NNT : 2016SACLX087

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
L'ECOLE POLYTECHNIQUE

ECOLE DOCTORALE N° 573
INTERFACES

BIOLOGIE

Par

Floriane DELPECH

Dynamique cellulaire des protéines de la réplication chez l'archée halophile *Haloferax volcanii*

Thèse présentée et soutenue à l'Ecole POLYTECHNIQUE, le 17 novembre 2016 :

Composition du Jury :

M. FRANZETTI Bruno	DR, CNRS	Président
M. ALLERS Thorsten	Pr. associé, Université de Nottingham	Rapporteur
M. FLAMENT Didier	Chercheur, Directeur adjoint, IFREMER	Rapporteur
M. CHASTANET Arnaud	CR, INRA	Examinateur
M. MYLLYKALLIO Hannu	DR, CNRS	Co-directeur de thèse
Mme LESTINI Roxane	Maitre de conférences, Ecole polytechnique	Co-directrice de thèse

REMERCIEMENTS

Tout d'abord je tiens à remercier les membres de mon jury d'avoir accepté de lire et d'évaluer ma thèse. Je remercie Bruno Franzetti d'avoir accepté de présider ce jury de thèse. Je remercie Didier Flament et Thorsten Allers d'avoir accepté d'être rapporteurs pour mon manuscrit, merci de l'avoir lu avec tant d'attention et de m'avoir éclairé par vos commentaires et questions. Je remercie Arnaud Chastanet qui bien que non spécialiste des archées a bien voulu être mon tuteur et a permis de porter un œil nouveau sur mon travail.

Je remercie Roxane Lestini pour ces presque quatre années à travailler ensemble, ça a été un réel plaisir. Merci de m'avoir encadrée en master et en thèse, de m'avoir fait confiance tout en m'aiguillant, d'avoir su me remonter le moral dans les moments où « ça marche pas…» et de m'avoir fait rentrer dans la communauté archée. Merci pour tout ce que tu m'as appris.

Je remercie Hannu Myllykallio pour son encadrement en thèse. Merci pour nos discussions au cours desquelles tu me faisais part de tes idées originales et parfois farfelues qui m'ont amenées à pousser la réflexion plus loin. Je te remercie aussi d'avoir toujours su faire ressortir le positif dans le négatif.

Je vous remercie tous les deux de m'avoir co-encadrée tout au long de ma thèse et de m'avoir formée à la recherche.

Merci aux directeurs du laboratoire Jean-Louis Martin et François Hache d'avoir veillé au bon déroulement de ma thèse pendant ces trois années.

Je voudrais également remercier l'équipe du pôle Biologie et quelques intrus du pole microscopie pour leur aide quotidienne et d'avoir supporté mes maugréassions sur des sujets divers et variés, plus particulièrement merci à Stéphane Skouloubris et Hubert Becker pour leurs conseils avisés, à Marie-Claire Schanne-Klein et Pierre Mahou pour leur aide sur la partie microscopie de mon travail.

Je remercie toute l'équipe d'enseignement de polytechnique qui m'a permis de profiter de cette expérience et de passer des moments agréables.

Je remercie tous les membres du LOB de faire de ce labo un lieu où il est si agréable de travailler. Mon séjour n'aurait jamais été aussi plaisant sans l'équipe des « non-permanents », les pauses café-gâteaux-thé-gâteaux risquent de me manquer. Je remercie Claire, Pierre et les deux Guillaume d'avoir souvent affronté la passerelle pour nous rendre visite. Merci à mes collègues de bureau Margaux, Yoann, Marianne et Chao pour leur soutien lors de la rédaction.

Je voudrais remercier aussi mes parents et ma sœur de m'avoir épaulée pendant mes études et de m'avoir appris à prendre du recul. Merci d'être là pour moi.

Et enfin je voudrais remercier Alexis pour son soutien et son optimisme indéfectible. Merci de m'aider à réaliser mes rêves.

Table des matières

LISTE DES FIGURES 4

LISTE DES TABLEAUX 7

1 INTRODUCTION 1

1.1 Les Archées 2

1.1.1 La découverte des archées 2

1.1.2 Les archées dans l'environnement 3

1.1.3 Le placement des archées au sein du domaine du vivant 4

1.2 La réplication chez les archées 9

1.2.1 Les trois étapes conservées : la pré-initiation, l'initiation et élongation 9

1.2.2 Les origines de réplication et leurs protéines associées 10

1.2.3 La fourche de réplication 15

1.2.4 La maturation des fragments d'Okazaki 20

1.2.5 *Haloferax volcanii* comme organisme modèle pour l'étude de la réplication chez les archées 20

1.3 L'imagerie 22

1.3.1 Microscopie à contraste de phase et contraste d'interférence pour observer les archées 22

1.3.2 La microscopie électronique pour étudier les structures des cellules 24

1.3.3 Etat de l'art sur la microscopie à fluorescence 26

1.3.4 La microscopie à fluorescence par immunomarquage ou marquage de l'ADN 28

1.3.5 Microscopie à fluorescence par expression de la GFP dans les cellules 30

1.4 Problématique de la thèse 34

2 MATERIEL ET METHODE 37

2.1 Milieux et souches d'*Haloferax volcanii* 38

2.2 Milieux et souches d'*Escherichia coli* 39

2.3 Expression physiologique d'une protéine fusionnée à la GFP chez *H. volcanii* .. 40

2.3.1 Construction dans *E. coli* du plasmide intégratif 42

2.3.2 Intégration et remplacement au locus chromosomique chez H. volcanii 48

2.4 Localisation du signal GFP dans les cellules vivantes et analyse quantitative ... 54

2.5 Colocalisation du signal GFP avec l'ADN dans les cellules vivantes et analyse quantitative ... 56

2.6 Fluorescence Recovery After Photobleaching (FRAP) ... 58

2.7 Western-blot ... 60

3 RESULTATS ... 61

3.1 Etude de la protéine PCNA ... 62

3.1.1 Clonage du plasmide intégratif ... 62

3.1.2 Construction d'une souche *Haloferax volcanii* permettant l'expression de PCNA fusionnée à la GFP ... 65

3.2 Etude de la protéine FEN1 ... 67

3.2.1 Clonage du plasmide intégratif ... 67

3.2.2 Construction d'une souche *Haloferax volcanii* permettant l'expression de FEN1 fusionnée à la GFP ... 69

3.2.3 Etude de la fonctionnalité de la souche exprimant FEN1 fusionnée à la GFP ... 71

3.2.4 Localisation *in vivo* de FEN1 ... 73

3.3 Etude de la protéine Orc1 ... 78

3.3.1 Clonage du plasmide intégratif ... 78

3.3.2 Construction des souches d'*H. volcanii* permettant l'expression de Orc1 fusionnée à la GFP ... 80

3.3.3 Caractéristiques des souches *gfp::orc1* et Δ*ori gfp::orc1* ... 82

3.3.4 Localisation *in vivo* de ORC1 ... 84

3.4 Etude de la protéine Rpa2 ... 90

3.4.1 Clonage du plasmide intégratif ... 90

3.4.2 Construction des souches d'*H. volcanii* permettant l'expression de Rpa2 fusionnée à la GFP ... 92

3.4.3 Etude de la fonctionnalité de la protéine fusion GFP::Rpa2 ... 93

3.4.4 Localisation *in vivo* de GFP::RPA2 selon la phase de croissance ... 96

3.4.5 Localisation *in vivo* de RPA2 après exposition à différents agents génotoxiques ... 99

3.4.6 Quantification de GFP::Rpa2 : western blot anti-GFP ... 105

3.4.7 Colocalisation de l'ADN avec les foyers GFP::Rpa2 ... 107

3.4.8 Dynamique de Rpa2 : Fluorescence Recovery After Photobleaching (FRAP) ... 109

4 CONCLUSION ET PERSPECTIVES ... 114

LISTE DES FIGURES

Figure 1. La découverte des archées. 2

Figure 2. Arbre phylogénétique des différentes espèces d'archées. 3

Figure 3. Les archées et leurs environnements extrêmes. 4

Figure 4. Homologies des protéines de la réplication archéennes avec les protéines bactériennes et eucaryotes. 7

Figure 5. Les origines de réplication chez un groupe représentatif des euryarchées, les halobactériales, et chez un groupe représentatif des crénarchées, les sulfolobales. 11

Figure 6. Schéma d'une fourche de réplication archéenne. 16

Figure 7. Les microscopes à contraste de phase et contraste interférentiel. 23

Figure 8. Images en microscopie optique. 24

Figure 9. Images d'enveloppe cellulaire en microscopie électronique. 25

Figure 10. Images en cryotomographie des cellules de *sulfolobus.* 26

Figure 11. Trajets optiques des deux types de microscopie: par champs large et confocale. 27

Figure 12. Image de fluorescence d'*H. volcanii.* 30

Figure 13. Structure de la GFP. 31

Figure 14. Localisation *in vivo* des protéines CetZ1-GFP. 33

Figure 15. Les différents motifs de structure aboutissant à la structure tridimensionnelle. 34

Figure 16. Carte du plasmide pTA131. 41

Figure 17. Méthodes de clonage utilisées. 43

Figure 18. Le pop-in / pop-out. 50

Figure 19. Localisation du signal GFP. 55

Figure 20. Schéma des spectres d'excitation et d'émission de l'Hoechst 33258. 57

Figure 21. Schéma des spectres d'excitation et d'émission de la GFP utilisée. 57

Figure 22. Photoblanchiment d'une zone comprenant deux foyers de fluorescence au sein d'une cellule. 60

Figure 23. Structure de la protéine PCNA d'*H. volcanii.* 62

Figure 24. Vérification par restriction du plasmide pFD4. 64

Figure 25. Schéma d'un pop-in/pop-out dans la région *« pcna ».* 66

Figure 26. Crible par PCR des colonies après pop-in/pop-out pour l'expression de GFP::PCNA. 66

Figure 27. Produit de la PCR 2 temps correspond à l'insert « *gfp-fen1* ». 68

Figure 28. Vérification par restriction du plasmide pFD1. 69

Figure 29. Crible par PCR des colonies après pop-in pour l'expression de GFP::Fen1. 70

Figure 30. Crible par PCR des colonies après pop-in/pop-out pour l'expression de la GFP. 70

Figure 31. Survie cellulaire des souches exprimant l'allèle *gfp::fen1*(HVFD19), l'allèle sauvage *fen1* (H26) et sans l'allèle *fen1 :* Δ*fen1* après exposition à différents agents génotoxiques. 72

Figure 32. Survie cellulaire après exposition à 2,5 mM d'hydroxyurée. 74

Figure 33. Analyses des cellules *gfp::fen1* (HVFD19) et *fen1* (H26) en phase exponentielle de croissance. 76

Figure 34. Analyse des cellules après 17h00 d'exposition à l'hydroxyurée. 77

Figure 35. Produit de la PCR 2 temps correspondant à l'insert « *gfp-orc1* ». 79

Figure 36. Vérification par restriction du plasmide pFD3. 80

Figure 37. Amplification par PCR de la région *orc1* après pop-in/pop-out pour intégrer l'allèle *gfp::orc1.* 81

Figure 38. Hybridation sur colonies pour tester la présence de l'allèle *gfp::orc1* réalisé sur les pop-out des cellules Δ*ori.* 82

Figure 39. Survie cellulaire des souches *gfp::orc1* après exposition à différents agents génotoxiques. 84

Figure 40. Images de microscopie de fluorescence des cellules exprimant l'allèle *gfp::orc1* selon la phase de croissance. 86

Figure 41. Analyses des cellules *gfp::orc1* (HVFD18) selon la phase de croissance. 87

Figure 42. Images de microscopie de fluorescence des cellules exprimant l'allèle *gfp::orc1* dans une souche Δ*ori* selon la phase de croissance. 88

Figure 43. Analyses des cellules *gfp::orc1* Δ*ori* (HVFD43) en phase exponentielle de croissance. 89

Figure 44. Produit de la PCR 2 temps correspondant à l'insert « *gfp-rpa2* ». 91

Figure 45. Vérification par restriction du plasmide pFD2. 92

Figure 46. Amplification par PCR de la région *orc1* après pop-in/pop-out pour intégrer l'allèle *gfp::rpa2.* 93

Figure 47. Survie cellulaire après exposition des cellules à différents agents génotoxiques. 94

Figure 48. Images de fluorescence des cellules exprimant *gfp::rpa2* selon la phase de croissance. .. 96

Figure 49. Analyse des cellules *gfp::rpa2* (HVFD1) selon la phase de croissance. 97

Figure 50. Images de fluorescence des cellules exprimant *gfp ::rpa2* après exposition à l'APD et au control (DMSO). 99

Figure 51. Analyse des cellules *gfp::rpa2* (HVFD1) après 17h00 d'exposition à l'APD et à son contrôle (DMSO). 100

Figure 52. Récapitulatif des images et analyses réalisées sur les cellules *gfp::rpa2* après 17h00 d'exposition à la MMC et à son contrôle (eau). 102

Figure 53. Analyses de l'imagerie réalisée sur les cellules *gfp ::rpa2* après 17h00 d'exposition différentes concentrations de phléomycine. 104

Figure 54. Intensité par unité de surface des cellules *gfp::rpa2* après exposition à différentes concentrations de phléomycine et d'APD. 105

Figure 55. Western blot anti-GFP réalisé sur les extraits protéiques des cellules *gfp::rpa2* (HVFD1) et des cellules sauvages (H26). 106

Figure 56. Analyse de la colocalisation de GFP ::Rpa2 où l'ADN est compacté après exposition à 0.1 mg/mL de phléomycine. 108

Figure 57. Trois cas de retour de fluorescence après photoblanchiment 110

Figure 58. FRAP réalisé sur les foyers des cellules *gfp::rpa2*. 111

LISTE DES TABLEAUX

Tableau 1. Les protéines de la réplication chez les archées comparées aux deux autres domaines. 6

Tableau 2. Tableau représentant le nombre d'origines de réplication chez les archées. 12

Tableau 3. Les protéines d'intérêt et leurs fonctions associées dans les trois domaines de la vie. 35

Tableau 4. Souches d'*Haloferax volcanii* utilisées. 39

Tableau 5. Souches d'Escherichia coli utilisées. 40

Tableau 6. Plasmides utilisés 40

Tableau 7. Caractéristiques des inserts utilisés pour le clonage 41

Tableau 8. Oligonucléotides utilisés. 47

Tableau 9. Les oligonucléotides utilisés pour le criblage des pop-out d'*H. volcanii* 52

Tableau 10. Conditions de PCR optimales pour générer l'insert « *gfp-pcna* » par PCR 2 temps. 63

Tableau 11. Conditions de PCR optimales pour générer l'insert « *gfp-pcna* » par SLIC. 63

Tableau 12. Les conditions de PCR optimales pour chaque générer l'insert « *gfp-fen1* » par PCR 2 temps. 67

Tableau 13. Temps de génération des souches exprimant l'allèle *gfp::fen1,* l'allèle sauvage *fen1* et délétées pour *fen1* : Δ*fen1* 71

Tableau 14. Tableau récapitulatif de la viabilité et du temps de génération des souches possédant l'allèle *gfp::orc1* et l'allèle sauvage *orc1* avec et sans origine de réplication 83

Tableau 15. Les conditions de PCR optimales pour chaque région constituant l'insert « *gfp-rpa2* » par PCR 2 temps 90

Tableau 16. Viabilité et temps de génération des souches possédant l'allèle *gfp::rpa2,* l'allèle sauvage *rpa2* et délétées pour *rpa2 :* Δ*rpa2.* 93

Tableau 17. Analyse du retour de fluorescence après exposition à l'aphidicoline, à la phléomycine et en phase stationnaire de croissance 110

Tableau 18. Quantification des fractions mobiles et immobiles 111

« Tous les dogmes finissent dans l'embarras »

- Roland Topor

1 INTRODUCTION

1.1 Les Archées

1.1.1 La découverte des archées

Les archées sont des microorganismes dépourvus de noyau et d'organites. Morphologiquement très semblables aux bactéries, elles ont donc longtemps été identifiées comme telles. Ce n'est que récemment, grâce au développement de marqueurs moléculaires et aux travaux réalisés entre 1977 et 1990 par Carl Woese et George Fox, que les archées ont été identifiées comme un domaine à part [1]. L'équipe de Carl Woese travaillait sur la petite sous-unité de l'ARN ribosomal (16S) de différents organismes, cet ARN 16s est essentiel et ubiquitaire. Ils ont mené la première étude quantitative phylogénétique basée sur la comparaison de séquences d'ARN 16S. Ils ont identifié certains organismes méthanogènes dont l'ARN 16S était différent de celui des bactéries mais aussi des eucaryotes, ce qui remettait en cause la classification et l'arbre phylogénétique établi qui partageait le monde du vivant en deux grands domaines : les bactéries et les eucaryotes [2]. Leurs travaux ont permis d'identifier trois grands types d'ARN 16S, ce qui a conduit à une nouvelle classification du monde vivant : les bactéries, les archées et les eucaryotes (Figure 1).

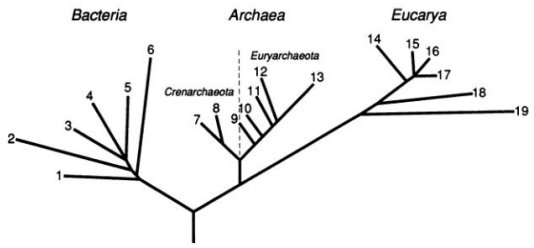


Figure 1. La découverte des archées. Arbre de la vie phylogénétique représentant les trois domaines du vivant. De 1 à 6 : les bactéries, de 7 à 13 : les archées, de 14 à 19 : les eucaryotes, d'après [3].

Les avancées dans des domaines tel que la métagénomique ont permis, par l'étude d'échantillons prélevés dans différents environnements, de découvrir d'autres archées. Des isolements de nouvelles souches en laboratoire ont aussi permis de diversifier le nombre de phylums et de caractéristiques disponibles pour classer les archées. Les archées sont classées en trois groupes principaux : les crenarchées, les euryarchées, les thaumarchées et deux groupes ne comportant chacun qu'un seul organisme identifié à ce jour : les aigarchées et les korarchées (Figure 2). Des études métagénomiques ont permis l'identification d'un autre genre d'archées : les lokiarcheota, ce genre reste néanmoins controversé.

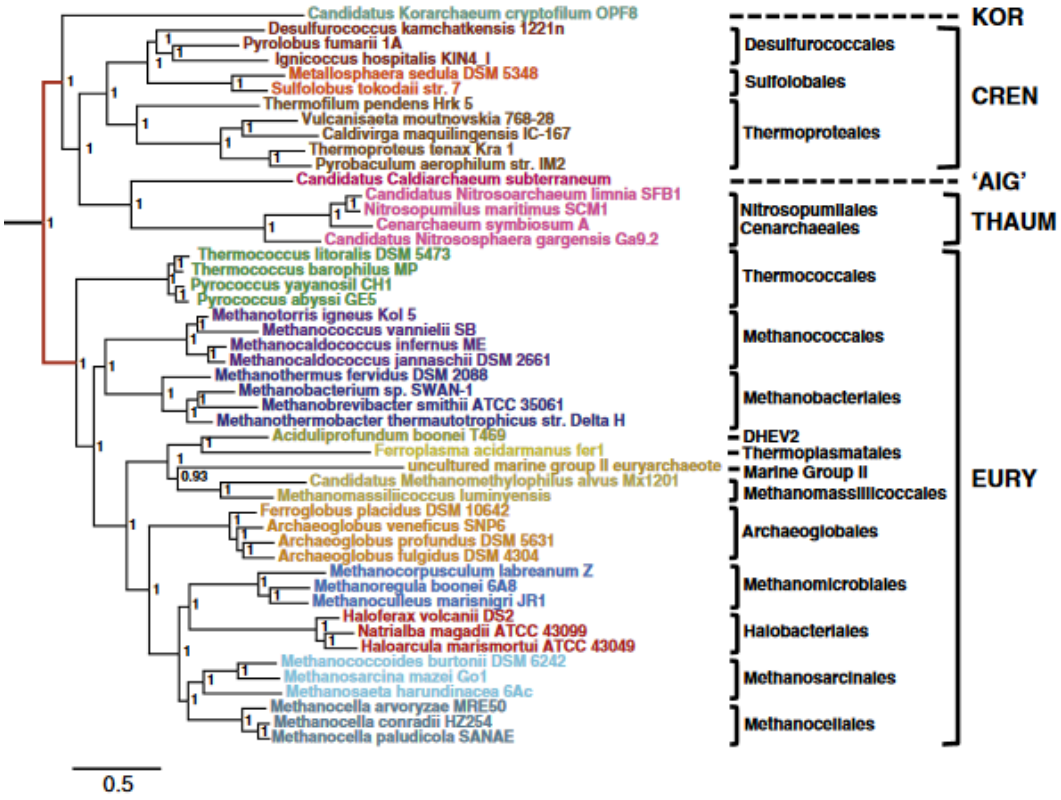


Figure 2. Arbre phylogénétique des différentes espèces d'archées.
Cette phylogénie est basée sur l'utilisation de marqueurs partagés entre les archées et les eucaryotes. D'après [4].

1.1.2 Les archées dans l'environnement

On trouve chez les archées beaucoup d'organismes extrêmophiles, et la diversité de leurs habitats est effectivement très importante. Les méthanogènes, premières archées identifiées, vivent le plus souvent en anaérobie stricte et produisent du méthane [5]. La production du méthane a un impact global sur l'environnement, et le réchauffement climatique en participant à l'effet de serre. Dans l'environnement, la voie métabolique des méthanogènes est spécifique aux archées. Les hyperthermophiles sont des archées supportant de très fortes températures, le record étant tenu par l'archée *Pyrolobus fumarii* qui croît jusqu'à 113°C [6]. Les *Sulfolobales* (Figure 3 A) sont rencontrées dans des environnements très acides, vivant à un pH < 1. Les *Halobacteriales* peuvent vivre à de très fortes concentrations en sel, proche de la saturation (Figure 3 B). D'autres archées ont été isolées près de cheminées hydrothermales comme *Pyroccosus abyssi* (Figure 3 C). Les archées peuvent s'adapter à des environnements extrêmes mais pas seulement, elles sont ubiquitaires. En effet, on retrouve des archées dans

notre tube digestif, sur notre plaque dentaire [7] mais également dans les jardins, les lacs et les forêts.



Figure 3. Les archées et leurs environnements extrêmes.
A. 'Grand prismatic spring' par David Mencin. **B.** Lac salé en Australie, la couleur rose est due aux halophiles y vivant, par Neilsphotography. **C.** Cheminées hydrothermales, fosse des caïmans près de la Jamaïque, par CNRS Phototèque / Jean Louis cheminée.

Jusqu'à ce jour, bien que nous sommes loin d'avoir appréhendé la diversité de ce domaine, aucune archée pathogène n'a été identifiée. Mais de par leur ubiquité, elles jouent un rôle prépondérant dans notre écosystème. A titre d'exemple, certaines archées sont impliquées dans le cycle de la nitrification et transforment l'ammonium en nitrite [8], or la nitrification est un processus qui est utilisé par des stations d'épuration. D'autres archées, faisant partie de notre microbiote, sont impliquées dans des processus biologiques de notre corps. Certaines études ont même proposé un usage thérapeutique des archées en prévention de maladies cardiovasculaires [9]. Ainsi les archées sont des organismes riches en diversité tant par leur environnement, leur morphologie ou leurs mécanismes intracellulaires. Leur vrai potentiel reste encore à découvrir et exploiter, notamment dans des domaines tel que l'écologie.

1.1.3 Le placement des archées au sein du domaine du vivant

La structure cellulaire

Les archées ont des morphologies variées, et certains sont même hautement pléiomorphique. Elles peuvent être de forme carrée comme *Haloquadratum walsbyi*, ronde comme *Sulfolobus solfataricus*, en disque plat ou en bâtonnet comme *Haloferax volcanii*. Certaines sont aussi très motiles et doivent leur mouvement à une structure baptisée « archaellum » qui leur permet de nager en surface ou dans une culture liquide [10]. L'archaellum est plus proche en termes d'évolution et de structure du pilli IV des bactéries qui assure une attache entre la bactérie dont il provient et une autre bactérie ou surface que de leur flagelle, responsable de la motilité chez les bactéries.

Bien que de morphologie bactérienne, leur enveloppe cellulaire diffère de celles des bactéries. En effet, à la différence des bactéries qui possèdent une paroi cellulaire riche en peptidoglycane [11], les archées n'ont qu'une couche S (S-layer) riche en glycoprotéines qui les protège des agressions extérieures [12]. En revanche, pour s'adapter à de fortes températures, qui dénaturent les biomolécules et liquéfient les membranes, les archées ne possèdent pas les mêmes lipides. En effet, les hydrocarbones sont liés entre eux *via* des liaisons ether qui en condition extrême sont plus stables chimiquement que les liaisons ester présentes chez les eucaryotes et la plupart des bactéries. Les membranes sont plus imperméables grâce à la structure de ces hydrocarbures qui sont formés de chaine de carbone branchées chez les archées au lieu de longue chaine linéaire chez les eucaryotes et les bactéries [13].

Réplication, transcription et traduction

Bien que les processus de réplication, transcription et traduction soient conservés au sein des trois domaines du vivant, ils présentent des divergences entre domaine (mais aussi au sein de chaque domaine). Cette partie décrit le placement des archées au sein du domaine du vivant à travers la comparaison de ces trois processus fondamentaux : la réplication, la transcription et la traduction, et à travers l'étude phylogénétique des protéines qui y sont impliquées.

Structure du génome et réplication chez les archées

Chez les bactéries, le chromosome est circulaire et présente une seule origine de réplication. Alors que chez les archées, sur un chromosome circulaire (comme pour les bactéries) il peut exister plusieurs origines de réplication (comme chez les eucaryotes mais dont le chromosome est linéaire).

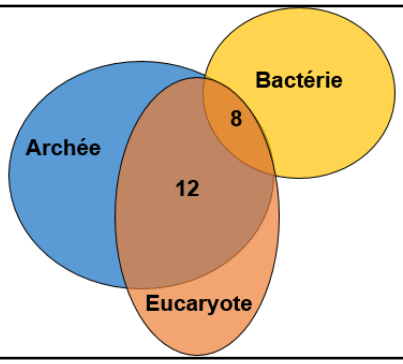
La réplication de l'ADN est le processus permettant la synthèse d'ADN et la production d'une molécule d'ADN identique à la molécule initiale. La réplication de l'ADN est un processus universel qui est essentiel pour toute cellule avant la division cellulaire. Elle est scindée en trois étapes : l'initiation, l'élongation et la terminaison. Ainsi les fonctions assurées par les protéines y participant sont similaires chez les trois domaines du vivant et peuvent donc être comparées. De récentes études ont permis de répertorier le rôle des protéines de la réplication (établi par des études fonctionnelles) et de déterminer par étude *in silico* leur homologie entre les trois domaines du vivant (Figure 4) [14].

		Bacteria	Archaea	Eukarya
Initiation				
	Origin recognition	*Dna A*	Cdc6-Orc1	ORC(1), *ORC(2-6)*, Cdc6
	Helicase loading	*Dna C*	Cdc6-Orc1	Cdc6
				Cdt1
	Replicative helicase	*Dna B (E. Coli)*	MCM	MCM(2-7)
			GINS-51, GINS-23	Gins (Sld5, Psf1), GINS [Psf2, Psf3]
			RecJ(Cdc45)?	Cdc45
Elongation				
	Single Stranded binding protein	*SSB*	SSB	*
			RPA1	RPA70, *RPA34, RPA14*
	Polymerase/exonuclease	*Pol III (Family C)*	Pol B (Family B)	Pol δ, Pol e (Family B)
			Pol D (DPL and DPS)	
	Clamp loader	*Y complex*	RFC-L, RFC-S	RFC-L, RFC-S (1-4)
	Sliding clamp	*B clamp*	PCNA	PCNA
	Primase	Dna G	Primase (PriS, PriL), DnaG ?	Primase (PriS, PriL)
				B subunit
				DNA pol α
	Primer excision (lagging)	*RNase HI*	RNase HII	RNase H2A , *RNase H2B, RNase H2C*
		DNA Pol I	FEN-1	FEN-1
			Dna2?	Dna2
	Maturation (lagging)		ATP-dependent DNA ligase	ATP-dependent DNA ligase
		NAD+-dependent DNA ligase	NAD+-dependent DNA ligase	
DNA relaxation				
	Topoisomerases		Topo VI (Type JJB)	*
			Topo IB (Type IB)	Topo IB (Type IB)
		DNA gyrase, Topo IV (Type IIA)	DNA gyrase (Type IIA)	Topo IIA (Type IIA)

Tableau 1. Les protéines de la réplication chez les archées comparées aux deux autres domaines.
La même couleur désigne une homologie, le surlignage grisé désigne les versions bactériennes qui n'ont de similitude avec les autres domaines que par leur structure, les «*» montrent qu'il existe une homologie chez les eucaryotes mais qu'elle n'est pas impliquée dans la réplication et les «?» désignent les protéines archéennes dont le rôle dans la réplication n'est pas clair. D'après [14].

Ainsi, il existe cinq protéines archéennes impliquées dans l'initiation de la réplication, c'est-à-dire dans le processus de préparation à la synthèse de l'ADN. Parmi ces cinq protéines, toutes sont homologues aux protéines eucaryotes uniquement, aucune n'est homologue à la fois aux bactéries et aux eucaryotes. Il existe quatorze protéines archéennes impliquées dans l'élongation qui assure la synthèse de l'ADN. Parmi ces quatorze protéines, sept sont homologues à la fois aux protéines bactériennes et eucaryotes, cinq ne sont homologues qu'aux protéines eucaryotes et pour les deux restantes l'une est spécifique aux archées (l'ADN polymérase) et l'autre n'intervient pas dans la réplication (DnaG). Et enfin, il existe trois protéines archéennes impliquées dans l'arrêt de la réplication. Parmi ces protéines, une est homologue à la fois aux protéines bactériennes et eucaryotes, les deux autres sont uniquement homologues aux protéines eucaryotes.

Il est important de noter que toutes les similitudes entre protéines archéennes et protéines eucaryotes sont des homologies de séquence, ce qui n'est pas le cas pour les protéines bactériennes pour lesquelles les outils utilisés n'ont pas permis d'obtenir d'homologies. Ces protéines bactériennes ne possèdent souvent qu'une structure similaire. De plus, lorsqu'une protéine bactérienne de la réplication est homologue à une protéine archéenne, elle l'est aussi avec une protéine eucaryote : c'est le cas de huit protéines sur vingt-deux. En revanche, il existe des protéines archéennes qui ne sont homologues qu'aux protéines eucaryotes, c'est le cas de douze d'entre elles (Figure 4). De plus, les archées ont des caractéristiques qui leur sont propres comme l'existence d'une ADN polymérase D.



<u>Figure 4.</u> Homologies des protéines de la réplication archéennes avec les protéines bactériennes et eucaryotes.
Les protéines archéennes sont en bleues, eucaryotes en orange et bactériennes en jaune.

Ainsi, il est clair que chez les archées, en dépit d'une structure du génome similaire aux bactéries, les mécanismes de la réplication et plus particulièrement les protéines impliquées lors de la réplication sont davantage de type eucaryote que bactérien. Ceci est en accord avec les études de phylogénie réalisées qui ont pu déterminer que les archées étaient plus proches phylogénétiquement des eucaryotes qu'elles ne le sont des bactéries.

Caractéristiques de la transcription chez les archées

La transcription est le processus permettant la synthèse d'une molécule d'ARN à partir d'une molécule d'ADN. Elle est scindée en trois étapes : (i) l'initiation, permettant la fixation de l'ARN polymérase, (ii) l'élongation, permettant la synthèse de l'ARN à partir d'une séquence d'ADN et (iii) la terminaison, permettant l'arrêt de la synthèse de l'ARN lorsque l'ARN polymérase rencontre une séquence de terminaison. Le processus de transcription au sein des archées semble être un mélange entre celui des bactéries et celui des eucaryotes. En effet, les promoteurs, l'ARN polymérase, les facteurs de transcription globaux, les régulateurs de la transcription et la maturation de l'ARN sont soit de type eucaryote soit de type bactérien.

Les promoteurs sont plutôt de type eucaryote, ils ont des séquences TATA box auxquels vient se fixer la TBP (Tata Box binding Protein) et des séquences BRE (B Recognition Element) auxquels se fixent les TFB (transcription factor B). Des études portant sur la comparaison des structures de la machinerie de transcription, rendues possible par la résolution de la structure de l'ARN polymérase archéenne, ont permis d'affirmer que l'ARN polymérase et les facteurs généraux de la transcription archéens sont plutôt de type eucaryote que bactérien [15, 16]. En revanche, les régulateurs de la transcription : (activateurs ou répresseurs) sont davantage de type bactérien qu'eucaryote [17, 18], quant à la maturation des ARN messagers chez les eucaryotes, elle implique un mécanisme d'épissage des introns et d'ajout de séquences protectrices : la coiffe en 5' et la queue polyA en 3'. Les archées ont, comme les eucaryotes, une possibilité d'épissage, en effet il existe des introns qui restent rares [19] et certains ARN peuvent être épissés comme les précurseurs des ARNt et des ARNr. En revanche, une autre forme d'épissage peut être rencontrée : les intéines qui sont des segments de protéines pouvant être « épissés » après la traduction. La polyadénylation des extrémités 3' des ARN n'est pas un processus systématique chez les archées : elle est présente chez certaines archées comme l'archée *Sulfolobus solfataricus* mais est absente chez l'archée halophile *Haloferax volcanii* [20]. En revanche, chez les archées, les ARNs qui possèdent une queue polyA n'ont pas de coiffe en 5'. Par ailleurs, les archées possèdent, comme les bactéries, des unités de transcription polycistroniques.

Ainsi l'appareil de transcription archéen possède un type de transcription eucaryote qui est régulé par un système de type bactérien.

Caractéristiques de la traduction chez les archées

La traduction est le processus permettant la synthèse d'une chaine polypeptidique à partir d'une séquence d'ARN. Elle est scindée en trois étapes : (i) l'initiation, c'est la reconnaissance par le ribosome d'une séquence spécifique sur l'ARNm, (ii) l'élongation, la

synthèse de la protéine, à partir de la séquence d'ARN, par assemblage d'acides aminés apportés par les ARNt, et (iii) la terminaison, lorsque le ribosome reconnaît le codon stop. La traduction est assurée pendant toutes ces étapes par les protéines ribosomales et les protéines de la traduction, comme les facteurs d'initiation (IF pour les bactéries, eIF pour les eucaryotes et aIF pour les archées), les facteurs d'élongation (EF pour les bactéries, eEF pour les eucaryotes et aEF pour les archées) et enfin les facteurs de terminaison (RF pour les bactéries, eRF pour les eucaryotes et aRF pour les archées). Quelles sont les caractéristiques de la traduction que les archées partagent avec les deux autres domaines ?

Lors de l'étape de l'initiation, chez les bactéries le ribosome se fixe sur une séquence spécifique : la séquence Shine-Dalgarno. Il a été montré que les archées ont des séquences de fixation du ribosome de type Shine-Dalgarno [21]. En ce qui concerne les protéines de la traduction, certains facteurs d'initiation sont communs aux archées et aux eucaryotes, et d'autres facteurs d'élongation sont partagés par les trois domaines [22].

Il semblerait donc que l'appareil de traduction archéen soit un hybride entre les deux appareils de traduction bactérien et eucaryote.

Ainsi, cette grande hétérogénéité remet en cause le lien de parenté entre les archées et les eucaryotes. C'est un sujet de débat très actif, certains favorisent l'hypothèse selon laquelle les archées et les eucaryotes ont un ancêtre commun, ce qui fait des archées nos lointaines cousines, tandis que d'autres favorisent l'hypothèse selon laquelle les eucaryotes ont un ancêtre archéen. Si l'on considère la forte hétérogénéité des mécanismes moléculaires qui existe au sein des archées ainsi que leur capacité d'adaptation, il est très probable que les eucaryotes aient un ancêtre archéen.

1.2 La réplication chez les archées

1.2.1 Les trois étapes conservées : la pré-initiation, l'initiation et élongation

Les trois domaines du vivant répliquent leur ADN de manière semi-conservative et bidirectionnelle [23, 24]. Diverses protéines sont impliquées : les protéines de la chromatine qui participent à sa protection et sa structure, et les facteurs de la réplication qui permettent la synthèse de l'ADN.

On distingue trois étapes :

- la formation du complexe pré-réplicatif implique la reconnaissance des séquences d'ADN à partir desquelles la synthèse de l'ADN est initié. Ces séquences d'ADN sont appelées

origine de réplication, elles sont reconnues par les protéines d'initiation qui vont ainsi recruter l'hélicase réplicative responsable de l'ouverture de la molécule d'ADN servant de matrice.

- l'initiation de la réplication qui correspond au recrutement du complexe réplicatif contenant notamment la/les ADN polymérase(s) et les primases responsables de la synthèse des amorces ARN nécessaires aux ADN polymérases pour la synthèse de l'ADN.

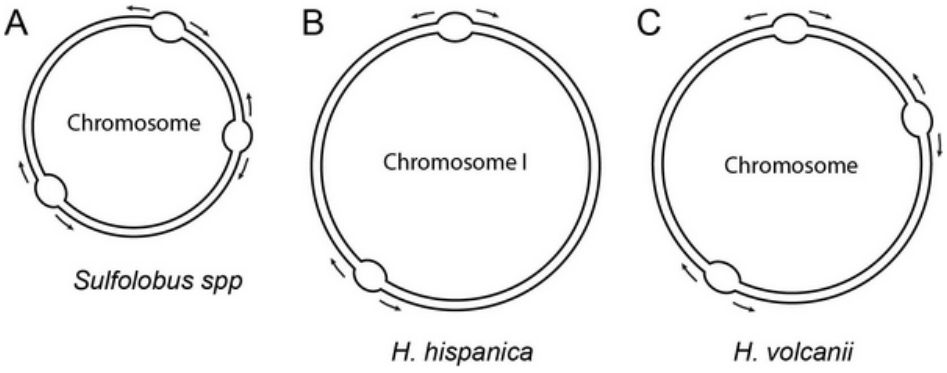
- l'élongation de la réplication qui correspond à la progression de l'ADN polymérase sur le brin matrice et donc à la formation du brin d'ADN néosynthétisé complémentaire. La progression de la synthèse est unidirectionnelle de 5' en 3', de sorte qu'un brin est synthétisé de façon continue (brin sens), l'autre de façon discontinue (brin antisens). La synthèse de l'ADN à partir du brin discontinu conduit à la formation de fragments d'ADN appelés fragments d'Okazaki qui sont ensuite maturés pour permettre l'obtention d'un brin d'ADN continu : les amorces d'ARN en début de chaque fragment sont excisées ou déplacées, une ADN polymérase synthétise l'ADN pour combler la région puis une ADN ligase lie les fragments d'Okazaki entre eux {Barry, 2006 #24;Henneke, 2012 #766}.

1.2.2 Les origines de réplication et leurs protéines associées

Les origines de réplication uniques ou multiples

Les origines de réplication chez les archées sont généralement des séquences riches en A-T qui ont des séquences conservées : les ORB (Origin Recognition Box). Il existe une énorme diversité concernant leurs nombres en fonction des phylums mais aussi des organismes. Les crénarchées possèdent plusieurs origines de réplication. Au sein des euryarchées on observe deux types d'organismes : les halobactériales qui ont généralement plus d'une origine de réplication, et les autres qui possèdent une origine de réplication unique (Figure 5 et Tableau 2). La différence entre le nombre d'origines de réplication chez les bactéries et chez les eucaryotes a longtemps été expliquée par la taille de leur génome. Or, les archées possèdent un génome circulaire et de taille comparable à celui des bactéries. A quoi servent donc ces origines de réplication multiples chez les archées ? Des études portant sur l'espèce *Haloarcula hispanica* ont montré qu'elle possède deux origines de réplication [26], que la présence d'au moins une origine de réplication est nécessaire pour la viabilité [26] et enfin que chaque origine semble présenter une régulation propre, ce qui suggère un niveau de complexité et de coordination élevé entre les différentes origines de la réplication. D'autres études menées chez l'archée *Haloferax mediterranei*, une halobacteriale, ont montré qu'elle possédait trois origines de réplication [27] et que la délétion des trois origines de réplication conduisait à l'activation d'une origine dormante [28]. D'autre part des études réalisées chez *Haloferax volcanii,* un autre organisme appartenant aux *halobacteriales,* ont montré qu'elle

possède trois origines de réplication [27, 29, 30], que la délétion de toutes les origines de réplication était non seulement viable mais qu'elle accélérait la croissance de l'archée [29]. Ceci pourrait être dû à l'activation d'origines de réplication dormantes chez *H. volcanii* qui ne sont pas détectées ou à une initiation de la réplication à partir d'intermédiaires de recombinaison, cette dernière hypothèse étant celle favorisée par les auteurs. L'ensemble de ces données suggère que les mécanismes régissant la réplication en présence et en absence d'origines de réplication est différent. De plus, la réplication semble hautement coordonnée entre les différentes origines de réplication. Puisque le nombre d'origines de réplication diffère d'organismes à organismes et que leur présence n'est pas forcément un caractère essentiel pour certains organismes : pourquoi certaines archées possèdent plusieurs origines de réplication et d'autres une seule, pourquoi la présence de ces origines de réplication non essentielles a-t-elle été sélectionnée au cours du temps? La diversité du nombre d'origines de réplication peut s'expliquer par le fait que les archées peuvent vivre dans des environnements extrêmes : on peut imaginer qu'un nombre élevé d'origines de réplication y est un avantage pour répliquer plus rapidement. Le caractère essentiel des origines de réplication a été déterminé en condition de laboratoire. Mais il est possible que dans leur habitat naturel, chez certaines archées, ces origines de réplication qui seraient spécifiques et très actives lors de variations de leur environnement aient été sélectionnées au cours du temps.



<u>Figure 5.</u> Les origines de réplication chez un groupe représentatif des euryarchées, les halobactériales, et chez un groupe représentatif des crénarchées, les sulfolobales.
D'après [31].

Phylum	Organismes	Nombre d'origines de réplication	Références
Euryarcheota	*Pyrococcus abyssi*	1	[32]
	Haloferax volcanii	3	[27, 29, 30]
	Haloferax mediterranei	2	[27]
	Archaeoglobus fulgibus	1	[33]
	Halobacterium sp. NRC1	4	[34, 35]
	Haloarcula hispanica	2	[26]
	Methanothermobacter thermoautotrophicus	1	[36]
Crenarcheota	*Sulfulobus acidocaldarius*	3	[37]
	Sulfulobus solfataricus	3	[37]
	Sulfulobus islandicus	3	[38]
	Pyrobaculum calidifontis	4	[39]
	Aeropyrum pernix	Au moins 2	[40]
Thaumarcheota	*Nitrosopumilus maritimus*	1	[27]

Tableau 2. Tableau représentant le nombre d'origines de réplication chez les archées. D'après [31].

Les protéines Cdc6/Orc1

Les origines de réplication sont reconnues par des protéines qui se fixent à ces séquences et recrutent l'hélicase réplicative. Chez les bactéries, cette protéine est DnaA. Chez les eucaryotes, c'est un complexe protéique appelé Orc (origin recognition complex) qui est constitué de six protéines. Tandis que chez les archées il s'agit des protéines Cdc6/Orc qui présentent des homologies à la fois avec la protéine eucaryote Orc, et avec la protéine Cell division cycle 6 (Cdc6) qui est impliquée dans le recrutement de l'hélicase réplicative. Leurs séquences codantes sont souvent proches des origines de réplication, elles s'y fixent en reconnaissant les séquences conservées ORB de ces origines de réplication (études menées chez l'euryarchée *Pyrococcus abyssi* et la crenarchée *Sulfolobus solfataricus*) [41-43].

La plupart des archées possèdent une à trois protéines homologues des protéines Cdc6/Orc, sauf les *Halobactériales* qui en possèdent entre trois et seize en fonction des espèces, certaines comme les *Methanococcales* n'en ont aucune [24]. De façon surprenante, ces protéines ne sont pas nécessairement essentielles à la viabilité. La crénarchée *S. islandicus* possède trois gènes *cdc6/orc* : aucun des trois n'est essentiel à la viabilité cellulaire mais le triple mutant n'est pas viable. L'euryarchée *Halobacterium. sp. NRC1* code pour dix protéines homologues de Cdc6/Orc, dont seulement deux d'entre elles sont essentielles pour la viabilité cellulaire. L'euryarchée *H. volcanii* possède seize gènes *cdc6/orc*, au moins trois d'entre eux ne sont pas essentiels [44]. Parmi ces seize gènes, leurs séquences ne sont pas nécessairement situées à côté des origines.

Comme le nombre de protéines Cdc6/Orc1, le nombre d'origines de réplication varie au sein des archées et la corrélation entre le nombre d'origines de réplication et le nombre de leurs protéines d'initiation associées n'est pas clair. En effet, chez *Halobacterium sp. NRC1,* il existe quatre origines de réplication et quatre gènes *cdc6/orc1* qui se fixent spécifiquement à leur origine de réplication. Chez *S. solfataricus,* il existe trois gènes *cdc6/orc1* et trois origines de réplication. Deux de ces origines de réplication sont reconnues par les protéines Cdc6/Orc1, la troisième origine est reconnue par une autre protéine de l'initiation : la protéine Whip (cf ci-dessous[40]). Tandis que l'archée *H. volcanii* possède quatorze protéines Cdc6/Orc et seulement trois origines de réplication. Ainsi, il existe une grande complexité concernant la régulation de l'initiation qui est encore mal comprise. Quel est donc le rôle de ces protéines et pourquoi sont-elles si nombreuses chez certaines archées ?

Les protéines WhiP

Les protéines WhiP sont des protéines impliquées dans l'initiation de la réplication. Elles sont structurellement proches des protéines eucaryotes Cdt1 et formeraient une structure en tandem « d'hélice ailée ». Elles ne sont présentes que chez les crénarchées et plus spécifiquement, les *desulfoloballes* et les *sulfolobales* chez qui des homologues des protéines WhiP ont été retrouvées. Cette protéine assurerait la même fonction que la protéine Cdc6/Orc en activant les origines de réplication. Chez les *sulfolobus,* il existe trois protéines Cdc6/Orc, les protéines Orc1-1 et 2 ainsi que la protéine WhiP peuvent se fixer aux trois origines de réplication, Orc1-3 se fixe à *oriC1* et à *oriC3*. De plus, il a été montré que les protéines WhiP se fixent sur des séquences spécifiques de la troisième origine de réplication : *oriC3* [40]. Il est important de noter que la séquence du gène *whip* est proche de l'origine de réplication comme c'est le cas pour certaines protéines Cdc6/Orc vis à vis de leur origine de réplication. Des études utilisant des techniques d'empreinte d'ADN ont montré qu'il existait une interaction entre le complexe Cdc6/Orc1, les protéines WhiP et les origines de réplication. Même si les trois protéines Cdc6/Orc se fixent sur la troisième origine de réplication, aucune n'est essentielle pour l'activer. Ce n'est pas le cas pour WhiP : elle est essentielle pour l'activation d'*oriC3* [38]. La redondance de la fonction d'activation des origines de réplication portée par les protéines Cdc6/Orc1 et les protéines WhiP renforce l'idée que le rôle des protéines Cdc6/Orc ne serait pas limité à l'initiation de la réplication.

L'hélicase réplicative MCM et le complexe GINS

Après reconnaissance des origines de réplication, une hélicase réplicative est chargée sur les origines de réplication pour séparer les deux brins d'ADN. Chez les bactéries ce rôle est assuré par une seule protéine : l'homohexamère DnaB. Chez les eucaryotes, cette hélicase réplicative est formée par le complexe d'hétérohexamère MCM, qui est assisté et

activé par d'autres protéines réplicatives : les protéines Cdc45 et le complexe hétérotetramérique GINS. Chez les archées, ce rôle serait assuré par un homocomplexe multimérique de type MCM, similaire aux eucaryotes, c'est le cas chez l'archée *T. kodakaraensis* [45]. En effet la plupart des archées possède dans leur génome un homologue de MCM. Chez les *Methanococcales,* il existe plusieurs homologues de MCM. Cependant, comme pour les homologues des protéines Cdc6/orc1 des halobactéries, la plupart ne sont pas essentielles comme c'est le cas chez l'archée *T. kodakaraensis* [46] et très probablement chez l'archée *M. maripaludis* [47]. L'activité hélicase de MCM a été montrée *in vitro* chez plusieurs archées notamment l'archée *Methanothermobacter thermautotrophicus* ΔH [48, 49]. L'interaction *in vivo* des protéines Cdc6/Orc1 avec la protéine MCM a été montrée chez la crénarchée *Sulfolobus solfataricus* [50] et chez l'euryarchée *Methanothermobacter thermautotrophicus* [51]. Chez cette archée, en utilisant la méthode du double hybride, il a été montré que la fixation de Cdc6 ou MCM à l'ADN n'est pas nécessaire pour l'interaction entre les deux protéines [52]. De plus, de façon intéressante, l'activité hélicase de MCM est inhibée en présence d'homologues de Cdc6 [51]. Cependant, des études de recrutement *in vitro* réalisées chez *P. furiosus* ont proposé les protéines Cdc6/Orc comme chargeurs du complexe MCM sur les origines de réplication [53].

Il est possible que les protéines Cdc6/Orc chargent l'hélicase aux origines de réplication mais que d'autres complexes protéiques soient recrutés pour activer l'activité hélicase de MCM, comme c'est le cas chez les eucaryotes avec le complexe GINS et Cdc45. Chez les crenarchées et certaines euyarchées, il existe deux homologues des protéines GINS, qui peuvent s'assembler et former le complexe GINS structuré en hétérotetramère [54, 55]. Chez d'autres archées il n'y a qu'un homologue (formant un homocomplexe) [56] et d'autres n'en n'ont aucun [24]. Cependant, en dépit de la présence de cette protéine GINS, la nature de l'interaction GINS/MCM ainsi que son rôle varie entre archées. En effet, chez *P. furiosus* et *T. kodakaraensis* l'activité hélicase est stimulée par l'interaction avec GINS [45, 57]. Tandis que chez *S. solfataricus* MCM et GINS interagissent *in vitro* sans pour autant activer l'activité hélicase de MCM [54]. Il a également été montré que MCM peut être chargé directement sur la chromatine en l'absence d'origines de réplication.

Une autre protéine assistant l'hélicase réplicative chez les eucaryotes est la protéine Cdc45. Son rôle exact n'a pas encore été élucidé chez les eucaryotes. Il n'existe pas d'homologue direct chez les archées, il existerait cependant une protéine candidate : RecJ. Mais son rôle chez les archées reste à déterminer.

1.2.3 La fourche de réplication

Après l'initiation de la réplication qui a lieu au niveau des origines de réplication et qui fait intervenir les protéines de reconnaissance de l'origine Cdc6/Orc et Whip ainsi que l'hélicase réplicative, c'est l'étape d'élongation de l'ADN. La duplication de l'ADN s'effectue pendant la phase d'élongation au niveau des fourches de réplication. Afin que la machinerie de réplication puisse progresser le long de l'ADN, les problèmes de torsion de l'ADN sont résolus par l'action des topoisomérases. L'hélicase réplicative (MCM) dénature les deux brins d'ADN, l'ADN simple brin ainsi formé est protégé par les protéines de protection de l'ADN simple brin : les SSB. Les ADN polymérases (B ou D) vont ensuite synthétiser l'ADN à partir d'amorces d'ARN, générées par les primases (PriS et PriL). Les ADN polymérases synthétisent l'ADN dans le sens 5'->3'. Pour l'un des deux brins de l'ADN la polymérase progresse dans le même sens que la fourche, elle peut synthétiser l'ADN de façon continue ; c'est le brin sens. Pour l'autre brin, l'ADN polymérase ne peut pas synthétiser l'ADN de façon continue, c'est le brin antisens. Pour ce brin antisens, l'ADN est répliqué par fragments : les fragments d'Okazaki. L'assemblage des différents fragments d'ADN nécessite l'excision des amorces, la synthèse de l'ADN puis la ligation des fragments entre eux par une ADN ligase. Le facteur de processivité de l'ADN, PCNA, stabilise les ADN polymérases sur l'ADN tout au long de la réplication. Il sert également de plateforme de recrutement pour certaines protéines comme la ligase. Afin que l'hélicase, les primases et les polymérases agissent sur chacun des brins de façon coordonnée, un pont protéique est formé par la protéine GINS.

Nous détaillerons ici les protéines intervenant lors de l'étape d'élongation de la réplication. D'autres protéines faisant partie de la fourche de réplication et intervenant dans la maturation des fragments d'Okazaki sont détaillées dans le paragraphe suivant (Figure 6).

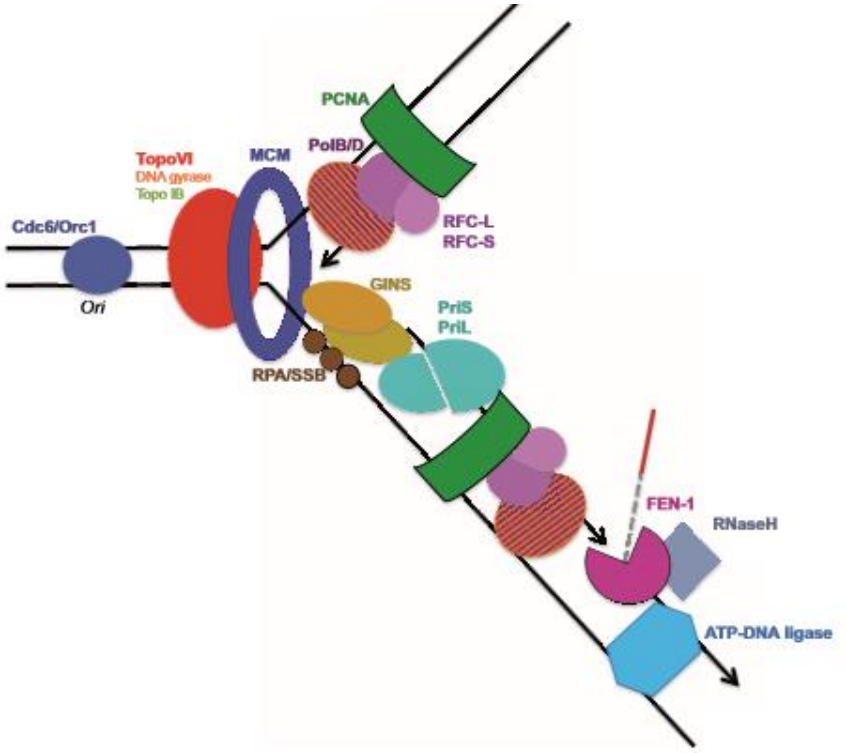


Figure 6. Schéma d'une fourche de réplication archéenne.
La protéine PCNA est représentée en vert, les protéines primases PriS et PriS sont en turquoise, l'ADN polymérase B ou D est représentée en orange striée violet et les protéines SSB sont en marron. D'après [14].

Les topoisomérases

La progression de la fourche de réplication induit des supertours positifs en amont et des supertours négatifs en aval. Les topoisomérases maintiennent un état de surenroulement qui permet à la machinerie de réplication de progresser. Elles sont présentes chez les trois domaines du vivant [58]. Il en existe deux types : les topoisomérases I et les topoisomérases II. Les topoisomérases I relâchent les supertours d'ADN en coupant une liaison phosphodiester sur seulement un des deux brins de l'ADN. Cette liaison lui procure l'énergie nécessaire pour son activité catalytique, elle ne nécessite pas d'ATP. La reverse gyrase en fait partie, elle est notamment capable de produire des supertours positifs au sein d'un ADN circulaire, cette caractéristique nécessite de l'ATP qui lui est procurée par un module hélicase [59, 60]. Les topoisomérases II résolvent les supertours en coupant les deux brins d'ADN pour y insérer un duplex d'ADN. A la différence des types I cette action consomme de l'énergie. Ces dernières sont également impliquées dans les processus de séparation des chromosomes circulaires qui ont lieu après la réplication de l'ADN. La gyrase en fait partie et est capable, à l'inverse de la reverse de gyrase, de produire des supertours négatifs au sein d'un ADN circulaire [61].

Les protéines de protection de l'ADN : les SSB (Single Stranded Binding Protein)

Lors de la réplication les brins d'ADN ne sont pas appariés entre eux, ils sont donc plus « fragiles » car plus exposés aux nucléases. Les protéines SSB ont pour rôle de protéger l'ADN simple brin lors de cette étape. Ce processus de protection est universel, ces SSB forment des homodimères ou homotétramères chez les bactéries [62, 63], elles forment des heterotrimères chez les eucaryotes et portent le nom de RPA [64]. Elles peuvent se fixer à l'ADN par l'intermédiaire de liaison oligonucléotide/oligosaccharide [65], qui sont regroupées dans des domaines structurellement conservés : les OB-fold. Ces OB-fold sont constitués d'un tonneau formé de feuillet β. Chez les archées, il existe une grande diversité sur le nombre d'homologues présents et sur la structure de ces protéines, elles peuvent être rencontrées sous forme de monomère, d'hétérotrimère, d'hétérotétramère. En effet, chez la crénarchée *S. solfataricus,* SSB est monomérique [66], tandis que chez les euryarchées leur organisation semble plus variée mais proche de la structure en hétérotrimère des RPA eucaryotes [67]. Dans le cas où plusieurs homologues sont présents, tous ne sont pas essentiels à la viabilité. En effet chez *H. volcanii,* qui possède trois homologues de RPA, une seule est essentielle [68]. Chez *M. maripaludis,* parmi les trois homologues deux semblent essentielles [69]. D'autres rôles de ces protéines seront développés plus tard.

Les primases

Les primases synthétisent les amorces d'ARN servant d'amorces à l'ADN polymérase sur le brin sens et le brin antisens. Chez les bactéries il s'agit de DnaG, une ARN polymérase ADN-dépendante qui interagit avec l'hélicase réplicative [70]. Cette interaction lui permet d'interagir avec le brin antisens. Chez les eucaryotes, il s'agit d'une enzyme formée par deux sous unités de type ADN polymérase et de deux sous-unités primase ; la sous-unité PriS, portant l'activité catalytique, et la sous unité PriL, régulant son activité. Des homologues des deux sous unités PriS et PriL ont été trouvés chez les euryarchées et chez les crenarchées. Son mode de fonctionnement est le même que celui décrit pour les eucaryotes. Certaines archées possèdent également des homologues de la protéine bactérienne DnaG. Jusqu'à aujourd'hui, les études fonctionnelles réalisées n'ont permis d'identifier qu'une fonction mineure de cette protéine dans la réplication. En revanche, cette enzyme pourrait être un composant de l'exosome qui interviendrait dans la dégradation des ARNs. Chez *M. maripaludis,* celle-ci semble non essentielle pour la survie cellulaire [47], c'est également le cas chez *H. volcanii* [71].

Sa capacité de polymérisation, de combler les trous, font d'elle une bonne candidate pour être également impliquée dans la réparation de l'ADN.

Les ADN polymérases

Les ADN polymérases utilisent l'ADN simple brin comme matrice pour allonger le brin d'ADN complémentaire : elles ont pour rôle de polymériser l'ADN. Elles interviennent au cours des étapes de réplication de l'ADN et peuvent avoir une activité de correction des erreurs mais elles interviennent également dans les processus de réparation de l'ADN.

Les ADN polymérases sont rencontrées chez les trois domaines du vivant et peuvent être classées selon leur séquence en plusieurs familles. Chez les bactéries, les ADN polymérases appartiennent aux classes A, B, C et Y [72]. Chez les eucaryotes, les ADN polymérases appartiennent aux classes A, B et Y. Les ADN polymérases nécessaires pour la réplication appartiennent à la famille B. Toutes les archées possèdent des ADN polymérase de la famille B et la majorité des archées (sauf les crénarchées) possèdent des ADN polymérases de la famille D [73], famille uniquement présente chez les archées [74, 75]. D'autres archées possèdent aussi des ADN polymérases de la famille Y, mais leur rôle dans la réplication n'a pas encore été mis en évidence [76]. Certaines archées ont besoin des deux familles d'ADN polymérases B et D, il a été proposé qu'elles interviendraient préférentiellement sur un brin particulier [77]. La majorité des ADN polymérases de la famille D ont deux sous unités : la sous unités DP1, qui serait impliquée dans la correction des mésappariements [78], et DP2. Les ADN polymérases de la famille B sont très variées. L'ADN polB1 est présente chez les euryarchées et crénarchées (qui possèdent également l'ADN polB2). Elle assure une activité 3'5' exonucléasique permettant de corriger les mésappariements. Chez la majorité des euryarchées étudiées, l'ADN PolB ne serait pas essentielle et pourrait avoir un rôle dans la réparation des dommages de l'ADN, la réplication serait alors assurée par l'ADN polymérase PolD qui est essentielle [47, 79, 80]. Ce n'est pas le cas chez *H. volcanii* chez qui *polB1* est essentiel.

La protéine PCNA

La protéine PCNA est une protéine essentielle à la réplication : elle augmente la processivité de l'ADN polymérase et joue le rôle de plateforme de recrutement d'autres protéines intervenant dans la réplication. Cette protéine est présente au sein des trois domaines du vivant, dans un état d'oligomérisation différent mais conservant la même structure en forme d'anneau, interagissant avec l'ADN en son centre. Chez les bactéries, cette protéine est un anneau homodimère. Chez les eucaryotes, c'est un homotrimère. Chez la majorité des crenarchées, il existe plusieurs homologues de PCNA qui forment un anneau heterotrimèrique [81, 82]. En revanche, la majorité des euryarchées n'ont qu'un homologue de PCNA. Les protéines interagissent avec PCNA *via* un motif PIP (PCNA interacting protein). Il a été montré chez *T. kodakarensis,* une des seules archées possédant deux homologues de PCNA, que PCNA1 et PCNA2 sont capables de se lier aux ADN polymérases B et D. En

revanche, *in vitro* seule PCNA1 stimule la synthèse d'ADN réalisée par les ADN polymérases B et D, PCNA2 stimule uniquement la synthèse de l'ADN polymérase B. De plus aucun mutant n'a pu être obtenu pour le gène *pcna1,* tandis que cela a été possible pour le gène *pcna2* [83]. Ainsi, PCNA1 serait essentiel pour la survie cellulaire et non PCNA2, reflétant un rôle majeur de PCNA1 dans la réplication.

Mais la fonction de PCNA ne se restreint pas à ses interactions avec l'ADN polymérase. Son rôle de plateforme de recrutement serait également utilisé lors de la maturation des fragments d'Okazaki. De récentes études ont proposé un modèle, se basant sur des reconstitutions *in vitro* de ce processus de maturation, dans lequel PCNA serait responsable de la coordination des différentes enzymes (DNA ligases, Pol B1 et Fen1) intervenant dans la maturation des fragments d'Okazaki [84].

Son rôle ne se limite pas à la réplication, chez les eucaryotes et les archées PCNA est une plateforme de recrutement de facteurs essentiels pour la réparation de l'ADN. Il a été montré chez les archées que PCNA interagit avec la nucléase NucS [85, 86], les protéines impliquées dans la réparation des dommages à l'ADN comme NreA/NreB et les protéines TIP dont le rôle n'est pas certain mais qui est supposé être un inhibiteur de PCNA [87]. A travers l'étude du réseau d'interaction de protéines, il a été montré que PCNA interagit avec une protéine jusque-là inconnue; la protéine Pab043, endoQ, qui possède une activité nucléasique [88]. Ces données renforcent un rôle potentiel de PCNA dans la réparation. De plus, chez *H. volcanii,* les cellules qui portent la délétion Δ*nreA* ou la délétion du motif PIP (responsable de l'interaction avec PCNA) de la protéine NreA sont plus sensibles à l'agent dommageant MMC. La protéine NreA serait dépendante de PCNA pour réparer les dommages de l'ADN provoqués par la MMC [89].

