise.

4. Reconnaître les séquences à éliminer et cibler la machinerie d'excision

S'il est désormais acquis que l'élimination de l'ensemble des séquences MIC spécifiques dépend de la machinerie d'excision Pgm, il reste encore à comprendre comment une si grande diversité de séquences est reconnue et excisée. Comment éliminer avec une telle précision pour les IES tout particulièrement et de manière si reproductible au cours des générations quand les séquences ne partagent pas de motif de séquence strictement conservé et ne sont pas apparentées à la transposase PiggyBac ?

4.1 Contrôle épigénétique de l'élimination

L'étude chez *P. tetraurelia* de l'épimutant d48 a permis de montrer que le contenu des différents noyaux gouvernait les profils d'élimination d'ADN. Cet épimutant ne parvient pas à exprimer l'antigène de surface A délété de l'extrémité d'un chromosome macronucléaire. Cette délétion est héritée de générations en générations sexuelles alors même que le MIC contient une version sauvage de ce gène (Epstein and Forney, 1984; Forney and Blackburn, 1988). Si maintenant, on introduit dans le MAC végétatif de l'épimutant d48, le gène A ou un fragment de ce gène, on restaure à la génération suivante le profil sauvage et l'expression du gène A dans le nouveau MAC. Ces premières observations indiquent que chez la paramécie il existe un profil de réarrangements hérité de manière non mendélienne. La présence d'un gène dans le MAC maternel est requise pour sa maintenance dans le MAC de la génération suivante indépendamment du contenu du MIC (Jessop-Murray et al., 1991; Koizumi and Kobayashi, 1989; You et al., 1991).

L'hérédité maternelle s'observe aussi pour les IES. Certaines IES, dites contrôlées maternellement (mclES), une fois introduite dans le MAC maternel, conduisent à la rétention de l'IES MIC homologue dans le nouveau MAC à l'issue des événements sexuels. Les mclES représentent 1/3 des IES testées (5/13) expérimentalement. On ne détecte pas aujourd'hui de particularités ou de points communs concernant la séquence ou la taille de ces mclES (Duhaucourt et al., 1995, 1998).

Le contenu du MAC maternel influence donc le contenu du nouveau MAC indépendamment du MIC. L'élimination d'ADN serait ainsi contrôlée par un mécanisme épigénétique reposant sur l'homologie entre le contenu du MAC maternel toujours présent dans le cytoplasme lorsque se développe le nouveau MAC et le contenu du MIC. On envisage alors que le profil d'élimination soit obtenu par comparaison entre les génomes MIC et MAC et que cette comparaison transnucléaire s'opère par exemple par appariement de séquences homologues. Etant donné qu'aucune interaction physique n'est détectée entre les MAC maternels et zygotiques, il est nécessaire que des acteurs en trans permettent le dialogue entre ces deux génomes.

4.2 ARN non codants et genome scanning model

Les acteurs permettant le dialogue entre les génomes germinaux et somatiques sont constitués par différentes classes d'ARN non codants permettant la régulation épigénétique des éliminations d'ADN.

4.2.1 Petits ARN non codants : scnARN et iesARN

Les scnARN sont une classe de petits ARN non codants de 25 nucléotides de long produits spécifiquement pendant la méiose à partir du génome MIC (Lepère et al., 2009; Sandoval et al., 2014) (Figure 11). Ces petits ARN sont produits par une voie de l'ARN interférence, par les protéines Dicer like RNase III, Dicer like 2 et 3 (Dcl2 et Dcl3). Le séquençage des scnARN a montré qu'ils portent en majorité une signature 5'UNG et pour une minorité la séquence complémentaire en position 21-23. Ces observations indiquent qu'ils ont été clivés à partir de précurseurs double brins donnant des duplexes d'ARN avec des extrémités libres de 2 nucléotides en 3'. L'étude *in vitro* de l'activité de clivage de Dcl2 et Dcl3 purifiés sur des ARN double brins a montré que les deux enzymes coopèrent pour produire les scnARN. Le séquençage des produits de clivage par Dcl2 montre des petits ARN d'une taille de 25nt avec une signature 5'UNN quand le séquençage des produits de clivage par Dcl3 montre des petits ARN de tailles variables avec une signature 5'UNG. Ces résultats montrent que Dcl2 est responsable de la taille des scnARN et Dcl3 de leur signature en 5' (Hoehener et al., 2018). Les scnARN sont requis pour les éliminations d'ADN. En effet, l'injection de duplexes d'ARN synthétiques de 25 nucléotides mimant les scnARN entraînent

l'élimination des séquences homologues dans le nouveau MAC (Lepère et al., 2008). L'inactivation par ARN interférence des gènes *DCL2/3* entraîne la létalité de la progéniture sexuelle, la rétention de l'ensemble des séquences éliminées imprécisément et d'une portion des IES (10%) (Lhuillier-Akakpo et al., 2014; Sandoval et al., 2014).

Les iesARN sont de petits ARN non codants d'une taille variant entre 26 et 31nt de long avec une extrémité 5'UAG et une extrémité 3' NUAUC produits par Dcl5 à la fin du développement MAC (Hoehener et al., 2018; Sandoval et al., 2014). Ces petits ARN sont produits à partir des IES excisées et circularisées par la Ligase IV (Allen et al., 2017). L'inactivation de *DCL5* n'entraîne la rétention que d'un petit nombre d'IES (Lhuillier-Akakpo et al., 2014; Sandoval et al., 2014)). La part des IES retenues augmente lorsque l'on inactive *DCL2/3* conjointement avec *DCL5*, le gène codant pour l'enzyme produisant une autre classe de petits ARN non codants les iesARN (Swart et al., 2017). Ces résultats tendent à montrer que les iesARN et les scnARNs coopèrent pour exciser les séquences germinales et que les iesARN amplifieraient l'élimination d'ADN induite par les scnARN (Sandoval et al., 2014; Swart et al., 2017).

Ce sont les protéines Piwi de la famille Argonaute qui prennent en charge ces petits ARN non codants pour assurer leur production et leur transport (Figure 11). 14 homologues de protéine Piwi sont retrouvés chez la paramécie dont certains exprimés durant le cycle végétatif sont probablement impliqués dans les voies d'ARN interférence (Bouhouche et al., 2011). L'absence de deux de ces protéines Ptiwi01 et 09 diminue la quantité de scnARN et provoque la rétention d'un certain nombre d'IES (Bouhouche et al., 2011; Furrer et al., 2017). Par RNA Immunoprécipitation (RIP), il a été montré que Ptiwi9 lie les scnARN (Furrer et al., 2017). Deux autres protéines Ptiwi10 et 11 exprimées tardivement pendant le développement MAC sont par contre requises pour la production d'iesARN et par RIP il est observé que Ptiwi10 lie les iesARN (Furrer et al., 2017).

4.2.2 Longs ARN non codants maternels

De longs ARN non codants sont transcrits à partir des deux brins du génome MAC maternel à bas niveau et de manière constitutive (Figure 11). Ils protégeraient le génome de l'élimination d'ADN. Par ARNi, la dégradation des longs ARN non codants maternels provoque l'élimination des séquences homologues dans le nouveau MAC (Lepère et al.,

2008). Il est aussi possible d'aboutir à la délétion d'un gène non essentiel dans le nouveau MAC lorsque celui-ci est ciblé par ARN interférence (Garnier et al., 2004). Même si une petite portion du gène est ciblée, il est possible de provoquer sa délétion. Plus encore que sa présence dans le MAC maternel, c'est la transcription d'une séquence qui assure sa présence dans le nouveau MAC.

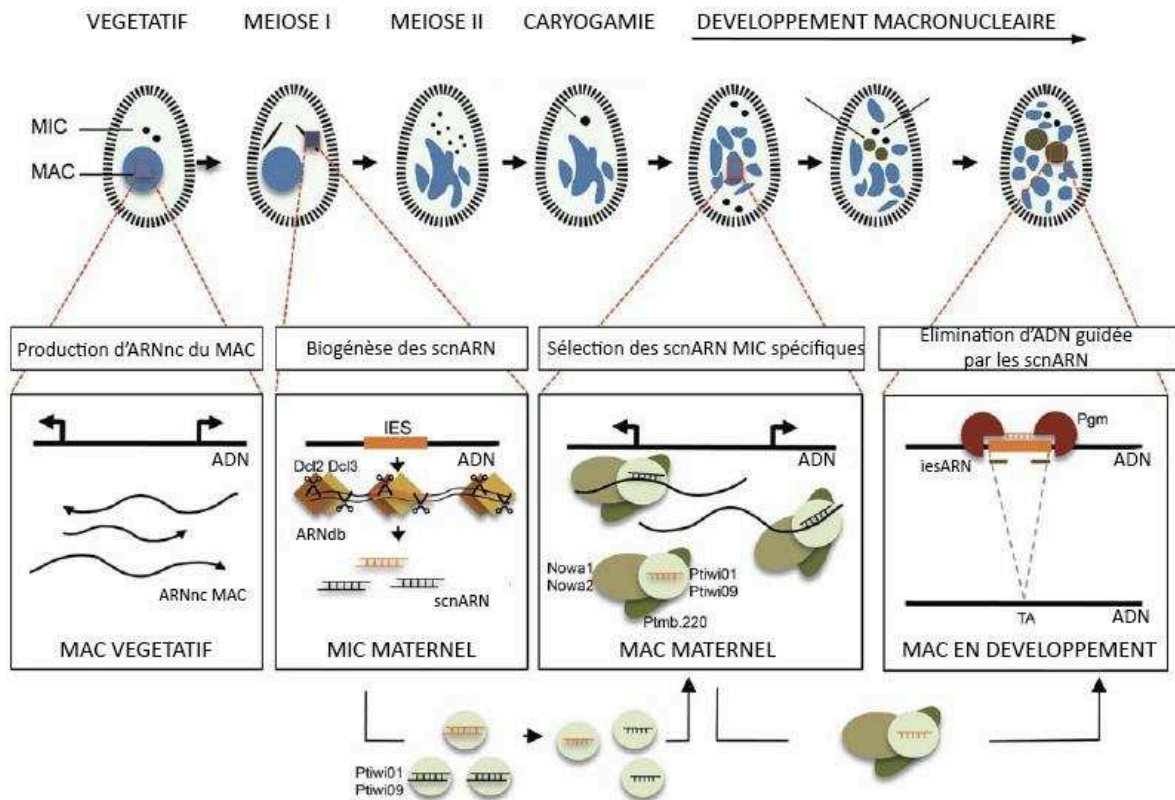


Figure 11 : Modèle de Genome Scanning

En haut sont représentées les étapes du cycle sexuel et schématisées en dessous. Tout à gauche, de longs transcrits non codants sont produits à partir de tout le MAC maternel. Puis lors de la méiose I, des duplexes de scnARN de 25nt sont produits par les protéines Dicer like 2 et 3 à partir de l'ensemble du MIC, transportés dans le cytoplasme par les protéines Ptiwi01 et 09 conduisant à l'élimination du brin passager. Dans le MAC maternel, les complexes Ptiwi-scnARN s'apparient sur les long ARN non codants maternels grâce à l'hélicase putative PtmB.220 et les protéines de liaison à l'ARN Nowa1 et 2, pour sélectionner les scnARN spécifiques des séquences germinales transportés ensuite dans les MAC en développement. Pendant les éliminations d'ADN les scnARN guident la machinerie d'excision aux séquences à éliminer.

4.2.3 Genome scanning model

Afin de réconcilier les rôles antagonistes de ces différentes classes d'ARN non codants fut proposé le modèle de genome scanning (Figure 11). Il repose sur une soustraction naturelle des génomes grâce aux scnARN et aux longs transcrits non codants maternels pour sélectionner les scnARN spécifiques des séquences du MIC à éliminer et permettre in fine l'excision.

Selon le modèle de genome scanning, les scnARN sont produits par les protéines Dcl2 et 3 à partir du génome MIC en méiose et chargés sur les protéines Ptiwi01 et 09 pour être transportés au MAC maternel. Les scnARN peuvent alors scanner le génome maternel à la recherche de séquences homologues parmi les longs transcrits non codants maternels produits à partir du MAC maternel. Les scnARN qui s'apparient avec un transcrit maternel correspondent aux séquences MAC qui doivent être conservées dans le nouveau MAC. Ces scnARN sont alors inactivés. A contrario, les scnARN qui ne trouvent pas de transcrits maternels pour s'apparier correspondent aux séquences MIC spécifiques absente du MAC maternel et qui doivent être éliminées du nouveau MAC. Ces scnARN sont transférés dans le nouveau MAC grâce aux protéines Ptiwi01/09 pour éliminer les séquences MIC (Figure 11).

Plusieurs faisceaux d'arguments vont dans le sens de ce modèle. Les protéines fusionnées à la GFP Ptiwi09 et Ptiwi01 se localisent au début du développement dans le cytoplasme puis dans le MAC maternel et enfin dans les MAC en développement (Bouhouche et al., 2011; Furrer et al., 2017). Ce type de localisation est aussi retrouvé chez le cilié *Tetrahymena* où les scnARN sont produits par Dicer-like 1 (Dcl1p) et pris en charge par la protéine Piwi Twi1p (Malone et al., 2005; Mochizuki et al., 2002). Cette dernière est capable de cliver le duplexe de scnARN pour enlever le brin passager et permettre l'import du petit ARN dans le nouveau MAC (Noto et al., 2010). Le séquençage à haut débit des populations de petits ARN dans un contexte sauvage confirme la dynamique des scnARN décrite dans le modèle. Au cours du développement, très précocement au moment de la méiose, on observe une production de petits ARN non codants de 25 nucléotides s'enrichissant progressivement en scnARN spécifiques des séquences MIC. Cet enrichissement reflète la dégradation des scnARN spécifiques du MAC et la sélection des scnARN spécifiques du MIC (Ignarski et al., 2014; Lhuillier-Akakpo et al., 2014; Maliszewska-Olejniczak et al., 2015; Sandoval et al., 2014; Singh et al., 2014). Le mécanisme permettant l'élimination des scnARN spécifiques du MAC s'étant appariés avec les transcrits maternels est encore mystérieux. Il est possible qu'il s'agisse soit d'une dégradation active suite à l'appariement ou bien simplement de la séquestration des scnARN dans le nouveau MAC.

D'autres acteurs ont été récemment mis en évidence pour leur implication dans le processus. La protéine Ptmb.220, une hélicase DEAH-Box putative, requise pour les éliminations d'ADN se localise dans l'ancien MAC puis le nouveau MAC. On l'imagine

participer au complexe scnARN/Ptiwi01/09 (Nowak et al., 2011). De même la protéine Nowa01 fusionnée à la GFP, une protéine se liant potentiellement aux ARN et requise pour l'élimination d'ADN, suit la même localisation au cours du développement (Nowacki et al., 2005). La protéine TFIS4 est l'homologue du facteur d'élongation de la transcription TFIS et est nécessaire pour les événements d'élimination. Elle est requise pour la production de transcrits non codants contenant des IES dans le MAC en développement. Ces transcrits somatiques permettraient de cibler les séquences homologues à éliminer dans le nouveau MAC en s'appariant aux scnARN spécifiques du MIC (Maliszewska-Olejniczak et al., 2015). La protéine Spt5 est un autre facteur de régulation de l'élongation de la transcription. Son homologue chez *P. tetraurelia* est nécessaire pour la biogénèse des scnARN et son inactivation entraîne la rétention de l'ensemble des séquences MIC spécifiques dans le nouveau MAC (Gruchota et al., 2017). On propose enfin que plus tard dans le développement du nouveau MAC, les iesARN et les protéines Ptiwi10/11 les accompagnant amplifient l'élimination d'ADN initiée par les scnARN à partir des produits d'excision (Furrer et al., 2017; Swart et al., 2017).

Le modèle de génome scanning permet d'expliquer les phénomènes d'hérédité maternel chez la paramécie et de réguler l'expression de certains gènes. Le signe sexuel chez *Paramecium tetraurelia* suit une transmission maternelle. Il n'est pas génétiquement déterminé par le MIC mais hérité et contrôlé maternellement par le MAC maternel. Le signe sexuel est déterminé par le profil de réarrangement alternatif du gène mtA pendant le développement. La délétion du promoteur du gène mtA dépendante des scnARN empêche son expression et permet le type sexuel O. Dans le cas du type sexuel E, le promoteur est retenu et le gène mtA est exprimé produisant la protéine transmembranaire mtA (Singh et al., 2014). Pour autant, ce modèle ne permet pas d'expliquer l'ensemble des événements d'élimination d'ADN. En effet si l'inactivation des protéines Dcl2/3 entraîne la rétention de l'ensemble des éléments répétés éliminés imprécisément, seules 10% des IES sont dépendantes des scnARN et environ 40% à la fois des scnARN et iesARN (Lhuillier-Akakpo et al., 2014; Swart et al., 2017). Il existe donc d'autres facteurs épigénétiques permettant d'expliquer la reconnaissance des séquences à éliminer et le ciblage de la machinerie d'excision.

4.3 L'histone méthyltransferase *Ezl1* est requise pour le développement MAC

La découverte récente dans le laboratoire de l'histone méthyltransferase *Ezl1* a permis de mieux comprendre comment sont ciblées les séquences germinales à éliminer (Frapporti et al., In submission; Lhuillier-Akakpo et al., 2014). *Ezl1* exprimée spécifiquement pendant le développement est responsable du dépôt des modifications d'histones H3K9me3 et H3K27me3 sur le génome de la paramécie dans les MAC en développement (Figure 12). Son inactivation provoque la létalité de la progéniture sexuelle et elle est nécessaire à l'élimination d'ADN. Elle est requise pour l'élimination de l'ensemble des séquences répétées et de 70% des IES. Plus les IES sont longues et récentes, plus elles sont sensibles à *Ezl1*. L'ensemble des IES sensibles à *Dcl2* et 3 le sont aussi à *Ezl1* indiquant une interdépendance entre les voies de petits ARN et le dépôt des modifications d'histone. Des expériences de Chip-qPCR montrent que H3K9me3 et H3K27me3 sont enrichies sur des éléments transposables et participent à leur régulation : en l'absence d'*Ezl1*, ces éléments sont surexprimés (Frapporti et al., In submission). En parallèle des facteurs en trans que sont les ARN, sont donc aussi impliqués des facteurs en cis, les modifications d'histones, qui assurent elles-aussi la reconnaissance et l'excision correcte des séquences à éliminer. Chez le cilié apparenté *Tetrahymena*, le dépôt d'H3K9me3 et H3K27me3 permettrait de guider l'endonucléase *Tpb2p*, homologue de *Pgm*, pour exciser les séquences germinales (Cheng et al., 2010; Vogt and Mochizuki, 2013). Chez d'autres organismes eucaryotes, on observe également la formation d'hétérochromatine guidée par de petits ARN (Martienssen and Moazed, 2015). On imagine alors que chez la paramécie comme chez *Tetrahymena*, la formation d'hétérochromatine en aval de la voie des scnARN permette la reconnaissance et l'élimination des séquences germinales au moins pour les éléments transposables éliminés imprécisément.

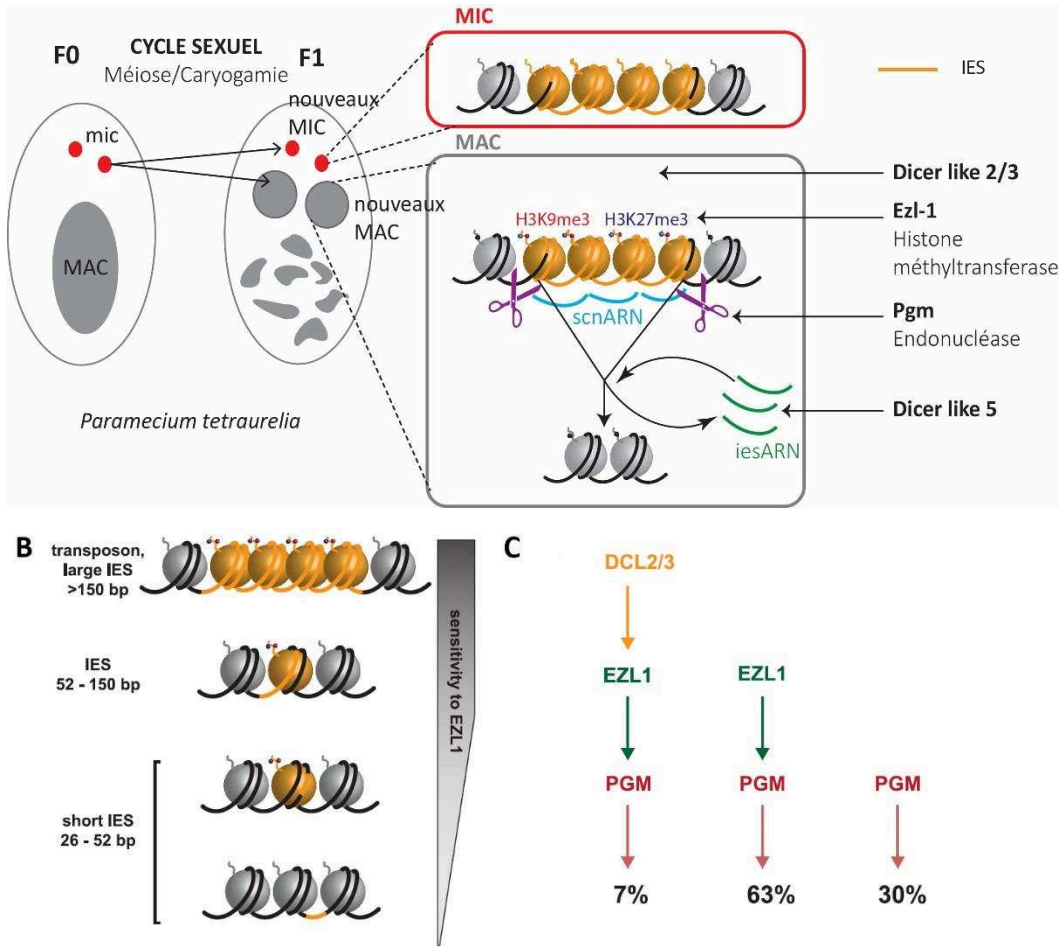


Figure 12 : Modèle d'élimination d'ADN et voies d'excision des IES

A. Modèle synthétisant l'ensemble des résultats acquis et permettant d'expliquer comment sont excisées les séquences germinales. Dans le cadre rouge, est représenté un chromosome MIC avec une séquence germinale en orange. Dans le cadre gris, dans le nouveau MAC les scnARN MICs spécifiques produits par Dcl2 et 3 reconnaissent les séquences germinales et permettent le dépôt des modifications d'histones H3K9me3 et H3K27me3 par l'histone méthyltransférase Ezl1. La machinerie d'excision est recrutée aux séquences marquées pour les éliminer permettant l'obtention du génome MAC final. Les iesARN sont produits lors de l'excision par la protéine Dcl5.

B. Sensibilité des IES à Ezl1 en fonction de leur longueur. Les IES les plus longues et les plus sensibles à Ezl1 sont compatibles avec la résolution de l'ordre du nucléosome apportée par les modifications d'histones. Les IES de taille intermédiaire entre 52 et 150pb sont toujours incluses au moins partiellement dans le nucléosome et toujours sensibles à Ezl1. Les plus petites IES sont par contre soit localisées à l'intérieur du nucléosome soit dans le linker d'ADN et la résolution apportée par les modifications d'histone n'est plus suffisante pour leur excision.

C. Représentation schématique des différentes voies partiellement chevauchantes pour l'excision des 45000 IES. Bien que toutes soient finalement excisées par Pgm, seule une petite portion, 7% dépend à la fois des scnARN et d'Ezl1, la majorité ne requiert qu'Ezl1 et une portion ne dépend d'aucun de ces deux facteurs.

4.4 Les mécanismes d'élimination des IES reposent sur des voies chevauchantes

L'ensemble des IES est excisé par l'endonucléase Pgm mais elles ne dépendent pas toutes des mêmes mécanismes de reconnaissance. Il a été montré que 5 des 13 IES testées dépendent de la présence d'une séquence homologue dans le génome maternel et sont

ainsi nommées mcIES pour IES maternellement contrôlées (Duharcourt et al., 1995, 1998). Le reséquençage du génome de cellules inactivées pour *DCL2/3* a montré que 10% des IES sont dépendantes des scnARN. Une portion encore plus faible d'IES mais différentes de celles retenues en RNAi *DCL2/3*, est dépendante des iesARN (Lhuillier-Akakpo et al., 2014). Il avait été proposé que le faible nombre d'IES dépendantes des scnARN pouvait en réalité être dû à une efficacité moindre de l'ARN interférence ou bien à la compensation par une autre voie. L'inactivation conjointe des voies de scnARN et iesARN entraîne la rétention de 40% des IES suggérant que les deux voies de petits ARN coopèrent pour exciser une partie des IES (Swart et al., 2017). Parmi ces IES, sont retrouvées les 5 mcIES dont l'excision dépend de la présence d'une séquence homologue dans le MAC maternel. Il est supposé que les IES qui dépendent des petits ARN correspondent à des mcIES mais rien ne permet de le déterminer à l'échelle du génome. L'inactivation d'*EZL1* montrait de son côté que 70% des IES dépendaient des modifications d'histones et qu'il s'agissait des IES les plus longues qui comprenaient les IES dépendantes de *Dcl2* et 3 (Lhuillier-Akakpo et al., 2014). Il subsiste donc une portion d'IES, 30%, les plus courtes qui ne dépendent ni des petits ARN ni d'*Ez11* pour leur élimination. Il faut noter parallèlement que l'ensemble des éléments transposables et répétés dépendent à la fois des petits ARN et des modifications d'histone pour leur élimination (Lhuillier-Akakpo et al., 2014). Les IES les plus courtes qui sont aussi les plus anciennes ne dépendent ni des petits ARN ni d'*Ez11* et seraient ainsi l'ultime étape dans l'évolution de ces séquences. Cette hypothèse est appuyée par l'observation que les éléments les plus longs et les plus récents dans le génome sont aussi ceux qui ont le plus haut score de rétention lorsque *DCL2/3* sont inactivés indiquant que plus les éléments sont grands et introduits récemment dans le génome plus ils requièrent les petits ARN et *Ez11* (Lhuillier-Akakpo et al., 2014). L'élimination de toutes ces séquences fait donc appel à plusieurs voies chevauchantes dont le recrutement semble dépendre de la taille des séquences et donc de leur ancienneté dans le génome. Pour autant il reste encore difficile de comprendre comment les séquences les plus courtes sont reconnues et éliminées. On suppose que d'autres voies encore non identifiées pourraient être impliquées.

4.5 De nouveaux acteurs pour comprendre comment la machinerie d'excision est ciblée à l'ensemble des séquences ?

Dans l'état actuel de nos connaissances, il apparaît que l'élimination d'ADN dépend en réalité d'une multiplicité de voies qui se chevauchent pour permettre de cibler correctement la machinerie d'excision aux séquences germinales. Pour autant, seul un petit nombre d'acteurs est connu et l'élimination d'un certain nombre d'IES, les plus petites et anciennes, est encore mal compris. Chez les eucaryotes, la formation d'hétérochromatine médiée par les petits ARN partage de fortes similarités avec notre processus d'élimination d'ADN chez la paramécie à la différence qu'au lieu d'être seulement transcriptionnellement inactivées les séquences sont physiquement éliminées. En considérant de plus les résultats obtenus chez le cilié *Tetrahymena*, nous sommes parvenus à un premier modèle expliquant l'élimination de la majorité des séquences germinales (Figure 12). Les scnARN une fois sélectionnés pour les séquences germinales permettent leur reconnaissance dans le génome du nouveau MAC et promeuvent le dépôt d'H3K9me3 et H3K27me3 sur les séquences à exciser. Ces modifications recruteraient et guideraient la machinerie d'excision sur les séquences marquées, mais si ce modèle explique l'élimination de la majorité des séquences, une portion des IES ne semble pas en dépendre.

L'élimination d'ADN requièrent une variété de complexes protéiques pour leur élimination illustrée par l'existence de différentes voies chevauchantes. Si l'ensemble des séquences éliminées imprécisément dépendent à la fois des scnARN et des modifications d'histones, l'élimination de certaines IES ne dépend que d'Ez1 (60%) et pour une portion (30%), les plus petites, elles ne dépendent ni des scnARN ni d'Ez1 (Lhuillier-Akakpo et al., 2014). Il reste à comprendre comment la machinerie d'excision est ciblée pour ces 30% d'IES et comment de manière générale les 45,000 IES sont excisées avec une telle précision au nucléotide près. Les modifications d'histone n'offrent qu'une résolution de l'ordre du nucléosome compatible avec l'élimination des éléments répétés et transposables d'une taille largement supérieure à celle du nucléosome et dont l'élimination ne semble pas nécessiter de précision de l'ordre du nucléotide. Les plus petites IES ont une taille inférieure à celle des linker d'ADN, jusqu'à 26pb. Il a ainsi été envisagé que la méthylation de l'ADN qui elle offre une résolution de l'ordre du nucléotide puisse constituer un marqueur des séquences à éliminer et/ou de leurs bornes (Coyne et al., 2012). Elle pourrait ainsi expliquer la

reconnaissance et l'élimination des séquences le plus courtes et pourrait constituer la voie d'élimination « par défaut ». Par ailleurs la chromatine constitue une barrière freinant l'interaction de la machinerie d'excision avec l'ADN. Il est ainsi très probable que d'autres acteurs promeuvent une conformation de la chromatine plus relâchée et compatible avec l'interaction de la machinerie d'excision avec l'ADN aux bornes des séquences germinales.

Objectifs de thèse :

- 1) Déterminer le rôle d'un homologue du chaperon d'histone Spt16 dans les réarrangements programmés d'ADN chez *Paramecium tetraurelia*

Chez *Paramecium tetraurelia*, nous avons identifié un homologue du chaperon d'histone Spt16 exprimé spécifiquement pendant le développement macronucléaire. Partant de ces constatations, nous avons fait l'hypothèse que Spt16 chez *P. tetraurelia* puisse avoir un rôle dans les réarrangements d'ADN. Pour tester notre hypothèse, nous avons analysé les effets de l'inactivation de SPT16 sur le développement macronucléaire et l'élimination d'ADN et cherché à comprendre à quel moment du développement Spt16 était requis. Ce travail nous a conduit à identifier un nouvel acteur majeur des réarrangements du génome chez *Paramecium tetraurelia*.

- 2) Décrire la méthylation des adénines chez *Paramecium tetraurelia*

Nous avons vu précédemment que *P. tetraurelia* était un eucaryote particulièrement riche en adénines méthylées et constituait un modèle de choix pour l'étude de la méthylation chez les eucaryotes. Nous avons émis l'hypothèse que la méthylation puisse permettre de définir les séquences à éliminer. Nous avons analysé la présence des différentes modifications dans l'ADN des paramécies grâce à des techniques d'HPLC-MS et décrit la méthylation des adénines dans la cellule grâce à un anticorps spécifique. Nous avons tenté de localiser sur le génome la méthylation des adénines par la technique de MedIP. Nos résultats préliminaires confirment la forte abondance de la méthylation des adénines dans le génome des paramécies et son établissement au cours du développement du macronoyau.

Résultats

Partie I : Le chaperon d'histone Spt16 chez *Paramecium tetraurelia* est un composant essentiel des réarrangements programmés du génome

Article de recherche

Dans cette étude, nous avons identifié un homologue du chaperon d'histone Spt16 chez *Paramecium tetraurelia* impliqué dans les réarrangements du génome. Spt16-1 est exprimé spécifiquement pendant le développement et localisé dans les nouveaux macronoyaux. Bien que divergents de ses homologues eucaryotes, la protéine Spt16-1 conserve les domaines fonctionnels classiquement retrouvés dans le chaperon d'histones Spt16. Nous avons montré que cette protéine est requise, comme l'endonucléase Pgm, pour l'élimination de l'ensemble des séquences MIC spécifiques : les 45000 IESs et l'ensemble des éléments répétés éliminés imprécisément. En revanche, elle n'est pas nécessaire à la biogénèse et dynamique des scnARN, ni au dépôt des modifications post-traductionnelles d'histones H3K9me3 et H3K27me3 dans les macronoyaux en développement. Spt16-1 est cependant requise pour la production des iesARN, la formation de foci d'H3K9me3 et H3K27me3 et leur disparition. Nous avons conclu que Spt16-1 agissait indépendamment des scnARNs ou de la déposition des marques d'hétérochromatine H3K9me3 et H3K27me3. L'analyse des localisations réciproques de Pgm et Spt16-1 dans les conditions où l'une ou l'autre des deux protéines est déplétée montre que Pgm n'est pas requise pour la localisation de Spt16-1, mais que Spt16-1 est requise pour la localisation de Pgm. Par Western Blot, nous avons confirmé que l'expression de Pgm n'était pas altérée par l'absence de Spt16-1. Nous avons

conclu que Spt16-1 agissait en amont de l'endonucléase Pgm et permet sa localisation dans les MACs en développement.

L'ensemble de nos résultats montre que le chaperon d'histone putatif Spt16-1 est un acteur majeur des réarrangements du génome et est requis pour permettre une activité correcte de l'endonucléase Pgm dans les MACs en développement. Compte tenu du rôle de chaperon d'histones de Spt16 dans d'autres espèces eucaryotes, nous imaginons que Spt16-1 puisse promouvoir une conformation plus ouverte de la chromatine permettant à la machinerie d'excision et à Pgm d'interagir avec l'ADN. Nous avons ainsi mis en évidence chez *P. tetraurelia* une protéine Spt16 spécifique du développement, avec des fonctions peu ou pas explorées chez ce chaperon d'histones parmi les autres eucaryotes : son implication dans les réarrangements programmés du génome et sa capacité à permettre la localisation nucléaire d'une autre protéine.

***Paramecium tetraurelia* histone chaperone Spt16 is an essential component of programmed genome rearrangements**

Amandine Touzeau¹, Augustin de Vanssay¹, Olivier Arnaiz², Andrea Frapporti¹ and Sandra Duharcourt¹

1- Institut Jacques Monod, CNRS UMR7592, Université Paris-Diderot, Paris, France

2- Institut de Biologie Intégrative de la Cellule (I2BC), UMR9198 CNRS CEA Université Paris-Sud, Gif-sur-Yvette, France

ABSTRACT

The highly conserved histone chaperone Spt16 is involved in a variety of processes such as transcription, replication or repair. In *P. tetraurelia*, a large proportion of the germline genome is reproducibly removed from the somatic genome after sexual events via a multi-step process that involves small RNA-directed heterochromatin formation followed by DNA excision and repair. The molecular mechanisms underlying the specific recognition of germline limited DNA sequences in the context of chromatin remain elusive. Here, we identify *P. tetraurelia* Spt16-1, a homolog of the Spt16 subunit of the FACT complex, as an essential player of programmed genome rearrangements. A GFP-Spt16-1 fusion protein was shown to exclusively localize in the nuclei where genome rearrangements take place. Gene silencing experiments showed that Spt16-1 is required for the elimination of all germline-limited sequences and for the survival of the sexual progeny. *SPT16-1* knockdown did not impair the production of 25nt-long scanRNAs nor the deposition of heterochromatin H3K9me3 and H3K27me3 histone marks, indicating that Spt16-1 acts downstream the small RNA-directed heterochromatin pathway. We further showed that Spt16-1 is required for the correct nuclear localization of the Pgm endonuclease responsible for the introduction of DNA double-strand breaks. We propose a model in which Spt16-1 mediates interactions between the excision machinery and chromatin that facilitate access to DNA cleavage sites for the Pgm endonuclease during genome rearrangements.

INTRODUCTION

Chromatin organization maintains genome stability in eukaryotes. Key players of the maintenance of this organization are histone chaperones, proteins that bind highly charged and hydrophobic surfaces of histone proteins (Hammond et al., 2017). Chaperones cooperate to promote assembly or disassembly of nucleosomes and allow access to DNA or maintain a repressive and stable chromatin barrier in a large variety of processes (Adam et al., 2015; Eitoku et al., 2008; Hammond et al., 2017; Ransom et al., 2010; Venkatesh and Workman, 2015).

The highly conserved histone chaperone FACT is a heterodimer composed of two essential proteins Spt16 and Ssrp1 (Pob3 in *S. pombe*) that interact with free histones or nucleosomes (Formosa, 2012; Jeronimo and Robert, 2016). The domain organization of these proteins is highly conserved among eukaryotes and allows multiple contacts with H2A-H2B dimer and H3-H4 tetramer in the nucleosome (Hondele et al., 2013; Hsieh et al., 2013; Kemble et al., 2013, 2015; VanDemark et al., 2008). *In vitro* analyses of isolated Spt16 or Ssrp1/Pob3 domains provide models to interpret the interaction of the FACT complex with the nucleosome (Hondele et al., 2013). The Spt16 and Pob3 proteins are thought to compete with DNA to interact with histones and promote an open configuration of the nucleosome. This chromatin organization would allow machineries to interact with DNA and if necessary could remove or replace histones (Hondele et al., 2013; Hsieh et al., 2013; Jeronimo and Robert, 2016; Kemble et al., 2015).

In *S. cerevisiae* or *H. sapiens*, mutations of both Spt16 and Pob3/Ssrp1 lead to defects that have been linked to replication or transcription (Malone et al., 1991; Prendergast et al., 1990; Schlesinger and Formosa, 2000). However, it seems that the FACT complex is more widely involved in regulating the access to DNA or chromatin. FACT indeed regulates transcription initiation allowing transcription factors and RNA polymerase II to interact with DNA by disassembling nucleosomes (Biswas et al., 2005; Jeronimo et al., 2015). Spt16 has also been identified as a transcription elongation factor allowing RNA polymerase II to process DNA through chromatin by disassembling and reassembling nucleosomes (Belotserkovskaya et al., 2003; Biswas et al., 2005; Hsieh et al., 2013; Jamai et al., 2009; Orphanides et al., 1998; True et al., 2016). Spt16 through the FACT complex is an essential replication factor required both for initiation and elongation, physically interacting with

subunit Pol1 of DNA polymerase alpha (Abe et al., 2011; Gambus et al., 2006; Han et al., 2010). Not only it ensures correct assembly and deposition of nucleosomes on new replicated DNA, but FACT also promotes accessibility to DNA for the X polymerase and the MCM helicase (Yang et al., 2016). In addition to these most extensively studied roles, FACT has been shown to be required for the recruitment of diverse chromatin proteins or repair machineries (Frost et al., 2018; Heo et al., 2008; Jeronimo et al., 2015; Prendergast et al., 2016). A recurrent theme in all these processes is the ability of FACT complex to regulate access to DNA and chromatin by relaxing the nucleosome. In this report, we use *Paramecium tetraurelia* as a model to study the involvement of Spt16 in developmental programmed DNA elimination, which requires heterochromatin formation and access to DNA for the elimination and repair machineries.

Paramecium tetraurelia, like in all ciliates, separates germline and somatic functions in two distinct nuclei within the same cytoplasm. Germline functions are supported by two small, diploid micronuclei (MICs) that are transcriptionally silent during vegetative growth. The highly polyploid somatic macronucleus (MAC) ensures gene expression and is destroyed at each sexual cycle (Betermier and Duharcourt, 2014). During sexual events, the MICs undergo meiosis and transmit the germline genome to the zygotic nucleus. New MICs and new MACs develop from mitotic copies of the zygotic nucleus. Development of the new MAC involves massive and highly reproducible genome rearrangements. It includes the elimination of approximately 30 Mbp of germline DNA that represents a third of the complexity of the germline genome (Guérin et al., 2017). In addition to the imprecise elimination of large genomic regions, up to several kb in length, containing repeated sequences, 45,000 short, single-copy Internal Eliminated Sequences (IESs) (3.5 Mbp) are precisely excised from intergenic and coding regions (Arnaiz et al., 2012). DNA elimination is initiated by DNA double strand breaks (DSB) mediated by the domesticated transposase PiggyMac (Pgm) (Baudry et al., 2009; Dubois et al., 2017). The DSB are then repaired precisely by the classical non-homologous repair pathway (Kapusta et al., 2011; Marmignon et al., 2014). The question of how so many diverse sequences are recognized and excised remains elusive. Different classes of small RNAs are required for elimination of most germline-limited sequences. The 25 nt-long scanRNAs are produced from the germline genome during meiosis by Dicer-like proteins (Lepère et al., 2008, 2009; Lhuillier-Akakpo et al., 2014; Sandoval et al.,

2014). They are loaded onto Piwi proteins and transported to the maternal MAC, where selection of MIC-specific scnRNAs is thought to take place (Furrer et al., 2017; Lepère et al., 2009). scanRNAs specific to germline sequences then guide the deposition of histone H3 post-translational modifications (H3K9me3 and H3K27me3) in the developing MAC and tether the Pgm elimination machinery to specific-germline sequences (Coyne et al., 2012; Duharcourt et al., 2009; Furrer et al., 2017) (Lhuillier-Akakpo et al., 2014). A second class of developmental-specific sRNAs, the iesRNAs, which map exclusively to IESs are 26 and 31 nt in length, is likely produced in the developing MAC after excision by another Dicer-like protein (Sandoval et al., 2014). The iesRNAs are thought to participate in a positive feedback that allows efficient IES excision (Sandoval et al., 2014).

We report here the identification and functional analysis of a novel major actor of genome rearrangements in *Paramecium tetraurelia*. *P. tetraurelia* Spt16-1 is a homolog of the Spt16 subunit of the FACT complex, whose expression is upregulated during MAC development, and is specifically localized in the new developing MAC. Resequencing the genome upon *SPT16-1* knockdown showed that Spt16-1 is required for the elimination of all germline specific sequences. Spt16-1 is required neither for the production of scanRNAs nor for the deposition of H3K9me3 or H3K27me3, indicating that it acts in parallel or downstream of the small RNA-directed heterochromatin pathway. We further showed that Spt16-1 is required for the correct nuclear localization of Pgm in the new developing MAC but not for its expression, suggesting that Spt16-1 is indispensable for the association of the Pgm elimination machinery to chromatin.

RESULTS

Identification of *Paramecium tetraurelia* Spt16 homologs

In order to find putative *Paramecium* homologs, blast searches against the *Paramecium tetraurelia* somatic genome were performed using the human or yeast SPT16 protein sequence as a query (Arnaiz and Sperling, 2011). We identified four genes encoding Spt16-like proteins grouped in two distinct families (Figure 1). Three proteins (Spt16 2a, 2b and 2c) belong to the same family and are close paralogs issued from whole genome duplications of the *Paramecium* genome (Aury et al., 2006) (Figure 1B-C). Spt16-1 is the only member of the most evolutionary distant

family that shares only 33% amino acid identity (Figure 1B). In contrast to most other eukaryotes with a single Spt16 representative, *Paramecium* species harbor two Spt16 families. In *P. caudatum* and *P. polycarium*, two genes are found, one in each family, while in other *Paramecium aurelia* species, four paralogs are found in the Spt16-2 family (Figure S1). In all *Paramecium* species however, Spt16-1 is encoded by a single gene.

The four *P. tetraurelia* Spt16 proteins harbor characteristic domains of Spt16 (Figure 1A): the N domain, composed of a lobe domain and of a non-catalytic peptidase M24 domain, which is non-essential but known to interact genetically with histones H2A and H3-H4; the dimerization D domain, the M domain homologous to Rtt106, known to interact with H3-H4, and the C domain that contains the Minimal Binding Domain (MBD) required for H2A-H2B interaction (Hondele et al., 2013; Kemble et al., 2013, 2015; VanDemark et al., 2008). The aromatic and acidic amino acids required for interaction with histones in MBD appear to be conserved in *P. tetraurelia* (Figure S2). The domain organization is thus conserved in all four *Paramecium* Spt16 proteins.

The expression of *SPT16* genes was examined using RNA-Seq data generated during *P. tetraurelia* life cycle (Arnaiz et al., 2017) (Figure 1C). *SPT16-2a*, *2b* and *2c* are expressed constitutively at significant levels, during vegetative growth and during the sexual process of auto-fertilization (autogamy) (Figure 1C). In contrast, no expression is detected during vegetative growth for *SPT16-1*, while it is specifically upregulated during autogamy, in a pattern that is very similar to that of the *PGM* gene (Baudry et al., 2009) (Figure 1C). The induction of its expression coincides with development of the new MAC after sexual events. Based on this peculiar expression profile, we decided to focus on Spt16-1.

Spt16-1 is required for post-zygotic development

To test whether Spt16-1 is required during sexual events when its expression is specifically induced, we knocked-down the *SPT16-1* gene by RNA interference during autogamy using the “feeding” technique, which consists of feeding *Paramecium* with *Escherichia coli* engineered to produce dsRNA from the target gene (Galvani and Sperling, 2002) (Figure 2A). Upon *SPT16-1* silencing, autogamy started normally and cells eventually harbored two new developing MACs, as expected. Yet post-autogamous cells were unable to resume vegetative growth when returned to normal food medium, as observed upon *PGM* silencing (Figure 2A). As

for *PGM* silencing, post-autogamous cells displayed two large macronuclei and died before the first cellular division (Figure 2B). To control for possible off-targets, two non-overlapping fragments of the *SPT16-1* gene were used to generate constructs for *SPT16-1* RNAi and both constructs gave similar results (6% of sexual progeny is viable) (Figure 2A).

The lethality observed in the sexual progeny could, in principle, be due to a defect in any events that occur during autogamy: MIC meiosis, karyogamy, new MAC development. Given its expression profile, we decided to knockdown *SPT16-1* specifically during the developmental window that corresponds to new MAC differentiation. To reach this goal, we used conjugation, a cross-fertilization process between two mating partners that allows us to initiate *SPT16-1* silencing when conjugating pairs separate, which coincides with the onset of MAC development (Figure 2C) (Berger, 1973). Reactive cells from opposing mating types were crossed in standard food medium then the mating pairs were transferred to *SPT16-1* silencing medium to allow pre-zygotic events as well as karyogamy and mitoses of the zygotic nucleus to occur normally before cells begin to ingest *SPT16-1* dsRNA from bacteria as soon as the conjugating pairs separate. The next day, cells were transferred to standard food medium to resume vegetative growth. Strong lethality was observed in *SPT16-1* knockdown, as for *PGM* knockdown used as positive control, while no effect was observed in the control RNAi as expected (Figure 2C). We conclude that *Spt16-1*, like *Pgm*, is essential for MAC development and production of viable sexual progeny.

Spt16-1 is localized in developing new somatic macronuclei

To gain further insight into the role of *Spt16-1*, we examined the subcellular localization of the *Spt16-1* protein. We constructed a *GFP-SPT16-1* transgene, in which the GFP coding sequence was inserted into the *SPT16-1* coding sequence, immediately downstream the start codon. The transgene was designed to be RNAi-resistant (Materials and Methods). Expression of the fusion gene was under the control of the natural *SPT16-1* up- and downstream sequences. Following microinjection of the transgene into the MAC of vegetative cells, no detectable GFP fluorescence was observed in vegetative cells or during the sexual events of autogamy (Figure 3A). Consistent with *SPT16-1* mRNA expression profile, GFP fluorescence accumulated in the new developing MAC during autogamy, where

genome rearrangements take place (Figure 3A). To check whether the GFP-Spt16-1 fusion protein is functional, we performed an *in vivo* complementation assay. No lethality was observed in the post-autogamous progeny of injected cells, indicating that the GFP-Spt16-1 fusions did not interfere with normal progression of autogamy. Transformed cells were submitted to *SPT16-1* RNAi to deplete the endogenous Spt16-1 protein. In non-injected cells, strong lethality (21% of viable progeny) was observed in sexual progeny, indicating that the endogenous *SPT16-1* gene was efficiently silenced. High recovery of viability of the sexual progeny (83%) was observed in the *GFP-SPT16-1* transformed cells, indicating that the fusion protein complements the lethality caused by the depletion of the endogenous Spt16-1 protein. GFP-Spt16-1 fusion protein is detected in new developing MACs of cells inactivated for *SPT16-1* (Figure S3). Thus, the RNAi-resistant *GFP-SPT16-1* transgene encodes a functional protein that is exclusively localized in the developing MAC.

Spt16-1 is essential for programmed genome rearrangements

To get insight into the role of Spt16-1 during MAC development, we assessed whether it is required for programmed genome rearrangements. We first investigated the role of the *SPT16-1* gene in IES excision. Excision was analyzed by PCR on genomic DNA, extracted after *SPT16-1*, *PGM*, or control silencing at a time when IES excision is normally finished. In control RNAi experiments, the 7 IESs analyzed were completely excised from the new developing MACs, as expected. In contrast, IES-retaining forms accumulated in the new MACs of *PGM* or *SPT16-1* knockdown cells (Figure 4A). Similar results were obtained for both *SPT16-1* silencing constructs (Figure 4A). The amplification of non-rearranged DNA indicates that programmed genome rearrangements do not proceed normally after Spt16-1 depletion. Because of the presence of rearranged forms in the fragments of the maternal MAC, this approach makes it difficult to assess whether new excision junctions are formed in the new developing MAC. We took advantage of a cell line (delta A) that has a wild type MIC genome, but carries a maternally inherited deletion of the non-essential A gene in the MAC. During autogamy, all IESs are normally excised in this cell line before the entire locus is deleted in the new MAC. Using PCR primers located on flanking sequences of two IESs located in the A locus, transient excision junctions corresponding to *de novo* IES excision could be detected, as expected (Figure 4B).

