



HAL
open science

Exploration du réseau d'interactions impliqué dans la maintenance génomique de l'Archaea hyperthermophile

Pyrococcus abyssi

Pierre-François Pluchon

► **To cite this version:**

Pierre-François Pluchon. Exploration du réseau d'interactions impliqué dans la maintenance génomique de l'Archaea hyperthermophile *Pyrococcus abyssi*. Biologie moléculaire. Université de Bretagne occidentale - Brest, 2012. Français. NNT : 2012BRES0070 . tel-01148105

HAL Id: tel-01148105

<https://theses.hal.science/tel-01148105>

Submitted on 4 May 2015

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

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



université de bretagne
occidentale



THÈSE / UNIVERSITÉ DE BRETAGNE OCCIDENTALE

sous le sceau de l'Université européenne de Bretagne

pour obtenir le titre de

DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

Mention : Biologie Moléculaire & Cellulaire

École Doctorale: EDSM

présentée par

Pierre-François Pluchon

Préparée au Laboratoire de Microbiologie
des Environnements Extrêmes - UMR6197

Exploration du réseau d'interactions impliqué dans la maintenance génomique de l'Archaea hyperthermophile *Pyrococcus abyssi*

Thèse soutenue le 18 décembre 2012

devant le jury composé de :

Docteur Hannu MYLLYKALLIO

Laboratoire d'Optique et Biosciences. Ecole Polytechnique.
Rapporteur.

Docteur Laurent TERRADOT

Biologie structurale des complexes macromoléculaires
bactériens, IBCP. Rapporteur

Professeur Mohammed JEBBAR

Laboratoire de Microbiologie des Environnements
Extrêmes, IUEM. Examineur.

Professeur Viannet PICHERAU

Laboratoire des Sciences de l'Environnement Marin,
IUEM. Examineur.

Docteur Didier FLAMENT

IFREMER, Laboratoire de Microbiologie des
Environnements Extrêmes. Directeur de thèse.

Ce travail a été réalisé au sein du Laboratoire de Microbiologie des Environnements Extrêmes, affilié à l'Ifremer, au CNRS et à l'Université de Bretagne Occidentale. Je souhaite remercier chaleureusement tous les membres du LMEE pour leur accueil et tout particulièrement l'équipe maintenance génomique au sein de laquelle j'ai passé 3 années très agréables et très enrichissantes. Merci également à toute ma famille et tous mes amis. LVS.

1	Les Archaea, troisième branche du vivant	8
1.1	Découverte des Archaea.....	8
1.2	La cellule Archaea.....	10
1.3	Biotopes.....	11
1.4	Les écosystèmes hydrothermaux profonds.....	13
1.5	<i>Pyrococcus Abyssii</i> , modèle de la maintenance génomique	14
2	Maintenance génomique.....	17
2.1	Réplication de l'ADN.....	17
2.1.1	Initiation	19
2.1.1.1	Reconnaissance de l'origine de réplication	19
2.1.1.2	Ouverture de la double hélice : l'hélicase réplicative.....	20
2.1.1.3	Le complexe GINS.....	21
2.1.1.4	Synthèse de l'amorce : L'ADN Primase.....	23
2.1.1.5	Les protéines de liaison à l'ADN simple brin (SSBs) : SSB et RPA.....	24
2.1.2	Synthèse du brin complémentaire	27
2.1.2.1	ADN Polymérase.....	27
2.1.2.2	PCNA, <i>maestro</i> de la réplication.....	31
2.1.2.3	Facteur de chargement du clamp.....	34
2.1.3	Maturation	36
2.1.3.1	Flap endonucléase 1.....	36
2.1.3.2	RNase H II.....	37
2.1.3.3	ADN Ligase.....	37
2.2	Réparation de l'ADN	39
2.2.1	Réparation par réversion du dommage	40
2.2.2	Réparation par Excision de Nucléotide NER.....	40
2.2.3	Réparation par Excision de Base BER	43
2.2.4	Réparation des mésappariements (MisMatch Repair, MMR).....	44
2.2.5	Réparation après une cassure double brin	46
2.2.5.1	Recombinaison Homologue	47

2.2.6	Non Homologous End Joining (NHEJ).....	50
2.3	Régulation de la topologie de l'ADN.....	50
3	Biologie des systèmes, La cellule micromachine et microordinateur.....	52
3.1	Biologie des systèmes.....	52
3.2	Des protéines complexes et des complexes protéiques.....	54
3.2.1	Réseau d'interactions protéine protéine	55
3.2.2	Propriétés des PPIN.....	56
3.2.2.1	Orientation du réseau	56
3.2.2.2	Degré, distribution des degrés	56
3.2.2.3	Path Length.....	56
3.2.2.4	Clustering coefficient.....	56
3.2.3	Topologie d'un réseau.....	57
3.2.4	De la topologie à la fonction	58
3.2.5	Techniques	59
3.3	PPIN de la maintenance génomique	60
4	Présentation de l'étude	64
5	Résultats et Discussion	66
5.1	Production et purification des protéines appâts.....	66
5.2	Mise au point de la purification d'affinité	70
5.3	Préparation de l'extrait cellulaire	71
5.4	Mise au point du protocole de purification d'affinité des complexes.....	72
5.5	Mise au point des conditions d'élutions.....	75
5.6	Analyse en spectrométrie de masse.....	78
5.7	Filtre des données	80
5.8	Partenaires spécifiques de 23 protéines de la maintenance génomique.....	87
5.9	Analyse topologique du réseau d'interaction.....	97
5.10	Présentation de l'article	102
5.11	Resultats complementaires sur l'interaction RPA-RNAP	140
5.12	RPA n'influence pas l'initiation de la transcription, in vitro	140

5.13	RPA n'influence pas la terminaison de la transcription, in vitro.....	141
6	Réseau Etendu, la partie immergée de l'iceberg.....	144
6.1.1	Problématique.....	144
6.1.2	Nucleotide Excision Repair	146
6.1.3	RcrA et ses complexes.....	149
7	Conclusions et perspectives.....	151
8	Matériels et Méthodes	161
8.1	Culture continue de <i>P. abyssi</i>	161
8.2	Préparation de l'extrait cellulaire	163
8.3	Clonage des gènes cibles	163
8.4	Production et purification des protéines appâts.....	164
8.5	Purification d'affinité des complexes protéiques de <i>P. abyssi</i>	164
8.6	Spectrométrie de masse.....	165
8.7	Identification des interactions spécifiques.....	165
8.8	Construction du réseau et analyse topologique.....	166
9	Bibliographie	167
10	Données Supplémentaires.....	167

Liste des figures :