Le facteur de chargement de la protéine PCNA : la protéine RF-C (replication factor C)

Le facteur de processivité est chargé sur l'ADN par son facteur de chargement : RF-C (replication factor C). Chez les eucaryotes, RF-C ouvre l'anneau et y introduit la molécule d'ADN pour charger PCNA aux sites d'initiation de la réplication [90]. Chez les bactéries le facteur de processivité est chargé par le complexe-γ de façon similaire [91]. Chez les archées, il est formé de deux sous unités, la sous unité RFCL et la sous unité RFCS. Il a été montré que PCNA peut se charger aux sites d'initiation de la réplication sans facteur de chargement [81, 92, 93]. Néanmoins, la présence de RF-C augmente significativement l'efficacité de chargement de PCNA. Le facteur RF-C archéen peut être organisé en pentamère ou en hexamère formé des sous unités RfcS et RfcL [94, 95]. Ce complexe appartient à la famille des ATPases AAA+. Pour ouvrir et charger PCNA, chacun des sites de fixation de l'ATP doivent être occupés et les sous unités du complexe doivent être associées [96, 97].

1.2.4 La maturation des fragments d'Okazaki

La maturation des fragments d'Okazaki repose sur deux fonctions essentielles : la fonction ligase et la fonction nucléase. Les petites amorces d'ARN sont clivées par l'activité nucléase de la protéine RnaseH chez les bactéries. Chez les eucaryotes sont impliquées les protéines FEN1, Dna2 et RnaseH. Puis une ADN polymérase synthétise l'ADN qui va combler ces 'trous'. Enfin l'ADN ligase lie les fragments d'ADN entre eux pour former un seul brin d'ADN [98].

Chez les archées, il existe deux protéines candidates pour le remplacement de ces amorces. Il peut être assuré par la protéine FEN1 (Flap Endonuclease 1), qui est une protéine homologue de la protéine eucaryote FEN1, toutes les archées en possèdent un homologue dans leur génome. C'est une endonucléase séquence spécifique qui reconnait les queues 5' (flap) non appariées sur l'ADN double brin. Au sein des archées, la fonction de FEN1 n'est pas exactement la même. En effet, chez *H. volcanii,* la délétion de *fen1* est viable mais conduit à un ralentissement de croissance et une réduction de la capacité de la cellule à incorporer des nucléotides dans l'ADN [99]. En revanche, chez une autre espèce halophile, *Halobacterium sp.* NRC-1 *fen1* est un gène essentiel. La RNAse Rnh2 aurait également un rôle lors de cette étape mais le gène *rnh2* n'est pas essentiel chez *H. volcanii.* Il est important de noter qu'il existe cinq copies du gène *RnaseH* chez *H. volcanii.* En revanche, des études réalisées sur *rnh1 et 2* indiquent davantage un rôle de ces protéines dans la réparation de l'ADN que dans la réplication {Meslet-Cladiere, 2007 #687}. De plus, la reconstruction *in vitro* de l'élimination des amorces a permis d'identifier deux voies de dégradation des amorces ; l'une repose sur l'excision des amorces par la protéine Rnh2, l'autre repose sur le déplacement de l'amorce d'ARN par l'ADN polymérase D sans Rnh2. La protéine Rnh2 n'est donc pas essentielle pour cette étape *in vitro* {Henneke, 2012 #766} et il est proposé que RnaseH2 soit en fait impliquée dans la réparation de l'ADN après incorporation de ribonucléotides dans l'ADN.

Toutes les archées possèdent dans leur génome des séquences codant pour la ligase ATP dépendante (LigA). De plus la majorité des halophiles possèdent une autre ligase en plus ; la ligase NAD dépendante (LigN). Chez *H. volcanii*, il a été montré que la délétion individuelle des gènes est viable mais la double délétion n'est pas possible. Les enzymes LigA et LigN partageraient donc une fonction essentielle chez *H.volcanii* [44].

1.2.5 *Haloferax volcanii* comme organisme modèle pour l'étude de la réplication chez les archées

Les caractéristiques générales d'*H. volcanii*

H. volcanii est l'organisme modèle choisi pour cette étude. C'est une archée appartenant au phylum des euryarchées. En dépit de son nom, cette archée n'a pas été isolée dans un volcan mais dans la mer morte et dans les grands lacs salés. Le nom de volcanii lui a été donné en hommage à M. Volcani dont les travaux de recherche ont porté sur les halophiles. Le génome *d'H. volcanii a* été publié en 2010 [100]. Elle possède les caractéristiques suivantes :

Les caractéristiques d'intérêt pour cette étude

- halophile stricte; elle pousse à de très hautes concentrations en NaCl (4M)
- généralement aérobe
- mésophile qui croit à 45°C à pH 7,5

Autres caractéristiques

- de couleur rougeâtre due aux pigments caroténoïdes, qui auraient un rôle dans la protection contre les UVs
- pléiomorphe, forme de disque plat ou bâtonnet
- polyploïde, entre 13 et 18 copies de son génome en fonction de la phase de croissance
- un chromosome principal de 2,848 Mb, trois méga plasmides pHV1, 3, 4 (85, 438, 636 kb, respectivement) et un plasmide plus petit pHV2 (6,4 kb) [100]
- trois origines de réplication sur le chromosome majeur [29, 30] et trois autres localisées sur les plasmides

Outils génétiques et imagerie *in vivo*

Divers protocoles de manipulation génétiques ont été mis en place chez cette archée, dont le développement de la technique du pop-in / pop-out (cf matériel et méthode) qui repose sur l'utilisation d'un même marqueur (*pyrE2*) de sélection et de contre sélection pour intégrer au locus chromosomique une séquence d'ADN donnée ou délèter un gène [101]. D'autres marqueurs de sélection ont été développés chez *H. volcanii* comme *leuB* et *trpA* disponibles sur plasmides réplicatifs élargissant les possibilités d'études des différents gènes et protéines chez *H. volcanii* [102]. D'autre part *H. volcanii* est un organisme généralement aérobe, cette caractéristique est importante pour la suite de notre étude, puisque pour la microscopie à fluorescence nous utilisons la GFP, dont la maturation du chromophore est dépendante de l'oxygène. Ces caractéristiques rendent sa manipulation en laboratoire aisée et font d'elle un organisme modèle pertinent.

1.3 L'imagerie

L'imagerie peut être réalisée au moyen de deux types principaux de microscope : le microscope électronique et le microscope photonique. Bien que les deux types de microscopie soient construits selon le même principe et qu'ils partagent les mêmes grands principes tels que la transmission, la réflexion, la diffusion; leur principe de fonctionnement et leur utilisation sont très différents. Dans ce chapitre nous détaillerons les différents outils de microscopie et leurs techniques associées qui ont été utilisés pour notre étude ainsi que ceux qui sont disponibles pour observer les archées. Nous détaillerons à la fois les méthodes de microscopie permettant la détection des cellules et de leurs structures, puis celles permettant la détection des processus biologiques (comme la réplication et la réparation) qui ont lieu au sein de ces cellules, et enfin celles permettant l'étude de la dynamique de ces processus biologiques *in vivo.*

1.3.1 Microscopie à contraste de phase et contraste d'interférence pour observer les archées

L'observation des microorganismes *in vivo* se fait au moyen de la microscopie photonique qui repose sur l'utilisation des photons. Elle possède une résolution moins grande que celle de la microscopie électronique, mais présente l'avantage de pouvoir imager les microorganismes vivants sans traitement des cellules. Afin de localiser les cellules qui sont dans le plan focal, deux techniques de microscopie optiques sont possibles : la microscopie à contraste interférentiel (DIC) et la microscopie à contraste de phase. Ces deux techniques permettent d'observer les cellules sans nécessiter aucun traitement (coloration…). Elles reposent sur des principes d'optique différents mais génèrent des images semblables.

La microscopie à contraste de phase permet de transformer des différences d'indice de réfraction en différence de phase. Lorsqu'un échantillon est imagé par un microscope de phase, une onde lumineuse traverse les différentes structures de l'échantillon. Comme les cellules ne possèdent pas le même indice de réfraction que le milieu dans lequel elles sont imagées, cette onde lumineuse est ainsi diffractée par ces cellules. Grâce aux deux anneaux de phase présents dans le microscope, cette lumière diffractée est déphasée. Ce phénomène de contraste de phase est observable par des niveaux d'intensités différents. En revanche,

cette méthode a l'inconvénient de créer un halo autour des structures qui peuvent rendre la détection de petits objets plus difficile.

La microscopie à contraste interférentiel repose sur l'interférence de deux faisceaux d'une même onde lumineuse qui traverse un échantillon. Il existe plusieurs types de microscopes à contraste interférentiel, le plus utilisé est le microscope interférentiel selon Nomarski, qui par abus de langage désigne le microscope à contraste interférentiel. Lorsqu'un échantillon est imagé par un microscope à contraste interférentiel, l'onde lumineuse est d'abord dédoublée par un prisme puis les deux faisceaux traversent l'échantillon et les cellules. Ces rayons sont diffractés par les cellules, car à cet endroit l'échantillon n'a pas la même épaisseur, et sont déphasés. Un deuxième prisme fait interférer les deux faisceaux, générant un contraste artificiel. La méthode de contraste interférentiel selon Nomarski génère ainsi des images en pseudo relief.

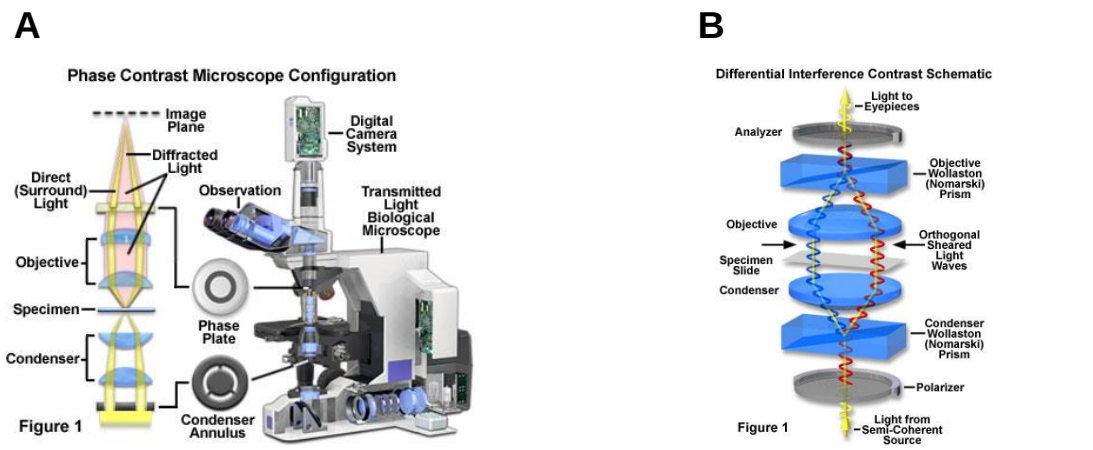


Figure 7. Les microscopes à contraste de phase et contraste interférentiel.
A. La microscopie de contraste de phase. D'après [103]. **B.** La microscopie à contraste interférentiel [104].

Les premières images des archées

L'Une des premières images d'archées a été obtenue en 1975 par l'équipe de Larsen H, ce qui a constitué une avancée majeure dans le monde de la microscopie des archées. Cependant, ces travaux ont eu lieu avant la découverte des archées par Carl Woese en 1977. Ainsi cette étude ne se réfère pas à *Haloferax volcanii* mais à *Halobacterium volcanii* (qui sera rebaptisée *Haloferax volcanii* par la suite) et décrivent ainsi la morphologie de ces cellules comme celles d'une bactérie halophile. Ils ont néanmoins noté que ces cellules poussaient à des concentrations en sel deux fois inférieures à celles habituellement rencontrées chez les halobactéries. Ces acquisitions ont été réalisées sur des cellules d' *H. volcanii* au moyen d'un microscope de phase. Ils ont pu montrer qu'une cellule d'*H. volcanii* mesure 2 µm de largeur et 2 µm d'épaisseur. Ces études ont permis de mettre en évidence le caractère hautement

pléiomorphique de ces cellules, aussi bien entre cellules provenant de culture exposées à différentes conditions qu'au sein d'un même échantillon. Deux principales formes d'*H. volcanii* ont été identifiées : une forme en disque plat et une forme en coupe qui semblait être obtenue lorsque les cellules étaient cultivées dans des conditions optimales [105]. Le caractère pléiomorphique des archées a également était rencontrée chez *Methanosarcina mazei* [106] ou encore *Thermoplasma acidophilum* [107]. Cette différence de morphologie par rapport aux bactéries a été rencontré chez une autre haloarchée *Haloquadratum walsbyi* dont les cellules sont plates et rectangulaires [108, 109]. D'autres morphologies ont été établies chez les archées méthanogènes, notamment chez l'archée *Methanocaldococcus jannaschii* [110] et chez les crenarchées, notamment l'archée hyperthermophile *Pyrolobus fumarii* [111].

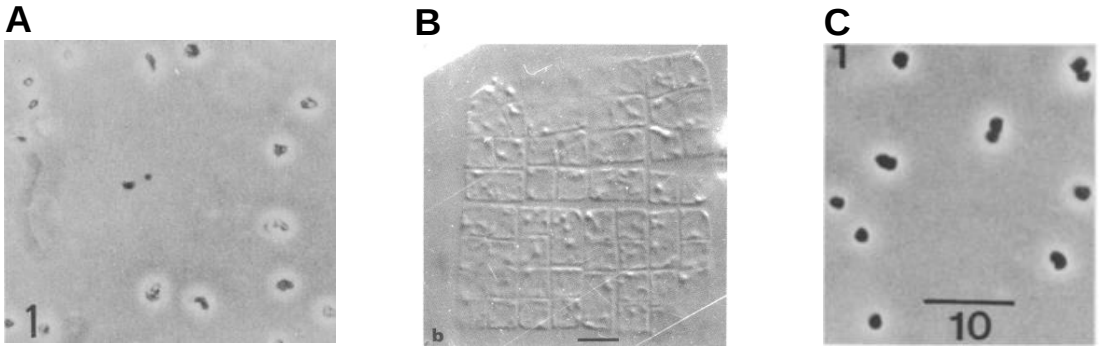


Figure 8. Images en microscopie optique.
A. Image d'*H. volcanii* en contraste de phase, le grossissement est de 1000, d'après [105]. **B.** Image d'*Haloquadratum walsbyi* en contraste de Nomarski. L'échelle est de 5 µm. D'après [108]. **C.** Image de *Methanocaldococcus jannaschii* en contraste de phase. L'échelle est de 10 µm. D'après [110].

1.3.2 La microscopie électronique pour étudier les structures des cellules

A l'échelle de la cellule, la microscopie optique est suffisante. En revanche, à l'échelle des structures intracellulaires par exemple, il est nécessaire d'utiliser un outil permettant d'obtenir une meilleure résolution : la microscopie électronique. Bien que ne permettant pas de réaliser des études sur des cellules vivantes, la microscopie électronique est un outil très utilisé en biologie car elle permet de visualiser à un niveau de détail plus précis les structures des tissus organiques, des enveloppes cellulaires ou de visualiser certains virus. En effet ses limites de détection sont plus basses (peut aller jusqu'à 50 pm) que celles de la microscopie optique (≈ 200 nm). Ceci est permis car elle repose sur l'utilisation d'électrons et non de photons. Il existe deux types principaux de microscopie électronique : la microscopie

électronique à transmission et la microscopie électronique à balayage. Dans le cas de la microscopie en transmission, le faisceau d'électrons traverse l'échantillon. Tandis que dans le cas de la microscopie à balayage, le faisceau d'électrons est réflichi par l'échantillon. Il existe différents traitements des échantillons biologiques pour observer différentes structures des microorganismes.

La microscopie électronique a été employée pour déterminer la structure de l'enveloppe cellulaire mais également pour étudier les phénomènes de division cellulaire. Ainsi la structure de la paroi des archées diffère de celles bactéries, en effet grâce à une approche par microscopie électronique et test enzymatique il a été montré que la plupart des archées (excepté les méthanobactériales et les haloccocales) n'ont pas de paroi rigide constituée de peptidoglycanes [11, 112]. En revanche, à l'instar des bactéries, les archées possèdent une S-layer, constituée de glycoprotéines, pour se protéger des agressions exogènes (variations de pH, stress osmotique [12, 113]).

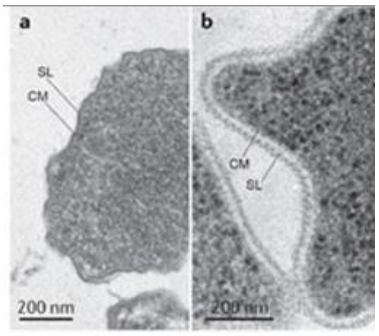


Figure 9. Images d'enveloppe cellulaire en microscopie électronique.
L'échelle est de 200 nm. La couche S (S-layer) est fléchée en SL et la membrane cellulaire est fléchée en CM. **A.** Section ultra fine de l'euryarchée *Methanocaldococcus villosus.* **B.** Section ultra fine de la crénarchée *Metallosphaera prunae.* D'après [12].

L'utilisation de la microscopie électronique et plus particulièrement de la cryotomographie, permettant de préserver la morphologie et la structure des échantillons, a permis de mieux comprendre comment s'opère la division cellulaire chez les archées, au moyen de quels acteurs et de quels complexes. Il a été montré que les archées appartenant au groupe des euryarchées possèdent des homologues des protéines eucaryotes Ftsz qui participent à la division cellulaire [114, 115]. En revanche, les crénarchées ne possèdent pas d'homologues des protéines Ftsz. Or, il a été montré par cryotomographie électronique que les archées du groupe crénarchées (excepté les thermoprotéales) possèdent un complexe de type Escrt-III qui pourrait intervenir dans la formation de la « ceinture » de division cellulaire [116]. Cette étude a été réalisée sur des cellules à différents stades de division, elles ont été préalablement synchronisées et triées par cytométrie de flux avant de les imager par microscopie électronique.

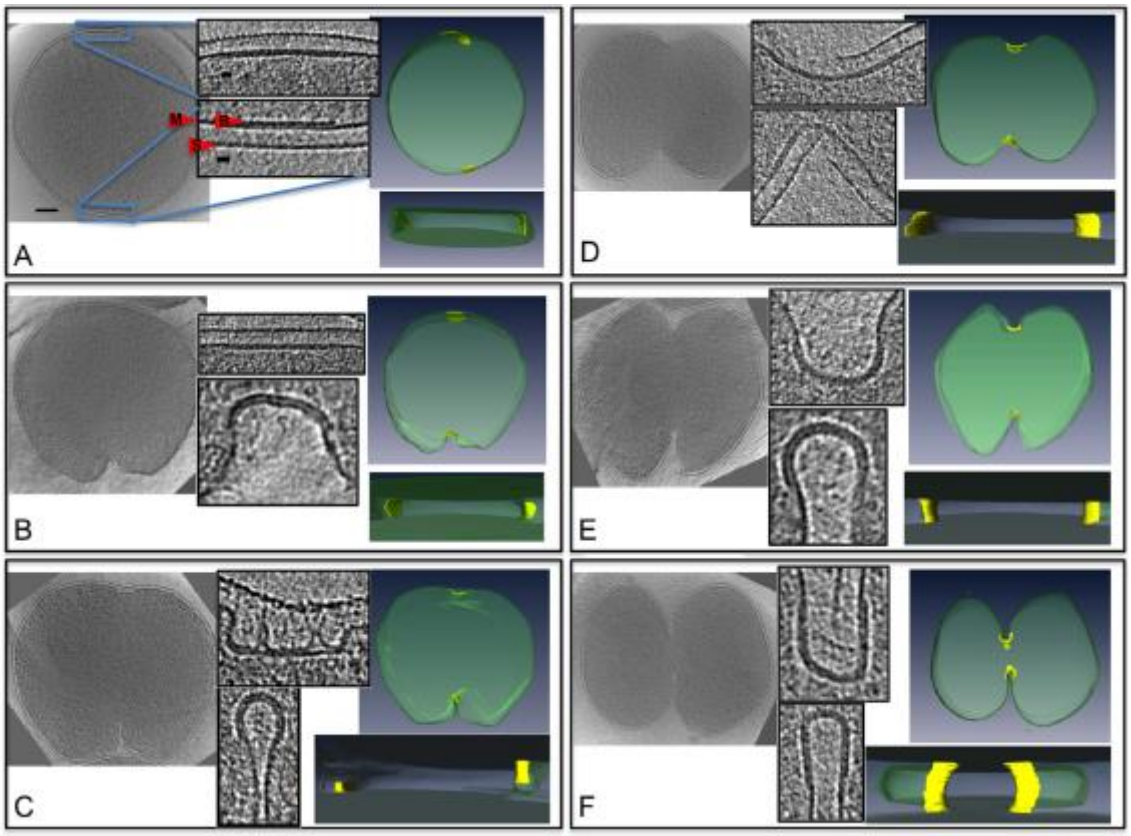


Figure 10. Images en cryotomographie des cellules de *sulfolobus.*
L'échelle est de 200 nm pour les images des cellules et de 20 nm pour les sélections. En jaune est représenté le complexe ESCRT-III. De A à F sont représentés différents états de la division cellulaire. D'après [116].

1.3.3 Etat de l'art sur la microscopie à fluorescence

En microscopie optique « classique », la résolution spatiale du microscope est limitée par sa diffraction : les objets visibles ont une taille apparente de 250 nm. Cette résolution est suffisante pour observer les cellules mais ne permet pas d'étudier la dynamique des processus biologiques fondamentaux. Le développement de la microscopie à fluorescence a permis d'étudier ces processus biologiques dans la cellule vivante. Il existe plusieurs sources de signal de fluorescence : l'autofluorescence des cellules, l'immunomarquage et l'utilisation de molécules fluorescentes comme la GFP. En effet, toutes les cellules possèdent une fluorescence naturelle. Celle-ci peut provenir des acides aminés tryptophane et tyrosine lorsque la cellule est excitée dans l'UV, ou des flavines ou NADH lorsque la cellule est excitée dans le visible. Cependant, cette fluorescence reste très faible. C'est pourquoi divers marqueurs fluorescents, couplés par exemple aux molécules d'ADN ou aux protéines, sont utilisés.

En microscopie à fluorescence, la détection de ces molécules et les acquisitions peuvent se faire selon deux modes : par microscopie en champs large et par microscopie par balayage.

Un exemple de la microcopie en champ large : l'épifluorescence

En ce qui concerne la microscopie en champs large, tout le champ est éclairé simultanément et l'acquisition est réalisée par l'intermédiaire d'une caméra. Afin d'obtenir une bonne qualité d'image sur des échantillons peu fluorescents, les caractéristiques de la caméra sont importants. En effet, celle-ci doit posséder un rendement quantique élevé qui lui permettra de détecter des signaux de fluorescence faibles (environ 100 photons). L'épifluorescence est une technique de microscopie qui utilise ces caractéristiques. Ce type de microscope présente l'inconvénient de ne pas être résolutif en z : la fluorescence de tous les plans de l'échantillon est collectée. De plus il n'est capable d'exciter des échantillons qu'à faible profondeur (environ 100 µm).

Un exemple de la microscopie par balayage : le confocale

Dans le cas de la microscopie par balayage, l'échantillon est balayé par le faisceau laser point par point. La fluorescence est ainsi collectée pour chaque point par un photomultiplicateur qui reconstitue l'image de l'échantillon. Cette technique permet de réduire le volume d'excitation et donc d'améliorer la netteté de l'image en augmentant le rapport signal collecté et bruits de fond. Cette méthode est utilisée pour la microscopie confocale et biphotonique. Pour notre étude, nous avons utilisé la microscopie confocale. A la différence de l'épifluorescence, ce type de microscope présente l'avantage d'être résolutif en z. En effet l'utilisation de deux trous sténopéïques pour l'émission et la détection restreignent la prise de vue à un seul point au plan focal, ainsi seule la fluorescence du point situé au plan focal est collectée. De plus, en excitant l'échantillon point par point, le volume excité est réduit et le rapport signal/bruit est augmenté, ce qui améliore la qualité de l'image [117]

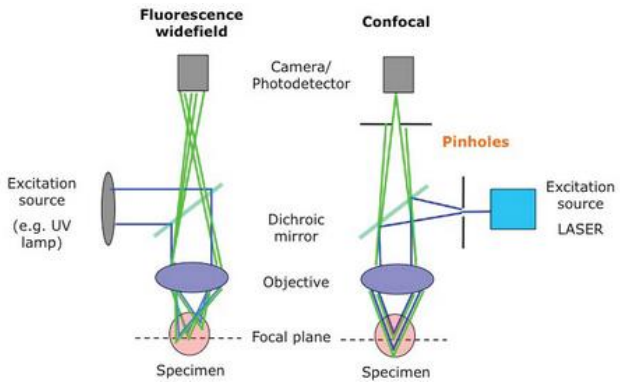


Figure 11. Trajets optiques des deux types de microscopie: par champs large et confocale.
D'après http://www.biotek.fr/fr/resources/articles/development-of-the-optical-microscope.html

L'épifluorescence et le confocale permettent d'étudier la localisation spécifique des protéines fluorescentes. Des techniques ont été développées dans le but d'étudier la dynamique des protéines fluorescentes, comme le FRAP (FRAP : Fluorescence Recovery After Photobleaching) ou l'interaction de protéines entre elles, comme le FRET (FRET : Fluorescence Resonance Energy Transfer).

1.3.4 La microscopie à fluorescence par immunomarquage ou marquage de l'ADN

Bien que la cellule soit auto-fluorescente, pour détecter les molécules au sein de ces cellules, il est nécessaire d'utiliser des marqueurs fluorescents couplés aux molécules d'intérêt qui vont augmenter le signal de fluorescence. La fluorescence repose sur deux phénomènes : l'excitation et l'émission. Les photons sont absorbés par le fluorochrome à une longueur d'onde donnée et le fluorochrome émet des photons à une autre longueur d'onde. Pour garantir une détection optimale et une bonne qualité d'image, les deux spectres d'excitation et d'émission doivent être suffisamment espacés pour ne collecter que les photons émis par le fluorochrome lors de l'acquisition des images (déplacement de strokes). Il existe trois principaux groupes de fluorochromes : les protéines fluorescentes, les fluorochromes organiques et les fluorochromes synthétiques. Chaque fluorochrome est caractérisé par sa durée de vie à l'état excité, son coefficient d'extinction, son rendement quantique et son photoblanchiment. Il est important de prendre en compte ; le type de fluorochrome et ses caractéristiques, pour choisir le fluorochrome le mieux adapté au système biologique dans lequel il doit fonctionner.

La microscopie à fluorescence par immunomarquage

Une des méthodes employées en biologie pour marquer des structures ou des protéines est la microscopie à fluorescence par immunomarquage qui peut être réalisée par des microscopes à épifluorescence ou par un microscope confocale. Cette technique repose sur le marquage des protéines par des techniques d'immunologie. Il en existe deux types : l'immunomarquage direct et indirect. Dans le cas de l'immunomarquage direct, un anticorps couplé à un fluorochrome se fixe à son antigène : la protéine d'intérêt. C'est le signal de fluorescence émis par l'anticorps primaire qui est détecté par le microscope. Dans le cas de l'immunomarquage indirect, un anticorps primaire reconnait la protéine d'intérêt puis un deuxième anticorps secondaire couplé à un fluorochrome vient se fixer à l'anticorps primaire. C'est alors le signal de fluorescence émis par l'anticorps secondaire qui est détecté par le

microscope. Cette méthode présente l'avantage de pouvoir réaliser des marquages différents au sein d'une même coupe et dans le cas de l'immunomarquage indirect le signal de fluorescence est plus intense car les anticorps secondaires peuvent se fixer à plusieurs endroits de l'anticorps primaire. Néanmoins cette méthode ne permet pas de réaliser des études *in vivo* : la reconaissance par l'anticorps se fait après la fixation des cellules et leur perméabilisation par traitement. Cette technique a été utilisée chez *sulfolobus* en utilisant des anticorps dirigés contre des protéines impliquées dans la réplication de l'ADN et dans la division cellulaire. L'immunolocalisation d'une protéine centrale de la réplication, PCNA1, a été utilisée après synchronisation des cellules pour localiser les réplisomes dans la cellule [118]. L'immunolocalisation du complexe « Endosomal sorting complex required for transport – III » (ESCRT III) a permis de mettre en évidence son rôle dans la division cellulaire et ainsi compléter les études réalisées en microscopie électronique [119].

La microscopie à fluorescence par marquage de l'ADN

L'ADN des cellules peut également être marqué et imagé par microscopie à fluorescence. Il existe plusieurs techniques de marquage de l'ADN : par colorant, par fluorochrome et par incorporation d'analogue marqué de la thymidine.

Plusieurs colorants marquant les acides désoxyribonucléiques sont disponibles comme le colorant DAPI, l'Hoechst et l'acridine orange. Par exemple le DAPI a été utilisé pour marquer l'ADN des cellules *S. acidocaldarius.* Cette coloration, combinée à la technique d'éléctrophorèse en 2 dimensions et de cytométrie en flux, a permis de mettre en évidence chez Orc1-1 le rôle de la fixation et de l'hydrolyse de l'ATP dans l'activation et la régulation des origines de réplication [38]. Ces colorants peuvent également être utilisés *in vivo,* et ont permis de mettre en évidence chez *H.volcanii* un processus dynamique : la compaction de l'ADN [120]. Chez les eucaryotes le noyau comprenant l'ADN est séparé du cytoplasme par une enveloppe nucléaire, tandis que chez les archées le nucléoide est compris dans le cytoplasme. La compaction du nucléoide chez les archées est un processus qui peut s'assimiler à l'état de condensation de l'ADN chez les eucaryotes lorsque l'ADN n'a pas besoin d'être répliqué ou transcrit, la molécule d'ADN est alors repliée sur elle-même, rendant difficile la fixation de protéines de la réplication ou de la transcription. Quelle est la fonction de la compaction de l'ADN chez les archées ? Pour répondre à cette question, des études de compaction de l'ADN ont été réalisées sur des cellules vivantes d'*H. volcanii* dont l'ADN a été coloré par un agent marquant tel que l'Hoechst ou l'acridine. Il a été montré que la compaction était induite après exposition aux UV et à la phléomycine, qui provoquent tous les deux des cassures double brin (Figure 12) [120]. Les mêmes expériences ont été réalisées sur des cellules dont le complexe Mre11-Rad50 a été muté, complexe requis pour la survie des cellules après exposition aux UVs, conduisant à une réduction de la compaction de l'ADN. Il a

également été observé dans les cellules dépourvues de RadA (impliquée dans la recombinaison homologue), une compaction de l'ADN après exposition aux UVs, mais sous une forme fragmentée.

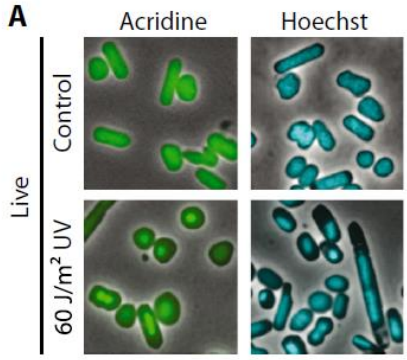


Figure 12. Image de fluorescence d'*H. volcanii.*
L'ADN des cellules contrôles et de celles exposées à 60 J/m^2 a été coloré au Hoechst ou à l'acridine. D'après [120].

La compaction de l'ADN chez les archées peut être expliquée par plusieurs hypothèses. La compaction de l'ADN pourrait être une réponse aux dommages de l'ADN qui favorise la récupération cellulaire en aidant les protéines de réparation de l'ADN à localiser leur cible, ou serait induite lorsque la cellule est en détresse. Elle constituerait une réponse de dernier recours visant à protéger l'ADN des agressions de l'extérieur après dommages de l'ADN. Chez les bactéries, il a été proposé que cette compaction favoriserait la réparation après cassures double brin.

Des études récentes ont montré que les analogues de la thymidine comme l'Edu peuvent être couplé à l'Alexa Fluor en utilisant la chimiclick : l'ADN ayant incrorporé l'Edu est ensuite détecté par microscopie de fluorescence [121]. Cette avancée a permis de localiser l'ADN néosynthétisé chez l'archée *Sulfolobus acidocaldarius.* Cette localisation combinée à l'immunolocalisation de PCNA1 (cf paragraphe précédent) a permis de montrer une localisation subcellulaire de l'ADN néosynthétisé et des réplisomes mais aussi que les replisomes ne sont pas nécessairement proches spatialement de leur origine de réplication [118].

1.3.5 Microscopie à fluorescence par expression de la GFP dans les cellules

La GFP (Green Fluorescent Protein) et ses propriétés de fluorescence

Il existe plusieurs protéines fluorescentes, les plus couramment utilisées sont la GFP et ses variants. Les séquences d'ADN de ces protéines fluorescentes sont fusionnées à celles d'autres protéines d'intérêt, en insérant la séquence d'ADN de la GFP dans le cadre de lecture.

Ceci aboutit à la production d'une seule protéine comprenant la protéine d'intérêt et « son étiquette » GFP. La GFP a été isolée de la méduse *Aequoria victoria* en 1962 [122]. Sa séquence code pour un polypetide de 28 kDa contenant le chromophore. Au sein de ce chromophore a lieu une réaction autocatalytique qui nécessite de l'oxygène. La GFP est excitée dans le bleu à une longueur d'onde maximale de 395 nm et émet dans le vert à 509 nm. Elle est de forme cylindrique, a une longueur de 5nm et un diamètre de 4 nm [123]. Les modifications génétiques de cette protéine ont permis de développer de nombreuses variantes de la GFP. Ces variants ont des spectres d'excitation et d'émission différents, une stabilité à la température et au pH différents, ou un état d'oligomérisation différents qui ouvrent un large éventail d'adaptation aux organismes dans lesquels ils peuvent être exprimés [124]. Dans cette étude et en accord avec des études précédentes [125], nous avons utilisé le variant smRS-GFP qui possède plusieurs mutations le rendant plus soluble (Phe99Ser, Met153Thr et Val63Ala) et une mutation au sein du chromophore (Ser65Thr) qui change son pic maximum d'absorption à 488 nm sans pour autant changer le spectre d'émission de la protéine [126]. Ce variant se rapproche de la protéine Superfolder GFP, dont 60 % est sous forme monomérique et 40 % sous forme dimérique [127] mais il présente l'avantage d'être plus soluble et ses propriétés de fluorescence sont actives chez *H. volcanii.* De plus, afin d'assurer un bon repliement de la protéine fusionnée nous avons utilisé un linker d'une longueur de 18 acides aminés (GEGQGQGQGPGRGYAYRS).



Figure 13. Structure de la GFP.
La structure de la GFP est en forme de tonneau. D'après [123].

Les outils génétiques pour exprimer la GFP dans les cellules vivantes

Le gène de la GFP est un gène rapporteur qui est utilisé pour étudier la localisation des protéines en les fusionnant à leurs gènes et/ou pour étudier le niveau d'expression de ces protéines en fusionnant le gène de la GFP au promoteur du gène d'intérêt. La fusion du gène de la GFP peut se faire du côté aminoterminal ou du côté carboxyterminal, ceci dépend des interactions potentielles de la protéine étudiée et des activités étudiées. Différentes approches génétiques répondant à des questions biologiques différentes peuvent être utilisées pour exprimer la GFP dans les cellules vivantes.

Dans le cas d'une expression transitoire de la protéine d'intérêt fusionnée à la GFP, la séquence d'ADN portant le gène de la GFP fusionné au gène d'intérêt est clonée dans un plasmide réplicatif mais non intégratif dans le génome de l'organisme hôte, sous contrôle d'un promoteur inductible permettant de réguler l'expression du gène fusionné. Ceci garantit des niveaux d'expression généralement élevés.

Dans le cas d'une expression physiologique de la protéine fusionnée, la séquence d'ADN portant le gène de la GFP fusionné au gène d'intérêt est clonée dans un plasmide intégratif. Le gène de la GFP peut être fusionné en amont ou en aval du gène d'intérêt et inséré au locus chromosomique, sous le control du promoteur du gène d'intérêt, pour un niveau d'expression égal au niveau natif d'expression de la protéine. Chez *H. volcanii*, ceci peut se faire par la méthode du pop-in/pop-out (cf partie matériel et méthode [101]).

Les techniques permettant la localisation de la GFP dans les cellules vivantes

En fonction des organismes étudiés et de la protéine étudiée, les conditions d'imagerie diffèrent. En effet, le niveau d'intensité de fluorescence produit sera différent en fonction des protéines étudiées qui ne seront pas exprimées au même niveau. De plus, pour être sûr que la protéine soit synthétisée correctement, à des niveaux d'expression reflétant les conditions natives, il est important d'adapter la séquence d'ADN au biais de codon de l'organisme hôte. Une des seules conditions requises pour la fluorescence de la GFP et de ses variantes est leur besoin d'oxygène : les cellules imagées doivent être imagées en aérobie. Les acquisitions peuvent se faire au moyen d'un microscope à fluorescence en épifluorescence, en confocale et en super-résolution comme le SIM (Microscopie par Illumination Structurée, cf page suivante). Pour notre système, nous avons employé la microscopie à fluorescence de type épifluorescence et confocale. L'utilisation combinée de la microscopie à fluorescence et de la super-résolution chez *H. volcanii* ont permis de mettre en évidence le rôle des protéines CetZ1 dans la différenciation des cellules et dans le contrôle de la forme des cellules (Figure 14 [115]).

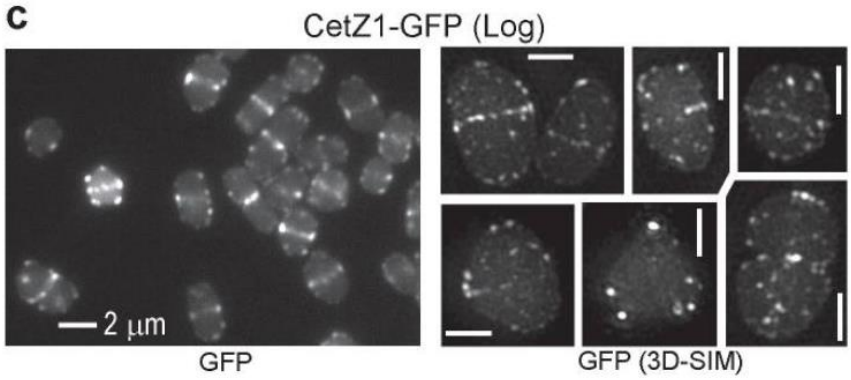


Figure 14. Localisation *in vivo* des protéines CetZ1-GFP.
L'échelle est de 2 µm. GFP : localisation *in vivo* en microscopie à fluorescence. GFP (3D-SIM) : localisation *in vivo* en super-résolution SIM. D'après [115].

Etude de la diffusion de la GFP au sein des cellules

La diffusion de la GFP au sein des cellules peut être étudiée par deux techniques reflétant la dynamique de ces protéines : le FRAP et le FLIP (Fluorescence Loss In Photobleaching). Ces deux méthodes permettent d'étudier le mouvement des molécules au sein des cellules et membranes. Elles reposent toutes les deux sur le photoblanchiment d'une zone définie. Dans le cas du FRAP, la zone est photoblanchie une seule fois et l'on observe le retour de la fluorescence dans cette zone due à l'arrivée de nouvelles molécules fluorescentes non photoblanchies. On obtient une courbe de retour de fluorescence en fonction du temps, grâce à laquelle deux populations de molécules peuvent être identifiées la population mobile et la population immobile. L'équation de la courbe de retour de la fluorescence permet d'estimer la constante de diffusion de la protéine [128]. Dans le cas du FLIP, la zone est photoblanchie mais cette fois à intervalle régulier et l'on observe la réduction du signal de fluorescence dans le reste de la cellule qui est due à la diffusion des molécules fluorescentes photoblanchies.

Une technique de super-resolution *in vivo* : la microscopie par illumination structurée (SIM)

Les limites de la diffraction en optique rendent difficile l'étude des processus biologiques à l'échelle moléculaire. Néanmoins il existe des techniques de super-résolution disponibles en microscopie optique. L'une d'entre elles est le SIM, c'est une méthode de microscopie à champs large qui double la résolution spatiale dans toutes les directions [129], de récentes avancées ont même permis de la diminuer jusqu'à 50 nm [130]. En revanche, cette méthode ne fonctionne pas en autofluorescence. C'est une des méthodes de super-résolution les moins invasives car elle fonctionne avec une puissance d'illumination

relativement faible, elle est donc particulièrement bien adaptée à l'imagerie *in vivo* pour différents systèmes, y compris les archées [115]. Elle est basée sur l'utilisation d'une illumination structurée de l'échantillon combinée à des analyses informatiques utilisant la transformation de fourrier qui permettent de reconstruire l'image 3 dimensions. Le motif de l'illumination structurée est projeté comme une grille sur l'échantillon qui possède lui aussi une structure particulière. L'interférence de ces deux motifs en produit un troisième : les franges Moiré (Figure 15). Des fréquences hautes d'un échantillon sont émises lorsque beaucoup de molécules fluorescentes sont localisées au même endroit, or ces hautes fréquences ne peuvent être collectées. Les structures de Moiré ont une fréquence plus basse qui permet de collecter les informations de ces hautes fréquences et permettent ainsi de générer des images super-résolues de l'échantillon.

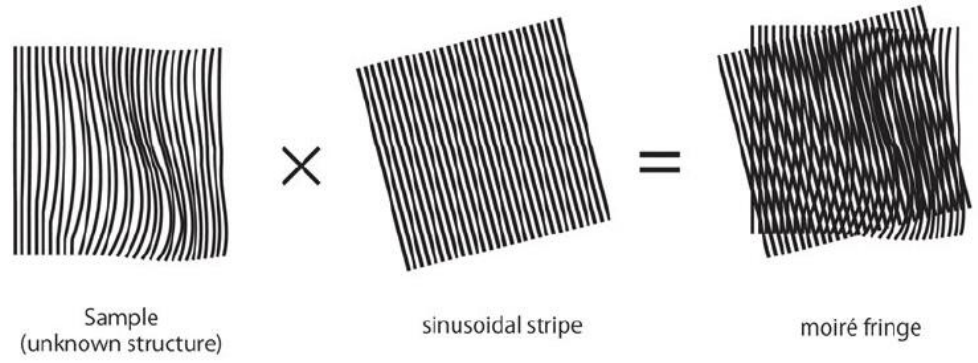


Figure 15. Les différents motifs de structure aboutissant à la structure tridimensionnelle. A gauche est représenté l'échantillon dont le motif de structure est inconnu. Au milieu est représentée la grille projetée (sinusoidal strip) : l'illumination structurée. A droite le résultat de l'interférence de ces deux motifs : le motif de moiré. D'après [130].

1.4 Problématique de la thèse

Nous avons étudié les processus de la réplication et de la réparation de l'ADN chez *H. volcanii* à travers l'étude des protéines suivantes : la protéine Cdc6-Orc1, la protéine RpA2, la protéine FEN1, et la protéine PCNA. Nous développerons dans cette partie la problématique de la thèse ainsi que le contexte dans laquelle cette dernière a été posée chez *H. voclanii.* Ci-joint, un tableau récapitulatif du rôle des protéines étudiées et de leurs homologues chez les bactéries et les eucaryotes.

Fonction assurée	Bactérie	Archées	Eucaryotes
Reconnaissance de l'origine	*Dna A*	Cdc6-Orc1	ORC(1), *ORC(2-6)*, Cdc6
Chargement de l'hélicase	*Dna C*	Cdc6-Orc1	Cdc6
			Cdt1
Protéines de protection de l'ADN simple brin	*SSB*	SSB	*
		RPA1	RPA70, *RPA34, RPA14*
Facteur de processivité de l'AND polymérase	*B clamp*	PCNA	PCNA
Excision des ARN sur les fragments d'Okazaki	*RNase HI*	RNase HII	RNase H2A , *RNase H2B, RNase H2C*
	DNA Pol I	FEN-1	FEN-1
		Dna2?	Dna2

Tableau 3. Les protéines d'intérêt et leurs fonctions associées dans les trois domaines de la vie.
Adapté de [14].

La problématique de cette thèse est la suivante : que peut nous apprendre la localisation intracellulaire des protéines de la réplication sur leur fonction et leur régulation de la réplication et/ou réparation ?

La protéine PCNA ou son homologue bactérien joue un rôle essentiel d'agent de processivité de l'ADN polymérase et de stabilisation du complexe formé par l'ADN, l'ADN polymérase et ses protéines associées chez les archées, les eucaryotes et les bactéries. Elle joue un rôle dans la réparation de l'ADN chez les archées en interagissant avec des protéines intervenant dans la réparation de l'ADN [89]. Elle interagit avec l'ADN en son centre et avec d'autres protéines qui contiennent généralement un motif PIP. Chez *H. volcanii,* il n'existe qu'un seul homologue de PCNA, qui est essentiel. La structure de cette protéine a été résolue, elle se constitue de trois unités qui forment un homotrimère [131, 132].

La protéine FEN1 ou son homologue bactérien, intervient dans la maturation des fragments d'Okazaki chez les trois domaines du vivant. Chez *H. volcanii*, il n'existe qu'un seul homologue mais biochimiquement la protéine RNH2 pourrait assurer la même fonction bien qu'il semble qu'elle soit davantage impliquée dans la réparation de l'ADN. La protéine FEN1 n'est pas essentielle, mais la délétion de son gène entraine un retard de croissance et une réduction de la capacité de la cellule à incorporer des nucléotides [30]. Elle interagit avec PCNA à son extrémité C-terminale par l'intermédiaire d'un PIP. La protéine Rnh2 n'est pas essentielle non plus chez *H. volcanii,* elle interagit également avec la protéine PCNA par son domaine PIP.

Les protéines Cdc6-Orc des archées sont homologues des protéines eucaryotes Cdc6 et ORC qui ont pour rôle respectivement de charger l'hélicase réplicative et de reconnaitre les origines de réplication. Chez *H. volcanii,* il existe seize protéines homologues de Cdc6/Orc, dont au moins trois d'entre elles ne sont pas essentielles [30]. Il est important de rappeler qu'il existe trois origines de réplication chez cette archée. Certains gènes codant pour ces protéines sont adjacents des origines de réplication, ce qui suggère un rôle fonctionnel. Néanmoins il n'est pas nécessaire que les gènes codant pour les protéines Cdc6/Orc soient adjacent des origines de réplication, en effet certaines protéines Cdc6/Orc1 peuvent agir sur plusieurs

origines de réplication. Nous avons choisi d'étudier la localisation de la protéine Cdc6/Orc1 qui n'est pas essentielle à la viabilité mais sa séquence codante est proche d'une origine de réplication.

Le rôle des protéines SSB ne se limite pas à la réplication, chez les bactéries elles sont assimilées à des plateformes permettant de lier les protéines intervenant dans le complexe de réparation, notamment les hélicases comme PriA, aux fourches de réplication [133, 134]. Elles jouent un rôle dans la réplication, la réparation et la recombinaison de l'ADN. Elles participent donc au maintien de la stabilité génomique. Chez les eucaryotes, les protéines assurant la protection de l'ADN sont les RPA. Mais récemment des protéines homologues aux protéines SSB ont été identifiées chez les eucaryotes, leur fonction n'est pas claire mais elles pourraient avoir un rôle dans le maintien de l'intégrité du génome [135, 136]. Chez l'archée *P. furiosus,* il existe une protéine RPA constituée de trois sous unités formant un hétérotrimère, de façon intéressante des études *in vitro* ont permis de montrer que cette protéine est impliquée dans la recombinaison homologue. [67]. Chez une autre archée *S. solfataricus,* la protéine SSB détecte les sites de dommages de l'ADN [137]. Chez *P. abyssi,* l'étude du réseau de protéines intervenant dans le maintien de la stabilité génomique ont permis d'identifier Rpa comme un des acteurs centraux dans ce processus [88]. Chez *H. volcanii,* il existe trois protéines RpA : RpA1, 2 et 3. Rpa1 et 3 sont associées à d'autres protéines : les Rpap. Une seule d'entre elle est essentielle : Rpa2 [68]. C'est celle-ci que nous avons étudiée. Elle intervient dans la protection de l'ADN simple brin et jouerait un rôle dans la réparation de l'ADN : lorsqu'elle est surexprimée, elle confère aux cellules une meilleure résistance aux UVs [138]. En revanche, il est important de noter que généralement le niveau d'expression des protéines RpA est constant. Que peut nous apprendre sa localisation intracellulaire sur sa fonction et la régulation de la réplication/réparation ?

2 MATERIEL ET METHODE

2.1 Milieux et souches d'*Haloferax volcanii*

Les souches d'*Haloferax volcanii* utilisées sont décrites dans le tableau 4.

Une solution d'eau salée (SW) 30 % entre dans la composition de tous les milieux et sert également à la dilution des cellules lors de la transformation ou des tests de viabilité. La solution d'eau salée est généralement utilisée à 18% pour atteindre une concentration en sel NaCl de 2,5 M.

Le YPC est un milieu riche utilisé à la fois pour ensemencer et strier des cellules. Il est composé de :

- eau milliQ,
- 2,5 M NaCl
- 90 mM $MgCl_2$
- 84 mM $MgSO_4$
- 54 mM KCl
- 3 mM $CaCl_2$
- 5 mg/mL d'extrait de levure
- 1 mg/mL de peptone
- 1 mg/mL de CASAmino acids
- 12 mM Tris HCl
- 1,8 mM KOH

A cette solution peut être ajouté 15 mg/mL d'agar, pour la préparation des milieux gélosés.

Le CASA est un milieu plus pauvre que le YPC. Il est composé de :

- 2,5 M NaCl,
- 90 mM $MgCl_2$
- 84 mM $MgSO_4$
- 54 mM KCl
- 1 mg/mL CASAmino acids,
- 2,4 mM KOH
- 3 mM $CaCl_2$
- 0,1 µg/mL Biotine
- 0,8 µg/mL Thiamine,
- 40 µg/mL Thymidine et Hypoxantine

A cette solution peut être ajoutés 15 mg/mL d'agar, 10 µg/mL d'uracile, 50 µg/mL de 5'FOA (l'acide 5-fluoroorotic).

Souche	Génotype	Source ou Référence	Utilisation
H26	Δ*pyrE2*	[102]	Souche sauvage
H1804	Δ*pyrE2*, Δ*trpA* Δ*oriC1* Δ*oriC2* Δ*oriC3* Δ*ori-pHV4*	[29]	Comparaison avec la souche HVFD43
H938	Δ*pyrE2* Δ*fen1*	[99]	Comparaison avec la souche HVFD19
HVFD1	Δ*pyrE2, gfp::rpa2*	Pop-in/pop-out pFD2 dans H26	Localisation de la protéine fusion GFP::RPA2
HVFD10	*pyrE2+, gfp::pcna*	Pop-in/Pop-out pFD4 dans H26	Pas de colonie viable exprimant *gfp ::pcna* obtenue après pop-out
HVFD18	Δ*pyrE2, gfp::orc1*	Pop-in/Pop-out pFD3 dans H26	Localisation de la protéine fusion GFP::ORC1
HVFD19	Δ*pyrE2, gfp::fen1*	Pop-in/Pop-out pFD1 dans H26	Localisation de la protéine fusion GFP::FEN1
HVFD43	Δ*pyrE2*, Δ*trp,A* Δ*oriC1* Δ*oriC2* Δ*oriC3* Δ*ori-pHV4, gfp::orc1*	Pop-in/Pop-out pFD3 dans H1804	Localisation de la protéine GFP::ORC1 dans un contexte Δ*ori*

Tableau 4. Souches d'*Haloferax volcanii* utilisées.

Après préculture dans du YPC, les souches sont conservées à -80°C dans un mélange 20 % glycérol-1,5 % eau salée.

2.2 Milieux et souches d'*Escherichia coli*

Les cellules d'*E. coli* ont été cultivées dans du milieu LB (Luria Butani Medium). Il est composé de :

- eau milliQ
- 5 mg/mL extraits de levure
- 10 mg/mL bacto-tryptone
- 10 mg/mL NaCl.

Lorsqu'il est utilisé sous forme solide, 15 mg/ml de bacto-agar est ajouté au milieu.

Le milieu SOC de clonetech est utilisé lors de l'étape de régénération après transformation pour une meilleure intégration des plasmides. Il est composé de :

- 2 % tryptone
- 0,5 % extraits de levure
- 10 mM NaCl
- 2,5 mM KCl
- 10 mM $MgCl_2$
- 10 mM $MgSO_4$
- 20 mM glucose

A ces milieux sont ajoutés pour la sélection 100 µg/mL d'ampicilline, 20 µg/ml de Xgal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) et 0,5 mM l'IPTG (isopropyl β-D-1-thiogalactopyranoside).

Souche	Génotype	Utilisation
Stellar, de Clonetech	*F⁻, endA1, supE44, recA1, relA1, gyrA96, phoA, ∅80dlacZΔM15, Δ(lacZYA-argF), U169, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, λ⁻.,*	Utilisées pour le clonage
dam⁻ GM121	*F⁻, dam-3, dcm-6, ara-14, fhuA31, galK2, galT22, hdsR3, lacY1, leu-6, thi-1, thr-1, tsx-78*	Utilisées pour préparer l'ADN plasmidique non méthylé avant la transformation d'*H. volcanii*

Tableau 5. Souches d'Escherichia coli utilisées.

2.3 Expression physiologique d'une protéine fusionnée à la GFP chez *H. volcanii*

Toutes les étapes de clonage ont été réalisées dans un contexte génétique *dam*$^+$. Les étapes de préparation des plasmides ont été réalisées dans des contextes *dam*$^-$ permettant de préparer l'ADN plasmidique non méthylé avant la transformation d'*H. volcanii*.

Plasmide	Spécificité	Utilisation	Référence
pTA131	Vecteur intégratif dérivé du pBluescript II. Possède un marqueur de sélection chez *E. coli* (*bla*), un marqueur de sélection chez *H. volcanii (pyrE2)* et une origine de réplication bactérienne	Vecteur intégratif chez *H. volcanii*	[102]
pTA535	pTA131 dans lequel sont insérées les régions amont et aval du gène *fen1* d'*H. volcanii*	Délétion du gène *fen1*	[99]
pRL18	pTA131 dans lequel est inséré le gène *fen1* d'*H. volcanii*	Matrice pour l'amplification de *fen1*	Cette étude
pRL29	pTA131 dans lequel est inséré le gène synthétique de la *gfp* codant pour la smRSGFP dont le biais de codon a été optimisé pour *H. volcanii* (Genecust)	Matrice pour l'amplification de *gfp*	[139]
pFD1	pTA131 dans lequel sont insérées les régions amont du gène *fen1,* le gène *gfp,* et les 400 premières paires de bases du gène *fen1*	Insertion au locus chromosomique de *gfp::fen1*	Cette étude
pFD2	pTA131 dans lequel sont insérées les régions amont du gène *rpa2,* le gène *gfp* et les 400 premières paires de bases du gène *rpa2*	Insertion au locus chromosomique de *gfp::rpa2*	Cette étude
pFD3	pTA131 dans lequel sont insérées les régions amont du gène *orc1,* le gène *gfp* et les 400 premières paires de bases du gène *orc1*	Insertion au locus chromosomique de *gfp::orc1*	Cette étude
pFD4	pTA131 dans lequel sont insérées les régions amont du gène *pcna,* le gène *gfp* et les 400 premières paires de bases du gène *pcna*	Insertion au locus chromosomique de *gfp::pcna*	Cette étude

Tableau 6. Plasmides utilisés

Insert	Amorces	Matrice	Utilisation	Méthode utilisée
US-FEN1	RL125-RL126	pRL18	Remplacement au locus chromosomique pour expression de FEN1 fusionnée à la GFP	PCR 2 temps
GFP	RL85bis-RL124	pRL29		
FEN1	RL127-RL128	pRL18		
US-Rpa2	RL129-RL130	ADN génomique de H26	Remplacement au locus chromosomique pour expression de Rpa2 fusionnée à la GFP	PCR 2 temps
GFP	RL85bis-RL124	pRL29		
RPA2	RL131-RL132	ADN génomique de H26		
US-Orc1	RL133-RL134	ADN génomique de H26	Remplacement au locus chromosomique pour expression de Orc1 fusionnée à la GFP	PCR 2 temps
GFP	RL85bis-RL124	pRL29		
ORC1	RL135-RL136	ADN génomique de H26		
US-PCNA	RL155-RL156	ADN génomique de H26	Remplacement au locus chromosomique pour expression de PCNA fusionnée à la GFP	SLIC
GFP	RL157-RL158	pRL29		
PCNA	RL159-RL160	ADN génomique de H26		

Tableau 7. Caractéristiques des inserts utilisés pour le clonage

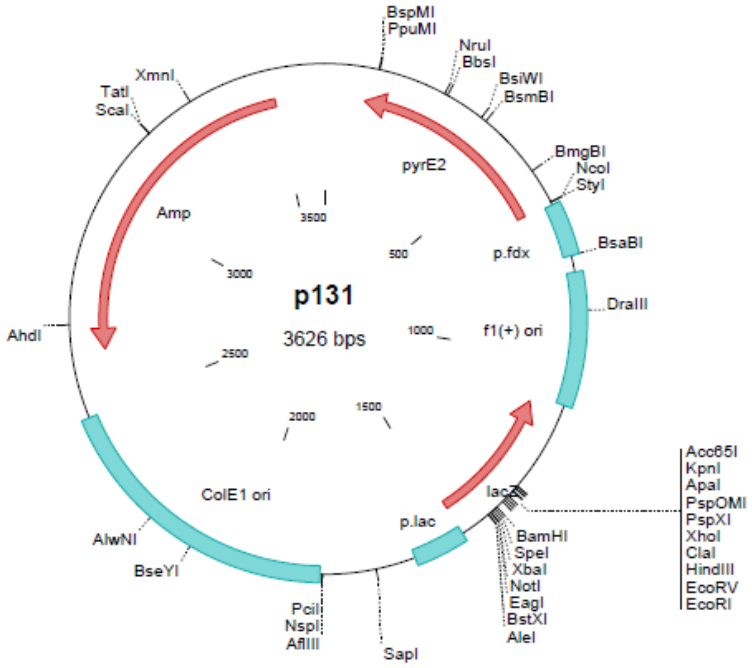


Figure 16. Carte du plasmide pTA131.
Ce vecteur est non réplicatif chez *H. volcanii*. Il possède une origine de réplication reconnue par *E. coli*, le gène de résistance à l'ampicilline et le gène *pyrE2* (orotate phosphoribosyltransferase), gène homologue du gène *ura* synthétisant l'uracile. Le site *Eco*RV se trouve dans le gène *lacZ*.

2.3.1 Construction dans *E. coli* du plasmide intégratif

La première étape pour exprimer les protéines fusionnées à la GFP chez *H. volcanii*, est de construire les plasmides permettant de remplacer au locus chromosomique l'allèle sauvage par l'allèle fusionné au gène *gfp*. Pour chaque plasmide construit, la GFP est flanquée de deux régions d'homologie avec le génome d'*H. volcanii* : celles-ci se situent en amont et au début du gène d'intérêt. Pour cela, les plasmides doivent contenir pour chaque gène : (i) une région d'environ 300 paires de bases correspondant à la région en amont du gène ii) le gène *gfp* et (iii) le début du gène d'intérêt (environ 300-400 pb) de sorte que le gène soit exprimé fusionné à la GFP. Deux approches ont été testées : une construction du plasmide 'pas à pas' *via* la méthode de PCR deux temps [140] et une construction du plasmide *via* la méthode SLIC (Sequence and Ligation Independent Cloning) [141] Les plasmides construits et utilisés sont listés dans les tableaux 6 et 7.

Principe de la méthode PCR 2 temps

La PCR deux temps se base sur une homologie entre chaque fragment (Figure 17 A). Celle-ci est portée par les deux régions flanquant la GFP : la région en amont du gène « US-gène » et la région du gène « gène ».

Dans un premier temps, une amplification par PCR classique de chacun des fragments à cloner est réalisée. Pour cela, un couple d'amorce sens et antisens est utilisé pour chaque fragment. Le fragment de la région en amont du gène « US-gène » est amplifié *via* les amorces « US-F » et « US-R ». US-R présente en son extrémité 5' 18 nucléotides homologues au fragment « GFP ». Le fragment de la région « GFP » est amplifié *via* les amorces « GFP-F » et « GFP-R ». Le fragment du gène d'intérêt est amplifié *via* les amorces « gène-F », qui présente en son extrémité 5' 18 nucléotides homologues au fragment « GFP », et « gène-R ». (Figure 17 A, PCR 1)

Ensuite, une deuxième PCR est réalisée en utilisant ces trois fragments. Le fragment « GFP » sert de matrice tandis que les fragments « US-gène » et « gène » jouent le rôle d'amorce en se fixant respectivement en 5' et en 3' de la séquence « GFP » par leur région d'homologie. Les fragments générés s'hybrident ensuite ensemble par la région « GFP » et la synthèse d'ADN continue en 3' de chaque fragment, aboutissant à la formation du fragment « US-GFP-gène » souhaité (Figure 17 A, PCR 2).

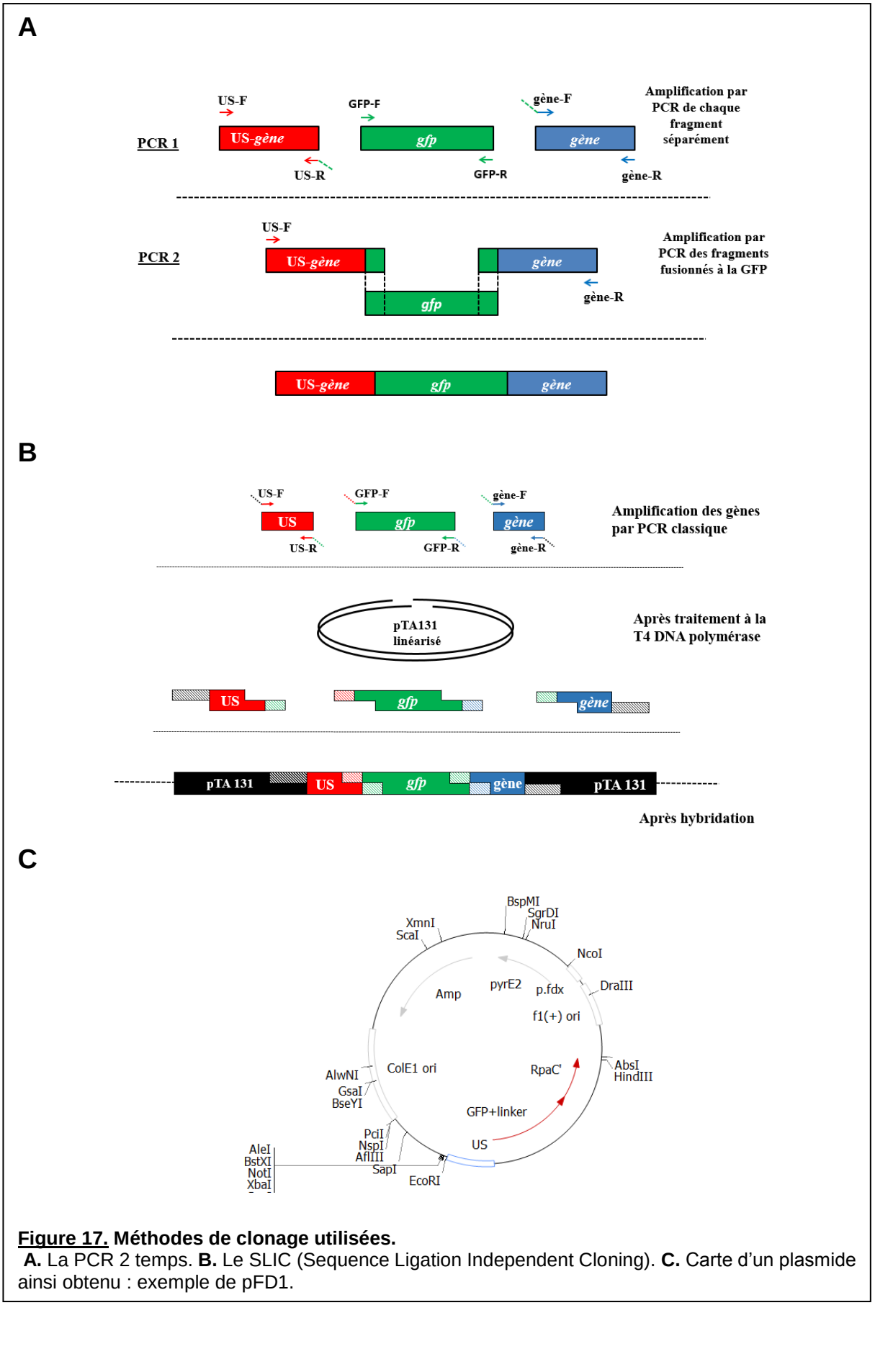


Figure 17. Méthodes de clonage utilisées.
A. La PCR 2 temps. **B.** Le SLIC (Sequence Ligation Independent Cloning). **C.** Carte d'un plasmide ainsi obtenu : exemple de pFD1.