In contrast, no excision junctions could be detected in *SPT16-1* RNAi (Figure 4B). To confirm that excision is impaired in Spt16-1 depleted cells, we monitored the appearance of excised IES circular molecules during autogamy. Covalently closed circles are formed from long excised linear IESs, following Pgm-dependent DNA cleavages at each IES boundary (Bétermier et al., 2000). Using divergent PCR primers internal to IES 51A4578, we could transiently detect IES circles in control RNAi, as expected, but failed to detect any in cells depleted for Spt16-1 (Figure 4C). These data indicate that Spt16-1 is essential for the recovery of both chromosomal junctions and excised IES circles.

In addition, we noted that the non-rearranged forms of the A locus accumulated at late time points during autogamy in Spt16-1-depleted cells (Figure 4B). Thus, maternally inherited deletion of the A gene is blocked in *SPT16-1* silenced cells, and the non-rearranged germline locus is retained in the developing new MACs, a phenotype that is very similar to that observed when DNA cleavage is inhibited (Baudry et al., 2009; Marmignon et al., 2014).

To expand these results genome wide, we performed high-throughput sequencing of DNA extracted from a nuclear preparation enriched for new MACs of *SPT16-1*-silenced autogamous cells. We checked that, in this autogamy time-course experiment, the levels of poly-adenylated mRNA corresponding to the *SPT16-1* gene were reduced upon *SPT16-1* RNAi at early stages of MAC development, when *SPT16-1* expression is normally switched on in control RNAi (data not shown). We next quantified the effects of *SPT16-1* knockdown on IES retention. We used sequencing data for the DNA sample obtained from autogamous cells of the same strain, but with no silencing, as a control (Lhuillier-Akakpo et al., 2014). Excision of all IESs is altered and 98.6% (44,334) of the IESs are significantly retained in the developing MAC after *SPT16-1* RNAi, very similar to 99% (44,491) observed after *PGM* RNAi (Denby Wilkes et al., 2015). Using the boundary score method (Denby Wilkes et al., 2015) we measured IES retention of each IES boundary. As expected, in control, the score is 0 since all IESs are correctly excised. In *SPT16-1* RNAi, similar Gaussian distributions for each boundary are observed with a mean retention of 0.44, indicating that retention of two boundaries is highly correlated, as for *PGM* RNAi (Figure 4D). Excision is an efficient, robust process but rare excision errors can be detected (Duret et al., 2008). We catalogued these error events (deletion of a somatic DNA fragment, use of alternative boundary during IES elimination) and

evaluated their frequency as described in (Denby Wilkes et al., 2015). No difference could be observed between cells depleted for Spt16-1, Pgm or control cells (Figure S4).

To estimate the effects of Spt16-1 depletion on the retention of germline-limited sequences other than IESs, we compared the sequencing read coverage of different genomic compartments in *SPT16-1* and *PGM* RNAi (Arnaiz et al., 2012) (Guérin et al., 2017) (Figure 4E). As expected, the part of the germline genome that is collinear with MAC chromosomes (“MAC-destined”) is similarly covered in all datasets (Figure 4E). In contrast, the germline-limited part of the genome (MIC-limited), which contain repeated sequences including transposable elements, is not covered by MAC reads and is well-covered by MIC reads, PGM and SPT16-1 reads, indicating that MIC-limited sequences are retained in Spt16-1-depleted cells, as in Pgm-depleted cells (Figure 4E). Next, we assessed retention by mapping reads from each dataset to the known cloned copies of the Sardine transposon family. We found that all characterized Sardine copies are retained after *SPT16-1* silencing (Figure 4F). Altogether, these results show that Spt16-1, like Pgm, is essential for all DNA elimination events.

Spt16-1 is required for iesRNAs but not for scanRNAs accumulation

To determine at which step Spt16-1 is involved in the process of DNA elimination, we examined whether Spt16 depletion affects small RNA levels. We used high throughput sRNA sequencing to compare sRNA populations present in control and *SPT16-1*-silenced cells at three time-points during autogamy: precocious (app. 20% of cell population undergoes meiosis), early (10h after the first time point), and middle (20 h after the first time point) (Figure 5A). At early autogamy, 25 nt-scanRNAs represent the majority of sRNAs in the control sample, as reported previously (Gruchota et al., 2017; Lepère et al., 2009; Sandoval et al., 2014). The same is true in *SPT16-1* RNAi, indicating that Spt16-1 is not required for scanRNAs accumulation. As MAC development proceeds in control and *SPT16-1* RNAi, the proportion of scanRNAs corresponding to MAC sequences decreases, indicating that selection of MIC-specific scnRNAs occurs normally in Spt16-1-depleted cells. In contrast to control RNAi, *SPT16-1* RNAi almost completely abrogated iesRNA production (Figure 5A). This is consistent with the fact that iesRNA formation requires DNA elimination (Allen et al., 2017). Thus, the *SPT16-1* knockdown RNA sequencing

showed that Spt16-1 is essential for iesRNA accumulation but does not affect the biogenesis nor the process of selection of scanRNA.

Spt16-1 is required for H3K27me3 and H3K9me3 foci formation

Next, we investigated whether Spt16-1 depletion affects H3K9me3 and H3K27me3 silent heterochromatin marks. To this end, we performed indirect immunostaining experiments using H3K9me3 and H3K27me3 specific antibodies in control, Pgm and Spt16-1-depleted cells during autogamy. As previously reported, in control RNAi, H3K9me3 and H3K27me3 signals are detected in new developing MACs with a diffuse pattern that gradually formed nuclear foci before the signals eventually disappear (Figures 5B and S5) (Lhuillier-Akakpo et al., 2014). Depletion of Spt16-1 or Pgm do not affect the biogenesis of H3K27me3 and H3K9me3 in the developing MACs (Figures 5B and S5). This suggests that the Spt16-1 protein, like Pgm, acts downstream of H3K27me3 and H3K9me3 deposition. Yet H3K27me3 and H3K9me3 signals remained diffuse as development proceeds in Spt16-1 and Pgm-depleted cells and no foci could be detected (Figures 5B and S5). Spt16-1, like Pgm, is thus required for H3K27me3 and H3K9me3 foci formation, whose biological significance remains unclear. Based on our observations, we suggest that H3K9me3-H3K27me3- nuclear bodies are likely formed as a consequence of DNA double strand break formation. Altogether, our data suggest Spt16-1 acts downstream the scanRNA-directed heterochromatin pathway.

Spt16-1 is essential for the correct nuclear localization of the Pgm endonuclease

In order to determine whether Spt16-1 acts before or after the Pgm endonuclease in the process of programmed DNA elimination, we examined Pgm localization in Spt16-1-depleted cells, and reciprocally, Spt16-1 localization in Pgm-depleted cells. To examine Pgm localization, we performed immunofluorescence experiments with Pgm antibodies (Dubois et al., 2017) in Spt16-1-depleted cells and in control cells during autogamy. As previously described, Pgm progressively accumulates in the developing new MAC with a peak corresponding to the time when DNA DSBs are introduced at IES ends, and a decrease at later stages of development before it disappears from the developing MAC (Dubois et al., 2017). In control cells, confocal microscopy indicated that Pgm accumulated in the developing new MAC at early

stages of development (T5), as expected (Figure 6A). We quantified the fluorescence signal in the new developing MACs (see Materials and Methods) in the cell population at this early stage of MAC development. Because cells are not synchronized before autogamy, there is some heterogeneity in the cell population that is exemplified by the difference in MAC size. Indeed, as new MACs develop, the size of the nucleus enlarges, and this can be used as a proxy for developmental time. When we plot the fluorescence intensity as a function of new MAC area, we observe a progressive increase in fluorescence intensity that positively correlates with nucleus area (20-50 μm^2), confirming that the Pgm protein accumulates in the new MAC as it develops (Figures 6A and S6). In sharp contrast with control cells, we could not detect Pgm accumulation in the developing new MAC within the same nucleus size range, in the majority of Spt16-1-depleted cells (70%; 42/60) (Figures 6A and S6). We conclude that Spt16-1 is required for the accumulation of the Pgm protein in the developing new MAC.

We noticed that Pgm could still be detected for a small proportion of cells (30%), with a similar intensity as in control cells (Figures 6A and S6). Even though we cannot formally exclude that it may reflect an incomplete knockdown of *SPT16-1*, it appears rather unlikely to be the case since our molecular analyses of genome rearrangements indicated that DNA elimination is completely blocked in Spt16-1-depleted cells. We rather favor the idea that the Pgm protein may be detected in the nucleus in a subset of Spt16-1-depleted cells.

We wanted to rule out that the dramatic effect of Spt16-1 depletion on Pgm nuclear accumulation was due to the fact that Spt16-1 impacts the production of the Pgm protein itself. To evaluate Pgm protein levels, we performed Western blot analysis of total protein extracts at different time points during autogamy of control and Spt16-1-depleted cells, using Pgm antibodies. The Pgm protein was shown to progressively accumulate with a T5-T10 peak at similar levels in both control and Spt16-1-depleted cells (Figure 6B). Altogether these data showed that Spt16-1 does not affect the production of the Pgm protein, and instead, impair its accumulation in the developing new MAC.

Conversely, we examined Spt16-1 localization in Pgm-depleted cells. Cells transformed with the GFP-*SPT16-1* transgene were subjected to *PGM* or control RNAi during autogamy. In the control, GFP-Spt16-1 appeared in the developing new MAC and persisted throughout MAC development, as expected (Figure 6C). In *PGM*

RNAi, the fusion protein also localized in the developing new MAC (Figure 6C). This was further confirmed by quantification of GFP fluorescence from confocal images in both conditions (see Materials and Methods) (Figures 6C and S6). We noted that, in a fraction of Pgm-depleted cells (42,5%), a higher Spt16-1 fluorescence intensity is detected in the developing new MAC. This appears to be simply due to a larger nucleus area (Figures 6C and S6), because there is no difference between control and Pgm-depleted cells when the fluorescence intensity is normalized by the nucleus area (Figure 6C). These data indicate that the histone chaperone Spt16-1 predominantly localizes in the nucleus, independently of the presence of Pgm.

DISCUSSION

The histone chaperone Spt16 is conserved in most eukaryotes (Formosa, 2012). It is generally encoded by a single gene. During evolution, two families of *SPT16* genes have emerged in *Paramecium* (Figure 1). Family 1 is composed of a single gene, *SPT16-1*. Family 2 comprises one to four genes depending on the species, which may be explained by whole genome duplications that have arisen in the *Paramecium aurelia* species complex (Aury et al., 2006). All *Paramecium* proteins contain the four canonical domains found in Spt16 and represent the most divergent group. In *P. tetraurelia*, all members of Family 2 are constitutively expressed, while *SPT16-1* is specifically expressed during development. Given their distinct expression profiles, the two *SPT16* gene families may exert specialized functions. In this study, we showed that *SPT16-1* is essential for post zygotic development and assembly of the new MAC genome, where the Spt16-1 protein specifically accumulates (Figures 2, 3 and 4).

The functional analysis of *SPT16-1* demonstrated its role in developmental programmed genome rearrangements. By genome resequencing upon Spt16-1 depletion, we showed that Spt16-1 is required for all types of DNA elimination events: the precise excision of IESs, the imprecise elimination of repeated sequences and transposable elements, and the maternal inheritance of gene deletions (Figure 4). We did not find any difference between SPT16 and PGM datasets: in particular, no change in the frequency or in the types of excision errors could be detected (Figure S4). Previous work revealed the existence of partially overlapping pathways involved

in IES excision: some factors are required for the excision of a subset of IESs, while other factors are required for the excision of yet another overlapping subset of IESs (Furrer et al., 2017; Gruchota et al., 2017; Lhuillier-Akakpo et al., 2014; Maliszewska-Olejniczak et al., 2015; Sandoval et al., 2014). Spt16-1 does not belong to any of these specialized pathways required for the excision of subset of IESs but instead, appears to be a generic factor, which, like Pgm, is required for the elimination of all IESs (Arnaiz et al., 2012; Baudry et al., 2009)

The current model to explain the elimination of the vast majority of MIC-specific sequences proposes that 25 nt-long scnRNA produced by the entire MIC genome during meiosis are transferred from the maternal MAC where they are enriched in MIC-specific sequences to the new MAC where they lead to the deposition of post-translational modifications of histone H3 (H3K9me3 and H3K27me3) on homologous sequences and eventually trigger DNA elimination. High throughput sequencing of small RNA populations at different time points during sexual events and during MAC development showed that Spt16-1 is not required for scnRNA production nor for the selection process that allows enrichment in MIC-specific sequences (Figure 5). We further showed that Spt16-1 is not required for the deposition of H3K9me3 and H3K27me3 in the developing new MAC (Figure 5), similarly to Pgm (Lhuillier-Akakpo et al., 2014). Our analyses thus indicate that Spt16-1, like Pgm, is a downstream factor of the scanRNA-mediated heterochromatin pathway.

Consistent with Spt16-1 being essential for DNA elimination, 26-29 nt long iesRNAs that are produced in the new MAC from excised IESs and appear involved to complete IES excision (Allen et al., 2017; Sandoval et al., 2014) were not detected in Spt16-1-depleted cells. Heterochromatin H3K27me3 and H3K9me3 marks normally display a dynamic pattern of localization during MAC development: the staining observed in immunofluorescence experiments with specific antibodies is first diffuse then reorganizes into nuclear foci that collapse into a single focus before it eventually disappears (Lhuillier-Akakpo et al., 2014). We previously showed that Pgm is required for the formation of these nuclear bodies (Lhuillier-Akakpo et al., 2014). Similarly, we showed here that Spt16-1 was essential for their dynamic reorganization (Figure 5). The biological significance of the formation and aggregation of H3K9me3/H3K27me3 nuclear bodies is unclear. This is reminiscent of phase separation, a phenomenon that gives rise to non-membrane compartments that is suggested to be involved in heterochromatin formation (Larson et al., 2017;

Strom et al., 2017). The late timing of their formation relative to DNA elimination, their dependency toward Pgm and Spt16-1 factors suggest that H3K9me3/H3K27me3 nuclear bodies aggregate once DNA elimination events are completed. One could imagine that MIC-specific DNA associated with the factors needed for elimination (proteins and non-coding RNAs) are aggregated after excision from the chromosomes before degradation in the nucleus.

We showed that Spt16-1 is critical for the correct accumulation of the Pgm protein in the developing new MAC (Figure 6), providing an explanation for why DNA elimination is impaired in Spt16-1-depleted cells. We ruled out the possibility that Spt16-1 is involved in the production of the Pgm protein. Indeed, we showed, by Western blot analysis of whole cells extracts, that the levels of the Pgm protein are not affected upon Spt16-1 depletion (Figure 6). Our quantitative image analysis of localization experiments showed that, at steady state, the Pgm and Spt16-1 proteins both accumulate in the nucleus in normal conditions. However, a drastic reduction of the Pgm nuclear signal was observed in the absence of Spt16-1 (Figure 6). Conversely, Spt16-1 still accumulated in the nucleus, in the absence of Pgm (Figure 6). Altogether our data indicate that Spt16-1 impacts Pgm localization and presumably, Pgm predominantly accumulates in the cytoplasm in the absence of Spt16-1. Our inability to detect a cytoplasmic Pgm localization by immunolocalization may be easily explained by the diffusion of the signal in the large volume occupied by the cytoplasm in *P. tetraurelia* cells (130 μm large, 30 μm width).

Pgm accumulation in the nucleus of some Spt16-depleted cells appears to contradict our molecular analyses showing defective DNA elimination under this condition (Figure 4). In Spt16-1-depleted cells, the absence of precise *de novo* IES excision junctions could either reflect an inhibition of DNA cleavage, as shown for Pgm depletion (Baudry et al., 2009) or a problem in DSB repair, as established for Ligase IV depletion (Kapusta et al., 2011). In Ligase IV-depleted cells, however, the developing new MACs display a faint DAPI staining, likely due to lack of endoreplication caused by the accumulation of non-repaired DSBs (Kapusta et al., 2011). In Spt16-1-depleted cells, as in Pgm-depleted cells, the Hoechst staining of the developing new MACs appears very similar to normal cells (Figure 2). The similarities between the phenotypes observed in Pgm- and Spt16-1- depleted cells strongly support the idea that Spt16-1 depletion inhibits DNA cleavage, despite accumulation of Pgm.

Based on our localization observations, we hypothesize that Pgm is a nucleocytoplasmic shuttling protein constantly moving in and out of the nucleus. At steady-state, Pgm predominantly accumulates in the nucleus in normal conditions, while it predominantly accumulates in the cytoplasm in absence of Spt16-1 due to changes of the balance between import and export rates. To explain its mode of action, we propose that Spt16-1 mediates, directly or indirectly, the interaction of Pgm to chromatin or DNA, preventing its export from the nucleus. Spt16 in other organisms is well known for its ability to open chromatin and make the DNA accessible to the replication, transcription or repair machineries (Hondele et al., 2013; Kemble et al., 2015; Yang et al., 2016). Similarly, Spt16-1 could interact with histones and rearrange the chromatin to open it, in order for Pgm and the other components of the excision machinery to interact with DNA.

In conclusion, our work establishes that a developmental-specific Spt16-1 is essential for the massive and reproducible programmed genome rearrangements that occur during *Paramecium* development. There is evidence that histone chaperones can play an active role in the import of histones. Histone chaperones have also been shown to regulate the localization of important chromatin modifying enzymes (Keck and Pemberton, 2012). Here, we showed that the Spt16-1 histone chaperone promotes nuclear localization and activity of the Pgm endonuclease, providing a novel mechanism to control nuclear availability of the enzymatic complex responsible for DNA cleavages that orchestrates genome reorganization.

MATERIALS AND METHODS

***Paramecium* strains, cultivation and autogamy**

All experiments were carried out with the entirely homozygous wild type strain 51 of *P. tetraurelia* or with strain 51 cells deleted for genes A and ND7 in the macronucleus. Cells were grown in wheat grass powder (WGP) (Pines International) infusion medium bacterized the day before use with *Klebsiella pneumoniae* and supplemented with 0.8 mg/mL β -sitosterol (Merck). Cultivation and autogamy were carried out at 27 °C (Beisson et al., 2010a).

Gene silencing experiments

Plasmids used for T7Pol-driven dsRNA production in silencing experiments were obtained by cloning PCR products from each gene using plasmid L4440 and *Escherichia coli* strain HT115 DE3, as previously described (Galvani and Sperling, 2002). Sequences used for silencing of *SPT16-2a*, *SPT16-2b*, *SPT16-2c* were segments 1572-2217; 1572–2217 and 1583-2227 of PTET.51.1.G1210069, PTET.51.1.G1050068 and PTET.51.1.G0550178, respectively. For *SPT16-1* silencing, two non-overlapping gene fragments covering positions 1704-2165 (*SPT16_1#1*) and 165-902 (*SPT16_1#2*) of PTET.51.1.G0710091 were used. The fragments used for *ND7* (Garnier et al., 2004), *ICL7a* (Bouhouche et al., 2011), *DCL2*, *DCL3* (Lepère et al., 2009), *PGM-1* (Baudry et al., 2009) and *EZL1-1* (Lhuillier-Akakpo et al., 2014) are those previously published. Preparation of silencing medium and RNAi during autogamy was performed as described in (Baudry et al., 2009). Lethality of post-autogamous cells after silencing of *PGM* in GFP-*Spt16-1* transformed cells was 90-100%. As expected, Pgm depletion led to retention of all IESs tested.

Injection of GFP fusion transgenes

For the construction of in-frame *GFP-SPT16-1* fusion, a GFP coding fragment adapted to *Paramecium* codon usage (Nowacki et al., 2005) was added by PCR fusion to the 5' end of the *SPT16-1* gene. As a result, the GFP is fused to the N-terminus of *SPT16-1* and the fusion protein is expressed under the control of the *SPT16-1* transcription signals (promoter and 3'UTR). It contains the 119-bp genomic region upstream of the *SPT16-1* open reading frame, the 308-bp genomic region downstream. The plasmid carrying the *GFP-SPT16-1* fusion transgene was linearized by *RsrII* and microinjected into the MAC of vegetative cells. No lethality was observed in the post-autogamous progeny of injected cells, indicating that the GFP-*SPT16-1* fusions did not interfere with normal progression of autogamy. For the construction of in-frame RNAi resistant GFP-*SPT16-1* fusion, the original *NheI-SwaI* restriction fragment of the *SPT16-1* coding sequence that comprises the RNAi target region *SPT16-1#2* was replaced by a synthetic DNA sequence (Eurofins Genomics). The fragment was designed to maximize nucleotide sequence divergence with the endogenous genomic locus without modifying the amino acid sequence of the encoded *Spt16-1* protein. Plasmid carrying the RNAi resistant GFP-*SPT16-1* fusion transgene was linearized by *RsrII* and microinjected into the MAC of vegetative cells.

DNA and RNA extraction and PCR, qPCR

DNA samples were typically extracted from 200-400-ml cultures of exponentially growing cells at <1,000 cells/ml or of autogamous cells at $\leq 2,000$ –4,000 cells/ml as previously described (Beisson et al., 2010b). Small-scale DNA samples were prepared from 1,000 cells using the NucleoSpin Tissue kit (Macherey-Nagel). RNA samples were typically extracted from 200–400-ml cultures of exponentially growing cells at <1,000 cells/ml or of autogamous cells at 2,000–4,000 cells/ml as previously described (Beisson et al., 2010b).

PCR amplifications were performed in a final volume of 25 μ L, with 10 pmol of each primer, 10 nmol of each dNTP and 1.9 U of Expand Long Template Enzyme mix (Expand Long Template PCR system, Roche). PCR products were analyzed on 0.8%-3% agarose gels. Oligonucleotides were purchased from Eurofins MWG Operon or SIGMA (see Table S2).

H3K9me3/H3K27me3 immunofluorescence

Cells were fixed for 30 minutes in solution I (10 mM EGTA, 25 mM HEPES, 2 mM MgCl₂, 60m M PIPES pH 6.9 (PHEM 1X); paraformaldehyde 1%, Triton X-100 2.5%, Sucrose 4%) and for 10 minutes in solution II (PHEM 1X, paraformaldehyde 6.5%, Triton X-100 1.2%, Sucrose 4%). The primary antibodies used were rabbit polyclonal α -H3K27me3 (07-449, Millipore) and α -H3K9me3 (07-442, Millipore) at 1:500 dilution. After incubation with the primary antibodies, cells were washed in 1X phosphate-buffered saline (PBS), incubated with the secondary antibodies (Alexa Fluor 568-conjugated goat anti-rabbit IgG, A-11036, Invitrogen) at 1:500 for 1h, stained with 1 μ g/mL Hoechst, washed in 1X PBS, centrifuged on microscope slides with the CytoSpin™ 4 Cytocentrifuge (Thermo Scientific) and finally mounted in Citifluor AF2 glycerol solution (Citifluor Ltd, London).

Pgm Immunofluorescence

Pgm Immunofluorescence was performed as described in (Dubois et al., 2017). Briefly, cells were permeabilized for 4 minutes in PHEM 1X, Triton X-100 1% and fixed for 15 minutes in PHEM 1X, paraformaldehyde 2%, then washed twice in TBST (10 mM Tris pH 7.4, 0.15 M NaCl, 0.1% Tween 20) 3% BSA (bovine serum albumin, Sigma-Aldrich). Cells were incubated for 2 h at room temperature with anti-Pgm

2659 primary antibodies (1:300 in Signal + solution A for immunostaining, GenTex), and washed with TBST 3% BSA prior to 40 min incubation with Alexa fluor 488 goat anti-rabbit IgG secondary antibody (1:300, ThermoFisher Scientific), followed by DAPI staining (0.2 µg/ml) for 5 min before being mounted in Citifluor AF2.

Fixation of GFP-Spt16 and GFP-Spt16 injected cells

Cells were fixed and stained for 10 minutes in PHEM 1X (10 mM EGTA, 25 mM HEPES, 2 mM MgCl₂, 60m M PIPES pH 6.9), paraformaldehyde 2%, Hoechst 1 µg/ml, washed in 1X PBS and finally mounted in Citifluor AF2 glycerol solution (Citifluor Ltd, London).

Image acquisition and quantification

All images were acquired using a Zeiss LSM 710 laser-scanning confocal microscope and a Plan-Apochromat 63x/1.40 oil DIC M27 objective. Z-series were performed with Z-steps of 0.39 µm (or 0.5 µm for quantification). Quantification was performed using ImageJ. For each cell, a “Region Of Interest” (ROI) is drawn around the developing new MAC, using Z-stack projections of the Hoechst/DAPI channel. The area of the nucleus is given by the area of the ROI in µm². Using Z-stack projections of the GFP or Pgm channel in the corresponding image, ROI was used to measure the mean fluorescence intensity. In ROI area, integrated raw density was quantified and mean fluorescence intensity is calculated. An estimate of GFP/Pgm local background is defined by drawing a square outside of the cell in the corresponding channel and quantifying integrated raw density in the square selection. Fluorescence intensity in the nucleus was then defined by subtracting local background to mean fluorescence intensity of the corresponding nucleus.

Whole cell protein extracts and Western-Blot analysis

Cell pellets from autogamous cells at ≤2,000–4,000 cells/ml were frozen at -80°C in liquid nitrogen. Pellets were lysed 3 min in boiling SDS 5% and Protease Inhibitor Cocktail Set I - Calbiochem 5X (Merck) and centrifuged at maximum speed. Supernatant with proteins were collected and Laemmli buffer was added. 10 µg of protein extracts were used for Western blot. Electrophoresis and blotting were carried out according to standard procedures. The Pgm (1:500; Proteogenix 2659-Guinea Pig) primary antibody was used (Dubois et al., 2017). Secondary horseradish

peroxidase-conjugated donkey anti-rabbit IgG antibody (Promega) was used at 1:5000 dilution followed by detection by ECL (SuperSignal West Pico Chemiluminescent Substrate, Thermo Scientific). For normalization, the membranes probed with Pgm antibody were stripped in stripping buffer (ThermoScientific) and probed against with alpha-Tubulin antibody (TEU435 (Callen et al., 1994)).

DNA and sRNA sequencing

DNA for deep-sequencing was isolated from post-autogamous cells as previously described (Arnaiz et al., 2012) and sequenced by a paired-end strategy using Illumina GA-IIx and Hi-Seq next-generation sequencers. Accession numbers and sequencing data are displayed in Table S3.

Purification, sequencing and analysis of sRNAs from control and Spt16-depleted cells were carried out as previously described (Maliszewska-Olejniczak et al., 2015). Briefly, for sRNA library construction, the ~19–28 nt fraction was purified from total RNA by polyacrylamide gel electrophoresis (15%, 19:1 acrylamide:bisacrylamide) and gel-eluted with 0.3 M sodium chloride, followed by ethanol precipitation. The eluate was used for library construction using standard Illumina protocols. The 20–30 nt sRNA reads were filtered for known contaminants (*Paramecium* rDNA, mitochondrial DNA, feeding bacteria genomes and L4440 feeding vector sequences) (Table S4). In addition, the 23 nt siRNA reads that map to the RNAi targets were removed. The filtered reads were mapped to reference MAC and MAC+IES genomes. Read counts were normalized using the total number of filtered reads.

Reference genomes

The following reference genomes (Arnaiz et al., 2012) were used in the IES analyses and for read mapping.

Paramecium tetraurelia strain 51 MAC genome (v1.0) :
[https://paramecium.i2bc.paris-](https://paramecium.i2bc.paris-saclay.fr/download/Paramecium/tetraurelia/51/sequences/ptetraurelia_mac_51.fa)

[saclay.fr/download/Paramecium/tetraurelia/51/sequences/ptetraurelia_mac_51.fa](https://paramecium.i2bc.paris-saclay.fr/download/Paramecium/tetraurelia/51/sequences/ptetraurelia_mac_51.fa)
 Paramecium tetraurelia strain 51 MAC+IES genome (v1.0) :
[https://paramecium.i2bc.paris-](https://paramecium.i2bc.paris-saclay.fr/download/Paramecium/tetraurelia/51/sequences/ptetraurelia_mac_51_with_ies.fa)
[saclay.fr/download/Paramecium/tetraurelia/51/sequences/ptetraurelia_mac_51_with_ies.fa](https://paramecium.i2bc.paris-saclay.fr/download/Paramecium/tetraurelia/51/sequences/ptetraurelia_mac_51_with_ies.fa)

Paramecium tetraurelia strain 51 PGM contigs (v1): [https://paramecium.i2bc.paris-](https://paramecium.i2bc.paris-saclay.fr/download/Paramecium/tetraurelia/51/sequences/ptetraurelia_mac_51_with_ies.fa)

saclay.fr/download/Paramecium/tetraurelia/51/sequences/contigs_ABK_COSP_best_k51_no_scaf.fa

Paramecium tetraurelia strain 51 MIC contigs (v0): https://paramecium.i2bc.paris-saclay.fr/download/Paramecium/tetraurelia/51/sequences/ptetraurelia_mic2.fa

Macronuclear DNA reads for PiggyMac depleted cells were obtained from the European Nucleotide Archive (Accession number ERA137420) (PGM).

Genome-wide analysis of IES retention

Analysis was made using the multi-threaded Perl software PARTIES available at <https://github.com/oarnaiz/ParTIES> (Denby Wilkes et al., 2015).

ACKNOWLEDGMENTS

We wish to thank the members of the Duharcourt lab for fruitful discussions, Vinciane Régnier and Mireille Bétermier for the gift of the Pgm antibody, Caridad Miro Pina for her help in Western blot analysis, Arthur Abello for his help with image quantification. This work was supported by intramural funding from the CNRS, the Fondation de la Recherche Médicale (Equipe FRM DEQ20160334868), LABEX “Who am I?” (ANR-11-LABX-0071_WHOAMI), the Agence Nationale de la Recherche (ANR-14-CE10-0005-04 ‘PIGGYPACK’). A.T. was recipient of PhD fellowships from ‘Ministère de l’Enseignement Supérieur et de la Recherche’ and ‘Fondation ARC’. We acknowledge the ImagoSeine facility, member of the France BioImaging infrastructure supported by the ANR-10-INSB-04. The sequencing benefited from the facilities and expertise of the high-throughput sequencing platform of I2BC.

REFERENCES

Abe, T., Sugimura, K., Hosono, Y., Takami, Y., Akita, M., Yoshimura, A., Tada, S., Nakayama, T., Murofushi, H., Okumura, K., et al. (2011). The histone chaperone facilitates chromatin transcription (FACT) protein maintains normal replication fork rates. *J. Biol. Chem.* *286*, 30504–30512.

Adam, S., Dabin, J., and Polo, S.E. (2015). Chromatin plasticity in response to DNA damage: The shape of things to come. *DNA Repair (Amst.)* *32*, 120–126.

Allen, S.E., Hug, I., Pabian, S., Rzeszutek, I., Hoehener, C., and Nowacki, M. (2017). Circular Concatemers of Ultra-Short DNA Segments Produce Regulatory RNAs. *Cell* *168*, 990-999.e7.

Arnaiz, O., and Sperling, L. (2011). ParameciumDB in 2011: new tools and new data for functional and comparative genomics of the model ciliate *Paramecium tetraurelia*. *Nucleic Acids Res.* *39*, D632-636.

Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.-M., Wilkes, C.D., Garnier, O., Labadie, K., Lauderdale, B.E., Le Mouël, A., et al. (2012). The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet.* *8*, e1002984.

Arnaiz, O., Van Dijk, E., Bétermier, M., Lhuillier-Akakpo, M., de Vanssay, A., Duharcourt, S., Sallet, E., Gouzy, J., and Sperling, L. (2017). Improved methods and resources for *paramecium* genomics: transcription units, gene annotation and gene expression. *BMC Genomics* *18*, 483.

Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., et al. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* *444*, 171–178.

Baudry, C., Malinsky, S., Restituto, M., Kapusta, A., Rosa, S., Meyer, E., and Bétermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev.* *23*, 2478–2483.

Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., Jr, et al. (2010a). *Paramecium tetraurelia*: the renaissance of an early unicellular model. *Cold Spring Harb Protoc* *2010*, pdb.emo140.

Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., Jr, et al. (2010b). Mass culture of *Paramecium tetraurelia*. *Cold Spring Harb Protoc* *2010*, pdb.prot5362.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. *Science* *301*, 1090–1093.

Berger, J.D. (1973). Nuclear differentiation and nucleic acid synthesis in well-fed exconjugants of *Paramecium aurelia*. *Chromosoma* *42*, 247–268.

Betermier, M., and Duharcourt, S. (2014). Programmed Rearrangement in Ciliates: *Paramecium*. *Microbiol Spectr* *2*.

Bétermier, M., Duharcourt, S., Seitz, H., and Meyer, E. (2000). Timing of developmentally programmed excision and circularization of *Paramecium* internal eliminated sequences. *Mol. Cell. Biol.* *20*, 1553–1561.

Biswas, D., Yu, Y., Prall, M., Formosa, T., and Stillman, D.J. (2005). The yeast FACT complex has a role in transcriptional initiation. *Mol. Cell. Biol.* *25*, 5812–5822.

- Bouhouche, K., Gout, J.-F., Kapusta, A., Bétermier, M., and Meyer, E. (2011). Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res.* *39*, 4249–4264.
- Callen, A. M., Adoutte, A., Andrew, J. M., Baroin-Tourancheau, A., Bré, M. H., Ruiz, P. C., Clérot, J. C., Delgado, P., Fleury, A., Jeanmaire-Wolf, R. (1994) Isolation and characterization of libraries of monoclonal antibodies directed against various forms of tubulin in *Paramecium*. *Biology of the Cell* *81*, 95-119
- Coyne, R.S., Lhuillier-Akakpo, M., and Duharcourt, S. (2012). RNA-guided DNA rearrangements in ciliates: is the best genome defence a good offence? *Biol. Cell* *104*, 309–325.
- Denby Wilkes, C., Arnaiz, O., and Sperling, L. (2015). ParTIES : a toolbox for *Paramecium* interspersed DNA elimination studies. *Bioinformatics*.
- Dubois, E., Mathy, N., Régnier, V., Bischerour, J., Baudry, C., Trouslard, R., and Bétermier, M. (2017). Multimerization properties of PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements. *Nucleic Acids Res.*
- Duharcourt, S., Lepère, G., and Meyer, E. (2009). Developmental genome rearrangements in ciliates: a natural genomic subtraction mediated by non-coding transcripts. *Trends Genet.* *25*, 344–350.
- Duret, L., Cohen, J., Jubin, C., Dessen, P., Gout, J.-F., Mousset, S., Aury, J.-M., Jaillon, O., Noel, B., Arnaiz, O., et al. (2008). Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: A somatic view of the germline. *Genome Res* *18*, 585–596.
- Eitoku, M., Sato, L., Senda, T., and Horikoshi, M. (2008). Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly. *Cell. Mol. Life Sci.* *65*, 414–444.
- Formosa, T. (2012). The role of FACT in making and breaking nucleosomes. *Biochim. Biophys. Acta* *1819*, 247–255.
- Frost, J.M., Kim, M.Y., Park, G.T., Hsieh, P.-H., Nakamura, M., Lin, S.J.H., Yoo, H., Choi, J., Ikeda, Y., Kinoshita, T., et al. (2018). FACT complex is required for DNA demethylation at heterochromatin during reproduction in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* *115*, E4720–E4729.
- Furrer, D.I., Swart, E.C., Kraft, M.F., Sandoval, P.Y., and Nowacki, M. (2017). Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA. *Cell Rep* *20*, 505–520.
- Galvani, A., and Sperling, L. (2002). RNA interference by feeding in *Paramecium*. *Trends Genet.* *18*, 11–12.
- Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D., and Labib, K. (2006). GINS maintains association of Cdc45 with

MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat. Cell Biol.* *8*, 358–366.

Garnier, O., Serrano, V., Duharcourt, S., and Meyer, E. (2004). RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. *Mol. Cell. Biol.* *24*, 7370–7379.

Gruchota, J., Denby Wilkes, C., Arnaiz, O., Sperling, L., and Nowak, J.K. (2017). A meiosis-specific Spt5 homolog involved in non-coding transcription. *Nucleic Acids Res.*

Guérin, F., Arnaiz, O., Boggetto, N., Denby Wilkes, C., Meyer, E., Sperling, L., and Duharcourt, S. (2017). Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements. *BMC Genomics* *18*, 327.

Hammond, C.M., Strømme, C.B., Huang, H., Patel, D.J., and Groth, A. (2017). Histone chaperone networks shaping chromatin function. *Nat. Rev. Mol. Cell Biol.* *18*, 141–158.

Han, J., Li, Q., McCullough, L., Kettelkamp, C., Formosa, T., and Zhang, Z. (2010). Ubiquitylation of FACT by the cullin-E3 ligase Rtt101 connects FACT to DNA replication. *Genes Dev.* *24*, 1485–1490.

Heo, K., Kim, H., Choi, S.H., Choi, J., Kim, K., Gu, J., Lieber, M.R., Yang, A.S., and An, W. (2008). FACT-mediated exchange of histone variant H2AX regulated by phosphorylation of H2AX and ADP-ribosylation of Spt16. *Mol. Cell* *30*, 86–97.

Hondele, M., Stuwe, T., Hassler, M., Halbach, F., Bowman, A., Zhang, E.T., Nijmeijer, B., Kotthoff, C., Rybin, V., Amlacher, S., et al. (2013). Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. *Nature* *499*, 111–114.

Hsieh, F.-K., Kulaeva, O.I., Patel, S.S., Dyer, P.N., Luger, K., Reinberg, D., and Studitsky, V.M. (2013). Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* *110*, 7654–7659.

Jamai, A., Puglisi, A., and Strubin, M. (2009). Histone chaperone spt16 promotes redeposition of the original h3-h4 histones evicted by elongating RNA polymerase. *Mol. Cell* *35*, 377–383.

Jeronimo, C., and Robert, F. (2016). Histone chaperones FACT and Spt6 prevent histone variants from turning into histone deviants. *Bioessays*.

Jeronimo, C., Watanabe, S., Kaplan, C.D., Peterson, C.L., and Robert, F. (2015). The Histone Chaperones FACT and Spt6 Restrict H2A.Z from Intragenic Locations. *Mol. Cell* *58*, 1113–1123.

Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J.D., Malinsky, S., and Bétermier, M. (2011). Highly precise and developmentally programmed genome assembly in *Paramecium* requires ligase IV-dependent end joining. *PLoS Genet.* *7*, e1002049.

- Keck, K.M., and Pemberton, L.F. (2012). Histone chaperones link histone nuclear import and chromatin assembly. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* *1819*, 277–289.
- Kemble, D.J., Whitby, F.G., Robinson, H., McCullough, L.L., Formosa, T., and Hill, C.P. (2013). Structure of the Spt16 middle domain reveals functional features of the histone chaperone FACT. *J. Biol. Chem.* *288*, 10188–10194.
- Kemble, D.J., McCullough, L.L., Whitby, F.G., Formosa, T., and Hill, C.P. (2015). FACT Disrupts Nucleosome Structure by Binding H2A-H2B with Conserved Peptide Motifs. *Mol. Cell* *60*, 294–306.
- Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin. *Nature* *547*, 236–240.
- Lepère, G., Bétermier, M., Meyer, E., and Duharcourt, S. (2008). Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev.* *22*, 1501–1512.
- Lepère, G., Nowacki, M., Serrano, V., Gout, J.-F., Guglielmi, G., Duharcourt, S., and Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res.* *37*, 903–915.
- Lhuillier-Akakpo, M., Frapporti, A., Denby Wilkes, C., Matelot, M., Vervoort, M., Sperling, L., and Duharcourt, S. (2014). Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements. *PLoS Genet.* *10*, e1004665.
- Maliszewska-Olejniczak, K., Gruchota, J., Gromadka, R., Denby Wilkes, C., Arnaiz, O., Mathy, N., Duharcourt, S., Bétermier, M., and Nowak, J.K. (2015). TFIS-Dependent Non-coding Transcription Regulates Developmental Genome Rearrangements. *PLoS Genet.* *11*, e1005383.
- Malone, E.A., Clark, C.D., Chiang, A., and Winston, F. (1991). Mutations in SPT16/CDC68 suppress cis- and trans-acting mutations that affect promoter function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *11*, 5710–5717.
- Marmignon, A., Bischerour, J., Silve, A., Fojcik, C., Dubois, E., Arnaiz, O., Kapusta, A., Malinsky, S., and Bétermier, M. (2014). Ku-mediated coupling of DNA cleavage and repair during programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *PLoS Genet.* *10*, e1004552.
- Nowacki, M., Zagorski-Ostojka, W., and Meyer, E. (2005). Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in *Paramecium tetraurelia*. *Curr. Biol.* *15*, 1616–1628.
- Orphanides, G., LeRoy, G., Chang, C.H., Luse, D.S., and Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* *92*, 105–116.

- Prendergast, J.A., Murray, L.E., Rowley, A., Carruthers, D.R., Singer, R.A., and Johnston, G.C. (1990). Size selection identifies new genes that regulate *Saccharomyces cerevisiae* cell proliferation. *Genetics* *124*, 81–90.
- Prendergast, L., Müller, S., Liu, Y., Huang, H., Dingli, F., Loew, D., Vassias, I., Patel, D.J., Sullivan, K.F., and Almouzni, G. (2016). The CENP-T/-W complex is a binding partner of the histone chaperone FACT. *Genes Dev.* *30*, 1313–1326.
- Ransom, M., Dennehey, B.K., and Tyler, J.K. (2010). Chaperoning histones during DNA replication and repair. *Cell* *140*, 183–195.
- Sandoval, P.Y., Swart, E.C., Arambasic, M., and Nowacki, M. (2014). Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. *Dev. Cell* *28*, 174–188.
- Schlesinger, M.B., and Formosa, T. (2000). POB3 is required for both transcription and replication in the yeast *Saccharomyces cerevisiae*. *Genetics* *155*, 1593–1606.
- Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. *Nature* *547*, 241–245.
- True, J.D., Muldoon, J.J., Carver, M.N., Poorey, K., Shetty, S.J., Bekiranov, S., and Auble, D.T. (2016). The Mot1 ATPase and Spt16 Histone Chaperone Co-Regulate Transcription through Preinitiation Complex Assembly and Nucleosome Organization. *J. Biol. Chem.*
- VanDemark, A.P., Xin, H., McCullough, L., Rawlins, R., Bentley, S., Heroux, A., Stillman, D.J., Hill, C.P., and Formosa, T. (2008). Structural and functional analysis of the Spt16p N-terminal domain reveals overlapping roles of yFACT subunits. *J. Biol. Chem.* *283*, 5058–5068.
- Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* *16*, 178–189.
- Yang, J., Zhang, X., Feng, J., Leng, H., Li, S., Xiao, J., Liu, S., Xu, Z., Xu, J., Li, D., et al. (2016). The Histone Chaperone FACT Contributes to DNA Replication-Coupled Nucleosome Assembly. *Cell Rep* *16*, 3414.

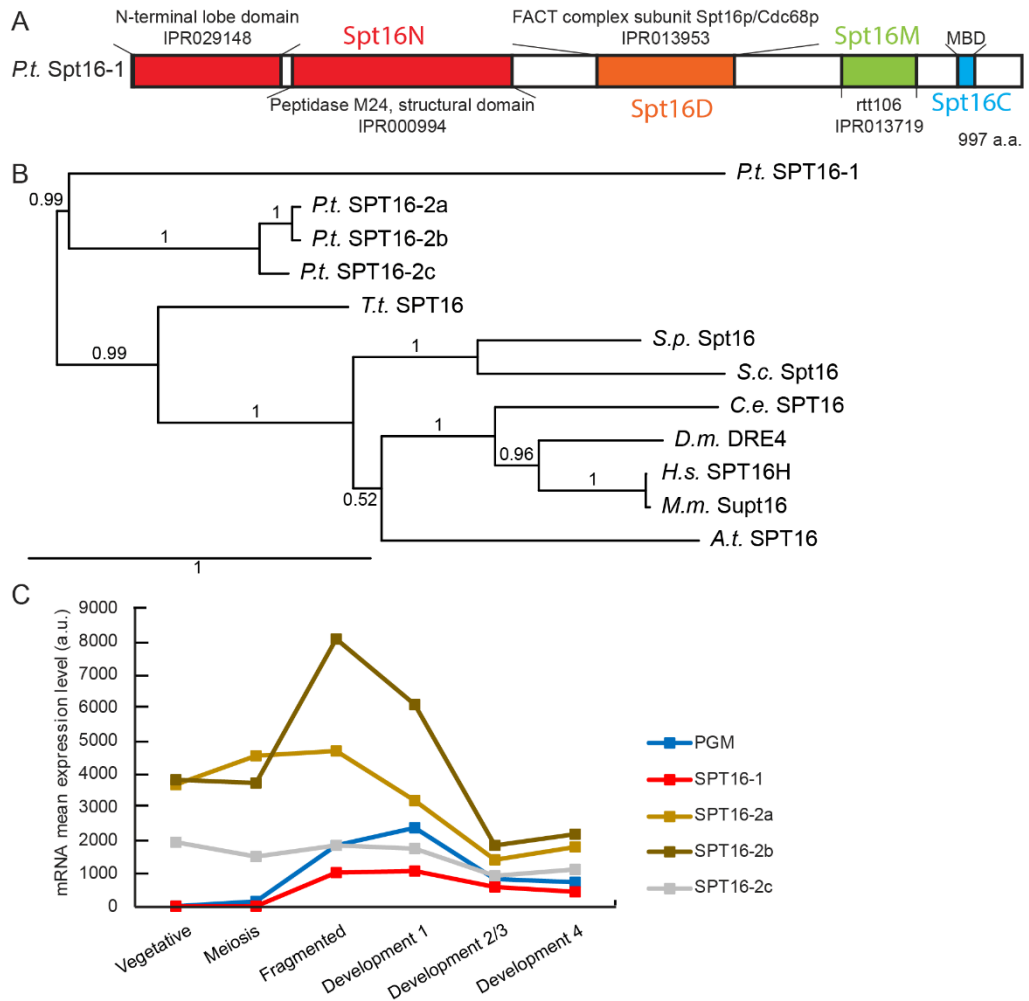


Figure 1. *Paramecium tetraurelia* Spt16 proteins.

A. Conserved domains (colored boxes) in *Paramecium tetraurelia* Spt16-1 protein. B. Phylogenetic tree of Spt16 proteins from *P. tetraurelia* (Pt), *Tetrahymena thermophila* (Tt), *Saccharomyces cerevisiae* (Sc), *Schizosaccharomyces pombe* (Sp), *Caenorhabditis elegans* (Ce), *Drosophila melanogaster* (Dm), mouse (Mm), human (Hs) and *Arabidopsis thaliana* (At) based on the alignment of full-length protein sequences. The tree was generated with PhyML 3.0 with bootstrapping procedure and visualized with Tree.Dyn 198.3. Accession numbers are provided in Table S1. C. *SPT16* gene expression profiles during the life cycle. Mean mRNA expression levels were determined by RNA sequencing during vegetative growth and at different time points during autogamy (Arnaiz et al. 2017). The vegetative point consists of two samples from mass cultures containing only log-phase cells showing no sign of meiosis. The meiosis time point consists of two samples containing 32-39% of cells undergoing meiosis and little fragmentation of the maternal MAC. The fragmented time point consists of three samples containing a similar proportion of meiotic cells (20-23%) as the meiosis time point, but also contained 40-47% of cells with a fragmented maternal MAC. The Development 1 time point consists of two samples containing 50-56% of cells with a fragmented maternal MAC and 35-37% of cells that contained two already visible new MACs. The development 2/3 time point consists of four samples containing 73-91% of cells with two new MACs. Development 4 time point consists of two samples taken 10 hours after the previous time point and consisting of 91-98% of cells with two new MACs and in which appear 1-9% of post-autogamous cells.

Figure 01

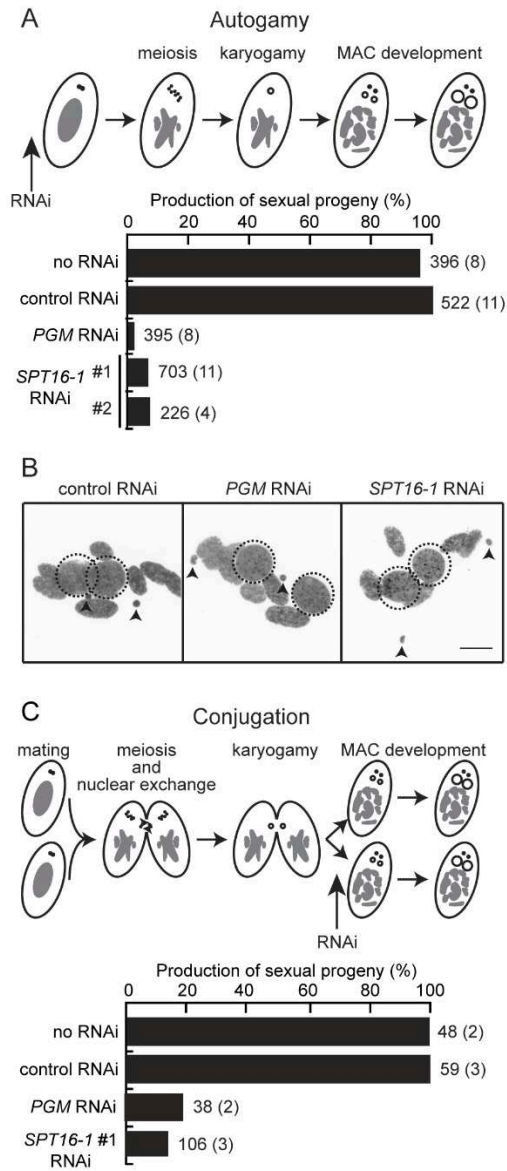


Figure 2. Spt16-1 is essential for development.