Figure 1: Arbre phylogénétique non - enraciné du vivant basé sur les séquences ARN ribosomales.	9
Figure 2: Arbre phylogénétique Bayésien non enraciné du domaine Archaea.....	9
Figure 3 : Répartition des protéines des systèmes informationnels chez les Archaea.	11
Figure 4 : Diagramme température / pH et dénomination des organismes adaptés.	12
Figure 5 : Hyperthermophiles et radiorésistance.....	13
Figure 6: Les sources hydrothermales profondes	14
Figure 7: Le chromosome de <i>P. abyssi</i> , une mosaïque génétique.	15
Figure 8: Comparaison des réplisomes Bactériens, Eucaryotes et Archaea.	18
Figure 9: Mécanismes possibles pour l'activité hélicase de MCM <i>in vivo</i>	21
Figure 10: Protéines et complexes à l'initiation.....	22
Figure 11: Complexe CMG Eucaryote et Archaea	23
Figure 12: Schéma de l'organisation des protéines SSB et RPA.....	25
Figure 13: Structure de l'ADN Polymérase B du bactériophage RB69 en complexe avec l'ADN.	29
Figure 14: ADN polymérase B de <i>Pyrococcus abyssi</i> liée à un ADN double brin.....	30
Figure 15: Structure cristalline des « sliding clamp » des trois domaines du vivant.....	32
Figure 16: Partenaires du PCNA Eucaryote.	33
Figure 17: Structure tridimensionnelle du complexe clamp-clamp loader du bactériophage T4.....	34
Figure 18: Modèle du chargement du clamp par son facteur de chargement.	35
Figure 19 : Modèle de la maturation des fragments d'Okazaki chez les Archaea.	36
Figure 20 : Fourches de réplication.....	38
Figure 21: Les grandes voies de réparation de l'ADN et types de lésions associés.....	39
Figure 22 : Réparation des résidus O6-méthylguanine par la protéine MGMT	40
Figure 23 : Protéines du NER Archaea et leur distribution	42
Figure 24: Base Excision Repair	43
Figure 25 : MMR chez les bactéries et les eucaryotes.	45
Figure 26: Dommages provoquant un DSB et voies de réparation.....	47
Figure 27 : Modèle de résection des DSB chez les Archaea.....	48
Figure 28: Protéines et voie de réparation par recombinaison homologue chez les Archaea.	50

Figure 29 : Le système cellulaire	53
Figure 30 : Structure du PPINet essentialité des protéines de <i>S. cerevisiae</i>	55
Figure 31 : Mesures élémentaires d'un graph.	57
Figure 32 : Modèles de réseaux.	58
Figure 33 : PPINde la réponse au dommages de l'ADN chez les Eucaryotes.	61
Figure 34 : PPIN des protéines dela réponse aux dommages de l'ADN chez les Bactéries.....	61
Figure 35 : PPIN de la réplication chez <i>A. fulgidus</i>	62
Figure 36 : PPIN de la réplication de <i>T. kodakarensis</i>	62
Figure 37 : PPIN de la maintenance génomique de <i>P. abyssi</i> proposé par J. Briffotiaux en 2008.	65
Figure 38 : Schéma général du protocole expérimental.	67
Figure 39 : Exemples de purifications des protéines appâts.....	69
Figure 40 : Mise au point des conditions de tampon.....	70
Figure 41 : Effet des nucléases sur l'extrait cellulaire de <i>P. abyssi</i>	72
Figure 42 : Purifications d'affinité des complexes, l'exemple PCNA.....	73
Figure 43: Effet de la déplétion sur un extrait cellulaire utilisé en AP.	74
Figure 44 : Conditions d'élution des complexes et intensité du signal en MS.....	76
Figure 45 : Nombre d'espèces protéiques identifiés selon le mode d'élution.....	78
Figure 46 : Schématisation du processus d'identification des protéines en MS.....	79
Figure 47 : Profil d'identification des protéines des témoins négatifs indépendants.....	81
Figure 48 : Scores Mascot de l'AP-MS PCNA.....	86
Figure 49 : Paramètres généraux du réseau.	98
Figure 50: Clusters au sein du réseau haute confiance.....	99
Figure 51 : Groupes de protéines optimaux sur la base du coefficient mutuel.	100
Figure 52: Epine dorsale du réseau haute confiance.	101
Figure 53 : Effet de RPA sur l'initiation de la transcription <i>in vitro</i>	140
Figure 54 : Effet de RPA sur la terminaison de la transcription.	141
Figure 55: modèle de l'interaction entre RPA et la RNA polymérase	142
Figure 56: Réseau étendu et position des clusters haute confiance.....	144
Figure 57: Acteurs établis du NER Archaea dans le réseau.....	148

Figure 58:Pab1036 nouvelle nucléase associée à la réparation ?.....	148
Figure 59 : RcrA et ses complexes.....	150
Figure 60: Union des clusters et de l'épine dorsale du réseau	151
Figure 61: Modèle de la synthèse des brin complémentaire par l'ADN primase lors de la recombinaison.	154
Figure 62: L'ARN polymérase dans le réseau étendu.....	156
Figure 63: Schématisation de la collaboration entre MCM et XPD.....	157
Figure 64: Modèle de l'action de pab0431	158
Figure 65 : Suivi de la concentration cellulaire lors de la culture continue	162
Figure 66 : Représentation schématique du système expérimental de culture en bioréacteur gazlift.....	162

Liste des tableaux:

Tableau 1: Protéines de la maintenance génomique interagissant avec le RPA Eucaryote.	25
Tableau 2: Distribution, Fonction et fidélité des différentes familles d'ADN Polymérases.	28
Tableau 3: Protéines interagissant avec le PCNA chez les Archaea.	33
Tableau 4 : Protéines du MMR et fonctions chez les Bactéries et les Eucaryotes	45
Tableau 5 : Gènes Archaea codants pour des homologues des protéines Mut. (Grogan 2004).....	46
Tableau 6 : Protéines de fusion utilisées en appâts dans les expériences d'AP-MS.	69
Tableau 7 : Protéines détectées dans les témoins négatifs et dans l'extrait brut.	83
Tableau 8 : Partenaires de PCNA identifiées en AP-MS.	85
Tableau 9 : Resultats des AP-MS, interactions haute confiance.....	87
Tableau 10 : Résultats de l'algorithme « operon » :	145

Liste des abreviations

ADN : Acide DéoxyRiboNucléique

ARN : Acide Ribonucléique

ATP : Adénosine Triphosphate

BER : Réparation par Excision de Base

dNTP : desoxyribo Nucleotide Tri Phosphate

DSB : Cassure double brin

EDTA : acide éthylènediaminotétraacétique sel disodique

HA: Archaea Hyperthermophile

HGT; Transfert Horizontal de Gène

ID Iniprot: Identifiant dans la base de données Uniprot

IPTG : Isopropyl- β -D-galactoside

LB : Luria-Bertani

LUCA : Last Universal Common Ancestor

MMR : Réparation des mésappariements

MS : Spectrométrie de Masse

NER : Réparation par Excision de Nucléotides

NHEJ: Non Homologous End Joining

ORC : Complexe de Reconnaissance de l'Origine

ORF : cadre ouvert de lecture)

pb: paires de bases

PBS : Phosphate Buffered Saline

PCR : Polymerase Chain Reaction

PEG: Poyléthylène Glycols

PIP : PCNA Interacting Protein

PPIN: Réseau d'interaction protéine -protéine

RH ou HR: Recombinaison Homologue

SPR : Résonance Plasmonique de Surface

SR: Significance Ratio

SSB : Protéine de liaison à l'ADN simple brin

TCR: Réparation Couplée à la Transcritpion

UV : Ultraviolets

1 LES ARCHAEA, TROISIEME BRANCHE DU VIVANT

1.1 DECOUVERTE DES ARCHAEA

Depuis les débuts de la microbiologie, notamment initiée par les travaux de Louis Pasteur et Robert Koch, la classification taxonomique des procaryotes était basée sur des critères phénotypiques tels que la morphologie, la physiologie et la pathogénicité. Cette classification ne tenait pas alors compte des relations évolutives complexes entre les différents groupes taxonomiques.