Les conditions de PCR utilisées (tableaux 6, 7) ont été optimisées pour chaque fragment. Toutes les PCR ont été réalisées avec l'enzyme Q5 de NEB. Le mix préparé contient la Q5 à 0,02 U/µL, 2 mM de Mg^{2+}, 0,5 µM de chaque amorce sens et antisens et 200 µM de dNTPs et 100 ng d'ADN ou 1 µL d'une colonie diluée dans 100µl d'eau milliQ. Les produits de la première PCR obtenus ont été purifiés sur gel d'agarose puis utilisés comme matrice pour réaliser la seconde PCR. Les trois fragments ont été ajoutés à quantité égale : 100 ng de chaque fragment, quantifié par absorbance, ont été ajoutés au mix de PCR. Avant d'ajouter les amorces et l'ADN polymérase Q5 (NEB), 20 cycles de dénaturation à 95°C (30 sec) et d'hybridation à 68°C (30 sec) sont réalisés. Cette étape permet d'augmenter la formation des fragments : « US-gène-*gfp* » et « *gfp*-gène » et ainsi favoriser l'appariement de ces deux fragments *via* la séquence « GFP » pour la production du fragment complet. Ensuite 25 cycles de PCR classique comprenant une étape de dénaturation à 98°C (10 sec), une étape d'hybridation optimisé pour chaque couple d'amorce (15 sec) et une étape d'élongation à 72 °C, sont réalisés avec comme amorces « US-F » et « gène-R » (Figure 17.A). Le produit de PCR ainsi obtenu, qui correspond à l'insert souhaité, est purifié sur gel.

Avant d'être inséré dans le vecteur pTA131 (Figure 16), l'insert obtenu par PCR 2 temps est traité par la polynucléotide kinase de NEB (1µl d'enzyme / 1550 ng d'ADN) utilisée avec son tampon (70 mM Tris-HCl,10 mM $MgCl_2$, 5 mM DTT) pendant 30 minutes à 37°C puis inactivés pendant 20 minutes à 60°C. Cette étape est nécessaire avant de procéder à la ligation car les fragments de PCR n'ont pas d'extrémité 5' phosphorylé, ce qui rend la ligation impossible.

Le plasmide est ensuite préparé pour la ligation : il est linéarisé par EcoRV puis déphosphorylé par l'antarctic phosphatase de NEB (1µl d'enzyme / 2000 ng d'ADN) utilisée dans son tampon (100 mM de NaCl, et 50 mM de Tris-HCl et 10 mM de $MgCl_2$ ainsi que 100 µg/ml de BSA) pendant 15 minutes à 37°C puis inactivée pendant 5 minutes à 65°C. Cette étape de déphosphorylation est nécessaire afin d'éviter une recircularisation du plasmide. Il est purifié sur gel d'agarose grâce au kit gel extraction PCR clean up de Macherey-Nigel.

Les inserts sont ligués dans le vecteur linéarisé au moyen de la T4 DNA ligase du kit « quick ligation » de NEB pendant 5 minutes à 25°C, en appliquant un ratio molaire insert : vecteur de 3:1 au moyen de la formule suivante :

$$quantité\ insert\ (g) = \frac{3}{1} \times \frac{\text{quantité de vecteur} \times \text{taille de l'insert (pb)}}{taille\ du\ vecteur\ (pb)}$$

Le produit de ligation est ensuite transformé dans *E. coli*. Pour cela, des cellules chimiocompétentes commerciales (Stellar, de Clonetech, tableau 5) sont transformées par choc thermique.

Principe de la méthode SLIC (Sequence Ligation Independent Cloning)

La méthode SLIC se base également sur l'homologie entre fragments. Cette méthode présente l'avantage de ne pas nécessiter de PCR à partir d'autres fragments de PCR. De plus, à la différence de la méthode PCR 2 temps, il n'y pas de ligation mais une hybridation entre tous les fragments constituant le plasmide recombinant : la région « US-gène », la région « GFP », la région « gène » et le plasmide pTA131 (Figure 17 B).

Afin d'obtenir les différents fragments, une étape d'amplification par PCR classique est réalisée pour les régions « US-gène », « GFP » et « gène ». Toutes les séquences amplifiées par PCR sont porteuses en 5' et en 3' d'une séquence de 30 paires de base d'homologie avec les deux régions qui les flanquent. L'oligonucléotide en 5' porte l'homologie avec la région située en amont, l'oligonucléotide en 3' porte l'homologie avec la région située en aval. Par exemple, les régions « US-gène » et « gène » sont porteuses de l'homologie avec le plasmide, conférée par les oligonucléotides en 5' de l'amplification « US-gène » et en 3' de l'amplification « gène ».

Ainsi, la région « US-gène » est amplifiée *via* l'amorce « US-F » qui porte 30 nucléotides d'homologie avec le plasmide pTA131 et l'amorce « US-R » qui porte 15 nucléotides d'homologie avec la « GFP ». La région « GFP » est amplifiée avec l'amorce « GFP-F » qui porte 15 nucléotides d'homologie avec la région « US-gène » et l'amorce « GFP-R » qui porte 15 nucléotides d'homologie avec la région « gène ». La région « gène » est amplifiée par l'amorce « gène-F » qui porte 15 nucléotides d'homologie avec la région « GFP » et avec l'amorce « gène-R » qui porte 30 nucléotides d'homologie avec le plasmide pTA131 (Figure 17 B).

Une fois les conditions de PCR optimisées pour chaque fragment (tableaux 6 et 7), les produits de PCR sont purifiés sur gel d'agarose.

Le vecteur est linéarisé par EcoRV puis déphosphorylé par l'antarctic phosphatase de NEB (1µl d'enzyme/30µl de mix) utilisée avec son tampon (100 mM de NaCl, et 50 mM de Tris-HCl et 10 mM de $MgCl_2$ ainsi que 100 µg/ml de BSA) pendant 15 minutes à 37°C puis inactivée pendant 5 minutes à 65°C.

Afin de procéder à l'hybridation des fragments entre eux, des extrémités simple brin, porteurs des homologies, doivent être générées. Pour cela, les fragments « US-gene », « GFP », « gène » et le plasmide sont traités à la T4 DNA polymerase à 1U/µl (0,5 µl d'enzyme/20 µl de mix) utilisée avec son tampon (50 mM de NaCl, 10 mM de Tris-HCl, 10 mM de MgCl2, 100 µg/ml de BSA) pendant 45 minutes à 22°C. L'activité enzymatique principale de polymérisation de cette enzyme est détournée en évitant d'ajouter des dNTPS au mix. Ainsi, l'activité 3'→5' exonucléasique est favorisée. Pendant ces 45 minutes d'incubation, l'enzyme digère en moyenne les 30 paires de base d'homologie. Cette réaction est arrêtée en ajoutant 2µl de dCTP à 10 mM dans chacun des tubes.

L’hybridation des différents fragments est réalisée avec un ratio molaire insert : vecteur de 3:1 en les incubant avec le tampon de la ligase (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT) pendant 30 minutes à 37 °C. Le produit de cette hybridation est ensuite directement utilisé pour la transformation des cellules chimiocompétentes commerciales d’*E. coli* (Stellar, de Clonetech), les clones ayant intégré la plasmide avec l’insert d’intérêt sont sélectionnés par criblage bleu/blanc.

Afin de contrôler l’insertion des fragments d’intérêt dans le vecteur, les plasmides recombinés ont été digérés par *Pvu*II-HF de BioLabs pendant 30 minutes à 37°C et par *Bst*B1 de NEB pendant 30 minutes à 65°C. Plusieurs clones ayant le profil de digestion attendu ont été séquencés dans le but de valider leur séquence avant toute insertion dans le génome (Figure 17 C).

Nom	Séquence (5'-3')	Tm°C	Fragment amplifié
Oligonucléotides utilisés pour la PCR 3 temps			
RL85 bis	ATGTCGAAAGGCGAGGAACTCTTC	58	GFP (avec RL124) pour construire pFD1, pFD2 et pFD3
RL124	CGACCGGTAGGCGTAGCCGC	58	GFP (avec RL85bis)
RL125	GCTACGAGACGGTCCGCATC	58	US Fen1 (avec RL126)
RL126	AGTTCCTCGCCTTTCGACATTTGCTCGGCTTTCGGGG	58	US Fen1 (avec RL125)
RL127	GCGGCTACGCCTACCGGTCGATGGGAAACGCAGACCTG	58	Fen1 (avec RL128)
RL128	TGTCCTGAATCGTCTCGGTG	58	Fen1 (avec RL127)
RL129	CGACGCGGAGAACTGCGCC	66	US Rpa2 (avec RL130)
RL130	AGTTCCTCGCCTTTCGACATCAGGCGTCACCTCCCGG	66	US Rpa2 (avec RL129)
RL131	GCGGCTACGCCTACCGGTCGATGGGCGTCATCCGGGAGG	66	Rpa2 (avec RL132)
RL132	AGGTTCTCCTTCGCGCCG	66	Rpa2 (avec RL131)
RL133	GACGACATCGCCGGGGGAGC	62	US Orc1 (avec RL134)
RL134	AGTTCCTCGCCTTTCGACATTGAATACCCCTTTGTTTC	62	US Orc1 (avec RL133)
RL135	GCGGCTACGCCTACCGGTCGATGGACGAGGAAGAGAATC	66	Orc1 (avec RL136)
RL136	TGGCGTGTACGACGGGCG	66	Orc1 (avec RL135)
Oligonucléotides utilisés pour le SLIC			
RL 155	CGACGGTATCGATAAGCTTGATATCGAATTCCATGACCTCGACGACCTCG	60	US PCNA pour p131-US-GFPopt-PCNA'
RL 156	CTCGCCTTTCGACATATGTGCCGTCGTACGACAGC	60	US PCNA pour p131-US-GFPopt-PCNA'
RL 157	CGTACGACGGCACATATGTCGAAAGGCGAGGAACT	60	GFP+linker pour p131-US-GFPopt-PCNA'
RL 158	GATGGCCTTGAACATCGACCGGTAGGCGTAGCCGC	60	GFP+linker pour p131-US-GFPopt-PCNA'
RL 159	TACGCCTACCGGTCGATGTTCAAGGCCATCGTGAG	60	PCNA' pour p131-US-GFPopt-PCNA'
RL 160	AAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCGGATGTGGTCCGAAACCATGT	60	PCNA' pour p131-US-GFPopt-PCNA'
Oligonucléotides utilisés pour le criblage			
RL28	TTAAGTTGGGTAACGCCAGGG	48	insert dans pTA131 et ses dérivés
RL29	ACCCCAGGCTTTACACTTTATGC	48	insert dans pTA131 et ses dérivés

Tableau 8. Oligonucléotides utilisés.

Transformation d'*E. coli dam*$^+$ et *dam*$^-$

Pour la construction du plasmide intégratif, deux types de cellules ont été utilisées : des cellules commerciales STELLAR de Clontech utilisées pour les étapes de clonage et les cellules *dam*$^-$ (*F-dam-3, dcm-6*) utilisées pour l'étape de préparation d'ADN non méthylé.

Pour les transformer, les cellules chimiocompétentes commerciales sont décongelées sur glace et 5 µl du mix de clonage est ajouté à 50 µl de cellules compétentes. Le mix est laissé reposer 30 minutes sur glace. Un choc thermique de 45 secondes à 42 °C est réalisé. Les tubes sont laissés dans la glace pendant 1 à 2 minutes et 500 µl de milieu SOC est ajouté

dans chaque tube. Les cellules sont ensuite incubées pendant 1h00 à 37 °C sous agitation puis étalées sur boite LB 100 µg/ml ampicilline, 20 µg/mL X-Gal et 0,5 mM IPTG. Les boites sont incubées à 37°C pour la nuit. Ainsi, seules les cellules ayant intégré le plasmide pTA131, porteur du gène de résistance à l'ampicilline peuvent pousser sur boite LB Amp. Les plasmides recombinants peuvent être sélectionnés en se basant sur le criblage « bleu-blanc ». Le produit de PCR est inséré dans une partie du gène *lacZ* qui est sous contrôle du promoteur lactose, inductible à l'IPTG, et code pour la partie α de la β-galactosidase. L'activité de cette enzyme permet l'hydrolyse du Xgal, composé incolore, en un galactoside bleu, ce qui confère leur couleur bleue aux bactéries *lacZ*$^+$. Or le gène *lacZ* est inactivé par l'insertion du fragment d'intérêt dans le plasmide. Ainsi les colonies transformées par un plasmide avec insert n'expriment pas la β-galactosidase. Le X-Gal n'étant plus transformé les colonies sont blanches.

L'ADN des cellules d'*E. coli* est méthylé par les Dam méthylases qui font partie du système de réparation des mésappariements. A l'opposé, l'ADN des cellules d'*H. volcanii* n'est pas méthylé, il s'agit également d'un système de défense contre l'insertion d'ADN étranger : si l'ADN est méthylé il sera digéré par des enzymes de restriction.

Pour la préparation de l'ADN non méthylé sont utilisées des cellules *dam*$^-$ electrocompétentes transformées par électroporation (tableau 5). Pour cela, 5 µl d'ADN plasmidique méthylé ont été ajoutés à 50 µl de cellules compétentes. Elles ont ensuite été soumises à un champ électrique de 2500 mV, puis 500 µl de SOC est ajouté. Les cellules sont incubées pendant 1h00 à 37 °C sous agitation puis étalées sur boite LB Ampicilline. Les boites sont incubées à 37°C pour la nuit.

L'ADN plasmidique est extrait et purifié à partir de précultures en LB-Amp en utilisant le kit 'NucleoBond Xtra plasmid purification' de Macherey-Nagel selon les instructions du fournisseur. La colonne retenant l'ADN est une résine de silice échangeuse d'anion. L'ADN plasmidique obtenu est ensuite quantifié par absorbance.

2.3.2 Intégration et remplacement au locus chromosomique chez H. volcanii

Principe général du pop-in / pop-out

Le système pop-in/pop-out repose sur le marqueur *pyrE2* utilisé à la fois pour la sélection et la contre sélection, et sur deux régions d'homologie. Le gène *pyrE2* code pour l'enzyme orotate phosphoribosyltransferase impliquée dans la synthèse d'uracile. *pyrE2* est présent dans le plasmide utilisé pour la transformation. L'étape de pop-in consiste en

l'intégration de ce plasmide dans l'ADN génomique d'*H. volcanii*. Lorsque le plasmide s'intègre dans le génome par recombinaison homologue, le génome *ΔpyrE2* devient *pyrE2*+. Les colonies issues de la transformation sont sélectionnées sur milieu CASA. A l'issu de cette étape seules les colonies ayant intégré le plasmide peuvent pousser.

Néanmoins, si le gène recombiné est essentiel et non fonctionnel lorsqu'exprimé en fusion avec la GFP, l'événement de recombinaison de pop-in aboutira pour 100 % des cas à un pop-in en US. En effet, lors du pop-in en US (Figure 18 B), le gène d'intérêt est sauvage. En revanche, si l'événement de recombinaison de pop-in a lieu en dans la région gène (Figure 18 A), le gène est uniquement présent dans le génome en fusion avec le gène *gfp*.

L'étape de pop-out repose sur le marqueur *pyrE2* utilisé cette fois-ci en tant que marqueur de contre-sélection. A partir d'une préculture d'une colonie issue du pop-in, les cellules sont étalées sur boite CASA 5'FoA (l'acide 5-fluoroorotic) avec de faibles concentrations d'uracile et laissées incuber pendant 7 jours à 45°C. L'activité de PyrE2 transforme le 5'FoA en un composé toxique pour la cellule : le 5-fluorouracil. Lors de cette étape, seules les cellules n'ayant plus le plasmide, donc n'exprimant pas le gène *pyrE2*, peuvent pousser. L'événement de recombinaison chromosomique aboutissant à la perte du plasmide conduit pour 50 % des cas à la conservation de la copie sauvage du gène et donc la perte de la copie du gène fusionné à la séquence de la GFP et pour 50 % à la conservation de la copie du gène fusionné au gène *gfp* et la perte de la copie sauvage du gène. En effet, si le deuxième évènement de recombinaison homologue a eu lieu dans la même région d'homologie que le premier évènement de recombinaison homologue, cela aboutit à un retour au contexte sauvage. En revanche, si le deuxième évènement de recombinaison homologue a lieu dans l'autre région d'homologie, cela aboutit à la conservation de la copie du gène fusionné à la GFP (Figure 18 A et 18 B).

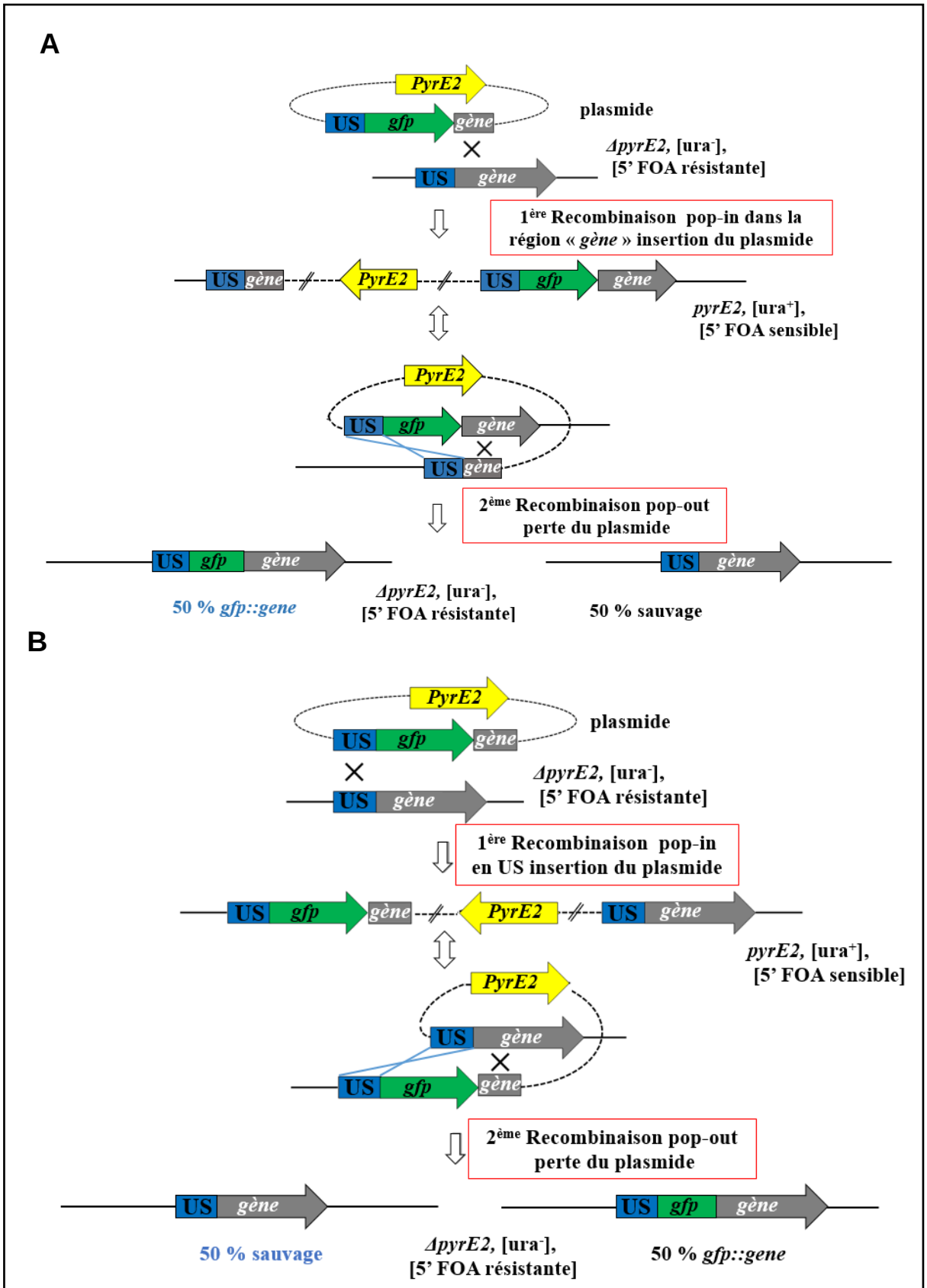


Figure 18. Le pop-in / pop-out.
A. Le pop-in/ pop-out dans la région « gène ». **B.** Le pop-in/pop-out en US.

La transformation d'*H. voclanii* : le pop-in

Pour la transformation d'*H. volcanii*, différentes solutions sont utilisées pour fragiliser la membrane, stabiliser les sphéroblastes ou en tant que solution de récupération avant étalement. Les solutions de sphéroblastes tamponnée et non tamponnée sont utilisées pour fragiliser la membrane d'*H. volcanii.* La solution de dilution des sphéroblastes sert à stabiliser les sphéroblastes et aide à leur récupération.

Une préculture de la souche à transformer est réalisée à partir d'une colonie isolée incubée sur la nuit dans 10 ml de YPC à 45 °C - 150 rpm. Toutes les étapes se font à température ambiante. La préculture est centrifugée 10 minutes à 4000 rpm. Le culot est resuspendu dans 2 mL de solution de sphéroblastes tamponnée (1M NaCl, 27 Mm KCl, 15 % sucrose, 50 mM Tris HCl à pH 8.5) et les cellules sont transférées dans un tube eppendorf 2 ml à fond rond. Une seconde centrifugation de 8 minutes à 6000 rpm permet d'obtenir un culot resuspendu dans 600 µl de solution de sphéroblastes tamponée. A 200 µl de ces cellules sont ajoutés de l'EDTA (0,05 M à pH8), ce mélange incube pendant 10 minutes. Un mix d'ADN contenant 1 à 2 µg de plasmide dam-, 15 µl de solution de sphéroblastes non tamponée (1M NaCl à, 27 mM KCl, 15 % sucrose) et de l'EDTA (0,083 M) est ajouté aux cellules, ce mélange incube pendant 5 minutes. A ce mélange « cellules + ADN » 250 µl de 60% PEG 600 sont ajoutés, et laissé incuber pendant 30 minutes. Une fois l'incubation terminée, 1.5 ml de solution de dilution de sphéroblastes (18% eau salée, 15% sucrose, 3 mM CaCl2) sont ajoutés, puis les cellules sont centrifugées 8 minutes à 6000 rpm. Le culot est récupéré avec 1 ml de solution de régénération (18 % eau salée, YPC, 15% sucrose, 3mM CaCl2) sans le resuspendre puis incubé à 45°C sans agitation pendant 1h30, puis sous agitation pendant 3h00. Enfin, les cellules sont centrifugées 8 minutes à 6000 rpm et sont resuspendues dans 1 ml d'eau salée 18 %. Des dilutions successives au 1/10ème sont réalisées en 18% eau salée puis étalées sur boites CASA puis incubées à 45 °C pendant 7 jours. Les transformants ainsi obtenus sont striés sur boîte CASA et incubés à 45°C pendant 7 jours. A partir d'une colonie ainsi isolée est réalisée l'étape de pop-out.

Le pop-out

Une préculture est réalisée à partir d'une colonie d'*H. volcanii* incubée dans 3 ml de YPC à 45°C et 150 rpm sur la nuit. A partir de cette préculture, une culture est ensemencée et incubée sur la nuit, ceci est répété une seconde fois. Puis les cellules sont culotées et resuspendues dans 1 mL de 18% eau salée. Des dilutions successives au 1/10ème sont réalisées dans 18% eau salée puis étalées sur boite CASA-5'FoA et incubées à 45°C pendant 7 jours. Les colonies obtenues sont repiquées sur YPC pour être testées par PCR. Les

colonies présentant le phénotype attendu sont alors striées sur YPC puis conservées à -80°C comme indiqué précédemment.

A l'issu du pop-out, afin de distinguer les cellules ayant le gène fusionné à la GFP des cellules ayant gardé la copie sauvage, un criblage par PCR amplifiant la GFP ainsi que le gène a été réalisé pour les quatre gènes : *orc1*, *fen1*, *rpa2* et *pcna*. Ces PCR ont été réalisées comme indiqué précédemment en utilisant une colonie diluée dans 100 µL d'eau milliQ comme matrice d'ADN, les différentes amorces utilisées et leurs températures d'hybridation sont listées dans le tableau ci-dessous (Tableau 9).

Nom	Séquence	Tm °C	Utilisation
RL142	CCTGCACGAACGGCTGTGCG	58	Contrôle de l'insertion au chromosome de GFP::Orc1 (avec RL164)
RL164	ATCGAGACGTTACACAAGGTC	58	Contrôle de l'insertion au chromosome de GFP::Orc1 (avec RL142)
FD22	GCCACCGCCGCGTCGGCGTGT		Contrôle de l'insertion au chromosome de GFP::FEN1 (avec FD23)
FD23	TCGACCTGCTCGGCTGGCCCG		Contrôle de l'insertion au chromosome de GFP::FEN1 (avec FD22)
RL165	GACGGTTACGATGTGAACTCC	58	Contrôle de l'insertion au chromosome de GFP::Rpa2 (avec RL141)
RL141	GCCAGTGTGGACTTCGACCG	58	Contrôle de l'insertion au chromosome de GFP::Rpa2 (avec RL165)
RL161	GATGGTCGGCCTGTCGTGGC	55	Contrôle de l'insertion au chromosome de GFP::PCNA (avec RL162)
RL162	GCAAGCGACGACCGACGCGC	55	Contrôle de l'insertion au chromosome de GFP::PCNA (avec RL161)

Tableau 9. Les oligonucléotides utilisés pour le criblage des pop-out d'*H. volcanii*

A l'issu du pop-out, un autre test permet de distinguer les cellules ayant le gène fusionné à la *gfp* des cellules ayant gardé la copie sauvage : le colony-lift. Ce test repose sur la détection des cellules qui ont intégré le gène fusionné à la *gfp* dans leur génome en utilisant une sonde qui s'hybride avec la séquence de la *gfp* et qui est ensuite détectée par immunofluorescence. Les différents clones de pop-out sont transférés sur membrane de nylon, puis les cellules sont préparées pour l'hybridation de la sonde au moyen de plusieurs étapes de lavage, l'ADN est fixé à la membrane par pontage aux UVs. La sonde de *gfp* est synthétisée par PCR et marqué à la digoxigenin (DIG) puis elle est dénaturée avant d'être incubée avec la membrane. La sonde, préalablement dénaturée, s'hybride aux séquences d'ADN de la *gfp* dénaturées et fixées à la membrane, puis est reconnue par un anticorps anti-DIG couplé à la phosphatase alkaline. La membrane est ensuite incubée avec le Disodium 3-dioxetane-phenyl phosphate (technique de CSPD = Chemiluminescent su*Bst*rate for alkaline phosphate) qui au contact de l'alkaline phosphatase est déphosphorylé en un anion phenolate émettant de la

lumière à 477 nm. La membrane est donc révélée au Chemidoc de Biorad grâce à une caméra de haute résolution CCD et les clones ayant la fusion du gène à la *gfp* sont sélectionnés.

Etude de la fonctionnalité des protéines par test de croissance, par mesure de la densité optique (DO) et par calcul de Colony Forming Unit (CFU/mL)

Le marquage des protéines à la GFP peut modifier leur structure et avoir un impact sur leur fonctionnalité. Afin de savoir si les protéines fusionnées sont fonctionnelles des suivis de croissance par mesure de la Densité Optique (DO) et par calcul de CFU/mL (Colony Forming Unit/mL) ont été réalisés. Pour chaque souche, à partir d'une préculture sur la nuit en YPC, une culture est ensemencée à DO=0,005. A différents temps un prélèvement de cellules est réalisé permettant (i) une mesure de la densité optique à 600 nm, et (ii) des étalements de dilution au $10^{ème}$ successives (réalisées dans du SW 18%) sur boîtes YPC. Les colonies individuelles sont comptées après 4 jours d'incubation à 45°C. La courbe de croissance est tracée, deux points situés sur la partie linéaire de la courbe de croissance (suffisamment éloignés l'un de l'autre et représentatifs de la tendance) sont sélectionnés : T1 et T2 à partir desquels le temps de génération est calculé *via* la formule suivante :

Temps de génération = temps (min) / [(log10 T1 – log10 T2) / (log10(2))]

Avec, Temps = temps en minutes entre T1 et T2

T1 = CFU/mL à T1

T2 = CFU/mL à T2

Etude de la fonctionnalité des protéines par test de viabilité cellulaire par CFU/mL après exposition aux UV

Un test de viabilité cellulaire après exposition aux UV a été réalisé. Pour chaque souche, à partir d'une préculture sur la nuit en YPC, une culture est ensemencée à DO=0,1. Après une incubation sur la nuit (environ 17h), des dilutions successives au $10^{ème}$, dans de la SW 18%, sont réalisées. Une goutte de 20 µl de chaque dilution est déposée sur boite YPC. Une fois la goutte séchée, les cellules sont exposées à 50 J/m^2, 100 J/m^2, 150 J/m^2 et 200 J/m^2 au moyen d'une lampe à 254 nm. La quantité totale de cellules viables avant irradiation est mesurée en réalisant une boîte non exposée aux UVs. Les colonies individuelles sont comptées après 4 jours d'incubation à 45 °C.

2.4 Localisation du signal GFP dans les cellules vivantes et analyse quantitative

La localisation cellulaire des différentes protéines fusionnées à la GFP a été effectuée sur des cellules en cultures liquides à différentes phases de croissance. Pour cela, à partir d'une préculture sur la nuit en YPC, une culture est ensemencée à DO=0,01; DO=0,005 et DO=0,001. Après une nuit d'incubation (environ 17h00) à 45°C – 150 rpm, les cellules sont prélevées à différents temps afin de les imager.

La localisation cellulaire des différentes protéines fusionnées à la GFP a également été effectuée sur des cellules en cultures liquides après exposition à différentes drogues :

- à l'aphidicoline (APD), qui inhibe la polymérase réplicative alpha entraînant des arrêts de fourches de réplication [142]. La solution stock de 1 mg/ml est réalisée dans du DMSO, de sorte que les cellules contrôle sont uniquement exposées à une concentration équivalente de DMSO

- à la mitomycine C (MMC), qui entraîne principalement des lésions pontantes de l'ADN. La solution stock à 20 µg/mL est réalisée dans de l'eau

- à la phléomycine, qui entraîne principalement des cassures double-brin de l'ADN. La solution stock à 20 mg/mL est réalisée dans de l'eau.

Pour cela, à partir d'une préculture sur la nuit en YPC, une culture est ensemencée à DO=0,1. Après deux heures d'incubation à 45°C - 150 rpm, le volume nécessaire de solution stock de la drogue (ou de DMSO pour le contrôle lors d'exposition à l'aphidicoline) est ajouté puis les cellules sont incubées sur la nuit (environ 17h) à 45°C - 150 rpm. Pour mesurer le taux de survie après exposition aux différentes drogues, des dilutions successives au $10^{ème}$ sont réalisées dans de la SW 18%. Une goutte de 20 µl de chaque dilution est déposée sur boite YPC. Les colonies individuelles sont comptées après 4 jours d'incubation à 45 °C.

Afin d'observer les cellules en microscopie de fluorescence, 5 µl de cellules sont déposés sur une fine couche d'agarose 1% - SW 18% déposée sur une lame de verre. Une fois la goutte séchée, elle est recouverte d'une lamelle d'épaisseur 0,17 mm et observée par microscopie de fluorescence à champ large à l'aide d'un microscope inversé ZEISS Axioobserver. Les objectifs utilisés pour les acquisitions sont des 40X à huile, avec des ouvertures numériques de 1,3 et 1,4. Ces caractéristiques permettent de collecter la fluorescence sur un maximum de cellules et de capter le plus de signal possible grâce à une ouverture numérique élevée (NA 1,4). Cette ouverture numérique combinée à l'utilisation de l'huile comme milieu d'immersion, qui possède un indice de réfraction proche de celui du verre, garantissent une haute puissance de résolution sur des échantillons peu profonds (Figure 19 A).

La mise au point est tout d'abord réalisée en observant les cellules en Contraste interférentiel différentiel (DIC). Puis le signal GFP est détecté en excitant l'échantillon à 470 nm. La fluorescence est collectée en z tous les 0,250 µm sur 30 coupes de l'échantillon, les 30 coupes étant centrées sur le focus déterminé en DIC [139].

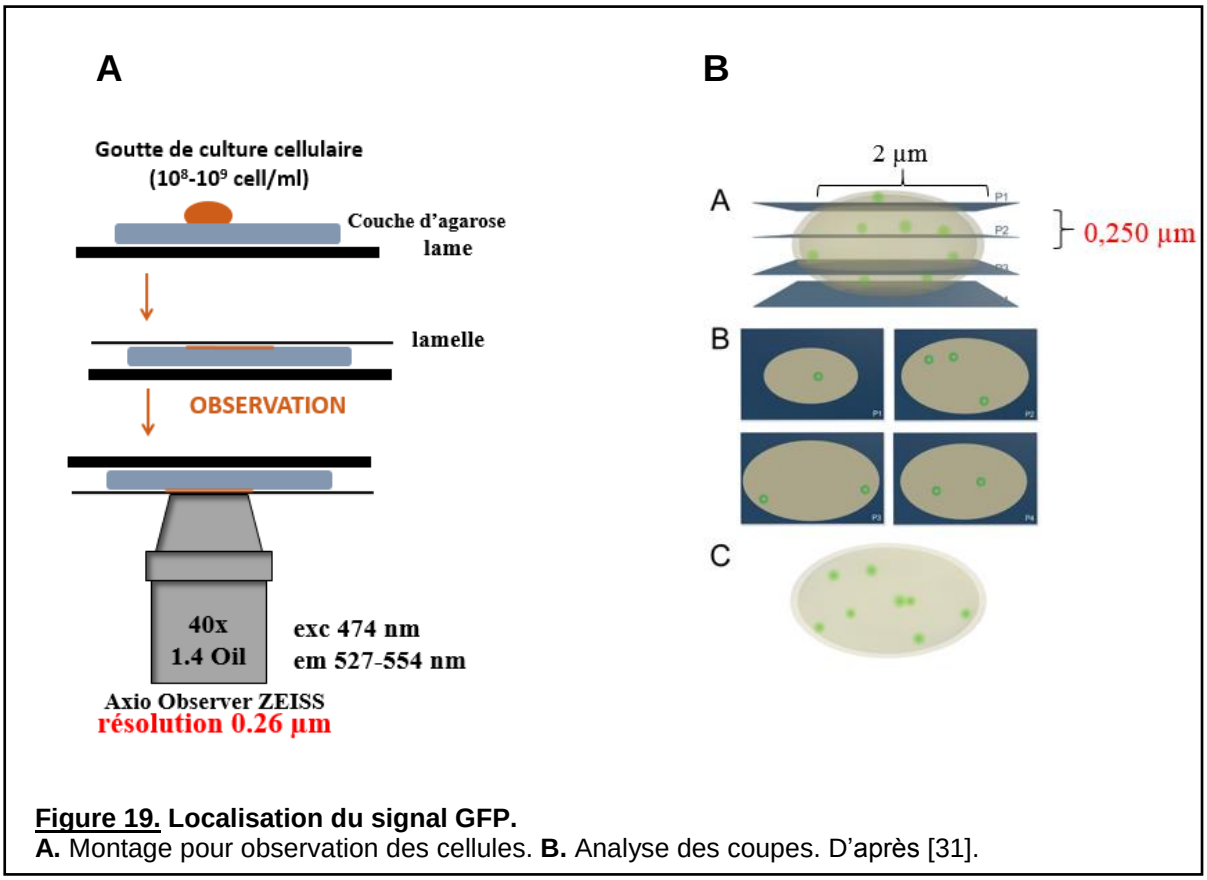


Figure 19. Localisation du signal GFP.
A. Montage pour observation des cellules. **B.** Analyse des coupes. D'après [31].

La première étape de l'analyse d'image consiste à réaliser une projection en z du maximum d'intensité en utilisant le logiciel ImageJ [31]. Parmi les 30 coupes réalisées, 4 à 6 coupes sont sélectionnées pour lesquelles les cellules sont dans le plan focal (Figure 19 B). Pour chaque projection en z ainsi obtenues, des analyses quantitatives sont réalisées en utilisant le logiciel Imaris 8.1. Chaque image est analysée en utilisant des paramètres de détection automatisés des cellules et des foyers de fluorescence ajustés pour chaque condition de culture. Pour toutes les conditions testées, les paramètres suivants sont appliqués : 1,5 µm de largeur du filtre lisse, 1µm de diamètre sphérique de soustraction d'arrière-plan, activation du fractionnement des cellules par points. Un diamètre cellulaire estimé à 2 µm est utilisé pour toutes les conditions sauf après exposition à 5 µg/ml et 10 µg/ml d'aphidicoline pour lesquelles le diamètre cellulaire augmente et est estimé à 4 µm. Les foyers de fluorescence dans les cellules sont automatiquement détectés en tant que 'vésicules' en utilisant les paramètres suivants : pas de soustraction du bruit de fond, un filtre de « qualité » sans seuillage automatique. Pour les foyers de fluorescence de GFP::RPA2, le diamètre

estimé est de 0,75 µm pour toutes les conditions sauf après exposition à 0,05 mg/ml et 0,1 mg/ml de phléomycine pour lesquelles les foyers sont détectés en tant que région nucléique grandissante seuillée à 1 µm, reflétant l'augmentation du diamètre du foyer de fluorescence lorsque les cellules sont traitées à de fortes concentrations en phléomycine. Pour les foyers de fluorescence de GFP::ORC1, pour toutes les conditions testées le diamètre estimé est de 0,75 µm.

Pour toutes les expériences, un grand nombre d'images regroupant plusieurs centaines de cellules a été analysé de façon automatisé sur Imaris nous garantissant une analyse non biaisée. Pour chaque condition testée au moins trois expériences indépendantes ont été analysées. Pour chaque cellule détectée, l'intensité totale des foyers de fluorescence, l'intensité totale de la cellule, la surface cellulaire, le nombre de foyers de fluorescence et le diamètre de chaque foyer de fluorescence sont déterminés.

Pour chaque expérience indépendante les données suivantes ont été déterminées :

- la surface cellulaire moyenne (surface cellulaire totale/nombre total de cellules),

- l'intensité cellulaire moyenne (intensité de fluorescence cellulaire totale/nombre total de cellules),

- le nombre moyen de foyers de fluorescence par cellule (nombre total de 'vésicules'/nombre total de cellules),

- la proportion de l'intensité de fluorescence au niveau du foyer de fluorescence (intensité totale des 'vésicules'/intensité de fluorescence totale cellulaire),

- le nombre de foyers de fluorescence par unité de surface (nombre total de 'vésicules'/surface cellulaire totale),

- l'intensité par μm^2 (intensité totale des cellules/surface cellulaire totale),

- la fréquence relative de cellules avec 0 à 20 foyers de fluorescence (nombre de cellules à 0 ou 1 ou 2…ou 20) spots / nombre total de cellules)

La représentation graphique des résultats a été réalisée *via* le logiciel GraPhpad Prism 5.

2.5 Colocalisation du signal GFP avec l'ADN dans les cellules vivantes et analyse quantitative

Pour la colocalisation de la GFP avec l'ADN, le spectre d'excitation de la GFP et le spectre d'excitation du marqueur utilisé pour marquer l'ADN doivent être suffisamment éloignés. Le DAPI fonctionne mal à de fortes concentrations en sel et l'acridine orange à un spectre d'excitation trop proche de celui de la GFP. L'Hoechst 33258 en revanche est excité

à 352 nm et émet à 461 nm (Figure 20), la GFP utilisée est excitée à 470 nm et émet à 509 nm (Figure 21), les deux spectres sont suffisamment éloignés.

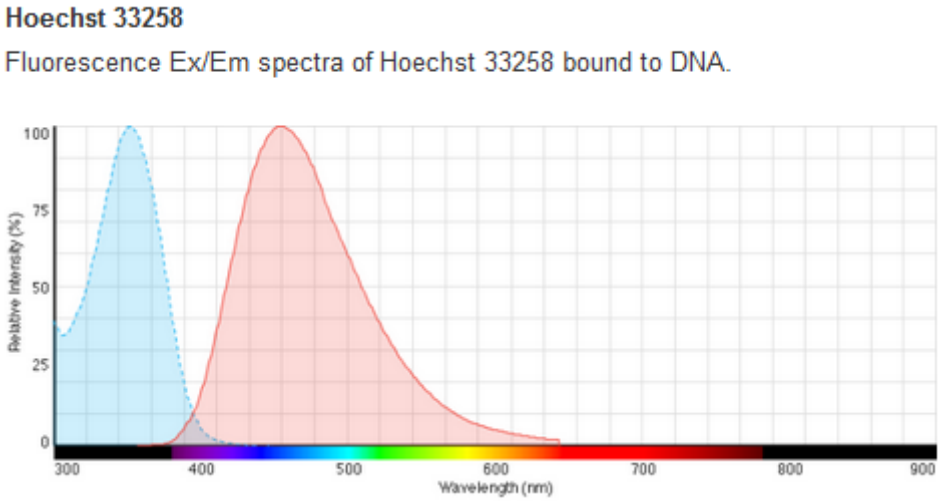


Figure 20. Schéma des spectres d'excitation et d'émission de l'Hoechst 33258.
De www.thermofisher.com.

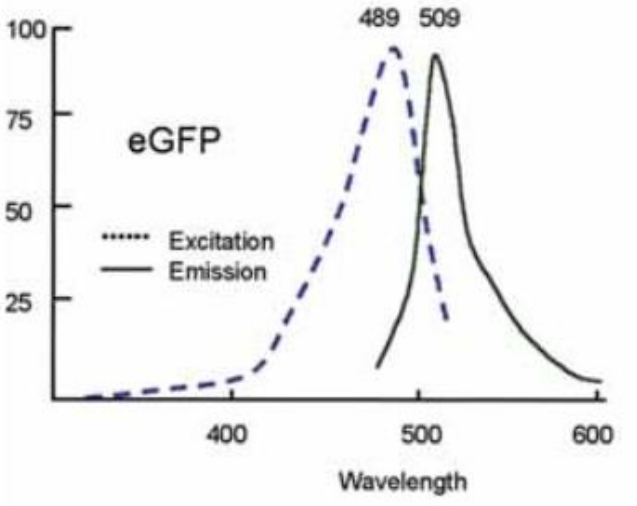


Figure 21. Schéma des spectres d'excitation et d'émission de la GFP utilisée.
De www.thermofisher.com.

Pour la colocalisation du signal GFP avec l'ADN dans les cellules vivantes le même microscope est utilisé que celui présenté précédemment (2.4).

A partir d'une préculture sur la nuit en YPC, une culture est ensemencée à DO=0,1. Après deux heures d'incubation à 45°C - 150 rpm, le volume nécessaire de solution stock de la drogue (ou de DMSO pour le contrôle lors d'exposition à l'aphidicoline et d'eau pour le contrôle lors de l'exposition à la MMC ou la phléomycine) est ajouté puis les cellules sont incubées sur la nuit (environ 17h) à 45°C - 150 rpm. Afin d'éliminer le YPC, 20 µl de chaque préculture sont centrifugées à 6000 rpm pendant 5 minutes puis resuspendues dans 20 µl de

18% d'eau salée. Un traitement des cellules au Hoechst 33258 (Cf=12 µg/mL) pendant 20 minutes, 150 rpm, à température ambiante, permet de marquer l'ADN. Afin d'éliminer le surplus d'Hoechst, les cellules sont centrifugées à 6000 rpm, 5 minutes puis resuspendues dans 20 µl de 18% d'eau salée. Les méthodes de préparation des échantillons et d'observation des cellules sont identiques à celles présentées dans la partie 2.4.

Pour toutes les expériences, un grand nombre d'images a été analysé par Image J, et pour chaque condition testée, au moins trois expériences indépendantes ont été analysées. Les images sont traitées *via* le logiciel Image J. Une coupe située au focus est sélectionnée pour étudier la colocalisation entre la GFP et l'ADN. Chaque image générée possède deux canaux : un correspondant à la fluorescence du Hoechst et un correspondant à la fluorescence de la GFP. Ces deux canaux sont scindés en deux images distinctes, et à chaque canal est attribué une couleur : la GFP est colorée en vert et l'Hoechst en rouge. Une image composite est ensuite créée en fusionnant les deux canaux ; les points de colocalisation apparaissent jaunes. La colocalisation au sein de chaque cellule est analysée individuellement : les cellules n'ayant pas intégré l'Hoechst de la même façon, le contraste et l'intensité doivent être ajustés pour chaque cellule.

Pour chaque image ont été déterminés manuellement :

- le nombre total de cellules,
- le nombre de cellules dont l'ADN est compacté,
- le nombre de cellules dont les foyers de fluorescence sont identifiés comme 'mega'.

Pour chaque cellule ont été déterminés:

- le nombre de 'mega' foyers de fluorescence,
- le pourcentage de cellules présentant un méga foyer,
- le pourcentage de cellules qui présentent un méga foyer colocalisant avec l'ADN,
- le nombre de foyers de fluorescence qui colocalisent avec l'ADN.

2.6 Fluorescence Recovery After Photobleaching (FRAP)

Afin de déterminer la dynamique de diffusion de la protéine Rpa2, le photoblanchiment de la protéine Rpa2 fusionnée à la GFP a été effectué sur des cellules en culture liquide dans différentes conditions : après exposition à la phléomycine ou à l'aphidicoline et au DMSO. Pour cela, à partir d'une préculture sur la nuit en YPC, une culture est ensemencée à DO=0,1. Après deux heures d'incubation à 45°C - 150 rpm, le volume nécessaire de solution stock de la drogue (ou de DMSO pour le contrôle lors d'exposition à l'aphidicoline) est ajouté puis les cellules sont incubées sur la nuit (environ 17h) à 45°C - 150 rpm.

Afin d'observer les cellules au confocal, les cellules doivent être immobilisées : 100 µl de cellules sont déposés sur une lamelle recouverte de 1 mg/mL poly-D-lysine chargée positivement pendant 30 minutes. Les cellules sont observées avec un microscope confocal inversé LEICA SP8-ST-WS. L'objectif utilisé pour les acquisitions est un 63X à huile, avec une ouverture numérique de 1,4. La mise au point est tout d'abord réalisée en observant les cellules en Contraste interférentiel différentiel (DIC). Puis le signal GFP est détecté en excitant l'échantillon à 470 nm au moyen d'un laser blanc dont la puissance laser est réglée pour la détection des cellules à 11,3 %. La taille des pixels collectés est de 144,50 nm * 144,50 nm.

Pour les trois conditions testées, le photoblanchiment des foyers de fluorescence a été mis au point. Chaque foyer de fluorescence présent au sein d'une cellule a été photoblanchi en excitant la GFP à 470 nm au moyen du laser blanc dont la puissance laser est cette fois-ci de 60 %. Le photoblanchiment du foyer de fluorescence s'effectue en ciblant trois points de photoblanchiment, qui font la taille du faisceau laser (environ 150 nm), dans le foyer à photoblanchir. Ceci aboutit au photoblanchiment de toute une zone (Figure 22).

Afin de pouvoir moyenner le signal de fluorescence collecté après photoblanchiment, le signal de fluorescence est collecté sur quinze images avant photoblanchiment et soixante après photoblanchiment. Une acquisition collectant le signal de fluorescence est effectuée toutes les 1 seconde 29 dans le cas d'un retour de fluorescence présumé rapide, contre 3 secondes 29 dans le cas d'un retour de fluorescence présumé plus lent.

L'analyse du retour de la fluorescence a été automatisée par MATLAB en tenant compte du bruit de fond et du photoblanchiment de la cellule due à l'imagerie en appliquant la formule suivante :

$$\frac{(signal\ de\ fluorescence\ des\ foyers - bruit\ de\ fond\ hors\ de\ la\ cellule)}{photoblanchiment\ de\ la\ cellule\ en\ imageant}$$

La soustraction du bruit de fond hors de la cellule au signal de fluorescence des foyers permet d'obtenir la véritable intensité du foyer. Et en divisant par le photoblanchiment de la cellule en imageant, on garantit d'obtenir l'intensité de fluorescence des foyers pour chaque temps, indépendamment du bruit de fond et de la perte du signal du à l'imagerie. Ceci permet de pouvoir quantifier le retour de fluorescence.

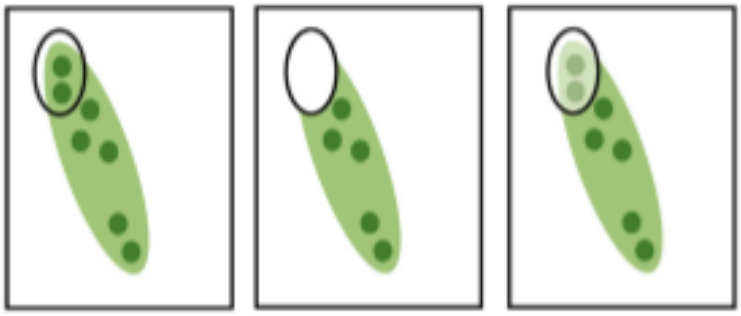


Figure 22. Photoblanchiment d'une zone comprenant deux foyers de fluorescence au sein d'une cellule.
D'après [31].

2.7 Western-blot

Pour cette quantification, 100 ml de cellules d'*Haloferax volcanii* à DO = 0,4 sont centrifugées. Les culots sont resuspendus dans 500 µL d'une solution de lyse constituée de 2M KCl et 10 Mm d'HEPES à pH 7 en alternant trois cycles de 15 secondes de sonication à 50% de puissance et 15 secondes de repos, dans un mélange éthanol et glace pour maintenir les cellules au froid. Les débris cellulaires et les cellules non lysées sont éliminées par une centrifugation de 60 minutes à 13 000 rpm effectuée à 4°C. Les lysats cellulaires sont séparés par éléctrophorèse en gel de polyacrylamide 4 à 12% contenant du sodium dodecyl sulphate (SDS-PAGE ; Invitrogen), pendant 35 minutes à 200 V dans du tampon TGS (Tris Glycine SDS). Ils sont ensuite transférés sur membrane de nitrocellulose en utilisant le Transblot, programme Turbo, gel mini TGX pendant 3 minutes à 25 V. Les membranes sont saturées avec un mélange de 50 % de solution de saturation de Licor (Sciencetec) et 50 % de PBS-T 1X (PBS = phosphate buffered saline) contenant 0,1% de Tween 20 pendant 2h00 sous agitation. Elles sont ensuite incubées pendant 1h00 avec l'anticorps primaire anti-GFP (ORIGENE) produit chez le lapin et dilué au 1:5000ème dans la solution de saturation de la membrane. Afin d'éliminer les anticorps ne s'étant pas fixés aux membranes, 4 lavages successifs de 5 minutes dans le PBS-T 1X sont effectués. Les membranes sont ensuite incubées pendant 45 minutes dans le noir avec l'anticorps secondaire anti-lapin couplé au marqueur IRDye 680 (Infra red fluorescend dye) dilué au 1:5000ème dans la solution de saturation. Avant de révéler les membranes, quatre lavages successifs de 5 minutes au PBS sont réalisés, puis les membranes sont excitées à 700 nm et analysées et quantifiées en utilisant le Licor Odyssey Imaging system.

3 RESULTATS

3.1 Etude de la protéine PCNA

Au sein de la fourche de réplication, PCNA assure deux fonctions : elle augmente la processivité de l'ADN polymérase et joue un rôle de plateforme de recrutement pour d'autres protéines de la réplication. PCNA étant un acteur clé de la réplication, nous souhaitions étudier la localisation cellulaire dans des cellules en prolifération pour comprendre sa dynamique de recrutement et la réponse cellulaire au stress réplicatif.

Pour cela, nous avons entrepris la construction d'une souche d'*H. volcanii* qui exprime la protéine PCNA fusionnée à la GFP. Afin de construire cette souche, nous avons choisi, en accord avec les études structurales sur PCNA d' *H. volcanii* [131], de réaliser la fusion à l'extrémité amino-terminale de la protéine. Chez *H. volcanii,* PCNA est un trimère en forme d'anneau qui interagit avec l'ADN en son centre et avec de nombreuses protéines *via* leurs motifs PIP vraisemblablement à la surface de la protéine. La fixation à l'ADN, qui a lieu au centre de la protéine, est une condition critique pour la fonction de PCNA. Or, les extrémités amino-terminales des trois sous-unités de la protéine se situent tous à la surface de PCNA (Figure 23), nous avons donc choisi de fusionner la GFP à l'extrémité amino-terminale de la protéine.

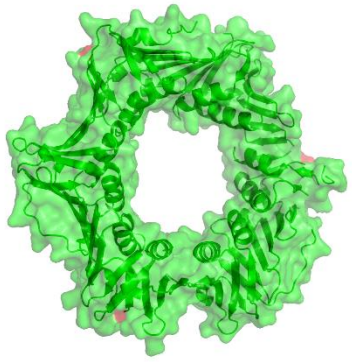


Figure 23. Structure de la protéine PCNA d'*H. volcanii.*
Les acides aminés phenylalanine à l'extrémité amino-terminale de chaque sous-unités sont en rouge. D'après, [131].

3.1.1 Clonage du plasmide intégratif

Une première étape de construction, à partir du plasmide intégratif (ou plasmide suicide) pTA131 permettant l'expression, au locus chromosomique, de PCNA fusionnée à la GFP par son extrémité N-terminale a été réalisée. Ce plasmide correspond au plasmide pFD4. Pour cela, deux méthodes ont été testées : la méthode de clonage par 'PCR 2 temps' et la méthode de clonage par 'SLIC' (3.3.1).

Dans un premier temps la méthode PCR 2 temps a été employée. L'amplification des régions « GFP » et « gène » a été optimisée en adaptant les températures d'hybridation des

amorces, indiquées dans le tableau ci-joint (Tableau 7). En revanche, l'amplification de la région « US-*pcna* » a abouti à une faible quantité d'ADN amplifiée. Pour cette PCR, trois températures d'hybridation ont été testées (58°C, 62°C, 64°C), toutes ont abouti à une quantité d'ADN après purification trop faible pour être détectée par gel d'agarose et très faible par mesure de l'absorbance (<5 ng/µL).

Afin de procéder à la PCR finale « US-pcna/GFP/*pcna* » au moyen des amorces RL137 et RL140, trois quantités égales de chaque fragment ont été testés : 12,5 ; 25 et 50 ng. Aucune des conditions testées n'a abouti à la formation de l'insert attendu. Ceci est probablement dû à l'efficacité réduite de l'amplification de la région « US-*pcna* » qui réduit les chances de produire le fragment de PCR finale.

Région à amplifier	Couple d'amorce	Tm°C	Longueur région amplifiée (pb)
«US-*pcna*»	RL137-RL138	64°C	390 pb
«GFP»	RL85bis-RL124	58°C	780 pb
«*pcna*»	RL139-RL140	62°C	380 pb
«US-GFP-*pcna*»	RL137-RL140	62°C	1 550 pb

Tableau 10. Conditions de PCR optimales pour générer l'insert « *gfp-pcna* » par PCR 2 temps.

Dans un deuxième temps une autre méthode de clonage permettant également de générer le plasmide intégratif a été testée : la méthode SLIC (II.3.a). Pour cette méthode d'autres amorces ont été utilisées. Les conditions de chaque PCR ont été optimisées (tableau 10). Le vecteur pTA131 est linéarisé par *Eco*RV puis déphosphorylé.

Région à amplifier	Couple d'amorce	Tm°C	Longueur région amplifiée (pb)
« US-*pcna* »	RL155-RL156	60°C	493 pb
« GFP »	RL157-RL158	60°C	762 pb
« *pcna* »	RL159-RL160	60°C	458 pb

Tableau 11. Conditions de PCR optimales pour générer l'insert « *gfp-pcna* » par SLIC.

Après purification des fragments sur gel d'agarose et traitement à l'activité exonucléasique 3'-5' de la T4 DNA polymérase, les fragments : « US-*pcna* », « GFP », « *pcna* » et pTA131 sont hybridés ensemble. Le produit de cette ligation est ensuite transformé dans des cellules compétentes commerciales d'*E. coli.* Après étalement des cellules sur boite LB ampicilline X-Gal-IPTG et incubation des boites à 37°C sur la nuit, les colonies sont comptées. Sur un total de 280 colonies, 6 colonies étaient blanches, indiquant l'insertion du fragment dans le vecteur.

Le contrôle des constructions s'est fait par PCR sur colonie et par digestion plasmidique. Deux des six colonies blanches et une colonie bleue ont été striées sur boite LB

Ampicilline. Une première étape de contrôle de ces constructions a été réalisée par PCR sur colonies au moyen du couple d'oligonucléotides servant à l'amplification du gène *gfp* : RL157-RL158. Une des deux colonies testées a permis l'obtention d'un produit de PCR de la taille attendue (700 pb). Ensuite, une purification de l'ADN plasmidique après préculture de cette colonie a permis de comparer les différents profils de digestion obtenus au profil de digestion théorique après digestion de l'ADN par deux enzymes de restriction : *Bst*B1 et *Pvu*II-HF.

	Fragments pour pTA131	**Fragments pour pFD4**
Digestion par *Pvu*II-HF	3 178 pb 448 pb	3 178 pb 2 132 pb
Digestion par *BstB*1	3626 pb	4 716 pb 594 pb

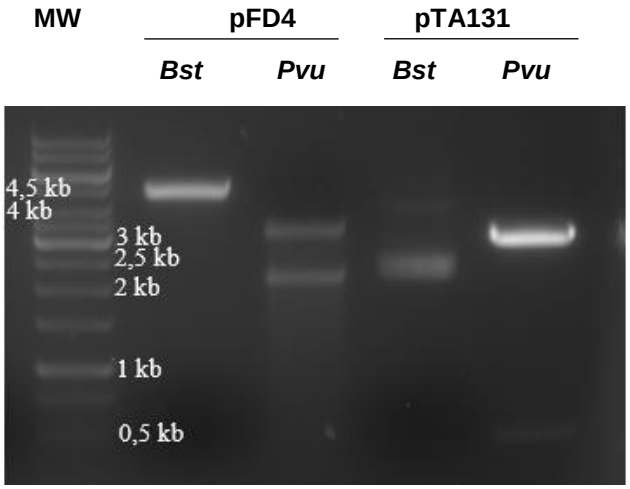


Figure 24. Vérification par restriction du plasmide pFD4.
Les pistes *Pvu* correspondent à la migration des fragments des plasmides pFD4 et pTA131 digérés par *Pvu*II-HF, les pistes *Bst* correspondent à la migration des fragments obtenus après digestion de pFD4 et pTA131 par *Bst*B1.

L'ADN de pFD4 digéré par *Bst*B1 migre entre 4500 pb et 5000 pb ce qui correspond à la longueur attendue mais on n'aperçoit pas la bande qui devrait migrer à 500 pb qui semble masquée par une « tache noire » correspondant au tampon de charge. Lorsqu'il est digéré par *Pvu*II-HF l'ADN du plasmide pFD4 migre à environ 3000 pb et 2000 pb, qui correspondent aux longueurs théoriques attendues.

Afin de s'assurer que pFD4 possède bien l'insert, une dernière étape de digestion par KpnI a été réalisée permettant de s'assurer du bon profil de digestion avant de séquencer le plasmide. Avant de procéder à la transformation d'*H. volcanii*, la séquence de ce plasmide a été validée par séquençage.

3.1.2 Construction d'une souche *Haloferax volcanii* permettant l'expression de PCNA fusionnée à la GFP

Afin d'exprimer la protéine GFP ::PCNA à partir du locus chromosomique, il faut remplacer l'allèle sauvage du gène *pcna* par l'insert cloné dans pFD4 par recombinaison homologue (technique du pop-in/pop-out). Il est important de noter que dans le cas d'un gène essentiel, l'insertion du plasmide par un évènement de recombinaison dans la région d'homologie du gène lui-même conduit à l'expression de la protéine fusionnée à la GFP uniquement, la protéine sauvage n'étant plus exprimée. A l'inverse, un évènement de recombinaison dans la région d'homologie en amont du gène (US) conduit à l'expression de la protéine sauvage uniquement (partie 2.4 et Figure 25).

La transformation de la souche sauvage (H26 : Δ*pyrE2*) par le plasmide pFD4 non méthylé a permis l'obtention de 6 colonies sur boîtes CASA (milieu sans uracile permettant de sélectionner l'intégration de *pyrE2* au chromosome). Seulement deux transformants ont pu être striés sur les quatre testés. Les deux transformants qui n'ont pas poussé étaient probablement Δ*pyrE2* et ont pu pousser sur les boites CASA grâce aux traces de milieu de régénération restant dans la suspension.

L'excision du plasmide résultant d'un second évènement de recombinaison (pop-out) est sélectionnée sur milieu CASA 5'FoA (+uracile) qui permet de contre-sélectionner la présence du marqueur *pyrE2* au chromosome. A partir des deux transformants, sélectionnés après pop-in et striés sur milieu CASA, ont été obtenues sur CASA 5'FoA 600 colonies pour l'un et 1300 colonies pour l'autre. La sélection des colonies ayant intégré *gfp::pcna* au locus chromosomique s'est faite par criblage des colonies par PCR au moyen du couple d'oligonucléotides RL161/RL162 qui amplifie le locus chromosomique comprenant les régions d'homologie utilisées pour l'insertion. Un produit de PCR de 2899 pb est attendu lorsque les cellules possèdent l'allèle *gfp::pcna* tandis qu'un produit de PCR de 2122 pb est attendu dans le cas où les cellules possèdent l'allèle *pcna* sauvage.

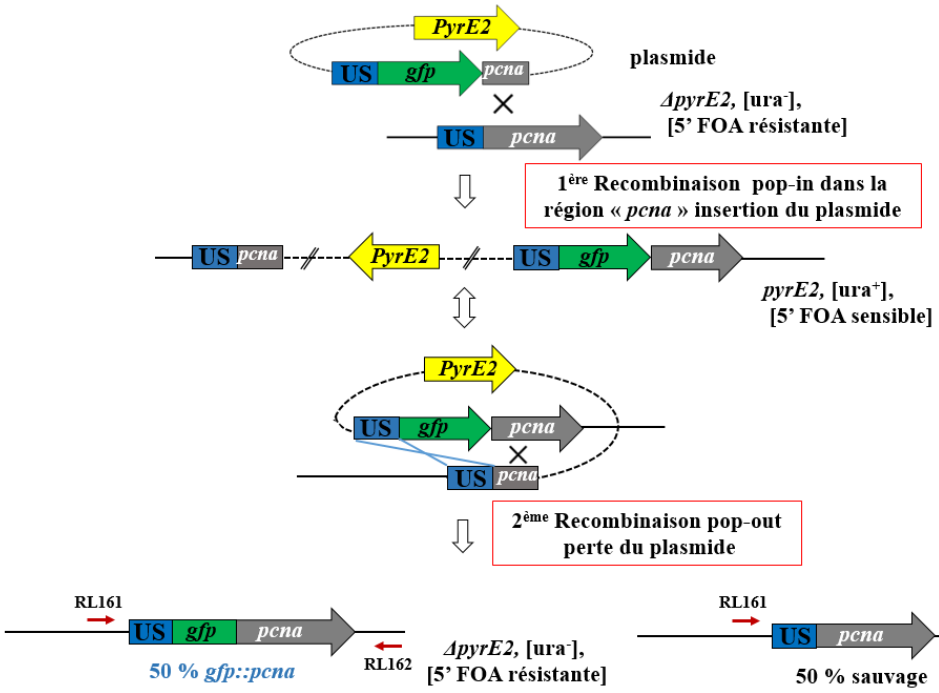


Figure 25. Schéma d'un pop-in/pop-out dans la région « *pcna* ».