A. Production of post-autogamous sexual progeny following *SPT16-1* gene silencing. The gene targeted in each silencing experiment is indicated. Two non-overlapping silencing fragments (#1 and #2) of the *SPT16-1* gene were used independently. The *ND7* or *ICL7* genes were used as control RNAi targets, since their silencing has no effect on sexual processes (Bouhouche et al., 2011). Autogamy was also performed in standard *K. pneumoniae* medium (no RNAi). Cells were starved in each medium to induce autogamy and, following 3–4 days of starvation, autogamous cells were transferred individually to *K. pneumoniae* medium to monitor growth of sexual progeny. The total number of autogamous cells analyzed for each RNAi and the number of independent experiments (in parenthesis) are indicated. Death in progeny after *SPT16-1* silencing was observed after less than three cell divisions, as for *PGM* silencing. B. Z-projections of Hoechst staining of control, *Pgm* or *Spt16-1*-depleted cells fixed at late stages of development during autogamy. Dashed black circles indicate the two developing MACs. Black arrows indicate the MICs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μ m. C. *SPT16-1* gene silencing following mating pair separation. Reactive cells were obtained in standard *K. pneumoniae* medium, and mating pairs were transferred to silencing medium. The gene targeted in each silencing experiment is indicated. Only silencing fragment #1 of the *SPT16-1* gene was used. The next day, individual sexual progenies were transferred individually to *K. pneumoniae* medium to monitor growth. The total number of cells analyzed for each RNAi and the number of independent experiments (in parenthesis) are indicated.

Figure 02

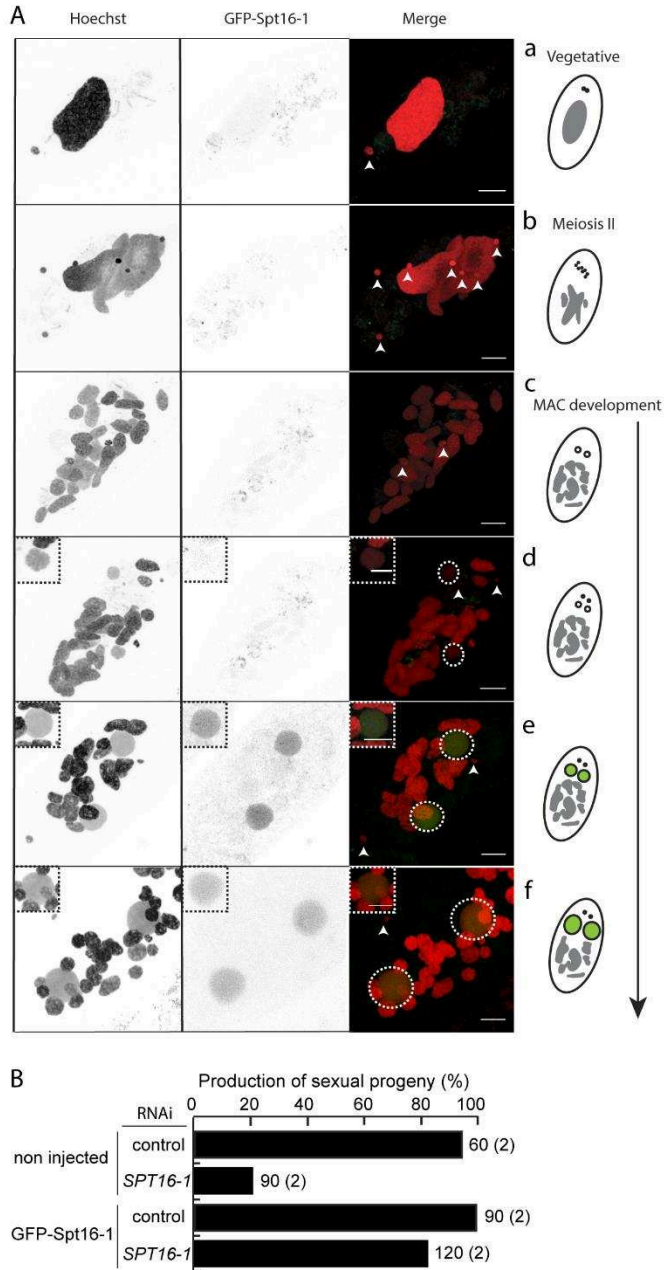


Figure 3. A functional GFP-Spt16-1 fusion protein is localized in the developing macronuclei.

A. Localization of the GFP-Spt16-1 fusion protein expressed from the RNAi-resistant *GFP-SPT16-1* transgene (see Materials and Methods). Overlay of Z-projections of magnified views of GFP-Spt16-1 (in green) and Hoechst (in red) are presented (a-f), aligned with their schematic representations on the right. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC or the MICs. Scale bar is 10 μ m. The inset displays one of the two developing new MACs. Transgene copy number per haploid genome was 19 as measured by qPCR from total genomic DNA extracted from vegetative cells and values were normalized to the endogenous *GAPDH* gene. B. *In vivo* genetic complementation with the RNAi-resistant *GFP-SPT16-1* transgene upon *SPT16-1* gene silencing during autogamy. Production of sexual of post-autogamous progeny following complementation of *SPT16-1* RNAi by GFP-Spt16-1 RNAi resistant is indicated. The gene targeted in silencing experiment is indicated. Silencing is realized in non-injected cells (NI) or GFP-Spt16-1 injected cells (GFP-Spt16-1). *ICL7* gene is used as control RNAi target, since its silencing has no effect on sexual processes (Bouhouche et al., 2011). Cells were starved in each medium to induce autogamy and, following 3-4 days of starvation, autogamous cells were transferred individually to *K. pneumoniae* medium to monitor growth of sexual progeny. The total number of autogamous cells analyzed for each RNAi and the number of independent experiments (in parenthesis) are indicated.

Figure 03

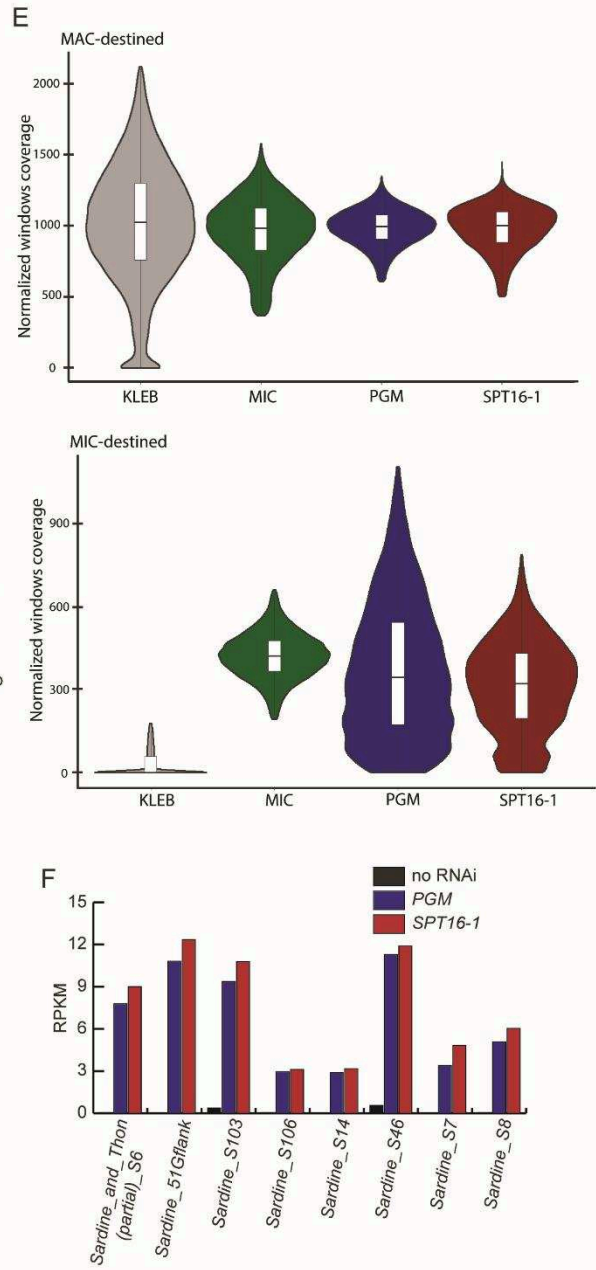
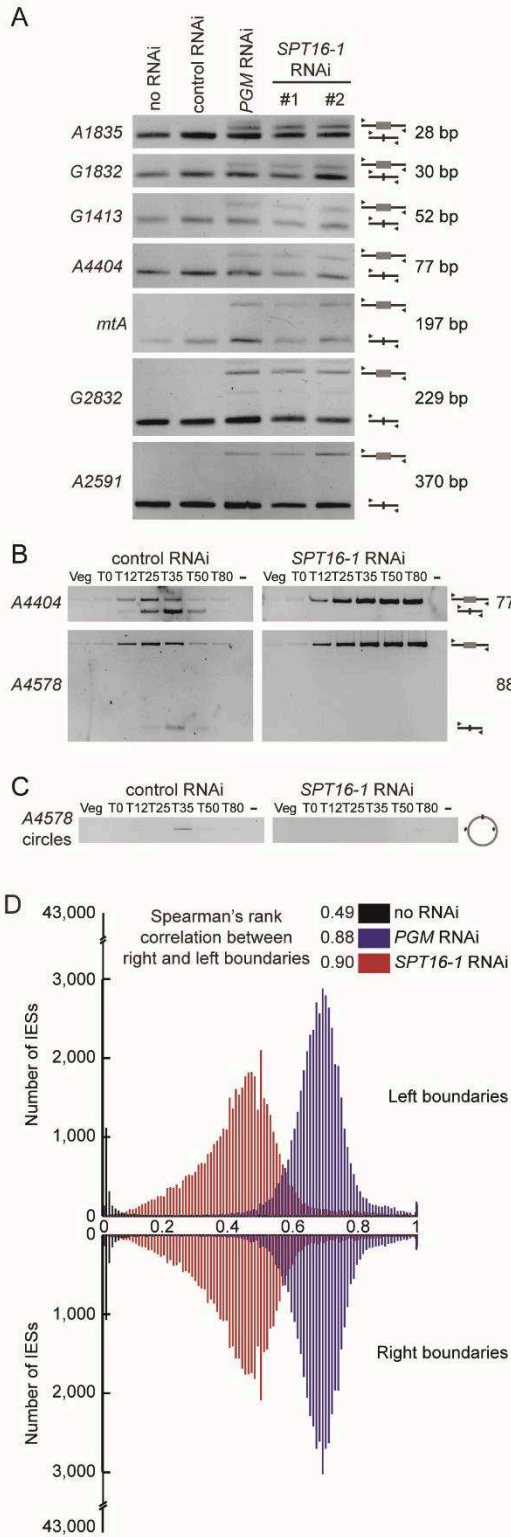


Figure 4. Spt16-1 is required for DNA elimination events

A. PCR analysis of the excision of some IEs using primers located each side of the IES after autogamy. On the left, are named each IES. On the right are listed corresponding size of each IES. The gene targeted in silencing experiment is indicated. As detailed in Figure 2, autogamy was also performed after no RNAi or control RNAi. As expected there was no IES retention for no RNAi or control RNAi but there was IES retention for *PGM* gene silencing. In each lane, upper band corresponds to the non-excised form whereas lower band corresponds to excised IES form and is mostly amplified from fragments of the maternal MAC. B. Maternal inheritance of macronuclear deletion is analyzed after Control or *SPT16-1* RNAi using a cell line reproducibly deleting the *ND7* gene and *A* gene from the MAC genome at each sexual generation. PCR is performed on DNA extracted during autogamy time course experiment with primers each side of IES included in *A* gene. Upper band corresponds to non deleted *A* gene with retained IESs. Lower band corresponds to non deleted *A* gene with excised IESs. In Control RNAi, *A* gene is deleted before autogamy and progressively deleted in last time points after IES have been excised. C. PCR detection of IES 51A4578 circles with divergent primers on genomic DNA at each time point during autogamy time course experiment silencing control or *SPT16-1* gene. D. IES retention scores after *PGM* and *SPT16-1* RNAi. Histograms of retention scores for left and right boundaries of all IESs, as described in Materials and Methods, for control (black), *PGM* RNAi (blue) and *SPT16-1* RNAi (red). Spearman's rank correlation between right and left boundaries is given for each RNAi. E. Violin plot superimposed with a boxplot of normalized coverage using DESeq2 (see Guerin et al.) of 1kb windows of the MIC assembly for a klebsiella (KLEB grey) sample, MIC sample (green), *PGM* RNAi sample (blue) and *SPT16* RNAi sample. The windowse are split into two categories defined in Guerin et al. 2017 : MAC-destined compartment and MIC-limited compartment. F. Retention of some class II DNA transposons imprecisely eliminated after *PGM* and *SPT16-1* RNAi. The bar plots represent read coverage of 8 individual copies of the Sardine transposon (GenBank Accession No. HE774468-HE774475). The coverage was determined by mapping reads for the conrai dataset (black), *PGM* silencing (blue) and *SPT16-1* silencing (red) datasets. The normalized units (RPKM) are reads per kilobase of the transposon sequence per million library read mapped against the MAC reference genome. As expected all transposable elements are retained in *PGM* RNAi (Arnal et al., 2012, Guerin et al., 2017)

Figure 04

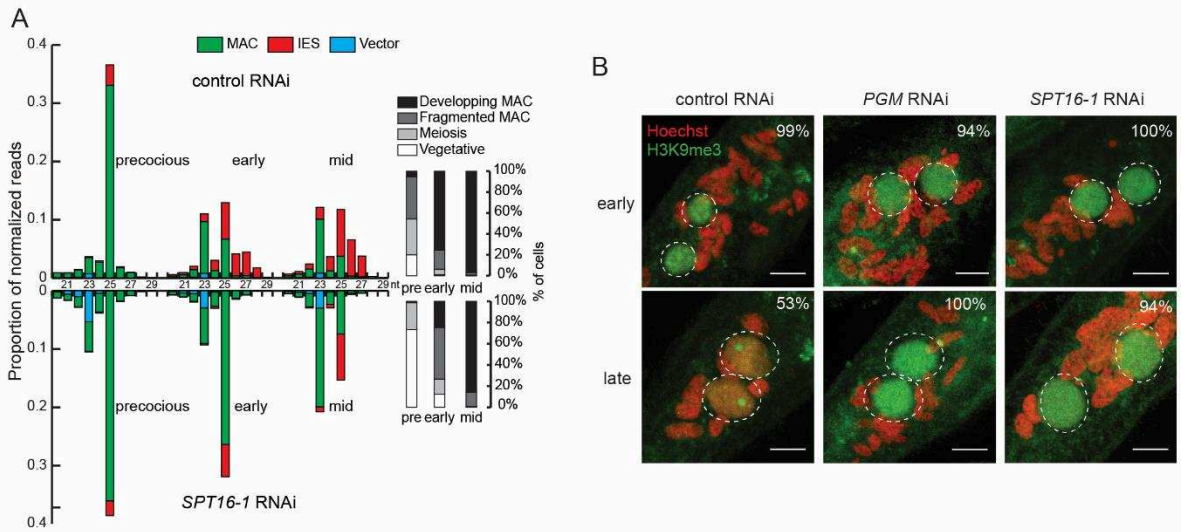


Figure 5. Spt16-1 is required for iesRNA accumulation and for H3K27me3 and H3K9me3 foci formation

A. Analysis of iesRNAs populations in *SPT16-1* RNAi. Small RNA libraries corresponding to three time-points during autogamy: precocious (app. 20% of cell population undergoes meiosis), early (10h after the first time point), and middle (20 h after the first time point), after control or *SPT16-1* RNAi were sequenced and mapped to the reference genomes (*Paramecium tetraurelia* MAC reference genome and MAC+IES reference genome). Bar plots show the normalized proportion of sRNA reads that match the MAC genome (green), annotated IESs (yellow) or to the feeding vector (purple). The histograms display the cytological stages of the cell population at each developmental time point. B. Overlay of z projections of magnified views of H3K9me3 specific antibodies (green) and Hoechst (red) in control or *SPT16-1* or *PGM* RNAi at early or late development time-points. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μ m. Percentage of H3K9me3 positive cells are indicated.

Figure 05

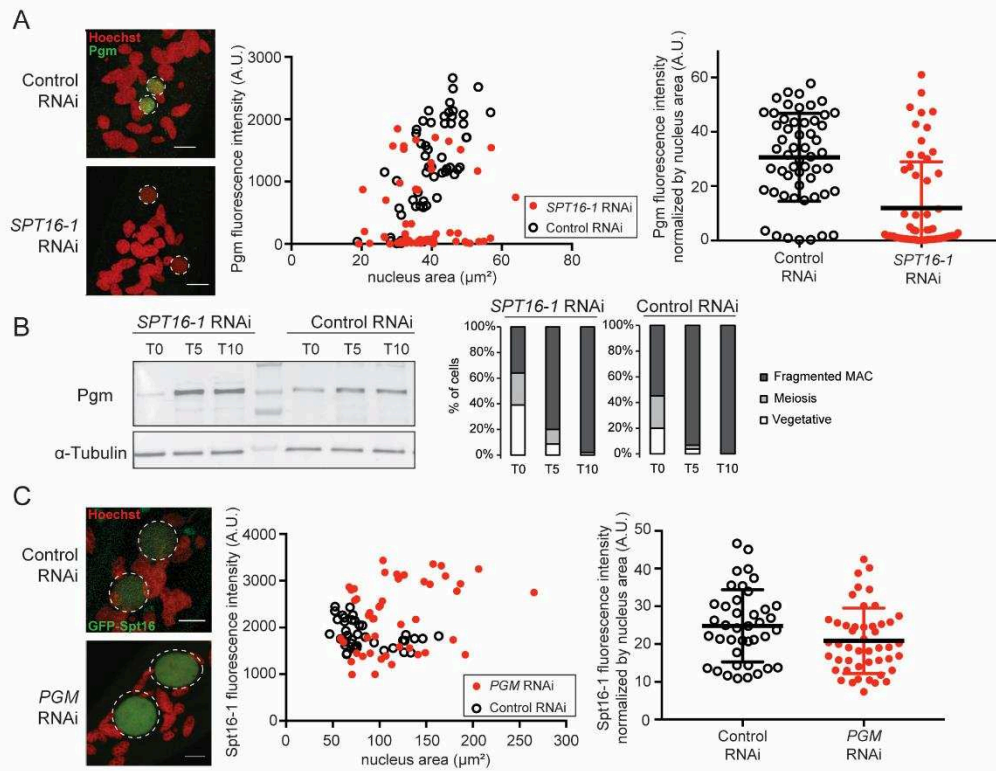


Figure 6. Spt16-1 is required for Pgm localization during development

A. Overlay of Z-projections of immunolabelling with Pgm antibody (green) and staining with Hoechst (red) in control or *SPT16-1* RNAi at T5. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μm . Next quantification by Image J macro of Pgm antibody fluorescence signal compared to nucleus size in μm^2 is shown during development for control RNAi (white circles) or *SPT16-1* RNAi (red circles). Quantification of the Pgm fluorescence signal in the developing new MAC (Arbitrary Units, A. U.) as a function of nucleus area (in μm^2) at T5 upon control RNAi (white circle) and *SPT16-1* RNAi (red circle). Bar plots represent Pgm fluorescence signal in the developing new MAC (Arbitrary Units, A. U.) normalized to the nucleus area upon control and *SPT16-1* RNAi.

B. Pgm expression during development in control or *SPT16-1* RNAi. Whole proteins are extracted at T0, T5 and T10 when Pgm is normally detected by immunofluorescence. Western blot of whole cell proteins extracted at T0, T5 and T10 during autogamy upon control and *SPT16-1* RNAi. Upper panel: Pgm antibodies were used for Pgm detection. Bottom panel: Western blot of the same membrane, following stripping and incubation with alpha tubulin antibodies. Cytology of the cell population at each developmental time point is displayed in the histograms. Pgm is quantified by normalization by Tubuline signal.

C. Localization of GFP-Spt16-1 fusion protein after *PGM* RNAi at late development time points (T50). Overlay of Z-projections of views of GFP-Spt16-1 (in green) and Hoechst (in red) are presented. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μm . Quantification of the GFP fluorescence signal in the developing new MAC (Arbitrary Units, A. U.) as a function of nucleus area (in microns μ^2) at a late developmental time point (T50) upon control RNAi (white circle) and *SPT16-1* RNAi (red circle).

Figure 06

Species	Name	Accession Number
<i>Paramecium tetraurelia</i>	Spt16-1	PTET.51.1.P0710091
<i>Paramecium tetraurelia</i>	Spt16-2a	PTET.51.1.P1210069
<i>Paramecium tetraurelia</i>	Spt16-2b	PTET.51.1.P1050068
<i>Paramecium tetraurelia</i>	Spt16-2c	PTET.51.1.P0550178
<i>Tetrahymena thermophila</i>	Spt16	TTHERM_00283330
<i>Schizosaccharomyces pombe</i>	Spt16	SPBP8B7.19
<i>Saccharomyces cerevisiae</i>	Spt16	YGL207W
<i>Caenorhabditis elegans</i>	Spt16	WP:CE17113
<i>Drosophila melanogaster</i>	DRE4	FBgn0002183
<i>Homo sapiens</i>	Spt16H	HGNC:11465
<i>Mus musculus</i>	Supt16	MGI:1890948
<i>Arabidopsis thaliana</i>	Spt16	AT4G10710
<i>Paramecium octaurelia</i>	Spt16-1	POCTA.138.1.P1140149
<i>Paramecium octaurelia</i>	Spt16-2a	POCTA.138.1.P1350085
<i>Paramecium octaurelia</i>	Spt16-2b	POCTA.138.1.P1280104
<i>Paramecium octaurelia</i>	Spt16-2c	POCTA.138.1.P1200107
<i>Paramecium tredecaurelia</i>	Spt16-1	PTRED.209.2.P71800001294160129
<i>Paramecium tredecaurelia</i>	Spt16-2a	PTRED.209.2.P71800001294140062
<i>Paramecium tredecaurelia</i>	Spt16-2b	PTRED.209.2.P71800001294090202
<i>Paramecium tredecaurelia</i>	Spt16-2c	PTRED.209.2.P71800001290980080
<i>Paramecium tredecaurelia</i>	Spt16-2d	PTRED.209.2.P71800001293940093
<i>Paramecium biaurelia</i>	Spt16-1	PBIA.V1_4.1.P01200073
<i>Paramecium biaurelia</i>	Spt16-2a	PBIA.V1_4.1.P00180125
<i>Paramecium biaurelia</i>	Spt16-2b	PBIA.V1_4.1.P00690070
<i>Paramecium biaurelia</i>	Spt16-2c	PBIA.V1_4.1.P04930005
<i>Paramecium biaurelia</i>	Spt16-2d	PBIA.V1_4.1.P02300049
<i>Paramecium primaurelia</i>	Spt16-1	PPRIM.AZ9-3.1.P1650085
<i>Paramecium primaurelia</i>	Spt16-2a	PPRIM.AZ9-3.1.P0700122
<i>Paramecium primaurelia</i>	Spt16-2b	PPRIM.AZ9-3.1.P0670199
<i>Paramecium primaurelia</i>	Spt16-2c	PPRIM.AZ9-3.1.P1030093
<i>Paramecium primaurelia</i>	Spt16-2d	PPRIM.AZ9-3.1.P0690041
<i>Paramecium pentaurelia</i>	Spt16-1	PPENT.87.1.P0520132
<i>Paramecium pentaurelia</i>	Spt16-2a	PPENT.87.1.P0610195
<i>Paramecium pentaurelia</i>	Spt16-2b	PPENT.87.1.P0440198
<i>Paramecium pentaurelia</i>	Spt16-2c	PPENT.87.1.P0970092
<i>Paramecium pentaurelia</i>	Spt16-2d	PPENT.87.1.P0510097
<i>Paramecium sexaurelia</i>	Spt16-1	PSEX.AZ8_4.1.P1620011
<i>Paramecium sexaurelia</i>	Spt16-2a	PSEX.AZ8_4.1.P0570148
<i>Paramecium sexaurelia</i>	Spt16-2b	PSEX.AZ8_4.1.P0960067
<i>Paramecium sexaurelia</i>	Spt16-2c	PSEX.AZ8_4.1.P1020092
<i>Paramecium sexaurelia</i>	Spt16-2d	PSEX.AZ8_4.1.P0830089
<i>Paramecium caudatum</i>	Spt16-1	PCAU.43c3d.1.P00040086
<i>Paramecium caudatum</i>	Spt16-2	PCAU.43c3d.1.P00700076

Table S1. Accession number of proteins used in the phylogenetic analysis

Accession number, name of the protein and species are given. *Paramecium* database: <https://paramecium.i2bc.paris-saclay.fr/>, *T. thermophila* database: <http://ciliate.org/index.php/home/welcome>, *S. pombe* database:

<https://www.pombase.org/>, *S. cerevisiae* database: <https://www.yeastgenome.org/>,
C. elegans database: <https://wormbase.org/#012-34-5>, *D. melanogaster* database:
<http://flybase.org/>, *M. musculus* database: <http://www.informatics.jax.org/>, *H. sapiens*
database: <https://www.genenames.org>, *A. thaliana* database:
<https://www.arabidopsis.org/>.

Name	Sequence (5' to 3')	Application
51A4578-2	TGGTTGTTAGTCTCAAAGAATTCTAAAGAC	PCR for IES circles
514578-7bis	AAATTCCAAAAAGTTTGGAA TA TCTTTTTGAG	PCR for IES circles
51A-1835-5'(3)	GGTTGCGTAAACACTTCCTCTTAAA GTGAG	PCR around IESA1835
51A-1835-3'	ATCCTAACATCCTTGAA TAGTTACTGATCC	PCR around IESA1835
51G1832up	GCTATAACTCTTGAAGCTGCTTGTAATG	PCR around IESG1832
51G1832do	TTGTCAATGAGCCATTAACAGTTGCTGGAT	PCR around IESG1832
51A-4404-5'(2)	TGGAATAGTGCTGCATCACCAGCTGCTTGC	PCR around IESA4404
51A-4404-3'(2)	CCAGTTATTGAACTGCAACTTACTGCAGTG	PCR around IESA4404
51G1413up	GAAGCTGCTTGTTAAGAA TTCTACTGG	PCR around IESG1413
51G1413do	GCATCCAGCACTAGTTGAA TTTACTGTA	PCR around IESG1413
MT1	TTCTAAGCTGATT TA TTCAA TTTCAACAGAAC	PCR around IES mtA
MT2	TTGAAAAAAGGTCA TCTCTTTCA TTAATTC	PCR around IES mtA
51G2832-6	CAAATACTGGTGGTGCAACATCCTCAACTGC	PCR around IESG2832
51G2832-8	GAACCATTTAA TGCGCAA TA TCCTGCATTTCC	PCR around IESG2832
51A2591-18	AAGTGCAACCTGTGCTGATGCTCCCGATGA	PCR around IESA2591
51A2591-20	AGTTCCTTTGAAAGATGTGCAAGCTCCAGA	PCR around IESA2591
51A-6649-5'	AAATGGTACTGTTTGTGCTTGGGATAGTGC	PCR around IESA6649
51A-6649-3'	CAGCAGTACATCCAGCTCTCTAAGTTTAGC	PCR around IESA6649
51A4578-5'	CACTGCAGTAAGTTGCAGTTCAATAACTGG	PCR around IESA4578
fed_SPT16_f1	AAATCACTATCAGATTTTGGACTAA TG	Insert feeding #1 for Spt16-1
fed_SPT16_r1	CATCTTAAAAATGAAATGAATGCATG	Insert feeding #1 for Spt16-1
fed2_SPT16_f2	TTGTGAATTGATGGATTGCATATTG	Insert feeding #2 for Spt16-1
fed2_SPT16_r2	TTAAAATAATAATGTCCTTGTGCAATATG	Insert feeding #2 for Spt16-1
51A4578-3'	GTAGTCTTAAAA TCTTAGCATGTTGTACC	PCR around IESA4578
SPT16_up1_F	ATTGCAATTGATTTACTGTGATGATGTG	PCR for SPT16-1 and regulatory sequences
SPT16_do1_R	TATCTCAATCAGAGGACTCAGAGGGA	PCR for SPT16-1 and regulatory sequences
pcrII SPT16_ATG_F	GATAAAGTG TAGATAGCA TTATAAAATAA	PCR to introduce GFP in GFP-Spt16-1
pcrII SPT16_ATG_R	TATTTAATA TTCTTTTACAATTGCCAAAC	PCR to introduce GFP in GFP-Spt16-1
qPCR_GFP_F1	CTTTTCACTGGTGTGTTGTTCC	qPCR to measure transgene copy number
qPCR_GFP_R1	TCCATAAGTTGCATCACCTT	qPCR to measure transgene copy number
Actin 1-1 for	AAGGTGAAATAATTGTCA TCATAATCA	qPCR to measure transgene copy number
Actin 1-1 rev	CAACA ACTGCGGGGAAG	qPCR to measure transgene copy number

Table S2. Oligonucleotides used in this study

Sample	ENA Accession	Insert size (bp)	Read length (bp)	Number of reads (M)	Aligned reads on the MAC	MAC coverage	Aligned reads on the MIC	MIC coverage
Control	ERX466735	212	74	106	98% (104M)	99%	99% (105M)	79%
<i>PGM</i> RNAi	ERA137444	341	108	128	87% (112M)	99%	97% (125M)	96%
<i>SPT16-1</i> RNAi	XXX	372	101	102	90,20% (92M)	99%	99% (101M)	96%
MIC	SAMN053236 60	265	101	92	84,78% (78M)	99%	98% (91M)	99%

Table S3. Sequencing and mapping statistics for DNaseq.

Read statistics are provided for the *SPT16-1* RNAi sample sequenced for this study.

Stage	Label	ENA Accession	Number of reads	Number of reads mapped on MAC	%	Number of reads mapped on IES	%	Number of reads mapped on vector	%
Precocious	Control	SR	14 215 837,00	6632869	46,66%	622 985,00	4,38%	3 108 206,00	21,86%
	<i>SPT16-1</i>	SR	23 133 590,00	13782467	59,58%	725 102,00	3,13%	4 387 066,00	18,96%
Early	Control	SR	20 966 122,00	4494664	21,44%	4 371 367,00	20,85%	4 789 590,00	22,84%
	<i>SPT16-1</i>	SR	29 561 256,00	13097835	44,31%	1 961 808,00	6,64%	9 039 867,00	30,58%
Mid	Control	SR	8 601 883,00	1629823	18,95%	2 173 420,00	25,27%	2 342 676,00	27,23%
	<i>SPT16-1</i>	SR	24 619 965,00	8643585	35,11%	2 366 803,00	9,61%	8 960 916,00	36,40%

Table S4. Description of sRNA-seq data

Read statistics are provided for the control RNAi and *SPT16-1* RNAi sample sequenced for this study.

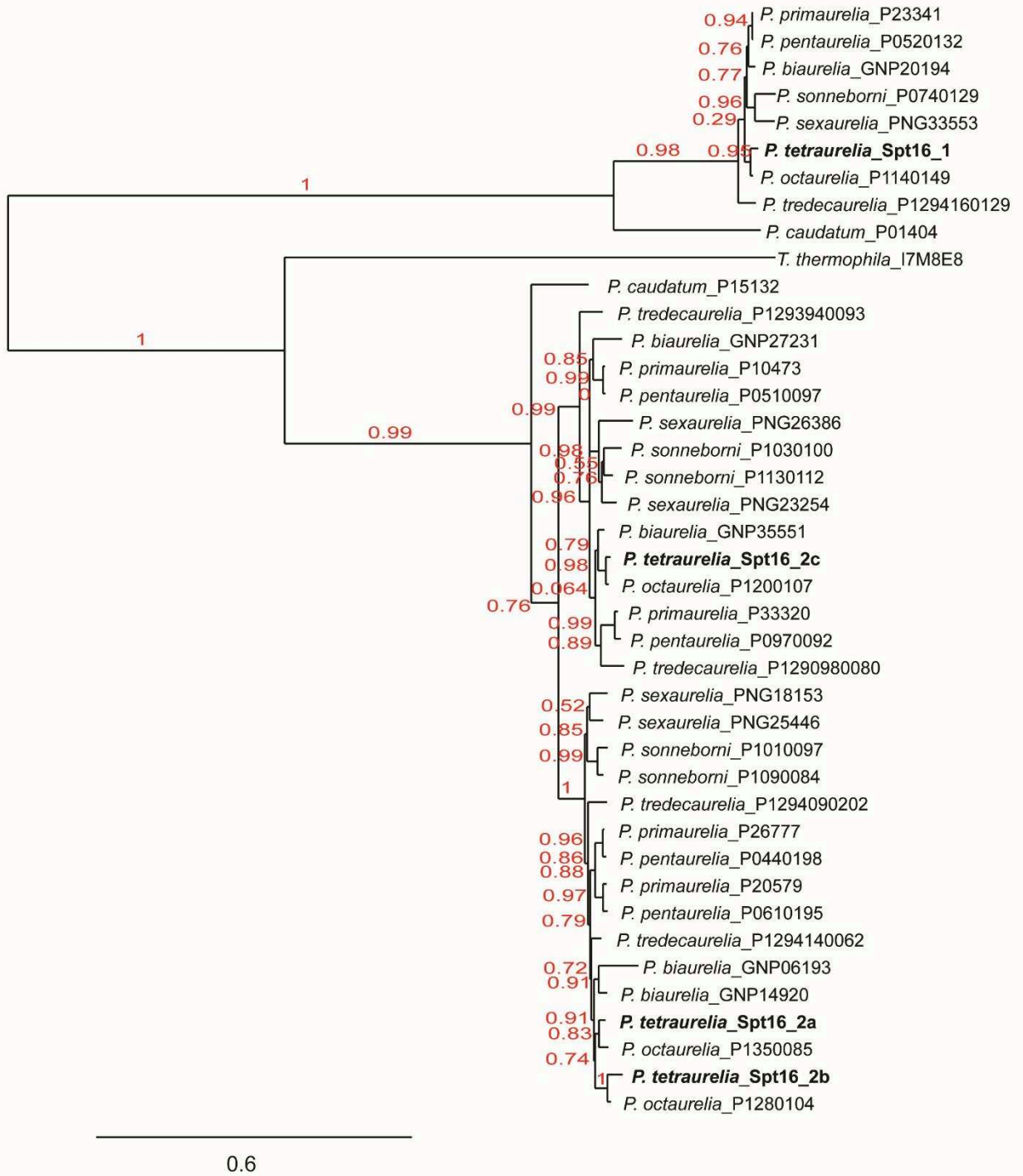


Figure S1. Identification of Spt16 homologues in *P. tetraurelia*

Phylogenetic tree of Spt16 proteins from *P. tetraurelia*, other *Paramecium* species and other eukaryotes based on alignment of full length protein sequence by MUSCLE. Phylogeny of the alignment was generated using PhyML 3.0 on Phylogeny.fr with bootstrapping procedure using 100 bootstraps. Tree has been viewed using Tree.Dyn 198.3. Protein sequences are indicated in (Table S4).

Figure S1

N Lobe

P.t._SPT16-1	1	-----MDRVCALQNKKANNGLAIVTLK-----ANQHCPLAVLTPSHTSNLGIQKRFQ
T.t._SPT16	1	-----MSKQKADLRYQHHCITLVMNK-----QPQYANTDAFTLNKPKHGGNKIKTSALS
P.t._SPT16-2c	1	-----MDKKIADLVGCHDQITSLND-----DIPACITLSEKELGSKPKKTKALF
P.t._SPT16-2a	1	-----MDKKIADLVGCHDQITSLNE-----DIPACITLSEKELGSKPKKTKALF
P.t._SPT16-2b	1	-----MDKKIADLVGCHDQITSLND-----DIPACITLSEKELGSKPKKTKALF
A.t._SPT16	1	MADSRNGNARAPPSGVPPKAGNTYSIVKNLISRAALAEVWWR--HS-ADLWGSADALATPPASDLRLVLSAALN
S.p._Spt16	1	-----MAEYDEEITGHRGGLLSWNEEDG-KTLFQDCSILVTVAHDLTNPVQKSSALH
S.c._Spt16	1	-----MEELNLELVGKRLLELYSKNE-----FEGSPNSLLELVSSNANLPPKTKLH
C.e._SPT16	1	-----MSGKRAVINRDLGRCRAERTLPEWFR--G-ADGLDSIKSLAFVY--HINFNPTKTSALF
D.m._Dre4	1	-----MSSFVLEKLAIVRRMRRLYEPWAPSIGHDDALRNLCITMSLVEV--DEIVVYSKSMALQ
H.s._SPT16H	1	-----MAVTLEKLAIVRRMRRLYSNWR--G-EDEYANVDALVSVGV--DEITVYAKSMALQ
M.m._Supt16	1	-----MAVTLEKLAIVRRMRRLYSNWR--G-EDEYASIDALVSVGV--DEITVYAKSMALQ

N Lobe

P.t._SPT16-1	52	QWYLCSEIMTCTITMTMCCIADVEVMFKIKH-----SDIKMKTFTFFLITNKKNNH--QCFQAFERPRKEYP
T.t._SPT16	52	MMYFGSEFIDTITLTKKTFALICGNKINMLKSYO-----EHAEAKYENVFEDDQANNSNQLQLEETDKDKNKSS-
P.t._SPT16-2c	48	WVLFGYMTEVFAIKKQIRYLASKKLQMMEEET-----KQRLCRYFHFYKGVADNRG-----FKLRQKIGN--
P.t._SPT16-2a	48	WVLFGYMTEVFAIKKQIRYLASKKLQMMEEET-----KQRLCRYFHFYKGVADNRG-----FKLRQKIGN--
P.t._SPT16-2b	48	WVLFGYMTEVFAIKKQIRYLASKKLQMMEEET-----KHKLSKFEHFYKGMNDNRDA-----FKLRQKIGN--
A.t._SPT16	77	WVLFGYEFPDITLVFKKQIRYLCSRNKASILEVV-----KKPAHDELKIDVYHVPKPGD--GTGLMADLFAHDLISR
S.p._Spt16	59	WVLFGYEFPSTLLEKIRITLISVNRANMLTK-----AETKCAADNLRKRDADEN--KKLTKLLEYRATN-
S.c._Spt16	53	NWVLSVFPATLALVPGKTIITTSARAKHTQKATDL--FKDPESKITLFWQNNKPEPL-NKKTPTDVALINSAG-
C.e._SPT16	56	WVLFGEHLDITVLLLDHLYVLSNKRVEFFGSITGD--NQSSGKVPSTLIRDTDKDA--GNFKTIDHLSAG-
D.m._Dre4	59	WVLFGEHLDITVLFCSDAVYVLSKKEEFLKQTP--NITERRPPENLVIRDTDKQ--GNFKTKAKQNSKK
H.s._SPT16H	54	WVLFGEHLDITLVFCDDKILFMASKRKEEFLKQANTKGNENANCAPATLIRDRNESNK--SSEDRMELARSRKN
M.m._Supt16	54	WVLFGEHLDITLVFCDDKILFMASKRKEEFLKQANTKGNENANCAPATLIRDRNESNK--SSEDRMELARSRKN

		N Lobe	Peptidase M24
P.t._SPT16-1	124	G---NNYRLAIVLSLQKSPLETEINQFID---QNHLYVTCISFTKELINNDNKDIFEYNTCGKINYYMKPVSQR	
T.t._SPT16	127	-----FNLGELAKQQVPEPTLEDSITKD---KNGKPAQSVFVQCICSWKQNEISLSEKAAKSVYBESLJLKE	
P.t._SPT16-2c	115	-----GKINPTTQQAGSLATEMYIY-----KGSSTVFAASQITSDVAVKDDDEQAFINQSSQITRIRFKKTLIKQ	
P.t._SPT16-2a	115	-----VKICNPPTTKQAGSLAAMWYIY-----KWCQIVLANQITSDVAVKDDDEQAFINQSSQITRIRFKKTLIKQ	
P.t._SPT16-2b	115	-----VKICNPPTTKQAGSLAAMWYIY-----KWCQIVLANQITSDVAVKDDDEQAFINQSSQITRIRFKKTLIKQ	
A.t._SPT16	151	GDGNSQVVEHIAEAPPEKLEETVTRPK---NANCFVDTGGSDLVAVKDDDEMSVHKAYIAYSVMKVWVFN	
S.p._Spt16	131	-----KKGVFKKTKCKEINENDSIFEPV--KSPENIVFASLGLAKCIAKDFEANTIGASRYSVAMMSYFIDE	
S.c._Spt16	129	-----KTYGPEKISYCKEIVTENPVWEAAVKENFNVITSLGLSKWPEFKVNEQAFISVSKCKDKFMDLISNE	
C.e._SPT16	130	-----GLVGNFVKKRFSSEVSSNNKALE---EGGVKNDVTLAPTHLFAVKDDKEDLIRKSAQAFASATAARAR	
D.m._Dre4	132	G---KRTGMFAKAYECPSEAKKSTT---ASKTEHVDISTLAYLMCKKDSINNTIKASLWSDMIRNVIKDE	
H.s._SPT16H	131	G---KKGVFSKRFKEDKKSNSLQIN---KGGDKLDSAWAYTAVKDDCENLMLKKAASITSEVNEKPFKER	
M.m._Supt16	131	G---KKGVFSKRFKEDKKSNSLQIN---KGGDKLDSAWAYTAVKDDCENLMLKKAASITSEVNEKPFKER	

Peptidase M24

P.t._SPT16-1	196	LELAIKFNENTNYSITQAKREKSSLNQMAIRKPEL-----Q-GNMDLS-STVQSGGYNVMSASE-STQSRIV
T.t._SPT16	197	LETTIDEGKPTHSQATMIFGITNPEFLKKSEFICG-----ESDNIDAYVPTVQSGGYNLKNPAQSNFVLS
P.t._SPT16-2c	182	LESLIVGTRITQDAKKVEQSLNDRQ--KIMKEIGL-----QDGLVFAVPTIQSGGNY--QVVDGPKRYLS
P.t._SPT16-2a	182	LESLIVGTRITQDAKKVEQSLNDRQ--KIMKEIGL-----QDGLVFAVPTIQSGGNY--QVVDGPKRYLS
P.t._SPT16-2b	182	LESLIVGTRITQDAKKVEQSLNDRQ--KIMKEIGL-----QDGLVFAVPTIQSGGNY--QVVDGPKRYLS
A.t._SPT16	227	LESALEKLVTHSALMLLEKALLDPTK---ASVRL-----KPEVDLGYPTIQSGGKDLKLSAASNDLIT
S.p._Spt16	203	LESTYIQGKKTHSKFSIQESLITNFAFQTKSLKEDI-----DLQTEWCTPTIQSGGSLDKSATIDDRNTH
S.c._Spt16	202	LVRAVEELKRTNARKSSEKFNKTDIVFLKQSPDLSALCPPNYKFNFDLHWYVPTIQSGKREDFVRSRSTNICTY
C.e._SPT16	199	YVITLQKRVRSVLSNEFAARFKLSIV-----QDGL-----AKYEALTYVPTIQSGGNYSKNHHSSESHLH
D.m._Dre4	203	IMVILSIRKVKNKSLSGCAALGKRY-----TSEI-----DPRLEMYVPTIQSGGYSKLSAVANLH
H.s._SPT16H	202	VMVIVAEKRVRSKLAASVEKALDGRY-----LARA-----DPSTVVEVYPTIQSGGNYMKLSVVSIRNHLH
M.m._Supt16	202	VMVIVAEKRVRSKLAASVEKALDGRY-----LARA-----DPSTVVEVYPTIQSGGNYMKLSVVSIRNHLH

Figure S2

Peptidase M24

P.t._SPT16-1 265 G---VITVSPCCQVMSQSYCTRTTIFQVNCLEOTFRVITNQHAFAG--LVKRDITQFKQVIRETONVETTFKDTPEY
 T.t._SPT16 269 YC---TITVSYGIRYMEYHIANVRTTFLDPTNDRKRIYRKYELNCAV-CLEKPLIKKIYVNAVNFN---EYVQI
 P.t._SPT16-2c 250 SE---VITISIGIQVNEYNINCLRTTIFINPTIEVKKLYNITIEVCSKILIG--LVVETPLINGVYKEAVPLIQ---QKICE
 P.t._SPT16-2a 250 SE---VITITIQIGTQVNEYNINCLRTTIFINPTIEVKKLYNITIEVCSKILIT--LVITIEVSNVYKESIQIQLQ---YKICE
 P.t._SPT16-2b 250 SE---VITITIQIGTQVNEYNINCLRTTIFINPTIEVKKLYNITIEVCSKILIT--LVITIEVSNVYKESIQIQLQ---YKICE
 A.t._SPT16 294 YCPASITTCALGARYNSYCSNVARTVITDAISTCSKAVFVITKAHEAATD-ALRSRKRNTVYQAAISVVE---KNAPPF
 S.p._Spt16 276 G---VITCSIGIRYKSYCSNVGRTVITFEDSECCQNSFVAACKKFE-YCDEAVVGDVYTKLIGLR---ARFEL
 S.c._Spt16 202 GN---GCLASGELRYNNYCSNIRRTFLIDPEEMANNVDFLITKELITNLEPFRTEKVVYKESVLEYE---KTKPEL
 C.e._SPT16 265 SQ-FGTITTSPEARLSEYQNTVITRTMLTTPSSLETPAIPATAELAVTA-ALMPEAKTSIVYKIGIDTIT---EYSEKI
 D.m._Dre4 269 FG---VITCSIGARYKSYCSNIRRTFLINPTIAMQENITFLVSCVEETIK--LVVPEKTKIIVYKRTIDFVK---KKEKSEY
 H.s._SPT16H 268 FG---ALTCALGIRKSYCSNIRRTLMVDPKCEVQENYNDLQKCEELK-ELEHVEKICIVYNSVLDVVK---KQKPEL
 M.m._Supt16 268 FG---ALTCALGIRKSYCSNIRRTLMVDPKCEVQENYNDLQKCEELK-ELEHVEKICIVYNSVLDVVK---KQKPEL

Peptidase M24

P.t._SPT16-1 341 ---KMKFV-TDLCMIFG----SCMLTDNHTE-TIDRMAVVIRFVNIIVQI,PFYPERTNIAICLADITFV---VSK
 T.t._SPT16 342 ---KDKLE-AMFSGIGLEFFRESNLYNAKNEE-ELEEMVFNVYVCEINIQSE----KGLAYAQISDVAIR-KRQNT
 P.t._SPT16-2c 323 NITNVLT-SSEFNGIGLELLEPPIALSERSTELVAVNLEVYVITLENISGQ---RGISYIISGIVLVV---TNC
 P.t._SPT16-2a 323 NLQNLGLE-TSFEVNGIGLELLESCITINERSTELVTKCFYVFNISLENISGQ---KNITVYIQVGVYI---TNC
 P.t._SPT16-2b 323 NLLNLGLE-NSFEVNGIGLELLESNLITINERSTELVAKCFYVFNVMNENISGQ---KNIVYTIQVGVYI---TNC
 A.t._SPT16 370 ---VDKLT-KASAGTIGLEFFRESNLTNAKNDV-VRPKAFNVSIGENICESESRSKAKKISLLEADIVLV---TDQ
 S.p._Spt16 349 ---EPIFV-ENECAGIGLEFFRESSEVNAKNPE-VLQASVFLNISEGPNLIPHPKNSQSKYALLIDITQI---TRS
 S.c._Spt16 357 ---VPEFT-KNLESLIGLEFFRESNLTNKNDYRKIQKDCENISFENNKDSQ---SANNVQLADIVVQLPLETE
 C.e._SPT16 340 ---AETNKRPESSAGLEFFRESNLTSAKCDV-KAKAMVFTVYVSVSIFKKN-KGEKCPAVALADIVLV---KEE
 D.m._Dre4 342 ---VDKLT-KSFEVAMGIFEPRESNLTGPRCQA-LTKNMVFNHWGTSNITVPEATDKKQVYALFIDVIV---GPQS
 H.s._SPT16H 341 ---LNKLT-KNLESLIGLEFFRESNLTNSKNQY-KKKCMVSNLIGSDITKKEGKPEKLYALFICDVLV---DEDC
 M.m._Supt16 341 ---LNKLT-KNLESLIGLEFFRESNLTNSKNQY-KKKCMVSNLIGSDITKKEGKPEKLYALFICDVLV---DEDC