En 1977, les microbiologistes Woese et Fox utilisent la séquence de l'ARN ribosomal (ARNr) comme un marqueur de l'évolution et identifient deux microorganismes méthanogènes phylogénétiquement distincts des Bactéries et des Eucaryotes (Woese and Fox 1977). La comparaison de leurs séquences d'ARN 16S ribosomiaux indique que ces deux espèces font partie du même phylum mais que celui-ci est différent de celui des procaryotes connus à l'époque. Ce ne sont pas des Bactéries mais des Archaea (initialement Archaeobacteria). En augmentant le nombre d'organismes étudiés ils proposent la même année une nouvelle classification du vivant : eubactérie, archéobactérie et Eucaryote (Woese and Fox 1977). En 1990, C. Woese et ses collaborateurs proposent finalement un nouvel arbre du vivant composé de 3 domaines : Archaea, Bacteria et Eukarya (Woese, Kandler et al. 1990). La phylogénie est une science en perpétuelle évolution mais, à ce jour, la découverte réalisée par ces pionniers de la microbiologie constitue toujours le premier niveau de classification du vivant (Pace 2009; Joyce 2012). La phylogénie définit ainsi trois lignées cellulaires ayant un ancêtre commun. La définition du dernier ancêtre universel commun (« Last Universal Common Ancestor », LUCA) est bien évidemment un enjeu primordial pour la compréhension de l'origine de la vie (Ledford 2012). L'enracinement profond de l'arbre phylogénétique du vivant suggère une divergence cellulaire très ancienne. La séparation des Bactéries de la branche partagée par les Archaea et les Eukarya représenterait le point d'embranchement le plus ancien.

Afin de comprendre l'origine et la filiation des 3 lignées cellulaires il est primordial de comprendre la phylogénie des Archaea. Les conditions de vies inhabituelles et la complexité des communautés a souvent rendu difficile la culture et l'isolement des Archaea. L'amélioration des techniques de cultures (Ben Hania, Ghodbane et al. 2011; Yan, Haaijer et al. 2012) et l'émergence des techniques de séquençage haut débit ont permis une description plus complète des relations phylétiques au sein des Archaea. Aujourd'hui, plus de 100 génomes Archaea ont été séquencés (Brochier-Armanet, Forterre et al. 2011) (source GOLD, génome online Database : 171 complet, 244 incomplet, 10 en projet). Initialement, deux groupes avaient été identifiés : Crenarchaeota et Euryarchaeota. Depuis 3 autres groupes supplémentaires ont complété la classification des Archaea ; Thaumarchaeota, Aigarchaeota et Korarchaeota (Figure 2).

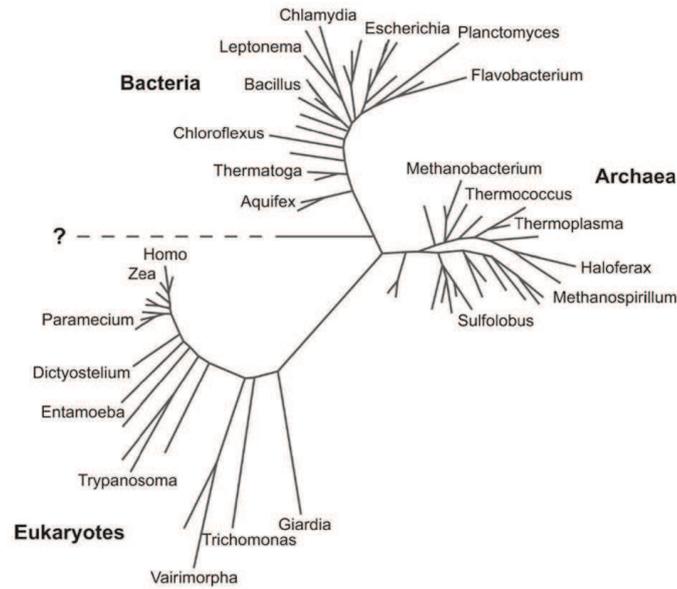


Figure 1: Arbre phylogénétique non - enraciné du vivant basé sur les séquences ARN ribosomales.

(Joyce 2012)

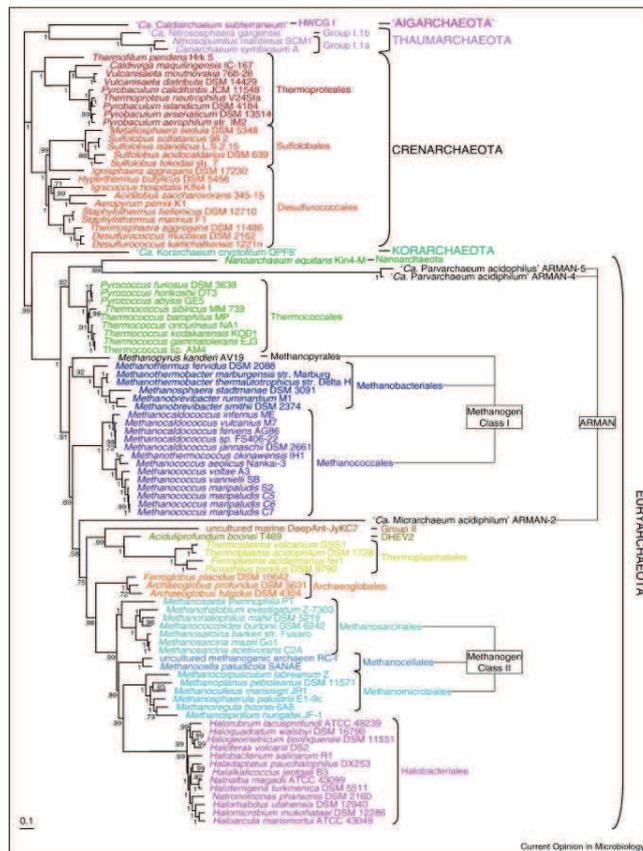


Figure 2: Arbre phylogénétique Bayésien non enraciné du domaine Archaea.