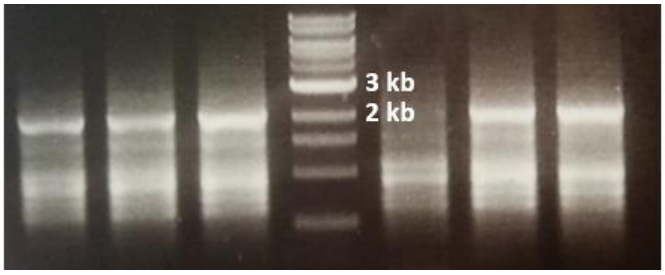


Figure 26. Crible par PCR des colonies après pop-in/pop-out pour l'expression de GFP::PCNA.
Sont représentés en exemple les PCR de 5 pop-out issus des pop-in 1 et 2 ainsi que la PCR de la souche sauvage, contrôle négatif de la construction

L'amplification de la région comprise entre RL161 et RL162 a été effectuée sur 137 colonies, dont 43 issues d'un transformant et 94 issues de l'autre. Toutes ont conduit à l'amplification d'une bande migrant sur gel d'agarose à 2 122 pb qui correspond à l'amplification de l'allèle sauvage de *pcna*. Ainsi, 100 % des colonies testées possèdent la copie du gène sauvage. Ces résultats permettent de conclure que la protéine GFP ::PCNA n'est pas fonctionnelle *in vivo* ou qu'elle n'est pas exprimée. Il est possible que la fusion de la GFP à l'extrémité N-terminale de la protéine empêche un repliement correct de la protéine, ou que la protéine soit correctement repliée mais ne puisse plus former un hétérotrimère

fonctionnel. Nous n'avons donc pas pu étudier la localisation dans la cellule vivante de la protéine PCNA en la fusionnant à la GFP.

3.2 Etude de la protéine FEN1

La protéine Flap Structure-specific endonuclease 1 (FEN1) est une exonucléase qui reconnait les structures de l'ADN 5' flap et clive la région simple-brin. Elle intervient dans la maturation des fragments d'Okazaki qui seront ensuite assemblés entre eux par la ligase et formeront le brin néosynthétisé antisens. Cette protéine fait partie des protéines clés de la réplication.

Pour étudier sa localisation cellulaire, nous avons entrepris la construction d'une souche d'*H. volcanii* qui exprime la protéine FEN1 fusionnée à la GFP par son extrémité amino-terminale au locus chromosomique.

3.2.1 Clonage du plasmide intégratif

Une première étape de construction à partir du plasmide intégratif pTA131 a été réalisée. Ce plasmide correspond à pFD1. Pour cela, la méthode de clonage PCR 2 temps a été employée (2.3.1). Les conditions de PCR ont été optimisées pour chaque fragment pour la première étape de PCR : « US-*fen1* », « *gfp* » et « *fen1* », ainsi que pour la dernière étape permettant la synthèse du fragment complet (Tableau 12 et Figure 27).

Région à amplifier	Couple d'amorce	Tm°C	Longueur région amplifiée (pb)
« US-*fen1* »	RL125-RL126	58°C	300 pb
« GFP »	RL85bis-RL124	58°C	780 pb
« *fen1* »	RL127-RL128	58°C	400 pb
« US-GFP-*fen1* »	RL125-RL128	55°C	1 480 pb

Tableau 12. Les conditions de PCR optimales pour chaque générer l'insert « *gfp-fen1* » par PCR 2 temps.

Afin de procéder à la PCR finale « US-*fen1*/*gfp*/*fen1*» au moyen des amorces RL125 et RL128, quatre quantités égales d'ADN de chaque fragment ont été testées : 12,5; 25; 50 et 100 ng. La PCR effectuée avec la quantité d'ADN de 100 ng a été sélectionnée (Figure 27).

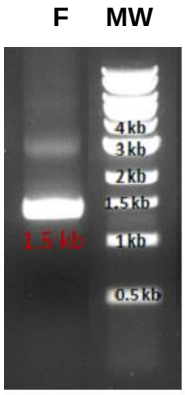


Figure 27. Produit de la PCR 2 temps correspond à l'insert « *gfp-fen1* ».
Piste F: PCR 2 temps *gfp::fen1*, MW : marqueur moléculaire de 1 kb

Le clonage suit le même principe que celui décrit dans la partie matériel et méthode (3.3.1). Le plasmide pTA131 a été linéarisé par l'enzyme *Eco*RV, dont le site restriction se trouve dans une partie du gène *lacZ* : *lacZα.* Il a ensuité été déphosphorylé pour éviter toute recircularisation du plasmide. Avant d'être inséré dans le vecteur pTA131, l'insert de 1,5 kb est traité par la polynucléotide kinase pour phosphoryler ses extrémités 5'. Il est ensuite purifié sur gel d'agarose, puis inséré dans la partie du gène *lacZ,* au sein du plasmide pTA131.

Les cellules compétentes d'*E. coli* ont été transformées avec le produit de ce clonage et étalées sur boite LB ampicilline. Les boites ont été incubées à 37 °C sur la nuit. A l'issu de cette étape, 800 colonies blanches et 70 colonies bleues ont poussé. Quinze colonies blanches ont été testées pour la présence de l'insert d'intérêt par PCR sur colonie au moyen des oligonucléotides RL28/RL29. La PCR conduit à l'amplification d'une séquence de 1812 pb si l'insertion est réussie contre l'amplification d'une séquence de 339 pb lorsqu'aucun insert n'est présent dans le plasmide. Une seule colonie parmi les quinze testées possède l'insert d'intérêt.

Afin de confirmer l'insertion du fragment « US-*fen1*/*gfp*/*fen1* » dans le plasmide pTA131, l'ADN plasmidique issu de cette colonie a été extrait puis digéré par les enzymes *Bst*B1 et *Pvu*II-HF.

L'ADN de pFD1 digéré par *Bst*B1 migre à environ 4500 pb ce qui correspond à la longueur attendue mais on n'aperçoit pas la bande qui devrait migrer à 639 pb qui semble masquée par une tache noire correspondant à la migration du BET (Figure 28). Lorsqu'il est digéré par *Pvu*II-HF l'ADN du plasmide pFD1 migre à environ 3000 pb et 2000 pb, qui correspondent aux longueurs théoriques attendues.

Afin de s'assurer que pFD1 possède bien l'insert, une dernière étape de digestion par KpnI a été réalisée, permettant de s'assurer du bon profil de digestion avant de séquencer le plasmide. Avant de procéder à la transformation d'*H. volcanii*, la séquence de ce plasmide a été validée par séquençage.

	Fragments pour pTA131	Fragments pour pFD1
Digestion par *Pvu*II-HF	3 178 pb 448 pb	3 178 pb 1 921 pb
Digestion par *Bst*B1	3 626	4 460 pb 639 pb

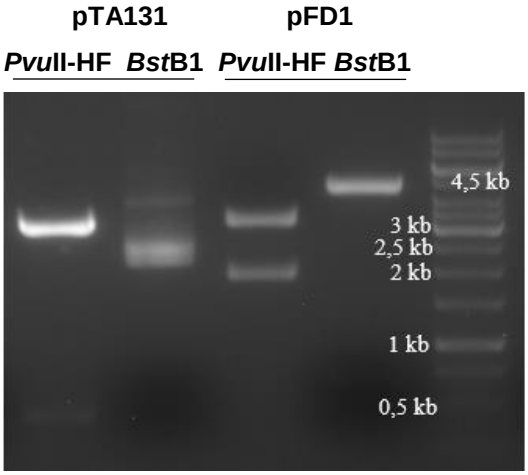


Figure 28. Vérification par restriction du plasmide pFD1.
Les pistes *Pvu*II HF correspondent à la migration des fragments des plasmides pFD1 et pTA131 digérés par *Pvu*II-HF, les pistes *Bst*b1 correspondent à la migration des fragments obtenus après digestion de pFD1 et pTA131 par *Bst*B1.

3.2.2 Construction d'une souche *Haloferax volcanii* permettant l'expression de FEN1 fusionnée à la GFP

Afin de pouvoir localiser et étudier *in vivo* Fen1 chez *H. volcanii,* il faut intégrer la séquence construite contenant FEN1 fusionnée à la GFP dans le génome.

Le plasmide pFD1 a été intégré au génome d' *H. volcanii* par pop-in (2.4). A l'issue de cette transformation dix colonies ont poussé. Parmi ces dix colonies quatre ont été striées sur milieu CASA. Afin de savoir si l'intégration du plasmide s'est effectuée par un évènement de recombinaison homologue dans la région en amont du gène (US) ou dans le gène lui-même (« *gène* »), les colonies ont été criblées par PCR sur colonie au moyen des oligonucléotides RL28 et RL76 qui s'hybrident de part et d'autre de la région d'homologie. Cette PCR conduit à l'amplification d'un fragment de 1007 pb si le pop-in s'est fait en US contre 1757 pb si le pop in s'est fait dans la région « *gène* ». Pour les colonies testées : le clone 1 a intégré le plasmide en US et le clone 2 a intégré le plasmide dans le gène lui-même (Figure 29).

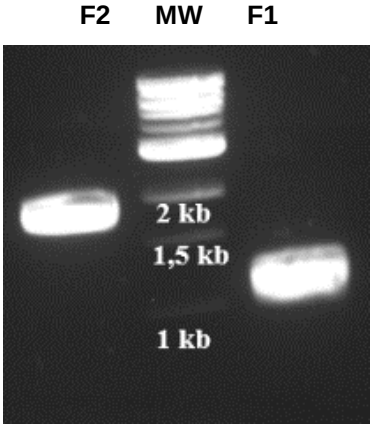


Figure 29. Crible par PCR des colonies après pop-in pour l'expression de GFP::Fen1.
La piste F2 correspond à l'amplification de la région contenant l'insert pour le clone 2, celui-ci migre à environ 1,8 kb ce qui correspond donc à un pop-in ayant eu lieu dans le gène lui-même «*gène*». La piste F1 correspond à l'amplification de la région contenant l'insert pour le clone 1, celui-ci migre à environ 1 kb ce qui correspond à un pop-in en US.

L'excision du plasmide résultant d'un second évènement de recombinaison (pop-out) est sélectionnée sur milieu CASA 5'FoA (+uracile) qui permet de contre-sélectionner la présence du marqueur *pyrE2* au chromosome. Cela a été réalisé pour les deux transformants sélectionnés après pop-in. A partir des colonies sélectionnées sur CASA + 5'FoA, 100 ont été repiquées sur YPC. La sélection de l'allèle *gfp::fen1* au locus chromosomique s'est faite par criblage des colonies par PCR au moyen du couple d'oligonucléotides RL85bis/RL124 qui amplifie la séquence de la *gfp.* Un produit de PCR d'environ 800 pb est attendu lorsque les cellules possèdent l'allèle *gfp::fen1* tandis qu'aucun produit de PCR n'est attendu lorsque les cellules possèdent l'allèle *fen1* sauvage (Figure 30).

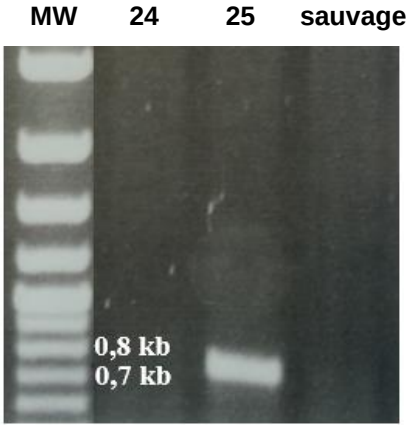


Figure 30. Crible par PCR des colonies après pop-in/pop-out pour l'expression de la GFP.
La piste 24 correspond à la PCR réalisée sur un pop-out qui a excisé son plasmide dans la région « gène » et donc qui n'a pas intégré la GFP. La piste 25 correspond à la PCR réalisée sur un pop-out qui a excisé le plasmide en US, et qui donc possèdent la protéine FEN1 fusionnée à la GFP. La piste sauvage correspond à la PCR réalisée sur la souche sauvage (H26) qui n'a pas la GFP

Parmi les 18 colonies issues du pop-out testées, trois colonies possèdent l'allèle *gfp::fen1* (toute issues du transformant 2). Le pourcentage de réussite d'intégration au chromosome de la séquence de la protéine fusionnée à la *gfp* est de 18%, ce qui représente un faible pourcentage d'intégration. Néanmoins, peu de colonies ont été testées. Un clone a été mis en collection correspondant à la souche HVFD19.

3.2.3 Etude de la fonctionnalité de la souche exprimant FEN1 fusionnée à la GFP

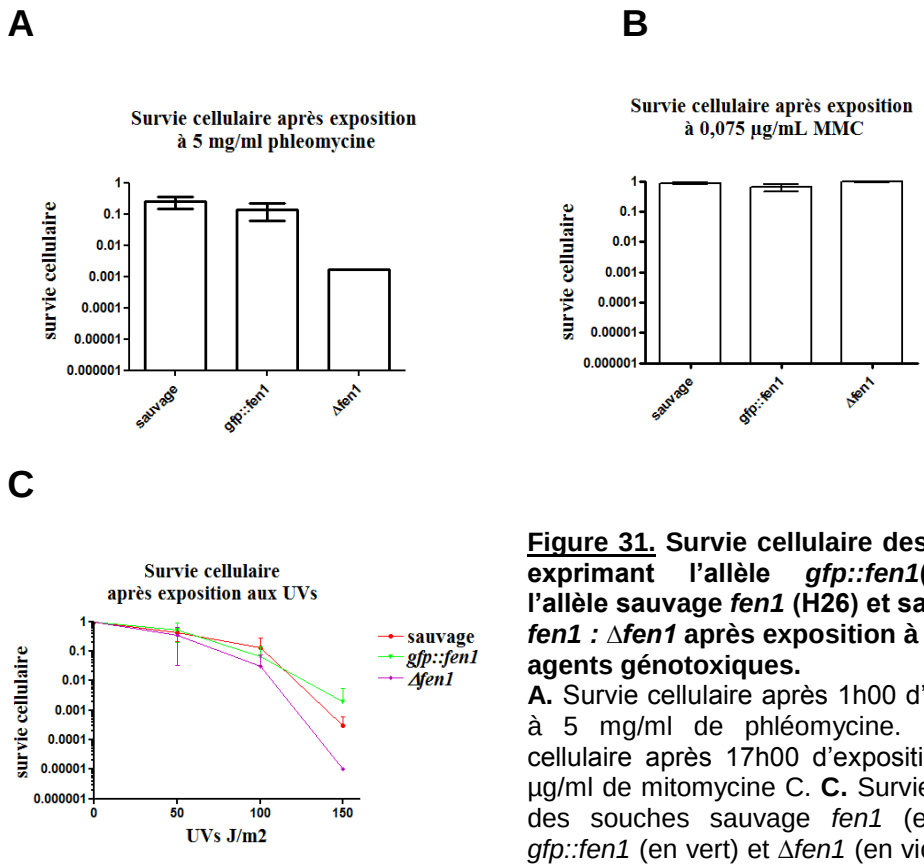
La souche d'*H. volcanii* exprimant FEN1 fusionnée à la GFP a été caractérisée. Chez les eucaryotes, FEN1 est impliquée dans des mécanismes de réparation de l'ADN par excision de base. Le gène *fen1* n'est pas essentiel mais son absence augmente la sensibilité aux agents mutagènes. Chez *H. volcanii, fen1* n'est pas un gène essentiel, mais sa délétion entraine un retard de croissance significatif comparé à la souche sauvage. Il a également été montré que la délétion du gène *fen1* diminue la survie de cellules en réponses aux dommages de l'ADN après exposition à des doses d'UVs importantes [99]. Des études génétiques réalisées chez *H. volcanii* ont montré que les protéines UvrA, UvrB et UvrC, homologues aux protéines bactériennes, sont responsables de la réparation de l'ADN par excision de nucléotides (NER) (Lestini et al., 2010). La sensibilité du mutant Δ*fen1* à des fortes doses d'UVs pourrait refléter un rôle de Fen1 dans la réparation de dommages de l'ADN non réparés par UvrABC. Ainsi j'ai déterminé le temps de génération ainsi que la survie des cellules après exposition à différents agents génotoxiques pour la souche *gfp*::*fen1,* que j'ai comparés à la souche exprimant le gène sauvage et le mutant de délétion Δ*fen1*.

Souche	Temps de génération
Δ*pyrE2* (H26)	2,1 h ± 0,6
Δ*pyrE2 gfp::fen1* (HVFD19)	2,6 h ± 0,4
Δ*pyrE2* Δ*fen1* (H938)	3,8 h

Tableau 13. Temps de génération des souches exprimant l'allèle *gfp::fen1,* l'allèle sauvage *fen1* et délétées pour *fen1* : Δ*fen1*

Le temps de génération de la souche exprimant FEN1 fusionnée à la GFP (2,6 h) n'est pas significativement différent de celui de la souche exprimant FEN1 sous sa forme native (2,1 h) (t-test, non significatif, p value =0,353)

La résistance à différents agents génotoxiques provoquant des lésions de l'ADN a ensuite été testée, notamment à la phléomycine, la mitomycine C et les UVs (Figure 31).



<u>Figure 31.</u> Survie cellulaire des souches exprimant l'allèle *gfp::fen1*(HVFD19), l'allèle sauvage *fen1* (H26) et sans l'allèle *fen1 :* Δ*fen1* après exposition à différents agents génotoxiques.
A. Survie cellulaire après 1h00 d'exposition à 5 mg/ml de phléomycine. **B.** Survie cellulaire après 17h00 d'exposition à 0,02 µg/ml de mitomycine C. **C.** Survie cellulaire des souches sauvage *fen1* (en rouge), *gfp::fen1* (en vert) et Δ*fen1* (en violet) après exposition à trois différentes doses d'UV : 50 J/m^2, 100 J/m^2 et 150 J/m^2.

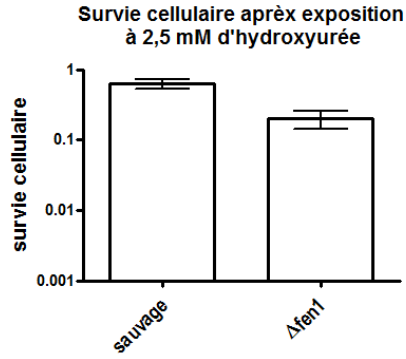
La phléomycine provoque des lésions à l'ADN : elle induit des cassures double brin qui sont réparées par recombinaison homologue. L'exposition à une concentration élevée de phléomycine (5 mg/ml) de la souche exprimant l'allèle *gfp::fen1* n'entraine pas de défaut de survie comparé à la souche sauvage. En revanche, on peut observer un défaut de survie important de la souche Δ*fen1* comparé à la souche sauvage, (noter que cette expérience n'a été réalisée qu'une seule fois pour la souche Δ*fen1*). Bien qu'il faille reproduire cette expérience, ce résultat suggère que la protéine FEN1 est impliquée dans des processus de réparation requis après exposition à la phléomycine. La mitomycine C réagit avec l'ADN induisant des pontages covalents d'ADN interbrins qui requiert leur suppression par la voie de réparation par excision de nucléotides (NER) ou par recombinaison homologue (HR). L'exposition à la mitomycine C de la souche exprimant l'allèle *gfp::fen1* n'entraine pas de défaut de survie comparé à la souche sauvage et à la souche Δ*fen1*.

L'exposition aux UVs induit des dimères de pyrimidines qui perturbent la structure double hélice de l'ADN et provoquent au cycle suivant des cassures double brin. Ces dommages de l'ADN peuvent être réparés par la voie de réparation par excision de nucléotides (NER). L'exposition aux UVs de la souche exprimant l'allèle *gfp::fen1* n'entraine pas de défaut de survie comparé à la souche sauvage. En revanche, on peut observer un défaut de survie pour la souche n'exprimant pas *fen1 (Δfen1)* comparée à la souche sauvage à partir d'une exposition élevée aux UVs (150 J/m^2), qui pourrait refléter un rôle de Fen1 dans la réparation de dommages de l'ADN non réparés par UvrABC..

Les cellules Δ*fen1* semblent plus sensibles aux agents génotoxiques conduisant à des lésions de l'ADN comme les UV et la phléomycine que les cellules exprimant l'allèle sauvage *fen1* (H26). Ceci suggère un rôle de FEN1 dans le processus de réparation de l'ADN. Ainsi FEN1 interviendrait potentiellement dans la maturation des fragments d'Okazaki mais aussi dans les processus de réparation de l'ADN par excision de nucléotides et par recombinaison homologue. Les cellules exprimant l'allèle *gfp::fen1* présentent une sensibilité comparable aux cellules exprimant l'allèle sauvage *fen1*. Ceci suggère que la protéine GFP::FEN1 permet la survie après dommages de l'ADN par les UVs et l'exposition à la phléomycine. L'ensemble de ces résultats suggèrent que la protéine FEN1 fusionnée à la GFP est fonctionnelle *in vivo.*

3.2.4 Localisation *in vivo* de FEN1

FEN1 intervient dans la maturation des fragments d'Okazaki, lors de la phase d'élongation de la réplication. La localisation *in vivo* de cette protéine a donc été observée selon la phase de croissance, et après exposition à l'hydroxyurée, qui induit un stress réplicatif. L'argument d'une action de l'hydroxyurée sur la ribonucléotide réductase chez les archées est controversé. En effet, cette enzyme induit des arrêts de la fourche de réplication en inhibant la ribonucléotide réductase chez les eucaryotes et la majorité des bactéries. Or chez les archées, les résidus clés du site catalytique inhibé par l'hydroxyurée ne sont pas complétement identiques. Néanmoins, une étude très récente menée chez l'hyperthermophile *S. solfataricus* montre que l'hydroxyurée inhibe la réplication, non pas en ciblant la ribonucléotide réductase mais en ciblant une sous-unité régulatrice de la primase (Liew et al, 2016). Or nous avons pu observer une sensibilité à 2,5 mM d'hydroxyurée pour les cellules Δ*fen1* (H938) (Figure 32) qui suggère que son action est possible chez *H. volcanii.*



<u>Figure 32.</u> Survie cellulaire après exposition à 2,5 mM d'hydroxyurée.

La Figure 33 A montre le signal d'autofluorescence observé dans des cellules sauvages (H26) n'exprimant pas la GFP. Le signal de fluorescence des cellules *gfp*::*fen1* a été observé à différents temps lors de la phase de croissance, correspondant à des densités optiques mesurées de 0,1 ; 0,3 ; 0,6 et 1 (Figure 33 B, C**,** D et E). Les cellules exprimant FEN1 fusionnée à la GFP (HVFD19) présentent un signal de fluorescence plus important, et des foyers de fluorescence peuvent être observés dans certaines cellules (jamais observé dans des cellules n'exprimant pas la GFP) (Figure 33 B –flèches jaunes). Le signal de fluorescence observé dans les cellules *gfp::fen1* ne semble pas varier au cours de la phase de croissance.

Afin de récolter toute la fluorescence de la cellule, trente images ont été prises tous les 0,250 µm de part et d'autre de la coupe au focus. Les images au focus ont ensuite été projetées en z sur une seule image qui récolte le maximum d'intensité pour chaque coupe au focus. Ces expériences ont été répétées trois fois et analysées quantitativement par le logiciel IMARIS. Les paramètres de détection ont été ajustés pour la détection des cellules mais la détection des foyers de fluorescence n'a pu être optimisée. En effet les foyers de fluorescence ne sont pas assez intenses pour être détectés de façon fiable par le logiciel IMARIS. Néanmoins, une analyse manuelle montre que le nombre de foyer de fluorescence reste constant selon la phase de croissance. En revanche, différentes données ont pu être collectées et analysées concernant les cellules : l'intensité de fluorescence moyenne par cellule (Figure 33 F) et la surface cellulaire moyenne (Figure 33 G). Les analyses quantitatives réalisées *via* IMARIS relatives aux images de microscopie de fluorescence des cellules *gfp::fen1* sont présentées Figure 33 F et G.

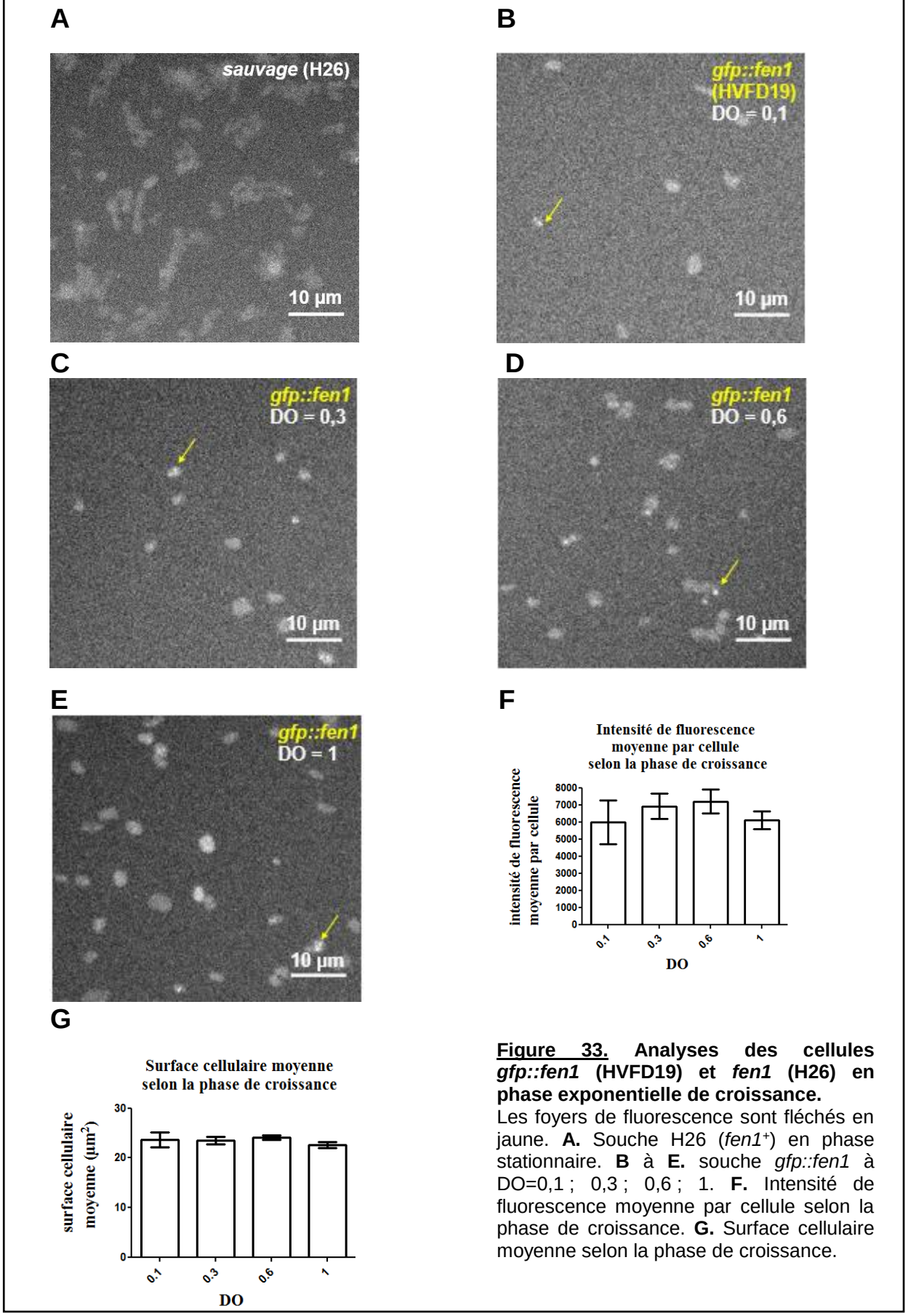
Une surface cellulaire moyenne d'environ 23 µm^2 est observée de façon constante tout au long de la phase de croissance (Figure 33 G). L'intensité de fluorescence moyenne par cellule est également constante (Figure 33 F). Nous n'avons pas observé de localisation spécifique de la protéine GFP::FEN1 selon la phase de croissance, ni de variation de son niveau d'expression.

N’ayant pas observé de localisation spécifique de la protéine GFP::FEN1 dans des cellules en prolifération, nous avons étudier la localisation de la protéine GFP::FEN1 lorsque la cellule est soumise à un stress réplicatif. Pour cela, les cellules ont été exposées à l’hydroxyurée qui entraine une diminution du pool de nucléotides disponibles et ainsi un ralentissement de la réplication.

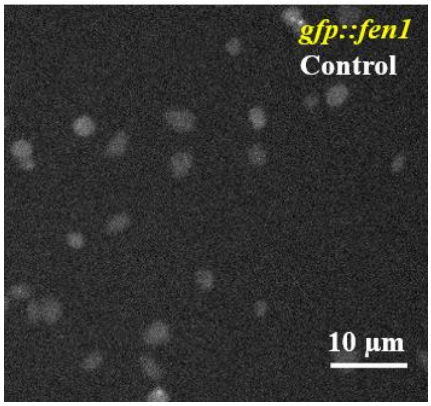
Sur la Figure 34 figurent les images de microscopie de fluorescence des cellules *gfp::fen1*, les analyses quantitatives réalisées *via* IMARIS associées ainsi que la survie cellulaire des cellules exposées à 5 mM d’hydroxyurée pendant 17h. Les mêmes traitements des données et paramètres de détection et d’analyse ont été employées que celles expliquées au paragraphe précédent. Cela a permis de quantifier : la surface cellulaire moyenne (Figure 34 C) et l’intensité de fluorescence moyenne par cellule (Figure 34 D).

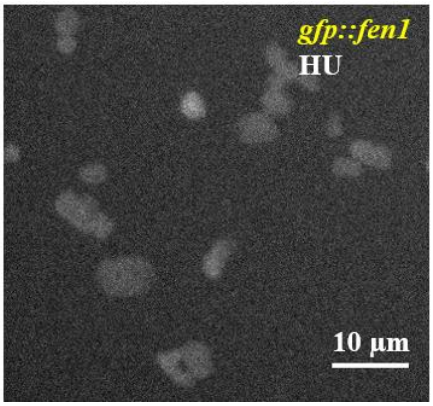
L’analyse quantitative des images montre une faible augmentation non significative de la surface cellulaire moyenne lorsque les cellules ont été traitées à l’hydroxyurée (Figure 34 C). Par ailleurs, l’intensité moyenne du signal de fluorescence des cellules traitées à l’hydroxyurée augmente légèrement mais de façon non significative par rapport aux cellules contrôles (Figure 34 D). Ceci suggère une expression constante de la protéine FEN1 dans ces deux conditions ou mis en perspective avec l’absence de sensibilité à l’hydroxyurée (Figure 34 E) un faible effet sur ces cellules.

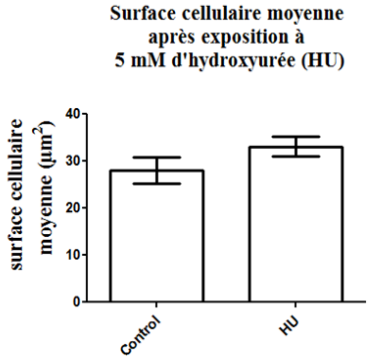
L’ensemble de ces résultats nous permet de conclure que le signal de fluorescence des cellules exprimant GFP::Fen1 est faible, reflétant une expression de la protéine relativement faible dans la cellule. Néanmoins ce niveau d’expression apparaît constant lors de la phase de croissance. Cependant, aucune localisation spécifique de la protéine GFP::FEN1 n’est observée dans des cellules en cours de prolifération comme dans des cellules exposées à l’hydroxyurée. Ceci serait-il le reflet d’une détection des foyers trop tardive ou d’une dynamique de la protéine Fen1 trop rapide ?

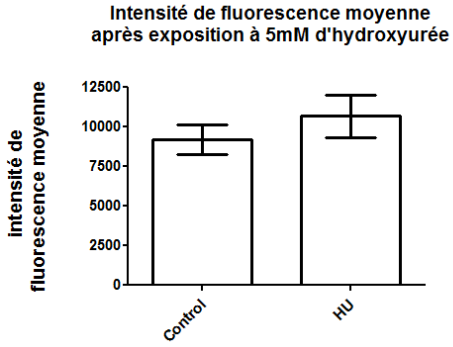


<u>Figure 33.</u> Analyses des cellules *gfp::fen1* (HVFD19) et *fen1* (H26) en phase exponentielle de croissance.
Les foyers de fluorescence sont fléchés en jaune. **A.** Souche H26 (*fen1*$^{+}$) en phase stationnaire. **B** à **E.** souche *gfp::fen1* à DO=0,1 ; 0,3 ; 0,6 ; 1. **F.** Intensité de fluorescence moyenne par cellule selon la phase de croissance. **G.** Surface cellulaire moyenne selon la phase de croissance.









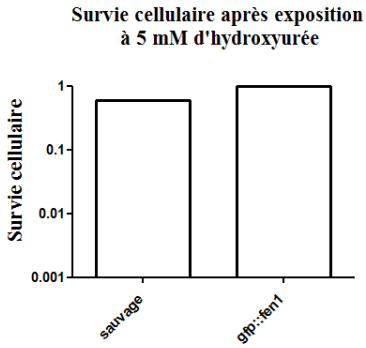


Figure 34. Analyse des cellules après 17h00 d'exposition à l'hydroxyurée.
Les expériences ont été réalisées après 17h00 d'exposition à 5mM d'hydroxyurée (HU) ou à l'eau dans la condition contrôle. **A.** Maximum d'intensité pour les cellules *gfp::fen1* en phase stationnaire (DO=1,7) exposées à l'eau. **B.** Maximum d'intensité pour les cellules *gfp::fen1* exposées à l'HU. **C.** Surface cellulaire moyenne. **D.** Intensité de fluorescence moyenne. **E.** Viabilité des cellules *gfp ::fen1* et sauvages après 17h00 d'exposition à 5 mM d'hydroxyurée.

3.3 Etude de la protéine Orc1

Les protéines Orc sont des protéines intervenant lors de l'initiation de la réplication. Elles se fixent sur les origines de réplication et participent à la formation du complexe pré-réplicatif. Ceci conduit à l'assemblage du réplisome qui est nécessaire pour l'initiation de la réplication et le chargement de l'hélicase réplicative [143]. Chez les archées, les gènes codant pour les protéines Orc sont toujours proches des origines de réplication et sont essentiels à la réplication. Chez *H. volcanii,* les séquences codant pour les seize protéines putatives Cdc6/Orc ne sont pas nécessairement proches des quatre origines de réplication présentes, et jusqu'à ce jour les méthodes de génétique inverse n'ont pas réussi à montrer le caractère essentiel des gènes *cdc6/orc*, ni le rôle précis de ces protéines [30] et leur régulation. Des études, réalisées chez les haloarchées, portant sur la régulation du transcriptome par des changements de pH extrêmes, ont montré que les gènes individuels faisant partie des familles de gènes multi-paralogue comme *cdc6/orc* sont régulés par des stress environnementaux [144]. Qu'en est-il pour Cdc6/Orc1 spécifiquement ? Le gène *cdc6/orc1* est localisé près de l'OriC1 et il a été montré que la protéine est exprimée. De plus, cette protéine a déjà été étudiée au sein du laboratoire. C'est pourquoi j'ai choisi cette protéine pour étudier les processus d'initiation de la réplication.

Afin d'étudier *in vivo* le rôle de Cdc6/Orc1 (nommé Orc1 pour le reste de l'étude) chez *H. volcanii,* j'ai réalisé la construction de souches exprimant, au locus chromosomique, Orc1 fusionnée à la GFP par son extrémité N-terminale. Ceci a été réalisé dans deux contextes génétiques différents : dans un contexte sauvage (H26) et dans un contexte sans origine de réplication (H1804). De façon surprenante, il a été noté que la souche ne possédant aucune origine de réplication pousse plus rapidement que la souche sauvage [29]. Ceci permettra de répondre aux questions suivantes : Quelle est sa localisation cellulaire ? Quelle est sa dynamique de recrutement ? Comment sa dynamique de recrutement et sa localisation sont affectées dans une souche sans origine de réplication [137, 138, 145] ?

3.3.1 Clonage du plasmide intégratif

Une première étape de construction, à partir du plasmide intégratif pTA131, permettant l'expression, au locus chromosomique, de Orc1 fusionnée à la GFP par son extrémité N-terminale a été réalisée. Ce plasmide correspond au plasmide pFD3. Pour cela la technique de clonage par PCR 2 temps a été employée (cf. 3.3.1). Les conditions de PCR ont été optimisées pour chaque fragment : « US-*orc1* », « gfp » et « *orc1* » lors de la première étape

de PCR, ainsi que pour la dernière étape permettant la synthèse du fragment complet (Tableau 13 et Figure 35).

Région à amplifier	Couple d'amorce	Tm°C	Longueur région amplifiée (pb)
«US-*orc1*»	RL133-RL134	62°C	270 pb
«GFP»	RL85bis-RL124	58°C	780 pb
«*orc1*»	RL135-RL136	66°C	407 pb
«US-GFP-*orc1*»	RL133-RL136	62°C	1 457 pb

Tableau 13. Les conditions de PCR optimales pour générer l'insert « *gfp-orc1*» par PCR 2 temps

Afin de procéder à la PCR finale « US-*orc1*/*gfp*/*orc1*» au moyen des amorces RL133 et RL136, quatre quantités égales d'ADN de chaque fragment ont été testées : 12; 25; 50; 100 ng. La PCR effectuée avec la quantité d'ADN de 100 ng a été sélectionnée (Figure 35).

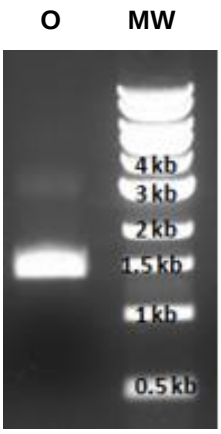


Figure 35. Produit de la PCR 2 temps correspondant à l'insert « *gfp-orc1* ».
Piste O: PCR 2 temps *gfp::orc1*, MW : marqueur moléculaire de 1 kb.

Le clonage suit le même principe que celui décrit dans la partie matériel et méthode (3.3.1). Le plasmide pTA131 a été linéarisé par l'enzyme *Eco*RV, dont le site restriction se trouve dans une partie du gène *lacZ : lacZα.* Il a ensuite été déphosphorylé pour éviter toute recircularisation du plasmide. Avant d'être inséré dans le vecteur pTA131, l'insert de 1,5 kb est traité par la polynucléotide kinase pour phosphoryler ses extrémités 5'. Il est ensuite purifié sur gel d'agarose, puis inséré dans la partie du gène *lacZ,* au sein du plasmide pTA131.

Les cellules compétentes d'*E. coli* ont été transformées avec le produit de ce clonage et étalées sur boite LB ampicilline. Les boites ont été incubées à 37 °C sur la nuit. A l'issu de cette étape, 800 colonies blanches et 80 colonies bleues ont poussé. Huit de ces colonies blanches ont été testées pour la présence de l'insert d'intérêt par PCR au moyen des oligonucléotides RL28/RL29. La PCR conduit à l'amplification d'une séquence de 1755 pb si l'insertion est réussie contre l'amplification d'une séquence de 339 pb lorsqu'aucun insert n'est

présent dans le plasmide. Six colonies parmi les huit colonies blanches testées possèdent l'insert migrant à environ 1,5 kb.

Afin de confirmer l'insertion du fragment « US-*orc1*/*gfp*/*orc1*» dans le plasmide pTA131, l'ADN plasmidique contenu dans ces colonies a été extrait puis digéré par les enzymes *Bst*B1 et *Pvu*II-HF. Les trois plasmides testés présentent le profil de digestion attendu (Figure 36). Avant de procéder à la transformation d'*H. volcanii,* les séquences de ces plasmides ont été validées par séquençage.

	Fragments pour pTA131	**Fragments pour pFD3**
Digestion par *Pvu*IIHF	3178 pb 448 pb	3 178 pb 1864 pb
Digestion par *Bst*b1	Ne coupe pas	-linéarisé - 5042 pb

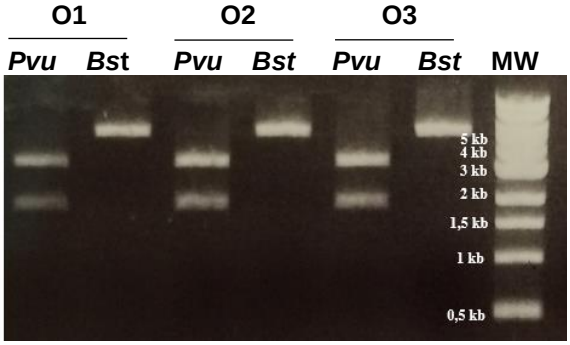


Figure 36. Vérification par restriction du plasmide pFD3.
Les pistes O1, O2 et O3 correspondent aux trois plasmides extraits des clones sélectionnés digérés. Les pistes *Pvu* correspondent à la migration des fragments de chaque plasmide digéré par *Pvu*II-HF, les pistes *Bst* correspondent à la migration du fragment obtenu après digestion du plasmide par *Bst*B1.

3.3.2 Construction des souches d'*H. volcanii* permettant l'expression de Orc1 fusionnée à la GFP

Afin d'exprimer la protéine GFP::Orc1 au locus chromosomique, il faut remplacer l'allèle sauvage du gène *orc1* par l'insert cloné dans pFD3 par recombinaison homologue (technique du pop-in/pop-out). Comme expliqué précédemment, ceci a été réalisé au sein de deux souches : la souche sauvage de laboratoire Δ*pyrE2* (H26) et la souche sans origines de réplication Δ*pyrE2* Δ*trpA* Δ*oriC1* Δ*oriC2* Δ*oriC3* Δ*ori-pHV4* (H1804) que j'appellerai Δ*ori* dans la suite de ce manuscrit.

Dans le contexte sauvage (H26) :

La transformation de la souche sauvage Δ*pyrE2* par le plasmide pFD3 non méthylé a permis l'obtention de 83 colonies sur boite CASA (milieu sans uracile permettant de sélectionner l'intégration de *pyrE2* au chromosome). Quatre transformants ont été striés sur milieu CASA afin d'obtenir des colonies isolées, tous ont poussé.

L'excision du plasmide résultant d'un second événement de recombinaison (pop-out) est sélectionnée sur milieu CASA + 5'-FoA (+uracile) qui permet de contre-sélectionner la présence du marqueur *pyrE2* au chromosome. Cela a été réalisé pour deux des quatre transformants sélectionnés après pop-in. A partir des colonies sélectionnées sur CASA + 5'FoA, 100 ont été repiquées sur YPC. Pour toutes, la sélection de l'allèle *gfp::orc1* s'est faite par criblage des colonies par PCR au moyen du couple d'oligonucléotides RL85 bis et RL124 amplifiant la séquence de la *gfp*. Un produit de PCR d'environ 800 pb est attendu lorsque les cellules possèdent l'allèle *gfp::orc1* tandis qu'aucun produit de PCR n'est attendu dans le cas où les cellules possèdent l'allèle *orc1* sauvage .

Parmi les colonies issues du pop-out du transformant 1, huit colonies sur douze possèdent l'allèle *gfp::orc1,* tandis que parmi les colonies issues du pop-out du transformant 2, une colonie sur dix possède l'allèle *gfp::orc1.* Ainsi, sur vingt-deux colonies testées, neuf colonies possèdent l'allèle *gfp::orc1.* Le pourcentage de réussite d'intégration au chromosome de la séquence *gfp::orc1* est de 41 %. Un clone a été mis en collection correspondant à la souche HVFD18. Afin de vérifier si l'insertion s'est faite au locus pour les clones sélectionnés, une PCR amplifiant la région intégrée en dehors des régions d'homologie a été faite au moyen des oligonucléotides RL141 et RL164. Elle conduit à l'amplification d'une séquence de 2988 pb lorsque les cellules possèdent l'allèle *gfp::orc1* contre 2238 pb lorsque les cellules possèdent l'allèle *orc1* (Figure 37).

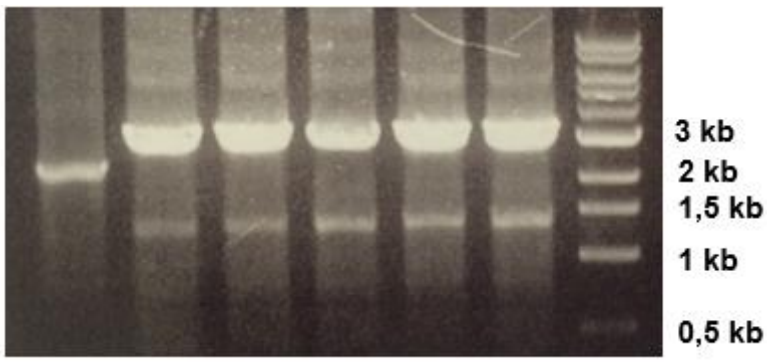


Figure 37. Amplification par PCR de la région *orc1* après pop-in/pop-out pour intégrer l'allèle *gfp::orc1.*
Les cellules exprimant *orc1* sont les cellules contrôles (sauvage), les différents clones validés par la PCR amplifiant la *gfp* et exprimant *gfp::orc1* au locus chromosomique sont les clones *gfp::orc1.*

Dans le contexte sans origines de réplication :

La transformation de la souche ∆*ori* (H1804) par le plasmide pFD3 non méthylé a permis l'obtention de quatre colonies sur boite CASA. Ils ont ensuite été striés sur milieu CASA afin d'obtenir des colonies isolées. L'excision du plasmide résultant d'un second événement de recombinaison (pop-out) est sélectionné sur milieu CASA + 5'-FoA. Ceci a été réalisé pour deux des quatre transformants sélectionnés après pop-in. A partir des colonies sélectionnées sur CASA + 5'FoA, 100 ont été repiquées sur YPC. Pour toutes la sélection de l'allèle *gfp::orc1* s'est faite par colony lift en utilisant une sonde *gfp*. Cette technique permet de discriminer les colonies avec l'allèle *gfp::orc1* des colonies avec l'allèle sauvage *orc1.* Les colonies sont déposées sur membrane qui après plusieurs lavages est incubée avec la sonde GFP s'hybridant avec les séquences GFP présentes. Parmi les pop-out testés, huit colonies sur cinquante possèdent l'allèle *gfp::orc1.* Le pourcentage de réussite d'intégration au chromosome de la séquence *gfp::orc1* est de 16 % (Figure 38). Un clone a été mis en collection correspondant à la souche HVFD43.

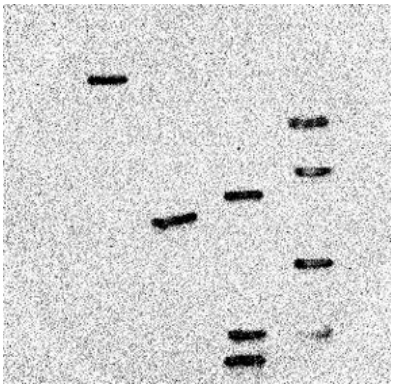


Figure 38. Hybridation sur colonies pour tester la présence de l'allèle *gfp::orc1* réalisé sur les pop-out des cellules ∆*ori.* Les colonies apparaissant en noir sont les colonies avec lesquelles la sonde *gfp* s'est hybridée. Huit colonies possèdent donc l'insert *gfp::orc1* au locus chromosomique.

3.3.3 Caractéristiques des souches *gfp::orc1* et ∆*ori gfp::orc1*

Orc1 n'est pas un gène essentiel chez *H.* volcanii, et la délétion de ce gène n'entraine pas de phénotype [71]. Néanmoins nous souhaitions répondre à la question suivante : est-ce que l'insertion de l'allèle *gfp::orc1* a un impact sur le temps de génération des souches ∆*pyrE2 gfp::orc1* et ∆*ori gfp::orc1* ?

J'ai mesuré le temps de génération pour les souches possédant l'allèle *gfp::orc1* dans la souche ayant conservé ses origines de réplication et dans la souche sans origine de réplication que j'ai comparé respectivement à la souche sauvage (H26) et à la souche sans

origine de réplication (H1804). J'ai également déterminé la survie des cellules après expositions à différents agents génotoxiques.

Souche	Temps de génération
Sauvage, Δ*pyrE2* (H26)	2,1 h ± 0,6
Δ*pyrE2, gfp::orc1* (HVFD18)	1,8 h ± 0,6
Δ*ori* (H1804)	1 h [29]
Δ*ori, gfp::orc1* (HVFD43)	1,28 h ± 0,55

Tableau 14. Tableau récapitulatif de la viabilité et du temps de génération des souches possédant l'allèle *gfp::orc1* et l'allèle sauvage *orc1* avec et sans origine de réplication

Le temps de génération de la souche exprimant l'allèle *gfp::orc1* (HVFD18) est de 1,8 h, celui-ci est comparable à celui de la souche sauvage (H26) qui est de 2,1 h. Le temps de génération de la souche exprimant l'allèle *gfp::orc1* sans origine de réplication (HVFD43) est de 1,28 h, également comparable à celui de la souche exprimant l'allèle sauvage *orc1* sans origine de réplication (H1804) qui est de 1 h. Ainsi, l'insertion de l'allèle *gfp::orc1* ne semble pas avoir d'impact sur le temps de génération des cellules. La protéine GFP::Orc1 ne semble pas avoir d'effet dominant négatif si l'on considère qu'elle fait partie d'un complexe.

La survie cellulaire après exposition à la phléomycine, à la mitomycine C et aux UVs a ensuite été déterminée (Figure 39). L'exposition des cellules à la phléomycine provoque des dommages à l'ADN en induisant des cassures double brin. L'exposition des cellules à la mitomycine C réagit avec l'ADN et génère des pontages interbrins covalents qui requière l'action du complexe RH (recombinaison homologue) et NER (réparation par excision des nucléotides). La survie cellulaire de la souche sauvage possédant l'allèle *orc1* et de la souche possédant l'allèle *gfp::orc1* après exposition à la phléomycine (5 mg/mL) (Figure 39 A) et à la mitomycine C de 0,075 µg/mL (Figure 39 B) sont semblables. L'exposition aux UVs, ici aux UVs-C, induit des dimères de pyrimidines qui perturbent la structure double hélice de l'ADN et provoque au cycle suivant des cassures double brin. Ces dommages de l'ADN peuvent être réparés par la voie de réparation par excision de nucléotides (REN). En accord avec les résultats précédents, l'exposition aux UVs (Figure 39 C) de la souche exprimant l'allèle *gfp::orc1* n'entraine pas de défaut de survie comparé à la souche sauvage, et ce même après une exposition à une dose élevée (150 J/m^2).

L'ensemble de ces résultats permettent de conclure que la protéine GFP::Orc1, si elle est exprimée, n'a pas d'effet négatif sur la survie des cellules exposées aux différents agents génotoxiques, sans pouvoir conclure quant à la fonctionnalité de la fusion *gfp::orc1* de par l'absence de phénotype du mutant Δorc1.

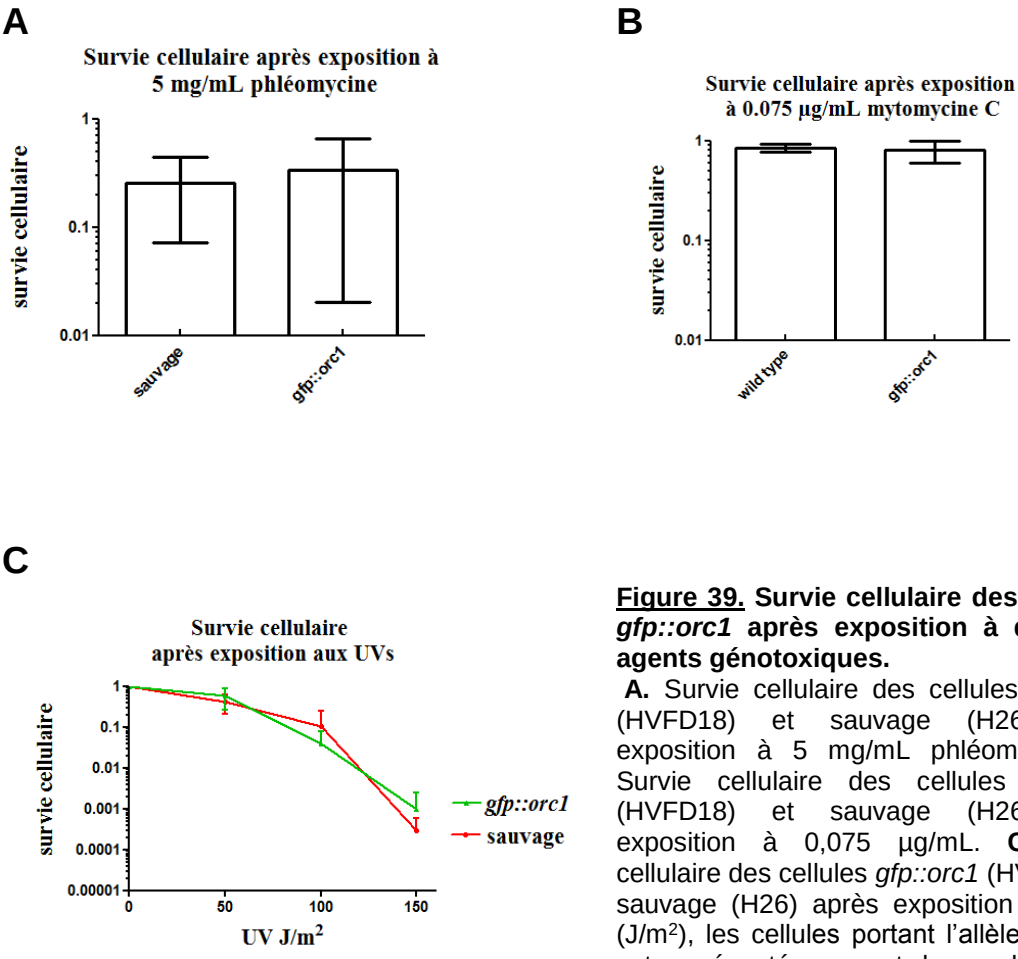


Figure 39. Survie cellulaire des souches *gfp::orc1* après exposition à différents agents génotoxiques.
A. Survie cellulaire des cellules *gfp::orc1* (HVFD18) et sauvage (H26) après exposition à 5 mg/mL phléomycine. **B.** Survie cellulaire des cellules *gfp::orc1* (HVFD18) et sauvage (H26) après exposition à 0,075 µg/mL. **C.** Survie cellulaire des cellules *gfp::orc1* (HVFD18) et sauvage (H26) après exposition aux UVs (J/m^2), les cellules portant l'allèle *gfp::orc1* est représentée en vert, la souche portant l'allèle sauvage *orc1* est représentée en rouge. Trois différentes doses d'UV ont été testées : 50, 100 et 150 J/m^2.

3.3.4 Localisation *in vivo* de ORC1

Orc1 intervient dans la reconnaissance des origines de réplication et participe à la formation du complexe pré-réplicatif. La localisation de cette protéine a donc été observée en phase exponentielle de croissance puis en phase stationnaire dans deux souches différentes : dans la souche sauvage (H26) possédant les quatre origines de réplication, et dans la souche ne possédant plus aucune origine de réplication (Δ*ori,* H1804).

Concernant les cellules *gfp::orc1*:

Le signal de fluorescence des cellules *gfp::orc1* a été observé a différents temps lors de la phase exponentielle de croissance, correspondant à des densités optiques mesurées de 0,1 ; 0,3 ; 0,6 et 1 (Figure 40).

Les paramètres de détection ont été ajustés pour la détection des cellules et des foyers de fluorescence afin d'analyser les images de façon quantitative *via* IMARIS (cf matériel et méthode). Les données concernant les cellules (nombre, surface, intensité…) et les foyers de fluorescence (nombre, répartition, intensité) ont été collectées.

Au cours de la phase de croissance, pour les cellules exprimant la protéine GFP::ORC1 des foyers de fluorescence ont pu être observés et analysés. Les analyses quantitatives correspondant aux images de fluorescence Figure 40 sont présentées Figure 41. Le nombre moyen de foyers de fluorescence diminue faiblement en début de phase de croissance entre DO=0,1 (2 foyers) et DO=0,6 (1,5 foyers) (p-value = 0,048), puis augmente en fin de phase de croissance où il est à nouveau de 2 (Figure 41 A). Cette variation étant très faible, nous pouvons considérer que le nombre de foyers de fluorescent est constant au cours de la phase de croissance. Afin d'approfondir cette observation et ne pas masquer une baisse significative du nombre de foyers de fluorescence par cellule par l'utilisation d'une moyenne, nous nous sommes intéressés à la répartition des cellules selon le nombre de foyer de fluorescence observé individuellement. Or le pourcentage de cellules à un foyer de fluorescence augmente de façon non significative entre DO=0,1 et D0=0,6 pour diminuer de nouveau à DO=1,6 (Figure 41 B). L'ensemble de ces données permet de conclure que le nombre de foyers de fluorescence ne varie pas significativement selon la phase de croissance. Il en est de même pour l'intensité de fluorescence qui reste constante tout au long de la phase de croissance, suggérant une expression constante de Orc1::GFP (Figure 41 C). On peut donc se demander quel rôle d'Orc1 reflètent ces foyers de localisation ?

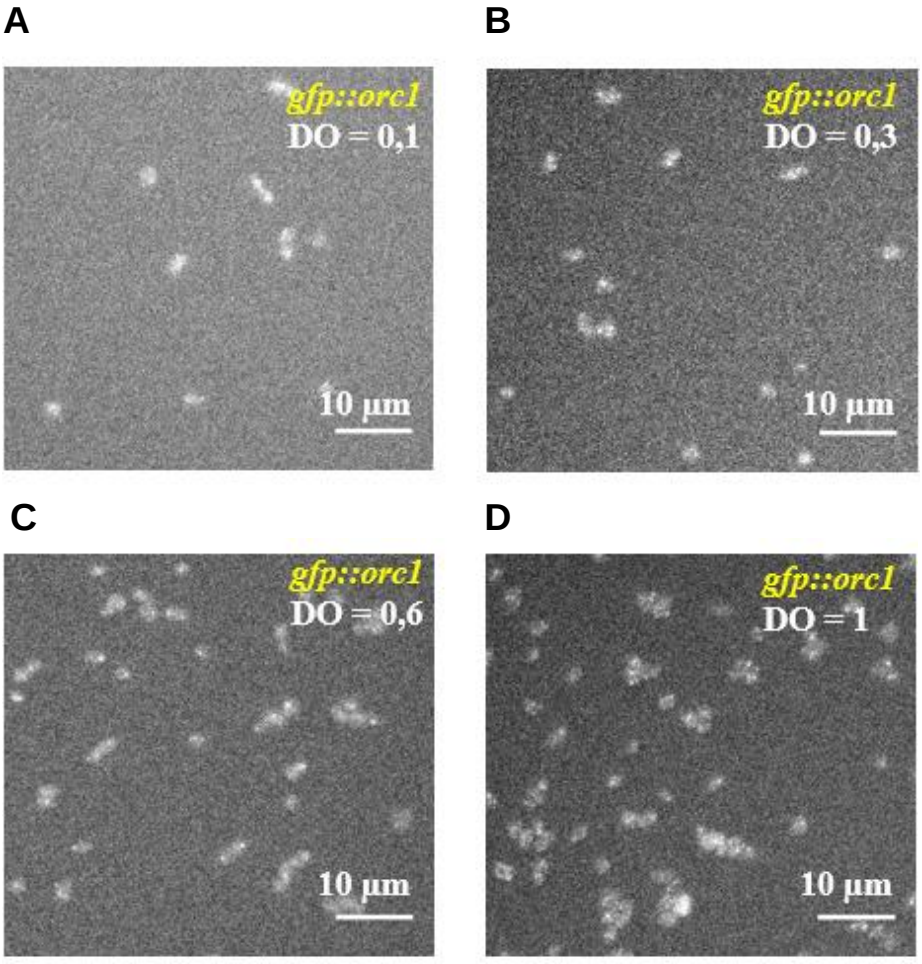


Figure 40. Images de microscopie de fluorescence des cellules exprimant l'allèle *gfp::orc1* selon la phase de croissance.
Les images sont le résultat de la projection sur une seule image du maximum d'intensité récolté pour chaque coupe au focus. L'échelle est de 10 µm. **A.** Souche *gfp::orc1* en début de phase exponentielle de croissance (DO=0.1). **B.** Souche *gfp::orc1* à DO=0.3. **C.** Souche *gfp::orc1* à DO=0.6. **D.** Souche *gfp::orc1* en fin de phase de croissance (DO=1). **E**. Souche *gfp::orc1* en phase stationnaire de croissance (DO=1,6).

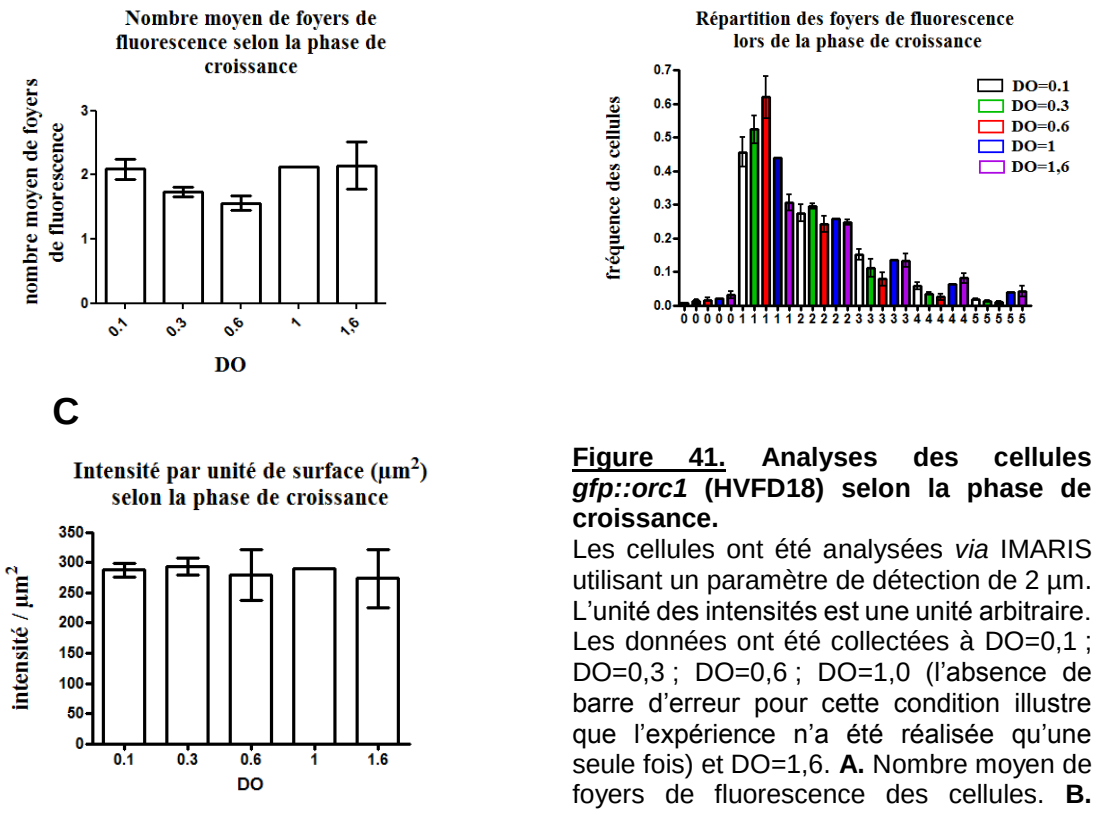


Figure 41. Analyses des cellules *gfp::orc1* (HVFD18) selon la phase de croissance.
Les cellules ont été analysées *via* IMARIS utilisant un paramètre de détection de 2 µm. L'unité des intensités est une unité arbitraire. Les données ont été collectées à DO=0,1 ; DO=0,3 ; DO=0,6 ; DO=1,0 (l'absence de barre d'erreur pour cette condition illustre que l'expérience n'a été réalisée qu'une seule fois) et DO=1,6. **A.** Nombre moyen de foyers de fluorescence des cellules. **B.** Répartition du nombre de foyers de fluorescence par cellule. **C.** Intensité par unité de surface ($\mu m^{2)}$ des cellules

Concernant les cellules *gfp::orc1* Δ*ori* :

Les mêmes expériences ont été réalisées pour les cellules *gfp::orc1* Δ*ori.* De façon surprenante, les foyers de fluorescence correspondant à la protéine GFP::Orc1 sont également observés dans une souche dépourvue des origines de réplication (Figure 42). Dans cette souche, le nombre moyen de foyers de fluorescence est plus important en début de phase de croissance (2,5 à DO=0,1) qu'à la fin (1,8 à DO=1) (Figure 43 A). En accord avec ces données, le pourcentage de cellules à un foyer de fluorescence augmente de façon significative entre DO=0,1 (35 %) et DO=0,6 (60%) pour ensuite diminuer à DO=1 (Figure 43 B). En revanche au cours de la phase de croissance, l'intensité par unité de surface diminue de façon non significative (pvalue=0,11) (Figure 43 C), suggérant une expression à un niveau constant de GFP::Orc1. En revanche, pour ces cellules il serait intéressant de quantifier la

fluorescence en fin de phase de croissance à DO=1,6 pour pouvoir conclure sur l'évolution de la fluorescence au cours de la phase de croissance.