Peptidase M24

P.t._SPT16-1 408 IEDCVITKABEFTFVSYQPTPEGE-----RFFK--STFQVENSND-VIHQSEFVITR---EQPEQALNLIKND
 T.t._SPT16 411 PNAVNFKYSKYEDISYIQLFQDVEQEEEDDLEKE-NITQVCHTRNAYHKNITIVS-----KKEKOKH--
 P.t._SPT16-2c 394 AANVTNSIPRAYKISYQLEDEDPAQKQ---QIKQ--TDAPLEGTRAPRNOQIQIQR-----ENKQKQIHL--
 P.t._SPT16-2a 394 AATVITQCPRAYKISYQLEDEDEPKRP---ALVQ--TDKLPITAR-PRNOQIQIQR-----ENKQKQIHL--
 P.t._SPT16-2b 394 AATVITQCPRAYKISYQLEDEDEPKRP---AVVQ--TDKLPITAR-PRNOQIQIQR-----ENKQKQIHL--
 A.t._SPT16 442 KPELTK-CSISVNDVAISFKEDELE-----KLRK--KARTSSSENY-VTKTALSISD---HVVSKVLELQCL--
 S.p._Spt16 421 DPIVITD-SIPAAQGLISVFPGLDSSLDGV---KLRK--PPTVCTATISSHGKQVRSIT---RDLDDSAKRRVLE--
 S.c._Spt16 429 PPRFTIN-YTRAKSISYFNNEDEENKKNKS---SEAT--KIPSPDRNSKILRTKIRGARGGAEDAQKQIRIEN--
 C.e._SPT16 413 DNEVTEKAKSRLSNVIKFRPCCNRAEKD---NDQK--KVLGCSQSV-VLTDQVINKT-----TNEDEKMER--
 D.m._Dre4 415 PASVYFP-SKIKIKVGTFRKISSEFVDDKKTAKEDQGTETLGSKNA-VESKIRNIT-----NTPEKRFH--
 H.s._SPT16H 414 PATVITIS-VKRVKVGITFRNDFEPEPEEK---DEAE--DILGCSRAA-VLPTERTNIM-----TAPKRFAL--
 M.m._Supt16 414 PATVITIS-VKRVKVGITFRNDDLEDEEEK---DEAE--DILGCSRAA-VLPTERTNIM-----TAPKRFAL--

Spt16

P.t._SPT16-1 471 EKKKEIKQYLEVRINDQ--TRQPKLLVKDQQAQQLVDFDQ---YFGEIADQDKSALEPILTHYPHALI
 T.t._SPT16 478 QLEREVLRKLEQRYNNGFLSNKINSALEDKVQYGGPQDIPKE--YKINQIHDAAHNALLEPWNELVPPHISL
 P.t._SPT16-2c 459 EKLAKDKQLEQREDDQVFNQSOETVALEDKLQYQVSEVYKPE--IQGGYIINQKCAVLEPILTHVPPHISC
 P.t._SPT16-2a 457 EKLAKDKQLEQREDDQVFNQSOETVALEDKLQYQVSEVYKPE--IQGGYIINQKCAVLEPILTHVPPHISC
 P.t._SPT16-2b 457 EKLAKDKQLEQREDDQVFNQSOETVALEDKLQYQVSEVYKPE--IQGGYIINQKCAVLEPILTHVPPHISC
 A.t._SPT16 504 QALARCKNDITARVAGDSSGAGDSRSRTAKTSADVAVKVNNDVHK----EIMQVITENAVLLEPILTHVPPHIAL
 S.p._Spt16 489 EKLASFKADLQVPAQGSVPSSGLEK--TVKSEFYKDSGLQA--GEIRLVLYAQSLLEPILTHVPPHISL
 S.c._Spt16 501 EKKHEKLEKNELLSAADA--GPDSEPRQYRKYESTVDSQVETN--RDRRHVDMVQSITLEPILTHVPPHISL
 C.e._SPT16 478 EKETVQLNELKARISKGGTDEKKS---KSNVYKTEERFQADQKMLTVDPRYDSVVPVPEKIVPPHISM
 D.m._Dre4 484 QRELAQQLNERKIDRABGNSKEVEKV---KNVSYKSISSMPPEPEKELKIVYDKKYETVIMPEVETVPPHISL
 H.s._SPT16H 478 ERELAQQLNERKIRTEKGEVQKAE---KSNVYKNSLMPKPEHREKLYDKKYETVIMPEVETVPPHIAL
 M.m._Supt16 478 ERELAQQLNERKIRTEKGEVQKAE---KSNVYKNSLMPKPEHREKLYDKKYETVIMPEVETVPPHIAL

Spt16

P.t._SPT16-1	546	LKNSVKLPLNGAGEHTFVWNEFHID---TR-FPEFM-LQDMFKEITLNQEFIK-----LQDIPNEINVCP
T.t._SPT16	556	LKNYSKNDGKTIHFRNHNHPCGSGN-LAN-TPEKI-DQIVFQKDTFRSKN-----AKNLETIHPTK
P.t._SPT16-2c	537	LKNMG-----SRINFTSEITAG---QI-QFQI-DDETIHKEQYRSK-----SDPNNLILQIKO
P.t._SPT16-2a	535	LKNYSKIDGKMGSRINFTSEITAG---QI-QFQV-DDETIHKEQYRSK-----SDPNNLILQIKS
P.t._SPT16-2b	535	LKNYSKIDGKMGSRINFTSEITAG---QI-QFQV-DDETIHKEQYRSK-----SDPNNLILQIKS
A.t._Spt16	580	HTVYSGNITNRNCRIRLINVETIPFNPHSN-SLKNQ-GAL-VEKVTSTRIRD-----SRISSEVTOQLT
S.p._Spt16	565	LKNYSKNDGKTIHFRNHNHPCGSGN-LAN-TPEKI-DQIVFQKDTFRSKN-----AKNLETIHPTK
S.c._Spt16	578	YKNSKNDGKTIHFRNHNHPCGSGN-LAN-TPEKI-DQIVFQKDTFRSKN-----AKNLETIHPTK
C.e._SPT16	554	LKNYSKNDGKTIHFRNHNHPCGSGN-LAN-TPEKI-DQIVFQKDTFRSKN-----AKNLETIHPTK
D.m._Dre4	560	LKNYSKNDGKTIHFRNHNHPCGSGN-LAN-TPEKI-DQIVFQKDTFRSKN-----AKNLETIHPTK
H.s._SPT16H	554	LKNYSKNDGKTIHFRNHNHPCGSGN-LAN-TPEKI-DQIVFQKDTFRSKN-----AKNLETIHPTK
M.m._Supt16	554	LKNYSKNDGKTIHFRNHNHPCGSGN-LAN-TPEKI-DQIVFQKDTFRSKN-----AKNLETIHPTK

P.t._SPT16-1	613	DARRKQIEKLLVDVDFVLEKTLVLPKN--TH-CLSKYVMRDTQSQKHSPEKRVICLHNGFRYK-ARGL-VIDFTIT
T.t._SPT16	621	LCARVQKTDQANNDDEIVQKQCL-ENITRE-ALNFKVREPAISK-CAVN-GLDELHNGFRYMI-TIND-KVDLTK
P.t._SPT16-2c	593	LQKAVRANQVDFEERQVGLLHLLSISGKRK-IFKDLVRRTPFGS-GAA-GLDMHNGFRYIII-SNKI-QDDIVK
P.t._SPT16-2a	599	LQKAVRANQVDFEERQVGLLHLLSISGKRK-IFKDLVRRTPFGS-GAA-GLDMHNGFRYIII-SNKI-QDDIVK
P.t._SPT16-2b	599	LQKAVRANQVDFEERQVGLLHLLSISGKRK-IFKDLVRRTPFGS-GAA-GLDMHNGFRYIII-SNKI-QDDIVK
A.t._SPT16	645	LRQVMARESERADATVQKTLGAGNFKLLELSELVRRPFGSKTIP-SLEBAHNGFRYSII-TREPRVLDLA
S.p._Spt16	629	YKAAATKRETRRFADVQKTFEIKNKPATVNRVAVRPA-DG-KRTP-SLEBAHNGFRYSII-TREPRVLDLA
S.c._Spt16	646	YKAAATKRETRRFADVQKTFEIKNKPATVNRVAVRPA-DG-KRTP-SLEBAHNGFRYSII-TREPRVLDLA
C.e._SPT16	628	YKAAATKRETRRFADVQKTFEIKNKPATVNRVAVRPA-DG-KRTP-SLEBAHNGFRYSII-TREPRVLDLA
D.m._Dre4	634	YKAAATKRETRRFADVQKTFEIKNKPATVNRVAVRPA-DG-KRTP-SLEBAHNGFRYSII-TREPRVLDLA
H.s._SPT16H	628	YKAAATKRETRRFADVQKTFEIKNKPATVNRVAVRPA-DG-KRTP-SLEBAHNGFRYSII-TREPRVLDLA
M.m._Supt16	628	YKAAATKRETRRFADVQKTFEIKNKPATVNRVAVRPA-DG-KRTP-SLEBAHNGFRYSII-TREPRVLDLA

P.t._SPT16-1	688	SIKLCEVSPEDVLIACIHLLHMPKCKGIMFSQIQVYRQIEGAEQDAR-----SKVLEHLLHFFLKKVQDFSL
T.t._SPT16	695	NIKHAFOPCGEMTILHFLHKNPIMGKKVMDVQVYRQIEGAEQDAR-----SKVLEHLLHFFLKKVQDFSL
P.t._SPT16-2c	668	NIKHYIQCSPQDILAHFLHLSPIVGGKHTHDVQVYRQIEGAEVHEDER-----KTNRNDDEDEDERLNMR
P.t._SPT16-2a	674	NIKHYIQCSPQDILAHFLHLSPIVGGKHTHDVQVYRQIEGAEVHEDER-----KTNRNDDEDEDERLNMR
P.t._SPT16-2b	674	NIKHYIQCSPQDILAHFLHLSPIVGGKHTHDVQVYRQIEGAEVHEDER-----KTNRNDDEDEDERLNMR
A.t._SPT16	723	NIKHAFOPCGEMTILHFLHKNPIMGKKVMDVQVYRQIEGAEVHEDER-----KTNRNDDEDEDERLNMR
S.p._Spt16	706	NIKHAFOPCGEMTILHFLHKNPIMGKKVMDVQVYRQIEGAEVHEDER-----KTNRNDDEDEDERLNMR
S.c._Spt16	723	NIKHAFOPCGEMTILHFLHKNPIMGKKVMDVQVYRQIEGAEVHEDER-----KTNRNDDEDEDERLNMR
C.e._SPT16	703	NIKHAFOPCGEMTILHFLHKNPIMGKKVMDVQVYRQIEGAEVHEDER-----KTNRNDDEDEDERLNMR
D.m._Dre4	709	NIKHAFOPCGEMTILHFLHKNPIMGKKVMDVQVYRQIEGAEVHEDER-----KTNRNDDEDEDERLNMR
H.s._SPT16H	703	NIKHAFOPCGEMTILHFLHKNPIMGKKVMDVQVYRQIEGAEVHEDER-----KTNRNDDEDEDERLNMR
M.m._Supt16	703	NIKHAFOPCGEMTILHFLHKNPIMGKKVMDVQVYRQIEGAEVHEDER-----KTNRNDDEDEDERLNMR

Rtt106

P.t._SPT16-1	761	EELRNDESFQQSEQY----YKRFNKKFERLEQYSSENYARERVREOPTOSCLVNIVOPEFDTLENVDMCCERV
T.t._SPT16	767	KLNKIDRDAIKDVENQ----LGDKIKFVEMANLGVYSSESSTCLLOPTONILNIIIDPPTFMSLAEVEACFERM
P.t._SPT16-2c	742	KVAKREVEFKTDEE----GADYKISPKKPRDLGRNWNRARIFLOPRTDILNIVDSPEFFITSEVEICCPERV
P.t._SPT16-2a	748	KVAKREVEFKTDEE----GADYKISPKKPRDLGRNWNRARIFLOPRTDILNIVDSPEFFITSEVEICCPERV
P.t._SPT16-2b	748	KVAKREVEFKTDEE----GADYKISPKKPRDLGRNWNRARIFLOPRTDILNIVDSPEFFITSEVEICCPERV
A.t._SPT16	796	KLNKIDRDAIKDVENQ----LGDKIKFVEMANLGVYSSESSTCLLOPTONILNIIIDPPTFMSLAEVEACFERM
S.p._Spt16	782	KVAKREVEFKTDEE----GADYKISPKKPRDLGRNWNRARIFLOPRTDILNIVDSPEFFITSEVEICCPERV
S.c._Spt16	803	KLNKIDRDAIKDVENQ----LGDKIKFVEMANLGVYSSESSTCLLOPTONILNIIIDPPTFMSLAEVEACFERM
C.e._SPT16	775	KVAKREVEFKTDEE----GADYKISPKKPRDLGRNWNRARIFLOPRTDILNIVDSPEFFITSEVEICCPERV
D.m._Dre4	781	KVAKREVEFKTDEE----GADYKISPKKPRDLGRNWNRARIFLOPRTDILNIVDSPEFFITSEVEICCPERV
H.s._SPT16H	775	KVAKREVEFKTDEE----GADYKISPKKPRDLGRNWNRARIFLOPRTDILNIVDSPEFFITSEVEICCPERV
M.m._Supt16	775	KVAKREVEFKTDEE----GADYKISPKKPRDLGRNWNRARIFLOPRTDILNIVDSPEFFITSEVEICCPERV

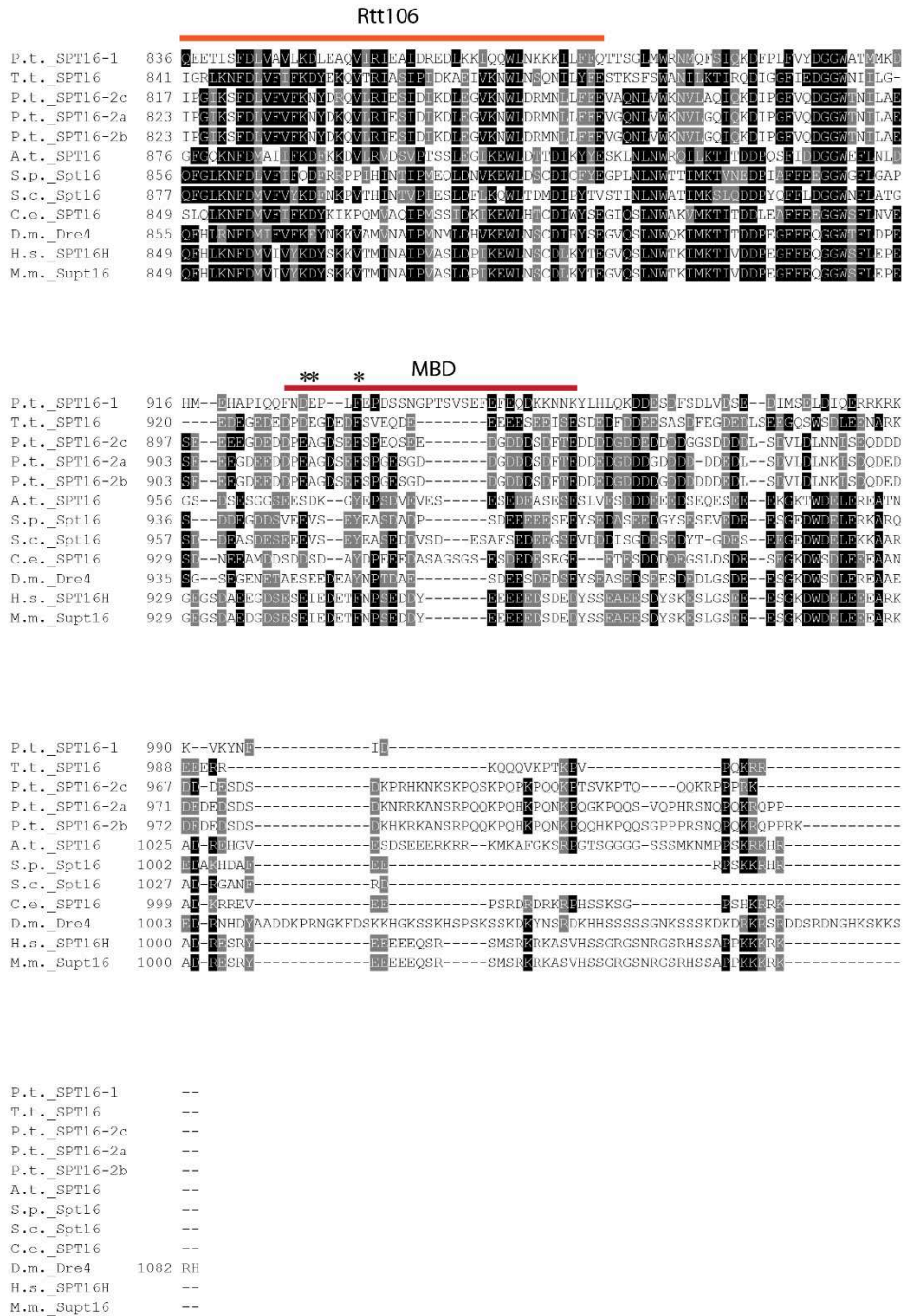


Figure S2. Spt16 is a conserved protein among eukaryotes
 MUSCLE alignment of Spt16 homologues from *P. tetraurelia* and other eukaryotes. Domains identified by Interpro are indicated in colors. Conserved aromatic and acidic residues in MBD are shown with asterisk (*). Conserved residues are highlighted in black and grey (amino acid with same physical and chelical properties). Pt. *Paramecium tetraurelia*, Tt *Tetrahymena thermophila*, Sp. *Schyzosaccharomyces pombe*, Sc. *Saccharomyces cerevisiae*, Ce. *Caenorhabditis elegans*, Dm. *Drosophila melanogaster*, Hs. *Homo sapiens*, Mm. *Mus musculus*, At. *Arabidopsis thaliana*.

Figure S2

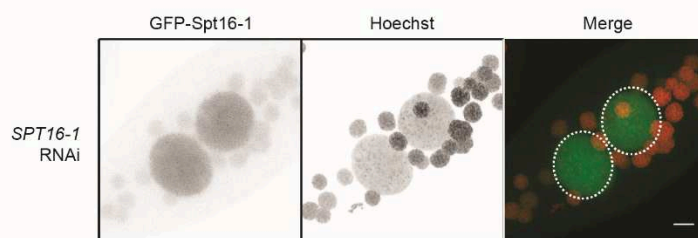


Figure S3. The GFP-Spt16-1 fusion protein is functional

Overlay of Z-projections of magnified views of GFP-SPT16-1 (green) and Hoechst (red) in *SPT16-1* RNAi during MAC development. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μ m.

Figure S3

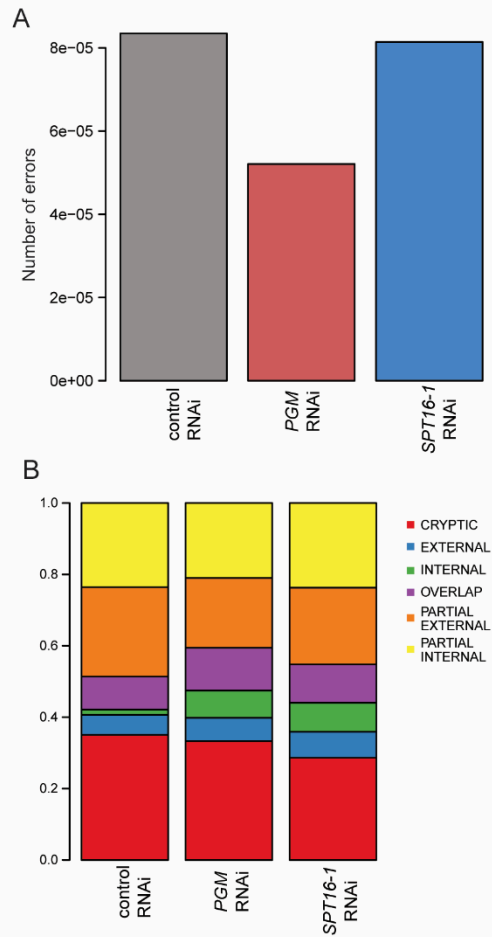


Figure S4. SPT16-1 RNAi do not lead to increased excision errors

A. Quantification of excision errors between control, *SPT16-1* and *PGM* RNAi. There is no significant difference between the three conditions. B. Quantification of IES excision errors by type of error between control, *SPT16-1* and *PGM* RNAi. Quantification and identification of error type is made with PARTIES software described in (Denby Wilkes et al., 2016) by type of errors. There is not significant difference between the three conditions.

Figure S4

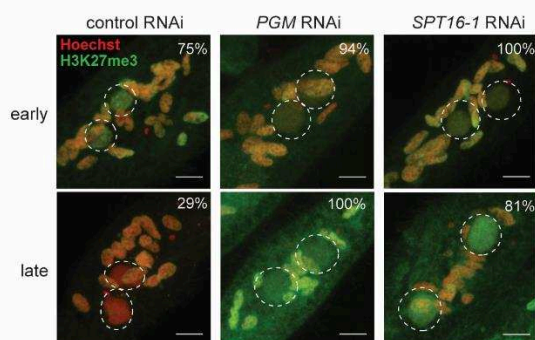


Figure S5. Spt16-1 is not required for H3K27me3 deposition in new developing MACs

Overlay of Z-projections of immunolabelling with H3K27me3 antibodies (green) and staining with Hoechst (red) in control, *SPT16-1* or *PGM* RNAi at early and late development time-points. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μ m. Percentage of H3K27me3 positive cells are indicated.

Figure S5

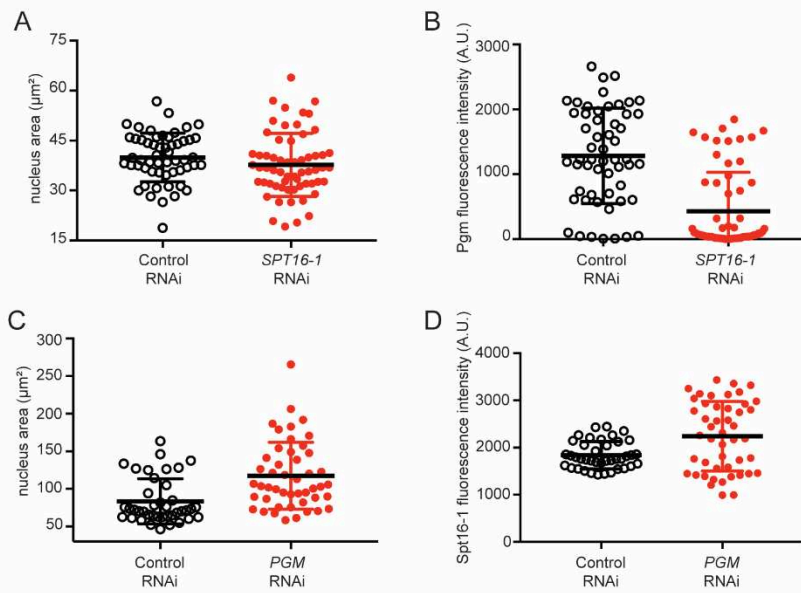


Figure S6. Distribution analysis of Pgm or Spt16-1 fluorescence quantification in developing macronuclei after *SPT16-1* or *PGM* RNAi

A. Comparison of nuclei area distributions after control or *SPT16-1* RNAi in T5 after quantification of Pgm fluorescence signal in developing MACs. B. Comparison of Pgm fluorescence intensity distributions after control or *SPT16-1* RNAi in T5 after quantification of Pgm fluorescence signal in developing MACs. C. Comparison of nuclei area distributions after control or *PGM* RNAi in late time point of development (T50) after quantification of GFP-Spt16-1 fluorescence signal in developing MACs. D. Comparison of Spt16-1 fluorescence intensity distributions after control or *PGM* RNAi in late time point of development (T50) after quantification of GFP-Spt16-1 fluorescence signal in developing MACs.

Figure S6

Figure legends

Figure 1. *Paramecium tetraurelia* Spt16 proteins.

A. Conserved domains (colored boxes) in *Paramecium tetraurelia* Spt16-1 protein. B. Phylogenetic tree of Spt16 proteins from *P. tetraurelia* (Pt), *Tetrahymena thermophila* (Tt), *Saccharomyces cerevisiae* (Sc), *Schizosaccharomyces pombe* (Sp), *Caenorhabditis elegans* (Ce), *Drosophila melanogaster* (Dm), mouse (Mm), human (Hs) and *Arabidopsis thaliana* (At) based on the alignment of full-length protein sequences. The tree was generated with PhyML 3.0 with bootstrapping procedure and visualized with Tree.Dyn 198.3. Accession numbers are provided in Table S1. C. SPT16 gene expression profiles during the life cycle. Mean mRNA expression levels were determined by RNA sequencing during vegetative growth and at different time points during autogamy (Arnaiz et al. 2017). The vegetative point consists of two samples from mass cultures containing only log-phase cells showing no sign of meiosis. The meiosis time point consists of two samples containing 32-39% of cells undergoing meiosis and little fragmentation of the maternal MAC. The fragmented time point consists of three samples containing a similar proportion of meiotic cells (20-23%) as the meiosis time point, but also contained 40-47% of cells with a fragmented maternal MAC. The Development 1 time point consists of two samples containing 50-56% of cells with a fragmented maternal MAC and 35-37% of cells that contained two already visible new MACs. The development 2/3 time point consists of four samples containing 73-91 % of cells with two new MACs. Development 4 time point consists of two samples taken 10 hours after the previous time point and consisting of 91-98% of cells with two new MACs and in which appear 1-9% of post-autogamous cells.

Figure 2. Spt16-1 is essential for development.

A. Production of post-autogamous sexual progeny following *SPT16-1* gene silencing. The gene targeted in each silencing experiment is indicated. Two non-overlapping silencing fragments (#1 and #2) of the *SPT16-1* gene were used independently. The *ND7* or *ICL7* genes were used as control RNAi targets, since their silencing has no effect on sexual processes (Bouhouche et al., 2011). Autogamy was also performed in standard *K. pneumoniae* medium (no RNAi). Cells were starved in each medium to induce autogamy and, following 3-4 days of starvation, autogamous cells were

transferred individually to *K. pneumoniae* medium to monitor growth of sexual progeny. The total number of autogamous cells analyzed for each RNAi and the number of independent experiments (in parenthesis) are indicated. Death in progeny after *SPT16-1* silencing was observed after less than three cell divisions, as for *PGM* silencing. B. Z-projections of Hoechst staining of control, Pgm or Spt16-1-depleted cells fixed at late stages of development during autogamy. Dashed black circles indicate the two developing MACs. Black arrows indicate the MICs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μ m. C. *SPT16-1* gene silencing following mating pair separation. Reactive cells were obtained in standard *K. pneumoniae* medium, and mating pairs were transferred to silencing medium. The gene targeted in each silencing experiment is indicated. Only silencing fragment #1 of the *SPT16-1* gene was used. The next day, individual sexual progenies were transferred individually to *K. pneumoniae* medium to monitor growth. The total number of cells analyzed for each RNAi and the number of independent experiments (in parenthesis) are indicated.

Figure 3. A functional GFP-Spt16-1 fusion protein is localized in the developing macronuclei.

A. Localization of the GFP-Spt16-1 fusion protein expressed from the RNAi-resistant *GFP-SPT16-1* transgene (see Materials and Methods). Overlay of Z-projections of magnified views of GFP-Spt16-1 (in green) and Hoechst (in red) are presented (a-f), aligned with their schematic representations on the right. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC or the MICs. Scale bar is 10 μ m. The inset displays one the two developing new MACs. Transgene copy number per haploid genome was 19 as measured by qPCR from total genomic DNA extracted from vegetative cells and values were normalized to the endogenous *GAPDH* gene. B. *In vivo* genetic complementation with the RNAi-resistant *GFP-SPT16-1* transgene upon *SPT16-1* gene silencing during autogamy. Production of sexual of post-autogamous progeny following complementation of *SPT16-1 RNAi* by *GFP-Spt16-1 RNAi* resistant is indicated. The gene targeted in silencing experiment is indicated. Silencing is realized in non-injected cells (NI) or GFP-Spt16-1 injected cells (GFP-Spt16-1). *ICL7* gene is used as control RNAi target, since its silencing has no effect on sexual processes (Bouhouche et al., 2011). Cells were starved in each medium to induce

autogamy and, following 3-4 days of starvation, autogamous cells were transferred individually to *K. pneumoniae* medium to monitor growth of sexual progeny. The total number of autogamous cells analyzed for each RNAi and the number of independent experiments (in parenthesis) are indicated.

Figure 4. Spt16-1 is required for DNA elimination events

A. PCR analysis of the excision of some IES using primers located each side of the IES after autogamy. On the left, are named each IES. On the right are listed corresponding size of each IES. The gene targeted in silencing experiment is indicated. As detailed in Figure 2, autogamy was also performed after no RNAi or control RNAi. As expected there was no IES retention for no RNAi or control RNAi but there was IES retention for *PGM* gene silencing. In each lane, upper band corresponds to the non-excised form whereas lower band corresponds to excised IES form and is mostly amplified from fragments of the maternal MAC. B. Maternal inheritance of macronuclear deletion is analyzed after Control or *SPT16-1* RNAi using a cell line reproducibly deleting the *ND7* gene and *A* gene from the MAC genome at each sexual generation. PCR is performed on DNA extracted during autogamy time course experiment with primers each side of IES included in *A* gene. Upper band corresponds to non-deleted *A* gene with retained IESs. Lower band corresponds to non-deleted *A* gene with excised IESs. In control RNAi, *A* gene is deleted before autogamy and progressively deleted in last time points after IES have been excised. C. PCR detection of IES 51A4578 circles with diver-gent primers on genomic DNA at each time point during autogamy time course experiment silencing control or *SPT16-1* gene. D. IES retention scores after *PGM* and *SPT16-1* RNAi. Histograms of retention scores for left and right boundaries of all IESs, as described in Materials and Methods, for control (black), *PGM* RNAi (blue) and *SPT16-1* RNAi (red). Spearman's rank correlation between right and left boundaries is given for each RNAi. E. Violin plot superimposed with a boxplot of normalized coverage using DESeq2 (see Guerin et al.) of 1 kb windows of the MIC assembly for a Klebsiella (KLEB grey) sample, MIC sample (green), *PGM* RNAi sample (blue) and *SPT16-1* RNAi sample. The windows are split into two categories defined in Guerin et al. 2017: MAC-destined compartment and MIC-limited compartment. F. Retention of some class II DNA transposons imprecisely eliminated after *PGM* and *SPT16-1* RNAi. The bar plots represent read coverage of 8 individual copies of the Sardine transposon

(GenBank Accession No. HE774468-HE774475). The coverage was determined by mapping reads for the control dataset (black), *PGM* silencing (blue) and *SPT16-1* silencing (red) datasets. The normalized units (RPKM) are reads per kilobase of the transposon sequence per million library read mapped against the MAC reference genome. As expected all transposable elements are retained in *PGM* RNAi (Arnaiz et al., 2012, Guerin et al., 2017)

Figure 5. *Spt16-1* is required for iesRNA accumulation and for H3K27me3 and H3K9me3 foci formation

A. Analysis of iesRNAs populations in *SPT16-1* RNAi. Small RNA libraries corresponding to three time-points during autogamy: precocious (app. 20% of cell population undergoes meiosis), early (10h after the first time point), and middle (20h after the first time point), after control or *SPT16-1* RNAi were sequenced and mapped to the reference genomes (*Paramecium tetraurelia* MAC reference genome and MAC+IES reference genome). Bar plots show the normalized proportion of sRNA reads that match the MAC genome (green), annotated IESs (yellow) or to the feeding vector (purple). The histograms display the cytological stages of the cell population at each developmental time point. B. Overlay of z projections of magnified views of H3K9me3 specific antibodies (green) and Hoechst (red) in control or *SPT16-1* or *PGM* RNAi at early or late development time-points. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μ m. Percentage of H3K9me3 positive cells are indicated.

Figure 6. *Spt16-1* is required for Pgm localization during development

A. Overlay of Z-projections of immunolabelling with Pgm antibody (green) and staining with Hoechst (red) in control or *SPT16-1* RNAi at T5. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μ m. Next quantification by Image J macro of Pgm antibody fluorescence signal compared to nucleus size in μ m² is shown during development for control RNAi (white circles) or *SPT16-1* RNAi (red circles). Quantification of the Pgm fluorescence signal in the developing new MAC (Arbitrary Units, A. U.) as a function of nucleus area (in μ m²) at T5 upon control RNAi (white circle) and *SPT16-1* RNAi (red circle). Bar plots represent Pgm fluorescence

signal in the developing new MAC (Arbitrary Units, A. U.) normalized to the nucleus area upon control and *SPT16-1* RNAi. B. Pgm expression during development in control or *SPT16-1* RNAi. Whole proteins are extracted at T0, T5 and T10 when Pgm is normally detected by immunofluorescence. Western blot of whole cell proteins extracted at T0, T5 and T10 during autogamy upon control and *SPT16-1* RNAi. Upper panel: Pgm antibodies were used for Pgm detection. Bottom panel: Western blot of the same membrane, following stripping and incubation with alpha tubulin antibodies. Cytology of the cell population at each developmental time point is displayed in the histograms. Pgm is quantified by normalization by Tubuline signal. C. Localization of GFP-Spt16-1 fusion protein after *PGM* RNAi at late development time points (T50). Overlay of Z-projections of views of GFP-Spt16-1 (in green) and Hoechst (in red) are presented. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μm . Quantification of the GFP fluorescence signal in the developing new MAC (Arbitrary Units, A. U.) as a function of nucleus area (in microns μ^2) at a late developmental time point (T50) upon control RNAi (white circle) and *SPT16-1* RNAi (red circle).

Table S1. Accession number of proteins used in the phylogenetic analysis

Accession number, name of the protein and species are given. *Paramecium* database: <https://paramecium.i2bc.paris-saclay.fr/>, *T. thermophila* database: <http://ciliate.org/index.php/home/welcome>, *S. pombe* database: <https://www.pombase.org/>, *S. cerevisiae* database: <https://www.yeastgenome.org/>, *C. elegans* database: <https://wormbase.org/#012-34-5>, *D. melanogaster* database: <http://flybase.org/>, *M. musculus* database: <http://www.informatics.jax.org/>, *H. sapiens* database: <https://www.genenames.org>, *A. thaliana* database: <https://www.arabidopsis.org/>.

Table S2. Oligonucleotides used in this study

Table S3. Sequencing and mapping statistics for DNaseq.

Read statistics are provided for the *SPT16-1* RNAi sample sequenced for this study.

Table S4. Description of sRNA-seq data

Read statistics are provided for the control RNAi and *SPT16-1* RNAi sample sequenced for this study.

Figure S1. Identification of Spt16 homologues in *P. tetraurelia*

Phylogenetic tree of Spt16 proteins from *P. tetraurelia*, other *Paramecium* species and other eukaryotes based on alignment of full length protein sequence by MUSCLE. Phylogeny of the alignment was generated using PhyML 3.0 on Phylogeny.fr with bootstrapping procedure using 100 bootstraps. Tree has been viewed using Tree.Dyn 198.3. Protein sequences are indicated in (Table S4).

Figure S2. Spt16 is a conserved protein among eukaryotes

MUSCLE alignment of Spt16 homologues from *P. tetraurelia* and other eukaryotes. Domains identified by InterPro are indicated in colors. Conserved aromatic and acidic residues in MBD are shown with asterisk (*). Conserved residues are highlighted in black and grey (amino acid with same physical and chemical properties). P.t. *Paramecium tetraurelia*, Tt *Tetrahymena thermophila*, S.p. *Schyzosaccharomyces pombe*, S.c. *Saccharomyces cerevisiae*, C.e. *Caenorhabditis elegans*, D.m. *Drosophila melanogaster*, H.s. *Homo sapiens*, M.m. *Mus musculus*, A.t. *Arabidopsis thaliana*.

Figure S3. The GFP-Spt16-1 fusion protein is functional

Overlay of Z-projections of magnified views of *GFP-SPT16-1* (green) and Hoechst (red) in *SPT16-1* RNAi during MAC development. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μ m.

Figure S4. *SPT16-1* RNAi do not lead to increased excision errors

A. Quantification of excision errors between control, *SPT16-1* and *PGM* RNAi. There is no significant difference between the three conditions. B. Quantification of IES excision errors by type of error between control, *SPT16-1* and *PGM* RNAi. Quantification and identification of error type is made with PARTIES software described in (Denby Wilkes et al., 2016) by type of errors. There is not significant difference between the three conditions.

Figure S5. Spt16-1 is not required for H3K27me3 deposition in new developing MACs

Overlay of Z-projections of immunolabelling with H3K27me3 antibodies (green) and staining with Hoechst (red) in control, *SPT16-1* or *PGM* RNAi at early and late development time-points. Dashed white circles indicate the two developing MACs. The other Hoechst-stained nuclei are fragments from the maternal somatic MAC. Scale bar is 10 μ m. Percentage of H3K27me3 positive cells are indicated.

Figure S6. Distribution analysis of Pgm or Spt16-1 fluorescence quantification in developing macronuclei after *SPT16-1* or *PGM* RNAi

A. Comparison of nuclei area distributions after control or *SPT16-1* RNAi in T5 after quantification of Pgm fluorescence signal in developing MACs. B. Comparison of Pgm fluorescence intensity distributions after control or *SPT16-1* RNAi in T5 after quantification of Pgm fluorescence signal in developing MACs. C. Comparison of nuclei area distributions after control or *PGM* RNAi in late time point of development (T50) after quantification of GFP-Spt16-1 fluorescence signal in developing MACs. D. Comparison of Spt16-1 fluorescence intensity distributions after control or *PGM* RNAi in late time point of development (T50) after quantification of GFP-Spt16-1 fluorescence signal in developing MACs.

Résultats supplémentaires

Dans l'étude précédente (Touzeau et al., In preparation), nous avons montré que l'homologue du chaperon d'histones Spt16-1 était un acteur majeur des réarrangements du génome chez *Paramecium tetraurelia* requis pour la localisation correcte de l'endonucléase Pgm. Notre étude n'a cependant pas permis de démontrer si Spt16-1 est bien un chaperon d'histone et si son interaction avec les histones est nécessaire pour son rôle dans les réarrangements du génome et la localisation de Pgm.

I. Le domaine MBD de Spt16-1 semble essentiel à l'activité de Spt16-1 pendant le développement MAC

Dans (Touzeau et al., In preparation), nous avons proposé un modèle dans lequel, Spt16-1 en tant que chaperon d'histone réarrangerait la chromatine pour permettre l'interaction de la machinerie d'excision et Pgm avec l'ADN. La caractéristique commune aux chaperons d'histones est la présence d'un domaine riche en résidus acides. Spt16 ne fait pas exception avec son domaine C à l'extrémité C-terminal requis pour l'interaction avec les histones H2A-H2B du nucléosome et essentiel pour sa fonction (Kemble et al., 2015). Le sous-domaine minimum requis pour l'interaction avec le nucléosome au sein du domaine C est nommé MBD (Minimal Binding Domain). Nous avons montré par Blast et InterPro, que Spt16-1 disposait bien de cette extrémité C-terminal riche en résidus acides dans le domaine C et des acides aminés conservés F931 et D927 requis pour l'interaction avec les histones H2A-H2B (Figure 13A). Ces acides aminés conservés correspondent aux résidus aromatiques D968 et Y972 chez *S. cerevisiae* (Kemble et al., 2015).

Afin de confirmer que c'est la fonction de chaperon d'histones qui est requise pour Spt16-1 dans le développement macronucléaire, nous avons voulu tester si le domaine C de Spt16-1 était requis pour le développement MAC et l'élimination d'ADN. Nous avons tiré avantage de la protéine de fusion à la GFP résistante au RNAi, GFP-Spt16-1 et en avons délété le sous-domaine MBD au sein du domaine C. Les transgènes *GFP-SPT16-1* et *GFP-SPT16-1_ΔMBD* ont été injectés dans le MAC végétatif de cellules et nous avons inactivé le gène endogène *SPT16-1* à l'autogamie pour réaliser des expériences de complémentation génétique (Figure 13B). En RNAi contrôle, aucune létalité n'est observée dans la progéniture

sexuelle de cellules injectées avec la construction sauvage GFP-Spt16-1 ou avec la construction mutante *GFP-Spt16-1_ΔMBD*. GFP-Spt16-1_ΔMBD n'a donc pas d'effet dominant négatif ou toxique à l'autogamie, comme c'est également le cas pour la construction sauvage. Lorsque l'on inactive le gène endogène *SPT16-1*, comme attendu, la construction GFP-Spt16-1 complémente presque totalement le phénotype de létalité dans la progéniture sexuelle (80%) (Figure 1B). En revanche, la présence du transgène *GFP-SPT16-1_ΔMBD*, présente dans le MAC à un nombre de copies équivalent à celui de *GFP-SPT16-1* (Figure 13D), ne permet pas de sauver ce phénotype de létalité, puisque 60% de la progéniture sexuelle meurt à l'issue de l'autogamie (Figure 13B). On en conclut que le domaine MBD est nécessaire à l'obtention d'une progéniture sexuelle viable.

Afin de confirmer que le domaine MBD est aussi requis pour l'élimination d'ADN, nous avons analysé par PCR l'excision de deux IES dans l'ADN de la progéniture sexuelle des cellules transformées GFP-Spt16-1_ΔMBD, après inactivation du gène *SPT16-1* endogène. Les réactions PCR sont réalisées sur l'ADN extrait de cellules à la fin du développement MAC en utilisant des amorces de part et d'autres des extrémités de chaque IES (Figure 13C). En RNAi contrôle, comme attendu, seule la forme correspondant à l'IES excisée est détectée dans la progéniture sexuelle des cellules non injectées, des cellules transformées par les transgènes *GFP-SPT16-1* et *GFP-SPT16-1_ΔMBD*. En RNAi *SPT16-1*, la forme non excisée de l'IES provenant des nouveaux MACs est observée, en plus de la forme excisée provenant des fragments du MAC maternel. Dans les cellules GFP-Spt16-1 en RNAi *SPT16-1*, on détecte faiblement la forme non excisée. On conclut que la construction GFP-Spt16-1 est capable de restaurer partiellement l'élimination correcte d'ADN, mais que la construction mutante délétée du MBD ne parvient pas à complémente ce phénotype. Il faut de plus noter que puisque la construction GFP-Spt16-1 sauve le phénotype de létalité après un RNAi *SPT16-1*, les défauts d'excision d'IES observés ne sont pas délétères pour la survie de la progéniture sexuelle. Nos résultats semblent indiquer que le domaine MBD est requis pour l'action de Spt16-1 dans l'élimination d'ADN.

Un contrôle important dans cette analyse consiste à s'assurer que la protéine GFP-Spt16-1_ΔMBD se localise, comme la protéine GFP-Spt16-1. Or, nous ne sommes pas parvenus à détecter le signal de fluorescence GFP de la protéine de fusion délétée pendant l'autogamie dans les cellules transformées, que ce soit en microscopie confocale sur cellules fixées ou en

analyse par Western Blot à l'aide d'anticorps anti-GFP sur des extraits protéiques totaux ou nucléaires en RNAi contrôle ou en RNAi *SPT16-1* (Figure 13D). Il nous est impossible de conclure si les phénotypes de létalité et de défauts moléculaires des réarrangements dans les cellules transformées par la version délétée du MBD sont dus à l'absence du MBD dans *Spt16-1* ou bien à un problème d'expression de la protéine de fusion.

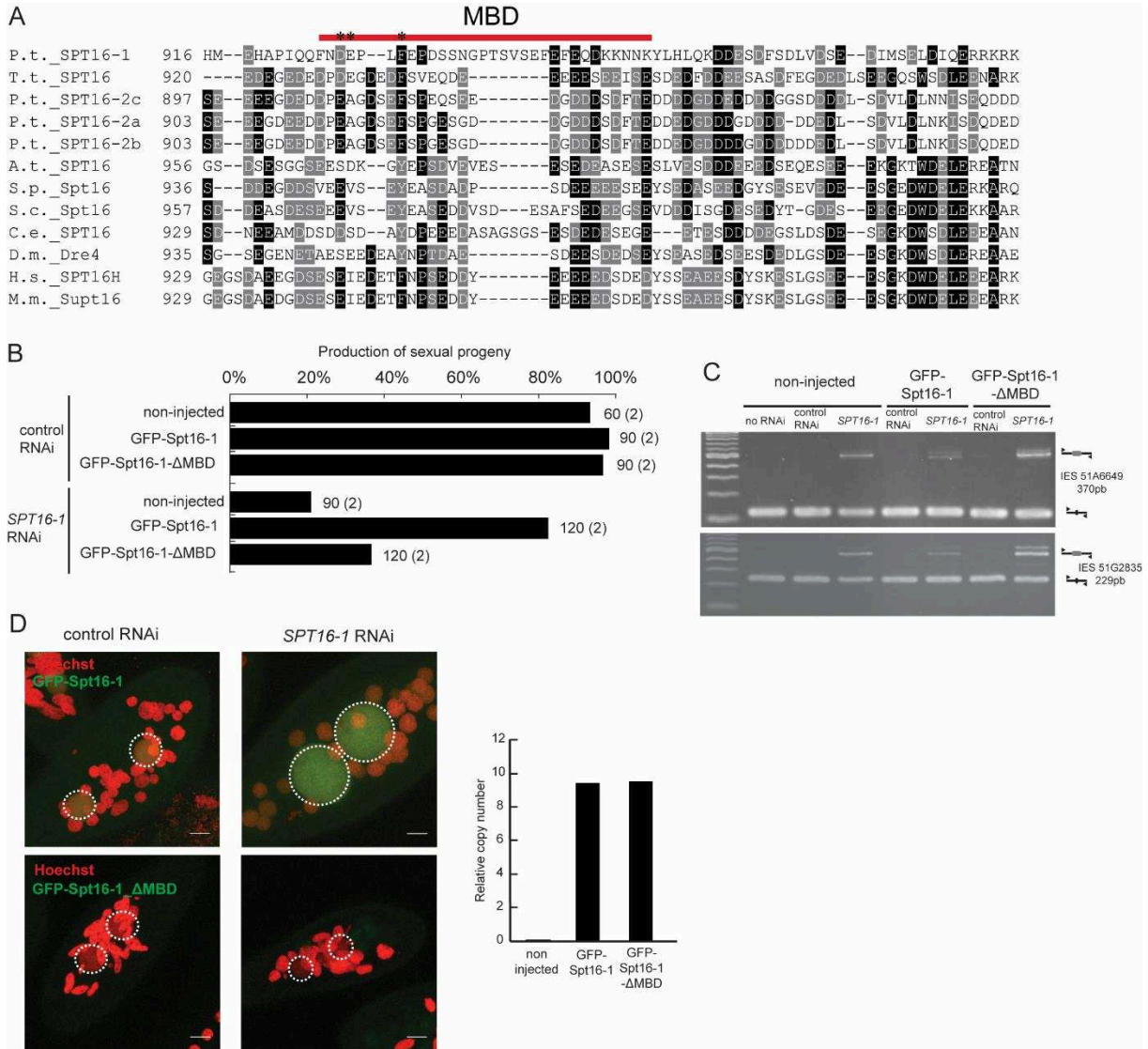


Figure 13 : Analyse fonctionnelle du domaine MBD de *Spt16-1*

A. Alignement par Muscle des domaines MBD des homologues de *Spt16* de *P. tetraurelia* et d'autres eucaryotes. Les domaines identifiés par Interpro sont indiqués en couleur avec leurs noms. Les résidus conservés sont indiqués par une *. Les couleurs noires signalent des acides aminés identiques. Les couleurs grises indiquent la conservation d'acides aminés avec des propriétés physico-chimiques identiques. P.t. *Paramecium tetraurelia*, T.t. *Tetrahymena thermophila*, S.p. *Schizosaccharomyces pombe*, S.c. *Saccharomyces cerevisiae*, C.e. *Caenorhabditis elegans*, D.m. *Drosophila melanogaster*, H.s. *Homo sapiens*, M.m. *Mus musculus*, A.t. *Arabidopsis thaliana*. **B.** Production de progéniture sexuelle viable à l'issue de l'autogamie après complémentation du RNAi *SPT16-1* par les transgènes RNAi résistants GFP-Spt16-1 et par GFP-Spt16-1-ΔMBD. Le gène ciblé dans chaque expérience est indiqué. L'inactivation est réalisée dans des cellules non injectées (NI) ou des cellules transformées par GFP-SPT16-1 ou des cellules transformées par GFP-SPT16-1-ΔMBD. Le gène *ICL7* est utilisé comme contrôle négatif de RNAi, puisque son inactivation n'a aucun effet sur les événements sexuels (Bouhouche et al.,

2011). Comme attendu, le RNAi *SPT16-1* réduit drastiquement la production de progéniture sexuelle viable. L'absence de létalité dans les cellules transformées par *GFP-SPT16-1* en RNAi *ICL7* indique que la présence du transgène n'a pas d'effet toxique. Les cellules sont mises en restriction alimentaire dans les milieux d'inactivation pour induire l'autogamie et après 3 ou 4 jours de restriction alimentaire, des cellules autogames sont transférées individuellement en milieu *K. pneumoniae* pour suivre la capacité de la production de la progéniture sexuelle à reprendre la croissance végétative. Le nombre total de cellules autogames analysées pour chaque RNAi et le nombre d'expériences indépendantes sont indiqués (entre parenthèses). La létalité dans la progéniture après le RNAi *SPT16-1* est observée après moins de trois divisions. C. Analyse par PCR de l'excision des IESs 51A6649 et 51G2835 à l'aide de d'amorces situées de part et d'autre de l'IES à la fin de l'autogamie. Le gène ciblé dans chaque expérience d'inactivation est indiqué. L'autogamie est aussi menée sans RNAi (no RNAi) ou en RNAi contrôle. L'inactivation est menée dans des cellules non injectées (NI), transformées par les transgènes RNAi-résistant *GFP-SPT16-1* ou *GFP-SPT16-1_ΔMBD*. Comme attendu, il n'y a pas de rétention d'IES en RNAi contrôle ou sans RNAi. Dans chaque piste, la bande du haut correspond à la forme non excisée de l'IES tandis que la bande du bas correspond à la forme excisée de l'IES et provient essentiellement des fragments du MAC maternel. D. Superposition de projections en Z de GFP-Spt16-1 et GFP-Spt16-1_ΔMBD (vert) et de Hoechst (rouge) pendant le développement macronudéaire de cellules transformées par GFP-Spt16-1 ou GFP-Spt16-1_ΔMBD en RNAi contrôle ou RNAi *SPT16-1*. Les cercles pointillés blancs indiquent les MACs en développement. Les autres noyaux colorés par le Hoechst correspondent aux fragments du MAC maternel. L'échelle est de 10 μm. A gauche sont donnés les nombres de copies relatifs des transgènes *GFP-SPT16-1* et *GFP-SPT16-1_ΔMBD* mesurés par qPCR à l'aide d'amorces spécifiques de la GFP. Les nombres de copies sont normalisés sur l'actine amplifiée à l'aide d'amorces spécifiques. Les nombres de copie relatifs des deux transgènes sont quasiment identiques.