L'analyse est basée sur la concaténation de 57 protéines ribosomales présentes dans au moins 89 des 99 génomes comparés. (Brochier-Armanet, Forterre et al. 2011)

1.2 LA CELLULE ARCHAEA

Les Archaea sont des organismes unicellulaires dépourvus de membranes nucléaires. Leurs tailles sont variables mais oscillent généralement entre 0.5 et 2 μ m. Elles arborent des formes diverses et variées au même titre que les Bactéries et elles possèdent également des structures flagellaires utilisées pour la motilité (Albers and Pohlschröder 2009). Les cellules Archaea présentent également pili, cannulae, hami, ou encore « bindosome » (Ng, Zolghadr et al. 2008). Dans leur environnement certaines archaea forment d'importants biofilms. Chez les Archaea méthanogènes ANME-1 le biofilm est associé à l'oxydation anaérobie du méthane. Les Archaea se différencient en plusieurs points fondamentaux des Eucaryotes et des Bactéries en termes métaboliques ou structuraux (voir la composition de la paroi à la fin du chapitre) mais cette différence s'exprime nettement au niveau de la composition du génome. Les chromosomes des archaea sont libres et circulaires et multi-copies (Malandrin, Huber et al. 1999) comme chez les Bactéries. Beaucoup d'Archaea possèdent au moins un plasmide. On retrouve dans les génomes Archaea des séquences codantes pour des homologues de protéines bactériennes mais également des homologues de protéines eucaryotes et évidemment des protéines spécifiques aux Archaea (Graham, Overbeek et al. 2000). La composition du génome Archaea est parfois comparée à une mosaïque (Kelman and White 2005). La distribution des gènes au sein de cette mosaïque est corrélée aux fonctions, conservées en systèmes. De manière générale, les processus dits « informationnels », réplication, transcription et traduction, sont assurés par des homologues de protéines eucaryotes alors que les homologues bactériens sont en charge des processus « opérationnels » (Kelman and White 2005; Makarova, Yutin et al. 2010). La présence ou l'absence d'homologues de protéines Eucaryotes est significativement variable selon la lignée phylogénétique considérée (Figure 3). Même si 20% à 30% des gènes Archaea ont été acquis par transfert horizontal (« Horizontal Gene Transfert », HGT) (Koonin, Makarova et al. 2001), la distribution éparse observée ne peut pas provenir uniquement de HGT entre lignées distinctes, mais elle suggère plutôt l'existence d'un ancêtre commun aux caractéristiques plus complexes. Cette hypothèse est supportée par la découverte d'un nombre significatif de protéines ribosomales ayant une distribution différente selon la lignée Archaea considérée (Desmond, Brochier-Armanet et al. 2011). Les phénomènes de HGT des protéines ribosomales étant des événements rares, leur distribution au sein des Archaea penche en faveur du scénario d'un organisme primaire plus complexe dont une partie du système a été perdue au cours de l'évolution (Desmond, Brochier-Armanet et al. 2011).

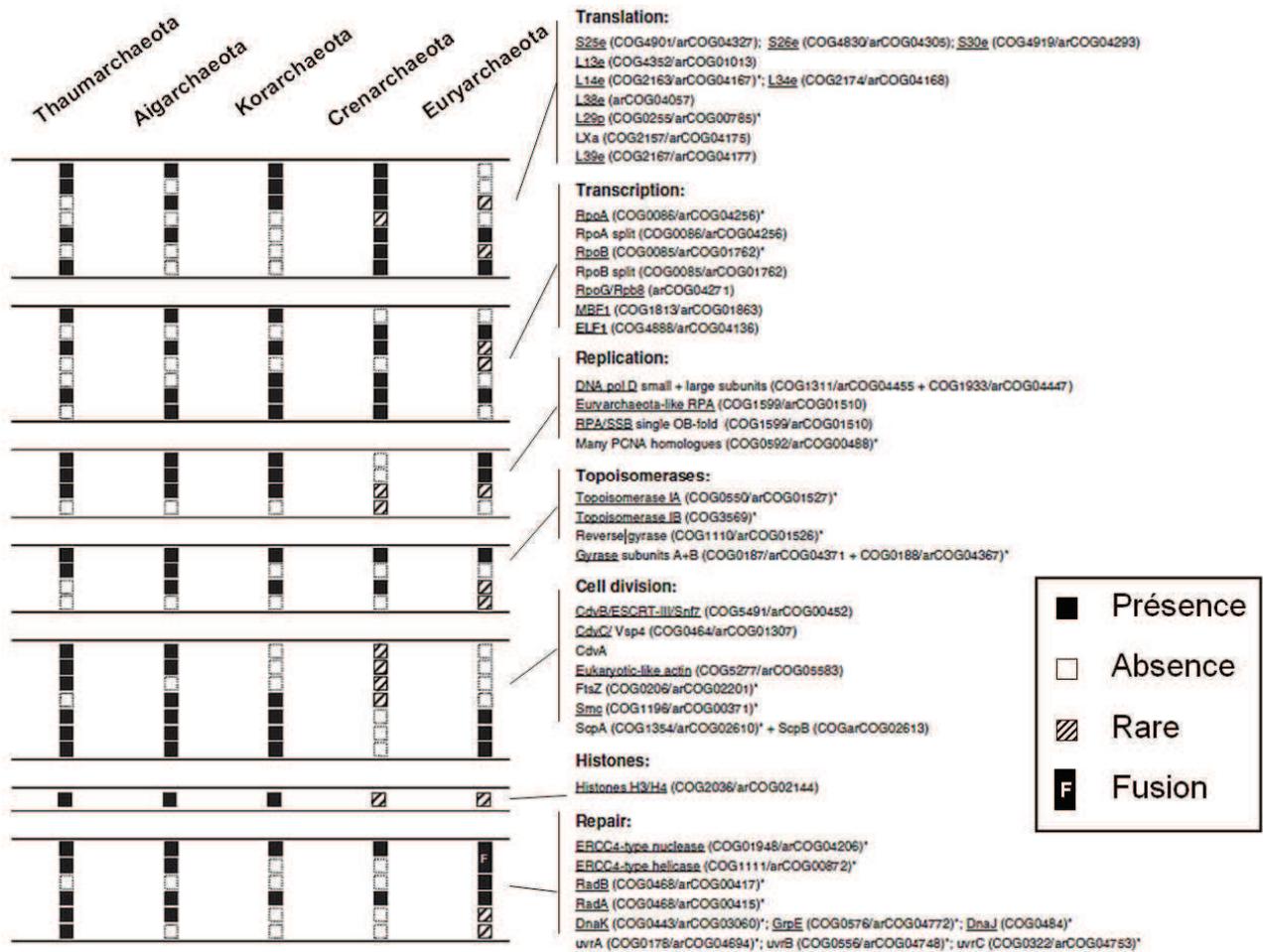


Figure 3 : Répartition des protéines des systèmes informationnels chez les Archaea.

(Brochier-Armanet, Gribaldo et al. 2012).