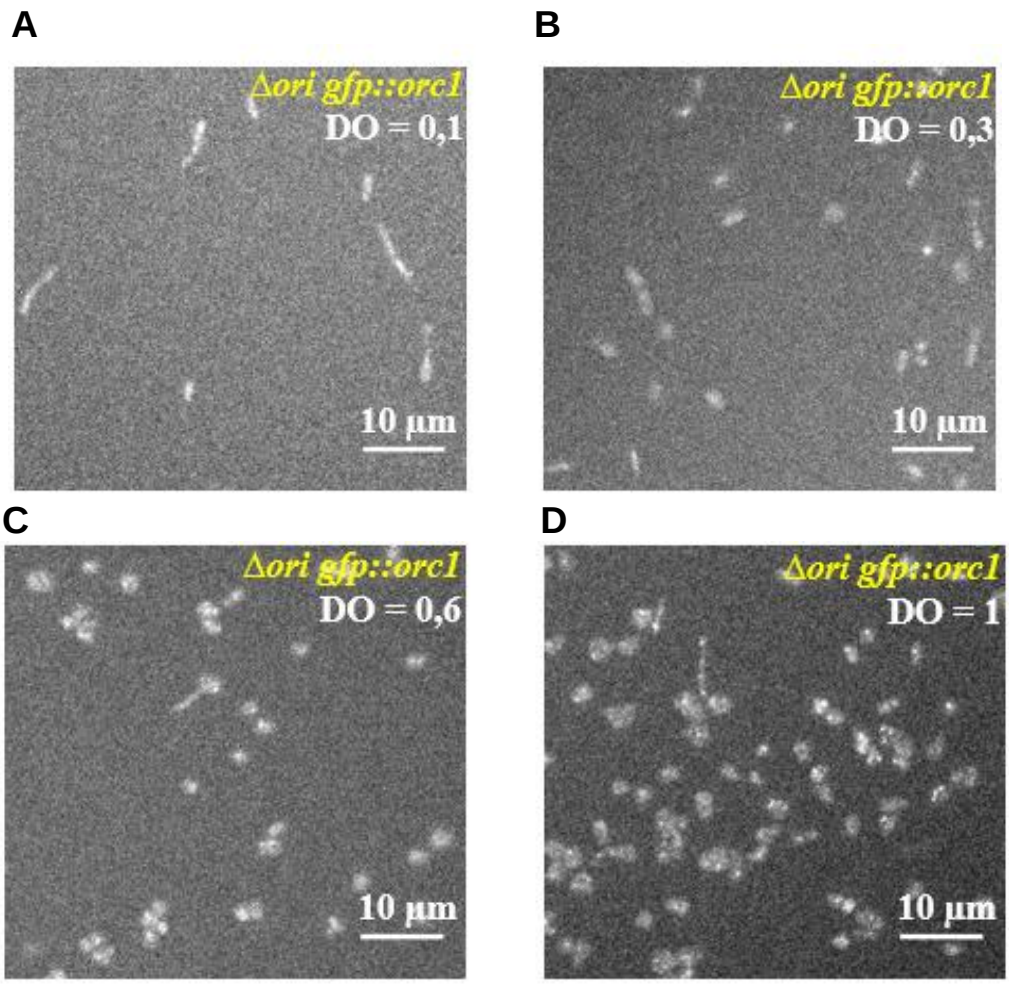


Figure 42. Images de microscopie de fluorescence des cellules exprimant l'allèle *gfp::orc1* dans une souche Δ*ori* selon la phase de croissance.
Les images sont le résultat de la projection sur une seule image du maximum d'intensité récolté pour chaque coupe au focus. L'échelle est de 10 µm. **A.** Souche *gfp::orc1*Δ*ori* en début de phase exponentielle de croissance (DO=0,1). **B.** Souche *gfp::orc1*Δ*ori* à DO=0,3. **C.** Souche *gfp::orc1*Δ*ori* à DO=0,6. **D.** Souche *gfp::orc1*Δ*ori* en fin de phase de croissance (DO=1).

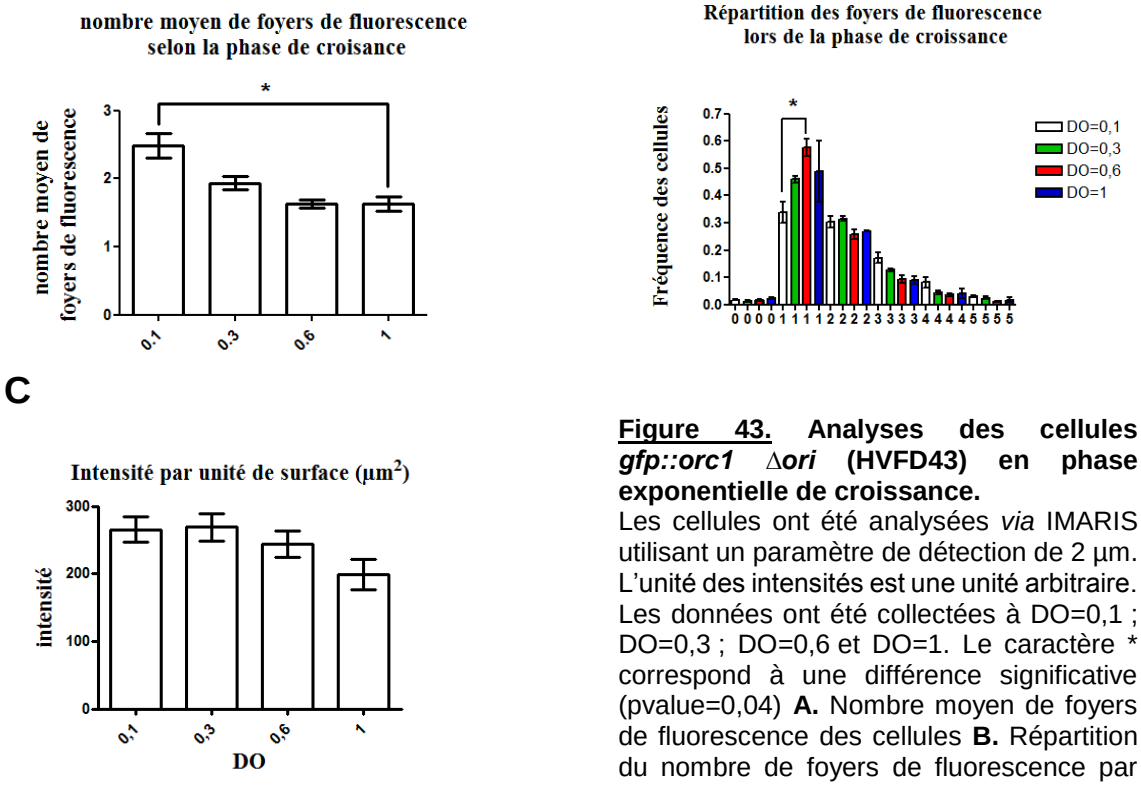


Figure 43. Analyses des cellules *gfp::orc1* Δ*ori* (HVFD43) en phase exponentielle de croissance.
Les cellules ont été analysées *via* IMARIS utilisant un paramètre de détection de 2 µm. L'unité des intensités est une unité arbitraire. Les données ont été collectées à DO=0,1 ; DO=0,3 ; DO=0,6 et DO=1. Le caractère * correspond à une différence significative (pvalue=0,04) **A.** Nombre moyen de foyers de fluorescence des cellules **B.** Répartition du nombre de foyers de fluorescence par cellule. **C.** Intensité par unité de surface (µm^2) des cellules.

Les cellules exprimant *gfp::orc1* avec et sans origine de réplication

La différence concernant la répartition et le nombre de foyers de fluorescence des cellules *gfp::orc1* pourrait être le reflet d'un recrutement de Orc1 à l'ADN plus important en début de phase exponentielle de croissance qui serait plus marqué dans la souche Δ*ori* dont le temps de génération est plus court. Néanmoins, cela implique que Orc1 peut être recrutée sur l'ADN lors de la réplication indépendamment de la présence d'origine de réplication. Ces foyers de fluorescence correspondent bien à Orc1 fixée à l'ADN puisque pour qu'il y ait formation d'un foyer de fluorescence il est nécessaire de fixer les protéines GFP::Orc1. Se pose donc la question du rôle de Orc1 reflété par ces foyers de localisation? Correspondent-ils à une fixation de Orc1 sur des séquences d'ADN aléatoires puisque la formation des foyers de fluorescence est indépendante de la présence d'origines de réplication, ou à une fixation de Orc1 à un complexe protéique comme avec MCM ?

3.4 Etude de la protéine Rpa2

Les protéines RPA sont des protéines clés de la réplication et de la réparation de l'ADN chez les trois domaines du vivant. La protéine Rpa (Replication Protein A) intervient dans la protection de l'ADN simple brin matrice sur lequel elle se fixe. Les protéines RPA se fixent à l'ADN par des domaines oligosaccharide-oligonucléotide (OB). Ces domaines sont généralement formés de cinq brins de feuillets β enroulés en un tonneau. *H. volcanii* possède trois gènes *rpa* : *rpa1*, *rpa2, rpa3*. Rpa1 possède deux domaines OB, Rpa3 n'en a qu'un et Rpa2 en possède trois. Par ailleusr les protéines Rpa1 et Rpa3 ont chacune un motif en doigt de zinc. Il est important de noter que le gène *rpa2* (nommé *rpaC* dans l'étude en question) est le seul des trois qui est essentiel à la viabilité, ce qui suggère un rôle clé dans la réplication [68]. Il a également été montré qu'il intervient dans la réparation de l'ADN chez *H. volcanii* : les cellules exprimant Rpa2 de façon constitutive ou la surexprimant présentent une résistance plus importante aux agents génotoxiques que les cellules sauvages [138].

Nous avons donc choisi d'étudier Rpa2. Nous souhaitions répondre aux questions suivantes : Quelle est la localisation de Rpa2 ? Reflète-t-elle la localisation des fourches de réplication dans la cellule ? Quelle est la dynamique de recrutement et de localisation lors des différentes phases de croissance cellulaire de Rpa2 ? Cette dynamique est-elle affectée lors des perturbations de la réplication ou après dommages de l'ADN ?

Pour répondre à ces questions, nous avons entrepris la construction d'une souche d'*H. volcanii* qui exprime la protéine Rpa2 fusionnée à la GFP. Afin de construire cette souche, nous avons choisi de réaliser la fusion à l'extrémité amino-terminale de la protéine.

3.4.1 Clonage du plasmide intégratif

Une première étape de construction à partir du plasmide intégratif pTA131 a été réalisée. Ce plasmide correspond au plasmide pFD2. Pour cela, la méthode de clonage PCR 2 temps a été employée (3.3.1). Les conditions de PCR ont été optimisées pour chaque fragment : «US-*rpa2*», «*gfp*» et «*rpa2'*» lors de la première étape de PCR, ainsi que pour la dernière étape permettant la synthèse du fragment complet (Tableau 15 et Figure 44).

Région à amplifier	Couple d'amorce	Tm°C	Longueur région amplifiée (pb)
«US-*rpa2*»	RL129-RL130	66°C	350 pb
«*gfp*»	RL85bis-RL124	58°C	780 pb
«*rpa2'*»	RL131-RL132	66°C	407 pb
«US-*gfp-rpa2'*»	RL129-RL132	66°C	1 457 pb

Tableau 15. Les conditions de PCR optimales pour chaque région constituant l'insert « *gfp-rpa2* » par PCR 2 temps

Afin de procéder à la PCR finale « US-*rpa2*/*gfp*/*rpa2*» au moyen des amorces RL129 et RL132, quatre quantités égales d'ADN de chaque fragment ont été testées : 12; 25; 50; 100 ng. La PCR effectuée avec la quantité d'ADN de 100 ng a été sélectionnée (Figure 44).

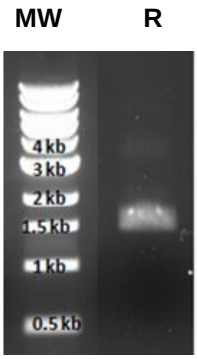


Figure 44. Produit de la PCR 2 temps correspondant à l'insert « *gfp-rpa2* ».
Piste MW : marqueur moléculaire de 1 kb.
Piste R : PCR 2 temps *gfp::rpa2*.

Le clonage suit le même principe que celui décrit dans la partie matériel et méthode (2.3.1). Le plasmide pTA131 a été linéarisé par l'enzyme *Eco*RV, dont le site de restriction se trouve dans le gène *lacZ*. Il a ensuite été déphosphorylé pour éviter toute recircularisation du plasmide. L'insert de 1,5 kb est purifié sur gel d'agarose puis ses extrémités sont traitées à la polynucléotide kinase avant d'être inséré dans le gène *lacZ,* au sein du plasmide pTA131.

Les cellules compétentes d'*E. coli* ont été transformées avec le produit de ce clonage et étalées sur boite LB ampicilline. Les boites ont été incubées à 37°C sur la nuit. A l'issu de cette étape, 480 colonies blanches et 56 colonies bleues ont poussé. Le contrôle des constructions s'est fait par PCR sur colonie et par digestion plasmidique. Seize des colonies blanches ont été testées pour la présence de l'insert d'intérêt par PCR au moyen des oligonucléotides RL28/RL29. La PCR conduit à l'amplification d'une séquence de 1802 pb si l'insertion est réussie contre l'amplification d'une séquence de 339 pb lorsqu'aucun insert n'est présent dans le plasmide. Quatre colonies parmi les seize colonies blanches testées possèdent l'insert migrant à environ 1,5 kb.

Afin de confirmer l'insertion du fragment « US-*gfp-rpa2* » dans le plasmide pTA131, l'ADN plasmidique contenu dans ces colonies a été extrait puis digéré par les enzymes *Bst*B1 et *Pvu*II-HF. Parmi les quatre plasmides testés, trois présentent le profil de digestion attendu. Après digestion par *Bst*B1, le plasmide issu du clone 1 présente un fragment de 4000 pb ainsi que deux fragments d'intensité plus faible de 486 et 617 pb. Lorsqu'il est digéré par *Pvu*II-HF l'ADN du plasmide pFD2 migre à environ 3000 pb et 2000 pb. L'ADN du clone 2 de pFD2 digéré par *Bst*B1 migre plus haut qu'aux longueurs attendues, ce plasmide n'a donc pas été sélectionné (Figure 45).

Avant de procéder à la transformation d'*H. volcanii,* la séquences du clone1 de pFD2 a été validée par séquençage.

	Fragments pour pTA131	Fragments pour pFD2
Digestion par *Pvu*II-HF	3178 pb 448 pb	3 178 pb 1911 pb
Digestion par *Bst*B1	Ne coupe pas 3626 pb	3986 pb 486 pb 617 pb

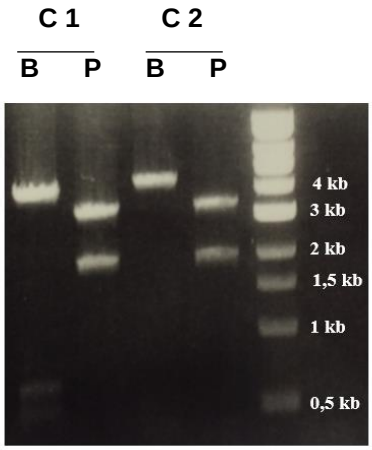


Figure 45. Vérification par restriction du plasmide pFD2.
Les pistes C1 et C2 correspondent aux deux plasmides extraits des clones sélectionnés digérés. Les pistes P correspondent à la migration des fragments de chaque plasmide digéré par *Pvu*II-HF, les pistes B correspondent à la migration des fragments obtenus après digestion des plasmides par *Bst*B1.

.

3.4.2 Construction des souches d'*H. volcanii* permettant l'expression de Rpa2 fusionnée à la GFP

Afin d'exprimer la protéine GFP::Rpa2 au locus chromosomique, il faut remplacer l'allèle sauvage du gène *rpa2* par l'insert cloné dans pFD2 par recombinaison homologue (technique pop-in/pop-out). La transformation de la souche sauvage (H26 : Δ*pyrE2*) par le plasmide pFD2 non méthylé a permis l'obtention de neuf colonies sur boite CASA (milieu sans uracile permettant de sélectionner l'intégration de pyrE2 au chromosome). L'excision du plasmide résultant d'un second événement de recombinaison (pop-out) est sélectionnée sur milieu CASA 5'FoA (+uracile) qui permet de contre-sélectionner la présence du marqueur pyrE2 au chromosome. Cela a été réalisé pour deux transformants sélectionnés après pop-in et striés sur milieu CASA. A partir des colonies sélectionnées sur CASA + 5'FoA, 100 ont été repiquées sur YPC, la sélection des colonies ayant intégré *gfp*::*rpa2* au locus chromosomique s'est faite par criblage des colonies par PCR au moyen du couple d'oligonucléotides RL85bis/RL124 qui amplifie la séquence de la *gfp*. Un produit de PCR d'environ 800 pb est attendu lorsque les cellules possèdent l'allèle *gfp*::*rpa2* tandis qu'aucun

produit de PCR n'est attendu lorsque les cellules possèdent l'allèle *rpa2* sauvage . Sur les 246 colonies issues du pop-out du transformant 1, 10 ont été testées pour l'amplification de la *gfp,* une seule colonie possède l'allèle *gfp::rpa2.* Sur les 83 colonies issues du pop-out du transformant 2, 10 ont été testées pour l'amplification de la *gfp,* deux colonies possèdent l'allèle *gfp::rpa2.* Plusieurs clones ont été testés par PCR au locus chromosomique, en dehors des régions d'homologie au moyen des oligonucléotides RL165/RL164. Cette PCR conduit à l'amplification d'une séquence de 2996 pb pour les cellules possédant l'allèle *gfp::rpa2,* contre 2246 pb lorsque les cellules possèdent l'allèle *rpa2* (Figure 46). Les cellules *d' H. volcanii* sont hautement polyploïdes, la sélection s'est faite suffisamment longtemps pour s'assurer qu'un maximum de copies du génome a l'allèle *gfp::rpa2.* Un clone a été mis en collection correspondant à la souche HVFD1.



Figure 46. Amplification par PCR de la région *orc1* après pop-in/pop-out pour intégrer l'allèle *gfp::rpa2.*
Le clone *rpa2* correspond aux cellules 'sauvages' et les clones *gfp::rpa2* (en vert) correspondent aux cellules exprimant l'allèle *gfp::rpa2.*

3.4.3 Etude de la fonctionnalité de la protéine fusion GFP::Rpa2

Le temps de génération de la souche exprimant Rpa2 fusionnée à la GFP est similaire (2,1h) à celui de la souche exprimant Rpa2 sous sa forme native (2,1 h). Sachant que le gène *rpa2* est essentiel, ces résultats nous permettent de conclure que la protéine Rpa2 fusionnée à la GFP permet la survie cellulaire ce qui indique que cette protéine fusion est fonctionnelle.

Souche	Temps de génération
Sauvage, Δ*pyrE2* (H26)	2,1 h ± 0,2
ΔpyrE2, *gfp::rpa2* (HVFD1) *	2,1 h ± 0,2

Tableau 16. Viabilité et temps de génération des souches possédant l'allèle *gfp::rpa2,* l'allèle sauvage *rpa2* et délétées pour *rpa2 :* Δ*rpa2.*
* : la souche Δ*pyrE2,* Δ*rpa2* ne peut pas être obtenue.

La résistance à différents agents génotoxiques provoquant des lésions de l'ADN a ensuite été testée (Figure 47).

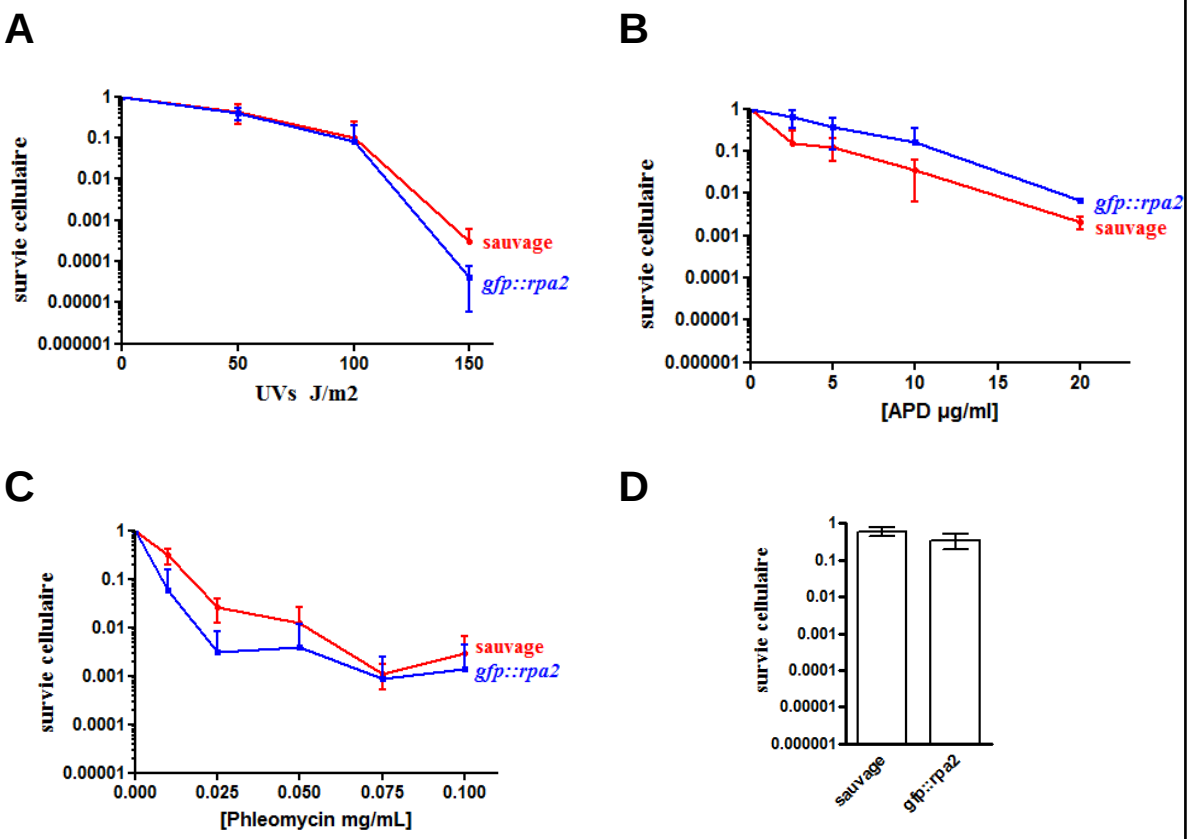


Figure 47. Survie cellulaire après exposition des cellules à différents agents génotoxiques. **A.** Survie cellulaire des cellules *gfp::rpa2* (en bleu) et sauvage (en rouge) après exposition à trois doses d'UVs en J/m^2. **B.** Survie cellulaire des cellules *gfp::rpa2* (en bleu) et sauvage (en rouge) après exposition à 2,5 ; 5 ;10 et 20 µg/mL d'APD. **C.** Survie cellulaire des cellules *gfp::rpa2* (en bleu) et sauvage (en rouge) après exposition à une concentration croissante de phléomycine en mg/mL. **D.** Survie cellulaire des cellules après exposition des cellules à une concentration en MMC 0,075 µg/mL.

Fonctionnalité après exposition aux agents génotoxiques

Il a été montré que la délétion du domaine amino-terminale de la protéine Rpa2 entraine une augmentation de la sensibilité des cellules aux UVs [68]. Or, l'exposition aux UVs de la souche exprimant l'allèle *gfp::rpa2* (HVFD1) n'entraine pas de défaut de survie comparé à la souche sauvage, et ce même après une exposition à une dose élevée (150 J/m^2) (Figure 47 A).

L'exposition des cellules à la mitomycine C entraine des pontages covalents inter brins qui nécessitent également l'action du complexe NER ou HR. Les survies des cellules exprimant l'allèle *gfp::rpa2* (HVFD1) et des cellules exprimant l'allèle sauvage *rpa2* (H26) après exposition à 0,02 µg/mL de mitomycine C sont semblables (Figure 47 D).

L'exposition des cellules à la phléomycine provoque des dommages à l'ADN en induisant des cassures double brin. Il a été montré que la délétion du domaine amino-terminale

de la protéine Rpa2 entraine une augmentation de la sensibilité des cellules à la phléomycine [68]. Les survies cellulaires des cellules exprimant l'allèle *gfp::rpa2* (HVFD1) et des cellules exprimant l'allèle sauvage *rpa2* (H26) après exposition à différentes concentrations de phléomycine (0,01 à 0,01 mg/mL) sont semblables (Figure 47 C).

Fonctionnalité après exposition à un stress réplicatif

De même, l'exposition des cellules à l'aphidicoline qui provoque des arrêts de la réplication en inhibant l'ADN polymérase n'entraine pas de défaut de survie de la souche exprimant l'allèle *gfp::rpa2* (HVFD1) comparé à la souche sauvage (Figure 47 B).

L'ensemble de ces résultats permet de conclure que la protéine GFP::Rpa2 permet la survie cellulaire même après exposition à des agents génotoxiques ou en réponse à un stress réplicatif (exposition à l'APD), et semble donc totalement fonctionnelle *in vivo.*

3.4.4 Localisation *in vivo* de GFP::RPA2 selon la phase de croissance

La localisation *in vivo* de la protéine GFP::Rpa2 a alors été observée selon la phase de croissance, et après exposition à différents agents génotoxiques induisant des arrêts de la réplication ou provoquant des cassures double brin.

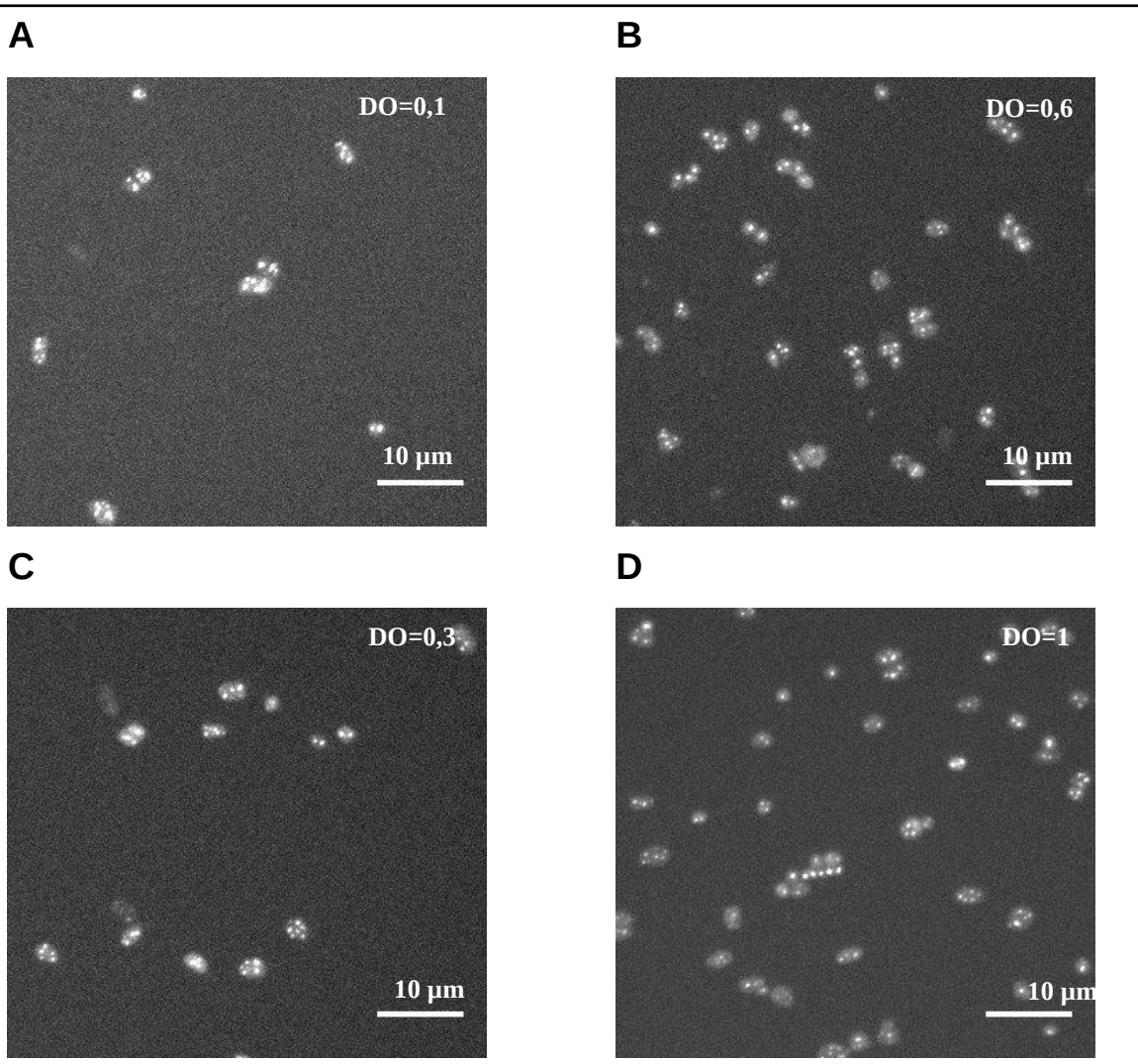


Figure 48. Images de fluorescence des cellules exprimant *gfp::rpa2* selon la phase de croissance.
Les images sont le résultat de la projection sur une seule image du maximum d'intensité récolté pour chaque coupe au focus. L'échelle est de 10 µm. **A**. Image à DO=0.1. **B**. Image à DO=0.3. **C**. Image à DO=0.6. D. Image à DO=1.

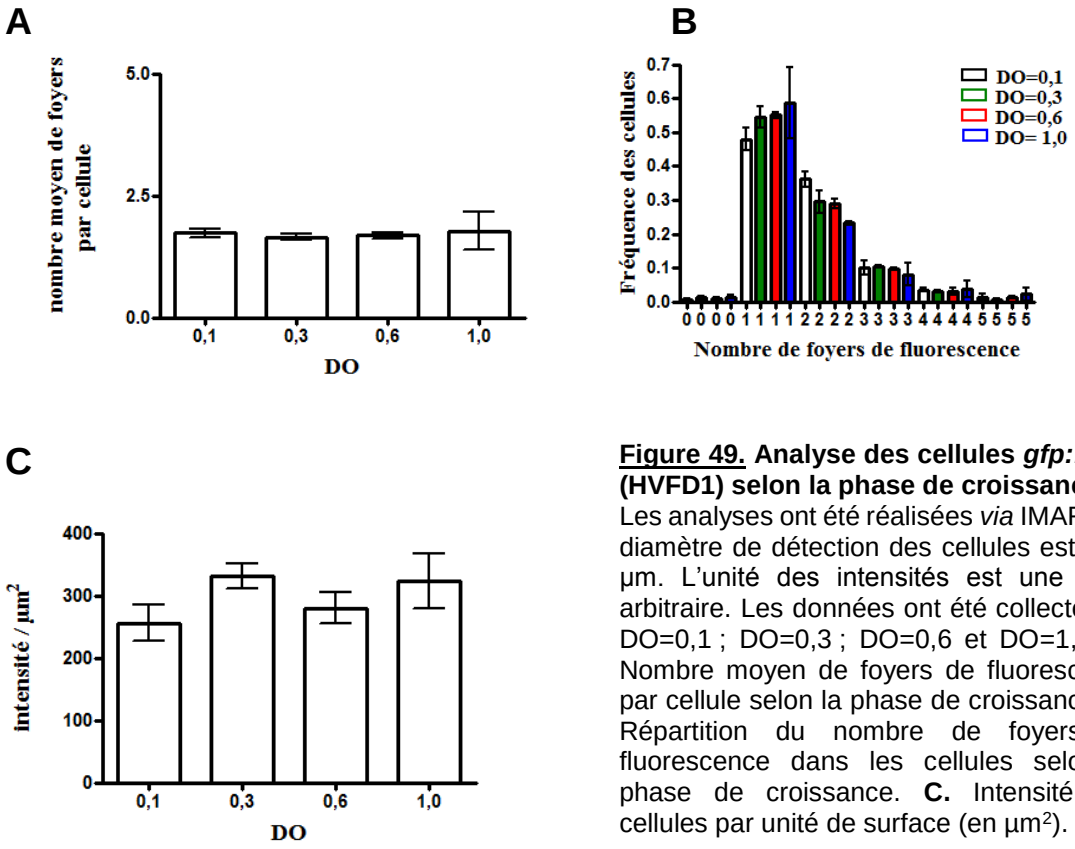


Figure 49. Analyse des cellules *gfp::rpa2* (HVFD1) selon la phase de croissance. Les analyses ont été réalisées *via* IMARIS le diamètre de détection des cellules est de 2 µm. L'unité des intensités est une unité arbitraire. Les données ont été collectées à DO=0,1 ; DO=0,3 ; DO=0,6 et DO=1,0. **A.** Nombre moyen de foyers de fluorescence par cellule selon la phase de croissance. **B.** Répartition du nombre de foyers de fluorescence dans les cellules selon la phase de croissance. **C.** Intensité des cellules par unité de surface (en µm^2).

Le signal de fluorescence des cellules *gfp*::*rpa2* a été observé à différents temps lors de la phase de croissance, correspondant à des densités optiques mesurées de 0,1 ; 0,3 ; 0,6 et 1,0 (Figure 48 A, B, C et D). Des foyers de fluorescence sont observés dans la plupart les cellules exprimant Rpa2 fusionnée à la GFP (HVFD1). Le signal de fluorescence observé dans les cellules *gfp::rpa2* ne semble pas varier au cours de la phase de croissance.

Comme expliqué précédemment (partie résultat de FEN1), les images analysées sont le résultat de la projection en z sur une seule image du maximum d'intensité récolté pour chaque coupe au focus. Ces expériences ont été répétées trois fois et analysées quantitativement par le logiciel IMARIS. Les paramètres de détection ont été optimisés : le diamètre de détection des cellules est de 2 µm, le diamètre de détection des foyers de fluorescence est de 0,75µm. Différentes données concernant à la fois la cellule et les foyers de fluorescence ont été collectées : le nombre moyen de foyers de fluorescence par cellule (Figure 49 A), la répartition du nombre de foyers de fluorescence dans les cellules (Figure 49 B) autrement dit la fréquence des cellules qui ont de 0 à 5 foyers de fluorescence par cellule et l'intensité de fluorescence des cellules par unité de surface (µm^2) (Figure 49 C). Les analyses quantitatives réalisées *via* IMARIS relatives aux images de microscopie de fluorescence des cellules *gfp::rpa2* présentées Figure 48 sont données Figure 49.

Les protéines GFP::Rpa2 forment des foyers de fluorescence individualisés. Or, le nombre moyen de foyers de fluorescence observé dans les cellules de DO=0,1 à DO=1,0 est constant (≈2 foyers de fluorescence par cellule) (Figure 49 A). Afin d'approfondir cette observation et ne pas masquer une baisse significative du nombre de foyers de fluorescence par cellule par l'utilisation d'une moyenne, nous nous sommes intéressés à la répartition des cellules selon le nombre de foyers de fluorescence observé individuellement. Or, le pourcentage des cellules ayant un foyer de fluorescence par cellule augmente de façon non significative de DO=0,1 à DO=1,0 : 48 % des cellules ont un foyer de fluorescence par cellule à DO=0,1 contre 58% à DO=1,0 (Figure 49 B). Par ailleurs, l'intensité totale des cellules rapportée à la surface cellulaire totale a été déterminée et ne varie pas de façon significative.

Nos résultats montrent que le nombre de foyers de fluorescence formés par GFP::Rpa2 ainsi que l'intensité du signal de fluorescence sont constants et ne semblent pas dépendre de la phase de croissance.

3.4.5 Localisation *in vivo* de RPA2 après exposition à différents agents génotoxiques

N'ayant pas observé de différence de localisation de la protéine GFP::Rpa2 dans des cellules en prolifération, et sachant que Rpa2 intervient également dans la réparation de l'ADN, nous avons étudié la localisation de GFP::Rpa2 lorsque la cellule est soumise à des agents génotoxiques provoquant des dommages de l'ADN, telle que la mitomycine C, ou à des arrêts de la réplication provoqués par l'exposition des cellules à l'aphidicoline (APD).

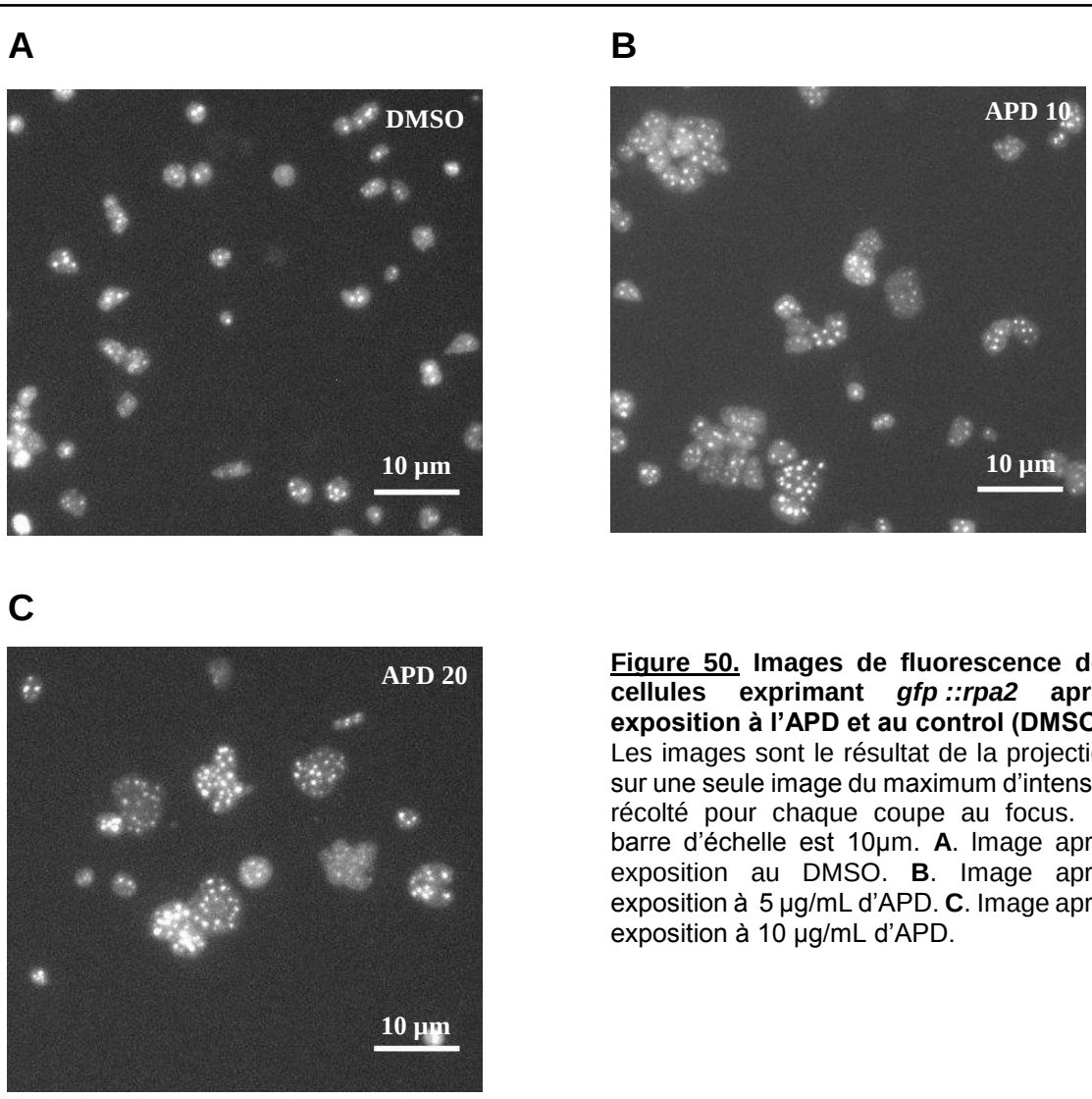


Figure 50. Images de fluorescence des cellules exprimant *gfp ::rpa2* après exposition à l'APD et au control (DMSO). Les images sont le résultat de la projection sur une seule image du maximum d'intensité récolté pour chaque coupe au focus. La barre d'échelle est 10µm. **A**. Image après exposition au DMSO. **B**. Image après exposition à 5 µg/mL d'APD. **C**. Image après exposition à 10 µg/mL d'APD.

Le stress réplicatif

Le signal de fluorescence des cellules *gfp*::*rpa2* a été observé après 17h00 d'exposition à différentes concentrations d'APD : 1 ;2.5 ;5 ;10 et 20 µg/mL et après exposition au DMSO (solvant de l'APD) qui correspond à la condition contrôle. Seules les images de fluorescence

correspondant aux concentrations d'APD 5 et 10 µg/mL et à la condition contrôle sont représentées Figure 50.

Le même protocole d'analyse d'image a été employé que celui expliqué dans la partie localisation en phase exponentielle de croissance (3.4.4). Les analyses quantitatives réalisées *via* IMARIS relatives aux images de microscopie de fluorescence des cellules *gfp::rpa2* illustrées Figure 50 sont présentées Figure 51.

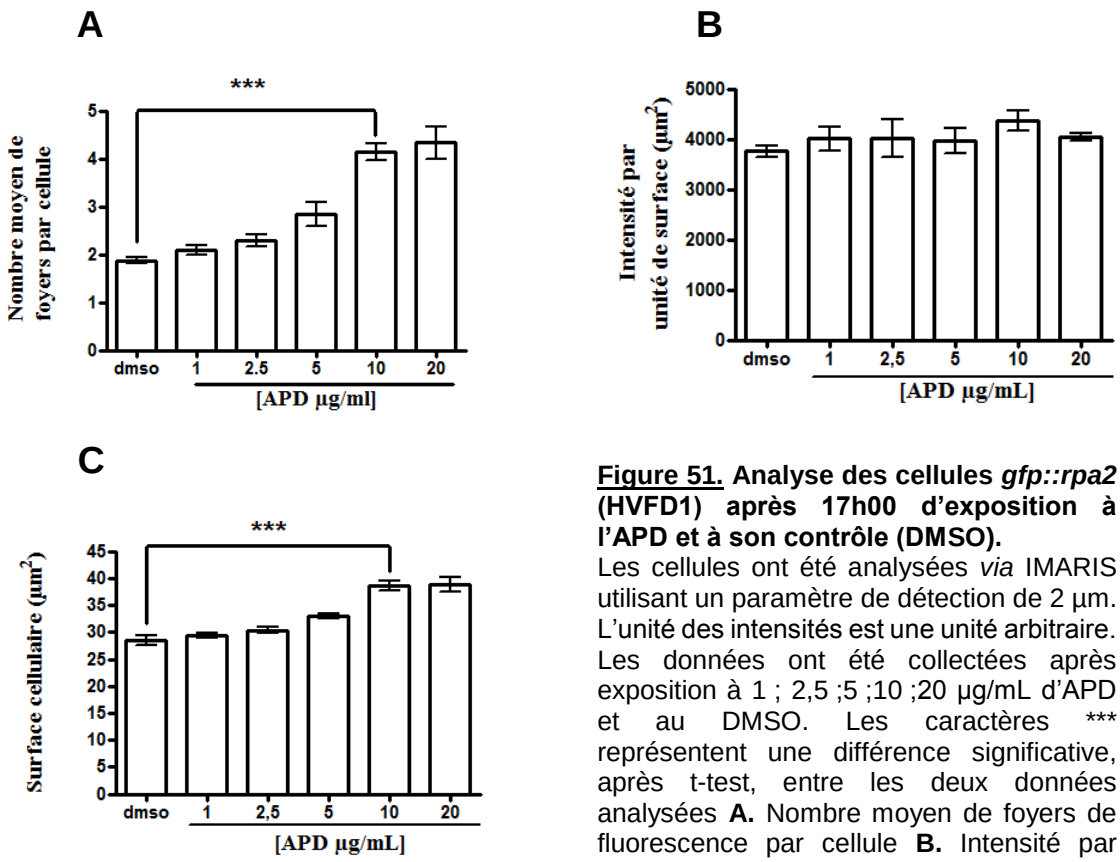


Figure 51. Analyse des cellules *gfp::rpa2* (HVFD1) après 17h00 d'exposition à l'APD et à son contrôle (DMSO).
Les cellules ont été analysées *via* IMARIS utilisant un paramètre de détection de 2 µm. L'unité des intensités est une unité arbitraire. Les données ont été collectées après exposition à 1 ; 2,5 ;5 ;10 ;20 µg/mL d'APD et au DMSO. Les caractères *** représentent une différence significative, après t-test, entre les deux données analysées **A.** Nombre moyen de foyers de fluorescence par cellule **B.** Intensité par unité de surface (µm²). **C.** Surface cellulaire moyenne (µm²).

Dans un premier temps, le nombre moyen de foyers de fluorescence observé dans les cellules a été déterminé. On observe un effet dose dépendant. Il augmente significativement entre la condition contrôle (DMSO), pour laquelle les cellules ont en moyenne 2 foyers par cellule, et la condition APD 10 µg/mL pour laquelle les cellules ont en moyenne 4,5 foyers par cellule (Figure 51 A). Afin de comprendre si cette augmentation de foyers de fluorescence est le reflet d'une augmentation de l'expression de GFP::Rpa2, nous avons déterminé l'intensité de fluorescence des cellules par unité de surface. Il est important de noter que la surface cellulaire augmente significativement lorsque les cellules sont exposées à l'APD, elle est de 28 µm² lorsque les cellules sont exposées au DMSO contre 40 µm² lorsque les cellules sont

exposées à 10 µg/mL d'APD (Figure 51 C). Il existe un couplage entre la réplication de l'ADN et la division cellulaire, celle-ci serait empêchée lorsque la réplication n'a pas pu se faire conduisant à une augmentation de la taille cellulaire comme c'est le cas ici. Ainsi, en rapportant les intensités de fluorescence totale des cellules à l'unité de surface, il est possible de comparer les intensités entre les différentes conditions. L'intensité totale des cellules rapportée à l'unité de surface est constante pour les cellules exposées aux différentes concentrations d'ADP et pour les cellules exposées au DMSO (Figure 51 B). Ainsi, l'exposition à l'APD conduit à une augmentation du nombre de foyers de fluorescence formés par les protéines GFP::Rpa2, ainsi qu'une augmentation de la surface cellulaire. L'intensité de fluorescence rapportée à l'unité de surface est identique pour les cellules exposées ou non à l'APD, témoignant d'un niveau d'expression de Rpa2 constant. Est-ce que ces foyers correspondent au recrutement de Rpa2 aux fourches arrêtées qui présentent des régions simple brins plus importantes et/ou sont-ils le reflet d'une situation figée facilitant leur détection ? Après exposition à l'APD, les fourches sont arrêtées et les molécules GFP::Rpa2 qui forment des foyers de fluorescence sont localisées au même endroit ce qui facilite leur détection. En effet la méthode d'acquisition utilisée (500 ms d'exposition par champ) favorise la détection des foyers de fluorescence stables. De plus, les arrêts des fourches de réplication peuvent provoquer un découplage de la synthèse d'ADN, ce qui engendrerait une augmentation de l'ADN simple brin. Pour répondre à cette hypothèse une quantification de l'ADN simple brin compris dans l'ensemble des cellules exposées à l'APD et à la condition contrôle serait intéressante.

La réparation de l'ADN

Afin de déterminer le rôle *in vivo* de Rpa2 dans la réparation de l'ADN, nous avons localisé GFP::Rpa2 dans les cellules après exposition à la mitomycine C (MMC).

Lorsque les cellules *gfp::rpa2* (HVFD1) sont exposées à 0,075 µg/mL de MMC, les protéines GFP::Rpa2 forment des foyers de fluorescence individualisés. Nous avons comparé la localisation des cellules contrôles (non exposées aux agents génotoxiques) aux cellules exposées à différentes concentrations de MMC. Le signal de fluorescence est similaire dans les deux conditions testées (Figure 52 A et B). Le nombre moyen de foyers de fluorescence, est également similaire pour les deux conditions testées (Figure 52 C). Il n'y a donc pas de localisation spécifique de la protéine GFP::Rpa2 après exposition des cellules à la MMC. D'autres protéines GFP::Rpa2 peuvent se trouver dans le cytoplasme sans former de foyers de fluorescence. L'intensité de fluorescence par unité de surface permet d'avoir une idée de la fluorescence totale au sein de la cellule comprenant l'intensité des foyers de fluorescence et du cytoplasme. Celle-ci est d'environ 4500 pour les cellules traitées ou non à la mitomycine

C. L'ensemble de ces résultats permet de conclure que l'exposition des cellules à la MMC n'a pas d'effet sur la localisation et le niveau d'expression de GFP::Rpa2.

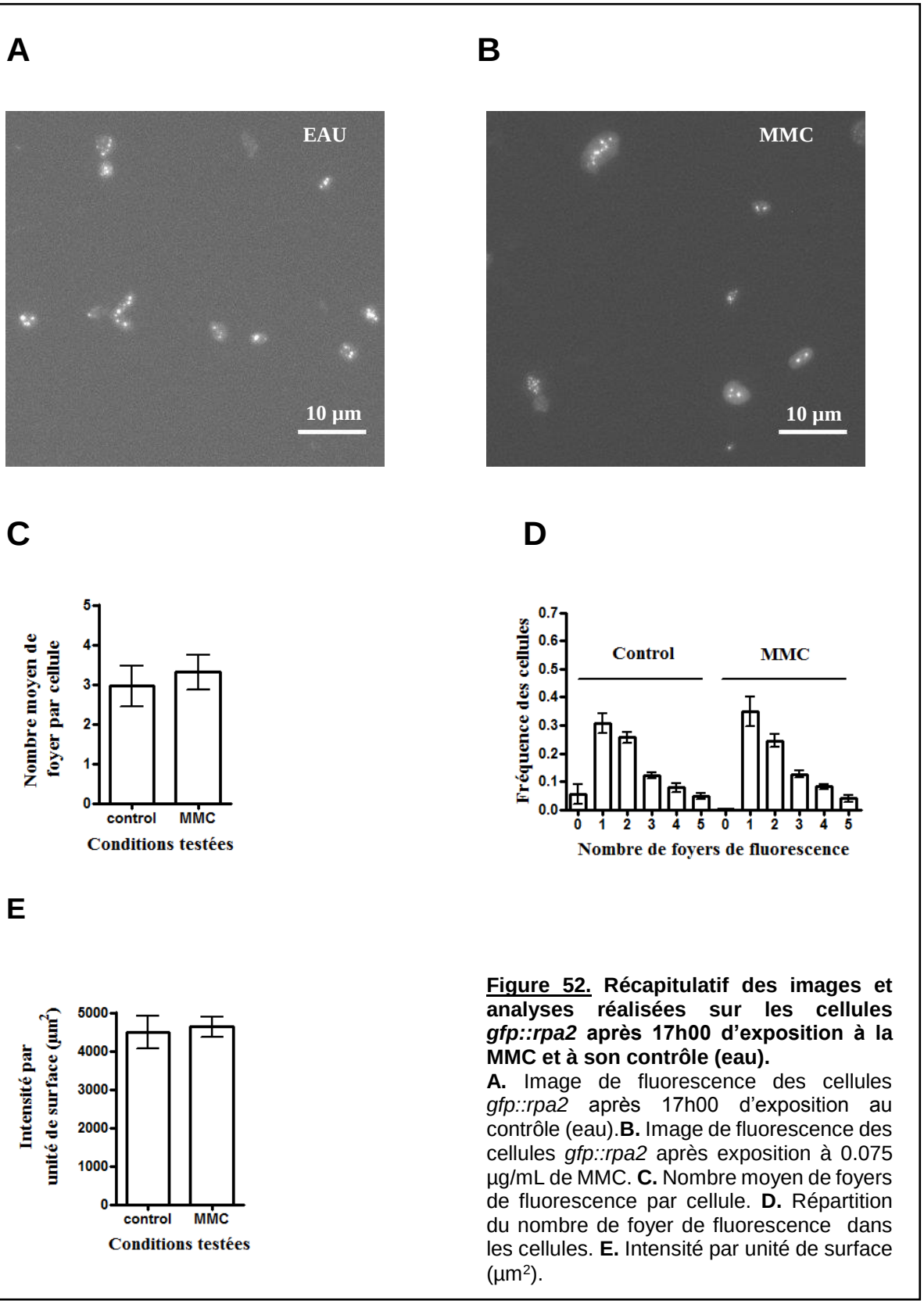
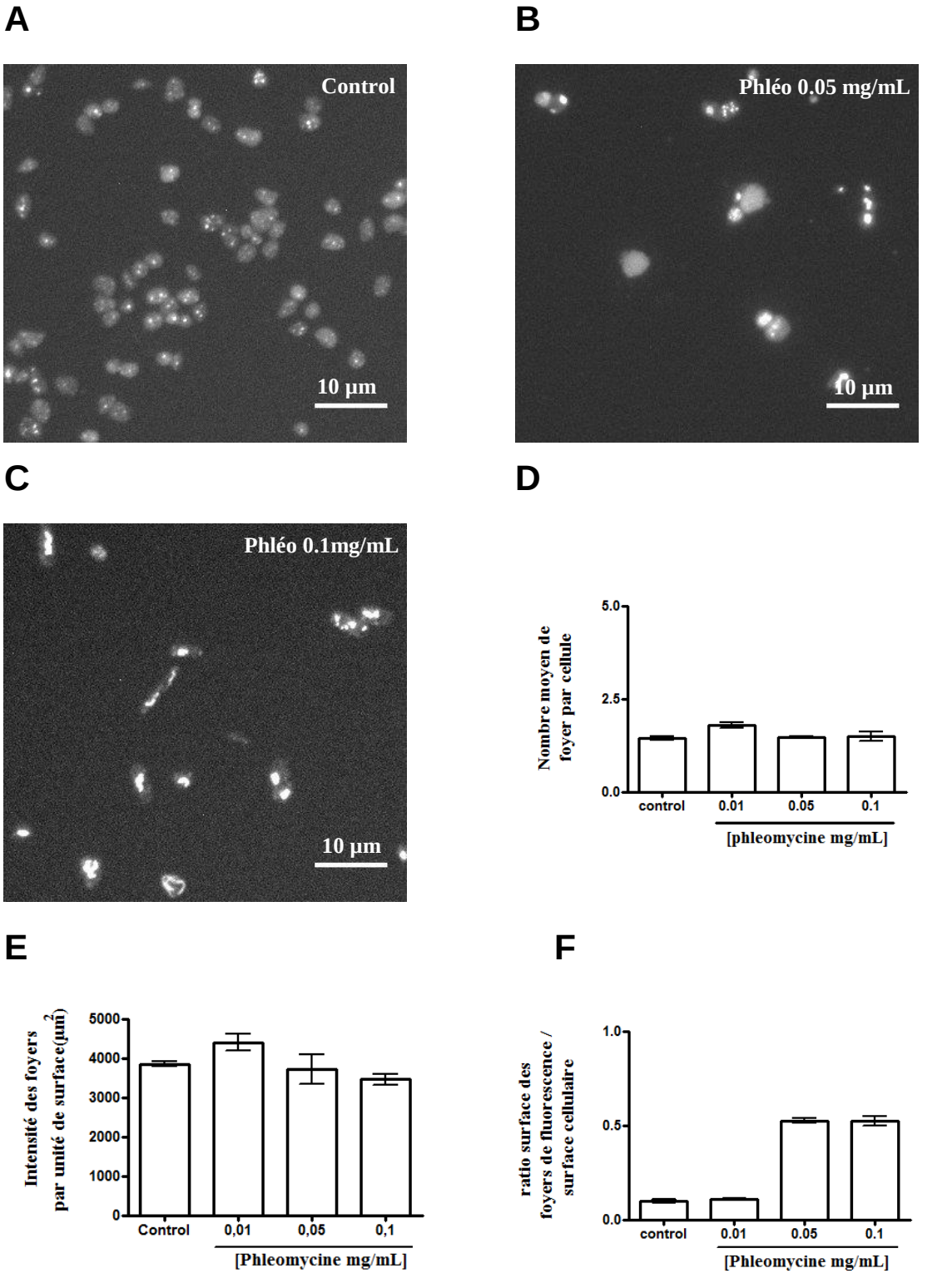


Figure 52. Récapitulatif des images et analyses réalisées sur les cellules *gfp::rpa2* après 17h00 d'exposition à la MMC et à son contrôle (eau).
A. Image de fluorescence des cellules *gfp::rpa2* après 17h00 d'exposition au contrôle (eau).**B.** Image de fluorescence des cellules *gfp::rpa2* après exposition à 0.075 µg/mL de MMC. **C.** Nombre moyen de foyers de fluorescence par cellule. **D.** Répartition du nombre de foyer de fluorescence dans les cellules. **E.** Intensité par unité de surface (μm^2).

La localisation de la protéine GFP::Rpa2 après exposition à la phléomycine a également été étudiée. Le signal de fluorescence des cellules *gfp*::*rpa2* a été observé à différentes concentrations de phléomycine : 0.01, 0,05 et 0,1 mg/mL puis comparé aux cellules contrôles sans agent génotoxiques (Figure 53 A, B et C). Des foyers de fluorescence sont observés dans la plupart des cellules. Après traitement à la phléomycine (Figure 53 B et C), les foyers de fluorescence semblent plus gros et compactés en comparaison des cellules exposées à l'eau (control, Figure 53 A) ou à l'APD, ces foyers ont été baptisés « mega spots » et leur paramètre de détection a été optimisé. Le nombre de foyers de fluorescence ne varie pas entre les cellules traitées à la phléomycine et les cellules non traitées, il est d'environ 2 foyers de fluorescence par cellule (Figure 53 D). En revanche, la taille du foyer est différente, ceci est reflété par le volume occupé par le foyer qui augmente significativement (Figure 53 F). Nous avons déterminé l'intensité de fluorescence comprise dans ces foyers rapportée à la surface qu'ils occupent, celle-ci reste constante (Figure 53 E), ainsi la formation des méga foyers n'est pas corrélée à une augmentation de l'expression de Rpa2.

L'exposition à la phléomycine provoque des cassures double brin qui sont réparées par recombinaison homologue, l'exposition à la mitomycine C provoque également des cassures double brin qui peuvent être réparées par recombinaison homologue mais également par la voie d'excision des nucléotides. L'absence de localisation spécifique après exposition à la mitomycine C suggère un rôle de Rpa2 dans la réparation de l'ADN par recombinaison homologue .



<u>Figure 53.</u> Analyses de l'imagerie réalisée sur les cellules *gfp ::rpa2* après 17h00 d'exposition différentes concentrations de phléomycine.
Trois concentrations ont été testées : 0.01 mg/Ml, 0.05 mg/Ml et 0.1 mg/Ml. L'échelle est de 10 µm. **A.** Maximum d'intensité des cellules contrôles (eau ajoutée).**B.** Maximum d'intensité des cellules traitées à 0.05 mg/Ml de phléomycine. **C.** Maximum d'intensité des cellules *gfp ::rpa2* traitées à 0.1 mg/Ml de phléomycine. **D.** Nombre moyen de foyers de fluorescence par cellule. **E.** Intensité des foyers de fluorescence par unité de surface (μm^2). **F.** Ratio entre la surface des foyers de fluorescence et la surface cellulaire.

3.4.6 Quantification de GFP::Rpa2 : western blot anti-GFP

L'intensité par unité de surface permet de comparer les intensités totales en s'affranchissant du biais apporté par la différence des surfaces cellulaires. Ces informations ont été complétées par l'analyse de l'expression des protéines par western-blot en utilisant des anticorps anti-GFP. Les extraits protéiques ont été préparés à quantité de cellules équivalentes pour tous les échantillons.

Les cellules ont été exposées aux drogues selon le même protocole que celui utilisé pour la microscopie de fluorescence. Les extraits protéiques des cellules en phase exponentielle de croissance ont été préparés lorsque les cellules ont atteint une DO=0,4, les extraits protéiques des cellules contrôle ou exposées à la phléomycine ou à l'APD ont été préparés après environ 17h00 d'exposition à ces drogues. L'intensité par unité de surface des cellules *gfp::rpa2* après exposition à différentes concentrations d'APD et de phléomycine ne varie pas significativement (Figure 54).

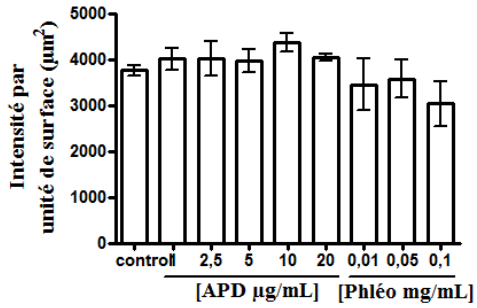


Figure 54. Intensité par unité de surface des cellules *gfp::rpa2* après exposition à différentes concentrations de phléomycine et d'APD.

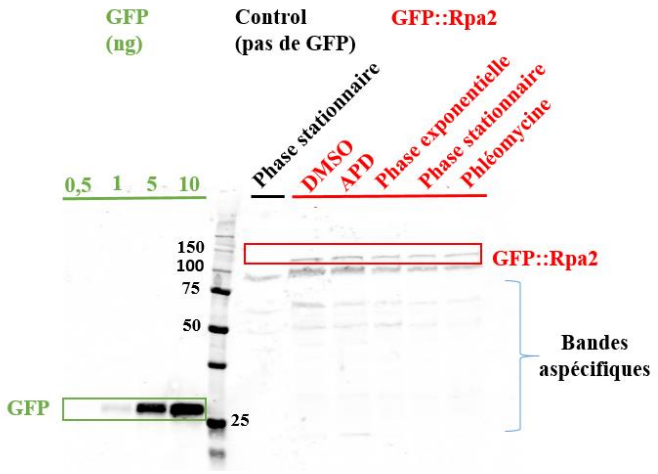


Figure 55. Western blot anti-GFP réalisé sur les extraits protéiques des cellules *gfp::rpa2* (HVFD1) et des cellules sauvages (H26).
La gamme de GFP est en vert (ng). La même quantité d'extrait protéique a été déposée pour toutes les conditions. Les extraits protéiques des cellules sauvage en phase stationnaire ont été déposés (en noir). Les extraits protéiques des cellules *gfp::rpa2* en phase exponentielle et stationnaire de croissance ont été déposés, ainsi que ceux obtenus après exposition des cellules à 0,1 mg/mL de phléomycine et 5 µg/mL d'APD et au DMSO (control) (en rouge).La migration de la protéine GFP::Rpa2 est entourée en rouge.