II. Identification d'homologues de Pob3 chez *Paramecium tetraurelia*

Chez tous les eucaryotes, le complexe FACT est composé de Spt16 et d'une protéine Pob3 chez *S. cerevisiae* et *S. pombe* ou Ssrp1 chez les autres espèces (Jeronimo and Robert, 2016). Ce partenaire de Spt16 de plus petit poids moléculaire interagit aussi avec le nucléosome et se fixe à Spt16 via le domaine de dimérisation qu'il possède en N-terminal (domaine D). Si Ssrp1 comme Spt16 sont essentielles chez les eucaryotes, chez certains champignons comme *S. pombe*, la délétion de POB3 ne provoque pas de létalité.

1. Identification d'homologues de la protéine Pob3 chez *Paramecium tetraurelia*

Afin d'identifier un éventuel partenaire de Spt16-1, nous avons entrepris de rechercher des homologues de Pob3 et Ssrp1 chez *Paramecium tetraurelia*. En utilisant comme requête la protéine Pob3 de *S. cerevisiae*, nous avons identifié par Blast dans le génome MAC de *P. tetraurelia* 3 homologues de Pob3 : Pob3-1, Pob3-2a et Pob3-2c qui présentent des domaines fonctionnels homologues de Pob3 (Figure 14A et B) (Aury et al., 2006). Comme pour Spt16, et contrairement à la majorité des eucaryotes, Pob3 est codée par plusieurs gènes. A l'aide des séquences des protéines Pob3 de *P. tetraurelia* et de leurs homologues eucaryotes, nous avons construit un arbre phylogénétique. Les protéines de paramécie groupent ensemble et sont relativement divergentes de leurs homologues eucaryotes. Les protéines Pob3-2a et 2b sont des duplicats issus de la dernière duplication globale du génome (Aury et al., 2006). Nous avons également observé que comme pour Spt16-1,

l'ensemble des espèces de *Paramecium* ont au moins une protéine de la famille Pob3-1 et une protéine de la famille Pob3-2 (Annexe 3/Figure S1). Le nombre de duplicats dans chaque famille varie cependant beaucoup plus que pour Spt16-1 avec certaines espèces de *Paramecium* ayant deux Pob3-1 ou trois Pob3-2. L'analyse des profils d'expression des gènes *POB3* par mesure des niveaux d'ARNm lors d'une cinétique d'autogamie en conditions sauvages montre que les gènes *POB3-2a* et *2c* sont hautement exprimés tout le long du cycle de vie y compris en conditions de croissance végétative (Figure 14C). Le gène *POB3-1* est exprimé uniquement pendant l'autogamie avec un profil d'expression très similaire à celui de *PGM* ou de *SPT16-1*. Ces observations suggèrent que Pob3-1 pourrait être un partenaire de Spt16-1, impliqué dans le développement et l'élimination d'ADN.

Nous avons analysé la présence de domaines protéiques à l'aide d'Interpro dans les protéines Pob3 de *P. tetraurelia* (Figure 14A). Elles présentent une organisation modulaire similaire à leurs homologues eucaryotes avec en C-terminal une extrémité riche en résidus acides correspondant au domaine C interagissant avec les histones H2A-H2B (Hoffmann and Neumann, 2015) et en amont un domaine M également nommé domaine Ssrp1 (Zunder et al., 2012). Cependant, on observe qu'il n'y a pas de domaine en N-terminal N/D qui permet normalement la dimérisation avec Spt16 (Winkler et al., 2011). Enfin en C-terminal, on ne détecte pas le domaine HMG présent dans Ssrp1, absent de Pob3 chez les champignons. L'organisation des protéines de paramécie ressemble plus à celle de Pob3 que de Ssrp1. Pourtant, l'absence de domaine putatif de dimérisation ne soutient pas l'hypothèse de formation de dimère Pob3-1/Spt16-1. Cette organisation sans domaine HMG ni domaine de dimérisation est conservée chez les autres espèces de *Paramecium* que nous avons observées.

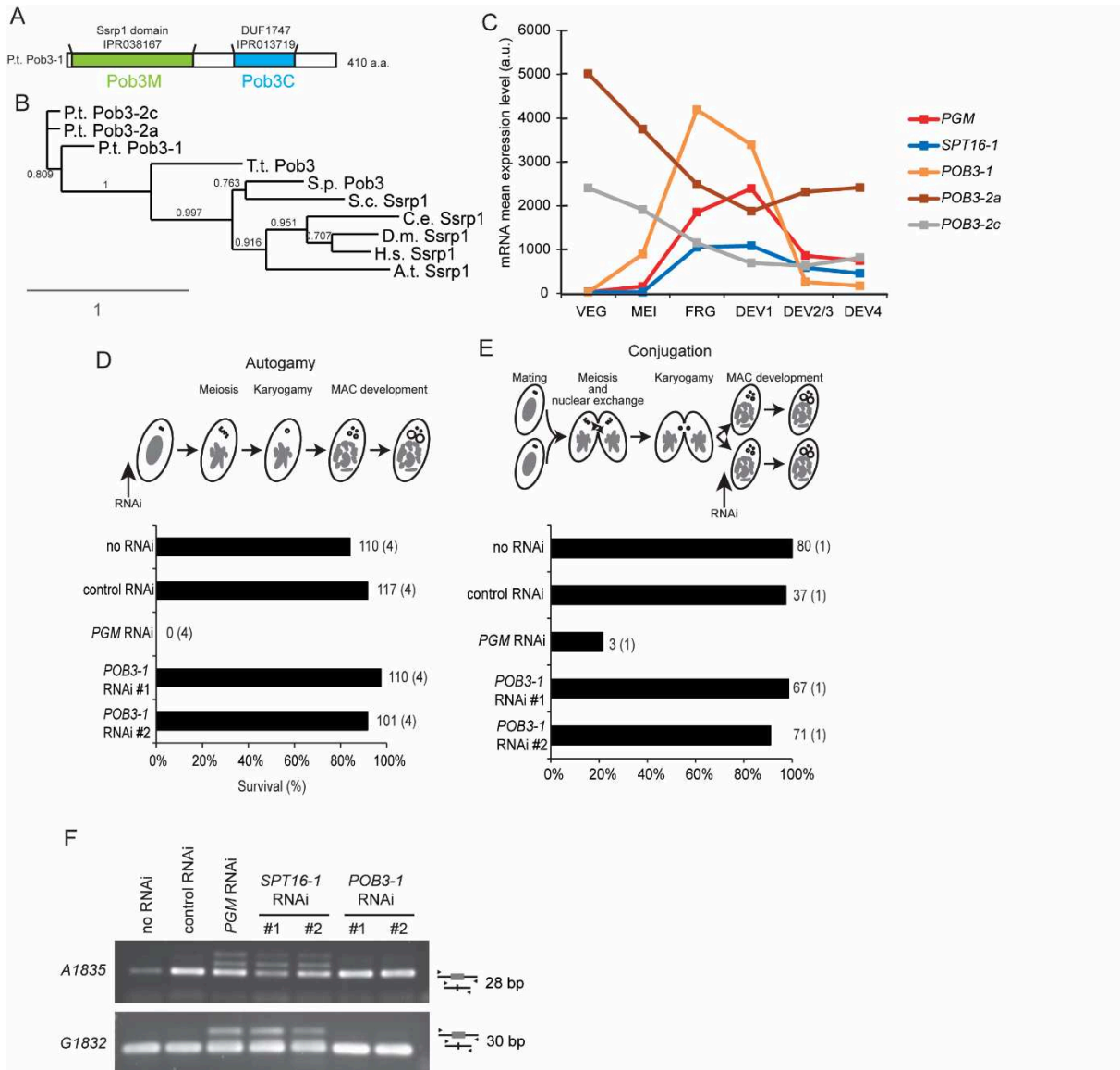


Figure 14 : Identification d'homologues de Pob3 chez *Paramecium tetraurelia*

A. Domaines prédits pour les protéines Pob3 de *Paramecium tetraurelia* (rectangles colorés). **B.** Arbre phylogénétique des protéines Pob3 chez *Paramecium tetraurelia* et d'autres eucaryotes à partir des alignements des séquences des protéines entières par MUSCLE. La phylogénie des alignements est réalisée via PhyML 3.0 sur Phylogeny.fr avec une procédure de bootstrap à 100 bootstraps. L'arbre est visualisé avec Tree.Dyn 198.3. Les séquences protéiques et leurs numéros d'accèsion sont indiqués en Annexe 3/figure S1. P.t. *Paramecium tetraurelia*, T.t. *Tetrahymena thermophila*, S.p. *Schizosaccharomyces pombe*, S.c. *Saccharomyces cerevisiae*, C.e. *Caenorhabditis elegans*, D.m. *Drosophila melanogaster*, H.s. *Homo sapiens*, M.m. *Mus musculus*, A.t. *Arabidopsis thaliana*. **C.** Profils d'expression des gènes *POB3* pendant une cinétique d'autogamie par mesure des niveaux d'expression des ARNm par RNAseq (Arnaiz et al., 2017). La description des points est disponible dans la légende de la figure 1 de (Touzeau et al., en prep). **D.** Létalité de la progéniture sexuelle après inactivation des gènes *POB3-1* et *PGM*. Le gène ciblé dans chaque expérience d'ARN interférence est indiqué. Deux fragments non chevauchants (#1 et #2) de *POB3-1* sont utilisés indépendamment. Les gènes *ND7* ou *ICL7* sont utilisés comme contrôles négatifs puisque leur inactivation n'a aucun effet sur les événements sexuels (Bouhouche et al., 2011). L'autogamie est aussi menée dans un milieu standard supplémenté en *K. pneumoniae* (no RNAi). L'inactivation de *PGM*, comme attendu, induit la létalité de la progéniture sexuelle (Baudry et al., 2009). Les cellules sont restreintes en nourriture pour induire l'autogamie et après 3 ou 4 jours de restriction alimentaire, les cellules autogames sont transférées individuellement en milieu *K. pneumoniae* pour observer la croissance de la progéniture sexuelle. Le nombre total de cellules autogames analysées dans chaque expérience d'ARN interférence et le nombre d'expériences indépendantes (entre parenthèses) sont indiqués. **E.** Létalité de la progéniture sexuelle après inactivation des gènes *POB3-1* et *PGM* pendant la conjugaison. Le gène ciblé dans chaque expérience d'ARN interférence est indiqué. Deux fragments non chevauchants (#1 et #2) de *POB3-1* sont utilisés indépendamment. Les gènes *ND7* ou *ICL7* sont utilisés comme contrôles négatifs puisque leur inactivation n'a aucun effet sur les événements sexuels (Bouhouche et al., 2011). L'autogamie est aussi menée dans un

milieu standard supplémenté en *K. pneumoniae* (no RNAi). L'inactivation de *PGM*, comme attendu, induit la létalité totale de la progéniture sexuelle (Baudry et al., 2009). Les cellules de signe sexuel opposé sont mélangées en milieu *K. pneumoniae* pour conjuguer et les paires en méiose/caryogamie sont mises en milieu RNAi pour inactiver les gènes pendant le développement macronucléaire. Après un jour de RNAi, les individus issus de la séparation des paires sont transférés individuellement en milieu *K. pneumoniae* pour observer la croissance de la progéniture sexuelle. Le nombre total de cellules autogames analysées dans chaque expérience d'ARN interférence et le nombre d'expériences indépendantes (entre parenthèses) sont indiqués. F. Analyse par PCR de l'excision de deux IES à l'aide d'amorces de chaque côté de l'IES à l'issue de l'autogamie. A droite sont listés les noms des IES et à gauche les tailles des IES correspondantes. Le gène ciblé par l'inactivation est indiqué. L'autogamie est aussi menée en condition sans RNAi ou en RNAi contrôle. Comme attendu, il n'y a pas de rétention d'IES en condition sans RNAi (no RNAi) ou en RNAi contrôle (control RNAi) mais les IES sont retenues quand *PGM* ou *SPT16-1* sont inactivés. Dans chaque piste, la bande du haut correspond à la forme non excisée et la bande du bas correspond à la forme excisée majoritairement amplifiée à partir des fragments du MAC maternel.

2. Pob3 n'est pas nécessaire aux événements sexuels ni au développement MAC

Pour déterminer l'implication de *Pob3-1* dans le développement et l'élimination d'ADN, nous avons inactivé le gène correspondant pendant l'autogamie par ARN interférence avec deux fragments non chevauchants du gène *POB3-1* (#1 et #2) (Figure 14D). Le protocole utilisé est le même que celui décrit dans (Touzeau et al., In preparation) pour l'inactivation de *SPT16-1*. Les cellules sont mises en restriction alimentaire et entament leur autogamie dans le milieu d'inactivation. Les cellules autogames sont ensuite isolées dans un milieu classique pour mesurer leur capacité à produire une progéniture sexuelle viable. Comme attendu, il n'y a pas production de progéniture sexuelle viable après l'inactivation de *PGM*. En revanche, on observe la production d'une progéniture sexuelle viable à l'issue de l'inactivation de *POB3-1*, comme observée en RNAi contrôle. On en conclut que *Pob3-1* n'est donc pas nécessaire à l'obtention de progéniture sexuelle viable à l'issue de l'autogamie. Afin de restreindre l'inactivation de *POB3-1* au développement MAC, nous avons inactivé le gène avec deux fragments non chevauchants (#1 et #2) pendant la conjugaison (Figure 14E). Des cellules sexuellement réactives de signes sexuels complémentaires sont mélangées et les couples sont isolés et transférés en milieu d'inactivation. Lorsque les couples se séparent, les exconjugants sont transférés individuellement dans un milieu classique sans RNAi et on suit la capacité de la progéniture sexuelle à reprendre la croissance végétative. Comme observé à l'autogamie, lorsque *POB3-1* est inactivé, on obtient une progéniture sexuelle viable comme en RNAi contrôle. Nous avons conclu que *POB3-1* n'était pas nécessaire au développement macronucléaire.

Il est possible que même si la protéine *Pob3-1* n'est pas essentielle à l'obtention d'une progéniture sexuelle viable, elle puisse être impliquée dans les réarrangements d'ADN. Pour tester cette hypothèse, nous avons inactivé *POB3-1* avec deux fragments non chevauchants

(#1 et #2) pendant l'autogamie et analysons par PCR l'excision de deux IES. Les réactions PCR sont réalisées sur l'ADN extrait de cellules à la fin du développement MAC en utilisant des amorces de part et d'autres des extrémités de chaque IES. En RNAi contrôle et en absence de RNAi, comme attendu, seule la forme correspondant à l'IES excisée est détectée dans la progéniture sexuelle. En RNAi *PGM* ou *SPT16-1*, comme attendu on détecte la forme excisée de l'IES provenant des fragments du MAC maternel et la forme non excisée de l'IES indiquant que Pgm et Spt16-1 sont bien nécessaires aux réarrangements d'ADN. Par contre lorsque l'on inactive *POB3-1* avec les deux fragments non chevauchants (#1 et #2), on observe que la forme non excisée des deux IES. On en conclue que *POB3-1* n'est pas nécessaire à l'excision de ces deux IES. Etant donné les résultats que nous avons observés, il semble que *POB3-1* ne soit pas acteur majeur des réarrangements d'ADN comme *SPT16-1*. L'ensemble de ces résultats étant négatifs il est difficile de donner une conclusion définitive tant que nous n'avons pas l'assurance que *POB3-1* a bien été inactivé en contrôlant par exemple le niveau d'ARNm *POB3-1*.

3. Pob3 est localisé dans les macronoyaux en développement

Afin d'examiner la localisation subcellulaire de la protéine Pob3-1 chez *P. tetraurelia*, nous avons inséré la GFP en amont du codon STOP du gène *POB3-1*. L'expression du gène est contrôlée par les séquences régulatrices endogènes amont et aval de *POB3-1*. Le transgène Pob3-1-GFP est micro-injecté dans le macronoyau végétatif (Figure 15A). Par fixation au cours du cycle de vie de cellules injectées Pob3-1-GFP, nous ne parvenons pas à détecter la protéine de fusion pendant la croissance végétative, ni pendant la méiose. La protéine de fusion s'accumule précocement dans les macronoyaux en développement et persiste de manière diffuse dans le nouveau MAC après la première division cellulaire. Elle est aussi détectée dans les MIC lorsqu'apparaissent les MAC en développement. Afin de déterminer si la protéine Pob3-1-GFP nucléaire est localisée dans le nucléoplasme et soluble ou bien associée à la chromatine, nous avons comparé deux conditions de fixation des cellules transformées avec Pob3-1-GFP au cours du développement. Nous avons perméabilisé avant de fixer les cellules transformées avec Pob3-1-GFP au cours du développement ou bien nous avons fixé (sans perméabilisation préalable) les mêmes populations de cellules (Figure 15B). On suppose que si la protéine Pob3-1 n'est pas associée fortement à la chromatine, elle

devrait être extraite des noyaux, et moins de fluorescence GFP serait détectée en présence des détergents utilisés pour perméabiliser (Triton X-100 à 2%). Nous avons mesuré l'aire des MAC en développement ainsi que la quantité de signal GFP dans le MAC au cours du développement à partir d'images acquises en microscopie confocale, pour 60 noyaux dans chaque condition. Cette analyse montre que l'aire des noyaux est plus petite lorsqu'on perméabilise avant de fixer les cellules (Figure 15C). En normalisant le signal de fluorescence par l'aire des noyaux (Figure 15C), on observe que le signal GFP dans les MAC en développement n'est pas modifié par la perméabilisation. Le signal de fluorescence est diminué par la perméabilisation des cellules mais cette diminution est la conséquence de la diminution de l'aire des noyaux. La concentration de Pob3-1-GFP dans les MAC est donc constante dans la fenêtre développementale analysée dans les deux types de conditions expérimentales. En revanche, le signal GFP dans les micronoyaux disparaît lorsque les cellules sont perméabilisées. Ces résultats indiquent que Pob3-1-GFP est plus facilement extraite des MIC que des MAC en développement. Ces observations suggèrent que Pob3-1-GFP est fortement associée à la chromatine dans les MAC en développement.

En conclusion, même si *POB3-1* a un profil d'expression très similaire à celui de *SPT16-1*, nos expériences n'ont pas permis de mettre en évidence un rôle de Pob3 dans le développement macronucléaire. Aucun domaine de dimérisation nécessaire pour l'interaction avec son partenaire Spt16-1 au sein du complexe FACT n'a pu être identifié dans la protéine Pob3-1. Nous n'avons donc pas d'évidence que Pob3-1 soit un partenaire de Sp16-1.

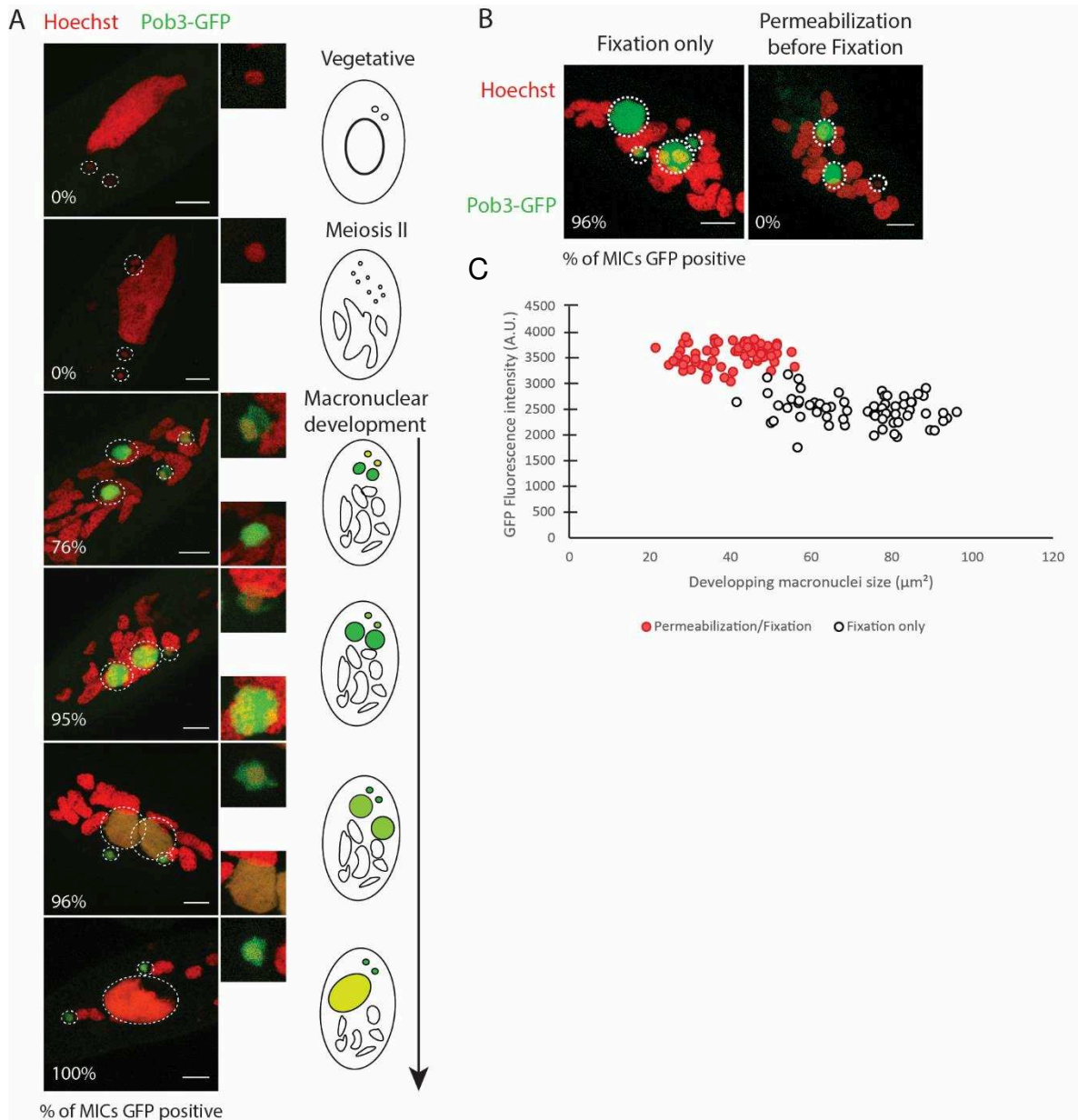


Figure 15 : Localisation de la protéine Pob3-1 à l'aide d'une fusion protéique à la GFP

A. Localisation dans la cellule de la protéine de fusion Pob3-1-GFP. Des superpositions des projections en Z de vues de Pob3-GFP (vert) et Hoechst (rouge) sur une sélection de projections à différents moments du cycle de vie sont présentées (a-f). Les cercles blancs en pointillés indiquent les deux nouveaux MACs se développant (M) et les deux MICs quand ils sont visibles. Les autres noyaux marqués au Hoechst sont des fragments du macronoyau maternel somatique. Sont indiqués en pourcentage le nombre de MIC marqués à la GFP. A droite, sont représentés les étapes correspondantes du cycle cellulaire avec l'intensité variable du vert symbolisant la présence de Pob3-1-GFP. L'échelle est de 10 μm . **B.** Effet de la perméabilisation sur la localisation de la protéine de fusion Pob3-1-GFP. Des superpositions des projections en Z de vues de Pob3-1-GFP (vert) et Hoechst (rouge) sur une sélection de projections sont présentées. Les cercles blancs en pointillés indiquent les deux nouveaux MACs se développant (M) et les deux MICs quand ils sont visibles. Les autres noyaux marqués au Hoechst sont des fragments du macronoyau maternel somatique. Le pourcentage du nombre de MIC marqués à la GFP est indiqué. L'échelle est de 10 μm . Quantification de l'intensité du signal GFP dans les nouveaux MACs en développement en fonction de leur aire. En rouge les cellules perméabilisées avant d'être fixées et en blanc les cellules fixées (sans perméabilisation). En dessous, la distribution d'aire des MAC en développement en μm^2 dans lesquels a été quantifié le signal fluorescent pour Pob3-1-GFP est montrée en condition de fixation seule (blanc) ou de perméabilisation avant fixation (rouge). En dessous, la distribution de l'intensité du signal GFP dans les nouveaux MACs en développement normalisée par l'aire des MACs en μm^2 est montrée en condition de fixation seule (blanc) ou de perméabilisation avant fixation (rouge).

Partie II : Analyse de la méthylation des adénines chez *Paramecium tetraurelia*

La présence d'adénines méthylées dans le génome de *Paramecium tetraurelia* est observée pour la première fois en 1974 à partir des expériences de (Cummings et al., 1974). L'ADN extrait des macronoyaux, micronoyaux et mitochondries est analysé par chromatographie après ingestion par les paramécies de méthionine, adénine et thymine radioactives présentes dans le milieu de culture. L'étude concluait à la présence d'environ 2% d'adénines méthylées (6mA) dans le génome macronucléaire et aussi dans le génome micronucléaire. Selon cette étude, 6mA était absent du génome mitochondrial. Les expériences de Cummings montraient aussi que la méthylation des cytosines était absente du génome. Aucun rôle n'a encore été attribué à la méthylation des adénines chez *Paramecium tetraurelia*.

I. Détection de bases modifiées dans le génome de *Paramecium tetraurelia*

1. Détection de nucléosides modifiés par HPLC-MS dans le génome

Partant de ces résultats, nous avons décidé de confirmer la présence de 6mA et l'absence de 5mC dans le génome de *P. tetraurelia*. Nous nous sommes en particulier attachés à obtenir de l'ADN génomique de cultures de paramécie nourries avec des bactéries qui sont dépourvues de méthylation de 6mA et 5mC. Nous voulons également examiner la présence éventuelle d'autres modifications de nucléotides dans le génome.

Afin de détecter l'ensemble de ces modifications, nous avons choisi d'analyser l'ADN génomique de *P. tetraurelia* par HPLC-MS. Cette approche permet de visualiser théoriquement l'ensemble des modifications portées par les nucléotides à condition que nous possédions des molécules standards pour chacune des modifications à tester (Figure 16). De l'ADNg extrait d'*E. coli* Dam-Dcm⁻ est utilisé comme contrôle négatif, cette souche bactérienne étant incapable de méthyler 5mC et 6mA. Comme contrôle positif, on utilise de l'ADN extrait de la souche SURE d'*E. coli* abondamment méthylée pour 5mC et 6mA (2%)

(Fang et al., 2012). On a choisi de comparer la présence de modifications de nucléotides dans le génome de chez *Paramecium tetraurelia* dans différentes conditions physiologiques. Dans ce but, des échantillons d'ADN génomique ont été préparés à partir de cultures de *P. tetraurelia* en croissance exponentielle (Veg), à la fin de l'autogamie lorsque les nouveaux MAC sont formés (T50) et en condition de restriction alimentaire avant l'entrée en autogamie (starved). Nous avons aussi analysé de l'ADNg issu de l'espèce apparentée *Paramecium sexaurelia* extrait à la fin de l'autogamie (T50). Les ADN sont tous traités à la RNase pour éviter des contaminations par les molécules d'ARN qui sont connues pour contenir de nombreuses modifications chez les bactéries comme chez les eucaryotes (Dominissini et al., 2012). Pour s'affranchir de la contamination bactérienne du milieu de culture des paramécies (*K. pneumoniae*), les cellules de paramécie sont nourries avec la souche Dam-Dcm- d'*E. coli*. Cette différence est cruciale par rapport aux expériences de Cummings. Puisqu'on extrait l'ADN total des paramécies, l'ADN du MIC et MAC sont présents dans les cultures végétatives et en plus, pour le point à l'autogamie, sont aussi présents les fragments du macronoyau maternel. Etant donné le rapport de ploïdie MIC (2n)/MAC (800n), ce sont les modifications associées au génome MAC qui seront préférentiellement détectées. Après traitement à la RNase, les échantillons d'ADN sont précipités et digérés par la Degradase plus pour obtenir des nucléosides. Les standards utilisés sont les nucléosides non modifiés (A, T, G, C) mais également les nucléosides modifiés suivants : 6mA, 5mC, 5hmC, 4mC et 8oxoG. Les standards sont utilisés pour déterminer le temps de rétention propre à chaque molécule et sont systématiquement injectés en début et fin d'analyse des échantillons pour contrôler que les temps de rétention n'ont pas changé au cours de l'expérience. Les différents nucléosides sont d'abord séparés dans le temps selon leur affinité avec la phase C18 apolaire de la colonne d'HPLC et élués grâce à un solvant polaire, l'acétonitrile. Ils sont ensuite détectés par spectrométrie de masse assignant un temps de rétention à la masse de chaque molécule. L'abondance de chaque nucléoside est calculée à partir de l'aire sous la courbe des signaux enregistrés et normalisée par le signal du nucléoside non modifié correspondant.

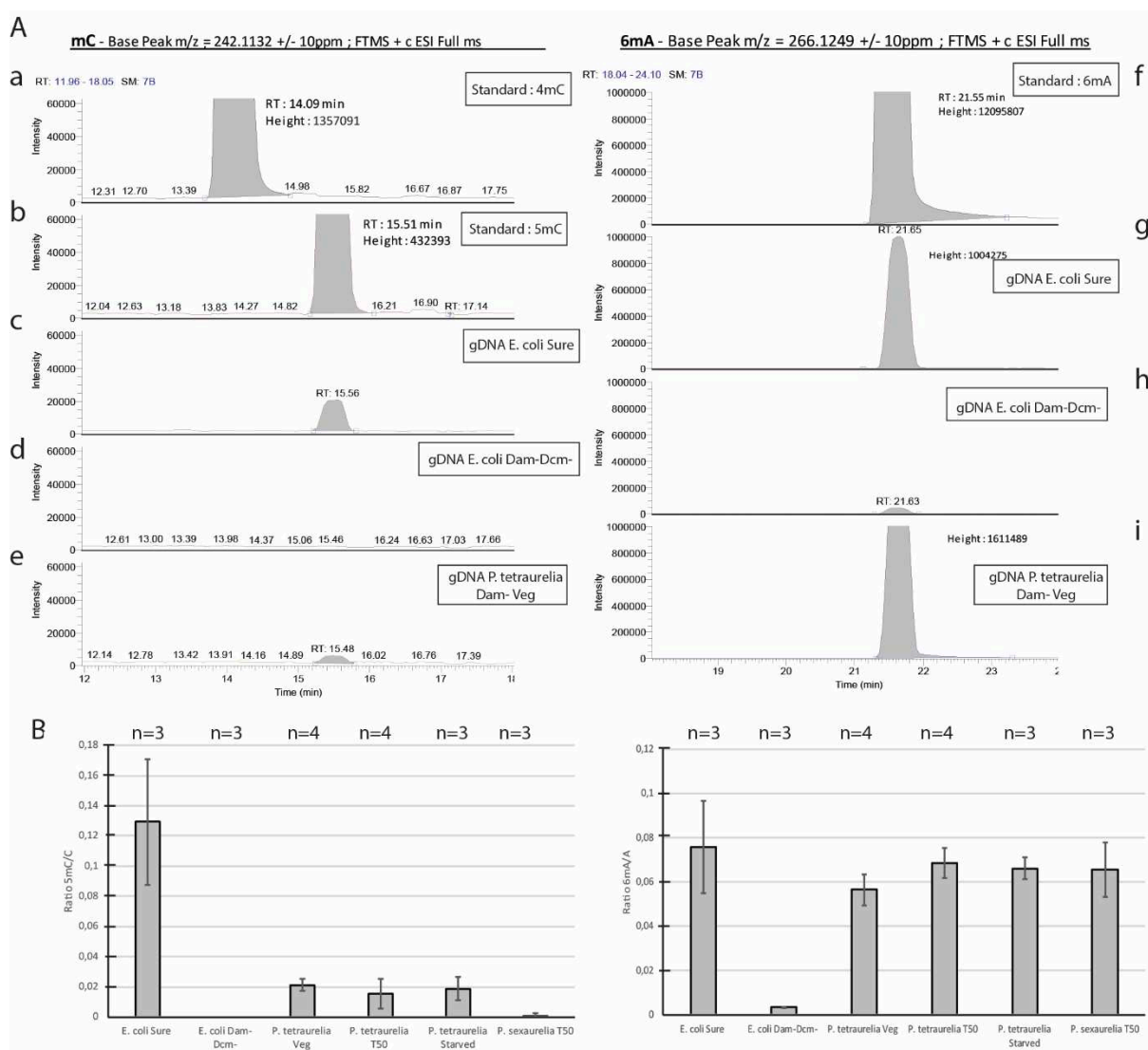


Figure 16 : Quantification par HPLC-MS des niveaux de méthylation des cytosines et des adénines de l'ADN génomique total de *P. tetraurelia* et *P. sexaurelia*

A. Rétenion des nucléosides 5mC, 4mC et 6mA des standards et échantillons d'ADN étudiés sur la colonne de séparation de l'HPLC. Le chromatogramme indique l'intensité détectée pour chaque nucléoside en fonction de son temps de rétenion sur la colonne. Sont représentés les chromatogrammes des standards pour 5mC (b) issu de la digestion par la Degradaase Plus d'un oligonucléotide 5mC, pour 4mC (a) issu de la digestion par la Degradaase plus et pour 6mA (f). Sont également représentés les chromatogrammes pour 5mC et 4mC des échantillons d'ADN génomique d'*E. coli* Dam-Dcm-, *E. coli* sure et *P. tetraurelia* nourries avec *E. coli* Dam-Dcm- (c-e). Sont représentés les chromatogrammes pour 6mA des échantillons d'ADN génomique d'*E. coli* Dam-Dcm-, *E. coli* sure et *P. tetraurelia* nourries avec *E. coli* Dam-Dcm- (g-i). **B.** Quantification des aires sous la courbe des chromatogrammes pour chaque échantillon d'ADN génomique analysé. L'aire correspondant à la densité de chaque nucléoside modifié est normalisée par le nucléoside standard correspondant. Les moyennes et déviations standards sont calculées à partir de 3 ou 4 expériences indépendantes pour chaque type d'échantillon, comme indiqué en dessous de l'axe des abscisses.

2. La méthylation des cytosines est absente du génome de *P. tetraurelia*

Bien que les molécules de 4mC et 5mC aient la même masse, le passage sur la colonne de l'HPLC a permis de les séparer parce que 4mC et 5mC ont des temps de rétenion

différents. Le nucléoside 4mC, visible dans le chromatogramme correspondant au standard (Figure 16Aa), est absent des chromatogrammes de l'ensemble de nos échantillons contrôle et de l'ADN de paramécies (Figure 16Ab-e). De la même manière, on détecte 5mC dans le chromatogramme du standard et du contrôle positif, la souche SURE d'*E. coli* (Figure 16Ab-c). Comme attendu, 5mC est absente de la souche *E. coli* Dam-Dcm- (Figure 16Ad). On ne détecte que des quantités très faibles voire quasi nulles de 5mC chez *P. tetraurelia* et *sexaurelia* comparables à ce qui est détecté pour *E. coli* Dam- Dcm- (Figure 16B). Les données ne sont pas présentées pour 5hmC et 8oxoG mais nos expériences de spectrométrie de masse n'ont pas détecté ces molécules dans l'ADN de *P. tetraurelia* et *sexaurelia*. En conclusion, les nucléosides modifiés 5mC, 4mC, 5hmC et 8oxoG ne sont pas présents dans le génome MAC de la paramécie ou dans des quantités trop faibles pour être détectées par cette méthode.

3. La méthylation des adénines est présente dans l'ADN de *P. tetraurelia*

Le nucléoside modifié 6mA est bien séparé du nucléoside non modifié par la colonne d'HPLC et bien détecté en spectrométrie de masse comme le montrent les chromatogrammes obtenus (Figure 16Af). Comme attendu, 6mA est observé dans le contrôle positif *E. coli* SURE et absent du contrôle négatif *E. coli* Dam-Dcm- (Figure 16Ag-h). Le nucléoside modifié 6mA est aussi fortement détecté dans l'ensemble des échantillons d'ADN de *P. tetraurelia* et de *P. sexaurelia* (Figure 16B). La présence de cette modification ne semble pas varier entre les différentes conditions physiologiques testées bien que nos résultats ne permettent pas de quantifier avec précision l'abondance de chaque modification. En conclusion, la méthylation des adénines est présente dans l'ADN génomique de *P. tetraurelia* et *P. sexaurelia*, quelles que soient les conditions physiologiques observées.

L'ensemble de ces résultats nous permet de confirmer les résultats obtenus par (Cummings et al., 1974) et de conclure que l'ADN de *P. tetraurelia* est méthylé sur les adénines. La méthylation des cytosines n'est pas détectée chez *P. tetraurelia* ni aucune autre des modifications étudiées (4mC, 5hmC, 8oxoG). Enfin, la méthylation des adénines est conservée chez *P. sexaurelia*, espèce qui n'avait pas été analysée dans (Cummings et al., 1974).

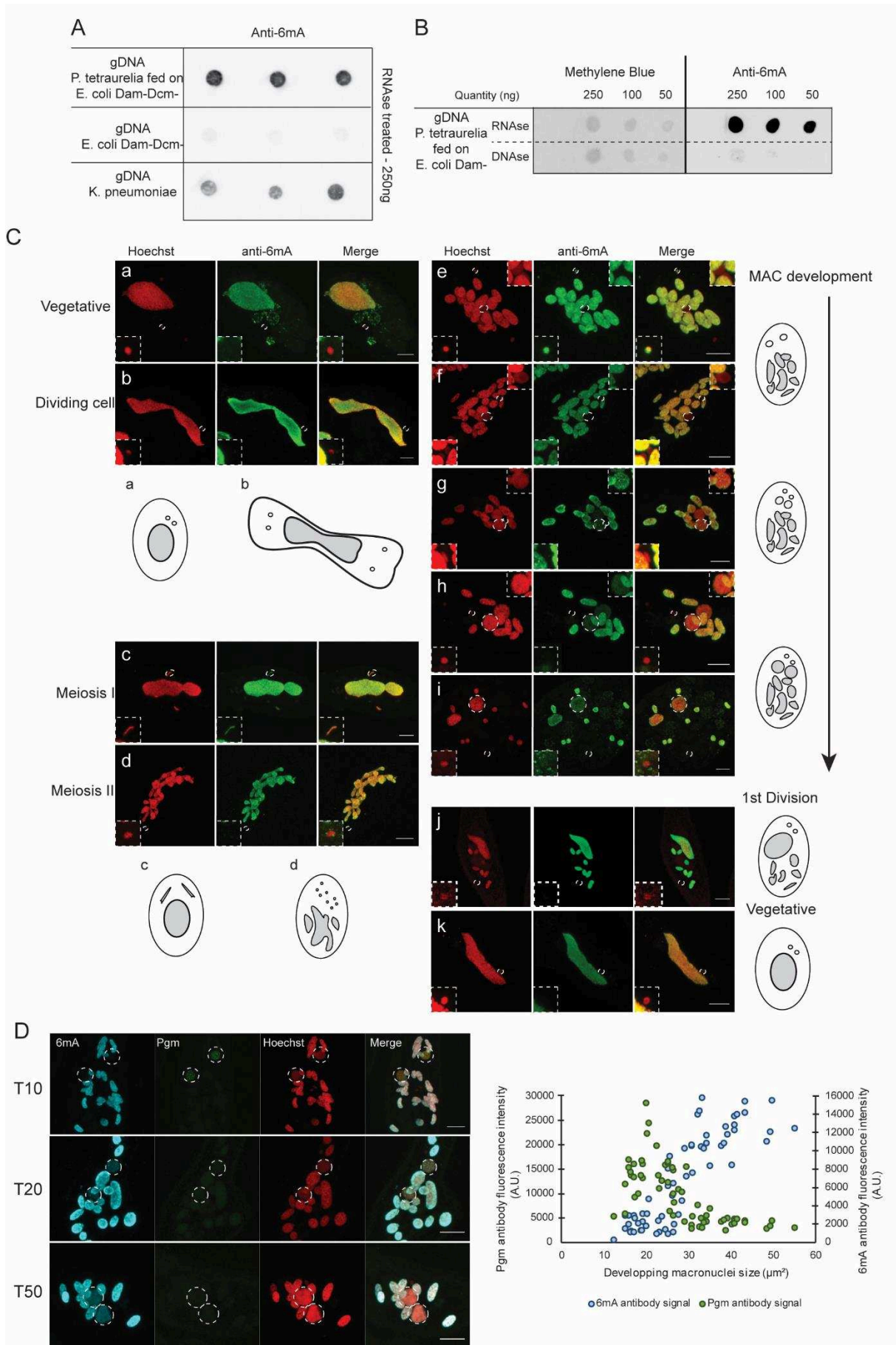


Figure 17 : Détection de 6mA dans le génome de *Paramecium tetraurelia* avec des anticorps anti-6mA.

A. Détection d'adénines méthylées dans l'ADN de *Paramecium tetraurelia* par Dot Blot à l'aide d'un anticorps anti-6mA. Des dépôts de 250ng d'ADN génomique traités à la RNase A et T1 de *Paramecium tetraurelia* nourries avec *E. coli* Dam-Dcm-, de *E. coli* Dam-Dcm- et de *K. pneumoniae* sont déposés sur une membrane en nylon chargée positivement et crosslinkés par UV. La méthylation est détectée après incubation avec un anticorps anti-6mA (Synaptic System 202 003) et par un anticorps secondaire couplé à la peroxydase révélée par ECL. **B.** Détection d'adénines méthylées dans l'ADN de *Paramecium tetraurelia* par Dot Blot avec un anticorps anti-6mA. Des dépôts d'ADN génomique de quantité croissante entre 50 et 250ng traités RNase puis DNase de *Paramecium tetraurelia* nourries *E. coli* Dam-Dcm- et d'*E. coli* Dam-Dcm- sont déposés sur une membrane en nylon chargée positivement et crosslinkés par UV. Les dépôts d'ADN sont révélés dans un bain de bleu de méthylène. La méthylation est détectée après incubation avec un anticorps anti-6mA et par un anticorps secondaire couplé à la peroxydase révélée par ECL. **C.** Description de la localisation cellulaire par immunofluorescence de 6mA chez *Paramecium tetraurelia* par immunofluorescence. Des superpositions d'une sélection de projections en Z de vues de 6mA (vert) et Hoecsht (rouge) à différents moments du cycle de vie sont présentées (a-l). Les cercles blancs en pointillés indiquent les deux MIC quand ils sont visibles. Le cadre en pointillés blancs en bas à gauche de chaque image montre un grossissement d'un MIC. Le cadre en pointillés blancs en haut à droite de chaque image montre un grossissement d'un nouveau MAC en développement. L'échelle est à 10µm. **D.** Localisation de la méthylation par immunofluorescence des adénines et de Pgm dans des cellules en développement à l'aide d'anticorps spécifiques. Des superpositions des projections en Z de vues de 6mA (cyan), Pgm (vert) et Hoecsht (rouge) à différents moments du développement micronucléaire sont présentées. Les cercles blancs en pointillés indiquent les deux nouveaux MAC en développement. A gauche sont indiqués les temps du développement auxquels ont été observées les cellules. A droite, quantification des signaux de fluorescence des anticorps anti 6mA (bleu) et anti Pgm (vert) en fonction de l'aire des MAC en développement en µm².

II. Description de la localisation de la méthylation des adénines chez *Paramecium tetraurelia*

Il n'est pour le moment pas possible de déterminer la présence de modifications dans l'ADN micronucléaire par HPLC-MS, en raison des trop faibles quantités obtenues après purification de l'ADN micronucléaire (Guérin et al., 2017). Pour étudier la dynamique de la méthylation au cours du cycle de vie et sa localisation précise dans la cellule, nous avons utilisé un anticorps spécifique de 6mA pour des expériences de dot blot et d'immunofluorescence.

1. La méthylation des adénines est majoritairement détectée sur l'ADN

Nous avons entrepris de confirmer les résultats obtenus en HPLC-MS par dot blot avec un anticorps spécifique pour 6mA (Synaptic system 202 003) (Figure 17A). De l'ADN génomique extraits de *P. tetraurelia*, *E. coli* Dam-Dcm- et *K. pneumoniae* digérés à la RNase sont déposés sur une membrane de nylon chargée positivement et réticulés par les UV. On dépose des quantités constantes (250 ng) d'ADN en triplicat. Comme attendu, on ne détecte pas voire très peu de signal 6mA dans l'ADN génomique extraits d'*E. coli* Dam-Dcm-. Au contraire, 6mA est bien détectée chez *Klebsiella pneumoniae* qui est utilisé pour nourrir

habituellement les paramécies. On détecte également 6mA dans l'ADN de *P. tetraurelia* nourries avec *E. coli* Dam-Dcm- confirmant les résultats obtenus par HPLC-MS.

Pour déterminer si la méthylation que nous détectons provient bien de l'ADN et non pas de l'ARN, on réalise ensuite des dépôts de quantité croissante d'ADN sur une membrane en nylon chargée positivement et réticulée par les UV (Figure 17B), après traitement RNase A et T1 ou traitement DNase I, pour s'assurer de la spécificité du signal anti-6mA. Après dépôts des acides nucléiques sur la membrane, leur présence est révélée grâce au bleu de méthylène. Comme précédemment (Figure 17B) et pour des quantités inférieures (dès 50 ng), on détecte 6mA dans l'ADN de *P. tetraurelia* traité RNase et ce signal disparaît lorsque l'ADN est ensuite traité DNase. Au bleu de méthylène on détecte cependant toujours un peu de signal qui provient probablement d'une digestion incomplète de l'ADN déposé. En conclusion, l'anticorps anti-6mA permet de détecter la méthylation des adénines dans l'ADN de *P. tetraurelia*.

2. La méthylation des adénines est établie pendant le développement des macronoyaux et perdure pendant tout le cycle cellulaire

Afin de localiser la méthylation des adénines entre les différents noyaux et déterminer quand elle est établie pendant le cycle de vie de *P. tetraurelia*, des expériences d'immunofluorescence ont été réalisées avec l'anticorps spécifique pour 6mA pendant le cycle cellulaire (Figure 17C). Pour s'affranchir des contaminations bactériennes et de la présence de 6mA sur les ARN (Dominissini et al., 2012; Fang et al., 2012), les cellules sont nourries avec *E. coli* Dam-Dcm- et traitées par la RNase A après avoir été perméabilisées et fixées. Le signal 6mA est détecté essentiellement dans le MAC tout au long du cycle de vie. Pendant le cycle végétatif, 6mA est détecté dans le MAC et est absent des MIC y compris lorsque les cellules se divisent. Lorsque s'engage l'autogamie, 6mA est présent dans le MAC maternel et apparaît dans les MIC de manière transitoire, d'abord en méiose I puis à l'apparition de jeunes MAC en développement. Le MAC maternel fragmenté reste méthylé au cours du développement du nouveau MAC. Le signal 6mA s'accumule progressivement dans les nouveaux MAC en développement. Il augmente à mesure que les MAC grossissent et le signal semble atteindre son maximum d'intensité dans le nouveau MAC après la première division cellulaire. L'ensemble de ces résultats nous permettent de conclure que la

méthylation des adénines est principalement localisée dans le MAC au cours du cycle de vie, et est présente de manière transitoire dans les MIC pendant et après les événements sexuels. Nous avons observé qu'il existe une méthylation *de novo* qui s'établit dans les MAC en développement et nous supposons qu'il existe une méthylation de maintenance conservée tout au long du cycle cellulaire.