Une autre spécificité des Archaea est la structure et la composition chimique unique de leurs parois. Les parois Archaea sont dépourvues de peptidoglycanes mais composées d'acide N-acetylmuramique. Certaines Archaea possèdent une paroi épaisse constituée de polysaccharides et de glycoprotéines. Le type de paroi le plus répandu chez les Archaea est la couche de surface para-cristalline constituée de sous unités protéiques présentant une symétrie hexagonale, appelé S-layer (Kandler and König 1998; Arbing, Chan et al. 2012). Les Bactéries et les Eucaryotes ont des membranes composées essentiellement de lipides glycérol-ester, tandis que les membranes Archaea sont des lipides glycérol-éther, cette différence de liaisons chimiques offrirait notamment une résistance élevée aux attaques hydrolytiques se déroulant à haute température (Stetter 1999). La membrane des Archaea a été récemment sujette à revue (Albers and Meyer 2011).

1.3 BIOTOPES

Les environnements dans lesquels on retrouve des Archaea sont très variés. Des espèces Archaea ont été identifiées dans les eaux océaniques Arctique et Pacifique. Dans les sols et rivières ainsi que dans les tubes digestifs de mammifères (DeLong 1998) Les Archaea ont aussi la particularité de coloniser des niches écologiques aux paramètres chimiques et physiques extrêmes. Il existe sur Terre des milieux dont les

paramètres physico-chimiques sont incompatibles avec la vie de la plupart des organismes classiquement décrits. La température, la salinité, le pH, la pression, les radiations ou l'absence d'oxygène sont des paramètres qui, lorsqu'ils atteignent des valeurs extrêmes, perturbent les structures et les fonctions cellulaires et les conduisent à la mort (Figure 4). Malgré leur inhospitalité, les milieux extrêmes se sont révélés accueillir de nombreuses espèces vivantes parmi lesquelles les Archaea prennent régulièrement une part significative. Les Archaea dites halophiles, notamment *Haloferax volcanii*, vivent dans les eaux saturées en sels, par exemple celles de la mer morte ou des grands lacs salés américains (Hartman, Norais et al. 2010). Des espèces psychrophiles telles que *Methanogenium frigidum* ont été isolées à des températures proches de 0°C (Cavicchioli 2006) et des hyperthermophiles (*Pyrococcus abyssi*, *P. furiosus* etc.) aux alentours 100°C (Gunbin, Afonnikov et al. 2009). Une archaea (strain 121) parente de *Pyrodictium occultum* et *Pyrobaculum aerophilum*, possède une température optimale de croissance de 121° et elle survit à un passage en autoclave (2 heures à 130°C) (Kashefi and Lovley 2003). *Sulfolobus acidocaldarius* est acidophile et thermophile et vit dans des environnements à la fois très acides (pH<2) et très chauds (Brock, Brock et al. 1972). *Thermococcus acidaminovorans*, alcalophile, possède un optimal de croissance à pH 9 (R., M. et al. 1998). Les Methanosarcinates du permafrost et les espèces *Pyrococcus* des sources hydrothermales profondes (dont *P. abyssi*) sont résistantes aux radiations gamma à des niveaux à peine inférieurs à ceux tolérés par *Deinococcus radiodurans* (Gérard, Jolivet et al. 2001) (Figure 5).

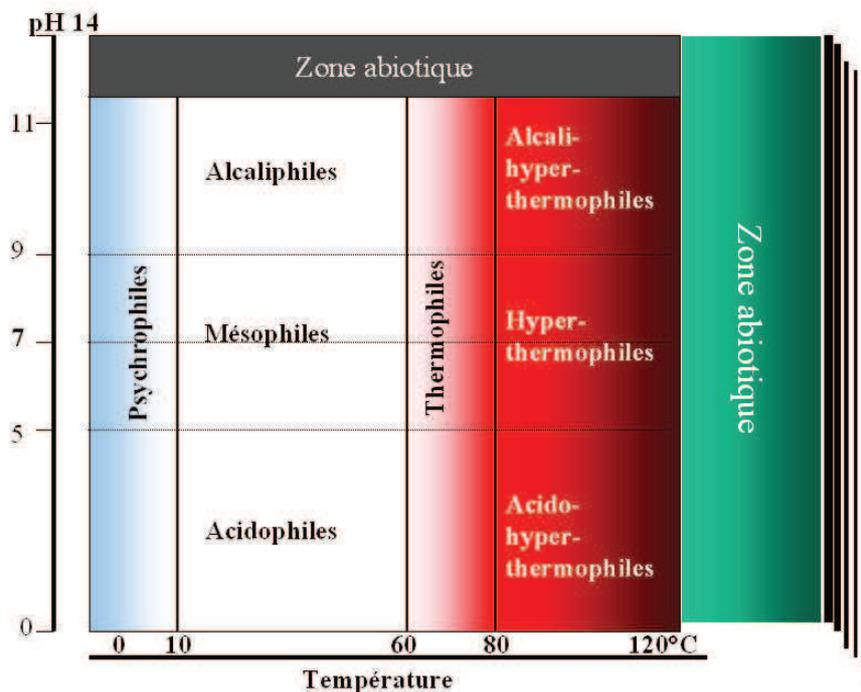


Figure 4 : Diagramme température / pH et dénomination des organismes adaptés.

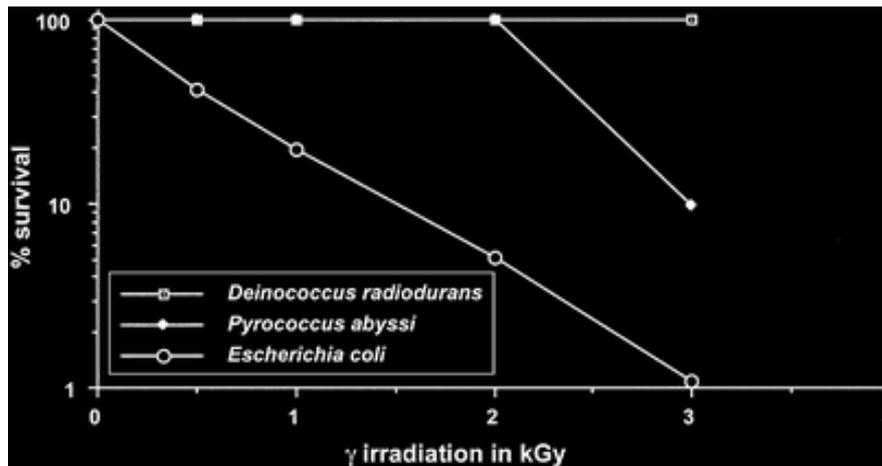


Figure 5 : Hyperthermophiles et radiorésistance.

Courbes de survies de *P. abyssi*, *D. radiodurans* et *E. coli* après exposition aux rayons gamma 137Cs.

Les résistances des hyperthermophiles aux radiations sont probablement liées au recouvrement du spectre des dommages à l'ADN provoqués par les hautes températures et les radiations (Gérard, Jolivet et al. 2001). Les Archaea des grands fonds marins supportent également une pression qui augmente proportionnellement à la profondeur et qui peut atteindre des valeurs de plusieurs dizaines de méga Pascal (25 MPa à -2500m). Certaines espèces Archaea dont la croissance nécessite une pression considérable sont dites piézophiles, comme par exemple *Pyrococcus yayanosii* (Birrien, Zeng et al. 2011). En terme de ratio température/pression, la limite haute de la vie a été observée jusqu'à 113°C à 52 méga Pascal, ce « record » est détenu par *Pyrococcus* CH1 (Zeng, Birrien et al. 2009).