La détection de la protéine GFP::Rpa2 par western-blot a été réalisée trois fois à partir de trois lots différents d'extraits protéiques. Les anticorps anti-GFP ont été utilisés pour la détection de la protéine GFP::Rpa2 dans tous les extraits protéiques. Une gamme de GFP contenant des quantités de GFP de 0,5 à 10 ng a été réalisée, permettant de quantifier par la suite les protéines GFP::Rpa2 présentes dans les différents échantillons. La GFP migre à 28 KDa et la quantité minimale détectable est de 1 ng. La détection de la GFP fusionnée à Rpa2 dans les extraits protéiques permet d'individualiser une bande migrant à environ 120 KDa qui n'est pas présente dans les extraits protéiques issus des cellules n'exprimant pas la GFP. La protéine GFP::Rpa2 migre plus haut que son poids théorique attendu (80 KDa), c'est le cas de beaucoup de protéines halophiles qui se chargeant mal en SDS et migrent souvent plus haut que leur poids théorique (Figure 55).

Les extraits protéiques des échantillons chargés sont issues de la même quantité de cellules, ainsi les quantités de la protéine GFP::Rpa2 détectée peuvent être comparées entre elles. Les intensités des bandes ont été déterminées *via* ImageJ. La quantité approximative en ng des protéines GFP::Rpa2 déposées pour chaque échantillon a été calculée *via* la formule suivante :

$$(intensité\ de\ la\ bande\ spécifique\ de\ GFP :: Rpa2)\ /\ (intensité\ de\ la\ bande\ GFP\ 1\ ng)$$

La moyenne des quantités en ng de GFP::Rpa2 détectées sont :

- 1,8 ng pour la phléomycine
- 2,3 ng pour l'aphidicoline
- 2,0 pour la condition contrôle en phase stationnaire de croissance
- 2,5 pour la phase exponentielle de croissance

Ces quantités ne varient pas significativement. L'ensemble de ces résultats permet de conclure qu'il n'y a pas de surexpression des protéines GFP::Rpa2 même si la localisation de la protéine GFP::Rpa2 est différente selon la phase de croissance et après exposition des cellules à l'APD ou à la phléomycine.

3.4.7 Colocalisation de l'ADN avec les foyers GFP::Rpa2

Des études précédentes ont montré qu'après exposition à la phléomycine, qui provoque des cassures doubles brins ou simple brin, l'ADN est compacté [120]. Or, nous avons montré ici que l'exposition des cellules exprimant la protéine GFP::Rpa2 à la phléomycine conduit à la formation de 'méga' foyers. Cette formation de 'mega' foyer est semblable à celle observée lorsque l'ADN est sous forme compactée. Ainsi, nous souhaitions colocaliser l'ADN des cellules présentant la localisation 'mega' foyer avec les protéines GFP::Rpa2 formant ces 'méga' foyers.

Pour cela, nous avons exposé les cellules *gfp::rpa2* à la phléomycine pour ensuite marquer l'ADN par le colorant Hoechst. Les images ont ensuite été réalisées en excitant d'abord l'Hoechst puis en excitant la GFP sur la même coupe au focus. Ceci a abouti à l'obtention de deux images : une correspond à l'excitation de l'Hoechst intercalé dans l'ADN, une correspond à l'excitation de la GFP des protéines GFP::Rpa2. Puis ces deux images sont superposées. Ces expériences ont été réalisées trois fois sur trois cultures de cellules différentes.

L'excitation de l'Hoechst compris dans l'ADN des cellules exposées à la phléomycine permet d'observer l'ADN compacté, représenté en rouge (Figure 56 A). L'excitation de la GFP des protéines GFP::Rpa2 des mêmes cellules montre une image pratiquement identique, pour laquelle les foyers de fluorescence formés par la protéine GFP::Rpa2, représentés en vert, semblent colocaliser avec l'ADN (Figure 56 B). Lorsque les deux images sont fusionnées, la superposition des foyers de fluorescence et de l'ADN forment un foyer jaune (Figure 56 C). Le pourcentage de cellules dont l'ADN est compacté a été comparé entre les cellules exposées à la phléomycine et les cellules contrôles. 80% des cellules exposées à 0,1 mg/mL de phléomycine ont leur ADN compacté contre moins de 10 % pour les cellules contrôles (Figure 56 D). Egalement 80% des cellules exposées à 0,1 mg/mL de phléomycine possèdent des

'mega' foyers (Figure 56 E). Parmi ces 'mega' foyers, 80% colocalisent avec l'ADN compacté (Figure 56 F), comme illustré Figure 56 C.

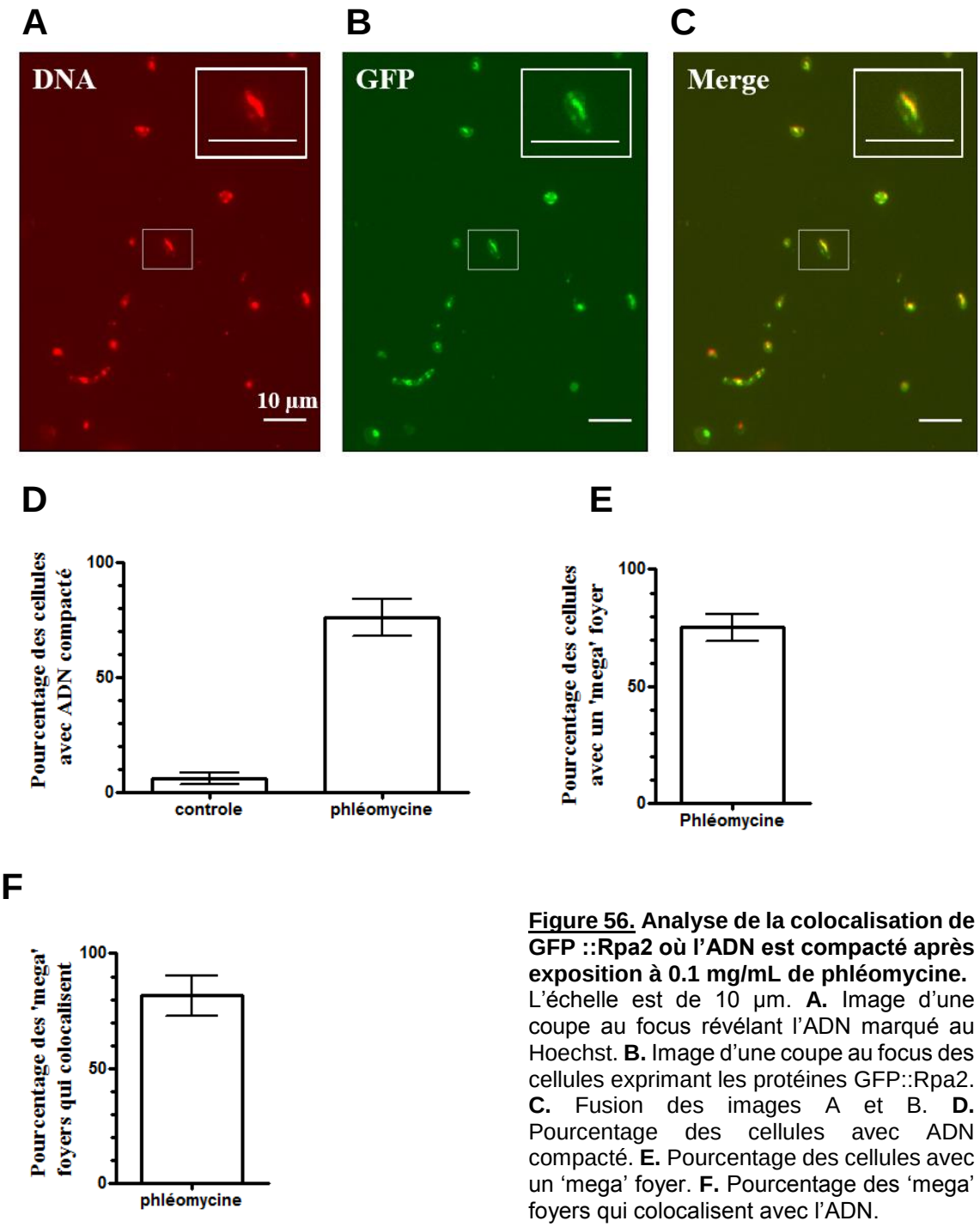


Figure 56. Analyse de la colocalisation de GFP ::Rpa2 où l'ADN est compacté après exposition à 0.1 mg/mL de phléomycine. L'échelle est de 10 µm. **A.** Image d'une coupe au focus révélant l'ADN marqué au Hoechst. **B.** Image d'une coupe au focus des cellules exprimant les protéines GFP::Rpa2. **C.** Fusion des images A et B. **D.** Pourcentage des cellules avec ADN compacté. **E.** Pourcentage des cellules avec un 'mega' foyer. **F.** Pourcentage des 'mega' foyers qui colocalisent avec l'ADN.

Ces expériences ont également été réalisées sur les cellules exposées 17h00 à 5 et 10 µg/mL d'APD. Aucune différence significative de compaction de l'ADN n'a été observée entre les cellules contrôles et les cellules exposées à l'APD.

L'ensemble de ces résultats permet de conclure que la compaction de l'ADN est spécifiquement induite par la phléomycine, et que 80 % des 'mega' foyers formés par GFP::Rpa2 colocalisent avec l'ADN compacté. La protéine GFP::Rpa2 semble donc recrutée spécifiquement au niveau de l'ADN compacté.

3.4.8 Dynamique de Rpa2 : Fluorescence Recovery After Photobleaching (FRAP)

L'observation de la localisation des protéines GFP::Rpa2 après exposition à différents agents génotoxiques a permis d'identifier trois types de localisation : une localisation obtenue lorsque les cellules sont exposées à l'aphidicoline, une localisation lorsque les cellules sont exposées à la phléomycine et une localisation de référence obtenue pour les cellules contrôles (exposées à l'eau ou au DMSO). Ces trois types de localisation reflètent trois recrutement de la protéine GFP::Rpa2 différents : (i) en condition contrôle, le nombre de foyers de fluorescence est de 2; (ii) avec l'APD, le nombre de foyers de fluorescence double en comparaison des cellules contrôles; (iii) avec la phléomycine, c'est la surface occupée par le foyer de fluorescence qui augmente ou les foyers de fluorescence sont trop nombreux et trop rapprochés qu'il n'est pas possible d'être suffisamment résolutif pour les distinguer. Afin de comprendre ces différences, nous nous sommes intéressés à la dynamique de diffusion qui existait au sein de ces trois types de foyers de fluorescence formés.

Pour cela, nous avons utilisé la technique de FRAP (Fluorescence Recovery After Photobleaching), qui consiste à photoblanchir une zone, comprenant tout le foyer de fluorescence, au moyen d'une exposition au laser et à mesurer le retour de la fluorescence au sein de la zone sélectionnée sur une période de temps définie.

Les cellules ont préalablement été immobilisées sur lame couverte de poly-D lysine. Les paramètres de photoblanchiment ont été optimisés, la puissance laser a été réglée à 60% et la zone a été photoblanchie en utilisant trois points de la taille du faisceau laser au sein du foyer de fluorescence. Ces expériences ont été réalisées trois fois de façon indépendante.

Dans le cas du photoblanchiment des foyers de fluorescence obtenus après exposition à 5 µg/mL d'APD : douze foyers ont été photoblanchis avec succès. Un type de population majoritaire est obtenu : pour 83% des foyers, le retour de la fluorescence se fait assez rapidement à 100% (Tableau 17 et Figure 58 A). Dans le cas du photoblanchiment des foyers de fluorescence obtenus après exposition à 0,1 mg/mL de phléomycine : onze foyers ont été photoblanchis avec succès. Un type majoritaire de population est obtenu : pour 90% des cas, la fluorescence ne revient pas (Tableau 17 et Figure 58 B). La plupart des molécules

GFP::Rpa2 présentes dans ces foyers ne sont donc pas renouvelées. Ainsi, la compaction freine la diffusion des protéines Rpa2. Dans le cas du photoblanchiment des foyers de fluorescence des cellules contrôles : vingt-quatre foyers ont été photoblanchis avec succès. Deux types de populations sont observés : pour 50 % des cas, la fluorescence revient à 100% et pour 50% des cas le retour de la fluorescence est faible voire nul.

	Aphidicoline	Phléomycine	Phase stationnaire
Nombre de foyers	4	2	2
Retour de la fluorescence après FRAP	Oui 83 % des foyers	Non 90 % des foyers	Oui - 50 %des foyers Non - 50 % des foyers

Tableau 17. Analyse du retour de fluorescence après exposition à l'aphidicoline, à la phléomycine et en phase stationnaire de croissance

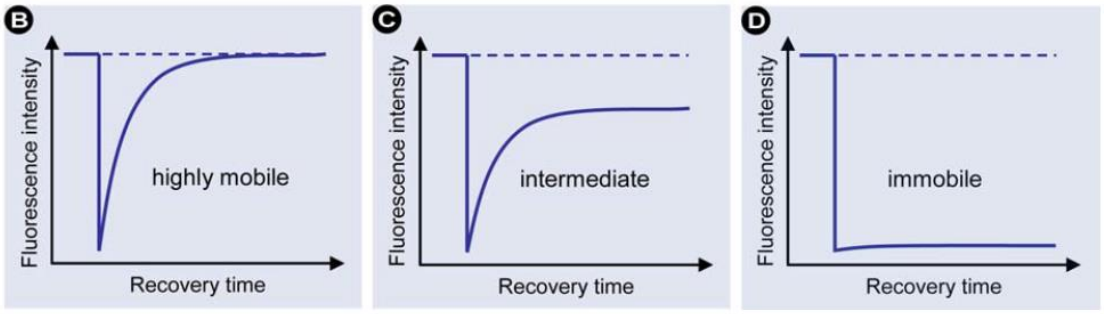


Figure 57. Trois cas de retour de fluorescence après photoblanchiment
B. Protéines très mobiles. **C.** Protéine à mobilité intermédiaire. **D.** Protéines immobiles. D'après [146].

Il existe trois types de réponse au photoblanchiment : (i) le retour de la fluorescence est rapide et complet, les protéines sont très mobiles, (ii) le retour de fluorescence est plus lent et incomplet, les protéines sont moyennement mobiles et (iii) le retour de fluorescence est nul, les protéines sont immobiles (Figure 57). Nous avons mesuré le retour de fluorescence et la fraction de protéines mobiles et immobiles pour les cellules en phase stationnaire de croissance, exposées à l'aphidicoline et à la phléomycine. La mobilité des protéines est complétement différente selon les conditions testées, les protéines sont très mobiles au sein des foyers photoblanchis après exposition à l'aphidicoline, elles sont très faiblement mobiles au sein des foyers photoblanchis après exposition à la phléomycine et peuvent être très mobiles et très peu mobiles au sein des foyers des cellules en phase stationnaire (Tableau 18).

	Fraction mobile	Fraction immobile
Aphidicoline (Fig. 58 A)	100 %	0 %
Phléomycine (Fig. 58 B)	15 %	85 %
Phase stationnaire - Avec retour de fluorescence (Fig. 58 C)	72 %	28 %
Phase stationnaire - Sans retour de fluorescence (Fig. 58 D)	7 %	93 %

Tableau 18. Quantification des fractions mobiles et immobiles
Les données sont relatives aux graphes présentés Figure 58.

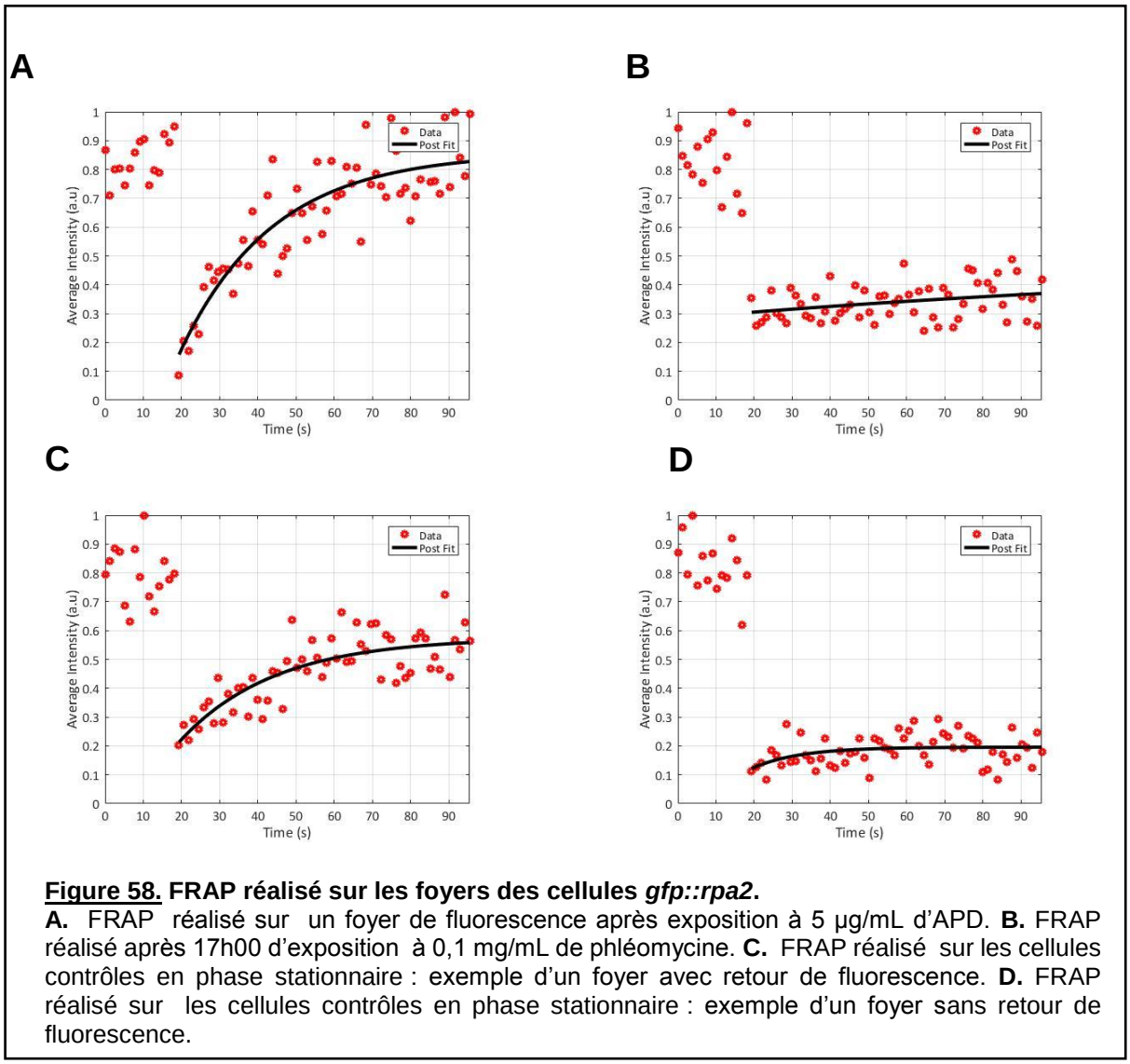


Figure 58. FRAP réalisé sur les foyers des cellules *gfp::rpa2*.
A. FRAP réalisé sur un foyer de fluorescence après exposition à 5 µg/mL d'APD. **B.** FRAP réalisé après 17h00 d'exposition à 0,1 mg/mL de phléomycine. **C.** FRAP réalisé sur les cellules contrôles en phase stationnaire : exemple d'un foyer avec retour de fluorescence. **D.** FRAP réalisé sur les cellules contrôles en phase stationnaire : exemple d'un foyer sans retour de fluorescence.

De façon intéressante, la réponse des cellules au photoblanchiment est différente en fonction des conditions testées. Si, après photoblanchiment des foyers de fluorescence donc des protéines GFP::Rpa2, on observe un retour de la fluorescence, cela signifie que les protéines GFP::Rpa2 diffusent. Cela illustre un processus dynamique qui a lieu au sein de ces

foyers de fluorescence. En revanche, lorsque il n'y a pas de retour de fluorescence, les protéines GFP::Rpa2 photoblanchies n'ont pas diffusées.

Plusieurs hypothèses peuvent expliquer cette différence de la dynamique de des protéines GFP::Rpa2. Ceci peut être expliqué par une différence d'accessibilité de la molécule d'ADN. En effet, lorsque les cellules sont exposées à l'aphidicoline, le processus de la réplication est arrêté mais la molécule d'ADN reste accessible à d'autres protéines permettant par exemple le redémarrage des fourches ou le renouvellement des protéines GFP::Rpa2. En revanche, lorsque les cellules sont exposées à la phléomycine, l'ADN est généralement sous forme compactée. Cette conformation de l'ADN peut rendre difficile son accès à d'autres protéines ou aux nouvelles Rpa2. Ou bien, le rôle de Rpa2 n'est peut-être pas le même aux fourches arrêtées qu'au niveau de l'ADN compacté et une fois photoblanchies les protéines Rpa2 peuvent rester sur l'ADN tout en assurant leur fonction de protection.

En conclusion, la protéine GFP::Rpa2 est fonctionnelle *in vivo.* De façon surprenante, les cellules en phase exponentielle de croissance ne présentent pas plus de foyers de fluorescence que celles en phase stationnaire et les intensités de fluorescence observées ne varient pas. Cependant lorsque les cellules sont exposées à un stress réplicatif (APD), on observe une augmentation significative du nombre de foyers de fluorescence en comparaison aux cellules en phase exponentielle de croissance. De plus, il existe une très forte dynamique au sein de ces foyers, les protéines GFP::Rpa2 présentent une mobilité importante. Lorsque les cellules sont exposées à des agents dommageant de l'ADN, comme la mitomycine C et la phléomycine, la localisation des foyers de fluorescence est différente. En effet, le nombre de foyers de fluorescence ne varie pas significativement après exposition à la MMC alors qu'il diminue après exposition à la phléomycine. De plus, l'aspect des foyers de fluorescence après exposition à la phléomycine est très différent de toutes les autres conditions testées. En effet, sa surface occupée augmente, ceci peut aussi être un biais de la technique utilisée : il existe une concentration plus importante de ces foyers au même endroit et la résolution n'est pas assez puissante pour les définir individuellement. Cette forte concentration au même endroit de ces foyers de fluorescence ou la présence d'un gros foyer de fluorescence colocalise avec l'ADN compacté. En dépit des limites de détection utilisées (résolution = 250 nm), cette colocalisation constitue un argument en plus pour justifier que ces foyers représentent bien les molécules de GFP::Rpa2 fixées à l'ADN. Ce méga foyer est le reflet d'un recrutement localisé à l'ADN compacté de GFP::Rpa2. De plus, la dynamique des protéines au sein de ces foyers diffère de celle obtenue avec l'APD. En effet, pour la plupart des foyers photoblanchis la fluorescence ne revient pas, les molécules ne diffusent pas. Ceci reflète-t-il un rôle de Rpa2 en tant que plateforme d'ancrage pour d'autres protéines intervenant dans la réparation par recombinaison homologue ou en tant que protéine de recombinaison homologue ? Ces différences de dynamique et de localisation qui existent entre l'APD et la phléomycine sont-

elles le reflet de différences concernant la topologie de l'ADN sur laquelle elles se fixent ? ou mettent-elles en évidence les différents rôles potentiels joués par cette même protéine ? Il est clair que l'état de la condensation de la chromatine peut avoir un rôle dans la localisation des protéines mais il est peu probable qu'elle provoque à elle seule une si grande différence de localisation après exposition à différents agents pour une même protéine. C'est pourquoi il est plus probable que la différence de localisation illustre les différents rôles possibles de Rpa2 *in vivo.* Dans le cas de l'exposition à l'APD, l'augmentation du nombre de foyer de fluorescence peut être le reflet d'une augmentation d'ADN simple brin sur lequel Rpa2 se fixe afin de le protéger ou le reflet d'un rôle de plateforme d'ancrage assurée par Rpa2 pour les protéines de redémarrage des fourches de réplication. Dans le cas de l'exposition à la phléomycine, l'augmentation de la surface du foyer peut davantage refléter le rôle de Rpa2 dans la recombinaison homologue nécessaire pour réparer les cassures double brin.

4 Conclusion et Perspectives

Les archées ont longtemps été assimilées à des bactéries, ce n'est que récemment, grâce au développement de marqueurs moléculaires, qu'elles ont été identifiées comme un domaine à part. Elles peuvent être classées en trois groupes principaux : les crenarchées, les euryarchées et les thaumarchées. Néanmoins cette classification est régulièrement mise à jour et remise en cause par la découverte de nouvelles archées comme c'est le cas avec la découverte des lokiarchées. La découverte de ces archées a remis en question l'évolution entre les archées et les eucaryotes. C'est un sujet de débat très actif, certains favorisent l'hypothèse selon laquelle les archées et les eucaryotes ont un ancêtre commun, ce qui fait des archées nos lointaines cousines, tandis que d'autres favorisent l'hypothèse selon laquelle les eucaryotes ont un ancêtre archéen. Les archées sont ubiquitaires, elles vivent à la fois dans des milieux extrêmes (cheminées hydrothermales, grands lacs salés…) mais aussi dans des milieux plus communs (lacs, système digestif humain…). Concernant la réplication, la transcription et la traduction, elles partagent des caractéristiques à la fois avec les eucaryotes mais aussi avec les bactéries. La diversité de leur habitat, de leur morphologie, de la structure de leur génome, des mécanismes intracellulaires de la réplication, la transcription et la traduction est élevée au sein des archées mais également au sein d'un même phylum. Le développement de la génétique et de la microscopie chez plusieurs organismes dont *Haloferax volcanii* ont permis l'étude *in vivo* de ces processus.

Notre objectif concernant ce travail de thèse était d'étudier la localisation et la dynamique de recrutement des protéines de la réplication pour mieux comprendre leur rôle *in vivo* chez l'archée halophile *Haloferax volcanii*. Pour répondre à cette question, nous avons choisi quatre protéines d'intérêt intervenant dans la fourche de réplication : l'anneau de processivité de l'ADN polymérase PCNA, Fen1 intervenant dans la maturation des fragments d'Okazaki, Orc1 reconnaissant les origines de réplication et Rpa2 intervenant dans la protection de l'ADN simple brin. Pour les localiser par microscopie à fluorescence, la même approche a été adoptée pour ces protéines : une expression physiologique de la protéine d'intérêt fusionnée à la protéine fluorescente GFP.

En ce qui concerne l'étude de la protéine PCNA, nous n'avons pu obtenir la construction souhaitée. La richesse de ses interactions à la fois protéiques et désoxyribonucléiques rend difficile son marquage par des protéines fluorescentes telle que la GFP. La structure de cette protéine est hautement conservée dans les trois domaines du vivant. Sa structure particulière en anneau lui permet d'interagir en son centre avec l'ADN et à l'extérieur avec d'autres protéines, jouant le rôle de plateforme de recrutement et de coordination. L'ajout d'une séquence protéique comme la GFP peut rendre difficile son repliement, l'empêchant d'assurer sa fonction ou gênant ses interactions protéiques qui peuvent être essentielles. Ceci peut expliquer le fait que nous n'avons pu obtenir cette construction. Une autre approche moins invasive peut être envisagée pour localiser cette

protéine *in vivo* comme la technique de marquage par des épitopes tags. La protéine d'intérêt est fusionnée à un épitope qui n'émet de la fluorescence que s'il interagit avec un autre peptide fluorescent pouvant être exprimé à partir d'un plasmide. Ces étiquettes portées par les protéines sont plus petites que celles de la GFP, elles constituent des versions tronquées des protéines sfcherry et GFP superfolder qui gardent leur propriété de fluorescence [147]. D'autres approches nécessitant des traitements de cellules peuvent être envisagées comme l'immunomarquage de PCNA, qui permettrait d'étudier la localisation de PCNA selon la phase de croissance ou en réponse à des agents génotoxiques. Cette technique nécessite une première étape de purification de PCNA d'*H. volcanii* pour ensuite synthétiser les anticorps anti-PCNA.

Fen1 est une protéine marqueur de la maturation des fragments d'Okazaki. Elle n'est pas essentielle chez *H. volcanii* mais sa délétion entraine un retard de croissance. Nous avons réussi à obtenir sa construction pour exprimer la protéine Fen1 fusionnée à la GFP, quelques foyers de fluorescence sporadiques ont pu être observés mais aucune localisation spécifique n'a pu être observée. Une des explications possible est que Fen1 n'est pas présente en assez grande quantité pour former un foyer de fluorescence détectable par microscopie de fluorescence. Pour répondre à cette question, il faudrait connaitre la quantité minimale de GFP nécessaire pour former un foyer de fluorescence détectable. De plus, il serait intéressant de pouvoir synchroniser les cellules dans le but d'identifier une apparition de signal à un instant précis.

Nous nous sommes également intéressés à une autre protéine intervenant lors de l'initiation de la réplication, la protéine Cdc6/Orc1. Le rôle de cette protéine, qui n'est pas essentielle, ainsi que de ses treize autres homologues chez *H. volcanii,* est mal compris. Nous avons donc localisé cette protéine lors de la phase de croissance, en absence et en présence d'origines de réplication. La différence concernant la répartition et le nombre de foyers de fluorescence des cellules *gfp::orc1* pourrait être le reflet d'un recrutement de Orc1 à l'ADN plus important en début de phase exponentielle de croissance, qui serait plus marqué dans la souche Δ*ori.* Cela implique que Orc1 peut être recrutée sur l'ADN lors de la réplication indépendamment de la présence d'origines de réplication. Il serait donc intéressant d'étudier plus particulièrement la localisation de Orc1 sur l'ADN. Nous avons pour cela réalisé des expériences de Chip-Seq au moyen d'anticorps dirigés contre la GFP. Le Chip-Seq comprend une première étape d'immunoprécipitation des séquences d'ADN fixées par la protéine d'intérêt puis une seconde étape de séquençage de cet ADN co-précipité. Malheureusement nous n'avons pas réussi à spécifiquement immunoprécipiter les protéines fusionnées à la GFP. Nous souhaiterions désormais développer cette technique en utilisant des anticorps dirigés contre Orc1 ce qui nécessite une synthèse des anticorps anti-Orc1. Cette approche, combinée à une synchronisation des cellules, permettrait de mettre en évidence une différence

de recrutement de Orc1 en fonction de la phase de croissance. Cela permettrait également de comprendre sur quelles régions du chromosome se fixent ces protéines Orc1 en absence d'origines de réplication.

La protéine Rpa2 est une protéine intervenant dans la protection de l'ADN simple brin et est essentielle pour la viabilité de la cellule. Elle joue un rôle dans la réplication mais aussi dans des processus de réparation. Son rôle n'est pas encore complétement compris. Nous avons localisé cette protéine selon la phase de croissance et après exposition à différents agents génotoxiques. De façon surprenante pour une protéine intervenant dans la fourche de réplication, le nombre de foyers de fluorescence correspondant à la protéine GFP::Rpa2 et l'intensité de fluorescence des cellules restent constants lors de la phase de croissance. Ceci peut être expliqué par le fait que les processus dans lesquels Rpa2 intervient (y compris la réparation) nécessitent une action immédiate de Rpa2 ; un recrutement à partir d'un pool disponible de molécules présentes dans le cytoplasme étant plus rapide qu'une augmentation de l'expression du gène en réponse à un stress. Lorsque les cellules sont exposées à un stress réplicatif, induit par une exposition à l'aphidicoline, des foyers de fluorescence sont observés. Ils sont semblables à ceux observés lors de la phase de croissance, mais leur nombre augmente. Or, la quantité de protéine GFP::Rpa2 est identique entre les cellules en phase de croissance et après exposition à l'aphidicoline. L'aphidicoline provoque un arrêt des fourches de réplication, il est donc possible que l'augmentation du nombre de foyers observé ne soit pas due à une augmentation de la quantité de Rpa2 produite mais à une meilleure détection des foyers puisque la réplication est « figée », ce qui faciliterait la localisation. De plus, il existe une très forte dynamique au sein des foyers après exposition à l'aphidicoline, les protéines GFP::Rpa2 présentent une mobilité importante. L'hypothèse testée pour expliquer cette différence de localisation est que lorsque la fourche de réplication est arrêtée, il y a plus d'ADN simple brin. Afin de comprendre si cette augmentation du nombre de foyers de fluorescence est corrélée avec une augmentation de l'ADN simple brin, nous souhaiterions quantifier l'ADN simple brin pour les cellules exposées à l'aphidicoline et les cellules en phase de croissance. Lorsque les cellules sont exposées à des agents dommageant de l'ADN, comme la phléomycine, des « mega spot » sont observés, ils sont plus gros que les foyers de fluorescence observés après exposition à l'aphidicoline et en phase de croissance. Ce « méga spot » serait le reflet d'une agrégation de plusieurs foyers de fluorescence au même endroit. Cette forte concentration au même endroit des foyers de fluorescence colocalise avec l'ADN compacté. En dépit des limites de détection utilisées (≈250 nm), cette colocalisation constitue un argument en plus pour justifier que ces foyers représentent bien les molécules de GFP::Rpa2 fixées à l'ADN. Une meilleure résolution par des techniques de super-résolution de microscopie à fluorescence comme le SIM (Microscopie à Structure Illuminée) pourrait nous aider à mieux comprendre le rôle joué par Rpa2 lors de la compaction. Les différentes

localisations observées après exposition à l'aphidicoline et à la phléomycine sont spécifiques des différents stress à savoir les arrêts de fourche de réplication et les dommages de l'ADN. La dynamique de GFP::Rpa2 n'est pas identique entre les cellules en phase de croissance et les cellules exposées aux agents génotoxiques, les protéines présentent une forte mobilité après exposition à l'aphidicoline, une mobilité réduite en phase stationnaire et une mobilité pratiquement nulle après exposition à la phléomycine. Ces différences de dynamique et de localisation sont-elles le reflet de différences concernant la topologie de l'ADN sur laquelle elles se fixent ? ou mettent elles en évidence les différents rôles potentiels joués par cette même protéine ?

L'approche utilisée par l'étude *in vivo* de la localisation et de la dynamique de recrutement des protéines de la réplication chez *Haloferax volcanii* qui a été utilisée tout au long de ce travail de thèse a permis de détecter une localisation spécifique de Rpa2 chez *H. volcanii*. Cette approche a également permis de détecter une localisation de Orc1 qui est indépendante de la présence des origines de réplication et qui ne semble pas refléter son rôle dans la réplication. Cette technique possède l'avantage de visualiser *in vivo* l'expression des protéines selon leur niveau d'expression physiologique. La taille de la GFP peut néanmoins gêner le repliement de la protéine ou ses interactions comme c'est probablement le cas pour la protéine PCNA. La localisation des protéines Fen1 et Orc1 combinée avec des techniques de Chip-Seq et de synchronisation apporteraient une réelle compréhension du rôle de ces protéines dans la cellule.

REFERENCES

1. Woese, C.R. and G.E. Fox, *Phylogenetic structure of the prokaryotic domain: the primary kingdoms.* Proc Natl Acad Sci U S A, 1977. **74**(11): p. 5088-90.
2. Fox, G.E., et al., *Classification of methanogenic bacteria by 16S ribosomal RNA characterization.* Proc Natl Acad Sci U S A, 1977. **74**(10): p. 4537-41.
3. Woese, C.R., O. Kandler, and M.L. Wheelis, *Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4576-9.
4. Raymann, K., C. Brochier-Armanet, and S. Gribaldo, *The two-domain tree of life is linked to a new root for the Archaea.* Proc Natl Acad Sci U S A, 2015. **112**(21): p. 6670-5.
5. Huber, R., H. Huber, and K.O. Stetter, *Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties.* FEMS Microbiol Rev, 2000. **24**(5): p. 615-23.
6. Rothschild, L.J. and R.L. Mancinelli, *Life in extreme environments.* Nature, 2001. **409**(6823): p. 1092-101.
7. Chaban, B., S.Y. Ng, and K.F. Jarrell, *Archaeal habitats--from the extreme to the ordinary.* Can J Microbiol, 2006. **52**(2): p. 73-116.
8. Martens-Habbena, W., et al., *Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria.* Nature, 2009. **461**(7266): p. 976-9.
9. Brugere, J.F., et al., *Archaebiotics: proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease.* Gut Microbes. **5**(1): p. 5-10.
10. Jarrell, K.F. and S.V. Albers, *The archaellum: an old motility structure with a new name.* Trends Microbiol, 2012. **20**(7): p. 307-12.
11. Kandler, O. and H. Hippe, *Lack of peptidoglycan in the cell walls of Methanosarcina barkeri.* Arch Microbiol, 1977. **113**(1-2): p. 57-60.
12. Albers, S.V. and B.H. Meyer, *The archaeal cell envelope.* Nat Rev Microbiol, 2011. **9**(6): p. 414-26.
13. Ulrih, N.P., D. Gmajner, and P. Raspor, *Structural and physicochemical properties of polar lipids from thermophilic archaea.* Appl Microbiol Biotechnol, 2009. **84**(2): p. 249-60.
14. Raymann, K., et al., *Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea.* Genome Biol Evol, 2014. **6**(1): p. 192-212.
15. Langer, D., et al., *Transcription in archaea: similarity to that in eucarya.* Proc Natl Acad Sci U S A, 1995. **92**(13): p. 5768-72.
16. Jun, S.H., et al., *Archaeal RNA polymerase and transcription regulation.* Crit Rev Biochem Mol Biol, 2011. **46**(1): p. 27-40.
17. Bell, S.D., *Archaeal transcriptional regulation--variation on a bacterial theme?* Trends Microbiol, 2005. **13**(6): p. 262-5.
18. Facciotti, M.T., et al., *General transcription factor specified global gene regulation in archaea.* Proc Natl Acad Sci U S A, 2007. **104**(11): p. 4630-5.
19. Watanabe, Y., et al., *Introns in protein-coding genes in Archaea.* FEBS Lett, 2002. **510**(1-2): p. 27-30.
20. Portnoy, V., et al., *RNA polyadenylation in Archaea: not observed in Haloferax while the exosome polynucleotidylates RNA in Sulfolobus.* EMBO Rep, 2005. **6**(12): p. 1188-93.
21. Sartorius-Neef, S. and F. Pfeifer, *In vivo studies on putative Shine-Dalgarno sequences of the halophilic archaeon Halobacterium salinarum.* Mol Microbiol, 2004. **51**(2): p. 579-88.
22. Hartman, H., P. Favaretto, and T.F. Smith, *The archaeal origins of the eukaryotic translational system.* Archaea, 2006. **2**(1): p. 1-9.
23. Meselson, M. and F.W. Stahl, *The replication of DNA.* Cold Spring Harb Symp Quant Biol, 1958. **23**: p. 9-12.

24. Sarmiento, F., et al., *Diversity of the DNA replication system in the Archaea domain.* Archaea, 2014. **2014**: p. 675946.
25. Barry, E.R. and S.D. Bell, *DNA replication in the archaea.* Microbiol Mol Biol Rev, 2006. **70**(4): p. 876-87.
26. Wu, Z., et al., *Multiple replication origins with diverse control mechanisms in Haloarcula hispanica.* Nucleic Acids Res, 2014. **42**(4): p. 2282-94.
27. Pelve, E.A., et al., *Mapping of active replication origins in vivo in thaum- and euryarchaeal replicons.* Mol Microbiol, 2013. **90**(3): p. 538-50.
28. Yang, H., et al., *Activation of a dormant replication origin is essential for Haloferax mediterranei lacking the primary origins.* Nat Commun, 2015. **6**: p. 8321.
29. Hawkins, M., et al., *Accelerated growth in the absence of DNA replication origins.* Nature, 2013. **503**(7477): p. 544-7.
30. Norais, C., et al., *Genetic and physical mapping of DNA replication origins in Haloferax volcanii.* PLoS Genet, 2007. **3**(5): p. e77.
31. Roxane Lestini, F.D., Hannu Myllykallio, ed. *DNA Replication Restart in Archaea*. Advances in DNA repair. 2015, Intech.
32. Myllykallio, H., et al., *Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon.* Science, 2000. **288**(5474): p. 2212-5.
33. Maisnier-Patin, S., et al., *Chromosome replication patterns in the hyperthermophilic euryarchaea Archaeoglobus fulgidus and Methanocaldococcus (Methanococcus) jannaschii.* Mol Microbiol, 2002. **45**(5): p. 1443-50.
34. Berquist, B.R. and S. DasSarma, *An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, Halobacterium sp. strain NRC-1.* J Bacteriol, 2003. **185**(20): p. 5959-66.
35. Coker, J.A., et al., *Multiple replication origins of Halobacterium sp. strain NRC-1: properties of the conserved orc7-dependent oriC1.* J Bacteriol, 2009. **191**(16): p. 5253-61.
36. Majernik, A.I. and J.P. Chong, *A conserved mechanism for replication origin recognition and binding in archaea.* Biochem J, 2008. **409**(2): p. 511-8.
37. Lundgren, M., et al., *Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination.* Proc Natl Acad Sci U S A, 2004. **101**(18): p. 7046-51.
38. Samson, R.Y., et al., *Specificity and function of archaeal DNA replication initiator proteins.* Cell Rep, 2013. **3**(2): p. 485-96.
39. Pelve, E.A., et al., *Four chromosome replication origins in the archaeon Pyrobaculum calidifontis.* Mol Microbiol, 2012. **85**(5): p. 986-95.
40. Robinson, N.P. and S.D. Bell, *Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes.* Proc Natl Acad Sci U S A, 2007. **104**(14): p. 5806-11.
41. Matsunaga, F., et al., *In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin.* Proc Natl Acad Sci U S A, 2001. **98**(20): p. 11152-7.
42. Matsunaga, F., et al., *Genomewide and biochemical analyses of DNA-binding activity of Cdc6/Orc1 and Mcm proteins in Pyrococcus sp.* Nucleic Acids Res, 2007. **35**(10): p. 3214-22.
43. Robinson, N.P., et al., *Identification of two origins of replication in the single chromosome of the archaeon Sulfolobus solfataricus.* Cell, 2004. **116**(1): p. 25-38.
44. MacNeill, S.A., *The haloarchaeal chromosome replication machinery.* Biochem Soc Trans, 2009. **37**(Pt 1): p. 108-13.
45. Ishino, S., et al., *Biochemical and genetical analyses of the three mcm genes from the hyperthermophilic archaeon, Thermococcus kodakarensis.* Genes Cells, 2011. **16**(12): p. 1176-89.
46. Pan, M., et al., *Thermococcus kodakarensis encodes three MCM homologs but only one is essential.* Nucleic Acids Res, 2011. **39**(22): p. 9671-80.

47. Sarmiento, F., J. Mrazek, and W.B. Whitman, *Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon Methanococcus maripaludis.* Proc Natl Acad Sci U S A, 2013. **110**(12): p. 4726-31.
48. Kelman, Z., J.K. Lee, and J. Hurwitz, *The single minichromosome maintenance protein of Methanobacterium thermoautotrophicum DeltaH contains DNA helicase activity.* Proc Natl Acad Sci U S A, 1999. **96**(26): p. 14783-8.
49. Shechter, D.F., C.Y. Ying, and J. Gautier, *The intrinsic DNA helicase activity of Methanobacterium thermoautotrophicum delta H minichromosome maintenance protein.* J Biol Chem, 2000. **275**(20): p. 15049-59.
50. De Felice, M., et al., *A CDC6-like factor from the archaea Sulfolobus solfataricus promotes binding of the mini-chromosome maintenance complex to DNA.* J Biol Chem, 2004. **279**(41): p. 43008-12.
51. Shin, J.H., et al., *Regulation of minichromosome maintenance helicase activity by Cdc6.* J Biol Chem, 2003. **278**(39): p. 38059-67.
52. Kasiviswanathan, R., J.H. Shin, and Z. Kelman, *Interactions between the archaeal Cdc6 and MCM proteins modulate their biochemical properties.* Nucleic Acids Res, 2005. **33**(15): p. 4940-50.
53. Akita, M., et al., *Cdc6/Orc1 from Pyrococcus furiosus may act as the origin recognition protein and Mcm helicase recruiter.* Genes Cells, 2010. **15**(5): p. 537-52.
54. Marinsek, N., et al., *GINS, a central nexus in the archaeal DNA replication fork.* EMBO Rep, 2006. **7**(5): p. 539-45.
55. Oyama, T., et al., *Architectures of archaeal GINS complexes, essential DNA replication initiation factors.* BMC Biol, 2011. **9**: p. 28.
56. Ogino, H., et al., *The GINS complex from the thermophilic archaeon, Thermoplasma acidophilum may function as a homotetramer in DNA replication.* Extremophiles, 2011. **15**(4): p. 529-39.
57. Yoshimochi, T., et al., *The GINS complex from Pyrococcus furiosus stimulates the MCM helicase activity.* J Biol Chem, 2008. **283**(3): p. 1601-9.
58. Forterre, P., et al., *Origin and evolution of DNA topoisomerases.* Biochimie, 2007. **89**(4): p. 427-46.
59. Borges, K.M., et al., *Characterization of the reverse gyrase from the hyperthermophilic archaeon Pyrococcus furiosus.* J Bacteriol, 1997. **179**(5): p. 1721-6.
60. Kikuchi, A. and K. Asai, *Reverse gyrase--a topoisomerase which introduces positive superhelical turns into DNA.* Nature, 1984. **309**(5970): p. 677-81.
61. Nollmann, M., N.J. Crisona, and P.B. Arimondo, *Thirty years of Escherichia coli DNA gyrase: from in vivo function to single-molecule mechanism.* Biochimie, 2007. **89**(4): p. 490-9.
62. Dabrowski, S., et al., *Identification and characterization of single-stranded-DNA-binding proteins from Thermus thermophilus and Thermus aquaticus - new arrangement of binding domains.* Microbiology, 2002. **148**(Pt 10): p. 3307-15.
63. Landwehr, M., U. Curth, and C. Urbanke, *A dimeric mutant of the homotetrameric single-stranded DNA binding protein from Escherichia coli.* Biol Chem, 2002. **383**(9): p. 1325-33.
64. Iftode, C., Y. Daniely, and J.A. Borowiec, *Replication protein A (RPA): the eukaryotic SSB.* Crit Rev Biochem Mol Biol, 1999. **34**(3): p. 141-80.
65. Murzin, A.G., *OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences.* EMBO J, 1993. **12**(3): p. 861-7.
66. Wadsworth, R.I. and M.F. White, *Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus.* Nucleic Acids Res, 2001. **29**(4): p. 914-20.
67. Komori, K. and Y. Ishino, *Replication protein A in Pyrococcus furiosus is involved in homologous DNA recombination.* J Biol Chem, 2001. **276**(28): p. 25654-60.

68. Skowyra, A. and S.A. MacNeill, *Identification of essential and non-essential single-stranded DNA-binding proteins in a model archaeal organism.* Nucleic Acids Res, 2012. **40**(3): p. 1077-90.
69. Walters, A.D. and J.P. Chong, *Methanococcus maripaludis: an archaeon with multiple functional MCM proteins?* Biochem Soc Trans, 2009. **37**(Pt 1): p. 1-6.
70. Frick, D.N. and C.C. Richardson, *DNA primases.* Annu Rev Biochem, 2001. **70**: p. 39-80.
71. NORAIS, C., *Etude de la réplication de l'ADN chez l'arhaea halophile Haloferax volcanii*, in *Biologie*. 2007, Université paris 11.
72. Ohmori, H., et al., *The Y-family of DNA polymerases.* Mol Cell, 2001. **8**(1): p. 7-8.
73. Ishino, Y. and I.K. Cann, *The euryarchaeotes, a subdomain of Archaea, survive on a single DNA polymerase: fact or farce?* Genes Genet Syst, 1998. **73**(6): p. 323-36.
74. Cann, I.K., et al., *A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase.* Proc Natl Acad Sci U S A, 1998. **95**(24): p. 14250-5.
75. Ishino, Y., et al., *A novel DNA polymerase family found in Archaea.* J Bacteriol, 1998. **180**(8): p. 2232-6.
76. Boudsocq, F., et al., *Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta.* Nucleic Acids Res, 2001. **29**(22): p. 4607-16.
77. Henneke, G., et al., *The hyperthermophilic euryarchaeota Pyrococcus abyssi likely requires the two DNA polymerases D and B for DNA replication.* J Mol Biol, 2005. **350**(1): p. 53-64.
78. Jokela, M., et al., *Characterization of the 3' exonuclease subunit DP1 of Methanococcus jannaschii replicative DNA polymerase D.* Nucleic Acids Res, 2004. **32**(8): p. 2430-40.
79. Cubonova, L., et al., *Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in Thermococcus kodakarensis.* J Bacteriol, 2013. **195**(10): p. 2322-8.
80. Li, Z., et al., *Affinity purification of an archaeal DNA replication protein network.* MBio, 2010. **1**(5).
81. Daimon, K., et al., *Three proliferating cell nuclear antigen-like proteins found in the hyperthermophilic archaeon Aeropyrum pernix: interactions with the two DNA polymerases.* J Bacteriol, 2002. **184**(3): p. 687-94.
82. Dionne, I., et al., *A heterotrimeric PCNA in the hyperthermophilic archaeon Sulfolobus solfataricus.* Mol Cell, 2003. **11**(1): p. 275-82.
83. Kuba, Y., et al., *Comparative analyses of the two proliferating cell nuclear antigens from the hyperthermophilic archaeon, Thermococcus kodakarensis.* Genes Cells, 2012. **17**(11): p. 923-37.
84. Beattie, T.R. and S.D. Bell, *Coordination of multiple enzyme activities by a single PCNA in archaeal Okazaki fragment maturation.* EMBO J, 2012. **31**(6): p. 1556-67.
85. Creze, C., et al., *Modulation of the Pyrococcus abyssi NucS endonuclease activity by replication clamp at functional and structural levels.* J Biol Chem, 2012. **287**(19): p. 15648-60.
86. Ren, B., et al., *Structure and function of a novel endonuclease acting on branched DNA substrates.* EMBO J, 2009. **28**(16): p. 2479-89.
87. MacNeill, S.A., *PCNA-binding proteins in the archaea: novel functionality beyond the conserved core.* Curr Genet, 2016. **62**(3): p. 527-32.
88. Pluchon, P.F., et al., *An extended network of genomic maintenance in the archaeon Pyrococcus abyssi highlights unexpected associations between eucaryotic homologs.* PLoS One, 2013. **8**(11): p. e79707.
89. Giroux, X. and S.A. MacNeill, *A novel archaeal DNA repair factor that acts with the UvrABC system to repair mitomycin C-induced DNA damage in a PCNA-dependent manner.* Mol Microbiol, 2016. **99**(1): p. 1-14.
90. Yao, N.Y., et al., *Mechanism of proliferating cell nuclear antigen clamp opening by replication factor C.* J Biol Chem, 2006. **281**(25): p. 17528-39.
91. Jeruzalmi, D., M. O'Donnell, and J. Kuriyan, *Clamp loaders and sliding clamps.* Curr Opin Struct Biol, 2002. **12**(2): p. 217-24.

92. Cann, I.K., et al., *Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea.* J Bacteriol, 1999. **181**(21): p. 6591-9.
93. De Felice, M., et al., *Two DNA polymerase sliding clamps from the thermophilic archaeon Sulfolobus solfataricus.* J Mol Biol, 1999. **291**(1): p. 47-57.
94. Kelman, Z. and J. Hurwitz, *A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon Methanobacterium thermoautotrophicum deltaH.* J Biol Chem, 2000. **275**(10): p. 7327-36.
95. Seybert, A., et al., *Biochemical characterisation of the clamp/clamp loader proteins from the euryarchaeon Archaeoglobus fulgidus.* Nucleic Acids Res, 2002. **30**(20): p. 4329-38.
96. Seybert, A., et al., *Communication between subunits within an archaeal clamp-loader complex.* EMBO J, 2006. **25**(10): p. 2209-18.
97. Seybert, A. and D.B. Wigley, *Distinct roles for ATP binding and hydrolysis at individual subunits of an archaeal clamp loader.* EMBO J, 2004. **23**(6): p. 1360-71.
98. Henneke, G., *In vitro reconstitution of RNA primer removal in Archaea reveals the existence of two pathways.* Biochem J, 2012. **447**(2): p. 271-80.
99. Meslet-Cladiere, L., et al., *A novel proteomic approach identifies new interaction partners for proliferating cell nuclear antigen.* J Mol Biol, 2007. **372**(5): p. 1137-48.
100. Hartman, A.L., et al., *The complete genome sequence of Haloferax volcanii DS2, a model archaeon.* PLoS One, 2010. **5**(3): p. e9605.
101. Bitan-Banin, G., R. Ortenberg, and M. Mevarech, *Development of a gene knockout system for the halophilic archaeon Haloferax volcanii by use of the pyrE gene.* J Bacteriol, 2003. **185**(3): p. 772-8.
102. Allers, T., et al., *Development of additional selectable markers for the halophilic archaeon Haloferax volcanii based on the leuB and trpA genes.* Appl Environ Microbiol, 2004. **70**(2): p. 943-53.
103. *microscope contraste de phase*. Available from: www.forum.mikroscopia.com.
104. *microscopie à contraste interférentiel.*
105. Mullakhanbhai, M.F. and H. Larsen, *Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement.* Arch Microbiol, 1975. **104**(3): p. 207-14.
106. Cairo, J.J., et al., *Methanosarcina mazei JC2, a new methanogenic strain isolated from lake sediments, that does not use H2/CO2.* Microbiologia, 1992. **8**(1): p. 21-31.
107. Hixon, W.G. and D.G. Searcy, *Cytoskeleton in the archaebacterium Thermoplasma acidophilum? Viscosity increase in soluble extracts.* Biosystems, 1993. **29**(2-3): p. 151-60.
108. Kessel, M. and Y. Cohen, *Ultrastructure of square bacteria from a brine pool in Southern Sinai.* J Bacteriol, 1982. **150**(2): p. 851-60.
109. Burns, D.G., et al., *Haloquadratum walsbyi gen. nov., sp. nov., the square haloarchaeon of Walsby, isolated from saltern crystallizers in Australia and Spain.* Int J Syst Evol Microbiol, 2007. **57**(Pt 2): p. 387-92.
110. W. J. Jones, J.A.L., F. Mayer, C. R. Woese, R. S. Wolfe, *Methanococcus jannaschii sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent* Archives of microbiology, 1983.
111. Blochl, E., et al., *Pyrolobus fumarii, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C.* Extremophiles, 1997. **1**(1): p. 14-21.
112. Kushner, D.J. and H. Onishi, *Absence of normal cell wall constituents from the outer layers of Halobacterium cutirubrum.* Can J Biochem, 1968. **46**(8): p. 997-8.
113. Claus, H., et al., *Molecular organization of selected prokaryotic S-layer proteins.* Can J Microbiol, 2005. **51**(9): p. 731-43.
114. Wang, X. and J. Lutkenhaus, *FtsZ ring: the eubacterial division apparatus conserved in archaebacteria.* Mol Microbiol, 1996. **21**(2): p. 313-9.
115. Duggin, I.G., et al., *CetZ tubulin-like proteins control archaeal cell shape.* Nature, 2015. **519**(7543): p. 362-5.

116. Dobro, M.J., et al., *Electron cryotomography of ESCRT assemblies and dividing Sulfolobus cells suggests that spiraling filaments are involved in membrane scission.* Mol Biol Cell, 2013. **24**(15): p. 2319-27.
117. Robert D. Goldmann, J.R.S., David L. Spector, *Live Cell Imaging*. 2010.
118. Gristwood, T., et al., *The sub-cellular localization of Sulfolobus DNA replication.* Nucleic Acids Res, 2012. **40**(12): p. 5487-96.
119. Samson, R.Y., et al., *A role for the ESCRT system in cell division in archaea.* Science, 2008. **322**(5908): p. 1710-3.
120. Delmas, S., I.G. Duggin, and T. Allers, *DNA damage induces nucleoid compaction via the Mre11-Rad50 complex in the archaeon Haloferax volcanii.* Mol Microbiol, 2013. **87**(1): p. 168-79.
121. Salic, A. and T.J. Mitchison, *A chemical method for fast and sensitive detection of DNA synthesis in vivo.* Proc Natl Acad Sci U S A, 2008. **105**(7): p. 2415-20.
122. Shimomura, O., F.H. Johnson, and Y. Saiga, *Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea.* J Cell Comp Physiol, 1962. **59**: p. 223-39.
123. Ormo, M., et al., *Crystal structure of the Aequorea victoria green fluorescent protein.* Science, 1996. **273**(5280): p. 1392-5.
124. Tsien, R.Y., *The green fluorescent protein.* Annu Rev Biochem, 1998. **67**: p. 509-44.
125. Reuter, C.J. and J.A. Maupin-Furlow, *Analysis of proteasome-dependent proteolysis in Haloferax volcanii cells, using short-lived green fluorescent proteins.* Appl Environ Microbiol, 2004. **70**(12): p. 7530-8.
126. Patterson, G.H., et al., *Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy.* Biophys J, 1997. **73**(5): p. 2782-90.
127. Cranfill, P.J., et al., *Quantitative assessment of fluorescent proteins.* Nat Methods, 2016. **13**(7): p. 557-62.
128. Reits, E.A. and J.J. Neefjes, *From fixed to FRAP: measuring protein mobility and activity in living cells.* Nat Cell Biol, 2001. **3**(6): p. E145-7.
129. Gustafsson, M.G., *Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy.* J Microsc, 2000. **198**(Pt 2): p. 82-7.
130. Hirano, Y., A. Matsuda, and Y. Hiraoka, *Recent advancements in structured-illumination microscopy toward live-cell imaging.* Microscopy (Oxf), 2015. **64**(4): p. 237-49.
131. Morgunova, E., et al., *Structural insights into the adaptation of proliferating cell nuclear antigen (PCNA) from Haloferax volcanii to a high-salt environment.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 10): p. 1081-8.
132. Winter, J.A., et al., *The crystal structure of Haloferax volcanii proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation.* BMC Struct Biol, 2009. **9**: p. 55.
133. Costes, A., et al., *The C-terminal domain of the bacterial SSB protein acts as a DNA maintenance hub at active chromosome replication forks.* PLoS Genet, 2010. **6**(12): p. e1001238.
134. Lecointe, F., et al., *Anticipating chromosomal replication fork arrest: SSB targets repair DNA helicases to active forks.* EMBO J, 2007. **26**(19): p. 4239-51.
135. Richard, D.J., et al., *Single-stranded DNA-binding protein hSSB1 is critical for genomic stability.* Nature, 2008. **453**(7195): p. 677-81.
136. Shi, W., et al., *Essential developmental, genomic stability, and tumour suppressor functions of the mouse orthologue of hSSB1/NABP2.* PLoS Genet, 2013. **9**(2): p. e1003298.
137. Cubeddu, L. and M.F. White, *DNA damage detection by an archaeal single-stranded DNA-binding protein.* J Mol Biol, 2005. **353**(3): p. 507-16.
138. Stroud, A., S. Liddell, and T. Allers, *Genetic and Biochemical Identification of a Novel Single-Stranded DNA-Binding Complex in Haloferax volcanii.* Front Microbiol, 2012. **3**: p. 224.
139. Lestini, R., et al., *Intracellular dynamics of archaeal FANCM homologue Hef in response to halted DNA replication.* Nucleic Acids Res, 2013. **41**(22): p. 10358-70.

140. Wach, A., *PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in S. cerevisiae.* Yeast, 1996. **12**(3): p. 259-65.
141. Li, M.Z. and S.J. Elledge, *SLIC: a method for sequence- and ligation-independent cloning.* Methods Mol Biol, 2012. **852**: p. 51-9.
142. Forterre, P., C. Elie, and M. Kohiyama, *Aphidicolin inhibits growth and DNA synthesis in halophilic arachaebacteria.* J Bacteriol, 1984. **159**(2): p. 800-2.
143. Samson, R.Y. and S.D. Bell, *Archaeal DNA Replication Origins and Recruitment of the MCM Replicative Helicase.* Enzymes, 2016. **39**: p. 169-90.
144. Moran-Reyna, A. and J.A. Coker, *The effects of extremes of pH on the growth and transcriptomic profiles of three haloarchaea.* F1000Res, 2014. **3**: p. 168.
145. Rolfsmeier, M.L. and C.A. Haseltine, *The single-stranded DNA binding protein of Sulfolobus solfataricus acts in the presynaptic step of homologous recombination.* J Mol Biol, 2010. **397**(1): p. 31-45.
146. Hellen C. Ishikawa-Ankerhold, R.A.a.G.P.C.D., *Advanced Fluorescence Microscopy Techniques—FRAP, FLIP,FLAP, FRET and FLIM* Molecules, 2012.
147. Kamiyama, D., et al., *Versatile protein tagging in cells with split fluorescent protein.* Nat Commun, 2016. **7**: p. 11046.

ANNEXES

Biochimie 118 (2015) 254–263

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Review

DNA replication restart and cellular dynamics of Hef helicase/nuclease protein in *Haloferax volcanii*

Roxane Lestini*, Floriane Delpech, Hannu Myllykallio

Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS UMR7645–INSERM U1182, 91128, Palaiseau Cedex, France

ARTICLE INFO

Article history:
Received 5 March 2015
Accepted 22 July 2015
Available online 26 July 2015

Keywords:
Archaea
DNA replication restart
GFP
Hef

ABSTRACT

Understanding how frequently spontaneous replication arrests occur and how archaea deal with these arrests are very interesting and challenging research topics. Here we will described how genetic and imaging studies have revealed the central role of the archaeal helicase/nuclease Hef belonging to the XPF/MUS81/FANCM family of endonucleases in repair of arrested replication forks. Special focus will be on description of a recently developed combination of genetic and imaging tools to study the dynamic localization of a functional Hef::GFP (Green Fluorescent Protein) fusion protein in the living cells of halophilic archaea *Haloferax volcanii*. As Archaea provide an excellent and unique model for understanding how DNA replication is regulated to allow replication of a circular DNA molecule either from single or multiple replication origins, we will also summarize recent studies that have revealed peculiar features regarding DNA replication, particularly in halophilic archaea.

We strongly believe that fundamental knowledge of our on-going studies will shed light on the evolutionary history of the DNA replication machinery and will help to establish general rules concerning replication restart and the key role of recombination proteins not only in bacteria, yeast and higher eukaryotes but also in archaea.

© 2015 Elsevier B.V. and Société Française de Biochimie et Biologie Moléculaire (SFBBM). All rights reserved.

1. Introduction

All organisms must accurately replicate their genetic material for proliferation. This task is performed by conserved core machineries in all three domains of life: bacteria, archaea and eukaryotes [1–3]. In bacteria, the circular DNA chromosome is replicated from a single replication origin. Two replication forks assembled at replication origins progress in the opposite direction with similar average speeds, explaining why DNA replication terminates in the chromosomal region opposite the origin (Fig. 1A). Because each replication fork is required to duplicate approximately one half of the genome, replication fork arrests will clearly prevent replication of the entire chromosome. Studies in *Escherichia coli* strongly suggest that many replication forks encounter DNA damage or roadblocks leading to replisome inactivation under normal growth conditions, in the absence of any DNA damaging agents. These studies have revealed that several replication restart pathways have evolved depending on the cause of arrest. However, the common feature of all pathways repairing arrested replication forks is the involvement of bacterial recombination proteins such as the RecA recombinase and the PriA protein responsible for the loading of a replisome at recombination intermediates [4–7].

On the contrary, most of the eukaryotic linear chromosomes contain numerous replication origins. As dormant replication origins can be used as backup origins to rescue arrested replication forks, the importance of replication restart pathways in eukaryotes has long been ignored. However, recent studies have demonstrated that fork restart pathways operate also in eukaryotic cells and are important for cell viability under replication stress conditions. Studies of eukaryotic replication restart pathways also reveal that recombination proteins are involved, further indicating that recombination proteins play key roles in the repair of arrested replication forks [8–12].

As relatively little is still known regarding repair of arrested replication forks in archaeal cells, it is less clear whether involvement of DNA recombination in repair of arrested replication forks is a universally conserved feature. Interestingly, regarding DNA replication and replication restart, a hybrid situation between bacteria and eukarya is found in archaea. Like most bacterial

* Corresponding author.
E-mail address: roxane.lestini@polytechnique.edu (R. Lestini).

http://dx.doi.org/10.1016/j.biochi.2015.07.022
0300-9084/© 2015 Elsevier B.V. and Société Française de Biochimie et Biologie Moléculaire (SFBBM). All rights reserved.