3. La méthylation des adénines s'établit en aval des réarrangements d'ADN dans les nouveaux macronoyaux

Pour déterminer plus précisément le moment d'apparition de 6mA dans les MAC en développement et le situer par rapport aux événements d'élimination d'ADN, nous avons réalisé des expériences d'immunofluorescence pour détecter simultanément l'endonucléase Pgm et 6mA avec des anticorps spécifiques (Figure 17D). Nous avons quantifié les signaux de fluorescence Pgm et 6mA dans les MAC en développement en fonction de l'aire de ces noyaux. La protéine Pgm s'accumule dans les MAC en développement puis disparaît à des stades plus tardifs du MAC en développement (Dubois et al., 2017). Nous observons que le signal Pgm est présent dans les MAC en développement alors que 6mA est encore indétectable (Figure 17D). La fluorescence 6mA augmente progressivement dans les MAC, alors que le signal Pgm a déjà disparu. Le signal 6mA augmente en même temps que l'aire des noyaux augmente. Le pic d'intensité de fluorescence de Pgm est détecté lorsque le signal 6mA est encore faible. Le pic d'intensité de 6mA est atteint lorsque le signal Pgm a déjà disparu. Même si nos résultats sont sujets à caution dans la mesure où ils dépendent du seuil de détection des anticorps utilisés, nos observations suggèrent que la méthylation des adénines s'établit dans les MAC en développement postérieurement à la présence de Pgm dans les MAC.

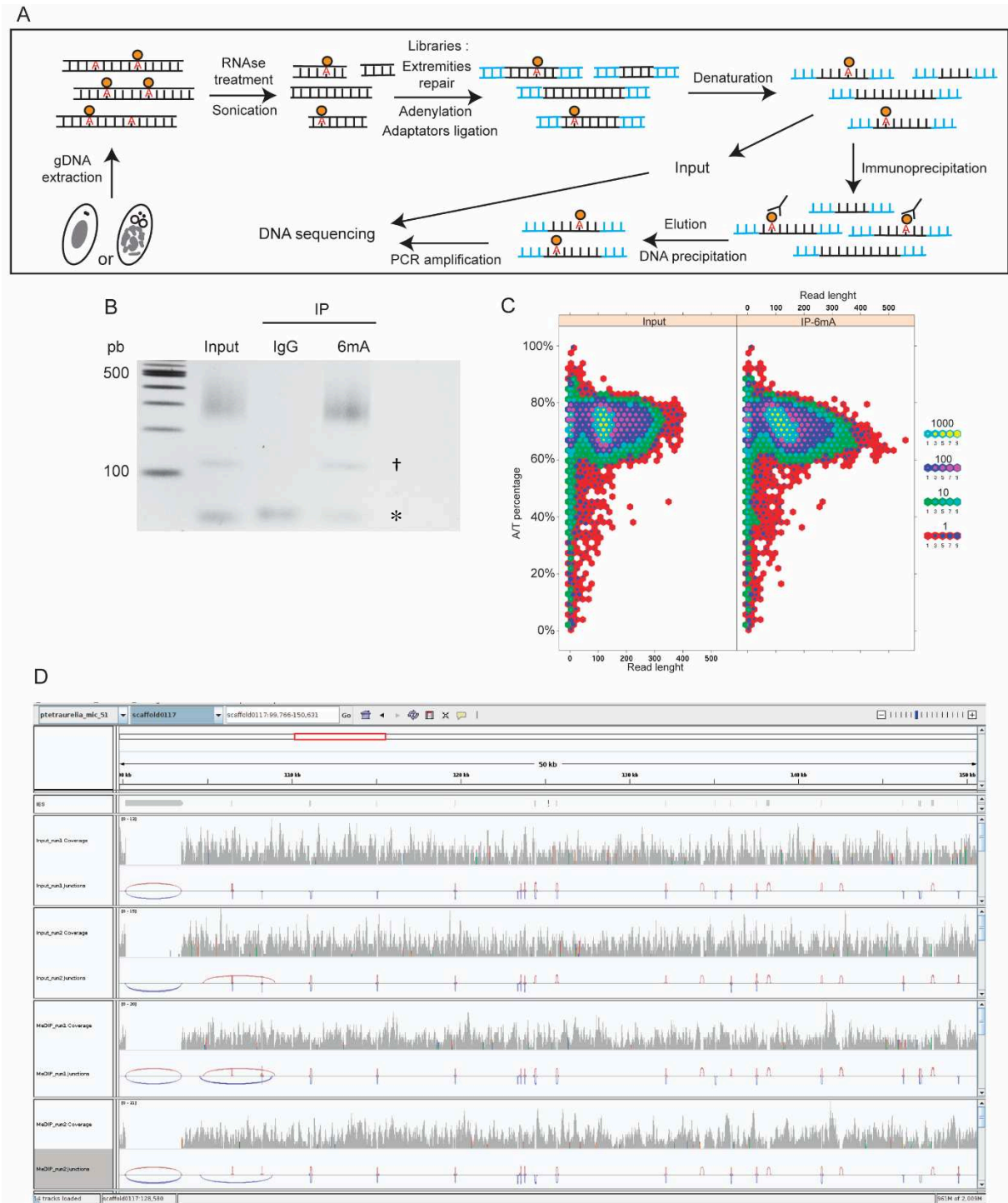


Figure 18 : Localisation de 6mA dans le génome de *P. tetraurelia* par MedIP

A. Représentation schématique des étapes du protocole de MedIP. De l'ADN génomique total est extrait de cellules en croissance exponentielle ou en fin d'autogamie (T50), traité par la RNase A et la RNase T1 et soniqué pour obtenir des fragments de 200 pb. Des bibliothèques sont préparées à partir des fragments soniqués. Les fragments sont ensuite dénaturés (Input) et utilisés pour l'immunoprécipitation avec les anticorps anti IgG et anti-6mA. Après lavage et élution, les fragments d'ADN immunoprécipités sont précipités et amplifiés par PCR avant séquençage Illumina. B. Dépôt sur gel d'agarose 1,2% des amplifications par PCR de l'Input et des immunoprécipitations par l'anticorps anti-IgG ou anti-6mA. Rien n'est détecté dans la piste correspondant à l'IP IgG hormis les dimères d'amorces utilisées pour la PCR. Dans les pistes Input et 6mA, la traînée correspond à l'ADN immunoprécipité, la bande inférieure (indiquée par †) aux adaptateurs non ligués et amplifiés et la bande indiquée par * aux dimères d'amorces de PCR. C. Comparaison des couvertures de séquençage entre l'Input et l'IP 6mA. On observe la taille des lectures en fonction de la proportion en AT dans chaque lecture. À droite sont données les proportions de chaque type de lecture dans l'échantillon séquençé, de bleu à jaune = milliers de lectures, de violet à rose =

centaines de lectures, de vert à bleu = dizaines de lectures, de rouge à bleu = lectures uniques. La majorité des lectures sont d'une taille allant de 100 à 200 pb compatible avec les fragments que nous avons produits par sonication et les bibliothèques que nous avons construites, avec une proportion en AT entre 60 et 80% similaire à celle trouvée dans l'ensemble du génome (Aury et al., 2006). Les profils de couverture ne diffèrent pas entre l'Input et l'IP 6mA. D. Comparaison entre la couverture dans le génome browser d'une région particulière du scaffold 117 entre les coordonnées 99,766 et 150,631 entre l'Input et l'IP 6mA alignés sur le génome *ptetraurelia_mic_51*. Sont indiqués les IES et autres séquences MIC spécifiques en haut en gris. On ne détecte pas de pic d'enrichissement pour une séquence particulière dans l'IP 6mA.

III. Localisation de 6mA dans le génome de *P. tetraurelia*

Ayant établi que la méthylation des adénines est présente de manière significative et facilement détectable dans l'ADNg de *P. tetraurelia*, nous avons choisi d'utiliser la technique de MedIP reposant sur le séquençage à haut débit de l'ADNg immunoprécipité par l'anticorps spécifique de 6mA pour localiser 6mA le long du génome.

1. Adaptation d'un protocole d'immunoprécipitation de l'ADNg méthylé sur 6mA chez *P. tetraurelia*

Le MedIP (Methylation DNA ImmunoPrecipitation) repose sur le séquençage d'ADN enrichi en séquences méthylées grâce à un anticorps spécifique de la méthylation étudiée. La technique est traditionnellement utilisée pour la méthylation des cytosines 5mC (Plongthongkum et al., 2014) mais elle a aussi été adaptée pour l'étude de la méthylation des adénines dans l'ARN et plus récemment pour la méthylation des adénines dans l'ADN chez quelques eucaryotes (Greer et al., 2015; Zhang et al., 2015).

Je me suis basée sur les protocoles développés pour le MeRIP, immunoprécipitation d'ARN méthylés 6mA (Dominissini et al., 2013, 2015) et sur ceux ayant servi à l'étude de la localisation de 6mA sur les génomes de *C. elegans* et *C. reinhardtii* (Fu et al., 2015; Greer et al., 2015). Les principales étapes du protocole sont rappelées dans la figure 18A.

- Echantillons étudiés et extractions d'ADN : Nous sommes partis, comme pour l'ensemble des études que j'ai présentées précédemment, de cellules de *P. tetraurelia* cultivées dans un milieu bactérisé avec la souche Dam-Dcm- d'*E. coli* incapable de méthyler 6mA et 5mC. Le MedIP permet d'identifier par le séquençage la contamination éventuelle en 6mA de d'autres organismes et notamment bactériens chez qui cette modification est très abondante. Pour garantir un enrichissement suffisant en séquences paramécie qui sont potentiellement autant méthylées en 6mA que chez les bactéries et une couverture suffisante par le séquençage, nous avons préféré réduire la contamination bactérienne. Une

première tentative de MedIP à partir d'ADN extraits de cellules en croissance végétative exponentielle a d'abord été tentée. Il s'est avéré par Blast des lectures de l'Input que l'ADN immunoprécipité était beaucoup trop contaminé par des souches bactériennes (75% des lectures de l'Input) pour offrir une couverture suffisante du génome de *P. tetraurelia* (5% des lectures de l'Input). Il était alors impossible d'utiliser cette faible couverture pour détecter des pics d'enrichissement dans l'IP par rapport à l'Input. On peut expliquer cette contamination par le stade du cycle de vie auquel l'ADN est extrait. En croissance végétative, les cellules sont pleines de vacuoles digestives riches en bactéries. Il se peut aussi que le protocole de lavages des cellules n'ait pas été suffisamment intensif pour se débarrasser des bactéries présentes dans le milieu de culture lui-même. Nous avons ensuite décidé d'extraire de l'ADN de cellules en fin d'autogamie, 50h après le début de l'autogamie. A ce stade, il n'y a pas de bactéries dans le milieu de culture et les cellules sont en carence alimentaire depuis plusieurs jours. Dans ces cellules, les réarrangements se sont déjà terminés ou se terminent dans les nouveaux MAC en développement. Sur la base de nos expériences d'immunofluorescence, la méthylation 6mA est détectée dans les nouveaux MAC en plus des fragments du MAC maternel encore présents dans la cellule. On suppose que l'on aura ainsi accès à la fois à la méthylation de maintenance (dans le MAC maternel) que l'on aurait observée dans des cellules en croissance exponentielle et à la méthylation déposée de novo dans les MAC en développement. L'ADN extrait est traité à la RNase A (Sigma 50U/mg) à 20µg/ml et 1U de RNase T1 (Ambion 1U/µl).

- Sonication : L'ADNg extrait est soniqué en fragments avoisinant les 200 pb. La sonication est réalisée sur un appareil Covaris M220 selon les recommandations et programmes du fabricant pour obtenir cette taille de fragment. Pour vérifier la taille des fragments obtenus, l'ADN soniqué est analysé sur un gel d'agarose 1,2%.
- Préparation des librairies : A la différence du protocole de MeRIP dans (Dominissini et al., 2013, 2015), et comme dans les protocoles de MedIP sur *C. elegans* ou *D. melanogaster* (Fu et al., 2015; Greer et al., 2015), nous réalisons les librairies pour le séquençage avant d'immunoprécipiter les fragments d'ADN. Dans le MeRIP, la préparation des librairies implique la conversion des ARN en ADN complémentaires et donc la perte de la méthylation par la RT-PCR. Il faut donc dans ce cas immunoprécipiter les ARN d'abord. En préparant les librairies avant l'immunoprécipitation, on s'affranchit au maximum de différences entre l'Input et l'IP qui ne proviendraient pas de l'immunoprécipitation. Les librairies sont préparées avec le kit NEBNext® DNA Library Prep Master Mix Set for Illumina®. Les extrémités des fragments d'ADN soniqués sont réparées, adénylées et liguées à des adaptateurs.

- Dénaturation : Les échantillons d'ADN fragmentés et liés aux adaptateurs sont dénaturés pour permettre l'immunoprécipitation et une petite quantité (Input) est mise de côté.
- Immunoprécipitation/Elution/Précipitation : Après dénaturation, les échantillons sont incubés avec des billes couplées à la protéine A liées aux anticorps spécifiques de la méthylation des adénines 6mA (Synaptic system). Seront donc immunoprécipités par les billes uniquement les portions du génome de *P. tetraurelia* de 200pb en moyenne portant 6mA. Les fragments d'ADN sont élués avec du N6MAmonop (N⁶-Methyladenosine 5'-monophosphate sodium salt), nucléotide entrant en compétition avec les adénines méthylées retenues par les anticorps anti-6mA puis précipités à l'éthanol. Comme contrôle négatif, on réalise aussi une immunoprécipitation avec un anticorps anti-IgG.
- Amplification par PCR : Suivant le protocole du kit NEBNext[®] DNA Library Prep Master Mix Set for Illumina[®], l'Input et les ADN immunoprécipités par les anticorps anti-6mA ou anti-IgG sont amplifiés par PCR. Une partie de l'Input et des IP sont déposés sur un gel d'agarose pour contrôler le nombre de cycles nécessaires en PCR (Figure 6B).
- Séquençage Illumina : L'Input et les ADN Immunoprécipités sont séquencés par séquençage Illumina (single end, reads 75pb)

2. L'immunoprécipitation par l'anticorps 6mA n'a pas permis d'enrichir en régions méthylées

A l'issue des PCR visant à enrichir l'Input et les immunoprécipitations par les anticorps anti-6mA et IgG, on a déposé une partie des amplifications sur un gel d'agarose 1,2% (Figure 18B). Dans l'IP IgG, on y observe que rien n'est détecté hormis des dimères d'amorces, comme attendu. Au contraire, en plus des dimères d'amorces, dans l'IP 6mA on observe entre 200 et 400 pb une traînée d'acides nucléiques similaire à l'Input correspondant aux fragments d'ADN immunoprécipités. On observe aussi au-dessus de 100 pb une bande correspondant aux adaptateurs des librairies, non éliminés par les différents lavages après l'immunoprécipitation. L'anticorps anti-6mA, au contraire de l'anticorps anti-IgG, permet l'immunoprécipitation de séquences d'ADN. La PCR n'étant pas quantitative il est difficile d'évaluer s'il y a un enrichissement entre l'Input et l'IP 6mA.

Nous avons confié l'analyse des données de séquençage à Reiner Schulz (King's College London, UK) qui utilise le pipeline suivant : Suppression des adaptateurs, Construction des index (portions des génomes de références qui rassemblées couvrent l'ensemble du

génomique) par Bowtie2, Hisat2, BWA, Alignements des lectures par Bowtie2, Hisat2 et BWA, Configuration de IGV (Integrative Genomics Viewer), Identification des contaminants, Analyse de la couverture, Détection des pics d'enrichissement en 6mA. En alignant les séquences de l'Input et de l'IP 6mA, on constate que 1,5% des lectures de l'input s'alignent sur des séquences MIC-spécifiques. Ce n'est pas le cas dans l'IP. Il est possible que l'IP 6mA soit déplétée en certaines séquences MIC-spécifiques. Au stade du développement que nous analysons, les réarrangements d'ADN ont déjà eu lieu et le génome ne contient quasiment plus de séquences MICs. C'est à ce moment-là aussi que s'établit la méthylation dans un génome qui contient majoritairement des séquences MACs si l'on tient compte de ce que l'on observe en Immunofluorescence (Figure 17C, 17D). Le petit enrichissement en séquences MICs dans l'Input serait alors cohérent avec une présence de 6mA majoritairement sur les séquences MACs et cet enrichissement reste faible étant donné que ces séquences sont peu présentes. Curieusement, l'analyse de la couverture de l'Input et de l'IP 6mA montre des distributions très similaires (Figure 18C). Ces distributions ne semblent pas corrélées avec la proportion d'A/T dans le génome. Ces observations amènent à se demander si l'IP 6mA est vraiment enrichie en séquences particulières par rapport à l'Input. Les programmes DESeq2 et MACS2 ont été utilisés pour détecter des pics d'enrichissement dans l'IP 6mA. DESeq2 ne détecte ni enrichissement ou déplétion en un type de séquence particulière. MACS2 détecte un enrichissement dans l'IP 6mA en séquences mitochondriales et ribosomales ou répétées à l'extrémité des chromosomes MAC mais pas d'enrichissement ou de déplétion dans le reste du génome. Il est possible que l'enrichissement en ces séquences provienne d'une couverture excessive due à la grande quantité de séquences répétées. En conclusion, l'immunoprécipitation par l'anticorps anti-6mA n'a pas permis d'enrichir ou de déléter en séquences spécifiques à ce stade du cycle de vie. On émet alors plusieurs hypothèses. Il est possible que la distribution de 6mA sur le génome ne soit pas du tout en cluster mais parsemée le long du génome et que la résolution du MedIP ne permette pas de détecter de pic d'enrichissement. Une autre hypothèse est que simplement l'anticorps anti-6mA n'est pas suffisamment spécifique en MedIP de 6mA. C'est l'analyse prochaine du génome de *P. tetraurelia* au même stade de développement par séquençage PacBio (SMRT sequencing Pacific Bioscience) qui devrait nous permettre de répondre à la première hypothèse et nous offrir une localisation plus précise de 6mA sur le génome de *P. tetraurelia* (Hammond et al., 2017).

Discussion

Chapitre 1 : Le chaperon d'histone Spt16 chez *Paramecium tetraurelia* est un composant essentiel des réarrangements programmés du génome

Chez le cilié *P. tetraurelia* des évènements d'élimination massifs et reproductibles d'ADN se produisent pendant le développement du noyau somatique et conduisent à la perte de près de 30% du génome germinale. Les régions éliminées incluent de longs éléments répétés et transposables ainsi que 45000 séquences uniques, de petite taille et non-codantes, les IESs (Arnaiz et al., 2012). Mon travail de thèse avait pour objectif de tester l'hypothèse que le chaperon d'histone Spt16-1 pouvait être impliqué dans l'élimination de ces séquences chez *P. tetraurelia*.

I. Démontrer l'activité chaperon d'histone de Spt16-1

Le travail présenté précédemment a permis de montrer que *Paramecium tetraurelia* disposait de plusieurs homologues du chaperon d'histone Spt16 dont une de ces protéines, Spt16-1, dont le gène est exprimé uniquement pendant le développement macronucléaire. Les expériences d'inactivation du gène *SPT16-1* ont montré que la protéine Spt16-1 est nécessaire au développement MAC et à l'ensemble des réarrangements du génome. Spt16-1 agit en aval des voies de scnARN (Lepère et al., 2009) et de dépôt des marques d'hétérochromatine, H3K9me3 et H3K27me3 (Lhuillier-Akakpo et al., 2014). Elle agit en amont de l'endonucléase Pgm (Baudry et al., 2009) et est nécessaire à sa localisation correcte dans les MAC en développement pour assurer les réarrangements du génome. Bien que la protéine Spt16-1 diverge plus que ses paralogues Spt16-2a/2b/2c de ses homologues

eucaryotes, elle conserve les quatre domaines retrouvés classiquement chez Spt16 : N, D, M et C (Formosa, 2012; Jeronimo and Robert, 2016). Les domaines M et C sont requis chez *S. cerevisiae* pour l'interaction avec les histones et essentiels à la protéine et à sa fonction de chaperon d'histones (Hondele et al., 2013; Kemble et al., 2013, 2015). L'ensemble des résultats que nous avons présentés posait la question de savoir si Spt16-1 était bien un chaperon d'histone et si son rôle pendant les réarrangements d'ADN était dû à son interaction avec les histones.

1. A la recherche des partenaires de Spt16-1

Afin de démontrer que Spt16-1 est un chaperon d'histone nous avons tenté en première approche d'immunoprécipiter GFP-Spt16-1 pour identifier ses partenaires. L'immunoprécipitation se fait par un système de GFP-trap. La fraction immunoprécipitée peut être analysée en premier temps par Western Blot pour y détecter par exemple les histones. Elle peut ensuite être analysée en spectrométrie de masse pour y détecter les partenaires de Spt16-1. Parmi les partenaires que nous imaginions identifier, on espérait trouver Pob3 ou Ssrp1 et les histones avec lesquelles Spt16 interagit normalement. Nous avons réalisé des extraits nucléaires à différentes salinités d'extractions de 15 à 900mM NaCl dans des cellules transformées par GFP-Spt16-1. La protéine fusionnée à la GFP était systématiquement retrouvée quelle que soit la salinité dans la fraction insoluble en Western Blot avec des anticorps anti-GFP. Il était attendu qu'une fraction non négligeable de GFP-Spt16-1 soit retrouvée dans la fraction insoluble contenant l'essentiel de la chromatine à chaque extraction si Spt16-1 est bien un chaperon d'histone. Cependant nous n'avons pas vu de solubilisation de la protéine en augmentant la salinité alors même que nous parvenions à extraire progressivement les histones. Nos tentatives de solubilisation à très forte salinité (2M NaCl), mécanique (par sonication) ou chimiques (Mnase, Benzonase) n'ont pas fonctionné. La faible fraction de protéines GFP-Spt16-1 que nous avons réussi à solubiliser ne permettait pas de réaliser d'immunoprécipitation par un système de GFP-trap. Nous avons tenté de produire des anticorps spécifiques de Spt16-1 mais les deux anticorps produits ne se sont pas révélés suffisamment spécifiques et nous avons dû abandonner la possibilité d'immunoprécipiter Spt16-1. Le remplacement de la GFP s'est aussi révélé ardue, la protéine FLAG-Spt16-1 par exemple étant toxique dans les bactéries lors du clonage.

2. Le domaine MBD pourrait être requis pour l'activité de Spt16-1 dans les réarrangements du génome

La caractéristique commune aux chaperons d'histones est l'existence d'un domaine riche en résidus acides en C-terminal permettant l'interaction avec les histones (Hammond et al., 2017). Chez Spt16, le domaine C assure cette fonction avec en particulier un sous-domaine dénommé MBD. Le domaine MBD est essentiel à l'activité de Spt16 et du complexe FACT chez *S. cerevisiae* et responsable de l'interaction de Spt16 avec les histones H2A-H2B (Kemble et al., 2015). L'analyse de la structure du domaine MBD de Spt16 de *S. cerevisiae* aux rayons X a montré qu'il disposait de deux résidus très conservés, acide et aromatique dont la mutation abroge la fonction de la protéine (Kemble et al., 2015). L'alignement des domaines MBD de Spt16-1 et de ses homologues montrent que les résidus acides et aromatiques sont conservés (Figure 13A). On peut donc supposer qu'ils sont responsables de l'interaction avec les histones chez *P. tetraurelia*. Afin de démontrer que le domaine MBD est essentiel à la fonction de Spt16-1, nous avons réalisé des expériences de complémentation fonctionnelle *in vivo* avec une protéine de fusion GFP-Spt16-1-ΔMBD dans des cellules inactivées pour *SPT16-1* (Figure 13B). La protéine de fusion délétée du domaine MBD ne sauve pas les phénotypes causés par la déplétion de la protéine endogène : la létalité dans la progéniture sexuelle et les défauts d'excision des IES. A première vue, nous pouvons conclure que le domaine MBD est requis pour l'obtention d'une progéniture sexuelle à l'issue de l'autogamie et aux événements d'élimination d'ADN. Cependant, n'étant pas capable de détecter la protéine de fusion GFP-Spt16-1_ΔMBD dans des cellules au cours du développement MAC, il est en l'état impossible de conclure si : 1) l'absence de sauvetage phénotypique provient bien de la délétion du domaine MBD ou 2) du fait que la protéine de fusion n'est pas correctement exprimée.

Pour infirmer ou confirmer la deuxième hypothèse, il est nécessaire de vérifier que la protéine de fusion est correctement produite dans la cellule et de déterminer dans quelle fraction cellulaire. Nous avons tenté d'extraire les protéines totales et les protéines nucléaires par fractionnement cellulaire de cellules transformées par GFP-Spt16-1_ΔMBD. Nous ne sommes pas parvenus à détecter en Western-Blot la protéine de fusion ni dans la fraction totale ni dans la fraction nucléaire ou cytoplasmique, avec un anticorps spécifique

de la GFP. Il est possible que la protéine GFP-Spt16-1_ΔMBD soit trop diluée dans les fractions totales ou cytoplasmiques pour être détectable en Western Blot. Il est aussi possible que la délétion du domaine MBD empêche une expression correcte de la protéine à partir du transgène ou nuise à la stabilité de la protéine reproduisant alors les phénotypes causés par l'absence de Spt16-1. Dans ce cas on ne peut affirmer que c'est le domaine putatif d'interaction avec les histones, MBD, qui est bien requis pour l'activité de Spt16-1 dans le développement. Le domaine MBD étant normalement requis pour l'interaction de Spt16-1 avec les histones, l'explication la plus probable est sans doute que son absence empêche la protéine de fusion d'interagir avec la chromatine et donc sa rétention dans les MAC en développement.

3. Démontrer les interactions entre le domaine MBD de Spt16-1 et les histones

Une expérience alternative consisterait à muter spécifiquement les résidus conservés nécessaires à l'interaction avec les histones. C'est cette stratégie qui a été utilisée pour caractériser le domaine MBD chez *S. cerevisiae* (Kemble et al., 2015). On imagine minimiser de cette manière les problèmes d'expression ou de stabilité de la protéine qu'on imagine plus conséquents après la délétion d'une portion de la protéine. L'idée serait de réaliser une expérience de complémentation *in vivo* avec un transgène *GFP-SPT16-1* dans lequel sont mutés les résidus suivants : acide aspartique et phénylalanine. Ce sont les deux résidus conservés nécessaires à l'interaction avec H2A et H2B chez *S. cerevisiae*. Dans l'étude par (Kemble et al., 2015), ils substituent les résidus correspondants chez *S. cerevisiae* par des alanines. Si la version mutée de GFP-Spt16-1 provoque les mêmes phénotypes de létalité et moléculaire que l'inactivation seule de *SPT16-1* et que l'on détecte la protéine de fusion en fixant des cellules en développement et /ou par Western-Blot, on en conclura que ces acides aminés dans le domaine MBD sont nécessaires à l'action de Spt16-1 pendant le développement MAC. Cependant il est fort probable que cette expérience produise les mêmes résultats que la délétion complète du domaine MBD puisque les mutations ciblées empêcheront aussi l'interaction avec les histones et probablement la rétention de la protéine mutée dans les MAC en développement.

L'ensemble de ces expériences ne nous permettraient cependant pas de montrer que Spt16-1 via ce domaine interagit avec les histones. Pour répondre à cette question, nous avons envisagé de produire *in vitro* les domaines M et C de Spt16-1 pour tester leur capacité à interagir respectivement avec les dimères d'histones H3-H4 et H2A-H2B comme l'ont montré de précédentes études dans d'autres organismes (Hondele et al., 2013; Kemble et al., 2013, 2015). Les domaines M et C sont fusionnés avec la GST et retenus sur une colonne avec du glutathion. On fait passer dans la colonne des histones purifiés, ou bien complexés en dimères. Les histones qui interagissent avec les domaines M ou C sont retenus avec eux sur la colonne et élués en même temps que les domaines par un excès de glutathion. Par Western Blot grâce à des anticorps spécifiques de chaque histone, on est capable de déterminer les histones retenues sur la colonne car interagissant avec les domaines M ou C.

II. Existence d'un complexe FACT chez *P. tetraurelia*

Le complexe FACT est un hétérodimère composé des protéines Spt16 et Pob3 ou Ssrp1 chez les eucaryotes (Formosa, 2012). L'identification d'un homologue de Spt16 et sa nécessité pour le développement de *P. tetraurelia* nous a amené à rechercher un partenaire de Spt16-1 homologue de Pob3 ou Ssrp1. Pour pallier à l'impossibilité d'immunoprécipiter Spt16-1 et d'analyser son domaine MBD, nous avons tenté d'utiliser la protéine Pob3-1.

1. A la recherche d'un partenaire de Spt16-1

1.1 *P. tetraurelia* dispose d'un homologue de Pob3

Comme pour Spt16-1, nous avons identifié des homologues de Pob3 chez *P. tetraurelia* à partir de la protéine de *S. cerevisiae* (Figure 2). Sur les trois homologues de Pob3 trouvés chez *P. tetraurelia*, l'un d'entre eux est exprimé spécifiquement pendant le développement quand les deux autres Pob3-2a et Pob3-2c sont exprimés pendant tout le cycle cellulaire. Nous avons émis comme hypothèse que le Pob3-1 exprimé spécifiquement pendant le développement pouvait être le partenaire de Spt16-1 dans un complexe FACT requis pour les réarrangements du génome. L'analyse des domaines protéiques des homologues de Pob3 par Interpro n'a pas détecté de domaine HMG-Box à l'extrémité C-terminale des protéines. Le domaine HMG-Box est un domaine en C-terminal de Ssrp1 mais

absent de Pob3 chez la levure où il est dans ce cas remplacé par la protéine Nhp6 qui s'associe avec le complexe FACT (Formosa, 2012). Le complexe FACT chez *P. tetraurelia* pourrait être un complexe analogue à celui retrouvé chez *S. cerevisiae* ou *S. pombe*.

1.2 *Pob3-1 a une localisation similaire à Spt16-1*

L'injection du transgène *POB3-1* fusionnée à la GFP en C-terminale a permis de localiser la protéine Pob3-1 dans les macronoyaux en développement (Figure 15). Pob3-1-GFP persiste dans le nouveau MAC après la première division cellulaire. Cette localisation est très semblable à celle de Spt16-1 à la différence que Pob3-1-GFP est aussi retrouvée dans les MIC pendant le développement macronucléaire. La localisation dans les MIC est perdue lorsque l'on perméabilise les cellules indiquant que la protéine est sans doute lavée par la perméabilisation. Il est surprenant qu'une protéine que l'on suppose être un chaperon d'histones et donc potentiellement interagissant avec les histones, soit lavée par la perméabilisation. Cette observation laisse penser que cette localisation n'est pas forcément celle de la protéine Pob3-1 endogène. Nous serions rassurés par la complémentation fonctionnelle par Pob3-1-GFP dans un RNAi *POB3-1*. Il nous est pour le moment impossible de réaliser ces expériences de complémentation puisque l'inactivation de *POB3-1* ne produit pas de phénotypes observables. La localisation dans les MAC en développement n'est par contre pas impactée par la perméabilisation et pourrait indiquer que Pob3-1-GFP y est plus fortement lié à la chromatine. Cette localisation dans les MAC en développement est compatible avec une fonction de chaperon d'histone qui interagit avec les histones. Pour des protocoles de fixation identiques, Pob3-1-GFP et GFP-Spt16-1 sont détectés dans des noyaux de tailles similaires indiquant que les protéines sont sans doute présentes au même moment pendant le développement.

1.3 *Dimérisation de Pob3-1 et Spt16-1*

L'analyse des domaines protéiques de Pob3-1 par Interpro a cependant révélé qu'un autre domaine caractéristique du chaperon d'histone était absent : le domaine N/D de dimérisation Figure 1 : Représentation schématique du complexe FACT et de ses domaines chez la levure (Figure 1). Ce domaine est requis pour former le complexe FACT en permettant à Pob3 de se lier au domaine de dimérisation D de Spt16 (Winkler et al., 2011). Il

semble dans ce cas peu probable que Pob3 puisse former un complexe FACT avec Spt16-1. Cette constatation amène à proposer plusieurs hypothèses : 1) Il est possible que la dimérisation se fasse malgré tout via une portion de la protéine Pob3-1 sans domaine caractérisé ou par son domaine M détecté par Interpro. 2) On peut aussi imaginer que le fonctionnement du couple Pob3-1/Spt16-1 se fasse sans dimérisation des deux protéines en complexe mais ne nécessite que l'interaction de chaque protéine avec les histones. 3) Il est enfin possible que Spt16-1 agisse seul dans le développement ou se complexe à d'autres partenaires non identifiés. Nous avons tenté de répondre à ces différentes hypothèses en immunoprécipitant soit GFP-Spt16-1 soit Pob3-1-GFP mais il n'a pas été possible de les immunoprécipiter pour analyser leur éventuels partenaires. Ces protéines de fusion à la GFP sont majoritairement insolubles. Une approche alternative consisterait à tester l'interaction entre Pob3-1 et Spt16-1 partir de protéines recombinantes Pob3-1 et Spt16-1 auxquelles on fusionne des tags différents (Histidine, Flag, GST etc). Par co-immunoprécipitation ou bien par des expériences de GST-Pulldown on devrait être capable de déterminer si Spt16-1 et Pob3-1 sont capables d'interagir. Il manquerait malgré tout la preuve que ce complexe Pob3-1/Spt16-1 existe réellement dans la cellule.

2. Implication de Pob3-1 dans le développement MAC

L'inactivation par ARN interférence de *POB3-1* pendant l'autogamie ou la conjugaison ne provoque aucune létalité dans la progéniture sexuelle. Pob3-1 ne semble donc pas requis pour l'obtention d'une progéniture sexuelle viable ni pour le développement macronucléaire. Comme nous n'avons pas examiné les niveaux des ARN messagers de *POB3-1*, on ne peut pas affirmer avec certitude que le RNAi *POB3-1* a été efficace. Il est possible que l'absence de phénotypes de létalité dans la progéniture sexuelle soit dû à l'inefficacité du RNAi. Pour autant, dans le contrôle positif, l'inactivation de *PGM*, il n'y a pas de production d'une progéniture sexuelle viable comme attendu et les expériences ont été répétées quatre fois de manière indépendante avec toujours les mêmes résultats. Il est plus raisonnable de penser que Pob3 n'est pas essentiel au développement MAC. Chez *S. pombe*, la délétion de *POB3-1* ne provoque pas de létalité contrairement à ces homologues *POB3* ou *SSRP1* chez d'autres eucaryotes (Formosa, 2008). L'absence de phénotype lié à l'inactivation de *POB3-1* chez *P. tetraurelia* pourrait très bien être similaire à ce que l'on observe chez *S.*

pombe. Il est donc envisageable que *POB3-1* chez *P. tetraurelia* ne soit pas non plus un gène essentiel tout en participant au complexe FACT.

3. Modèle d'action des différents Pob3 et Spt16 chez *P. tetraurelia*

Chez tous les eucaryotes où Spt16 est étudiée, la protéine est presque systématiquement retrouvée en complexe avec la protéine Pob3 ou Ssrp1 pour former le chaperon d'histone FACT. La littérature ne fournit pas particulièrement d'exemple de rôles de Spt16 ou Pob3/Ssrp1 seuls. Les analyses de transcriptome chez *H. sapiens* ou *A. thaliana* après inactivation de Spt16 ou Ssrp1 montrent que certains gènes sont spécifiques d'un des composants de FACT (Li et al., 2007; Pfab et al., 2018). Ces données laissent penser que Spt16 et Ssrp1 ont des rôles indépendants dans la transcription. Des complexes protéiques de Ssrp1 sans Spt16 ont été purifiés à partir de cellules Hela suggérant que Ssrp1 pourrait avoir un rôle indépendamment du complexe FACT. Observant l'absence de domaine de dimérisation chez Pob3-1, nous avons émis plusieurs hypothèses concernant l'existence d'un complexe FACT composé par Pob3-1 et Spt16-1. Dans le cas des deux premières hypothèses, Pob3-1 et Spt16-1 s'associent pour former un complexe FACT soit par une interaction directe soit par une interaction indirecte via le nucléosome par exemple. Les profils d'expression, les phénotypes dus aux inactivations et la littérature sur le complexe FACT nous laissent penser que, chez *P. tetraurelia*, les quatre Spt16 et les trois Pob3 pourraient s'associer pour former différents complexes FACT requis pendant le cycle végétatif et pendant le cycle sexuel (Figure 19). S'il s'avère que Spt16-1 interagit bien avec Pob3-1, le complexe FACT serait ainsi plus proche de celui que l'on retrouve chez *S. pombe* où *POB3* n'est pas un gène essentiel (Lejeune et al., 2007). Notre modèle proposerait que Spt16-1 et Pob3-1 chez *P. tetraurelia* réarrangent la chromatine spécifiquement dans les nouveaux MAC pour permettre à la machinerie d'excision d'interagir avec l'ADN pour assurer l'élimination efficace de l'ensemble des séquences germinales. Les autres Spt16-2a, 2b, 2c et Pob3-2a et 2c s'associeraient en complexe FACT mais requis pendant l'ensemble du cycle de vie de la cellule pour remplir des rôles plus canoniques pour le complexe FACT tels que régulation de la transcription, de la réplication, de la réparation, etc. Pour autant l'hypothèse la plus vraisemblable serait la première proposée dans les paragraphes précédents. L'absence de domaine de dimérisation et de phénotypes à l'inactivation

comparables à ceux provoqués par l'inactivation de SPT16-1 est plutôt en désaccord avec l'hypothèse d'un complexe FACT. Il est tout à fait possible d'imaginer que chez *P. tetraurelia*, Spt16-1 agit seule, en dehors du complexe FACT et a acquis un rôle spécifique du développement sans l'intervention de Pob3. La littérature propose très peu d'évidences de rôles de Spt16 isolée mais il est possible que ce rôle ait été acquis spécifiquement chez *Paramecium*.

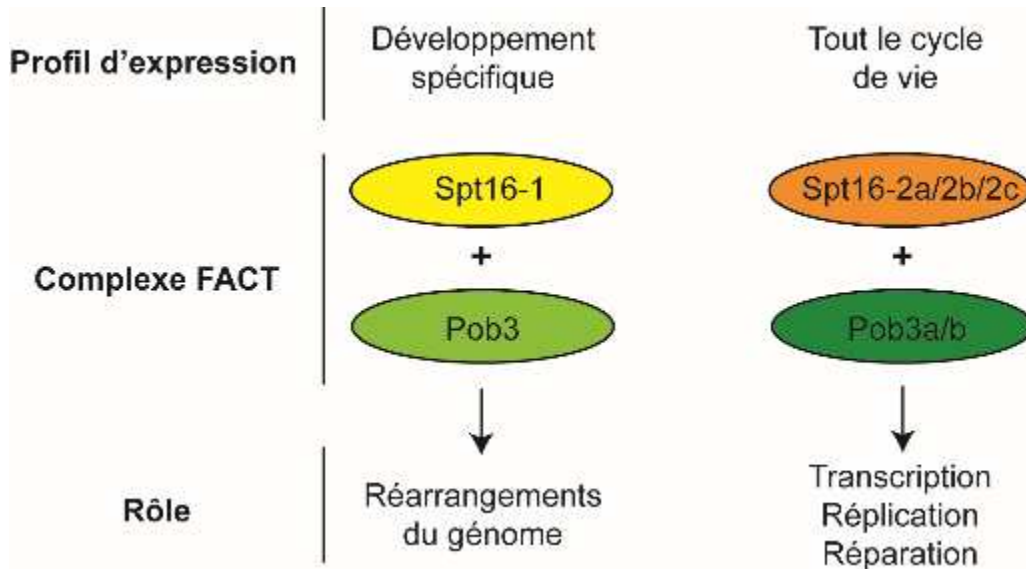


Figure 19 : Modèles d'action de Spt16-1 et de composition de complexes FACT chez *P. tetraurelia*

On propose que les protéines exprimées spécifiquement pendant le développement que sont Spt16-1 et Pob3-1 s'associent pour former un complexe FACT spécifique du développement et dont le rôle principal serait dédié à aux réarrangements du génome. Les autres Spt16-2a, 2b et 2c ainsi que Pob3-2a et 2b qui sont exprimées de manière constitutive pourraient former des complexes FACT requis pour des fonctions plus classiques comme la transcription, la réplication ou la réparation en dehors du développement.

III. Comprendre comment Spt16-1 permet les réarrangements du génome

Nous avons montré précédemment que Spt16-1, en plus d'être requis pour la production d'une progéniture sexuelle viable et le développement macronucléaire, est nécessaire à l'ensemble des événements d'élimination d'ADN se produisant pendant le développement MAC. La déplétion de Spt16-1 entraîne la rétention de l'ensemble des éléments répétés éliminés imprécisément et de l'ensemble des 45000 IESs excisées précisément. Nous nous

positions alors la question de savoir à quel moment des réarrangements, Spt16-1 intervient et si la protéine interagissait avec les voies déjà identifiées dans l'élimination d'ADN. Pour cibler le moment d'action de Spt16-1 dans les réarrangements nous avons observé les effets de la déplétion de Spt16-1 sur les populations de petits ARNs et les marques d'hétérochromatine H3K9me3 et H3K27me3. Nous avons conclu que Spt16-1 était un facteur général des réarrangements du génome agissant en aval des voies de petits ARN et de marques d'hétérochromatine.

1. L'action de Spt16-1 dans le développement macronucléaire n'est pas régulé par les scnARN ou les marques d'hétérochromatine

Nous avons montré que Spt16-1 agit en aval de la déposition des marques d'hétérochromatine H3K9me3 et H3K27me3 et de la production des scnARN. Spt16-1 n'est requis ni pour la production de scnARN ni pour l'enrichissement en séquences MIC spécifiques consécutives au processus de scanning auquel Spt16-1 ne participe donc pas (Coyne et al., 2012; Lepère et al., 2009). Dans des données que nous n'avons pas présentées, nous avons observé que la localisation de GFP-Spt16-1 n'est pas impactée par la déplétion en Dcl2 et 3 responsables de la production des scnARNs (Lepère et al., 2009). De la même manière, l'absence de Spt16-1 n'empêche pas le dépôt des marques d'hétérochromatines H3K9me3/H3K27me3 dans les MAC en développement requises pour l'élimination d'une majorité de séquences MIC spécifiques (Lhuillier-Akakpo et al., 2014). Nous avons également observé la localisation de GFP-Spt16-1 après l'inactivation de *EZL1* responsable du dépôt d'H3K9me3/H3K27me3 et conclu qu'Ezl1 n'était pas requis pour la localisation correcte de GFP-Spt16-1. Même si Spt16-1 est bien localisée, ces résultats ne permettent pas de conclure si elle est bien active en l'absence de Dcl2/3 ou d'Ezl1. Cependant, puisque les phénotypes moléculaires consécutifs à l'inactivation de *EZL1* et *DCL2/3* sont moins drastiques que ceux observés en inactivant *SPT16-1*, on peut supposer que Spt16-1 est toujours active dans les MAC en développement (Lhuillier-Akakpo et al., 2014). Ces données suggèrent que Spt16-1 agit non seulement en aval mais agit aussi indépendamment des voies d'élimination d'ADN nécessitant les scnARN et les marques d'hétérochromatine H3K9me3 et H3K27me3.

1.1 *Spt16-1 n'est pas recruté par l'intermédiaire des scnARNs ou des marques d'hétérochromatine*

La présence de H3K9me3 et H3K27me3 n'influe pas sur la localisation de Spt16-1 dans les nouveaux MAC. On sait par immunofluorescence et par CHIP que Spt16 chez *S. cerevisiae* ou *H. sapiens* est localisée dans le noyau et interagit avec la chromatine (Feng et al., 2016b; Mayer et al., 2010; True et al., 2016b). Si rien ne permet d'affirmer que Spt16-1 interagit avec la chromatine chez *P. tetraurelia*, on peut raisonnablement imaginer que dans notre cas l'absence d'Ezl1 ne modifiant pas la localisation de Spt16-1, le recrutement du chaperon d'histone ne semble pas influencé par les modifications d'histones H3K9me3 et H3K27me3 comme c'est parfois le cas dans d'autres organismes. Chez *S. cerevisiae*, les modifications H3K36me3 ou H3K4me3 altèrent la fonction de FACT dans la transcription (Biswas et al., 2005). Dans l'introduction, j'avais détaillé le fait que des protéines interagissant avec la chromatine comme HP1c assuraient le recrutement de FACT (Kwon et al., 2010). HP1c lie l'euchromatine chez *D. melanogaster*, interagit avec FACT et l'ARN polymérase II. Si nous étions parvenus à immunoprécipiter Spt16-1 nous aurions pu tenter de montrer si Spt16-1 est éventuellement localisé sur la chromatine en fonction d'autres marques des histones.

On note de la même manière que le recrutement de Spt16-1 ne dépend pas non plus des petits ARN non codants, les scnARN. Lorsque Dcl2 et 3 ne sont pas présents, Spt16-1 est toujours localisé dans les MAC en développement. Il ne semble pas y avoir pour le moment d'évidence dans la littérature chez d'autres eucaryotes que des petits ARN pourraient participer au recrutement de FACT ou Spt16 sur la chromatine et cela ne semble pas être le cas non plus pour *P. tetraurelia*.

1.2 *Spt16-1 permet de préciser comment sont reconnues et éliminées les séquences ne dépendant pas de Dcl2/3 ou Ezl1*

Si Spt16-1 est recrutée de manière spécifique à proximité des éléments à exciser, ce n'est vraisemblablement pas par l'intermédiaire des marques d'hétérochromatine et par les scnARN. Il est tout à fait possible d'imaginer que Spt16-1 se localise sur l'ensemble du génome indépendamment des séquences à éliminer et promeuve un relâchement général

de la chromatine pour rendre accessibles les séquences MIC spécifiques à la coupure. En partant de cette hypothèse, il, est alors facile d’imaginer comment sont rendues accessibles les séquences à éliminer pour la machinerie d’excision grâce au relâchement général de la chromatine. Un CHIP Spt16-1 nous aurait permis de déterminer où se localise Spt16-1 sur le génome. Malheureusement, il nous a été impossible d’utiliser la protéine GFP-Spt16-1 ou des anticorps contre Spt16-1 pour réaliser ces expériences de CHIP.

Cette hypothèse ne permet cependant pas de préciser plus ce que notre modèle proposait déjà : comment est signalée et reconnue la majorité des séquences MIC spécifiques (Figure 12, 20). J’avais expliqué dans l’introduction qu’une portion des IES ne dépend pas pour son excision ni des voies de scnARN ni d’Ezl1 (Lhuillier-Akakpo et al., 2014). Elles sont parmi les plus petites IES, avec une taille bien inférieure à celle du nucléosome ou du linker d’ADN et ne présentent pas de séquence consensus. L’intervention d’un chaperon d’histone comme Spt16-1 permet d’imaginer que certaines de ces très petites séquences qui pourraient être complètement masquées par le nucléosome seraient rendues accessibles à la coupure par l’ouverture de la chromatine. Les modifications portées par les histones ne peuvent par contre pas expliquer la précision de la coupure des IES à la base près et offrent une résolution non compatible avec la reconnaissance d’IES aussi petites. Il apparaît évident que d’autres acteurs sont requis pour marquer les séquences germinales et assurer le recrutement correct de Pgm. On émet alors l’hypothèse que la méthylation de l’ADN pourrait cibler précisément les sites de coupure. Le génome de *P. tetraurelia* est méthylé sur les adénines et je développerai dans le chapitre suivant les observations qui nous permettent de penser que la méthylation des adénines pourrait être impliquée dans les réarrangements du génome.

2. Spt16-1 pourrait faciliter l’accès à l’ADN de la machinerie d’excision

Nos travaux ont montré que Spt16-1 est requis pour la localisation correcte de Pgm dans les MAC en développement. A l’inverse, Pgm n’est pas requis pour la localisation de GFP-Spt16-1. Nous avons conclu que Spt16-1 agissait en amont de Pgm. Nous nous étions assurés que Spt16-1 n’était pas requis pour l’expression de Pgm. Nous avons constaté que lorsque Spt16-1 est déplétée, si Pgm est absente de la majorité des MAC en développement

(70%), dans une proportion de cellules (30%) un signal nucléaire d'intensité similaire aux conditions sauvages est observée. Nous avons suggéré que Pgm, en l'absence de Spt16-1, transite continuellement entre les noyaux et le cytoplasme et ne parvient pas à résider suffisamment longtemps dans les MAC. Plus que la localisation de Pgm, Spt16-1 permettrait la rétention de Pgm dans les MAC. Ces observations nous ont amené à imaginer un modèle dans lequel le chaperon d'histone Spt16-1 faciliterait les interactions entre Pgm et l'ADN des séquences germinales à éliminer en modifiant la chromatine localement (Figure 20). C'est l'interaction de Pgm avec la chromatine ou l'ADN qui permettrait sa rétention dans les MAC en développement.