Ces différentes conditions extrêmes ne le sont en fait que d'un point de vue anthropique puisque les organismes extrémophiles sont également extrêmement adaptés aux conditions de pression, de température ou de salinité. Cette adaptation est telle que ces conditions extrêmes sont le plus souvent indispensables au développement de l'organisme extrémophile. Les Archaea hyperthermophiles (HA) forment des communautés complexes dans des eaux réchauffées par des phénomènes géothermaux ou volcaniques. Au niveau des sources hydrothermales profondes, elles se développent dans des eaux dont la température excède le point d'ébullition à pression atmosphérique. Les HA ne croissent pas en dessous de 50°C mais sont en revanche capables de survivre à des températures ambiantes voire négatives (-140°C).

1.4 LES ECOSYSTEMES HYDROTHERMAUX PROFONDS

En 1977, une équipe de géologues dirigée par John Corliss, découvre au large de l'Equateur, à près de 2 600 m de profondeur, une forte activité volcanique et d'imposantes structures minérales aux allures de cheminées d'où s'échappent un fluide surchauffé. Autour de ces structures, ils observent à leur grande surprise une faune luxuriante (Corliss, Dymond et al. 1979). Des campagnes de plongée ont ultérieurement découvert des sites hydrothermaux semblables au niveau des dorsales des océans Atlantique, Pacifique et Indien. Les cheminées hydrothermales sont des structures dynamiques qui expulsent des fluides acides, riches en gaz et ions

métalliques (Hekinian, Fevrier et al. 1983). Ces fluides sont issus de la surpression engendrée par le réchauffement de l'eau de mer qui s'est infiltrée à proximité du magma, dont la température atteint 1 200°C. (Figure 6). Les sites hydrothermaux sont caractérisés par des conditions de hautes pressions, de forts gradients de température (2°C à 400 °C), de fortes teneurs en éléments toxiques (métaux lourds, méthane, sulfures, etc.), (Ishibashi, #160 et al. 1994) l'absence totale de lumière et de faibles apports nutritionnels essentiellement liés à la sédimentation de la matière organique des couches superficielles de l'océan.

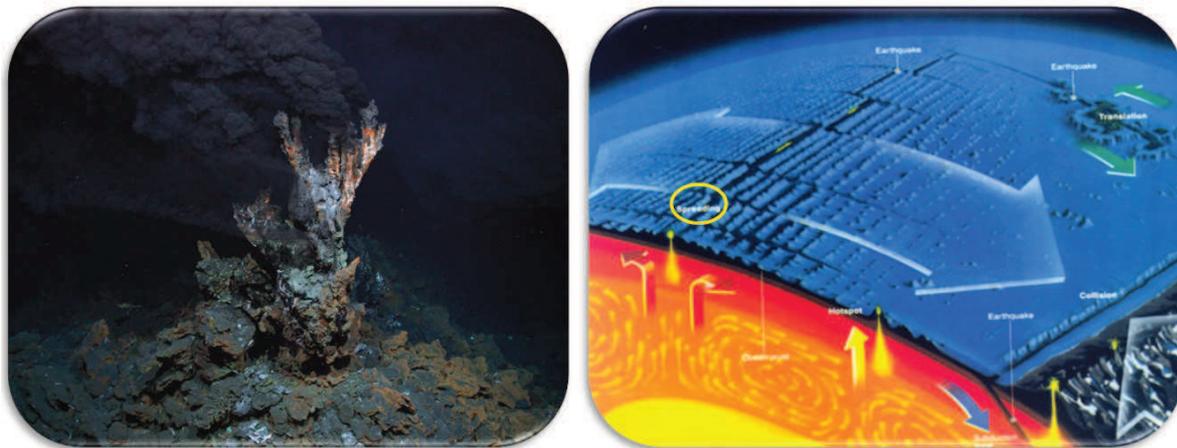


Figure 6: Les sources hydrothermales profondes

Cheminées hydrothermales (©University of Bremen) et schéma des phénomènes géologiques à l'origine de leur formation

Ces espaces que l'on imaginait initialement désertiques constituent en réalité des oasis pour la vie. Ils abritent une communauté complexe d'organismes, Archaea, Bactéries et Eucaryotes. Une des caractéristiques majeures de ces écosystèmes est l'absence de production primaire photosynthétique. Celle-ci est en fait assurée par des organismes autotrophes capables d'utiliser l'énergie de l'oxydation pour assimiler le dioxyde de carbone en molécule organique. Les sources hydrothermales profondes comportent des milieux extrêmement divers. Depuis l'eau de mer environnante et les sédiments superficiels jusqu'aux zones chaudes des fumeurs, il existe un gradient thermique (2 – 400°C) qui permet le développement d'une diversité d'organismes psychrophiles, mésophiles, thermophiles et hyperthermophiles. Il existe également une radioactivité naturelle environ 100 fois plus élevée que la radioactivité en surface. Cette radioactivité est causée par la présence de radionucléides ^{222}Rn , ^{210}Pb , ^{210}Po ainsi que par des isotopes de l'Uranium dans les cheminées, les fluides et donc dans l'ensemble du biotope (Charmasson, Sarradin et al. 2009). C'est sur l'un de ces édifices hydrothermaux profonds qu'a été isolé le modèle d'étude du laboratoire, *Pyrococcus abyssi*.

1.5 PYROCOCCUS ABYSSI, MODELE DE LA MAINTENANCE GENOMIQUE

P. abyssi est une Euryarchaeota affiliée à l'ordre des Thermococcales. La souche GE5 a été collectée au niveau d'une cheminée de sources hydrothermales à 2 000 m de profondeur, au nord du bassin fidjien, dans l'Océan Pacifique (G., CHARBONNIER et al. 1992). *P. abyssi* est un micro-organisme anaérobie strict, hyperthermophile, chimio-organotrophe, réduisant le soufre inorganique. A pression atmosphérique, la croissance est possible