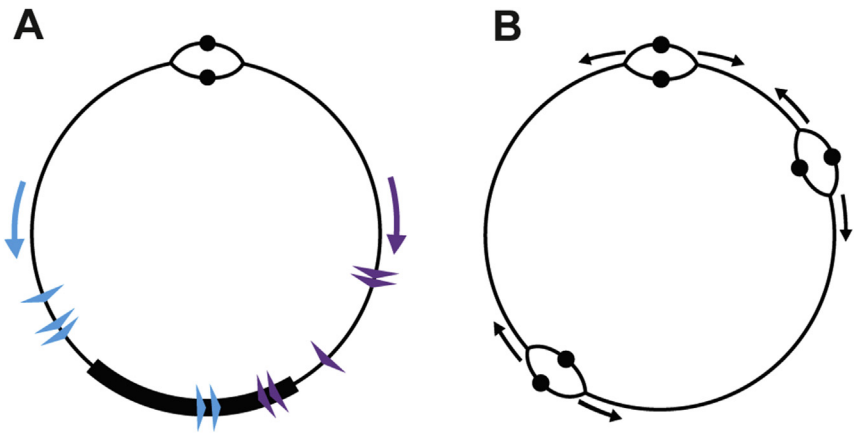


Fig. 1. Distribution of replication origins. Replication bubbles correspond to replication origin regions, with replication origins symbolized by black dots. Arrows represent replication fork direction on the chromosome. (A) in *Escherichia coli* replication is initiated from one replication origin and terminates in the terminus region represented by a black rectangle opposite the origin. The replication fork progressing clockwise (purple) is blocked by *ter* sites represented by blue triangles while the replication fork progressing counter-clockwise (blue) is blocked by *ter* sites represented by purple triangles. (B) the three replication origins found in the WT *Haloferax volcanii* strain DS2 are represented.

chromosomes, the archaeal chromosome also consists of circular DNA molecules. Experimental studies on replication initiation in the euryarchaea *Pyrococcus abyssi* [13] and *Archaeoglobus fulgibus* [14] reported the use of a single replication origin in these organisms. However, not all archaea initiate replication from a single origin. Several studies have shown that organisms from both euryarchaea and crenarchaea phyla often have multiple replication origins [15–21], revealing that these archaeal species have a circular bacterial-like chromosome with multiple replication origins (Fig. 1B). Thus understanding how replication is regulated to allow a single circular DNA molecule to be replicated from multiple origins, and how archaea deal with replication forks arrests, is a very interesting and challenging topic with potential to reveal evolutionary conserved and original features of archaeal DNA replication. Such knowledge will also shed light on the evolutionary history of replication restart mechanisms and its complex machinery that we are just starting to unravel in eukaryotes.

2. *Haloferax volcanii* as a model organism to study replication restart

H. volcanii is a halophilic euryarchaea first isolated in the Dead Sea [22]. It is a mesophile that grows aerobically at an optimal temperature of 45 °C in various defined media. Moreover, both *in vivo* studies based on well-developed genetic tools, as well as *in vitro* studies adapted to halophilic environments can be performed in this organism [23–26]. All these features make *H. volcanii* a relevant model organism to study archaeal cell biology.

The genome of the *H. volcanii* wild-type DS2 strain consists of a main circular chromosome of 2.85 megabases, as well as of four plasmids of 85 (pHV1), 6.4 (pHV2), 438 (pHV3) and 636 kilobases (pHV4) [27]. It has been shown that this organism has in average 18 copies of its genome during exponential growth [28]. This polyploidy does not hinder genetic manipulation of *H. volcanii*, as a highly efficient gene conversion mechanism allows isolation of genetically "homogenous" strains carrying only mutant alleles [29]. Multiple replication origins are present on the main chromosome of *H. volcanii* (Fig. 1B). Two of these replication origins were first described by genetic and physical mapping by Norais and colleagues [30]. Later marker frequency studies using next-generation sequencing identified additional replication origins [15,18]. Hawkins and colleagues have shown that *H. volcanii* strain H26, currently used in our laboratory, carries up to four replication origins on the main chromosome [15]. They also revealed a very striking feature of *H. volcanii* cells: this species is viable even after removal of the all identified replication origins [15]. Not only these "originless" cells can be isolated but they even have an increased fitness (exemplified by doubling time and competition experiments) under laboratory growth conditions when compared to wild type cells. Different situation is found in the closely related archaea *Haloarcula hispanica* by Wu and colleagues [31]. In this organism the two replication origins from the main replicon could be deleted separately, but the attempt to delete both replication origins at the same time was unsuccessful, suggesting that at least one of the identified replication origin is essential for viability in *H. hispanica*. How *H. volcanii* cells replicate their genome in the absence of any well-defined replication origin remains a puzzling question [32].

Replication in absence of mapped replication origins could rely on the activation of dormant origins [33]. Although no major dormant origins could be identified in *H. volcanii*, activation of dormant origins could occur randomly in cells, explaining why detection of alternative replication origins at the level of a cell population has not been possible. Another possibility is that replication is initiated from recombination intermediates randomly on the chromosome. In that scenario viability is expected to rely on recombination proteins such as the RadA recombinase (archaeal RecA orthologue). Indeed, RadA appears essential for cell viability in the absence of replication origins, suggesting that recombination is required for DNA replication initiation under these conditions. However, the deletion of *radA* alone in the wild type background markedly impairs *H. volcanii* growth [34], necessitating further studies on a possible roles of recombination in replication initiation. We propose that one interesting protein to study in this context is the helicase/nuclease Hef belonging to the XPF/MUS81/FANCM family of endonucleases.

3. Biochemical characterization of archaeal Hef/XPF

Proteins belonging to the XPF/MUS81/FANCM endonuclease family act on 3′-flap DNA structures that are formed during DNA repair or replication restart. They are found throughout eukarya and archaea, but to date have not been identified in bacteria. Eukaryotes have several XPF/MUS81/FANCM family members. They

assemble into heterodimeric complexes to form distinct active complexes involved in DNA repair, meiotic recombination and replication restart [35–43].

All archaea encode a protein of the XPF/MUS81/FANCM family of endonucleases. It exists in two forms. The long form, referred as Hef, consists of an N-terminal helicase fused to a C-terminal nuclease and is specific to euryarchaea. The short form, referred as XPF, lacks the helicase domain and is specific to crenarchaea and thaumarchaea (Fig. 2A). The long-form Hef protein was first identified in *Pyrococcus furiosus* for its activity on branched DNA structures, Hef meaning helicase-associated endonuclease fork-structure DNA [44]. Hef has the unique feature among XPF/MUS81/FANCM proteins of having both active helicase and nuclease domains. This domain configuration is highly similar with the human FANCM, a possible orthologue of archaeal Hef proteins [45,46].

The C-terminal region of Hef consists of a nuclease domain and a helix-hairpin-helix (HhH) domain as observed in archaeal XPF proteins. *In vitro* experiments on *Pfu*Hef have shown that a variety of branched DNA structures carrying single-strand DNA (ssDNA) regions, such as flapped and fork-structured DNAs (Fig. 2B), were recognized and cleaved by the C-terminal nuclease domain [47,48]. In the N-terminal domain of Hef two conserved helicase motifs from Super-Family 2 (SF2) helicases are found. They are separated by a third domain that shares structure similarity with the "Thumb" domain of polymerases involved in dsDNA binding [48]. Based on *in vitro* experiments on *Pfu*Hef it was proposed that the helicase and the nuclease domains of the entire *Pfu*Hef protein act synergistically, allowing the helicase domain to bind and process the fork-structured DNA to form a four-way structure that is then cleaved by the endonuclease domain [49]. Thus euryarchaeal Hef was proposed to be involved in the resolution of stalled replication forks, as suggested for Mus81 complexes in eukaryotes.

Consistent with a role of Hef in replication and/or repair, *Thermococcus kodakarensis* Hef has recently been shown to interact with *Tko*PCNA1 [50]. In the cell, the replication factor PCNA (Proliferating Cell Nuclear Antigen) encircles double-strand DNA (dsDNA) and enhances *e.g.* the processivity of the replicative polymerase. PCNA is a central protein of DNA replication and repair as it interacts with a large number of proteins. In *T. kodakarensis*, two PCNAs are found, but only PCNA1 is essential for viability in this organism [51]. Interaction with PCNA often involves a conserved motif known as PCNA-Interacting Protein (PIP) motif. Indeed *in vitro* studies from the crenarchaea *Sulfulobus solfataricus* have shown that the interaction of *Sso*XPF with *Sso*PCNA involves a conserved PIP motif [52,53] and is required for nuclease activity of this "short" XPF [53]. In the case of *Tko*Hef, the interaction with PCNA1 does not involve a canonical PIP motif [52], but a disordered region of Hef between the helicase and nuclease domains. Interestingly, these predicted long disordered regions connecting two catalytic domains are a common feature of euryarchaeal Hef and eukaryotic FANCM proteins [50]. So, both creanarchaeal XPF and euryarchaeal Hef proteins can interact with PCNA and display biochemical activities consistent with XPF/Hef being involved in DNA repair and/or replication restart.

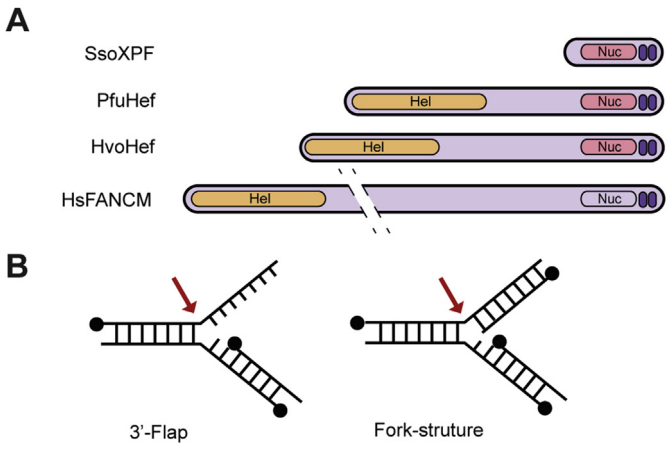


Fig. 2. Schematic representation of archaeal Hef protein and representative members of XPF/MUS81/FANCM family. (A) Schematic representation of crenarchaeal XPF from *Sulfulobus solfataricus (Sso)*, euryarchaeal Hef from *Pyrococcus furiosus* (*Pfu*) and *Haloferax volcanii* (*Hvo*), and human FANCM. Yellow-filled regions represent active helicase domains, pink-filled region represent active nuclease domains, and dark purple-filled ovals represent active HhH domains. (B) Substrate specificity of *Pfu*Hef. A red arrow indicates the approximate sites of cleavage within the DNA structure. Black circle indicate 5′ termini.

4. Genetic experiments to understand Hef function in *H. volcanii*

Next we will describe genetic studies that have been performed to understand the role of XPF/Hef proteins in archaeal cells.

The *hef* gene has been deleted in two different archaea: the hyperthermophile *T. kodakarensis* and in the halophile *H. volcanii*. In both organisms, Hef was non-essential for cell viability under normal laboratory growth conditions. An increased sensitivity to a variety of DNA damaging agents was observed in *T. kodakarensis* Δ*hef* cells [54], notably an increased sensitivity to UV radiation. DNA lesions caused by UV radiation are mainly repaired by the Nucleotide Excision Repair (NER) pathway. The existence of an archaeal NER pathway has not been established, although most archaea have eukaryal-type NER genes. As most of the eukaryal NER proteins have multiple cellular functions, their presence in archaea is insufficient for reaching a conclusion that a functional NER pathway is present in archaea [2,55]. Nevertheless, the sensitivity of Δ*hef* cells to UV radiation suggests that Hef is involved in Nucleotide Excision Repair (NER) in *T. kodakarensis*, as its human orthologue XPF-ERCC1 and its counterpart RAD1-RAD10 in the yeast *Saccharomyces cerevisae* [43,56]. This observation also suggests that an active NER pathway must exist in this archaea. Clearly, additional experiments are still needed to better understand the possible role of *Tko*Hef in NER and further dissect the pathway responsible for archaeal Nucleotide Excision Repair.

Phenotype of *H. volcanii* Δ*hef* strain is different from what was observed for *Thermococcus kodakarensis*. For instance, deletion of *hef* in *H. volcanii* cells does not bring about sensitivity to various DNA damaging agents or does not modulate recombination frequency. We could only observe a slow-growth phenotype of the Δ*hef* colonies when chronically exposed to mitomycin C (MMC) on plate [57]. Note that *H. volcanii* also possesses bacterial-like NER that were shown to be responsible for the repair of UV lesions [57]. This may explain why Hef in *H. volcanii* is not apparently involved in NER.

Further studies on the role of Hef in *H. volcanii* revealed that this protein is essential for viability in the absence of the Holliday junctions (HJs) resolvase Hjc. Holliday junctions are four-way branched DNA structures formed during homologous recombination strand exchange and recombination-dependent replication restart. Although Hjc is conserved throughout archaea, the *hjc* gene could be deleted in *H. volcanii* and *T. Kodakarensis* cells without affecting cell viability [54,57]. Hef and Hjc were first thought to act as redundant HJs resolvases. However, homologous recombination could be successfully inactivated (by deletion of the *radA* gene) in Δ*hjc* cells but not in Δ*hef* cells [57], revealing the distinct functional roles for Hjc and Hef. The key observation of this study was that Hef

is specifically required for cell viability in absence of recombination.

Which alternative pathways or additional functions could *i)* depend on Hef, *ii)* be essential for cell viability during normal growth condition, and *iii)* implicate recombination proteins and/or Hef? Clearly these observations are consistent with a role of Hef in replication restart, as suggested by earlier *in vitro* experiments. To further understand the cellular role of Hef proteins, we decided to investigate dynamic localization of fluorescently-labelled Hef proteins in living *H. volcanii* cells.

5. Strategy for using Green Fluorescent Protein (GFP) in living *H. volcanii* cells

Fluorescent Proteins (FP) are frequently used in a wide range of living cells/organisms to study protein localization, mobility, turnover, interactions, and much more [58–61]. Such approaches can reveal key features of proteins *in vivo* to complete our understanding of pathways, as illustrated for NER pathway in mammalian cells [62]. This is why we have constructed GFP-fusions of Hef and developed tools to investigate its behaviour in *H. volcanii* living cells. The fluorescence of several GFP variants had previously been tested in the halophilic cytoplasm of *H. volcanii* cells, revealing that the smRS-GFP variant could be successfully used [63]. This variant that we have used throughout our studies (referred to as GFP) has mutations increasing solubility (Phe99Ser, Met153Thr and Val63Ala) as well as a mutation in the chromophore (Ser65Thr) that redshifts the absorption maximum to 488 nm without changing the emission maximum at 509 nm. It was fused to the C-terminal end of the protein and the GFP-fused Hef protein expressed from the native chromosomal locus. To do so the *gfp* gene was inserted directly at the end of *hef* gene on a plasmid containing the whole *hef* gene flanked by up-stream and down-stream regions used to

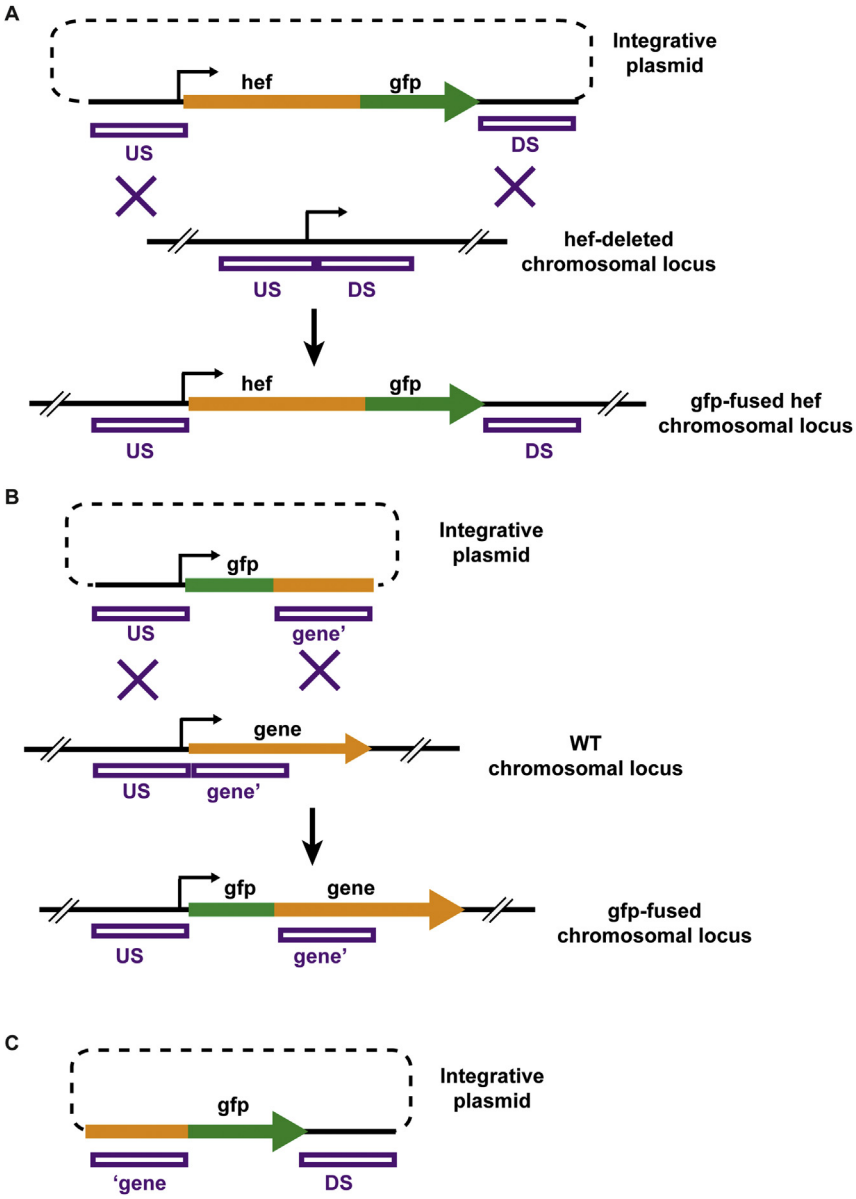


Fig. 3. Pop-In/pop-Out strategies to express a GFP-fused protein from its native chromosomal locus. Homologous regions between the integrative plasmid and the chromosomal locus are represented by purple rectangles. The two recombination events taking place during pop-in/pop-out are symbolized by purple crosses. (A) Strategy used to replace the *hef*-deleted chromosomal locus by a locus allowing the expression of a GFP-fused Hef protein from *hef* native promoter. (B) Strategy used to obtain a N-terminal GFP-fused protein essential for viability. (C) Strategy used to obtain a C-terminal GFP-fused protein essential for viability. Only the integrative plasmid is represented.

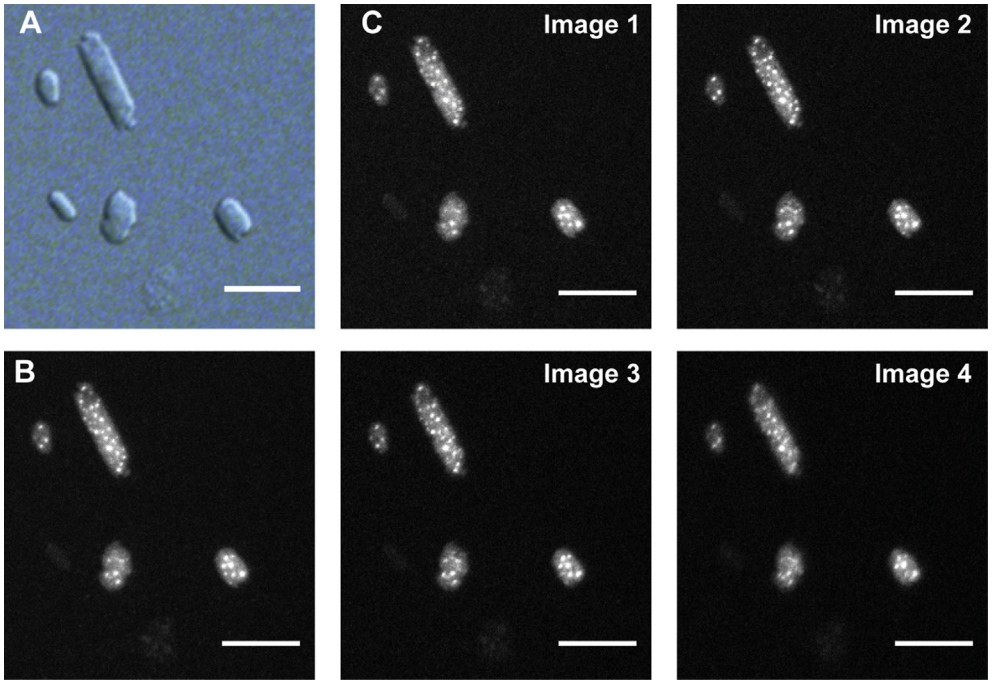


Fig. 4. Representative images of *hef*::*gfp* cells in response to aphidicolin exposure. (A) Picture of DIC signal. (B) Picture of GFP signal obtained from maximum intensity z-projection of the four images shown in panel C. (C) Pictures of GFP signal acquired at different focal planes on the z-axis. Bar equals 10 μm.

construct Δ*hef* mutant by pop-in/pop-out strategy. Using the same strategy, we replaced the *hef*-deleted chromosomal locus by a locus allowing the expression of a GFP-fused Hef protein from *hef* native promoter (Fig. 3A) [64]. We stress that this strategy cannot be applied for essential gene, as only the *gfp* gene would be inserted at the chromosomal locus. In this case, a translational in-frame fusion

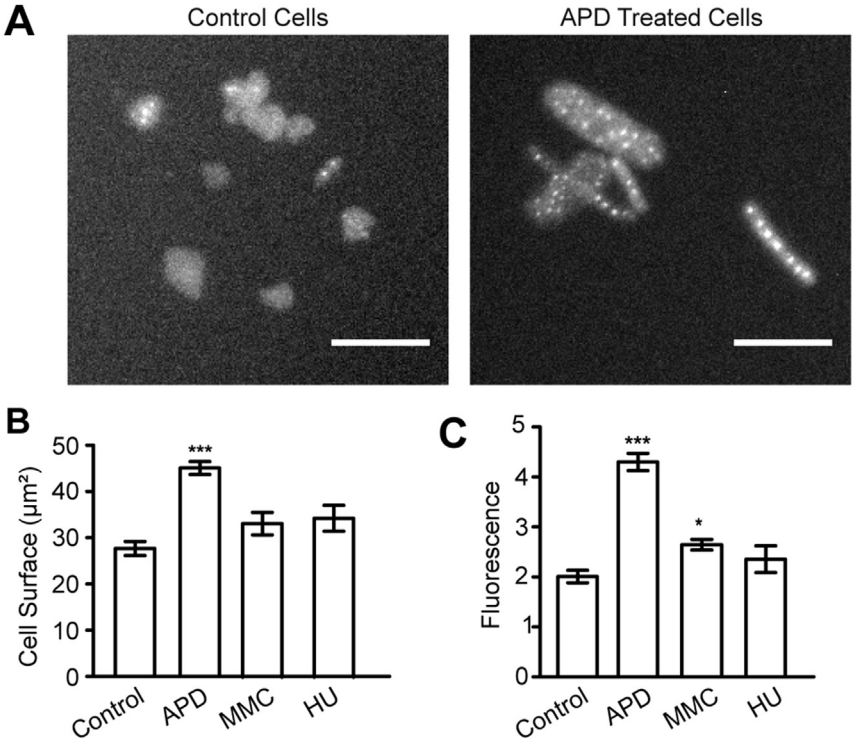


Fig. 5. *In vivo* localization of GFP-fused Hef protein. (A) GFP signal of *hef*::*gfp* cells under control conditions and after exposure to 5 μg/ml aphidicolin. Bar equals 10 μm. (B) Average cell surface of *hef*::*gfp* cells under control conditions and after exposure to 5 μg/ml aphidicolin, 0.075 μg/ml mitomycin or 5 mM hydroxyurea. (C) Mean number of fluorescence foci of *hef*::*gfp* cells under control conditions and after exposure to 5 μg/ml aphidicolin, 0.075 μg/ml mitomycin or 5 mM hydroxyurea. All errors bars represent SD. *t*-test are performed in comparison to control conditions. *Significantly different, $P < 0.05$; ***Significantly different, $P < 0.001$.

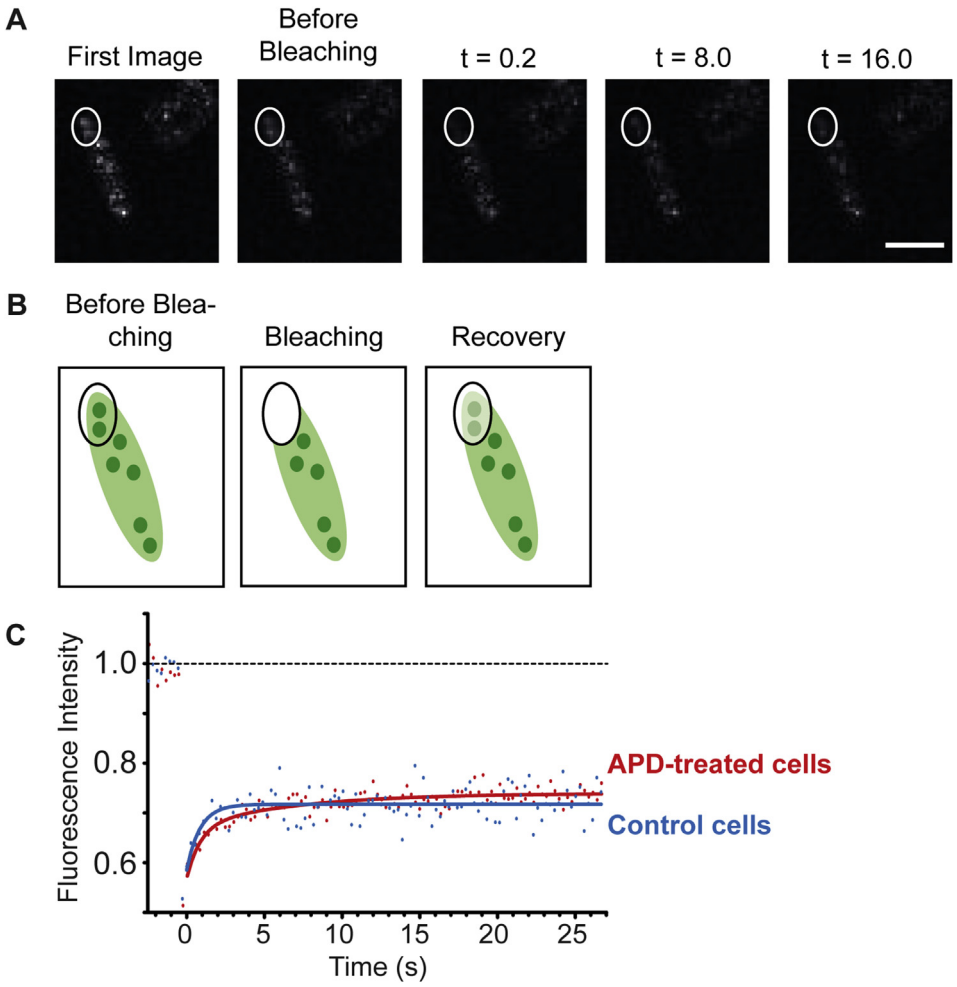


Fig. 6. Diffusion dynamic of GFP-fused Hef proteins by FRAP experiments. (A) Images of a representative cell after exposure to 5 μg/ml aphidicolin for FRAP analysis. Bar equals 5 μm. (B) Schematic representation of the cell in panel A. (C) Averaged fluorescence recovery curves for control cells (blue) and APD-treated cells (red).

of the gene/protein of interest must be constructed. For N-terminal fusion the *gfp* gene is flanked by up-stream region of the gene and the first 300–400 bp of the gene (Fig. 3B). For C-terminal fusion the *gfp* gene is flanked by the last 300–400 bp of the gene and its down-stream region (Fig. 3C). Such strategies have been applied to study N-terminal or C-terminal GFP-fused PCNA but we have not been able to obtain viable cells expressing GFP-fused PCNA (data not shown).

We next established that the obtained fusion protein was functional. A growth defect of Δ*hef* cells on plates containing the DNA damaging agent mitomycine C (MMC) was previously observed [57]. No growth delay was observed for *hef*::*gfp* cells on MMC plates, indicating functional complementation by the Hef::GFP construct. In agreement with this notion, we successfully deleted *hjc* in cells expressing Hef::GFP. Note that a growth defect was observed for *hef*::*gfp* Δ*hjc* cells, no longer measurable in cells expressing Hef fused by its N-terminal domain to the GFP *via* a linker (unpublished data). To test whether Hef could be implicated in replication restart, we increased replication fork arrest by exposing cells to aphidicolin (APD) as this DNA polymerase inhibitor had been shown to inhibit DNA synthesis in halophilic archaea [65]. A decrease in cell viability was observed upon exposure of *hef*-deleted cells to increasing concentration of APD. Such decrease in cell viability was not observed with cells expressing GFP-fused Hef [64]. From these results we deduced that APD treatment revealed a role of Hef upon replication arrest that could be investigated using the functional GFP-fused Hef.

6. Hef::GFP molecules are recruited at arrested replication forks

To observe the localisation of Hef::GFP proteins by fluorescence microscopy, a large number of individual agarose-embedded cells were imaged using a wide-field microscope. Differential Interference Contrast (DIC) [also known as Nomarski Interference Contrast (NIC)], was first used to visualized the cells, followed by fluorescence imaging (Fig. 4). Because not all cells were in the same focal plane, fluorescence images were acquired at different focal planes on the z-axis. Consecutive slices of cells in focus were then selected and used to perform a maximum intensity z-projection. At each pixel, the highest fluorescence signal was kept. In this way we obtained a two-dimensional picture where the maximal fluorescence signals from different focal planes were recorded. Resulting images containing hundreds of cells were then analysed by quantitative image analyses using IMARIS software. To avoid user-bias several parameters were optimized to detect cells and fluorescence foci within cells using automatic thresholds. This approach allowed thousands of cells to be analysed in each condition tested, providing extremely high statistical power for these measurements.

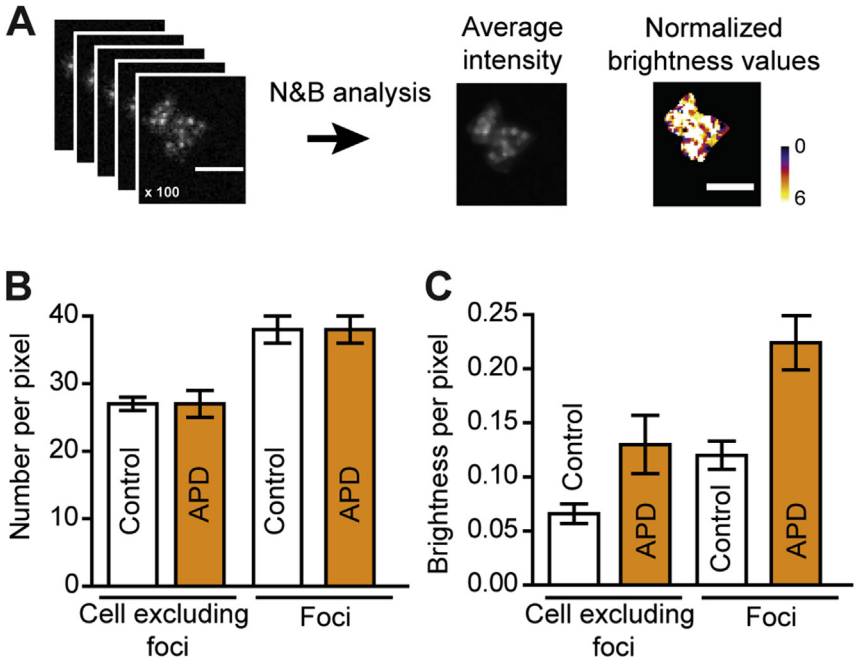


Fig. 7. Number and brightness analysis of diffusing Hef::GFP proteins. (A) Images of a representative cell after exposure to 5 μg/ml aphidicolin for N&B analysis. Bar equals 5 μm. (B) Average number of diffusing Hef::GFP proteins per pixel (±SEM) after N&B analysis. (C) Average brightness of diffusing Hef::GFP proteins per pixel (±SEM) after N&B analysis.

These results revealed that Hef::GFP proteins formed fluorescence foci even under normal growth conditions, in the absence of any DNA damaging agents. This observation demonstrated that replication arrests may occur naturally in actively growing cells, during the normal replication process. APD exposure resulted in a significant increase in cell size (from 28 to 45 μm^2) and in the average number of fluorescence foci (from two to four foci per cell). Exposing the cells to mitomycin or hydroxyurea, two DNA damaging agents, had no effect on both cell size and number of *hef*::*gfp* foci per cell. Thus the increased cell size and number of foci were specific to APD treatment (Fig. 5B and C), further indicating that *Hvo*Hef is recruited at arrested replication forks brought about by addition of aphidicolin.

7. Dynamic localization of Hef::GFP molecules at arrested replication forks

To further understand the physiological role of Hef at fluorescence foci, we aimed to investigate its diffusion dynamics in living cells. Fluorescence Correlation Spectroscopy technique could not been applied to *hef*::*gfp* cells because the physiological fluorescence signal is too low. However, we successfully analysed *hef*::*gfp* cells using two additional techniques that allow accessing the *in vivo* dynamics of fluorescent particles. These Fluorescence Recovery After Photobleaching (FRAP) and Number and Brightness (N&B) experiments were performed using a confocal microscope on cells immobilized on a poly-D-lysine coated cover-slip. Both of these techniques reveal behaviour of diffusing particles in living cells. In FRAP experiments, a region of the cells is photobleached using a 488 nm CW Argon laser at maximum intensity, and the speed of fluorescence recovery in that region is then measured. The obtained recovery reflects the diffusion rate of fluorescent molecules arriving from the non-photobleached part of the cell. Due to the small size of *H. volcanii* cells, in order to limit the size of photobleached area, the bleached region overlapped only partially with a peripheric region of the cell (Fig. 6A and B). Doing so, in control cells (no aphidicolin) we have measured a recovery of fluorescence in the photobleached region reflecting one major population of Hef::GFP diffusing molecules (Fig. 6C). The apparent two-dimensional diffusion rate of Hef::GFP was estimated at 0.8–2.3 μm^2 per second. This appeared markedly lower than expected for Hef dimer according to diffusion rate of fused-GFP proteins in *E. coli* cytoplasm [66]. Indeed analytical ultracentrifugation experiments performed on purified *Hvo*Hef had shown that *Hvo*Hef formed dimers in solution with a peculiar elongated shape [64]. Such non-globular structure may explain this limited diffusion, as well as physical constraints of the cytosol that could result from high DNA and salt concentration, and/or transient interactions with cellular components (DNA or proteins complexes). We next addressed whether diffusing Hef::GFP molecules could be observed in fluorescence foci. The same FRAP experiments were performed on cells exposed to aphidicolin, including foci in the photobleached region. In that case we also measured a recovery of fluorescence, reflecting the diffusion outside foci of photobleached-Hef::GFP molecules allowing the recruitment of fluorescent Hef::GFP molecules from outside the photobleached region. Interestingly, in addition to the previously observed slow-diffusing population, an even more slowly-diffusing population clearly induced in cells after APD exposure was observed (Fig. 6C).

These changes in the diffusion pattern of Hef::GFP molecules upon replication arrests were also independently observed using Number & Brightness technique (Fig. 7A). 100 images were collected at time intervals of 1.95 s per frame for N&B analysis that relies on fluctuation of fluorescence intensity in each analysed pixel [67] to determine the number of diffusing molecules and their brightness that reflects their oligomeric state. Thus, N&B analyses have potential to reveal changes in the oligomeric state of diffusing fluorescent molecules. When applied to diffusing Hef::GFP molecules, only changes in the oligomeric state (i.e. higher brightness) were observed upon APD treatment (Fig. 7C), as the number of diffusing Hef::GFP proteins per pixel was not increased (Fig. 7B). Those results indicated that Hef:GFP proteins at foci either have a higher oligomeric state, or that several molecules are co-diffusing. Both hypotheses are compatible with a slower-diffusing population

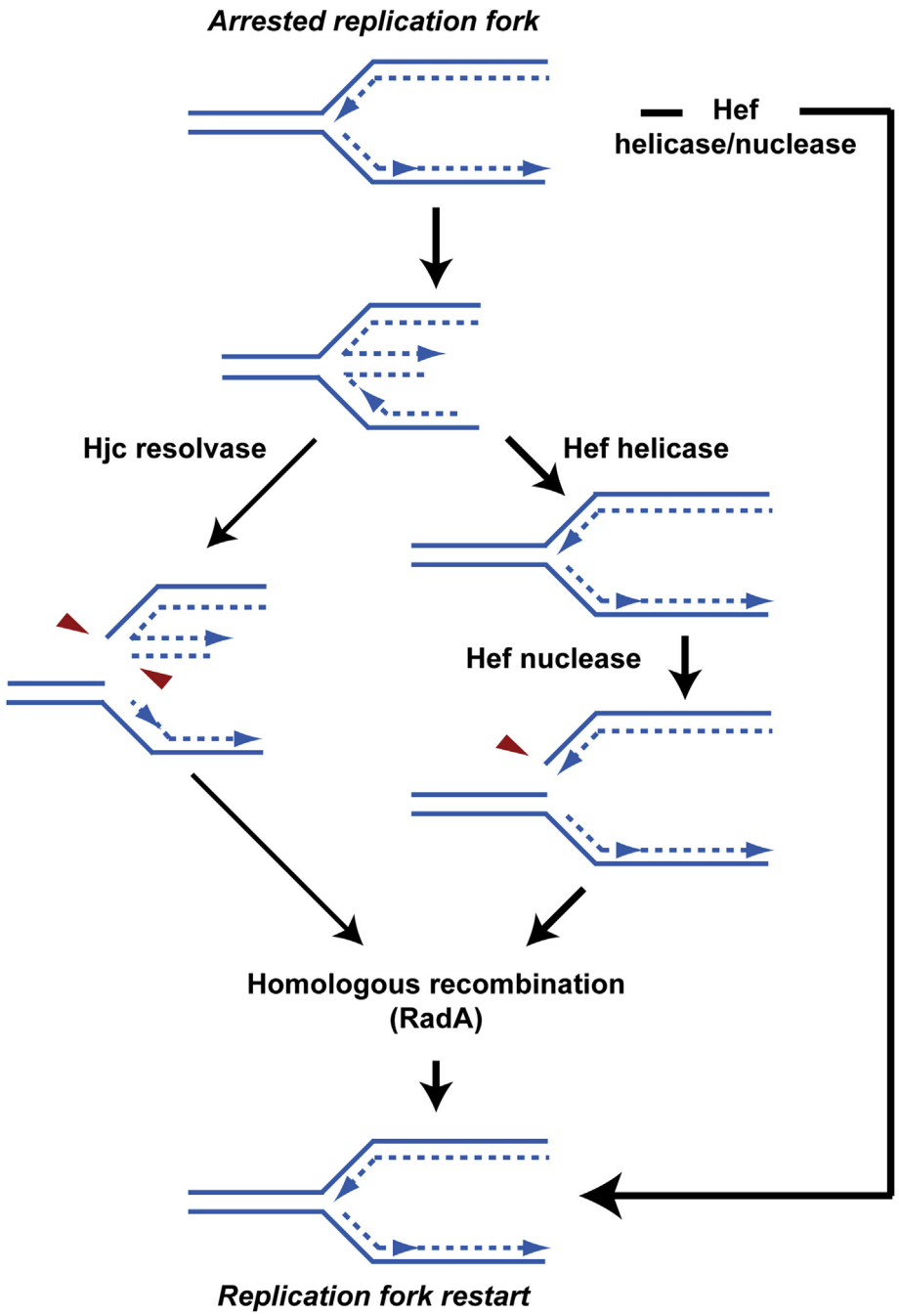


Fig. 8. Working model for replication restart in *Haloferax volcanii*. Three pathways allowing replication restart are proposed. The first pathway (left panels) involves Hjc resolvase and the homologous recombination protein RadA. The second pathway (right panels) involves both helicase and nuclease activities of Hef and the homologous recombination protein RadA. A third pathway involving Hef is proposed, independent of homologous recombination. We have symbolized the dominant role of Hef during replication restart using wider arrows.

revealed by FRAP in response to APD treatment.

Altogether, studies of the diffusion dynamics of Hef::GFP proteins in the cell has allowed us to conclude that the observed fluorescence foci correspond to arrested replication forks at which Hef proteins are recruited. Whether slow-diffusing Hef::GFP molecules reflect the recruitment of Hef directly on DNA and/or as part of protein complexes at arrested replication forks remains to be investigated.

8. What is the role of Hef at arrested replication fork?

We had previously shown that both the helicase and nuclease activities of Hef are required for functioning of this protein [57], strongly indicating that Hef is directly implicated in processing arrested replication forks. Can we further understand its role? Because *hjc* could be deleted in cells expressing Hef::GFP proteins, we investigate whether inactivating the alternative pathway for replication restart by deleting *hjc* had an effect on Hef::GFP localisation. Surprisingly, the absence of Hjc was shown to affect neither the cell size nor the number of Hef::GFP foci, even in cells exposed to APD [64]. As mentioned previously Δ*hjc hef*::*gfp* display a growth defect no longer observed in Δ*hjc* cells expressing Hef fused to the GFP by its N-terminal extremity. We have reproduced those experiments in Δ*hjc gfp*::*hef* cells and observed the same results

(unpublished data). So the absence of the alternative pathway involving Hjc (and RadA) does not increase the need for Hef at arrested replication forks. This raises the possibility that *Hvo*Hef may prevent access of recombination proteins to arrested replication forks. This hypothesis is compatible with the recent studies that have indicated that eukaryotic FANCM proteins can prevent homologous recombination [68–71]. Similar role had also been described in *E. coli* for the UvrD helicase that can act at arrested replication forks to prevent access of recombination protein [72]. We propose a model for replication restart in *H. volcanii* that accounts both *in vitro* and *in vivo* data on Hef protein (Fig. 8). One pathway would involve the processing of the arrested replication fork into a four-way branched DNA structure that can either be cleaved by HJC resolvase, or processed back to a three-way structure and then cleaved by Hef helicase/nuclease activities as proposed based on *in vitro* studies [49]. In any case, both pathways must rely on homologous recombination to repair the DNA, thus allowing replication restart from recombination intermediates. As we had shown that Hef can act independently of homologous recombination, a third pathway involving Hef but avoiding the need for homologous recombination is proposed. In this pathway Hef could only be involved in preventing access of recombination proteins and/or could process the DNA. Whether its nuclease and/or helicase activities are required in this pathway could be addressed by deleting *radA* in cells expressing a nuclease-deficient or a helicase-deficient Hef protein. Indeed future experimental work is needed to test this new working hypothesis, unravelling the specific action of Hef at arrested replication forks and pathways allowing replication restart in *H. volcanii*.

9. What have we learnt on replication (restart) in *H. volcanii*

In addition to what we have learnt on Hef, this work has also shown that replication forks arrests occur spontaneously in *H. volcanii* cells. It is currently unclear whether the role of Hef in DNA replication process could be more direct than suggested by previous biochemical studies. We have observed up to 15 to 20 arrested replication forks in some individual cells. As a typical *H. volcanii* cell contains 18 genome copies and that each molecule can be replicated by up to 8 replication forks, our imaging studies rather suggest that several copies of the genome may be replicated in one cell. Whether all individual genome copies are replicated simultaneously remains unknown; this striking question needs to be addressed in future work. Clearly, *H. volcanii* provides a unique model for studying replication dynamics in living cells in the absence and presence of replication origins using a combination of genetic and imaging techniques.

10. Concluding remarks

We believe that studies of FP-fused replication proteins provide a crucial tool to unravel proteins function *in vivo* in archaeal cells. Indeed, in a very elegant work, Duggin and colleagues studied tubulin-like proteins using mCherry-fused proteins in *H. volcanii* cells, further increasing the possibilities offered from studies of FP-fused proteins [73]. As archaea possess hallmarks of both bacterial and eukaryotic replication systems we believe that such fundamental knowledge will shed light on the evolutionary history of replication restart mechanisms and its complex machinery that we are just starting to unravel in eukaryotes and now archaea.

Conflict of interest

There is no conflict of interest.

Acknowledgements

We want to acknowledge the Agence Nationale de la Recherche for funding (ARCREP and RETIDYNA grants).

References

[1] E.R. Barry, S.D. Bell, DNA replication in the archaea, Microbiol. Mol. Biol. Rev. 70 (2006) 876–887.
[2] Z. Kelman, M.F. White, Archaeal DNA replication and repair, Curr. Opin. Microbiol. 8 (2005) 669–676.
[3] M. O'Donnell, L. Langston, B. Stillman, Principles and concepts of DNA replication in bacteria, archaea, and eukarya, Cold Spring Harb. Perspect. Biol. 5 (2013).
[4] G. Grompone, D. Ehrlich, B. Michel, Cells defective for replication restart undergo replication fork reversal, EMBO Rep. 5 (2004) 607–612.
[5] B. Michel, H. Boubakri, Z. Baharoglu, M. LeMasson, R. Lestini, Recombination proteins and rescue of arrested replication forks, DNA Repair (Amst.) 6 (2007) 967–980.
[6] C.J. Rudolph, A.L. Upton, R.G. Lloyd, Maintaining replication fork integrity in UV-irradiated Escherichia coli cells, DNA Repair (Amst.) 7 (2008) 1589–1602.
[7] L. Xu, K.J. Marians, PriA mediates DNA replication pathway choice at recombination intermediates, Mol. Cell 11 (2003) 817–826.
[8] A.M. Carr, S. Lambert, Replication stress-induced genome instability: the dark side of replication maintenance by homologous recombination, J. Mol. Biol. 425 (2013) 4733–4744.
[9] C. Lundin, N. Schultz, C. Arnaudeau, A. Mohindra, L.T. Hansen, T. Helleday, RAD51 is involved in repair of damage associated with DNA replication in mammalian cells, J. Mol. Biol. 328 (2003) 521–535.
[10] L. Roseaulin, Y. Yamada, Y. Tsutsui, P. Russell, H. Iwasaki, B. Arcangioli, Mus81 is essential for sister chromatid recombination at broken replication forks, EMBO J. 27 (2008) 1378–1387.
[11] Y. Saintigny, F. Delacote, G. Vares, F. Petitot, S. Lambert, D. Averbeck, B.S. Lopez, Characterization of homologous recombination induced by replication inhibition in mammalian cells, EMBO J. 20 (2001) 3861–3870.
[12] J.T. Yeeles, J. Poli, K.J. Marians, P. Pasero, Rescuing stalled or damaged replication forks, Cold Spring Harb. Perspect. Biol. 5 (2013) a012815.
[13] H. Myllykallio, P. Lopez, P. Lopez-Garcia, R. Heilig, W. Saurin, Y. Zivanovic, H. Philippe, P. Forterre, Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon, Science 288 (2000) 2212–2215.
[14] S. Maisnier-Patin, L. Malandrin, N.K. Birkeland, R. Bernander, Chromosome replication patterns in the hyperthermophilic euryarchaea Archaeoglobus fulgidus and Methanocaldococcus (Methanococcus) jannaschii, Mol. Microbiol. 45 (2002) 1443–1450.
[15] M. Hawkins, S. Malla, M.J. Blythe, C.A. Nieduszynski, T. Allers, Accelerated growth in the absence of DNA replication origins, Nature 503 (2013) 544–547.
[16] M. Lundgren, A. Andersson, L. Chen, P. Nilsson, R. Bernander, Three replication origins in sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 7046–7051.
[17] E.A. Pelve, A.C. Lindas, A. Knoppel, A. Mira, R. Bernander, Four chromosome replication origins in the archaeon Pyrobaculum calidifontis, Mol. Microbiol. 85 (2012) 986–995.
[18] E.A. Pelve, W. Martens-Habbena, D.A. Stahl, R. Bernander, Mapping of active replication origins in vivo in thaum- and euryarchaeal replicons, Mol. Microbiol. 90 (2013) 538–550.
[19] N.P. Robinson, S.D. Bell, Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 5806–5811.
[20] R.Y. Samson, Y. Xu, C. Gadelha, T.A. Stone, J.N. Faqiri, D. Li, N. Qin, F. Pu, Y.X. Liang, Q. She, S.D. Bell, Specificity and function of archaeal DNA replication initiator proteins, Cell Rep. 3 (2013) 485–496.
[21] Z. Wu, H. Liu, J. Liu, X. Liu, H. Xiang, Diversity and evolution of multiple orc/cdc6-adjacent replication origins in haloarchaea, BMC Genomics 13 (2012) 478.
[22] M.F. Mullakhanbhai, H. Larsen, Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement, Arch. Microbiol. 104 (1975) 207–214.
[23] T. Allers, S. Barak, S. Liddell, K. Wardell, M. Mevarech, Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in Haloferax volcanii, Appl. Environ. Microbiol. 76 (2010) 1759–1769.
[24] T. Allers, H.P. Ngo, M. Mevarech, R.G. Lloyd, Development of additional selectable markers for the halophilic archaeon Haloferax volcanii based on the leuB and trpA genes, Appl. Environ. Microbiol. 70 (2004) 943–953.
[25] G. Bitan-Banin, R. Ortenberg, M. Mevarech, Development of a gene knockout system for the halophilic archaeon Haloferax volcanii by use of the pyrE gene, J. Bacteriol. 185 (2003) 772–778.
[26] A. Large, C. Stamme, C. Lange, Z. Duan, T. Allers, J. Soppa, P.A. Lund, Characterization of a tightly controlled promoter of the halophilic archaeon Haloferax volcanii and its use in the analysis of the essential cct1 gene, Mol. Microbiol. 66 (2007) 1092–1106.
[27] A.L. Hartman, C. Norais, J.H. Badger, S. Delmas, S. Haldenby, R. Madupu,

J. Robinson, H. Khouri, Q. Ren, T.M. Lowe, J. Maupin-Furlow, M. Pohlschroder, C. Daniels, F. Pfeiffer, T. Allers, J.A. Eisen, The complete genome sequence of Haloferax volcanii DS2, a model archaeon, PLoS One 5 (2010) e9605.
[28] S. Breuert, T. Allers, G. Spohn, J. Soppa, Regulated polyploidy in halophilic archaea, PLoS One 1 (2006) e92.
[29] C. Lange, K. Zerulla, S. Breuert, J. Soppa, Gene conversion results in the equalization of genome copies in the polyploid haloarchaeon Haloferax volcanii, Mol. Microbiol. 80 (2011) 666–677.
[30] C. Norais, M. Hawkins, A.L. Hartman, J.A. Eisen, H. Myllykallio, T. Allers, Genetic and physical mapping of DNA replication origins in Haloferax volcanii, PLoS Genet. 3 (2007) e77.
[31] Z. Wu, J. Liu, H. Yang, H. Liu, H. Xiang, Multiple replication origins with diverse control mechanisms in Haloarcula hispanica, Nucleic Acids Res. 42 (2014) 2282–2294.
[32] B. Michel, R. Bernander, Chromosome replication origins: do we really need them? Bioessays 36 (2014) 585–590.
[33] C. Cvetic, J.C. Walter, Eukaryotic origins of DNA replication: could you please be more specific? Semin. Cell Dev. Biol. 16 (2005) 343–353.
[34] W.G. Woods, M.L. Dyall-Smith, Construction and analysis of a recombination-deficient (radA) mutant of Haloferax volcanii, Mol. Microbiol. 23 (1997) 791–797.
[35] A. Ahmad, A.R. Robinson, A. Duensing, E. van Drunen, H.B. Beverloo, D.B. Weisberg, P. Hasty, J.H. Hoeijmakers, L.J. Niedernhofer, ERCC1-XPF endonuclease facilitates DNA double-strand break repair, Mol. Cell Biol. 28 (2008) 5082–5092.
[36] A. Ciccia, N. McDonald, S.C. West, Structural and functional relationships of the XPF/MUS81 family of proteins, Annu. Rev. Biochem. 77 (2008) 259–287.
[37] S.J. Collis, A. Ciccia, A.J. Deans, Z. Horejsi, J.S. Martin, S.L. Maslen, J.M. Skehel, S.J. Elledge, S.C. West, S.J. Boulton, FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the Fanconi anemia core complex, Mol. Cell 32 (2008) 313–324.
[38] J.H. Enzlin, O.D. Scharer, The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif, EMBO J. 21 (2002) 2045–2053.
[39] K. Gari, C. Decaillet, M. Delannoy, L. Wu, A. Constantinou, Remodeling of DNA replication structures by the branch point translocase FANCM, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 16107–16112.
[40] K. Hanada, M. Budzowska, S.L. Davies, E. van Drunen, H. Onizawa, H.B. Beverloo, A. Maas, J. Essers, I.D. Hickson, R. Kanaar, The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks, Nat. Struct. Mol. Biol. 14 (2007) 1096–1104.
[41] P. Matulova, V. Marini, R.C. Burgess, A. Sisakova, Y. Kwon, R. Rothstein, P. Sung, L. Krejci, Cooperativity of Mus81.Mms4 with Rad54 in the resolution of recombination and replication intermediates, J. Biol. Chem. 284 (2009) 7733–7745.
[42] F. Osman, M.C. Whitby, Exploring the roles of Mus81-Eme1/Mms4 at perturbed replication forks, DNA Repair (Amst.) 6 (2007) 1004–1017.
[43] A.M. Sijbers, W.L. de Laat, R.R. Ariza, M. Biggerstaff, Y.F. Wei, J.G. Moggs, K.C. Carter, B.K. Shell, E. Evans, M.C. de Jong, S. Rademakers, J. de Rooij, N.G. Jaspers, J.H. Hoeijmakers, R.D. Wood, Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease, Cell 86 (1996) 811–822.
[44] K. Komori, R. Fujikane, H. Shinagawa, Y. Ishino, Novel endonuclease in Archaea cleaving DNA with various branched structure, Genes Genet. Syst. 77 (2002) 227–241.
[45] A.R. Meetei, A.L. Medhurst, C. Ling, Y. Xue, T.R. Singh, P. Bier, J. Steltenpool, S. Stone, I. Dokal, C.G. Mathew, M. Hoatlin, H. Joenje, J.P. de Winter, W. Wang, A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M, Nat. Genet. 37 (2005) 958–963.
[46] G. Mosedale, W. Niedzwiedz, A. Alpi, F. Perrina, J.B. Pereira-Leal, M. Johnson, F. Langevin, P. Pace, K.J. Patel, The vertebrate Hef ortholog is a component of the Fanconi anemia tumor-suppressor pathway, Nat. Struct. Mol. Biol. 12 (2005) 763–771.
[47] T. Nishino, K. Komori, Y. Ishino, K. Morikawa, Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms, Structure 13 (2005) 1183–1192.
[48] T. Nishino, K. Komori, D. Tsuchiya, Y. Ishino, K. Morikawa, Crystal structure and functional implications of Pyrococcus furiosus hef helicase domain involved in branched DNA processing, Structure 13 (2005) 143–153.
[49] K. Komori, M. Hidaka, T. Horiuchi, R. Fujikane, H. Shinagawa, Y. Ishino, Cooperation of the N-terminal helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks, J. Biol. Chem. 279 (2004) 53175–53185.
[50] S. Ishino, T. Yamagami, M. Kitamura, N. Kodera, T. Mori, S. Sugiyama, T. Ando, N. Goda, T. Tenno, H. Hiroaki, Y. Ishino, Multiple interactions of the intrinsically disordered region between the helicase and nuclease domains of the archaeal Hef protein, J. Biol. Chem. 289 (2014) 21627–21639.
[51] M. Pan, T.J. Santangelo, L. Cubonova, Z. Li, H. Metangmo, J. Ladner, J. Hurwitz, J.N. Reeve, Z. Kelman, Thermococcus kodakarensis has two functional PCNA homologs but only one is required for viability, Extremophiles 17 (2013) 453–461.
[52] L. Meslet-Cladiere, C. Norais, J. Kuhn, J. Briffotaux, J.W. Sloostra, E. Ferrari, U. Hubscher, D. Flament, H. Myllykallio, A novel proteomic approach identifies new interaction partners for proliferating cell nuclear antigen, J. Mol. Biol. 372 (2007) 1137–1148.
[53] J.A. Roberts, S.D. Bell, M.F. White, An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA, Mol. Microbiol. 48 (2003) 361–371.
[54] R. Fujikane, S. Ishino, Y. Ishino, P. Forterre, Genetic analysis of DNA repair in the hyperthermophilic archaeon, Thermococcus kodakaraensis, Genes Genet. Syst. 85 (2010) 243–257.
[55] C. Rouillon, M.F. White, The evolution and mechanisms of nucleotide excision repair proteins, Res. Microbiol. 162 (2011) 19–26.
[56] A.J. Bardwell, L. Bardwell, A.E. Tomkinson, E.C. Friedberg, Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease, Science 265 (1994) 2082–2085.
[57] R. Lestini, Z. Duan, T. Allers, The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in Haloferax volcanii, DNA Repair (Amst.) 9 (2010) 994–1002.
[58] M. Chalfie, Y. Tu, G. Euskirchen, W.W. Ward, D.C. Prasher, Green fluorescent protein as a marker for gene expression, Science 263 (1994) 802–805.
[59] D.M. Chudakov, M.V. Matz, S. Lukyanov, K.A. Lukyanov, Fluorescent proteins and their applications in imaging living cells and tissues, Physiol. Rev. 90 (2010) 1103–1163.
[60] O. Shimomura, F.H. Johnson, Y. Saiga, Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, aequorea, J. Cell Comp. Physiol. 59 (1962) 223–239.
[61] R.Y. Tsien, The green fluorescent protein, Annu. Rev. Biochem. 67 (1998) 509–544.
[62] W. Vermeulen, Dynamics of mammalian NER proteins, DNA Repair (Amst.) 10 (2011) 760–771.
[63] C.J. Reuter, J.A. Maupin-Furlow, Analysis of proteasome-dependent proteolysis in Haloferax volcanii cells, using short-lived green fluorescent proteins, Appl. Environ. Microbiol. 70 (2004) 7530–7538.
[64] R. Lestini, S.P. Laptenok, J. Kuhn, M.A. Hink, M.C. Schanne-Klein, U. Liebl, H. Myllykallio, Intracellular dynamics of archaeal FANCM homologue Hef in response to halted DNA replication, Nucleic Acids Res. 41 (2013) 10358–10370.
[65] P. Forterre, C. Elie, M. Kohiyama, Aphidicolin inhibits growth and DNA synthesis in halophilic arachaebacteria, J. Bacteriol. 159 (1984) 800–802.
[66] M.B. Elowitz, M.G. Surette, P.E. Wolf, J.B. Stock, S. Leibler, Protein mobility in the cytoplasm of Escherichia coli, J. Bacteriol. 181 (1999) 197–203.
[67] M.A. Digman, R. Dalal, A.F. Horwitz, E. Gratton, Mapping the number of molecules and brightness in the laser scanning microscope, Biophys. J. 94 (2008) 2320–2332.
[68] S.T. Bakker, H.J. van de Vrugt, M.A. Rooimans, A.B. Oostra, J. Steltenpool, E. Delzenne-Goette, A. van der Wal, M. van der Valk, H. Joenje, H. te Riele, J.P. de Winter, Fancm-deficient mice reveal unique features of Fanconi anemia complementation group M, Hum. Mol. Genet. 18 (2009) 3484–3495.
[69] W. Crismani, C. Girard, N. Froger, M. Pradillo, J.L. Santos, L. Chelysheva, G.P. Copenhaver, C. Horlow, R. Mercier, FANCM limits meiotic crossovers, Science 336 (2012) 1588–1590.
[70] A. Knoll, J.D. Higgins, K. Seeliger, S.J. Reha, N.J. Dangel, M. Bauknecht, S. Schropfer, F.C. Franklin, H. Puchta, The Fanconi anemia ortholog FANCM ensures ordered homologous recombination in both somatic and meiotic cells in Arabidopsis, Plant Cell 24 (2012) 1448–1464.
[71] A. Lorenz, F. Osman, W. Sun, S. Nandi, R. Steinacher, M.C. Whitby, The fission yeast FANCM ortholog directs non-crossover recombination during meiosis, Science 336 (2012) 1585–1588.
[72] R. Lestini, B. Michel, UvrD controls the access of recombination proteins to blocked replication forks, EMBO J. 26 (2007) 3804–3814.
[73] I.G. Duggin, C.H. Aylett, J.C. Walsh, K.A. Michie, Q. Wang, L. Turnbull, E.M. Dawson, E.J. Harry, C.B. Whitchurch, L.A. Amos, J. Lowe, CetZ tubulin-like proteins control archaeal cell shape, Nature 519 (2015) 362–365.

Chapter 13

DNA Replication Restart in Archaea

Roxane Lestini, Floriane Delpech and
Hannu Myllykallio

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/59994

1. Introduction

One fundamental challenge of cells is to accurately copy their genetic material for cell proliferation. This task is performed by core machineries considered conserved in all three domains of life: bacteria, archaea and eukaryotes [1].

For the vast majority of bacteria, the genome consists of one circular DNA molecule. Replication is initiated at a single replication origin from which two replication forks progress in the opposite direction. Replication termination takes place in the terminus region opposite the origin so that each replication fork has copied approximately one half of the genome. Studies of *Escherichia coli* mutants in key proteins for replication restart such as PriA strongly suggested that many replication forks encounter DNA damage or roadblocks leading to replisome inactivation under normal growth conditions. The reactivation of replication forks has been studied for several decades in bacteria. The picture that emerges is that bacterial proteins implicated in homologous recombination also play a key role in stabilizing and/or restoring blocked replication forks.

Unlike bacterial genomes, eukaryotic chromosomes contain numerous replication origins that can be used as backup origins to rescue arrested forks. Consequently, the importance of replication restart pathways in eukaryotes has long been ignored. However, recent studies have demonstrated that fork restart pathways operate also in eukaryotic cells and are important for cell viability under replication stress conditions. Eukaryotic replication restart pathways described also involve recombination proteins, as in bacteria. Thus it appears that general rules regarding replication restart and the key role of recombination proteins in these processes are conserved in bacteria, yeast and higher eukaryotes, but little is known in archaea, the third domain of life. This is of interest as archaea appear to be evolutionary hybrids between bacteria and eukaryotes.

INTECH
open science | open minds

© 2015 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and eproduction in any medium, provided the original work is properly cited.

Three main archaeal phyla are currently recognized: Crenarchaeota, Euryarchaeota [2] and Thaumarchaeota [3]. Similarly to most bacteria, archaeal genomes are also formed by a circular DNA molecule. However, unlike bacteria, some archaeal species have a single origin, whereas others have multiple origins per chromosome. Moreover, the ploidy of the genome in archaea varies considerably, with some species having one copy per cell whereas others have up to 25 copies of their genome in proliferating cells. As archaeal DNA replication consists both of evolutionary conserved as well as original features, understanding replication restart in these microorganisms will shed light on these fundamental but very complex pathways crucial to fulfill DNA replication. In this chapter we present recent advances on replication in archaea, followed by focused description of the Hef/XPF protein and its implication in replication restart in archaeal cells.

2. DNA replication origins in archaea

2.1. Multiple replication origins

Bacteria replicate their circular chromosome from a defined site called a replication origin. Two replication forks assemble at the replication origin and move in opposite directions. Each replication fork progresses at the same rate, and termination occurs at specific sites opposite the origin. Archaeal chromosomes are also circular, but whereas some archaea initiate replication from a single origin others replicate their chromosome from multiple replication origins, as observed for eukaryotic linear chromosomes (Table 1).