Le modèle proposé est similaire aux mécanismes observés chez d'autres eucaryotes pour assurer le rôle de Spt16-1 dans la réplication, dans la transcription ou encore dans la réparation (Formosa, 2012; Jeronimo and Robert, 2016; Yang et al., 2016). Malgré la diversité des processus dans lequel le chaperon d'histone FACT est impliqué, un point commun est sa faculté à recruter les machineries de transcription, réplication, réparation à l'ADN en réorganisant la chromatine. Des études ont montré que FACT est retrouvé conjointement avec la RNA polymérase II sur les corps des gènes transcrits chez *D. melanogaster*, *H. sapiens* ou *S. cerevisiae* (Feng et al., 2016b; Li et al., 2007; Mayer et al., 2010; Saunders et al., 2003; True et al., 2016b). En réorganisant la chromatine, il facilite l'interaction de la polymérase avec l'ADN et sa progression pour transcrire les gènes. Il a été montré chez *S. cerevisiae* que le désassemblage de la chromatine via la liaison de FACT aux histones était nécessaire à l'initiation de la transcription et à la liaison de facteurs de transcription comme TBP sur la TATA Box (Biswas et al., 2005). Lors de la réplication, le complexe FACT interagit avec la DNA polymérase 1 mais aussi avec des composants du replisome comme l'hélicase MCM (Yang et al., 2016). FACT désassemble alors la chromatine pour permettre l'interaction de la polymérase et de l'hélicase avec l'ADN et ainsi la réplication. On imagine que de manière similaire Spt16-1 chez *P. tetraurelia* pourrait réorganiser la chromatine aux sites d'excision pour faciliter l'interaction de la machinerie d'excision avec l'ADN et l'introduction de cassures double brin. Plus récemment, il a été montré que FACT interagissait avec l'intégrase du HIV-1 et promouvait l'intégration du HIV-1 *in vitro* sur de la chromatine. Les auteurs ont conclu comme dans notre cas, que FACT

favorise l'interaction de l'intégrase avec l'ADN en promouvant une conformation plus ouverte de la chromatine (Matysiak et al., 2017).

Il est raisonnable de se demander s'il y a une interaction dans notre système entre Pgm et Spt16-1. Pour tester cette hypothèse, nous pourrions produire les deux protéines auxquelles on fusionne des tags différents (Histidine, Flag, GST etc). Par co-immunoprécipitation ou bien par des expériences de GST-Pulldown on devrait être capable de déterminer si Spt16-1 et Pgm interagissent. Pour autant il est peu probable que Spt16-1 et Pgm interagissent directement dans les MACs en développement étant donné que l'inactivation de *PGM* n'a pas d'effet sur la localisation de GFP-Spt16-1. On suppose plutôt que Pgm et Spt16-1 se localisent à proximité sur le génome et que plus que l'interaction entre Pgm et Spt16-1 c'est l'ouverture de la chromatine qui promouvrait la rétention de Pgm dans le MAC en développement.

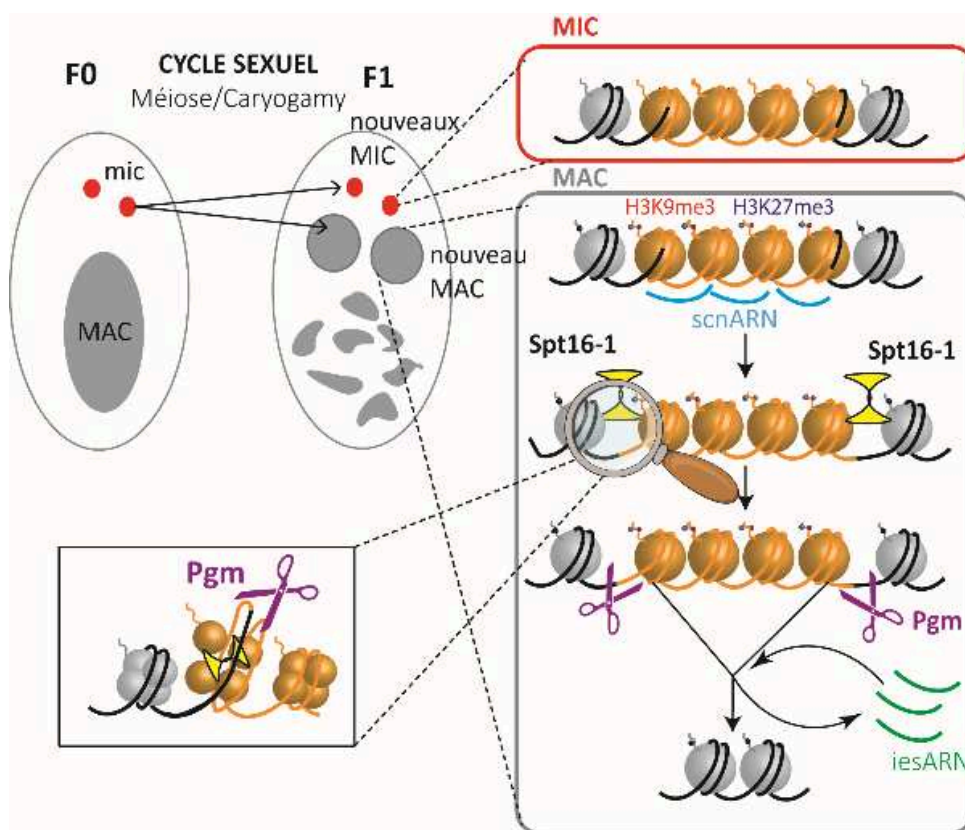


Figure 20 : Modèle d'action de Spt16-1 dans les réarrangements du génome

Au cours du développement des nouveaux MACs, la majorité des séquences MICs spécifiques seraient ciblées par les scnARN qui guident le dépôt des marques d'hétérochromatine. En aval le chaperon d'histone Spt16-1 viendrait relâcher la chromatine et libérer l'ADN afin que la machinerie d'excision interagisse avec la chromatine et introduise des cassures double brins.

3. Spt16-1 pourrait-il faciliter la réparation chez *P. tetraurelia* ?

On a observé que GFP-Spt16-1 persiste tardivement dans les MAC en développement y compris après la première division cellulaire dans le nouveau MAC. Cette localisation pourrait indiquer que Spt16-1 a un rôle supplémentaire plus tardif, par exemple dans la réparation. La machinerie de réparation accompagne la machinerie d'excision pendant les réarrangements d'ADN. Des expériences de co-immunoprécipitation montrent que le dimère Ku70/80 acteurs du NHEJ réparant les cassures double brins chez *P. tetraurelia* interagit avec Pgm dans des extraits cellulaires (Marmignon et al., 2014). On pourrait imaginer alors que Spt16-1 qui participerait au recrutement de la machinerie d'excision pourrait persister avec la machinerie d'excision pour réparer les cassures double brins. Pour répondre à cette hypothèse, nous avons comparé les phénotypes produits par l'inactivation de *SPT16-1*, *PGM* et des autres acteurs de la réparation.

Les phénotypes observés après l'inactivation des acteurs connus du NHEJ chez *P. tetraurelia* sont similaires à ceux que nous avons observés pour Spt16-1 et ne sont pas incompatibles avec un rôle dans la réparation. Les protéines Ku70a et Ku80c sont exprimées plus tardivement que Spt16-1 pendant le développement macronucléaire mais leur déplétion provoque aussi la rétention de l'ensemble des IES et des éléments répétés (Betermier and Duharcourt, 2014; Marmignon et al., 2014). Des fusions protéiques de Ku70a et 80c avec la GFP se localisent dans les MAC en développement comme on l'observe pour Spt16-1 (Marmignon et al., 2014). Une fusion protéique à la GFP de la ligase IV (*LIG4*), responsable de la réparation des cassures double brins dans le NHEJ, se localise dans les macronoyaux en développement et persiste dans le nouveau MAC et la cellule à la fin de l'autogamie comme Spt16-1 (Kapusta et al., 2011). La déplétion de la ligase IV entraîne l'accumulation de cassures double brins aux bornes des IES et empêche la réparation (Kapusta et al., 2011). Par ailleurs dans un RNAi *LIG4*, comme lorsque l'on inactive *SPT16-1* on n'observe pas de formation de jonction d'excision ni de production de cercles d'IES qui sont les produits d'élimination des IES quand bien même on parvient à détecter l'introduction de cassures double brins qui s'accumulent. Nous avons évoqué que l'absence de jonctions d'excisions pouvait tout autant s'interpréter comme une absence de cassures double brins ou l'absence de réparations de cassures double brins comme dans un RNAi

LIG4. Pour tenter de vérifier l'absence de cassures double brins dans un RNAi *SPT16-1*, nous avons cherché à observer dans le séquençage d'un point tardif du développement de l'ADNg après inactivation de *SPT16-1* si des lectures s'accumulaient au niveau des jonctions des éléments MIC spécifiques. Il semble que comme pour une inactivation *PGM* mais contrairement à une inactivation *LIG4*, nous n'observons pas de lectures s'accumulant aux bornes des éléments MIC spécifiques (Arnaiz, communication personnelle) indiquant que des cassures ne seraient pas introduites. De plus, l'accumulation des cassures dans un RNAi *LIG4* altère l'endoréplication produisant des nouveaux MACs avec un marquage au Hoecsht très faible dans les temps tardifs du développement. Quand *SPT16-1* est inactivé comme pour *PGM*, le marquage Hoecsht est similaire au contrôle indiquant que les cycles d'endoréplication ont bien eu lieu et qu'il n'y vraisemblablement pas eu d'accumulations de cassures. Quoiqu'il en soit, pour observer un effet de *Spt16-1* sur la réparation il faudrait parvenir à découpler les effets de la protéine sur les réarrangements et la réparation, ce qu'il nous est impossible de faire pour le moment. Il nous semble bien plus probable que *Spt16-1* empêche l'introduction de cassures double brins et il paraît peu probable que *Spt16-1* soit lié à la réparation des cassures.

4. *Spt16-1* et son effet sur la transcription chez *P. tetraurelia*

Le complexe FACT est principalement connu pour son rôle dans la transcription chez les eucaryotes. L'inactivation de *SPT16* ou de *POB3* provoque des modifications dans les niveaux d'expression de gènes aux fonctions variées chez *S. cerevisiae*, *A. thaliana* ou encore *H. sapiens* (Brewster et al., 1998; Li et al., 2007; Pfab et al., 2018). Au début de notre étude, nous avons émis la possibilité que *Spt16-1* puisse être un facteur de la transcription comme ses homologues eucaryotes et réguler l'expression de certains acteurs des réarrangements d'ADN comme l'endonucléase Pgm. Par Western Blot, nous avons montré que la protéine Pgm était bien produite lorsque l'on inactivait *SPT16-1* ce qui infirmait notre hypothèse. *Spt16-1* ne régule pas le niveau d'expression de Pgm. On ne peut cependant pas exclure que *Spt16-1* modifie le niveau d'expression d'autres gènes impliqués dans le développement. On ne peut pas exclure non plus que *Spt16-1* soit nécessaire à la transcription de gènes dont la protéine produite est requise pour la localisation correcte de Pgm dans les MAC en développement (molécules d'import/export nucléaires, acteurs des réarrangements,

protéines nucléaires...). Dans cette hypothèse Spt16-1 agirait alors de manière indirecte sur la localisation de Pgm.

Pour répondre à cette question nous avons analysé le transcriptome de cellules inactivées pour *SPT16-1*. Nous n'avons pas présenté ces données car elle ne montrait pas de modifications très importantes dans les niveaux d'expression des gènes en absence de Spt16-1. Nous n'avons pas identifié de modifications particulières dans les niveaux d'expression d'un certain nombre de gènes identifiés comme des acteurs des réarrangements. On dénombre 3338 gènes sous-exprimés et 3051 gènes surexprimés par rapport aux conditions contrôles, ce qui fait seulement 10% des gènes impactés par l'absence de *SPT16-1*. Ces observations sont plutôt cohérentes avec ce qui est observé dans les transcriptomes de d'autres espèces en absence de Spt16. Chez *H. sapiens* ou *A. thaliana*, la proportion de gènes qui voient leur niveau d'expression changé reste très faible (Li et al., 2007; Pfab et al., 2018). Si Spt16-1 a un rôle dans la transcription il est difficilement observable dans ces conditions. Il nous apparaît plus probable que Spt16-1 n'ait pas de rôle particulier dans la transcription au contraire de ces homologues eucaryotes.

Chapitre 2 : Méthylation des adénines

Un nombre croissant d'études parvient à détecter la méthylation des adénines chez des eucaryotes variés grâce à l'amélioration des méthodes utilisées. 6mA avait déjà été identifiée chez le cilié *P. tetraurelia* par chromatographie à l'aide de nucléotides marqués radioactivement (Cummings et al., 1974). Mon travail de thèse a permis de confirmer la présence de 6mA chez *P. tetraurelia* et de décrire sa localisation dans la cellule et au cours du cycle de vie.

I. Identification de modifications dans l'ADN de *P. tetraurelia*

Nous étions partis des résultats obtenus en 1974 par Cummings décrivant que les génomes MIC et MAC de *P. tetraurelia* mais aussi de plusieurs autres espèces du groupe aurelia, contiennent 6mA mais pas de 5mC (Cummings et al., 1974). Nous avons décidé d'utiliser la technique d'HPLC-MS pour confirmer ces résultats.

1. 5mC, 5hmC, 4mC et 8oxoG ne sont pas détectées chez *P. tetraurelia*

Nous avons montré par HPLC-MS que l'ADN génomique de *P. tetraurelia* ou *P. sexaurelia* ne contient pas ou très peu de méthylation des cytosines 5mC quel que soit le moment du cycle cellulaire. Nos analyses d'HPLC-MS ne permettent pas de quantifier chaque modification et nous sommes bien entendus dépendant de la sensibilité de la méthode de détection. Nous pouvons simplement en conclure que s'il existe du 5mC sur le génome de *Paramecium*, celui-ci est indétectable par HPLC-MS. Les chromatogrammes obtenus pour *E. coli* Dam-Dcm⁻ sont très similaires à ceux obtenus pour *P. tetraurelia* ou *sexaurelia*, ne montrant pas de pic au temps de rétention correspondant à 5mC. Nous avons utilisé comme contrôle positif l'ADNg d'*E. coli* sure pour lequel on détecte comme attendu un pic important au temps de rétention pour 5mC. C'était déjà le constat de Cummings en 1974 qui ne parvenait pas à détecter de 5mC et en concluait que la modification était absente du génome de *P. tetraurelia* (Cummings et al., 1974). D'autres eucaryotes comme *D. melanogaster*, *C. elegans* ou *S. cerevisiae* ne possèdent pas de 5mC dans leur génome

(Breiling and Lyko, 2015). Chez *P. tetraurelia* comme chez *C. elegans* ou *S. cerevisiae* par exemple, on ne parvient pas à détecter d'homologues de cytosine méthyltransférases responsables de la méthylation *de novo* ou de maintenance par exemple chez les mammifères (Ambrosi et al., 2017).

Nous avons aussi à notre disposition des molécules standards pour les nucléotides modifiés 5hmC, 4mC et 8oxoG. Nous ne sommes pas parvenus à détecter ces modifications dans l'ADN génomique de *P. tetraurelia* ou *P. sexaurelia* quelle que soit la condition à laquelle nous avons extrait l'ADN. Finalement l'absence de 5hmC nous paraît logique étant donnée l'absence de 5mC qui est le précurseur de 5hmC, dans le génome de *P. tetraurelia*. Les enzymes TET déméthylent 5mC et produisent l'intermédiaire 5hmC (Shi et al., 2017). Nous ne détectons pas non plus d'homologues d'enzymes TET dans le génome de *P. tetraurelia*. La 8oxoG est une modification résultant de lésions de l'ADN et de stress oxydatif et provoquant des mutations dans l'ADN (Kanvah et al., 2010). Nous n'avons pas non plus détecté par cette méthode de 8oxoG. Les molécules de 4mC et 5mC ont le même poids moléculaire mais nous sommes parvenus à les séparer sur notre colonne d'HPLC-MS. Pour autant nous ne parvenons pas à détecter de 4mC avec la sensibilité de notre technique dans nos échantillons. Ces résultats négatifs sont à prendre avec précaution. D'une part nous n'avons pas de contrôle positif de la détection de 4mC dans nos échantillons. *E. coli* ne contient pas de 4mC. Il aurait fallu utiliser de l'ADN de souches bactériennes comme *Pseudomonas alcaligenes* ou de bactéries thermophiles dans lesquelles la présence de 4mC est attestée (Morgan et al., 2016). Par ailleurs nous avons entamé l'analyse du séquençage par SMRT sequencing du génome de *P. tetraurelia* et avons détecté une grande quantité de 4mC, près de 8% des cytosines. Nous souhaitons valider de manière indépendante du séquençage la présence ou absence de 4mC. Il est possible que notre méthode d'HPLC-MS ne soit pas assez sensible pour détecter 4mC dans le génome de la paramécie mais vu les quantités détectées par le séquençage PacBio il semble plutôt que 4mC soit un artefact du séquençage par SMRT sequencing. Il est possible que ce signal produit par la polymérase soit mal interprété par l'analyse et nous nous posons la question de savoir si cet artefact est spécifique d'un contexte génomique, d'un type de séquence. La présence de 4mC a été récemment décrite dans l'ADN de certains eucaryotes par SMRT sequencing (Ye et al., 2017).

Cependant aucune technique indépendante n'est venue confirmer la présence de 4mC dans l'ADN de ces espèces (anticorps ou HPLC-MS).

2. Détection de 6mA dans l'ADN de *P. tetraurelia*

Les expériences d'HPLC-MS sur l'ADN génomique de *P. tetraurelia* et *P. sexaurelia* à différents temps du cycle cellulaire ont confirmé les résultats obtenus par Cummings (Cummings et al., 1974) : 6mA est abondamment détecté. Leurs niveaux de détection semblent comparables à ce que nous observons dans l'ADN génomique extrait de souches sure d'*E. coli*. 2% des adénines sont méthylées dans le génome d'*E. coli* comme l'atteste l'analyse du génome par SMRT sequencing (Fang et al., 2012). Il nous est impossible de quantifier le pourcentage d'adénines méthylées dans nos expériences d'HPLC-MS mais connaissant la proportion détectée dans l'ADNg d'*E. coli* et celui précédemment détecté par Cummings chez *P. tetraurelia* (2,5%), on imagine que les intensités de signal relevées par HPLC-MS sont comparables entre *P. tetraurelia* et *E. coli* (Cummings et al., 1974). Il est important de noter que le niveau de méthylation ne semble pas beaucoup varier entre les stades du cycle cellulaire que nous avons observés. On peut supposer que la méthylation est maintenue tout au long du cycle y compris pendant le cycle sexuel et le développement bien qu'il soit impossible de quantifier le niveau de méthylation par HPLC-MS.

Cummings précisait dans ses travaux qu'il parvenait à détecter près de 3% d'adénines méthylées dans le génome micronucléaire (Cummings et al., 1974). L'analyse du génome par HPLC-MS ne nous a pas permis de différencier entre ADN MIC et MAC puisque nous avons extrait l'ADN total de nos cellules. Il est possible que les échantillons de MIC préparés par Cummings par fractionnement cellulaire aient été contaminés en fragment de MAC hautement polyploïdes, ce qui aurait biaisé le pourcentage d'adénines méthylées détectées. Nous avons désormais la possibilité de purifier les MICs avec une grande pureté, en réduisant au maximum les contaminations en ADN MAC ou en ADN bactérien (Guérin et al., 2017). La technique mise au point au laboratoire permet de trier par FACS les MIC fluorescents. Cependant la quantité d'ADN récupérée est insuffisante pour réaliser des analyses avec notre protocole d'HPLC-MS : par tri des MIC environ 100ng sont obtenus quand 500ng de nucléosides sont nécessaires pour nos analyses en HPLC-MS à partir de 10µg d'ADN. On peut envisager de compiler des extractions d'ADN MIC pour obtenir une

quantité suffisante d'ADN à analyser mais cela semble difficilement réalisable. Les quantités d'ADN MIC que nous parvenons à obtenir sont par contre compatibles avec la détection de 6mA par un anticorps spécifique en Dot Blot puisque nous parvenons à détecter du signal dès 50ng d'ADN déposé. Cet ADN MIC pourrait aussi être utilisé à terme dans des analyses par séquençage de la méthylation.

Les résultats que nous avons obtenus par Dot Blot nous ont permis de confirmer par une méthode indépendante de l'HPLC que le génome de *P. tetraurelia* contient bien de la méthylation des adénines. L'anticorps anti-6mA est un anticorps initialement utilisé pour détecter la méthylation des adénines sur les ARN. Il est utilisé pour détecter 6mA dans l'ADN de *C. elegans* et sa spécificité a été démontrée sur des oligonucléotides simple ou double brins (Greer et al., 2015). Dans cette étude, les auteurs montrent que l'anticorps est aussi capable de reconnaître 1mA issus de dommages de l'ADN. Nous ne pouvons donc pas exclure que le signal que nous détectons par l'anticorps anti-6mA provienne aussi d'1mA. Cependant cette modification est généralement minoritaire chez les eucaryotes et il semble plus probable que la majorité du signal provienne bien de 6mA (Iyer et al., 2011). Même si nous avons montré que le signal que nous détectons pour 6mA en Dot-Blot provient exclusivement de l'ADN, il est tout à fait possible que la méthylation des adénines soit aussi retrouvée dans l'ARN chez *P. tetraurelia*. Nous ne sommes pas parvenus à détecter 6mA dans les ARN totaux de *P. tetraurelia* en Dot-blot. Pour répondre à cette question, il faudrait analyser par HPLC-MS des échantillons d'ARN pour y détecter 6mA.

II. La méthylation des adénines est établie pendant le développement macronucléaire

1. Méthylation *de novo* et de maintenance chez *P. tetraurelia*

L'anticorps que nous avons utilisé pour détecter 6mA nous a permis de décrire sa localisation dans la cellule et pendant le cycle cellulaire par immunofluorescence. Nos résultats nous ont permis de confirmer que quel que soit le moment du cycle cellulaire la méthylation des adénines est détectée dans le génome MAC. Elle s'établit progressivement dans les nouveaux MAC pendant le développement. Elle apparaît de manière transitoire

dans les MIC d'abord en méiose I puis très précocement dans le développement MAC. Bien que nous n'ayons pas quantifié précisément le signal 6mA, il semble que l'intensité du signal 6mA dans le nouveau MAC après la première division cellulaire soit comparable à l'intensité dans le MAC de cellules végétatives. On suppose qu'une fois établie à la fin du développement la méthylation est maintenue globalement au même niveau au fur et à mesure des divisions cellulaires. Le niveau de méthylation relativement constant dans les expériences d'HPLC-MS entre les différentes conditions s'expliquent relativement bien en Immunofluorescence. A chaque stade, l'ensemble des noyaux sont positifs à la méthylation et même lorsque 6mA est établie dans les nouveaux MAC les fragments du MAC maternel sont toujours abondamment méthylés. Ces résultats nous permettent de distinguer une méthylation dite de maintenance qui conserve 6mA au cours des divisions cellulaires dans le génome MAC et une méthylation *de novo* qui s'établit progressivement dans les nouveaux MAC au cours du développement. Chez les plantes et les mammifères, la méthylation 5mC de maintenance et *de novo* sont catalysées par des enzymes différentes. Chez les mammifères, DNMT1 catalyse la méthylation de maintenance de préférence sur l'ADN hémi-méthylé. Elle permet ainsi de maintenir la méthylation et l'information épigénétique au cours des divisions cellulaires et de la réplication (Breiling and Lyko, 2015). Les enzymes de la famille DNMT3 (A, B, C et L) assurent la méthylation *de novo* principalement au début de l'embryogénèse. On peut tout à fait imaginer un système similaire chez *P. tetraurelia* où des méthyltransférases différentes permettent la maintenance ou l'établissement de la méthylation.

2. A la recherche de méthyltransférases des adénines

L'hypothèse de plusieurs méthyltransférases pour des méthylation *de novo* et de maintenance est appuyée par la recherche d'homologues de l'adénine méthyltransférase Damt-1 de *C. elegans* (Greer et al., 2015) dans le génome de *P. tetraurelia* qui nous a permis d'identifier 12 méthyltransférases putatives. Nous avons réalisé au laboratoire une analyse phylogénétique des méthyltransférases à domaine MTA70 que nous avons identifiées et les avons comparées à un ensemble de protéines déjà identifiées pour certaines comme des ARN ou ADN méthyltransférases sur les adénines (Figure 21). L'analyse des profils d'expressions de chacun des gènes correspondants par RNAseq révèle que quatre d'entre

eux sont exprimés spécifiquement pendant le développement (*NMAD05,06,07,08* appartenant au clade 4 des ZZ domain) et huit autres de manière constitutive (*NMAD01,02,11,12* dans les clades 1 et 2, *NMAD03,04,09,10* dans le clade 5) (Figure 21) (Arnaiz et al., 2017). Les quatre spécifiques du développement pourraient correspondre à la méthylation *de novo* quand les autres se préoccuperaient de la méthylation de maintenance. L'ensemble de ces enzymes appartiennent à la famille MT-A70 des méthyltransférases putatives eucaryotes. Elles conservent toutes un domaine homologue du domaine catalytique MT-A70 et les acides aminés DPPY caractéristiques des adénines méthyltransférases. Cependant rien pour le moment ne nous permet de conclure que ces enzymes sont impliquées plutôt dans la méthylation des ADN ou des ARN. Il est possible que les enzymes des clades 1 et 2 où figurent déjà des ARN méthyltransférases le soient aussi. J'ai réalisé l'inactivation de l'ensemble des gènes codant ces protéines, seules ou en combinaison mais ne suis pas parvenu à observer de phénotypes ni de létalité, ni sur la méthylation, ni moléculaire sur les réarrangements dans la majorité des cas. Il est prévu d'inactiver l'ensemble de ces gènes d'adénines méthyltransférases putatives en extrayant systématiquement les ARN pour attester de l'efficacité de l'inactivation. Pour déterminer si les protéines codées par ces gènes sont bien nécessaires à la méthylation, nous pourrions détecter 6mA dans les cellules inactivées par HPLC-MS ou à l'aide de notre anticorps spécifique par Immunofluorescence. Lorsque nous aurons également localisé la méthylation dans le génome par séquençage nous pourrions aussi échantillonner et regarder sur des sites précis le niveau de méthylation après inactivation.

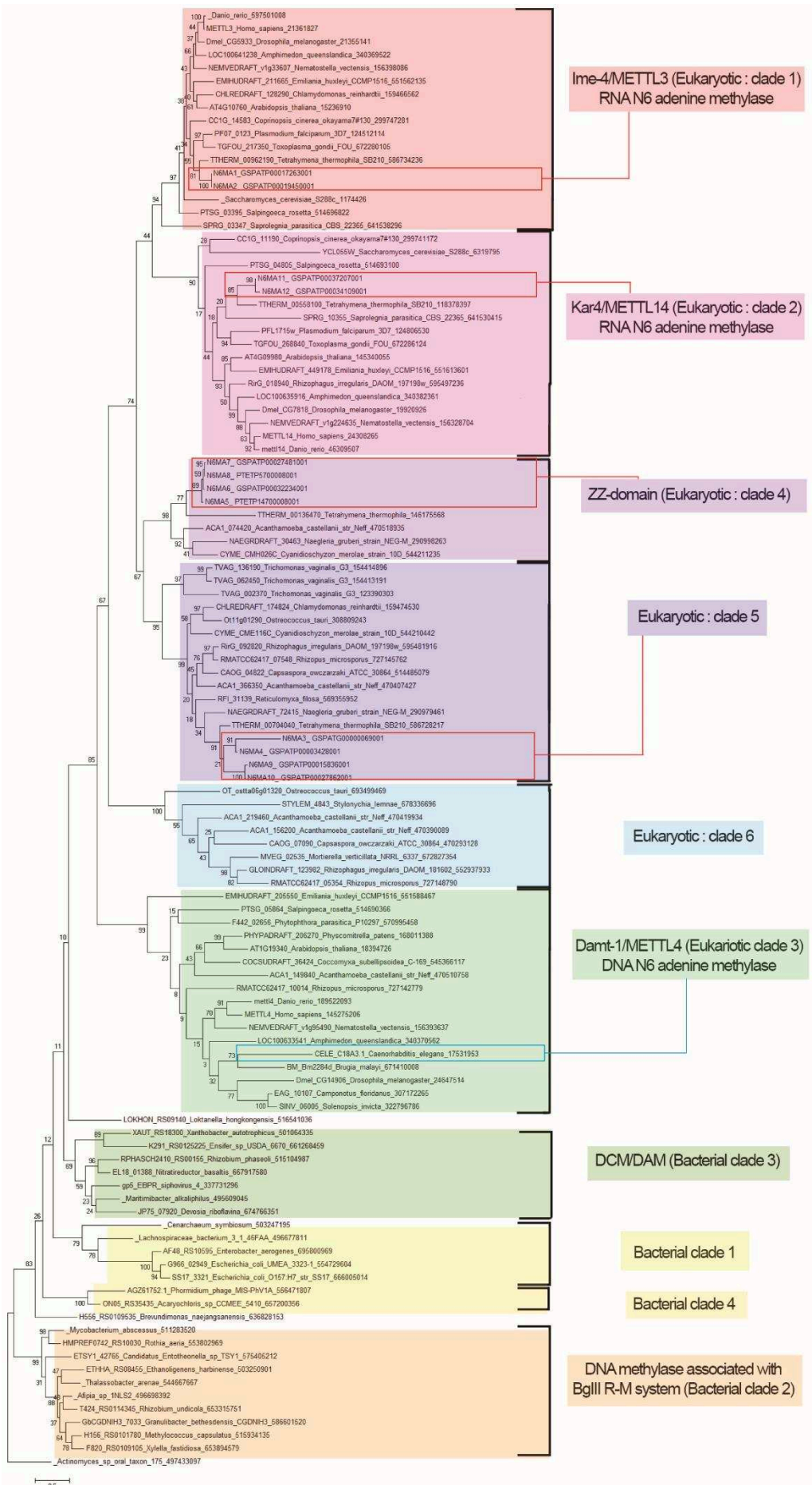


Figure 21 : Analyse phylogénétique des adénines méthyltransférases putatives à domaine MTA70 de *P. tetraurelia*.

Arbre phylogénétique des N6-Adénine Méthyltransférases du domaine MT-A70 obtenu par maximum de vraisemblance à l'aide du programme MEGA6. Les protéines sont indiquées par le nom des espèces et du numéro d'accèsion, et colorées en fonction des différents groupes. Les valeurs de bootstrap sont indiquées à chaque embranchement. Enfin, les groupes bactériens représentent des clades successifs permettant d'enraciner l'arbre. Réalisé par Romain Coppée (Etude phylogénétique des N6-Adénine Méthyltransférases chez l'espèce *Paramecium tetraurelia*, Projet long de master 2).

3. La méthylation des adénines dans les MIC

Par immunofluorescence, l'anticorps anti-6mA a permis de détecter de la méthylation dans les MIC à deux moments du cycle cellulaire : en méiose I et au début du développement macronucléaire. Ces résultats ne viennent pas confirmer la totalité des observations de Cummings où la méthylation est détectée dans les MIC de cellules en cycle végétatif (30 à 40 divisions après l'autogamie) (Cummings et al., 1974). Dans cette étude, les auteurs n'avaient pas quantifié la méthylation pendant les événements sexuels et les MIC extraits proviennent de *P. primaurelia* et non de *P. tetraurelia*. Il nous est difficile de comparer nos résultats avec les leurs. Le génome micronucléaire est transcriptionnellement inactif à l'exception de la méiose où il est intégralement transcrit pour produire les scnARN qui permettront les réarrangements d'ADN pendant le développement (Coyné et al., 2012; Lepère et al., 2009). Le signal 6mA détecté dans les MIC à ce moment du cycle pourrait indiquer que la méthylation des adénines est liée à la transcription des scnARN. Nous n'avons par contre pas connaissance d'un marquage particulier d'une autre protéine ou modification de l'ADN ou de la chromatine visible dans les MICs au début du développement MAC. La signification biologique de la méthylation dans les MICs à ce moment du cycle reste encore mystérieuse.

Il est possible que le protocole d'immunofluorescence ne permette pas à l'anticorps anti-6mA d'accéder au génome MIC à tout moment du cycle cellulaire. Le signal que nous détectons pour la méthylation des adénines en immunofluorescence est dépendant de la sensibilité de la technique et de la capacité de l'anticorps à accéder aux différents génomes MIC et MAC. Les fenêtres de temps pendant lesquelles nous détectons de la méthylation des adénines dans les MIC pourraient être liées à une plus grande accessibilité de l'ADN micronucléaire à l'anticorps anti-6mA. Il n'est pas impossible que bien que nous ne détectons pas de 6mA dans les MICs à d'autres moments du cycle de vie, la modification y soit quand même présente.

4. Méthylation des adénines et réarrangements du génome

L'une de nos hypothèses de travail proposait que la méthylation des adénines chez *P. tetraurelia* soit liée aux réarrangements d'ADN. Nous avons observé que la méthylation s'établissait dans les MAC en développement sans être sûrs que cela se fasse au moment où se produisent les réarrangements du génome. Pour répondre à notre hypothèse, nous avons localisé par immunofluorescence 6mA et l'endonucléase Pgm avec des anticorps spécifiques (Dubois et al., 2017). Nous en avons conclu que la méthylation des adénines apparaissait tardivement après que Pgm a disparu des MAC en développement. L'établissement de la méthylation serait donc postérieur aux réarrangements ce qui ne va pas dans le sens de notre hypothèse. L'augmentation progressive du signal 6mA dans les noyaux pourrait par contre refléter les cycles d'endoréplication. Ces résultats dépendent bien sûr de la sensibilité de l'anticorps et il est tout à fait possible que le peu de méthylation déposée précocement dans les MAC en développement dont la ploïdie est encore faible suffise pour aider aux réarrangements. Il devient important de localiser la méthylation des adénines dans le génome de *P. tetraurelia* par SMRT sequencing afin de voir si comme nous l'imaginons elle se localise sur des motifs particuliers comme sur les bornes des séquences germinales par exemple. Une fois que les adénine méthyltransférases de *P. tetraurelia* seront identifiées, il sera aussi possible de les inactiver pendant les réarrangements et d'étudier l'effet de leur inactivation sur l'élimination d'ADN.

III. Localiser dans le génome de *P. tetraurelia* la méthylation des adénines

1. Adaptation du protocole de MedIP

Le MedIP permet de localiser avec une résolution de l'ordre de 200pb une méthylation de l'ADN dont nous possédons un anticorps spécifique. Pour localiser 6mA par cette technique dans le génome de *P. tetraurelia*, nous nous sommes inspirés de l'analyse de la méthylation des adénines chez les eucaryotes *C. elegans* et *C. reinhardtii* (Fu et al., 2015; Greer et al., 2015). Dans ces études et en particulier chez *C. elegans* qui utilise le même anticorps commercial contre 6mA que nous, les auteurs se sont basés sur un protocole de

MeRIP, immunoprécipitation d'ARN méthylés 6mA que nous avons adapté à l'ADN de *P. tetraurelia* (Dominissini et al., 2015; Greer et al., 2015). Comme pour les autres expériences visant à détecter la méthylation des adénines chez *P. tetraurelia*, il nous a fallu nous affranchir de la contamination bactérienne riche en 6mA et de la contamination en ARN. La première tentative de MedIP que nous avons réalisée utilisait de l'ADN extrait en croissance végétative. Cependant la contamination bactérienne était beaucoup trop forte pour couvrir suffisamment le génome de *P. tetraurelia*. Nous avons donc décidé d'utiliser de l'ADN extrait en toute fin d'autogamie (T50) pour s'affranchir de la contamination bactérienne contenue dans les vacuoles digestives de cellules en croissance végétative. Les analyses que nous avons réalisées après séquençage de l'immunoprécipitation 6mA n'ont cependant pas permis d'identifier de zones du génome enrichies ou déplétées en 6mA. Nous avons utilisé comme contrôle une immunoprécipitation par l'anticorps anti IgG qui ne permettait pas d'immunoprécipiter d'ADN comme c'est le cas avec l'anticorps anti-6mA. Ces résultats semblaient indiquer que l'anticorps anti 6mA ciblait des séquences spécifiques dans le génome. Pour autant, les séquences obtenues après séquençage de l'input et de l'immunoprécipitation 6mA sont très similaires : couvertures comparables, pourcentage d'AT comparables et composition similaire. Plusieurs hypothèses peuvent expliquer ces résultats. Il est tout à fait possible que la méthylation soit répartie de manière très homogène sur le génome, présente en faible quantité mais parsemée partout. De fait, la résolution du MedIP de 200 pb ne permettrait pas de résoudre la localisation de 6mA et d'identifier de zones enrichies ou déplétées en 6mA. Dans ce cas il faut localiser 6mA dans le génome de *P. tetraurelia* par une autre méthode comme le SMRT sequencing par Pacific Bioscience. Finalement il est aussi possible que l'anticorps anti-6mA ne soit pas réellement spécifique de 6mA et pourrait par exemple reconnaître de manière générale dans le génome de *P. tetraurelia* les adénines y compris non méthylées. Cela expliquerait que l'immunoprécipitation 6mA produise des données hautement similaires à celles de l'input. Pour répondre à cette question il serait intéressant de reproduire les expériences de MedIP en introduisant comme contrôle positif des oligonucléotides méthylés sur les adénines dont nous sommes capables de contrôler la quantité par qPCR par exemple. Nous pourrions ainsi faire varier la quantité de ces séquences et contrôler que la technique de MedIP fonctionne bien et que l'anticorps permet bien d'immunoprécipiter l'ADN modifié par 6mA.

2. Localiser 6mA par SMRT sequencing

Nous avons préféré ne pas poursuivre l'analyse de 6mA par MedIP car nous avons entrepris d'analyser la localisation de 6mA par la technique de SMRT sequencing développée par Pacific Bioscience. Nous avons décidé d'utiliser le même ADN que celui sur lequel nous souhaitons faire le MedIP, de l'ADN à T50 en toute fin d'autogamie. Nous espérons obtenir une localisation au nucléotide près et également identifier des motifs et potentiellement voir si 6mA marque les éléments MIC spécifiques ou du moins si sa localisation semble corrélée à la localisation d'éléments MIC spécifiques. Les analyses viennent de débiter mais les résultats préliminaires confirment la présence de 6mA dans le génome de *P. tetraurelia* et indique un pourcentage d'environ 1,41% de 6mA/A. Il semble que la méthylation soit retrouvée essentiellement sur des motifs AT et en particulier dans les portions codantes du génome. Nous avons pour objectif de comparer ces données de séquençage à l'annotation du génome de *P. tetraurelia* et aux données concernant la localisation des nucléosomes par Mnase-seq. On imagine que pendant le développement 6mA pourrait marquer spécifiquement les séquences à éliminer et en particulier les petites IES qui ne dépendent ni des scnARNs ni d'*EZL1* pour leur élimination (Lhuillier-Akakpo et al., 2014). Ces marques guideraient alors la machinerie d'excision. Pour autant étant donné que 6mA est présent pendant tout le cycle de vie et dans tous les noyaux il est tout à fait possible que la méthylation ait d'autres fonctions chez *P. tetraurelia*. Il sera ainsi intéressant de voir si sa localisation est en lien avec le positionnement des nucléosomes comme chez *T. thermophila*, les sites d'initiation de la transcription chez *C. reinhardtii* ou encore la présence de transposons comme chez *D. melanogaster* (Fu et al., 2015; Greer et al., 2015; Wang et al., 2017; Zhang et al., 2015). L'abondance de la méthylation des adénines chez *P. tetraurelia* en fait un parfait modèle d'étude de cette modification chez les eucaryotes. Grâce à cette méthode on est capable de détecter et localiser la méthylation qui peut être analysée en lien avec l'annotation du génome et le positionnement des nucléosomes. Cette première expérience de SMRT seq que nous sommes en train d'analyser est la preuve de principe que la technique marche. Elle pourra être adaptée pour d'autres conditions physiologiques comme pendant les réarrangements du génome.

Bibliographie

- Abe, T., Sugimura, K., Hosono, Y., Takami, Y., Akita, M., Yoshimura, A., Tada, S., Nakayama, T., Murofushi, H., Okumura, K., et al. (2011). The histone chaperone facilitates chromatin transcription (FACT) protein maintains normal replication fork rates. *J. Biol. Chem.* 286, 30504–30512.
- Adam, S., Dabin, J., and Polo, S.E. (2015). Chromatin plasticity in response to DNA damage: The shape of things to come. *DNA Repair* 32, 120–126.
- Adams, R.L., McKay, E.L., Craig, L.M., and Burdon, R.H. (1979). Methylation of mosquito DNA. *Biochim. Biophys. Acta* 563, 72–81.
- Allen, S.E., Hug, I., Pabian, S., Rzeszutek, I., Hoehener, C., and Nowacki, M. (2017). Circular Concatemers of Ultra-Short DNA Segments Produce Regulatory RNAs. *Cell* 168, 990–999.e7.
- Ambrosi, C., Manzo, M., and Baubec, T. (2017). Dynamics and context-dependent roles of DNA methylation. *J. Mol. Biol.*
- Are, V.N., Ghosh, B., Kumar, A., Gadre, R., and Makde, R.D. (2016). Crystal structure and dynamics of Spt16 N-domain of FACT complex from *Cicer arietinum*. *Int. J. Biol. Macromol.*
- Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.-M., Wilkes, C.D., Garnier, O., Labadie, K., Lauderdale, B.E., Le Mouël, A., et al. (2012). The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet.* 8, e1002984.
- Arnaiz, O., Van Dijk, E., Bétermier, M., Lhuillier-Akakpo, M., de Vanssay, A., Duharcourt, S., Sallet, E., Gouzy, J., and Sperling, L. (2017). Improved methods and resources for *paramecium* genomics: transcription units, gene annotation and gene expression. *BMC Genomics* 18, 483.
- Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., et al. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444, 171–178.
- Baranyi, U., Klein, R., Lubitz, W., Krüger, D.H., and Witte, A. (2000). The archaeal halophilic virus-encoded Dam-like methyltransferase M. phiCh1-I methylates adenine residues and complements dam mutants in the low salt environment of *Escherichia coli*. *Mol. Microbiol.* 35, 1168–1179.
- Baudry, C., Malinsky, S., Restituto, M., Kapusta, A., Rosa, S., Meyer, E., and Bétermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev.* 23, 2478–2483.
- Beaulaurier, J., Zhang, X.-S., Zhu, S., Sebra, R., Rosenbluh, C., Deikus, G., Shen, N., Munera, D., Waldor, M.K., Chess, A., et al. (2015). Single molecule-level detection and long read-based phasing of epigenetic variations in bacterial methylomes. *Nat. Commun.* 6, 7438.

- Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., Jr, et al. (2010). *Paramecium tetraurelia*: the renaissance of an early unicellular model. *Cold Spring Harb. Protoc.* 2010, pdb.emo140.
- Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. *Science* 301, 1090–1093.
- Berger, J.D. (1973). Nuclear differentiation and nucleic acid synthesis in well-fed exconjugants of *Paramecium aurelia*. *Chromosoma* 42, 247–268.
- Betermier, M., and Duharcourt, S. (2014). Programmed Rearrangement in Ciliates: *Paramecium*. *Microbiol. Spectr.* 2.
- Bétermier, M., Duharcourt, S., Seitz, H., and Meyer, E. (2000). Timing of developmentally programmed excision and circularization of *Paramecium* internal eliminated sequences. *Mol. Cell. Biol.* 20, 1553–1561.
- Biswas, D., Yu, Y., Prall, M., Formosa, T., and Stillman, D.J. (2005). The yeast FACT complex has a role in transcriptional initiation. *Mol. Cell. Biol.* 25, 5812–5822.
- Biswas, D., Dutta-Biswas, R., Mitra, D., Shibata, Y., Strahl, B.D., Formosa, T., and Stillman, D.J. (2006). Opposing roles for Set2 and yFACT in regulating TBP binding at promoters. *EMBO J.* 25, 4479–4489.
- Bouhouche, K., Gout, J.-F., Kapusta, A., Bétermier, M., and Meyer, E. (2011). Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res.* 39, 4249–4264.
- Breiling, A., and Lyko, F. (2015). Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* 8, 24.
- Brewster, N.K., Johnston, G.C., and Singer, R.A. (1998). Characterization of the CP Complex, an Abundant Dimer of Cdc68 and Pob3 Proteins That Regulates Yeast Transcriptional Activation and Chromatin Repression. *J. Biol. Chem.* 273, 21972–21979.
- Brewster, N.K., Johnston, G.C., and Singer, R.A. (2001). A Bipartite Yeast SSRP1 Analog Comprised of Pob3 and Nhp6 Proteins Modulates Transcription. *Mol. Cell. Biol.* 21, 3491–3502.
- Brygoo, Y., Sonneborn, T.M., Keller, A.M., Dippell, R.V., and Schneller, M.V. (1980). Genetic Analysis of Mating Type Differentiation in *PARAMECIUM TETRAURELIA*. II. Role of the Micronuclei in Mating-Type Determination. *Genetics* 94, 951–959.
- Capowski, E.E., Wells, J.M., Harrison, G.S., and Karrer, K.M. (1989). Molecular analysis of N6-methyladenine patterns in *Tetrahymena thermophila* nuclear DNA. *Mol. Cell. Biol.* 9, 2598–2605.
- Capuano, F., Mülleder, M., Kok, R., Blom, H.J., and Ralser, M. (2014). Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. *Anal. Chem.* 86, 3697–3702.

- Casadesús, J., and Low, D. (2006). Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.* MMBR 70, 830–856.
- Chen, K., Zhao, B.S., and He, C. (2016). Nucleic Acid Modifications in Regulation of Gene Expression. *Cell Chem. Biol.* 23, 74–85.
- Cheng, C.-Y., Vogt, A., Mochizuki, K., and Yao, M.-C. (2010). A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in *Tetrahymena thermophila*. *Mol. Biol. Cell* 21, 1753–1762.
- Clark-Adams, C.D., Norris, D., Osley, M.A., Fassler, J.S., and Winston, F. (1988). Changes in histone gene dosage alter transcription in yeast. *Genes Dev.* 2, 150–159.
- Corpet, A., and Almouzni, G. (2009). Making copies of chromatin: the challenge of nucleosomal organization and epigenetic information. *Trends Cell Biol.* 19, 29–41.
- Coyne, R.S., Lhuillier-Akakpo, M., and Duharcourt, S. (2012). RNA-guided DNA rearrangements in ciliates: is the best genome defence a good offence? *Biol. Cell Auspices Eur. Cell Biol. Organ.* 104, 309–325.
- Cummings, D.J., Tait, A., and Goddard, J.M. (1974). Methylated bases in DNA from *Paramecium aurelia*. *Biochim. Biophys. Acta* 374, 1–11.
- Derelle, R., and Lang, B.F. (2012). Rooting the eukaryotic tree with mitochondrial and bacterial proteins. *Mol. Biol. Evol.* 29, 1277–1289.
- Dinant, C., Ampatzidis-Michailidis, G., Lans, H., Tresini, M., Lagarou, A., Grosbart, M., Theil, A.F., van Cappellen, W.A., Kimura, H., Bartek, J., et al. (2013). Enhanced chromatin dynamics by FACT promotes transcriptional restart after UV-induced DNA damage. *Mol. Cell* 51, 469–479.
- Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature* 485, 201–206.
- Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N., and Rechavi, G. (2013). Transcriptome-wide mapping of N(6)-methyladenosine by m(6)A-seq based on immunocapturing and massively parallel sequencing. *Nat. Protoc.* 8, 176–189.
- Dominissini, D., Moshitch-Moshkovitz, S., Amariglio, N., and Rechavi, G. (2015). Transcriptome-Wide Mapping of N⁶-Methyladenosine by m⁶A-Seq. *Methods Enzymol.* 560, 131–147.
- Drabløs, F., Feyzi, E., Aas, P.A., Vaagbø, C.B., Kavli, B., Bratlie, M.S., Peña-Díaz, J., Otterlei, M., Slupphaug, G., and Krokan, H.E. (2004). Alkylation damage in DNA and RNA--repair mechanisms and medical significance. *DNA Repair* 3, 1389–1407.
- Dubois, E., Mathy, N., Régnier, V., Bischerour, J., Baudry, C., Trouslard, R., and Bétermier, M. (2017). Multimerization properties of PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements. *Nucleic Acids Res.*

Duharcourt, S., Butler, A., and Meyer, E. (1995). Epigenetic self-regulation of developmental excision of an internal eliminated sequence on *Paramecium tetraurelia*. *Genes Dev.* 9, 2065–2077.