entre 67 et 102°C. A la température optimale de croissance à pression atmosphérique (96°C), le temps de génération est d'environ 30 min. Lorsque les cellules sont cultivées à la pression *in situ* (20 MPa), la température optimale de croissance atteint 100°C. Les cellules se présentent sous la forme de coques de 0,8 à 2 µm de diamètre, mobiles grâce à de nombreux flagelles. Sur le plan métabolique, *P. abyssi* est hétérotrophe, elle nécessite des composés organiques comme des peptides ou des acides aminés comme source de carbone. Dans ces conditions, *P. abyssi* peut produire de l'acétate, de l'isovalérate, du propionate, H₂ et CO₂, ainsi que H₂S lorsqu'elle est cultivée en présence de soufre élémentaire ou de cystéine (Godfroy, Raven et al. 2000). La culture de *P. abyssi* peut être réalisée en milieu liquide et en milieu solide (Erauso, Reysenbach et al. 1993). Le génome de *P. abyssi*, souche GE5, a été séquencé au Génoscope en 1998. L'annotation du génome (Cohen, Barbe et al. 2003) a permis de déterminer les acteurs génétiques et moléculaires en charge de la vie de cette Archaea hyperthermophile (Figure 7). La séquence comprend un chromosome (44,7% GC) de 1 765 118 paires de bases ainsi qu'un plasmide multi copies de 3 444 paires de bases (pGT5). La séquence chromosomique comporte 1 764 cadres de lecture ouvert (seuil de détection : 100 acides aminés), codant pour autant de gènes dont à peu près un tiers correspond à des séquences orphelines. Une deuxième annotation du génome a permis d'assigner une fonction à 200 protéines hypothétiques supplémentaires (Gao and Wang 2012). Dans le génome, 73 gènes ont été initialement identifiés comme apparentés aux fonctions de réplication, recombinaison et réparation de l'ADN et 13 à la division cellulaire et séparation des chromosomes. Une grande partie des protéines impliquées dans la réplication de l'ADN chez *Pyrococcus abyssi* sont des homologues de protéines Eucaryotes ce qui fait de *P. abyssi* un excellent modèle pour étudier la réplication.

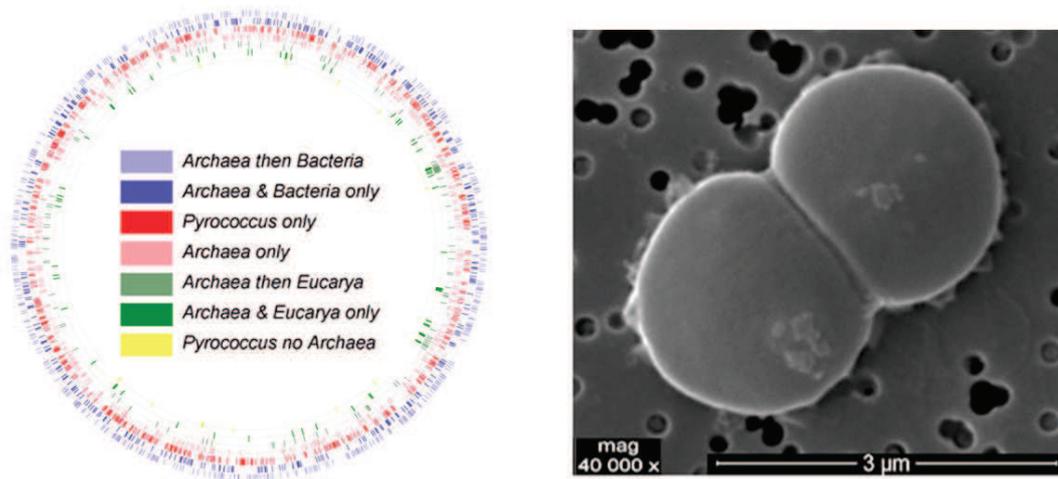


Figure 7: *P. abyssi*, une mosaïque génétique.

A gauche ; Résultats du blast du génome de *P. abyssi* vs GenBank NR database. ($P < 1e^{-05}$). (<http://www-archbac.u-psud.fr/>). A droite : *P. abyssi* observée en microscopie électronique à balayage (Ifremer – Nicolas Gayet).

Pyrococcus abyssi est la première Archaea dont l'origine de réplication (*oriC*) a été déterminée expérimentalement (Myllykallio *et al.* 2000 ; Matsunaga *et al.* 2001). Les protéines en charge de la réplication du chromosome de *P. abyssi* ont contribué à la description de plusieurs protéines et complexes de la réplication et de la réparation chez les Archaea (Gueguen, Rolland et al. 2001; Henneke, Flament et al. 2005; Le

Breton, Henneke et al. 2007; Meslet-Cladière, Norais et al. 2007; Rouillon, Henneke et al. 2007; Palud, Villani et al. 2008; Castrec, Rouillon et al. 2009; Castrec, Laurent et al. 2010; Le Laz, Le Goaziou et al. 2010; Creze, Ligabue et al. 2012; Gouge, Ralec et al. 2012). En dépit de la température élevée qui catalyse considérablement les dommages physiques et chimiques sur l'ADN (Lindahl 1993), *P. abyssi* présente un niveau de sites apuriniques, dommage de l'ADN dû à l'élévation de la température, du même ordre que celui des organismes mésophiles (Palud, Villani et al. 2008) suggérant l'existence d'un système de prise en charge des dommages extrêmement efficace.



2 MAINTENANCE GENOMIQUE

Les organismes vivants doivent reproduire et transmettre *ne variatur* l'information contenue dans les chromosomes. Ainsi, la conservation de l'intégrité du génome constitue le socle de la survie d'un organisme et de la persistance de son espèce. La maintenance génomique constitue l'ensemble des processus biologiques impliqués dans la conservation, la duplication et la transmission de l'information génétique contenue dans les chromosomes. L'étude des systèmes biologiques en charge de la réplication, de la réparation et de la recombinaison de l'ADN est l'une des thématiques majeures de la biologie moléculaire. Les études réalisées depuis les années 1960 pour les Bactéries et 1980 pour les Eucaryotes ont fourni une quantité d'information considérable sur les mécanismes complexes de la réplication, la réparation et la recombinaison de l'ADN. La grande variété des types de lésions pouvant survenir sur l'ADN et l'importance que revêt leur prise en charge par la cellule, reflète la diversité des mécanismes biochimiques de réparation de l'ADN découverts en étudiant ces organismes modèles. Dans ce contexte, les Archaea hyperthermophiles (HA) apparaissent comme d'intérêt pour étudier la réparation de l'ADN, puisqu'à la température de croissance le taux de décomposition de l'ADN est accéléré conduisant à des réactions élevées de déamination/dépurination des bases et des cassures des liaisons phosphodiester (Lindahl and Nyberg 1972; Lindahl 1993). Cependant, il a été montré que dans le génome de *P. abyssi*, cultivé à 96°C, le niveau de sites apuriques est seulement 10 fois plus élevé que chez *E. coli* (Palud, Villani et al. 2008) et pourtant, le taux de mutations spontanées observé chez les HA est similaire à celui mesuré chez les procaryotes mésophiles (Grogan, Carver et al. 2001). Ceci suggère donc que les Archaea hyperthermophiles doivent réparer de manière extrêmement active et efficace les dommages apparaissant au sein du chromosome. Comment ces organismes hyperthermophiles maintiennent l'intégrité de leurs génomes est donc une question doublement intéressante, à la fois du point de vue évolutif, étant donné la parenté des systèmes informationnels entre les Archaea et les Eucaryotes, mais aussi du point de vue physiologique. Pour autant, les voies de réparation de l'ADN sont peu décrites chez les Archaea, le retard dans le développement d'outils génétiques spécifiques des Archaea et le fait que l'on ne connaisse pas à ce jour d'Archaea pathogènes pour l'Homme sont en grande partie les causes de ce manque dans la connaissance de ces mécanismes essentiels chez les Archaea.