Phylum	Organism	No. of replication origins	References
Euryarchaeota	*Pyrococcus abyssi*	1	[4]
	Haloferax volcanii	3	[5-7]
	Haloferax mediterranei	2	[7]
	Archaeoglobus fulgibus	1	[8]
	Halobacterium sp. NRC1	4	[9, 10]
	Haloarcula hispanica	2	[11]
	Methanothermobacter thermoautotrophicus	1	[12]
Crenarchaeota	*Sulfulobus acidocaldarius*	3	[13]
	Sulfulobus solfataricus	3	[13]
	Sulfulobus islandicus	3	[14]
	Pyrobaculum calidifontis	4	[15]
	Aeropyrum pernix	At least 2	[16]
Thaumarchaeota	*Nitrosopumilus maritimus*	1	[7]

Table 1. Replication origins experimentally identified in archaeal chromosomes

Replication from a single replication origin was reported experimentally in the euryarchaea *Pyrococcus abyssi* [4] and in *Archaeoglobus fulgibus* [8]. But then several studies showed that various euryarchaea have multiple replication origins like the halophiles *Haloferax volcanii* [5-7] and *Haloarcula hispanica* [11]. Multiple replication origins have also been identified in *Sulfulobus solfataricus*, *Sulfulobus acidocaldarius* and *Sulfulobus islandicus* [13, 14] as well as in *Pyrobaculum calidifondis* [15] and *Aeropyrum pernix* [16] that belong to the crearchaeota phylum. Whether archaea from the recently discovered phylum thaumarchaeota have multiple origins remains unknown but a recent study in *Nitrosopumilus maritimus* identified a single replication origin in this organism that is conserved in the phylum, suggesting they have a single replication origin as also suggested by recent computational analysis [7, 17].

2.2. Archaeal replication initiator Orc1/Cdc6 proteins and origins recognition

How replication is regulated to allow a single circular DNA molecule to be replicated from uneven multiple origins is an ongoing question in archaea (Figure 1).

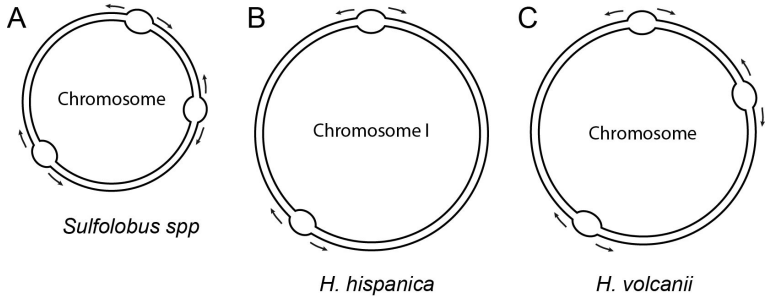


Figure 1. Uneven distribution of multiple replication origins in archaeal chromosomes. (A) in *Sulfolobus* species, (B) in *Haloarcula hispanica*, and (C) in *Haloferax volcanii*. Bubbles on chromosomes indicate replication origins, and arrows indicate bidirectional replication from each origin.

From each replication origin two replication forks are assembled and progress at the same rate so that termination of the replication is asynchronous. The origin region usually has a high content of adenine and thymine residues flanked by several conserved repeated motifs known as Origin Recognition Boxes (ORBs). In manycases archaeal replication origins are linked to replication genes [15, 18] and are located near genes coding initiator proteins. Despite the conservation of the replication origin-initiator structure, archaeal replication origins exhibit considerable diversity in terms of both ORB elements and their initiator genes [7, 11, 12]. Because replication origins can be dramatically diverse, it may facilitate differential usages by these microorganisms to adapt to various harsh environments.

All sequenced archaeal genomes encode proteins homologous to the eukaryotic initiator proteins Orc1 and Cdc6. Because the archaeal proteins are related both to the eukaryotic Orc1 subunit, involved in the replication origin recognition, and Cdc6, involved in the replicative helicase recruitment, they may combine both activities in a single polypeptide. Indeed, several

studies have shown that the archaeal replicative helicase MCM is recruited by Cdc6/Orc1 proteins at replication origins [14, 19, 20]. Archaeal Cdc6/Orc1 proteins also share mechanistic similarities with the bacterial initiator protein DnaA. *In vitro* studies on the binding of Cdc6/Orc1 proteins to ORBs in the *Methanothermobacter thermoautotrophicus* replication origin revealed that they bind cooperatively to the repeated sequences found in the vicinity of the *oriC*, as observed for the association of the bacterial initiator protein DnaA [12, 21].

How multiple replication origins are regulated by Cdc6/Orc1 proteins in archaeal cells is a complex question. The number of Orc1/Cdc6 proteins varies between species, and recent genetic studies attempting to delete *cdc6*/*orc1* genes revealed a complex regulation of replication, highlighting a specificity of initiator proteins at each origin.

For instance, four *cdc6*/*orc1* genes are found on *Halobacterium* NRC-1 chromosome that replicates from four distinct replication origins, but only two *cdc6*/*orc1* genes are located near a replication origin. Genetic studies of replication initiation at one of these origins showed that only the presence of the initiator protein associated was needed, revealing a specific binding of each replication origin by initiator proteins [10]. This is also the case in *S. Solfataricus* cells. Three *cdc6*/*orc1* genes are found in the chromosome. Out of the three replication origins, two were linked to a *cdc6*/*orc1* gene, and different subsets of the three Cdc6/Orc1 proteins recognized these replication origins [22]. The third origin was not recognized by Cdc6/Orc1 initiators. It was specifically recognized by the crenarchaeal-specific WhiP protein (for Winged-Helix initiator protein). WhiP proteins share sequence similarity with the essential eukaryal replication factor Cdt1 and display a domain organization reminiscent of bacterial plasmid initiator proteins. The conservation of WhiP-coding genes located near the replication origin in other crenarchaea suggested that this third replication origin was captured from extrachromosomal elements [16]. A similar situation is found in *Sulfolobus islandicus*. *Sis*OriC-1 was bound by Orc1-1, *Sis*OriC-2 by Orc1-3 while no association of any Cdc6/Orc1 protein was observed at *Ssi*OriC-3 specifically recognized by *Ssi*WhiP protein [14].

Moreover, additional role of initiator proteins independent of replication origins has recently been suggested by serial deletions of *cdc6*/*Orc1* genes in *S. Islandicus* and *H. hispanica*. In *S. Islandicus* none of the three *cdc6*/*Orc1* genes were essential for viability and all three possible double-mutants were viable. However, although one of the Cdc6/Orc1 proteins seemingly did not bind to any replication origin *in vivo*, the triple mutant could not be generated, further suggesting that the observed synthetic lethality may reflect additional role of replication initiator proteins [14]. Similarly, both replication origins in *H. hispanica* chromosome were shown to be controlled independently by specific *cdc6*/*orc1* genes. But while one of the replication origin could be deleted, the deletion of its associated *cdc6*/*orc1* gene lead to a severe growth defect, also suggesting a vital function of the protein outside replication initiation from its associated origin [23].

2.3. Are replication origins essential for viability in archaea?

The specific initiation sites, replication origins, on the chromosome of *H. hispanica* could be deleted separately but it was not possible to generate a mutant deleted for both origins at the same time. Attempt to delete also the replication origins of other replicons found in this

http://dx.doi.org/10.5772/59994

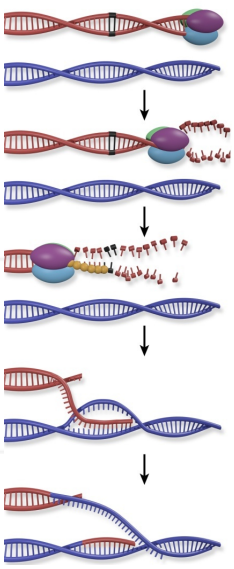


Figure 2. Model for double-strand break repair and replication restart in *E. coli*. RecBCD (purple, green and blue egg-shaped) degrades double-stranded end until it encounters a Chi site (black region). A switch in RecBCD activity produces a 3′-single-stranded DNA on which RecA (yellow ball) proteins are loaded. Homology search and strand exchange forms a Holliday junction (HJ) adjacent to a D-loop. The Holliday junction is resolved by RuvABC. PriA then load the replisome on the D-loop at which replication restarts.

organism suggested that one active *ori-cdc6* pairing on each replicon was essential for genome replication in *H. hispanica* [23]. But recent data obtained in the halophilic archaea *Haloferax volcanii* challenged the notion that replication origins are essential determinants of DNA replication. Indeed, Hawkins *et al*. revealed that not only cells were viable upon deletion of all known replication origins but they even grew faster than the corresponding wild-type cells [5].

How replication initiates in absence of replication origins? Because no evidence for activation of dormant origins has been found, authors favoured the hypothesis that replication initiation occurs randomly on the chromosome at recombination intermediates. Recombination-Dependent Replication (RDR) has first been observed in *E. coli* cells and extensively studied by various laboratories. From those studies it appears that replication fork inactivation occurs very frequently under normal growth conditions. Several replication restart pathways have evolved depending on the cause of arrest. They all share the common feature to involve recombination proteins such as the RecA recombinase and the PriA protein responsible for the loading of a replisome at recombination intermediates (Figure 2).

Indeed Hawkins *et al*. have shown that the archaeal RecA ortholog (RadA) is essential for viability in absence of replication origins.

But the deletion of *radA* alone impaired *H. volcanii* growth, highlighting that viability already relied on recombination [24]. Thus the essentiality of RadA in absence of replication origins may not reflect a direct need for recombination to start replication. Furthermore, in *E. coli* cells

RDR is deleterious for growth and viability. It also seems to be the case in yeast [25] and higher eukaryotes in which replication defects are linked to genome rearrangements and diseases [26, 27]. In that context the better fitness of origins-deleted cells observed in *H. volcanii* is puzzling. Hawkins *et al*. argued that replicative helicases MCM were more efficiently recruited at recombination intermediates as they were not sequestered at replication origin(s). Whether MCM is a limiting factor for replication initiation at replication origins in *H. volcanii* cells is currently not known and would have to be investigated. They also argued that the polyploidy of *H. volcanii* genome (18 copies of the genome in exponential phase [28]) allows viability to rely on stochastic partitioning. This argument implies that all chromosome dimers generated by recombination events (including RDR) do not have to be resolved to provide viable daughter cells, and that proteins involved in resolution of recombination intermediates such as the Holliday junction resolvase Hjc are not essential for viability. This hypothesis is clearly worth of experimental testing.

An alternative explanation for RDR is activation of dormant origins randomly in cells so that no preferential origin emerged at the level of a cell population [29, 30]. In this scenario the essentiality of RadA could imply that randomly-initiated replication forks more often collapse and have to be restarted. This notion would be consistent with an organization of archaeal genes on the genome preventing collision of replication machinery with transcription machinery [18] and physical connections recently suggested between replication and transcription machineries [31].

In conclusion, this study by Hawkins *et al*. raised many interesting questions that need to be experimentally addressed to understand how *H. volcanii* genome is replicated in absence of replication origins. Future work should aim at unravelling the molecular mechanisms that allow archaeal cells lacking origins to be viable and to even show increased fitness. In that context, one interesting protein to focus on might be the helicase/nuclease Hef. Indeed it has recently been shown that Hef (i) is genetically linked to the HJ resolvase Hjc and (ii) is recruited at arrested replication forks in living *H. volcanii* cells [32, 33].

3. Archaeal Hef/XPF proteins from the XPF/MUS81/FANCM family

Proteins belonging to the XPF/MUS81/FANCM endonuclease family act on 3′-flap DNA structures that are formed during DNA repair or replication restart. They are found throughout eukarya and archaea but to date have not been identified in bacteria. Eukaryotes have several XPF/MUS81/FANCM family members that all share a conserved nuclease domain [34] whereas MUS81 proteins possess only an active nuclease domain. In XPF, an active nuclease domain is fused to a SF2-helicase domain that is degenerated and appears to be inactive [35]. By contrast, FANCM consists of a helicase:nuclease fusion in which the nuclease domain is degenerated [36, 37]. Other members can be found that have a degenerated nuclease and/or helicase domain. They assemble into heterodimeric complexes with MUS81, XPF or FANCM proteins to form distinct active complexes involved in DNA repair, meiotic recombination and replication restart [38] (Figure 3).

All archaea encode a protein of the XPF/MUS81/FANCM family of endonucleases. It exists in two forms. The long form, referred as Hef, consists of an N-terminal helicase fused to a C-terminal nuclease and is specific to the euryarchaea. The short form, referred as XPF, lacks the helicase domain and is specific to the crenarchaea and the thaumarchaea (Figure 3).

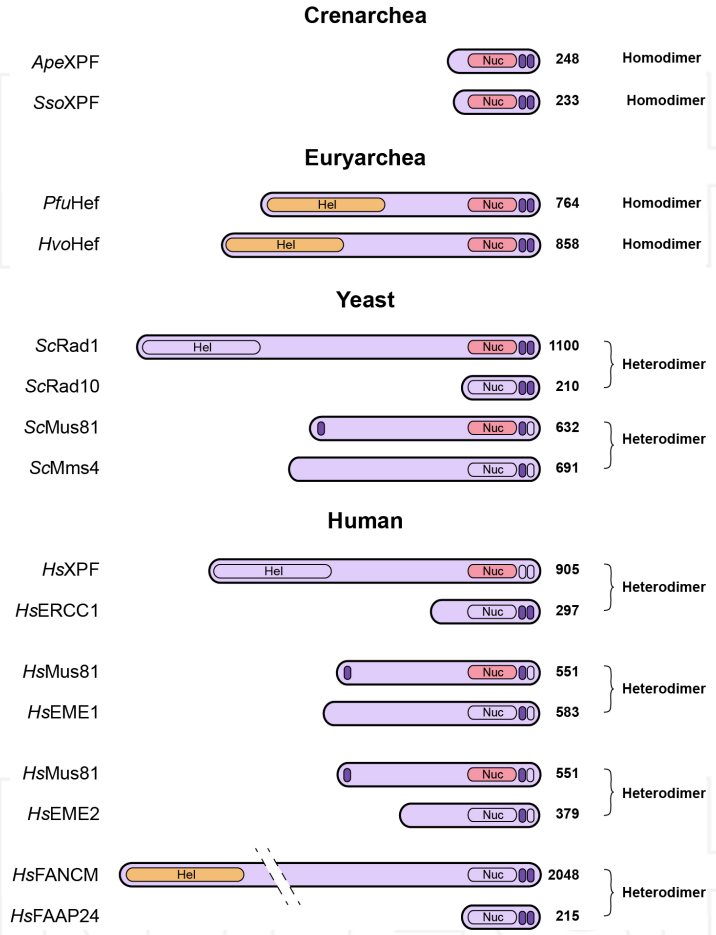


Figure 3. Schematic representation of archaeal and eukaryotic members of XPF/MUS81/FANCM family. Yellow-filled regions represent active helicase domains, pink-filled regions represent active nuclease domains while active HhH domains are represented by dark purple-filled ovals. Numbers of amino acids for each protein are indicated.

The long-formed Hef protein was first identified in *Pyrococcus furiosus* due to its activity on branched DNA structures, Hef meaning **h**elicase-associated **e**ndonuclease **f**ork-structure DNA [39]. Hef has the unique feature among XPF/MUS81/FANCM proteins of having both an active helicase domain and an active nuclease domain, allowing the identification of its human

ortholog FANCM protein also consisting of a helicase:nuclease fusion [37, 40]. What do we know about archaeal Hef/XPF function?

3.1. *In vitro* studies of crenarchaeal XPF proteins

The crystallographic structure of XPF from the crenarchaea *Aeropyrum pernix* was solved in presence and absence of double-stranded DNA [41, 42]. The protein has two domains, a N-terminal nuclease domain and an Helix-hairpin-Helix $(HhH)_2$ domain. *Ape*XPF formed homodimers. The interaction involved the two nuclease domains and the two $(HhH)_2$ domains from each monomer. However, only one monomer seemed catalytically active at a time when the homodimer was bound to DNA. The $(HhH)_2$ domain had a major role in interacting with DNA. This interaction triggered a domain movement coupling the $(HhH)_2$ domain to the nuclease domain to allow subsequent cleavage of the DNA substrate. The DNA was bent by around 90° upon interaction, suggesting that XPF binding causes distortion at double-strand/single-strand DNA junctions.

The nuclease activity of XPF from *Sulfulobus solfataricus* has been studied in more details. The replication factor PCNA (Proliferating Cell Nuclear Antigen) was required *in vitro* for nuclease activity of this "short" XPF [43]. In the cell, the trimeric PCNA ring encircles double-strand DNA (dsDNA) and firmly attaches the replicative polymerase to the template strand, enhancing its processivity. PCNA is a central protein as it also interacts with various proteins involved in replication and/or repair like Fen1. Interaction with PCNA often involves a conserved motif known as PCNA-Interacting Protein (PIP) motif conserved in XPF proteins. Indeed it was shown that *Sso*XPF interacts with PCNA through its conserved PIP motif [43, 44]. Intramolecular FRET experiments showed that the binding of *Sso*XPF to a 3′-flap indeed bent the DNA as observed in *Ape*XPF structure, but that the interaction with PCNA allowed *Sso*XPF to distort the DNA structure in a proper conformation for efficient cleavage [45, 46]. *Sso*XPF preferentially cleaved 3′-flap and processed them into gapped duplex products. It was also observed that *Sso*XPF can act on substrates containing a variety of DNA damages or modifications [47, 48].

3.2. *In vitro* characterization of euryarchaeal Hef proteins

As mentioned previously Hef was identified in *P. furiosus* due to its enzymatic activity on branched DNA structures [39]. *In vitro* experiments on *Pfu*Hef revealed a similar organization of the C-terminal region of archaeal XPF proteins, with a nuclease domain and a helix-hairpin-helix domain. Similarly, homodimers were observed with both the nuclease and the HhH domains forming domain-domain interfaces. Dimer formation appeared crucial for substrate recognition specificity [49]. A variety of branched DNA structures carrying single-strand DNA (ssDNA) portions, such as flapped and fork-structured DNAs, were recognized and cleaved by the C-terminal nuclease domain of *Pfu*Hef [50, 51]. The N-terminal domain of Hef displayed a structured-DNA specific helicase. Two conserved helicase motifs from Super-Family 2 (SF2) helicases were separated by a third domain that shares structure similarity with the "Thumb" domain of polymerases involved in dsDNA binding [51].

In vitro experiments suggested collaboration between the coupled helicase and the nuclease domains of *Pfu*Hef: the helicase domain binds and processes the fork-structured DNA, forming a four-way structure that is then cleaved by the endonuclease domain [52]. Yet replication fork restart can involve the formation of a four-way junction (Holliday junctions) from a three-way junction (fork-like structure). The four-way junction is then resolved by a Holliday junction resolvase. Thus euryarchaeal Hef could be involved in the resolution of stalled replication forks, as suggested for Mus81 complexes in eukaryotes. Indeed in both fission and budding yeast it seems that the primary function of Mus81 complexes is the restart of collapsed-replication forks by homologous recombination [53-57], a role that is functionally redundant with the helicase-nuclease Sgs1-Top3 and the Rqh1-Top3 complexes, respectively [53, 58]. The MUS81 complex is also found in humans [59], and promotes replication fork restart by homologous recombination [60-63].

More recently, *Thermococcus kodakarensis* Hef has been shown to interact with *Tko*PCNA1. The interaction with PCNA did not involve a canonical PIP motif but a disordered region of Hef between the helicase and nuclease domains. Interestingly, these long disordered regions connecting two catalytic domains are a common feature of euryarchael Hef and eukaryotic FANCM proteins [64].

These biochemical studies have indicated that both creanarchaeal XPF and euryarchaeal Hef proteins interact with PCNA and display biochemical activities consistent for being proteins involved in DNA repair and/or replication restart. Is this hypothesis supported by *in vivo* studies of Hef proteins?

3.3. What have we learned deleting *hef* gene in euryarchaea

The *hef* gene has been deleted in two different euryarchaea: in the hyperthermophile *Thermococcus kodakarensis* and in the halophile *Haloferax volcanii*. In both organisms, Hef was non-essential for cell viability under normal laboratory growth conditions.

T. kodakarensis cells deleted for *hef* showed increased sensitivity to a variety of DNA damaging agents [65], consistent with a role of Hef in the maintenance of genomic stability. The sensitivity to UV irradiation suggested that Hef was involved in the repair of UV lesions. Both helicase and nuclease domains of *Tko*Hef were needed as the same phenotype was observed upon deletion of the entire gene and deletion of the helicase-coding region or the nuclease-coding region of the gene [65]. Nucleotide Excision Repair (NER) is the major pathway to repair DNA lesions after UV radiation. But the existence of an archaeal NER pathway has not been established yet. Most archaea have eukaryal-type NER genes, but most of eukaryal NER proteins have multiple cellular functions so the presence of several NER-like protein is not enough to conclude that a functional NER pathway can be found in archaea [66]. Interestingly the sensitivity to UV radiations of *T. kodakarensis* cells deleted for *hef* suggested that Hef was involved in Nucleotide Excision Repair (NER), as its human ortholog XPF-ERCC1 and its counterpart RAD1-RAD10 in the yeast *Saccharomyces cerevisae* [67, 68]. And it also suggested that an active NER pathway exists in archaea. Clearly, additional experiments are now needed to better understand the role of *Tko*Hef in NER and, more generally, to further dissect the pathway responsible for archaeal Nucleotide Excision Repair.

In contrast, the deletion of *hef* in *H. volcanii* neither affected sensitivity to various DNA damaging agents nor recombination frequency. We could only observed a slow-growth phenotype of the Δ*hef* colonies when chronically exposed to mitomycin C (MMC) on plate [32]. Note that the direct comparison between these two studies on phenotypic analyses of Δ*hef* strains is difficult as Table 2 illustrates major differences regarding experimental conditions used including drug concentrations and cell treatment methods.

	H. volcanii			***T. kodakarensis***		
	Concentration	**Exposure**	**Sensitivity**	**Concentration**	**Exposure**	**Sensitivity**
UV irradiation	Up to 150 J/m^2	on plate	-	2 or 5 J/m^2	on plate	+
MMS	0,04%	1 hour in suspension	-	0,05%	4 hours in suspension	++
γ-rays	Up to 1000 Gy	on plate	-	1700 Gy	in suspension	++
Phleomycin	1 or 2 mg/ml	1 hour in suspension	-			
Mitomycin C	0,02 μg/ml	On plate	slow-growing	100 μg/ml	4 hours in suspension	+++

Table 2. Methods used for exposure of *H. volcanii* and *T. Kodakarensis* Δ*hef* cells to DNA damaging agents as reported in [32, 65, 69].

A possible explanation for these phenotypic differences is that NER proteins in *Thermococcus* and *Haloferax* species are very different. In fact *H. volcanii* also possesses bacterial-like NER proteins most probably acquired by lateral gene transfer [66], and it was shown that they were responsible for the repair of UV lesions [32].

To further investigate the role of Hef in *H. volcanii*, the observed lack of an obvious phenotype for Δ*hef* cells prompted us to combine *hef*-deletion with other endonuclease or helicase deletions that may encode redundant functions with Hef. Among several combinations tested, we demonstrated that Hef was essential for viability in the absence of the Holliday junctions (HJs) resolvase Hjc. Holliday junctions are four-way branched DNA structures formed during homologous recombination strand exchange and recombination-dependent replication restart. HJs resolvases are found in bacteria, archaea and eukarya, although they are not evolutionary related. Hjc is conserved throughout archaea. The single deletion of *hjc* gene in *H. volcanii* cells (as well as in *T. Kodakarensis* cells) did not affect growth rate, DNA repair or recombination [32, 65]. Co-lethality of Hef and Hjc could be explained by redundant roles of Hef and Hjc as HJs resolvase. In this scenario Hef could use its helicase activity on arrested replication forks to process them into four-way DNA structures that can be resolved by its nuclease activity. This scenario was compatible with the *in vitro* studies described above.

Indeed, point mutations inactivating the helicase activity (*Hvo*Hef-K48A) or the nuclease activity (*Hvo*Hef-D679A) of *Hvo*Hef resulted in the same phenotype observed in the absence of the entire protein. This nicely demonstrated that both helicase and nuclease activities of Hef were required for fulfilling its role in the absence of the Hjc resolvase [32]. To test the hypothesis that Hef and Hjc were both acting as HJs resolvase, we deleted *hef* or *hjc* in a strain carrying a *radA* deletion. In absence of RadA recombinase, HJs are no longer formed by homologous recombination so that deleting HJs resolvases should not have any affect. Indeed cells deleted for both *radA* and *hjc* were phenotypically similar to cells only deleted for *radA*. This observation was consistent with a role of Hjc in the resolution of HJs formed by RadA-mediated strand exchange during homologous recombination. However, *radA* gene could not be deleted in *hef*-deleted cells [32], strongly reflecting that functional roles of Hjc and Hef were distinct. This observation also suggested that Hef was required for cell viability in absence of recombination. Which alternative pathways or additional functions could (i) depend on Hef, (ii) be essential for cell viability during normal growth condition, and (iii) implicate recombination proteins and/or Hef? Replication restart is one possible pathway. But how could we obtain more detailed information on functional roles of Hef if *hef*-deleted mutant strains hardly shows any phenotype or cannot be combined with other deletions? We decided to develop tools to allow dynamic localization of fluorescently-labelled Hef proteins in living *Haloferax volcanii* cells.

Genotype	Growth phenotypes	References
WT	+++	
Δ*hef*	+++	[32]
Δ*hjc*	+++	[32]
Δ*radA*	+	[24]
Δ*hef* Δ*hjc*	-	[32]
Δ*radA* Δ*hjc*	+	[32]
Δ*radA* Δ*hef*	-	[32]

Table 3. Growth phenotypes of *H. volcanii* deletion mutants.

4. Dynamic localization of Hef proteins fused to the Green Fluorescent Protein (GFP) in living *H. volcanii* cells

The Green Fluorescent Protein (GFP) was originally isolated from the jellyfish *Aequora Victoria* [70]. It is encoded by a single polypeptide containing the chromophore. After translation of the protein, an autocatalytic process involving oxygen has to take place within the chromophore. Once active, the GFP has a major excitation peak at a wavelength of 395 nm and an emission peak at 509 nm. A deep understanding of the protein has enabled the development

of several GFP variants with modified spectral properties [71, 72]. All those FPs are now used in living cells/organisms to study protein localization, mobility, turnover, interactions, and much more [73]. Such approaches can reveal key features of proteins *in vivo* to complete our understanding of pathways, as illustrated for NER pathway in mammalian cells [74], but their use in archaeal cells has been rather limited until recently.

4.1. Fusion of the green fluorescent protein to the C-terminal end of Hef protein in *H. volcanii*

GFP has been used to investigate proteasome-dependant proteolysis and protein levels in *H. volcanii* [75, 76] as well as biofilm formation by the crenarchaea *Sulfulobus solfataricus* [77], two aerobic archaea. Because GFPs variant can differ not only by their fluorescence properties but also by their maturation rate of the fluorophore, temperature and pH stability or oligomeric state for instance, one has to carefully choose the variants that optimally fit the lifestyle of the organism being studied. We have recently used GFP-fusions to investigate protein localization and behaviour in archaeal cells for the first time. These studies were performed using the halophile *H. volcanii* that has a relatively high intracellular salt concentration (around 2,5M in laboratory growth conditions) and an optimal growth temperature of 45°C. Expression of several GFP variants were previously tested in this species, demonstrating that the smRS-GFP could be used for further studies [75]. This variant has mutations increasing solubility (Phe99Ser, Met153Thr and Val63Ala) as well as a mutation in one of the three amino acids of the chromophore (Ser65Thr) that redshifts the absorption maximum to 488 nm without changing the emission properties of the protein [78, 79]. Based on this observation we fused the smRS-GFP to the C-terminal end of *H. volcanii* Hef. The resulting fusion protein was expressed under physiological expression levels and conditions from the native chromosomal locus of the *hef* gene [33].

Whether GFP-fused Hef proteins remained functional was then tested by comparing cells deleted for *hef* with cells expressing the *hef*::*gfp* allele. No growth delay was observed for *hef*::*gfp* cells on MMC plates, indicating functional complementation by Hef::GFP construct. In agreement with this notion, *hjc* could be deleted, although a growth defect was measured for *hef*::*gfp* Δ*hjc* cells. Because we were interested in the localization of Hef in response to replication arrests, we exposed cells to aphidicolin (APD), an antibiotic that inhibits DNA synthesis in halophilic archaea [80], thus arresting replication forks. Exposing *hef*-deleted cells to increasing concentrations of APD decreased cell viability, showing that indeed Hef is involved in the genomic stability upon replication arrest. Such decrease in cell viability was not observed with cells expressing GFP-fused Hef.

4.2. Localising the fluorescence signal in *H. volcanii* living cells

We then observed the localisation of Hef::GFP proteins by fluorescence microscopy, comparing cells exposed to APD to non-treated control samples. Towards this goal, a drop of cells was spotted on an agarose slice placed on a glass slide. After allowing this drop to dry, the agarose pad was covered with a cover-slip for cell imaging studies using a wield-field microscope to visualize a large number of individual cells. Differential Interference Contrast (DIC) [also

known as Nomarski Interference Contrast (NIC)], was first used to visualized the cells as it enhances the contrast in unstained, transparent samples. Then fluorescence imaging was performed (exciting at 474 nm and collecting at 527-554 nm). Note however that due to the small cell size (around 1 to 2 μm) and the use of the soft agarose, not all cells were in the same focal plane. In order not to lose any information, fluorescence images were acquired at different focal planes on the z-axis. Consecutive slices of cells in focus were then selected and used to perform a maximum intensity z-projection. At each pixel, the highest fluorescence signal was kept when comparing the selected images. This maximum intensity z-projection resulted into a two-dimensional picture where the maximal fluorescence signals from different focal planes were recorded (Figure 4).

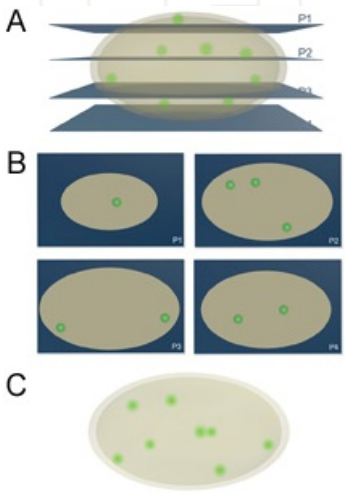


Figure 4. Schematic representation of fluorescence signal analysed in cells. (A) Representation of a cell with fluorescence foci and the different focal planes used for imaging. (B) Representation of fluorescence signal in each focal plane. (C) Resulting image after projection of the maximum fluorescence signal at each pixel for the four focal planes.

Resulting images contained hundreds of cells that were analysed by quantitative image analyses using IMARIS software. Different imaging parameters were optimized to detect cells and fluorescence foci within cells using automatic thresholds to avoid user-bias. This approch allowed thousands of cells to be analysed in each condition tested, providing extremely high statistical power.

4.3. Hef::GFP molecules are recruited at arrested replication forks

Using such approach, we have shown that Hef::GFP proteins formed fluorescence foci even under normal growth condition, in the absence of any DNA damaging agents. The number of these foci was significantly increased from two to four foci *per* cell in response to aphidicolin exposure. We also observed that the number of foci per individual cell changed significantly.

While the majority of cells had one or two foci in normal growth conditions, a higher proportion of cells having more than two foci was observed upon APD exposure (Figure 5).

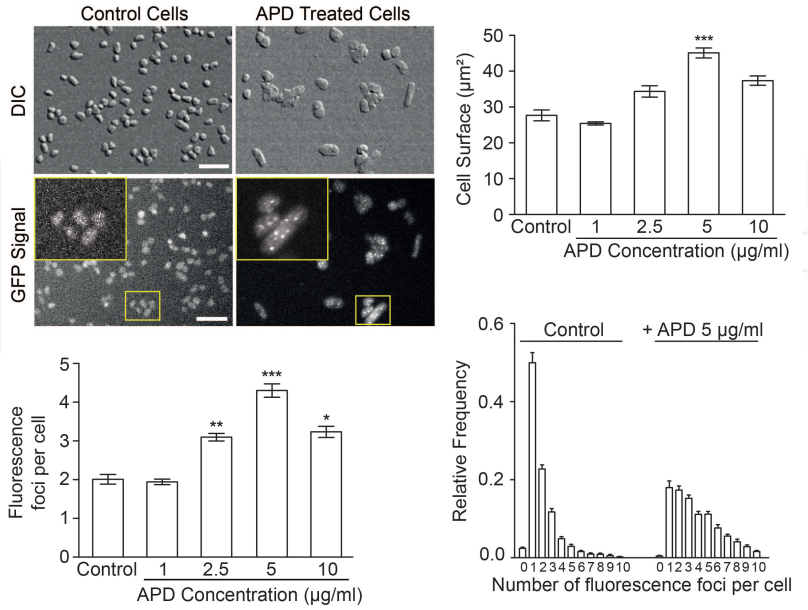


Figure 5. *In vivo* localization of GFP-labeled Hef in response to aphidicolin exposure. A total of 23760 spots within 13666 control cells and 15299 spots within 3721 APD-treated cells were analyzed. (A) Pictures of DIC and GFP signal of *hef::gfp* cells under control conditions and after exposure to 5 µg/ml aphidicolin. Bar equals 10 µm. (B) Average cell surface of *hef::gfp* cells in response to increasing concentrations of aphidicolin. (C) Mean number of GFP-Hef labeled fluorescence foci per cell in response to increasing concentrations of aphidicolin. (D) Relative frequency of number of foci per individual cell. All error bars represent SD. n > 3 experiments, t test are performed in comparison to control without aphidicolin. *** Significantly different, p<0.001. ** Significantly different, p<0.01. * Significantly different, p<0.05. From [33].

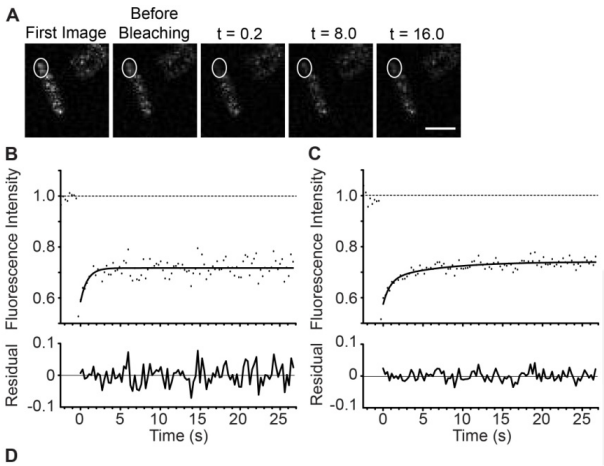
We have also observed that cell size was increased from 28 to 45 µm² in response to replication arrest (i.e. APD exposure). We have shown using other DNA damaging agents that increased cell size and number of foci were specific to APD treatment, suggesting that indeed *Hvo*Hef is recruited at arrested replication forks brought about by addition of aphidicolin.

4.4. Diffusing pattern of Hef::GFP molecule upon replication arrests

To investigate the diffusion of Hef::GFP molecules inside and outside fluorescence foci, we performed Fluorescence Recovery After Photobleaching (FRAP) and Number and Brightness (N&B) experiments. These experiments were performed using a confocal microscope on cells immobilized on a poly-D-lysine coated cover-slip.

In FRAP experiments a region of interest was photobleached in a cell. The speed of fluorescence recovery in that region was then measured, reflecting the diffusion of

http://dx.doi.org/10.5772/59994



Sample	k1 (1/s)	k2 (1/s)	n
Control Cells	1.153 [0.605, 1.702]		9
APD-treated Cells	2.750 [0.512,4.989]	0.240 [0.147,0.334]	8
	1.153 (fixed)	0.144 [0.057,0.232]	8

Figure 6. Fluorescence Recovery After Photobleaching experiments to study the dynamic localization of GFP-labeled Hef molecules at fluorescence foci. (A) Images of a representative cell in response to aphidicolin treatment for FRAP analysis. FRAP regions are shown by white circles. Time after photobleaching in seconds. Bar equals 5 μm. (B) Fluorescence recovery curve averaged for 9 control cells. C) Fluorescence recovery curve averaged for 8 aphidicolin treated-cells. (D) Diffusion constants [Confidence interval at 95%] calculated for GFP-labeled Hef diffusing molecules. From [33].

Hef::GFP fluorescent molecules arriving from the non-photobleached region of the cell. In control cells (no aphidicolin), one major population of Hef::GFP diffusing molecules was observed. From the fit of the recovery curve we obtained the recovery constant, allowing then the apparent two-dimensional diffusion rate of Hef::GFP to be estimated at 0.8 to 2.3 μm² per second. This appeared markedly lower than expected for Hef dimer, as revealed by analytical ultracentrifugation experiments on purified *Hvo*Hef further indicating that Hef has a peculiar elongated shape in solution. Several possibilities may explain this limited diffusion. In addition to the non-globular quaternary structure, physical constraints of the cytosol, possibly resulting from high DNA and salt concentration, and/or transient interactions with cellular components (DNA or proteins complexes) may explain this slow diffusion. But FRAP experiments performed on cells exposed to aphidicolin revealed an additional, even more slowly-diffusing population that was clearly induced by APD treatment (Figure 6).

Such changes in the diffusion pattern of Hef::GFP molecules upon replication arrests were also observed using N&B technique that measures fluctuation of fluorescence intensity in each analysed pixel [81]. These analyses were performed on one hundred images taken every 2 seconds, and cell regions including and excluding fluorescence foci were compared. Fluctuation of fluorescence intensity per pixel was then used to

determine the number of diffusing molecules and their brightness. This information can then be used to deduce changes in the oligomeric state of the fluorescent molecules. When N&B technique was applied to diffusing Hef::GFP molecules, changes in the oligomeric state (i.e. higher brightness) were observed upon APD treatment. This observation revealed oligomerization and/or co-localisation of several Hef::GFP molecules induced by APD exposure (Figure 7), and provided a feasible explanation for the slow diffusion in APD treated cells revealed by FRAP experiments.

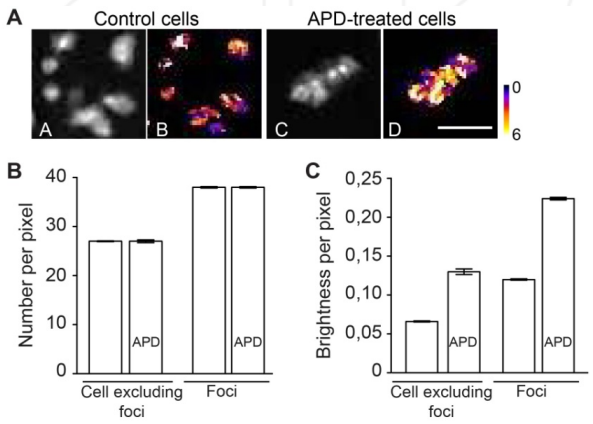


Figure 7. Number & Brightness experiments to study the oligomeric state of GFP-labeled Hef molecules diffusing at fluorescence foci. (A) Images of representative cells for N&B analysis. Average intensity (A and C) and pseudo-coloured normalized brightness values (B and D) for representative control cells (A and B) and cells exposed to 5 μg/ml aphidicolin (C and D). Bar equals 5 μm. (B) Average number of Hef::GFP diffusing molecules per pixel. (C) Average brightness of Hef::GFP diffusing molecules per pixel. From [33].

Overall, the results obtained from FRAP and N&B experiments were consistent with the notion that Hef::GFP molecules are actively recruited at arrested replication forks. Whether the slow-diffusion pattern of Hef::GFP molecules reflects their recruitment directly on DNA and/or as part of protein complexes at arrested replication forks are questions that remain to be addressed. Interestingly, *hjc* deletion had effect neither on cell size nor on the number of foci per cell in normal growth condition as well as in response to APD treatment. These observations showed that recruitment of Hef to arrested replication forks was not increased in the absence of the alternative pathway involving Hjc (and RadA), suggesting that Hef is recruited at arrested replication forks even in the presence of the alternative HR-dependent pathway. We also noted that in eukarya recent studies have indicated that FANCM proteins can prevent homologous recombination [82-85]. This raised the possibility that *Hvo*Hef may prevent access of recombination proteins to arrested replication forks (Figure 8) [33]. Because both the helicase and nuclease activities of Hef are presumably needed for biological function, we also assumed that Hef is directly implicated in processing of arrested replication forks. These hypotheses will be addressed in the future experimental work.

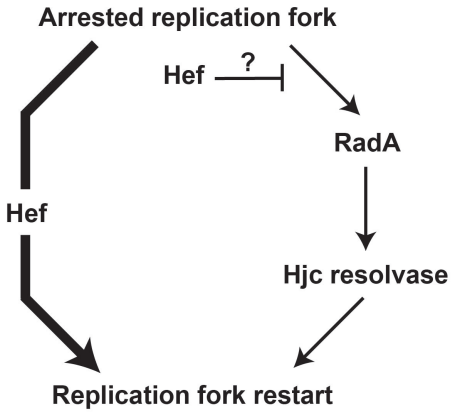


Figure 8. Model for replication restart in *Haloferax volcanii*. Two alternative pathways allow replication restart: one is dependent on the homologous recombination proteins Hjc and RadA (pathway on the right) and one is independent of homologous recombination (pathway on the left). Our data show that Hef has a dominant role during replication restart, even in the presence of Hjc, and are also compatible with Hef preventing the access of recombination proteins at arrested forks. From [33].

Moreover, our work has also shown that replication forks arrest spontaneously occured in *H. volcanii* cells that may contain a high number of replications forks. This is exemplified by the fact that a typical *H. volcanii* cell contains 18 genome copies and that each molecule can be replicated by up to 8 replication forks. Whether all genome copies are replicated simultaneously is not known and this striking question needs to be addressed in future work to better understand replication dynamic in archaea. But as we have observed up to 15 to 20 arrested replication forks in some individual cells, our imaging studies rather suggest that several copies of the genome may be replicated in one cell.

5. Concluding remarks

In vitro characterization of Hef/XPF proteins suggested a key role in genomic stability similarly to what was observed in Eukarya. Genetic experiments coupled to dynamic localization of Hef proteins fused to the Green Fluorescence Protein further revealed that Hef is recruited at arrested replication forks in *Haloferax volcanii* cells. Experimental approaches we have recently developed for halophiles provide a valuable tool for studying functional roles of Hef at arrested replication forks in living archaeal cells. Understanding how Hef is recruited at arrested replication forks and the processing taking place to allow replication restart is our next challenge. Whether XPF protein have similar role in crearchaeal cells remains to be addressed. As genetic tools have been developed for crenarchaeal organisms, future studies might tackle this issue [86, 87].

In conclusion, as archaea possess hallmarks of both bacterial and eukaryotic replication systems we believe that continuation of studies underlined will shed light on the evolutionary

history of replication restart mechanisms and its complex machinery that we are just starting to unravel in eukaryotes and now archaea.

Acknowledgements

We thank A. Lestini for figures 1, 2, 3 and 4. Figures 5, 6, 7 and 8 were reproduced with permission from Lestini *et al*. [33] published by Oxford Journals in Nucleic Acides Research. We also want to acknowledge the Agence Nationale de la Recherche for funding (ARCREP and RETIDYNA grants).

Author details

Roxane Lestini*, Floriane Delpech and Hannu Myllykallio

*Address all correspondence to: roxane.lestini@polytechnique.edu

Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS UMR–INSERM U696, Palaiseau Cedex, France

References

[1] O'Donnell M, Langston L, Stillman B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. Cold Spring Harb Perspect Biol2013 Jul;5(7).

[2] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A1990 Jun; 87(12):4576-9.

[3] Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat Rev Microbiol2008 Mar;6(3):245-52.

[4] Myllykallio H, Lopez P, Lopez-Garcia P, Heilig R, Saurin W, Zivanovic Y, et al. Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. Science2000 Jun 23;288(5474):2212-5.

[5] Hawkins M, Malla S, Blythe MJ, Nieduszynski CA, Allers T. Accelerated growth in the absence of DNA replication origins. Nature2013 Nov 28;503(7477):544-7.

[6] Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T. Genetic and physical mapping of DNA replication origins in Haloferax volcanii. PLoS Genet2007 May 18;3(5):e77.

[7] Pelve EA, Martens-Habbena W, Stahl DA, Bernander R. Mapping of active replication origins in vivo in thaum- and euryarchaeal replicons. Mol Microbiol2013 Nov; 90(3):538-50.

[8] Maisnier-Patin S, Malandrin L, Birkeland NK, Bernander R. Chromosome replication patterns in the hyperthermophilic euryarchaea Archaeoglobus fulgidus and Methanocaldococcus (Methanococcus) jannaschii. Mol Microbiol2002 Sep;45(5):1443-50.

[9] Berquist BR, DasSarma S. An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, Halobacterium sp. strain NRC-1. J Bacteriol2003 Oct;185(20):5959-66.

[10] Coker JA, DasSarma P, Capes M, Wallace T, McGarrity K, Gessler R, et al. Multiple replication origins of Halobacterium sp. strain NRC-1: properties of the conserved orc7-dependent oriC1. J Bacteriol2009 Aug;191(16):5253-61.

[11] Wu Z, Liu H, Liu J, Liu X, Xiang H. Diversity and evolution of multiple orc/cdc6-adjacent replication origins in haloarchaea. BMC Genomics2012;13:478.

[12] Majernik AI, Chong JP. A conserved mechanism for replication origin recognition and binding in archaea. Biochem J2008 Jan 15;409(2):511-8.

[13] Lundgren M, Andersson A, Chen L, Nilsson P, Bernander R. Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. Proc Natl Acad Sci U S A2004 May 4;101(18):7046-51.

[14] Samson RY, Xu Y, Gadelha C, Stone TA, Faqiri JN, Li D, et al. Specificity and function of archaeal DNA replication initiator proteins. Cell Rep2013 Feb 21;3(2):485-96.

[15] Pelve EA, Lindas AC, Knoppel A, Mira A, Bernander R. Four chromosome replication origins in the archaeon Pyrobaculum calidifontis. Mol Microbiol2012 Sep;85(5): 986-95.

[16] Robinson NP, Bell SD. Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. Proc Natl Acad Sci U S A2007 Apr 3;104(14):5806-11.

[17] Luo H, Zhang C-T, Gao F. Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. Frontiers in Microbiology. (Methods). 2014-September-15;5.

[18] Andersson AF, Pelve EA, Lindeberg S, Lundgren M, Nilsson P, Bernander R. Replication-biased genome organisation in the crenarchaeon Sulfolobus. BMC Genomics2010;11:454.

[19] Matsunaga F, Forterre P, Ishino Y, Myllykallio H. In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. Proc Natl Acad Sci U S A2001 Sep 25;98(20):11152-7.

[20] Miller JM, Arachea BT, Epling LB, Enemark EJ. Analysis of the crystal structure of an active MCM hexamer. elife2014;3.

[21] Capaldi SA, Berger JM. Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon Methanothermobacter thermoautotrophicus. Nucleic Acids Res2004;32(16):4821-32.

[22] Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, Bell SD. Identification of two origins of replication in the single chromosome of the archaeon Sulfolobus solfataricus. Cell2004 Jan 9;116(1):25-38.

[23] Wu Z, Liu J, Yang H, Liu H, Xiang H. Multiple replication origins with diverse control mechanisms in Haloarcula hispanica. Nucleic Acids Res2014 Feb;42(4):2282-94.

[24] Woods WG, Dyall-Smith ML. Construction and analysis of a recombination-deficient (radA) mutant of Haloferax volcanii. Mol Microbiol1997 Feb;23(4):791-7.

[25] Kolodner RD, Putnam CD, Myung K. Maintenance of genome stability in Saccharomyces cerevisiae. Science2002 Jul 26;297(5581):552-7.

[26] Abbas T, Keaton MA, Dutta A. Genomic instability in cancer. Cold Spring Harb Perspect Biol2013 Mar;5(3):a012914.

[27] Jackson AP, Laskey RA, Coleman N. Replication proteins and human disease. Cold Spring Harb Perspect Biol2014 Jan;6(1).

[28] Breuert S, Allers T, Spohn G, Soppa J. Regulated polyploidy in halophilic archaea. PLoS One2006;1:e92.

[29] Cvetic C, Walter JC. Eukaryotic origins of DNA replication: could you please be more specific? Semin Cell Dev Biol2005 Jun;16(3):343-53.

[30] Michel B, Bernander R. Chromosome replication origins: do we really need them? Bioessays2014 Jun;36(6):585-90.

[31] Pluchon PF, Fouqueau T, Creze C, Laurent S, Briffotaux J, Hogrel G, et al. An Extended Network of Genomic Maintenance in the Archaeon Pyrococcus abyssi Highlights Unexpected Associations between Eucaryotic Homologs. PLoS One2013;8(11):e79707.

[32] Lestini R, Duan Z, Allers T. The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in Haloferax volcanii. DNA Repair (Amst)2010 Sep 4;9(9):994-1002.

[33] Lestini R, Laptenok SP, Kuhn J, Hink MA, Schanne-Klein MC, Liebl U, et al. Intracellular dynamics of archaeal FANCM homologue Hef in response to halted DNA replication. Nucleic Acids Res2013 Dec 1;41(22):10358-70.

[34] Enzlin JH, Scharer OD. The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. EMBO J2002 Apr 15;21(8):2045-53.

[35] Sgouros J, Gaillard PH, Wood RD. A relationship betweena DNA-repair/recombination nuclease family and archaeal helicases. Trends Biochem Sci1999 Mar;24(3):95-7.

[36] Ciccia A, Ling C, Coulthard R, Yan Z, Xue Y, Meetei AR, et al. Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. Mol Cell2007 Feb 9;25(3):331-43.

[37] Meetei AR, Medhurst AL, Ling C, Xue Y, Singh TR, Bier P, et al. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. Nat Genet2005 Sep;37(9):958-63.

[38] Ciccia A, McDonald N, West SC. Structural and functional relationships of the XPF/MUS81 family of proteins. Annu Rev Biochem2008;77:259-87.

[39] Komori K, Fujikane R, Shinagawa H, Ishino Y. Novel endonuclease in Archaea cleaving DNA with various branched structure. Genes Genet Syst2002 Aug;77(4):227-41.

[40] Mosedale G, Niedzwiedz W, Alpi A, Perrina F, Pereira-Leal JB, Johnson M, et al. The vertebrate Hef ortholog is a component of the Fanconi anemia tumor-suppressor pathway. Nat Struct Mol Biol2005 Sep;12(9):763-71.

[41] Lally J, Newman M, Murray-Rust J, Fadden A, Kawarabayasi Y, McDonald N. Crystallization of the xeroderma pigmentosum group F endonuclease from Aeropyrum pernix. Acta Crystallogr D Biol Crystallogr2004 Sep;60(Pt 9):1658-61.

[42] Newman M, Murray-Rust J, Lally J, Rudolf J, Fadden A, Knowles PP, et al. Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. EMBO J2005 Mar 9;24(5):895-905.

[43] Roberts JA, Bell SD, White MF. An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. Mol Microbiol2003 Apr;48(2):361-71.

[44] Meslet-Cladiere L, Norais C, Kuhn J, Briffotaux J, Sloostra JW, Ferrari E, et al. A novel proteomic approach identifies new interaction partners for proliferating cell nuclear antigen. J Mol Biol2007 Oct 5;372(5):1137-48.

[45] Hutton RD, Craggs TD, White MF, Penedo JC. PCNA and XPF cooperate to distort DNA substrates. Nucleic Acids Res2010 Mar;38(5):1664-75.

[46] Hutton RD, Roberts JA, Penedo JC, White MF. PCNA stimulates catalysis by structure-specific nucleases using two distinct mechanisms: substrate targeting and catalytic step. Nucleic Acids Res2008 Dec;36(21):6720-7.

[47] Roberts JA, White MF. DNA end-directed and processive nuclease activities of the archaeal XPF enzyme. Nucleic Acids Res2005;33(20):6662-70.

[48] Roberts JA, White MF. An archaeal endonuclease displays key properties of both eukaryal XPF-ERCC1 and Mus81. J Biol Chem2005 Feb 18;280(7):5924-8.

[49] Nishino T, Komori K, Ishino Y, Morikawa K. X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes. Structure2003 Apr;11(4):445-57.

[50] Nishino T, Komori K, Ishino Y, Morikawa K. Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms. Structure2005 Aug;13(8):1183-92.

[51] Nishino T, Komori K, Tsuchiya D, Ishino Y, Morikawa K. Crystal structure and functional implications of Pyrococcus furiosus hef helicase domain involved in branched DNA processing. Structure2005 Jan;13(1):143-53.

[52] Komori K, Hidaka M, Horiuchi T, Fujikane R, Shinagawa H, Ishino Y. Cooperation of the N-terminal Helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks. J Biol Chem2004 Dec 17;279(51):53175-85.

[53] Doe CL, Osman F, Dixon J, Whitby MC. DNA repair by a Rad22-Mus81-dependent pathway that is independent of Rhp51. Nucleic Acids Res2004;32(18):5570-81.

[54] Froget B, Blaisonneau J, Lambert S, Baldacci G. Cleavage of stalled forks by fission yeast Mus81/Eme1 in absence of DNA replication checkpoint. Mol Biol Cell2008 Feb; 19(2):445-56.

[55] Kai M, Boddy MN, Russell P, Wang TS. Replication checkpoint kinase Cds1 regulates Mus81 to preserve genome integrity during replication stress. Genes Dev2005 Apr 15;19(8):919-32.

[56] Matulova P, Marini V, Burgess RC, Sisakova A, Kwon Y, Rothstein R, et al. Cooperativity of Mus81.Mms4 with Rad54 in the resolution of recombination and replication intermediates. J Biol Chem2009 Mar 20;284(12):7733-45.

[57] Roseaulin L, Yamada Y, Tsutsui Y, Russell P, Iwasaki H, Arcangioli B. Mus81 is essential for sister chromatid recombination at broken replication forks. EMBO J2008 May 7;27(9):1378-87.

[58] Kaliraman V, Mullen JR, Fricke WM, Bastin-Shanower SA, Brill SJ. Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease. Genes Dev2001 Oct 15;15(20):2730-40.

[59] Chen XB, Melchionna R, Denis CM, Gaillard PH, Blasina A, Van de Weyer I, et al. Human Mus81-associated endonuclease cleaves Holliday junctions in vitro. Mol Cell2001 Nov;8(5):1117-27.

[60] Ciccia A, Constantinou A, West SC. Identification and characterization of the human mus81-eme1 endonuclease. J Biol Chem2003 Jul 4;278(27):25172-8.

[61] Franchitto A, Pirzio LM, Prosperi E, Sapora O, Bignami M, Pichierri P. Replication fork stalling in WRN-deficient cells is overcome by prompt activation of a MUS81-dependent pathway. J Cell Biol2008 Oct 20;183(2):241-52.

[62] Hanada K, Budzowska M, Davies SL, van Drunen E, Onizawa H, Beverloo HB, et al. The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. Nat Struct Mol Biol2007 Nov;14(11):1096-104.

[63] Shimura T, Torres MJ, Martin MM, Rao VA, Pommier Y, Katsura M, et al. Bloom's syndrome helicase and Mus81 are required to induce transient double-strand DNA breaks in response to DNA replication stress. J Mol Biol2008 Jan 25;375(4):1152-64.

[64] Ishino S, Yamagami T, Kitamura M, Kodera N, Mori T, Sugiyama S, et al. Multiple interactions of the intrinsically disordered region between the helicase and nuclease domains of the archaeal Hef protein. J Biol Chem2014 Aug 1;289(31):21627-39.

[65] Fujikane R, Ishino S, Ishino Y, Forterre P. Genetic analysis of DNA repair in the hyperthermophilic archaeon, Thermococcus kodakaraensis. Genes Genet Syst2010;85(4):243-57.

[66] Rouillon C, White MF. The evolution and mechanisms of nucleotide excision repair proteins. Res Microbiol2011 Jan;162(1):19-26.

[67] Bardwell AJ, Bardwell L, Tomkinson AE, Friedberg EC. Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease. Science1994 Sep 30;265(5181):2082-5.

[68] Sijbers AM, de Laat WL, Ariza RR, Biggerstaff M, Wei YF, Moggs JG, et al. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. Cell1996 Sep 6;86(5):811-22.

[69] Duan Z. Genetic Analysis of Two Structure-specific Endonucleases Hef and Fen1 in Archaeon *Haloferax volcanii* (PhD thesis): University of Nottingham; 2008.

[70] Shimomura O, Johnson FH, Saiga Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. J Cell Comp Physiol1962 Jun;59:223-39.

[71] Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the Aequorea victoria green fluorescent protein. Science1996 Sep 6;273(5280):1392-5.

[72] Tsien RY. The green fluorescent protein. Annu Rev Biochem1998;67:509-44.

[73] Chudakov DM, Matz MV, Lukyanov S, Lukyanov KA. Fluorescent proteins and their applications in imaging living cells and tissues. Physiol Rev2010 Jul;90(3):1103-63.

[74] Vermeulen W. Dynamics of mammalian NER proteins. DNA Repair (Amst)2011 Jul 15;10(7):760-71.

[75] Reuter CJ, Maupin-Furlow JA. Analysis of proteasome-dependent proteolysis in Haloferax volcanii cells, using short-lived green fluorescent proteins. Appl Environ Microbiol2004 Dec;70(12):7530-8.

[76] Reuter CJ, Uthandi S, Puentes JA, Maupin-Furlow JA. Hydrophobic carboxy-terminal residues dramatically reduce protein levels in the haloarchaeon Haloferax volcanii. Microbiology2010 Jan;156(Pt 1):248-55.

[77] Henche AL, Koerdt A, Ghosh A, Albers SV. Influence of cell surface structures on crenarchaeal biofilm formation using a thermostable green fluorescent protein. Environ Microbiol2012 Mar;14(3):779-93.

[78] Cubitt AB, Woollenweber LA, Heim R. Understanding structure-function relationships in the Aequorea victoria green fluorescent protein. Methods Cell Biol1999;58:19-30.

[79] Patterson GH, Knobel SM, Sharif WD, Kain SR, Piston DW. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. Biophys J1997 Nov;73(5):2782-90.

[80] Forterre P, Elie C, Kohiyama M. Aphidicolin inhibits growth and DNA synthesis in halophilic arachaebacteria. J Bacteriol1984 Aug;159(2):800-2.

[81] Digman MA, Dalal R, Horwitz AF, Gratton E. Mapping the number of molecules and brightness in the laser scanning microscope. Biophys J2008 Mar 15;94(6):2320-32.

[82] Bakker ST, van de Vrugt HJ, Rooimans MA, Oostra AB, Steltenpool J, Delzenne-Goette E, et al. Fancm-deficient mice reveal unique features of Fanconi anemia complementation group M. Hum Mol Genet2009 Sep 15;18(18):3484-95.

[83] Crismani W, Girard C, Froger N, Pradillo M, Santos JL, Chelysheva L, et al. FANCM limits meiotic crossovers. Science2012 Jun 22;336(6088):1588-90.

[84] Knoll A, Higgins JD, Seeliger K, Reha SJ, Dangel NJ, Bauknecht M, et al. The Fanconi anemia ortholog FANCM ensures ordered homologous recombination in both somatic and meiotic cells in Arabidopsis. Plant Cell2012 Apr;24(4):1448-64.

[85] Lorenz A, Osman F, Sun W, Nandi S, Steinacher R, Whitby MC. The fission yeast FANCM ortholog directs non-crossover recombination during meiosis. Science2012 Jun 22;336(6088):1585-8.

[86] Wagner M, van Wolferen M, Wagner A, Lassak K, Meyer BH, Reimann J, et al. Versatile Genetic Tool Box for the Crenarchaeote Sulfolobus acidocaldarius. Front Microbiol2012;3:214.

[87] Zhang C, Tian B, Li S, Ao X, Dalgaard K, Gokce S, et al. Genetic manipulation in Sulfolobus islandicus and functional analysis of DNA repair genes. Biochem Soc Trans2013 Feb 1;41(1):405-10.

université PARIS-SACLAY

ÉCOLE DOCTORALE
INTERFACES
Approches Interdisciplinaires : Fondements, Applications et Innovation

Titre : Dynamique cellulaire des protéines de la réplication chez l'archée halophile *Haloferax volcanii*

Mots clés : Réplication de l'ADN, Archées, *Haloferax volcanii,* Microscopie de fluorescence, Protéine fluorescente verte (GFP)

Résumé : Ce travail de thèse porte sur l'étude de la réplication chez les archées, qui constituent le troisième domaine du vivant avec les bactéries et les eucaryotes. L'organisme modèle que nous avons utilisé est l'archée halophile *Haloferax volcanii* car les outils génétiques disponibles permettent d'exprimer des protéines fusionnées à la Protéine Fluorescente Verte (GFP) dans cet organisme mésophile et aérobe et ainsi de localiser les protéines d'intérêt dans des cellules vivantes. Nous nous sommes ainsi intéressés à la localisation cellulaire de quatre protéines de la réplication qui ont été fusionnées à la GFP et exprimées sous contrôle de leur propre promoteur : (i) la protéine 'Flap Endonuclease 1' (FEN1), qui intervient dans la maturation des fragments d'Okazaki, (ii) la protéine 'Origin Recognition Complex' (ORC1) impliquée dans la reconnaissance des origines de réplication, (iii) la protéine 'Proliferating Cellular Nuclear Antigen' (PCNA), anneau de processivité des ADN polymérases réplicatives, et (iv) la protéine de fixation à l'ADN simple-brin 'Replication Protein A' (RPA2) essentielle à la réplication chez *H. volcanii.*

Seule la protéine PCNA n'a pu être exprimée en fusion avec la GFP, suggérant que la protéine fusion n'est pas fonctionnelle. GFP::Orc1 et GFP::Fen1 ont été exprimées dans la cellule mais ne présentent pas de localisation spécifique reflétant un rôle de ces protéines dans la réplication de l'ADN. En revanche des foyers de fluorescence de la protéine fusion GFP::Rpa2 ont été observés, dont le nombre augmente significativement dans des cellules exposées à l'aphidicoline, drogue inhibant la synthèse de l'ADN et induisant ainsi un stress réplicatif. Cependant une localisation différente de la protéine GFP::Rpa2 a été observée lorsque les cellules sont exposés à la phléomycine, qui induit notamment des cassures double-brin de l'ADN. Dans ces cellules, GFP::Rpa2 forme un foyer de fluorescence massif qui colocalise avec l'ADN compacté dans la grande majorité des cellules observées. Nos résultats suggèrent donc que la localisation spécifique observée pour GFP::Rpa2 reflète son rôle dans la réparation de l'ADN et/ou le redémarrage des fourche de réplication arrêtées.

Title : Cellular dynamic of the replication proteins in the halophilic archaeon *Haloferax volcanii*

Keywords : DNA replication, archaea, *Haloferax volcanii,* Fluorescence microscopy, GFP (Green fluorescent protein).

Abstract : The aim of this thesis project was to improve our understanding of DNA replication in archaea, the third domain of life with bacteria and eukarya. The model organism chosen for these studies is the halophilic archaea *Haloferax volcanii*, a mesophilic aerobe for which genetics tools allow studying in living cells the localization of proteins fused to the Green Fluorescent protein (GFP). Four proteins involved in DNA replication were fused to the GFP and expressed under the control of their own promoter: (i) the 'Flap Endonuclease 1' (FEN1), involved in Okazaki fragments maturation, (ii) the 'Origin Recognition Complex' (ORC1), involved in DNA replication origin recognition, (iii) the 'Proliferating Cellular Nuclear Antigen' (PCNA), processivity factor of replicative DNA polymerases, and (iv) the 'Replication Protein A' (RPA2), single-stranded DNA binding protein essential for DNA replication in *H. volcanii*. Only the PCNA fusion to the GFP was not successful, suggesting that the GFP hinders essential roles of PCNA in DNA replication. Fen1 and Orc1 were successfully fused to the GFP and expressed in living cells, but specific localization in cells related to growth phase, reflecting different replication dynamics, were not observed. In contrast, we could observed fluorescent foci formed by the fully functional GFP::Rpa2 protein that actively responded to DNA damage in *H. volcanii* cells. The number of these fluorescent foci per cell was constant during cell growth but it significantly increased in cells exposed to aphidicoline, which inhibits DNA synthesis during replication. When cells were treated with phleomycine, a DNA damaging agent mainly causing double-strand breaks, formation of a massive fluorescent focus coinciding with DNA compaction was observed. Our results suggest that the specific cellular localization of GFP::Rpa2 observed reflects Rpa2 roles in DNA repair and/or DNA replication fork restart.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France