Duharcourt, S., Keller, A.M., and Meyer, E. (1998). Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. *Mol. Cell. Biol.* 18, 7075–7085.

Dunn, D.B., and Smith, J.D. (1958). The occurrence of 6-methylaminopurine in deoxyribonucleic acids. *Biochem. J.* 68, 627–636.

Edwards, J.R., Yarychivska, O., Boulard, M., and Bestor, T.H. (2017). DNA methylation and DNA methyltransferases. *Epigenetics Chromatin* 10, 23.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138.

Epstein, L.M., and Forney, J.D. (1984). Mendelian and non-mendelian mutations affecting surface antigen expression in *Paramecium tetraurelia*. *Mol. Cell. Biol.* 4, 1583–1590.

Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M.C., Jabado, O.J., et al. (2012). Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol.* 30, 1232–1239.

Feng, J., Gan, H., Eaton, M.L., Zhou, H., Li, S., Belsky, J.A., MacAlpine, D.M., Zhang, Z., and Li, Q. (2016a). Non-coding transcription is a driving force for nucleosome instability in *spt16* mutant cells. *Mol. Cell. Biol.*

Feng, J., Gan, H., Eaton, M.L., Zhou, H., Li, S., Belsky, J.A., MacAlpine, D.M., Zhang, Z., and Li, Q. (2016b). Noncoding Transcription Is a Driving Force for Nucleosome Instability in *spt16* Mutant Cells. *Mol. Cell. Biol.* 36, 1856–1867.

Fleming, A.B., Kao, C.-F., Hillyer, C., Pikaart, M., and Osley, M.A. (2008). H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. *Mol. Cell* 31, 57–66.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465.

Formosa, T. (2008). FACT and the reorganized nucleosome. *Mol. Biosyst.* 4, 1085–1093.

Formosa, T. (2012). The role of FACT in making and breaking nucleosomes. *Biochim. Biophys. Acta* 1819, 247–255.

Formosa, T., Eriksson, P., Wittmeyer, J., Ginn, J., Yu, Y., and Stillman, D.J. (2001). Spt16-Pob3 and the HMG protein Nhp6 combine to form the nucleosome-binding factor SPN. *EMBO J.* 20, 3506–3517.

- Formosa, T., Ruone, S., Adams, M.D., Olsen, A.E., Eriksson, P., Yu, Y., Rhoades, A.R., Kaufman, P.D., and Stillman, D.J. (2002). Defects in SPT16 or POB3 (yFACT) in *Saccharomyces cerevisiae* cause dependence on the Hir/Hpc pathway: polymerase passage may degrade chromatin structure. *Genetics* 162, 1557–1571.
- Forney, J.D., and Blackburn, E.H. (1988). Developmentally controlled telomere addition in wild-type and mutant paramecia. *Mol. Cell. Biol.* 8, 251–258.
- Frapporti, A., Arnaiz, O., Miro Pina, C., Eleftheriou, E., Holoch, D., Sperling, L., Margueron, R., and Duharcourt, S. (In submission). A Polycomb protein dually methylates H3 lysine 9 and 27 to mediate transposable element repression.
- Fu, Y., Luo, G.-Z., Chen, K., Deng, X., Yu, M., Han, D., Hao, Z., Liu, J., Lu, X., Doré, L.C., et al. (2015). N6-Methyldeoxyadenosine Marks Active Transcription Start Sites in *Chlamydomonas*. *Cell* 161, 879–892.
- Furrer, D.I., Swart, E.C., Kraft, M.F., Sandoval, P.Y., and Nowacki, M. (2017). Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA. *Cell Rep.* 20, 505–520.
- Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D., and Labib, K. (2006). GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat. Cell Biol.* 8, 358–366.
- Gao, Y., Li, C., Wei, L., Teng, Y., Nakajima, S., Chen, X., Xu, J., Legar, B., Ma, H., Spagnol, S.T., et al. (2017). SSRP1 Cooperates with PARP and XRCC1 to Facilitate Single-Strand DNA Break Repair by Chromatin Priming. *Cancer Res.* 77, 2674–2685.
- Garnier, O., Serrano, V., Duharcourt, S., and Meyer, E. (2004). RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. *Mol. Cell. Biol.* 24, 7370–7379.
- Gorovsky, M.A., Hattman, S., and Pleger, G.L. (1973). (6 N)methyl adenine in the nuclear DNA of a eucaryote, *Tetrahymena pyriformis*. *J. Cell Biol.* 56, 697–701.
- Grandchamp, S., and Beisson, J. (1981). Positional control of nuclear differentiation in *Paramecium*. *Dev. Biol.* 81, 336–341.
- Gratias, A., and Bétermier, M. (2003). Processing of double-strand breaks is involved in the precise excision of paramecium internal eliminated sequences. *Mol. Cell. Biol.* 23, 7152–7162.
- Gratias, A., Lepère, G., Garnier, O., Rosa, S., Duharcourt, S., Malinsky, S., Meyer, E., and Bétermier, M. (2008). Developmentally programmed DNA splicing in *Paramecium* reveals short-distance crosstalk between DNA cleavage sites. *Nucleic Acids Res.* 36, 3244–3251.
- Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corrales, D., Hsu, C.-H., Aravind, L., He, C., and Shi, Y. (2015). DNA Methylation on N6-Adenine in *C. elegans*. *Cell* 161, 868–878.
- Grosjean, H. (2013). Nucleic Acids Are Not Boring Long Polymers of Only Four Types of Nucleotides: A Guided Tour (Landes Bioscience).

- Gruchota, J., Denby Wilkes, C., Arnaiz, O., Sperling, L., and Nowak, J.K. (2017). A meiosis-specific Spt5 homolog involved in non-coding transcription. *Nucleic Acids Res.*
- Guérin, F., Arnaiz, O., Boggetto, N., Denby Wilkes, C., Meyer, E., Sperling, L., and Duharcourt, S. (2017). Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements. *BMC Genomics* 18, 327.
- Hammond, C.M., Strømme, C.B., Huang, H., Patel, D.J., and Groth, A. (2017). Histone chaperone networks shaping chromatin function. *Nat. Rev. Mol. Cell Biol.* 18, 141–158.
- Han, J., Li, Q., McCullough, L., Kettelkamp, C., Formosa, T., and Zhang, Z. (2010). Ubiquitylation of FACT by the cullin-E3 ligase Rtt101 connects FACT to DNA replication. *Genes Dev.* 24, 1485–1490.
- Harrison, G.S., and Karrer, K.M. (1989). Methylation of replicating and nonreplicating DNA in the ciliate *Tetrahymena thermophila*. *Mol. Cell. Biol.* 9, 828–830.
- Harrison, G.S., Findly, R.C., and Karrer, K.M. (1986). Site-specific methylation of adenine in the nuclear genome of a eucaryote, *Tetrahymena thermophila*. *Mol. Cell. Biol.* 6, 2364–2370.
- Hattman, S., Kenny, C., Berger, L., and Pratt, K. (1978). Comparative study of DNA methylation in three unicellular eucaryotes. *J. Bacteriol.* 135, 1156–1157.
- Heo, K., Kim, H., Choi, S.H., Choi, J., Kim, K., Gu, J., Lieber, M.R., Yang, A.S., and An, W. (2008). FACT-mediated exchange of histone variant H2AX regulated by phosphorylation of H2AX and ADP-ribosylation of Spt16. *Mol. Cell* 30, 86–97.
- Hoehener, C., Hug, I., and Nowacki, M. (2018). Dicer-like Enzymes with Sequence Cleavage Preferences. *Cell* 173, 234–247.e7.
- Hoffmann, C., and Neumann, H. (2015). In Vivo Mapping of FACT-Histone Interactions Identifies a Role of Pob3 C-terminus in H2A-H2B Binding. *ACS Chem. Biol.* 10, 2753–2763.
- Hondele, M., and Ladurner, A.G. (2013). Catch me if you can: how the histone chaperone FACT capitalizes on nucleosome breathing. *Nucl. Austin Tex* 4, 443–449.
- Hondele, M., Stuwe, T., Hassler, M., Halbach, F., Bowman, A., Zhang, E.T., Nijmeijer, B., Kotthoff, C., Rybin, V., Amlacher, S., et al. (2013). Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. *Nature* 499, 111–114.
- Hotchkiss, R.D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.* 175, 315–332.
- Hsieh, F.-K., Fisher, M., Ujvári, A., Studitsky, V.M., and Luse, D.S. (2010). Histone Sin mutations promote nucleosome traversal and histone displacement by RNA polymerase II. *EMBO Rep.* 11, 705–710.
- Hsieh, F.-K., Kulaeva, O.I., Patel, S.S., Dyer, P.N., Luger, K., Reinberg, D., and Studitsky, V.M. (2013). Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* 110, 7654–7659.

Hsu, H.-K., Weng, Y.-I., Hsu, P.-Y., Huang, T.H.-M., and Huang, Y.-W. (2014). Detection of DNA methylation by MeDIP and MBDCap assays: an overview of techniques. *Methods Mol. Biol.* Clifton NJ 1105, 61–70.

Huang, W., Xiong, J., Yang, Y., Liu, S.-M., Yuan, B.-F., and Feng, Y.-Q. (2015). Determination of DNA adenine methylation in genomes of mammals and plants by liquid chromatography/mass spectrometry. *RSC Adv.* 5, 64046–64054.

Ignarski, M., Singh, A., Swart, E.C., Arambasic, M., Sandoval, P.Y., and Nowacki, M. (2014). Paramecium tetraurelia chromatin assembly factor-1-like protein PtCAF-1 is involved in RNA-mediated control of DNA elimination. *Nucleic Acids Res.* 42, 11952–11964.

Iyer, L.M., Abhiman, S., and Aravind, L. (2011). Natural history of eukaryotic DNA methylation systems. *Prog. Mol. Biol. Transl. Sci.* 101, 25–104.

Iyer, L.M., Zhang, D., and Aravind, L. (2015). Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. *BioEssays News Rev. Mol. Cell. Dev. Biol.*

Jamai, A., Puglisi, A., and Strubin, M. (2009). Histone chaperone spt16 promotes redeposition of the original h3-h4 histones evicted by elongating RNA polymerase. *Mol. Cell* 35, 377–383.

Janulaitis, A., Klimasauskas, S., Petrusyte, M., and Butkus, V. (1983). Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. *FEBS Lett.* 161, 131–134.

Jeltsch, A., Christ, F., Fatemi, M., and Roth, M. (1999). On the substrate specificity of DNA methyltransferases. adenine-N6 DNA methyltransferases also modify cytosine residues at position N4. *J. Biol. Chem.* 274, 19538–19544.

Jeronimo, C., and Robert, F. (2016). Histone chaperones FACT and Spt6 prevent histone variants from turning into histone deviants. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 38, 420–426.

Jeronimo, C., Watanabe, S., Kaplan, C.D., Peterson, C.L., and Robert, F. (2015). The Histone Chaperones FACT and Spt6 Restrict H2A.Z from Intragenic Locations. *Mol. Cell* 58, 1113–1123.

Jessop-Murray, H., Martin, L.D., Gilley, D., Preer, J.R., and Polisky, B. (1991). Permanent rescue of a non-Mendelian mutation of Paramecium by microinjection of specific DNA sequences. *Genetics* 129, 727–734.

Johri, P., Krenek, S., Marinov, K., Doak, T.G., Berendonk, T.U., and Lynch, M. (2017). Population genomics of Paramecium species. *Mol. Biol. Evol.*

Jurkowska, R.Z., and Jeltsch, A. (2016). Mechanisms and Biological Roles of DNA Methyltransferases and DNA Methylation: From Past Achievements to Future Challenges. *Adv. Exp. Med. Biol.* 945, 1–17.

Kahramanoglou, C., Prieto, A.I., Khedkar, S., Haase, B., Gupta, A., Benes, V., Fraser, G.M., Luscombe, N.M., and Seshasayee, A.S.N. (2012). Genomics of DNA cytosine methylation in Escherichia coli reveals its role in stationary phase transcription. *Nat. Commun.* 3, 886.

- Kanvah, S., Joseph, J., Schuster, G.B., Barnett, R.N., Cleveland, C.L., and Landman, U. (2010). Oxidation of DNA: Damage to Nucleobases. *Acc. Chem. Res.* 43, 280–287.
- Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J.D., Malinsky, S., and Bétermier, M. (2011). Highly precise and developmentally programmed genome assembly in *Paramecium* requires ligase IV-dependent end joining. *PLoS Genet.* 7, e1002049.
- Karrer, K.M., and VanNuland, T.A. (2002). Methylation of adenine in the nuclear DNA of *Tetrahymena* is internucleosomal and independent of histone H1. *Nucleic Acids Res.* 30, 1364–1370.
- Keller, D.M., and Lu, H. (2002). p53 serine 392 phosphorylation increases after UV through induction of the assembly of the CK2.hSPT16.SSRP1 complex. *J. Biol. Chem.* 277, 50206–50213.
- Kemble, D.J., Whitby, F.G., Robinson, H., McCullough, L.L., Formosa, T., and Hill, C.P. (2013). Structure of the Spt16 middle domain reveals functional features of the histone chaperone FACT. *J. Biol. Chem.* 288, 10188–10194.
- Kemble, D.J., McCullough, L.L., Whitby, F.G., Formosa, T., and Hill, C.P. (2015). FACT Disrupts Nucleosome Structure by Binding H2A-H2B with Conserved Peptide Motifs. *Mol. Cell* 60, 294–306.
- Klobutcher, L.A., and Herrick, G. (1997). Developmental genome reorganization in ciliated protozoa: the transposon link. *Prog. Nucleic Acid Res. Mol. Biol.* 56, 1–62.
- Koizumi, S., and Kobayashi, S. (1989). Microinjection of plasmid DNA encoding the A surface antigen of *Paramecium tetraurelia* restores the ability to regenerate a wild-type macronucleus. *Mol. Cell. Biol.* 9, 4398–4401.
- Koziol, M.J., Bradshaw, C.R., Allen, G.E., Costa, A.S.H., Frezza, C., and Gurdon, J.B. (2016). Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* 23, 24–30.
- Kumar, R., Mukhopadhyay, A.K., Ghosh, P., and Rao, D.N. (2012). Comparative transcriptomics of *H. pylori* strains AM5, SS1 and their hpyAVIBM deletion mutants: possible roles of cytosine methylation. *PLoS One* 7, e42303.
- Kumar, S., Karmakar, B.C., Nagarajan, D., Mukhopadhyay, A.K., Morgan, R.D., and Rao, D.N. (2018). N4-cytosine DNA methylation regulates transcription and pathogenesis in *Helicobacter pylori*. *Nucleic Acids Res.*
- Kwon, S.H., Florens, L., Swanson, S.K., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2010). Heterochromatin protein 1 (HP1) connects the FACT histone chaperone complex to the phosphorylated CTD of RNA polymerase II. *Genes Dev.* 24, 2133–2145.
- Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I., and Saenger, W. (1994). Three-dimensional structure of the adenine-specific DNA methyltransferase M.Taq I in complex with the cofactor S-adenosylmethionine. *Proc. Natl. Acad. Sci. U. S. A.* 91, 10957–10961.

- Laskey, R.A., Honda, B.M., Mills, A.D., and Finch, J.T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* 275, 416–420.
- Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220.
- Lejeune, E., Bortfeld, M., White, S.A., Pidoux, A.L., Ekwall, K., Allshire, R.C., and Ladurner, A.G. (2007). The chromatin-remodeling factor FACT contributes to centromeric heterochromatin independently of RNAi. *Curr. Biol. CB* 17, 1219–1224.
- Lepère, G., Bétermier, M., Meyer, E., and Duharcourt, S. (2008). Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev.* 22, 1501–1512.
- Lepère, G., Nowacki, M., Serrano, V., Gout, J.-F., Guglielmi, G., Duharcourt, S., and Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res.* 37, 903–915.
- Lhuillier-Akakpo, M., Frapporti, A., Denby Wilkes, C., Matelot, M., Vervoort, M., Sperling, L., and Duharcourt, S. (2014). Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements. *PLoS Genet.* 10, e1004665.
- Li, Y., Zeng, S.X., Landais, I., and Lu, H. (2007). Human SSRP1 has Spt16-dependent and -independent roles in gene transcription. *J. Biol. Chem.* 282, 6936–6945.
- Liang, Z., Shen, L., Cui, X., Bao, S., Geng, Y., Yu, G., Liang, F., Xie, S., Lu, T., Gu, X., et al. (2018a). DNA N6-Adenine Methylation in *Arabidopsis thaliana*. *Dev. Cell* 45, 406–416.e3.
- Liang, Z., Yu, G., Liu, J., Geng, Y., Mao, J., Wang, D., Zhou, J., and Gu, X. (2018b). The N6-adenine methylation in yeast genome profiled by single-molecule technology. *J. Genet. Genomics Yi Chuan Xue Bao*.
- Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 79, 181–211.
- Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A., Wilson, G.G., and Murray, N.E. (2014). Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.* 42, 3–19.
- Lu, M., Campbell, J.L., Boye, E., and Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* 77, 413–426.
- Maliszewska-Olejniczak, K., Gruchota, J., Gromadka, R., Denby Wilkes, C., Arnaiz, O., Mathy, N., Duharcourt, S., Bétermier, M., and Nowak, J.K. (2015). TFIIIS-Dependent Non-coding Transcription Regulates Developmental Genome Rearrangements. *PLoS Genet.* 11, e1005383.
- Malone, C.D., Anderson, A.M., Motl, J.A., Rexer, C.H., and Chalker, D.L. (2005). Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of *Tetrahymena thermophila*. *Mol. Cell. Biol.* 25, 9151–9164.

- Malone, E.A., Clark, C.D., Chiang, A., and Winston, F. (1991). Mutations in SPT16/CDC68 suppress cis- and trans-acting mutations that affect promoter function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 11, 5710–5717.
- Malone, T., Blumenthal, R.M., and Cheng, X. (1995). Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* 253, 618–632.
- Marmignon, A., Bischerour, J., Silve, A., Fojcik, C., Dubois, E., Arnaiz, O., Kapusta, A., Malinsky, S., and Bétermier, M. (2014). Ku-mediated coupling of DNA cleavage and repair during programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *PLoS Genet.* 10, e1004552.
- Martienssen, R., and Moazed, D. (2015). RNAi and heterochromatin assembly. *Cold Spring Harb. Perspect. Biol.* 7, a019323.
- Mason, P.B., and Struhl, K. (2003). The FACT Complex Travels with Elongating RNA Polymerase II and Is Important for the Fidelity of Transcriptional Initiation In Vivo. *Mol. Cell. Biol.* 23, 8323–8333.
- Matsuda, A., Mayer, K.M., and Forney, J.D. (2004). Identification of single nucleotide mutations that prevent developmentally programmed DNA elimination in *Paramecium tetraurelia*. *J. Eukaryot. Microbiol.* 51, 664–669.
- Matysiak, J., Lesbats, P., Mauro, E., Lapaillerie, D., Dupuy, J.-W., Lopez, A.P., Benleulmi, M.S., Calmels, C., Andreola, M.-L., Ruff, M., et al. (2017). Modulation of chromatin structure by the FACT histone chaperone complex regulates HIV-1 integration. *Retrovirology* 14, 39.
- Mayer, K.M., and Forney, J.D. (1999). A mutation in the flanking 5'-TA-3' dinucleotide prevents excision of an internal eliminated sequence from the *Paramecium tetraurelia* genome. *Genetics* 151, 597–604.
- Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* 17, 1272–1278.
- McCullough, L., Rawlins, R., Olsen, A., Xin, H., Stillman, D.J., and Formosa, T. (2011). Insight into the mechanism of nucleosome reorganization from histone mutants that suppress defects in the FACT histone chaperone. *Genetics* 188, 835–846.
- McCullough, L., Connell, Z., Petersen, C., and Formosa, T. (2015). The Abundant Histone Chaperones Spt6 and FACT Collaborate to Assemble, Inspect, and Maintain Chromatin Structure in *Saccharomyces cerevisiae*. *Genetics* 201, 1031–1045.
- McCullough, L.L., Connell, Z., Xin, H., Studitsky, V.M., Feofanov, A.V., Valieva, M.E., and Formosa, T. (2018). Functional roles of the DNA-binding HMGB domain in the histone chaperone FACT in nucleosome reorganization. *J. Biol. Chem.*
- McGrath, C.L., Gout, J.-F., Johri, P., Doak, T.G., and Lynch, M. (2014). Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome Res.* 24, 1665–1675.

- McIntyre, A.B.R., Alexander, N., Burton, A.S., Castro-Wallace, S., Chiu, C.Y., John, K.K., Stahl, S.E., Li, S., and Mason, C.E. (2017). Nanopore detection of bacterial DNA base modifications. *bioRxiv* 127100.
- Militeello, K.T., Mandarano, A.H., Varchtchouk, O., and Simon, R.D. (2014). Cytosine DNA methylation influences drug resistance in *Escherichia coli* through increased *sugE* expression. *FEMS Microbiol. Lett.* 350, 100–106.
- Mitra, R., Fain-Thornton, J., and Craig, N.L. (2008). *piggyBac* can bypass DNA synthesis during cut and paste transposition. *EMBO J.* 27, 1097–1109.
- Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell* 110, 689–699.
- Mondo, S.J., Dannebaum, R.O., Kuo, R.C., Louie, K.B., Bewick, A.J., LaButti, K., Haridas, S., Kuo, A., Salamov, A., Ahrendt, S.R., et al. (2017). Widespread adenine N6-methylation of active genes in fungi. *Nat. Genet.* 49, 964–968.
- Morgan, R.D., Luyten, Y.A., Johnson, S.A., Clough, E.M., Clark, T.A., and Roberts, R.J. (2016). Novel m4C modification in type I restriction-modification systems. *Nucleic Acids Res.* 44, 9413–9425.
- Morillo-Huesca, M., Maya, D., Muñoz-Centeno, M.C., Singh, R.K., Oreal, V., Reddy, G.U., Liang, D., Géli, V., Gunjan, A., and Chávez, S. (2010). FACT prevents the accumulation of free histones evicted from transcribed chromatin and a subsequent cell cycle delay in G1. *PLoS Genet.* 6, e1000964.
- Le Mouël, A., Butler, A., Caron, F., and Meyer, E. (2003). Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. *Eukaryot. Cell* 2, 1076–1090.
- Nakayama, S., and Endoh, H. (2003). Preferential cleavage of *Paramecium* DNA mediated by the *C. elegans* Tc1 transposase in vitro. *Genes Genet. Syst.* 78, 391–398.
- Noto, T., Kurth, H.M., Kataoka, K., Aronica, L., DeSouza, L.V., Siu, K.W.M., Pearlman, R.E., Gorovsky, M.A., and Mochizuki, K. (2010). The *Tetrahymena* argonaute-binding protein *Giw1p* directs a mature argonaute-siRNA complex to the nucleus. *Cell* 140, 692–703.
- Nowacki, M., Zagorski-Ostojka, W., and Meyer, E. (2005). *Nowa1p* and *Nowa2p*: novel putative RNA binding proteins involved in trans-nuclear crosstalk in *Paramecium tetraurelia*. *Curr. Biol. CB* 15, 1616–1628.
- Nowak, J.K., Gromadka, R., Juszczuk, M., Jerka-Dziadosz, M., Maliszewska, K., Mucchielli, M.-H., Gout, J.-F., Arnaiz, O., Agier, N., Tang, T., et al. (2011). Functional study of genes essential for autogamy and nuclear reorganization in *Paramecium*. *Eukaryot. Cell* 10, 363–372.
- O’Brown, Z.K., and Greer, E.L. (2016). N6-Methyladenine: A Conserved and Dynamic DNA Mark. *Adv. Exp. Med. Biol.* 945, 213–246.

- O'Donnell, A.F., Brewster, N.K., Kurniawan, J., Minard, L.V., Johnston, G.C., and Singer, R.A. (2004). Domain organization of the yeast histone chaperone FACT: the conserved N-terminal domain of FACT subunit Spt16 mediates recovery from replication stress. *Nucleic Acids Res.* 32, 5894–5906.
- Okuhara, K., Ohta, K., Seo, H., Shioda, M., Yamada, T., Tanaka, Y., Dohmae, N., Seyama, Y., Shibata, T., and Murofushi, H. (1999). A DNA unwinding factor involved in DNA replication in cell-free extracts of *Xenopus* eggs. *Curr. Biol.* 9, 341–351.
- Oliveira, D.V., Kato, A., Nakamura, K., Ikura, T., Okada, M., Kobayashi, J., Yanagihara, H., Saito, Y., Tauchi, H., and Komatsu, K. (2014). Histone chaperone FACT regulates homologous recombination by chromatin remodeling through interaction with RNF20. *J. Cell Sci.* 127, 763–772.
- Orphanides, G., LeRoy, G., Chang, C.H., Luse, D.S., and Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* 92, 105–116.
- Orphanides, G., Wu, W.H., Lane, W.S., Hampsey, M., and Reinberg, D. (1999). The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* 400, 284–288.
- Pathak, R., Singh, P., Ananthakrishnan, S., Adamczyk, S., Schimmel, O., and Govind, C.K. (2018). Acetylation-Dependent Recruitment of the FACT Complex and Its Role in Regulating Pol II Occupancy Genome-Wide in *Saccharomyces cerevisiae*. *Genetics* 209, 743–756.
- Patterson, K., Molloy, L., Qu, W., and Clark, S. (2011). DNA methylation: bisulphite modification and analysis. *J. Vis. Exp.* JoVE.
- Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* 125, 703–717.
- Pfab, A., Breindl, M., and Grasser, K.D. (2018). The *Arabidopsis* histone chaperone FACT is required for stress-induced expression of anthocyanin biosynthetic genes. *Plant Mol. Biol.*
- Plasterk, R.H., Izsvák, Z., and Ivics, Z. (1999). Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet. TIG* 15, 326–332.
- Plongthongkum, N., Diep, D.H., and Zhang, K. (2014). Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat. Rev. Genet.* 15, 647–661.
- Preer, L.B., Hamilton, G., and Preer, J.R., Jr (1992). Micronuclear DNA from *Paramecium tetraurelia*: serotype 51 A gene has internally eliminated sequences. *J. Protozool.* 39, 678–682.
- Prendergast, J.A., Murray, L.E., Rowley, A., Carruthers, D.R., Singer, R.A., and Johnston, G.C. (1990). Size selection identifies new genes that regulate *Saccharomyces cerevisiae* cell proliferation. *Genetics* 124, 81–90.
- Prendergast, L., Müller, S., Liu, Y., Huang, H., Dingli, F., Loew, D., Vassias, I., Patel, D.J., Sullivan, K.F., and Almouzni, G. (2016). The CENP-T/W complex is a binding partner of the histone chaperone FACT. *Genes Dev.* 30, 1313–1326.

- Prescott, D.M. (1994). The DNA of ciliated protozoa. *Microbiol. Rev.* 58, 233–267.
- Proffitt, J.H., Davie, J.R., Swinton, D., and Hattman, S. (1984). 5-Methylcytosine is not detectable in *Saccharomyces cerevisiae* DNA. *Mol. Cell. Biol.* 4, 985–988.
- Ransom, M., Williams, S.K., Dechassa, M.L., Das, C., Linger, J., Adkins, M., Liu, C., Bartholomew, B., and Tyler, J.K. (2009). FACT and the proteasome promote promoter chromatin disassembly and transcriptional initiation. *J. Biol. Chem.* 284, 23461–23471.
- Ransom, M., Dennehey, B.K., and Tyler, J.K. (2010). Chaperoning histones during DNA replication and repair. *Cell* 140, 183–195.
- Ratel, D., Ravanat, J.-L., Berger, F., and Wion, D. (2006). N6-methyladenine: the other methylated base of DNA. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 28, 309–315.
- Rhoades, A.R., Ruone, S., and Formosa, T. (2004). Structural features of nucleosomes reorganized by yeast FACT and its HMG box component, Nhp6. *Mol. Cell. Biol.* 24, 3907–3917.
- Rowley, A., Singer, R.A., and Johnston, G.C. (1991). CDC68, a yeast gene that affects regulation of cell proliferation and transcription, encodes a protein with a highly acidic carboxyl terminus. *Mol. Cell. Biol.* 11, 5718–5726.
- Ruone, S., Rhoades, A.R., and Formosa, T. (2003). Multiple Nhp6 molecules are required to recruit Spt16-Pob3 to form yFACT complexes and to reorganize nucleosomes. *J. Biol. Chem.* 278, 45288–45295.
- Sánchez-Romero, M.A., Cota, I., and Casadesús, J. (2015). DNA methylation in bacteria: from the methyl group to the methylome. *Curr. Opin. Microbiol.* 25, 9–16.
- Sandoval, P.Y., Swart, E.C., Arambasic, M., and Nowacki, M. (2014). Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. *Dev. Cell* 28, 174–188.
- Saunders, A., Werner, J., Andrulis, E.D., Nakayama, T., Hirose, S., Reinberg, D., and Lis, J.T. (2003). Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. *Science* 301, 1094–1096.
- Schlesinger, M.B., and Formosa, T. (2000). POB3 is required for both transcription and replication in the yeast *Saccharomyces cerevisiae*. *Genetics* 155, 1593–1606.
- Sen, R., Ferdoush, J., Kaja, A., and Bhaumik, S.R. (2016). Fine-Tuning of FACT by the Ubiquitin Proteasome System in Regulation of Transcriptional Elongation. *Mol. Cell. Biol.* 36, 1691–1703.
- Shi, D.-Q., Ali, I., Tang, J., and Yang, W.-C. (2017). New Insights into 5hmC DNA Modification: Generation, Distribution and Function. *Front. Genet.* 8, 100.
- Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* 14, 407–410.

Singh, D.P., Saudemont, B., Guglielmi, G., Arnaiz, O., Goût, J.-F., Prajer, M., Potekhin, A., Przybòs, E., Aubusson-Fleury, A., Bhullar, S., et al. (2014). Genome-defence small RNAs exapted for epigenetic mating-type inheritance. *Nature advance online publication*.

Sonneborn, T.M. (1974). *PARAMECIUM AURELIA*.

Steele, C.J., Barkocy-Gallagher, G.A., Preer, L.B., and Preer, J.R. (1994). Developmentally excised sequences in micronuclear DNA of *Paramecium*. *Proc. Natl. Acad. Sci. U. S. A.* 91, 2255–2259.

Stuwe, T., Hothorn, M., Lejeune, E., Rybin, V., Bortfeld, M., Scheffzek, K., and Ladurner, A.G. (2008). The FACT Spt16 “peptidase” domain is a histone H3-H4 binding module. *Proc. Natl. Acad. Sci. U. S. A.* 105, 8884–8889.

Sulewska, A., Niklinska, W., Kozłowski, M., Minarowski, L., Naumnik, W., Niklinski, J., Dabrowska, K., and Chyczewski, L. (2007). Detection of DNA methylation in eucaryotic cells. *Folia Histochem. Cytobiol.* 45, 315–324.

Swart, E.C., Denby Wilkes, C., Sandoval, P.Y., Hoehener, C., Singh, A., Furrer, D.I., Arambasic, M., Ignarski, M., and Nowacki, M. (2017). Identification and analysis of functional associations among natural eukaryotic genome editing components. *F1000Research* 6, 1374.

Takahata, S., Yu, Y., and Stillman, D.J. (2009a). The E2F functional analogue SBF recruits the Rpd3(L) HDAC, via Whi5 and Stb1, and the FACT chromatin reorganizer, to yeast G1 cyclin promoters. *EMBO J.* 28, 3378–3389.

Takahata, S., Yu, Y., and Stillman, D.J. (2009b). FACT and Asf1 regulate nucleosome dynamics and coactivator binding at the HO promoter. *Mol. Cell* 34, 405–415.

Tam, L.W., and Ng, S.F. (1986). The role of the micronucleus in stomatogenesis in sexual reproduction of *Paramecium tetraurelia*: laser microbeam irradiation of the micronucleus. *J. Cell Sci.* 86, 287–303.

Tan, B.C.-M., Chien, C.-T., Hirose, S., and Lee, S.-C. (2006). Functional cooperation between FACT and MCM helicase facilitates initiation of chromatin DNA replication. *EMBO J.* 25, 3975–3985.

Touzeau, A., de Vanssay, A., Arnaiz, O., Frapporti, A., and Duharcourt, S. (In preparation). *Paramecium tetraurelia* histone chaperone Spt16 is an essential component of programmed genome rearrangements.

True, J.D., Muldoon, J.J., Carver, M.N., Poorey, K., Shetty, S.J., Bekiranov, S., and Auble, D.T. (2016a). The Mot1 ATPase and Spt16 Histone Chaperone Co-Regulate Transcription through Preinitiation Complex Assembly and Nucleosome Organization. *J. Biol. Chem.*

True, J.D., Muldoon, J.J., Carver, M.N., Poorey, K., Shetty, S.J., Bekiranov, S., and Auble, D.T. (2016b). The Modifier of Transcription 1 (Mot1) ATPase and Spt16 Histone Chaperone Co-regulate Transcription through Preinitiation Complex Assembly and Nucleosome Organization. *J. Biol. Chem.* 291, 15307–15319.

- Tsunaka, Y., Toga, J., Yamaguchi, H., Tate, S., Hirose, S., and Morikawa, K. (2009). Phosphorylated intrinsically disordered region of FACT masks its nucleosomal DNA binding elements. *J. Biol. Chem.* 284, 24610–24621.
- Tucker, J.B., Beisson, J., Roche, D.L., and Cohen, J. (1980). Microtubules and control of macronuclear “amitosis” in *Paramecium*. *J. Cell Sci.* 44, 135–151.
- Ujvári, A., Hsieh, F.-K., Luse, S.W., Studitsky, V.M., and Luse, D.S. (2008). Histone N-terminal tails interfere with nucleosome traversal by RNA polymerase II. *J. Biol. Chem.* 283, 32236–32243.
- VanDemark, A.P., Blanksma, M., Ferris, E., Heroux, A., Hill, C.P., and Formosa, T. (2006). The structure of the yFACT Pob3-M domain, its interaction with the DNA replication factor RPA, and a potential role in nucleosome deposition. *Mol. Cell* 22, 363–374.
- VanDemark, A.P., Xin, H., McCullough, L., Rawlins, R., Bentley, S., Heroux, A., Stillman, D.J., Hill, C.P., and Formosa, T. (2008). Structural and functional analysis of the Spt16p N-terminal domain reveals overlapping roles of yFACT subunits. *J. Biol. Chem.* 283, 5058–5068.
- Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* 16, 178–189.
- Vogt, A., and Mochizuki, K. (2013). A domesticated PiggyBac transposase interacts with heterochromatin and catalyzes reproducible DNA elimination in *Tetrahymena*. *PLoS Genet.* 9, e1004032.
- Wang, Y., Sheng, Y., Liu, Y., Pan, B., Huang, J., Warren, A., and Gao, S. (2017). N6-methyladenine DNA modification in the unicellular eukaryotic organism *Tetrahymena thermophila*. *Eur. J. Protistol.* 58, 94–102.
- Winkler, D.D., Muthurajan, U.M., Hieb, A.R., and Luger, K. (2011). Histone chaperone FACT coordinates nucleosome interaction through multiple synergistic binding events. *J. Biol. Chem.* 286, 41883–41892.
- Winston, F., and Carlson, M. (1992). Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends Genet.* 8, 387–391.
- Wion, D., and Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* 4, 183–192.
- Wittmeyer, J., and Formosa, T. (1997). The *Saccharomyces cerevisiae* DNA polymerase alpha catalytic subunit interacts with Cdc68/Spt16 and with Pob3, a protein similar to an HMG1-like protein. *Mol. Cell Biol.* 17, 4178–4190.
- Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., et al. (2016). DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature*.
- Xin, H., Takahata, S., Blanksma, M., McCullough, L., Stillman, D.J., and Formosa, T. (2009). yFACT induces global accessibility of nucleosomal DNA without H2A-H2B displacement. *Mol. Cell* 35, 365–376.

- Yang, J., Zhang, X., Feng, J., Leng, H., Li, S., Xiao, J., Liu, S., Xu, Z., Xu, J., Li, D., et al. (2016). The Histone Chaperone FACT Contributes to DNA Replication-Coupled Nucleosome Assembly. *Cell Rep.* 16, 3414.
- Ye, P., Luan, Y., Chen, K., Liu, Y., Xiao, C., and Xie, Z. (2017). MethSMRT: an integrative database for DNA N6-methyladenine and N4-methylcytosine generated by single-molecular real-time sequencing. *Nucleic Acids Res.* 45, D85–D89.
- You, Y., Aufderheide, K., Morand, J., Rodkey, K., and Forney, J. (1991). Macronuclear transformation with specific DNA fragments controls the content of the new macronuclear genome in *Paramecium tetraurelia*. *Mol. Cell. Biol.* 11, 1133–1137.
- Zagulski, M., Nowak, J.K., Le Mouël, A., Nowacki, M., Migdalski, A., Gromadka, R., Noël, B., Blanc, I., Dessen, P., Wincker, P., et al. (2004). High coding density on the largest *Paramecium tetraurelia* somatic chromosome. *Curr. Biol. CB* 14, 1397–1404.
- Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., et al. (2015). N6-Methyladenine DNA Modification in *Drosophila*. *Cell* 161, 893–906.
- Zheng, S., Crickard, J.B., Srikanth, A., and Reese, J.C. (2014). A highly conserved region within H2B is important for FACT to act on nucleosomes. *Mol. Cell. Biol.* 34, 303–314.
- Zunder, R.M., Antczak, A.J., Berger, J.M., and Rine, J. (2012). Two surfaces on the histone chaperone Rtt106 mediate histone binding, replication, and silencing. *Proc. Natl. Acad. Sci. U. S. A.* 109, E144–153.

Annexes 1

Nom	Séquence (5' - 3')	Application
fed1_SSRP1_f1	CAGAGTCAATATAGATGATAAGAATTTG	Insert feeding #1 pour Pob3-1
fed1_SSRP1_r1	ACTTCGTGAATGGTAAGTATGGAATCGC	Insert feeding #1 pour Pob3-1
fed2_SSRP1_f2	AAAATATACAATGGATTTCTTTAAAAAAG	Insert feeding #2 pour Pob3-1
fed2_SSRP1_r2	ATTTTTCTAATGGGAATAGGAATCCTG	Insert feeding #2 pour Pob3-1
51A-1835-5'(3)	GGTTGCGTAACACTTCCTCTTAAATGTGAG	PCR around IESA1835
51A-1835-3'	ATCCTAACATCCTTGAATAGTTACTGATCC	PCR around IESA1835
51G1832up	GCTATAACTCTTGAAGCTGCTTGTAAATATG	PCR around IESG1832
51G1832do	TTGTCAATGAGCCATTAACAGTTGCTGGAT	PCR around IESG1832
51G2832-6	CAAATACTGGTGGTGAACATCCTCAACTGC	PCR around IESG2832
51G2832-8	GAACCATTTAATGCGCAATATCCTGCATTTCC	PCR around IESG2832
51A-6649-5'	AAATGGTACTGTTTGTCTTGGGATAGTGC	PCR around IESA6649
51A-6649-3'	CAGCAGTACATCCAGCTCTCTAAGTTTAGC	PCR around IESA6649
pob3_up2_F	GAATTAGGTATTAAGGAAGTGGATG	PCR pour POB3-1 et les séquences régulatrices
pob3_do2_R	AAACGAAATCATCTAATTACGAGAG	PCR pour POB3-1 et les séquences régulatrices
pcrIIPOB3_TGA_F	ACAATAATATATATTTGAATATTTAAAATA	PCR pour introduire GFP en C-ter de Pob3-1
pcrIIPOB3_TGA_R	TTTTTTAACAGCTCTTGTTTTTTA	PCR pour introduire GFP en C-ter de Pob3-1
qPCR_GFP_F1	CTTTTCACTGGTGTGTTCC	qPCR to measure transgene copy number
qPCR_GFP_R1	TCCATAAGTTGCATCACCTT	qPCR to measure transgene copy number
GAPDH_qPCR_F1	TTTGGTATTGTTGAGGGTAAAA	qPCR to measure transgene copy number
GAPDH_qPCR_R1	AAAGGTGGAAAAGACTGGAG	qPCR to measure transgene copy number
deletionMBD_F1	TGATCAACTTATATAATATATTGTATAA	Délétion de MBD dans GFP-Spt16-1
deletionMBD_R1	GTCTTATTCAAATTCGAATTCAG	Délétion de MBD dans GFP-Spt16-1

Table S1 : Oligonucléotides utilisés dans les résultats supplémentaires

Annexe 2

Espèce	Nom	Numéro d'accèsion
<i>Paramecium tetraurelia</i>	Pob3_1	PTET.51.1.P0610231
<i>Paramecium tetraurelia</i>	Pob3_2a	PTET.51.1.P0610125
<i>Paramecium tetraurelia</i>	Pob3_2c	PTET.51.1.P0340178
<i>Tetrahymena thermophila</i>	Pob3	TTHERM_00049080
<i>Schizosaccharomyces pombe</i>	Pob3	SPBC609.05
<i>Saccharomyces cerevisiae</i>	Pob3	YML069W
<i>Caenorhabditis elegans</i>	Ssrp1	WP:CE08542
<i>Drosophila melanogaster</i>	Ssrp1	FBgn0010278
<i>Homo sapiens</i>	Ssrp1	HGNC:11327
<i>Mus musculus</i>	Ssrp1	MGI:107912
<i>Arabidopsis thaliana</i>	Ssrp1	AT3G28730
<i>Paramecium octaurelia</i>	Pob3_1	POCTA.138.1.P0760227
<i>Paramecium octaurelia</i>	Pob3_2a	POCTA.138.1.P0490118
<i>Paramecium octaurelia</i>	Pob3_2c	POCTA.138.1.P0760120
<i>Paramecium tredecaurelia</i>	Pob3_1	PTRED.209.2.P71800001293620112
<i>Paramecium tredecaurelia</i>	Pob3_2a	PTRED.209.2.P71800001293690223
<i>Paramecium tredecaurelia</i>	Pob3_2b	PTRED.209.2.P71800001293370204
<i>Paramecium tredecaurelia</i>	Pob3_2c	PTRED.209.2.P71800001293260140
<i>Paramecium biaurelia</i>	Pob3_1a	PBIA.V1_4.1.P00530042
<i>Paramecium biaurelia</i>	Pob3_1b	PBIA.V1_4.1.P00220183
<i>Paramecium biaurelia</i>	Pob3_2a	PBIA.V1_4.1.P00530142
<i>Paramecium biaurelia</i>	Pob3_2b	PBIA.V1_4.1.P00210081
<i>Paramecium biaurelia</i>	Pob3_2c	PBIA.V1_4.1.P00460005
<i>Paramecium primaurelia</i>	Pob3_1a	PPRIMP18748
<i>Paramecium primaurelia</i>	Pob3_1b	PPRIMP13357
<i>Paramecium primaurelia</i>	Pob3_2a	PPRIMP14276
<i>Paramecium primaurelia</i>	Pob3_2b	PPRIMP15045
<i>Paramecium primaurelia</i>	Pob3_2c	PPRIMP18654
<i>Paramecium pentaurelia</i>	Pob3_1	PPENT.87.1.P0800209
<i>Paramecium pentaurelia</i>	Pob3_2a	PPENT.87.1.P0010114
<i>Paramecium pentaurelia</i>	Pob3_2b	PPENT.87.1.P0530160
<i>Paramecium pentaurelia</i>	Pob3_2c	PPENT.87.1.P0470124
<i>Paramecium sexaurelia</i>	Pob3_1	PSEX.AZ8_4.1.P0470222
<i>Paramecium sexaurelia</i>	Pob3_2a	PSEX.AZ8_4.1.P0470114
<i>Paramecium sexaurelia</i>	Pob3_2b	PSEX.AZ8_4.1.P0580133
<i>Paramecium sexaurelia</i>	Pob3_2c	PSEX.AZ8_4.1.P0430154
<i>Paramecium caudatum</i>	Pob3_1	PCAU.43c3d.1.P00030119
<i>Paramecium caudatum</i>	Pob3_2	PCAU.43c3d.1.P00730044

Table S2 : Numéros d'accèsion des protéines utilisées dans l'analyse phylogénétique

Les numéros d'accèsion, nom des protéines et noms des espèces sont donnés. *Paramecium* database: <https://paramecium.i2bc.paris-saclay.fr/>, *T. thermophila* database: <http://ciliate.org/index.php/home/welcome>, *S. pombe* database: <https://www.pombase.org/>, *S. cerevisiae* database: <https://www.yeastgenome.org/>, *C. elegans* database: <https://wormbase.org/#012-34-5>, *D. melanogaster* database: <http://flybase.org/>, *M. musculus* database: <http://www.informatics.jax.org/>, *H. sapiens* database: <https://www.genenames.org/>, *A. thaliana* database: <https://www.arabidopsis.org/>.

Annexe 3

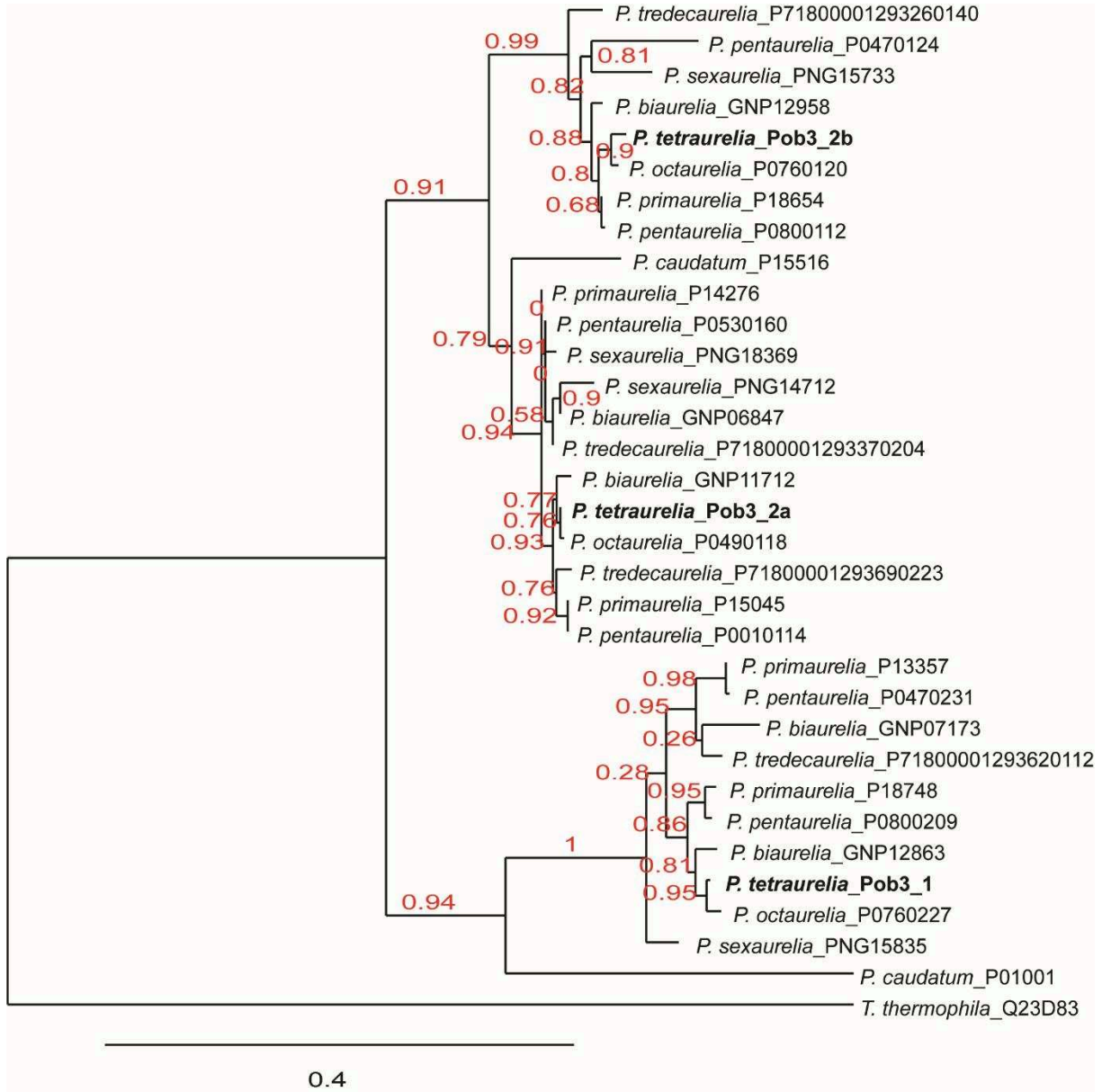


Figure S1 : Identification des homologues de Pob3 chez *P. tetraurelia*

Arbre phylogénétique des protéines Pob3 de *P. tetraurelia* et des autres espèces de Paramécies basé sur l'alignements des séquences protéiques totales par MUSCLE. La phylogénie de l'alignement est générée en utilisant PhyML 3.0 sur Phylogeny.fr avec une procédure de bootstrap de 100. L'arbre est observé grâce à Tree.Dyn 198.3. Les séquences protéiques et les numéros d'accèsion sont indiqués en Table S2/Annexe 2.