Je présenterai ci-dessous un aperçu des mécanismes moléculaires et des protéines impliquées dans la réplication et la réparation de l'ADN. J'exposerai succinctement différents acteurs protéiques et leurs fonctions vis-à-vis de l'ADN et discuterai plus en détails le fonctionnement de la machinerie moléculaire responsable de la maintenance génomique chez les Archaea.

2.1 REPLICATION DE L'ADN

En 1958, Matthew Meselson and Franklin Stahl établissent que la double hélice d'ADN est répliquée selon un processus semi conservatif où chacun des deux brins sert de matrice à la synthèse d'un brin complémentaire (Meselson and Stahl 1958). La réaction de polymérisation au sein de l'ADN polymérase est depuis peu décrite au niveau atomique (Nakamura, Zhao et al. 2012). L'ADN polymérase, isolée, est incapable de répliquer l'ADN. La réplication de l'ADN est prise en charge par un système protéique complexe et dynamique appelé réplisome

(Reyes-Lamothe, Possoz et al. 2008). Selon le type cellulaire, la composition du réplisome varie, cependant les mécanismes physico-chimiques permettant la duplication du génome sont identiques. La réplication est décomposable en trois étapes; l'initiation, l'élongation et la maturation. Les protéines essentielles à la réplication des Bactéries, des Eucaryotes et des Archaea sont présentées en figure 8. La réplication des virus ne sera pas traitée dans ce manuscrit hormis, ultérieurement, en figure 17 pour l'aspect structural d'un complexedu réplisome.

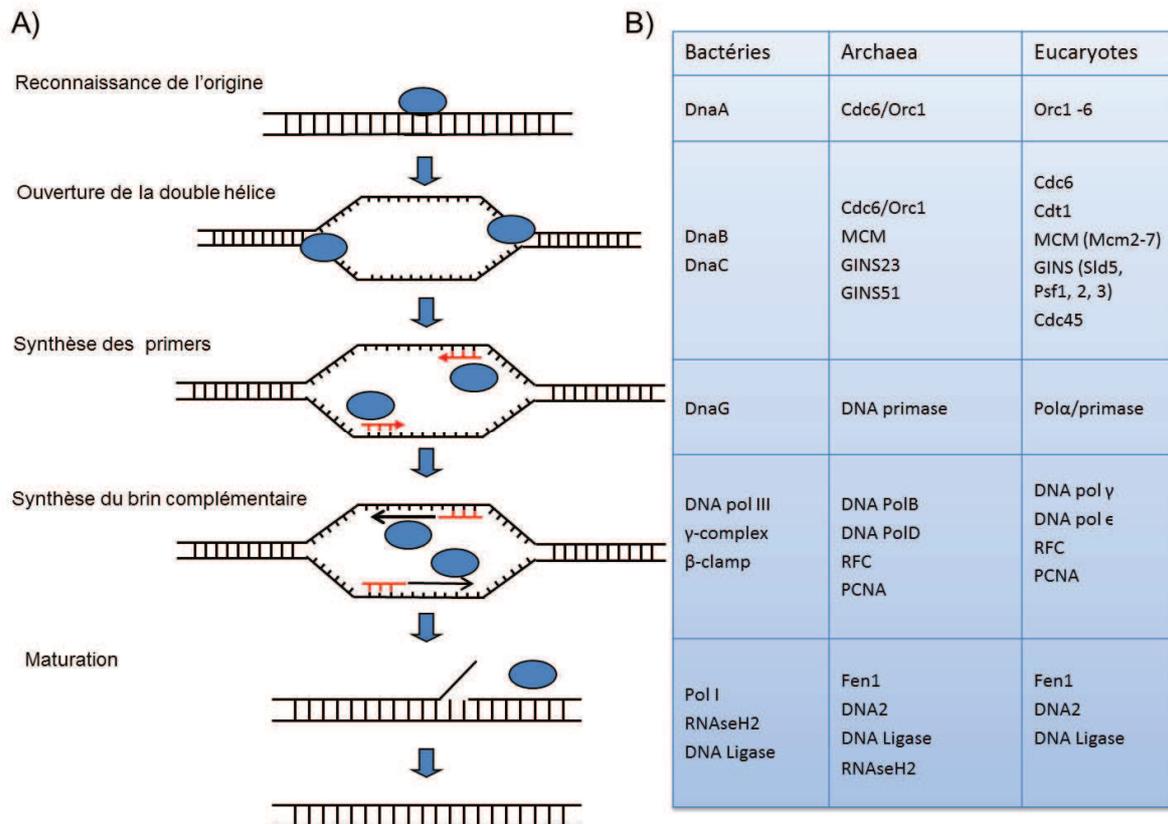


Figure 8: Comparaison des réplisomes Bactériens, Eucaryotes et Archaea.

A) Mécanisme de la réplication de l'ADN. De la reconnaissance de l'origine à la maturation des fragments d'Okazaki. B) Protéines impliquées à chaque étape dans les trois domaines du vivant.

Chez les Archaea 10 à 15 protéines sont considérées comme appartenant au cœur de la machinerie répllicative (Figure 8). Les protéines en charge de la reconnaissance de l'origine, du déroulement de l'ADN, de la synthèse du brin complémentaire et de sa maturation chez les Archaea sont des homologues de protéines Eucaryotes. Ces similitudes font de ces procaryotes d'excellents modèles pour l'étude des mécanismes de réplication (Kelman and White 2005).

2.1.1 INITIATION

L'une des étapes clé de la réplication et de sa régulation est l'initiation. L'initiation s'effectue en des points déterminés du génome appelés origines. Une famille de protéines, appelées OBP (Origin Binding Proteins), reconnaissent spécifiquement l'origine de réplication et s'y lient. L'interaction entre ces protéines et l'ADN est spécifique à la séquence de l'ADN et aux domaines présents chez cette famille de protéines. L'interaction de la protéine à l'ADN induit une déstabilisation locale de la double hélice qui servira de point de nucléation à un déroulement de l'ADN. Les protéines Orc participent au recrutement des hélicases répliquatives qui initieront, depuis l'origine de réplication, une ouverture processive et bidirectionnelle de l'ADN.

2.1.1.1 RECONNAISSANCE DE L'ORIGINE DE REPLICATION

Chez les Bactéries, il n'y a qu'une seule origine de réplication (OriC). Cette région, riche en A-T, est reconnue par la protéine DnaA. Cette protéine est, le plus souvent, codée de façon adjacente à l'origine de réplication. Les Eucaryotes possèdent quant à eux une multitude d'origines de réplication : les séquences ARS (Autonomous Replication Séquences) qui sont reconnues par un complexe de reconnaissance de l'origine (Origin Replication Complex). Les ORC sont constituées de 6 sous unités protéiques Orc1-6 (origin recognition complex 1-6).

Chez les Archaea le nombre d'origines de répliquions varie selon les espèces. *Pyrococcus abyssi* possède par exemple une origine de réplication unique tandis que *Haloferax volcanii*, *Sulfolobus solfataricus* ou encore *Pyrobaculum calidifontis* sont des espèces possédant des origines de réplication multiples (Pelve, Lindås et al. 2012). Le premier homologue Archaea de protéine OBP a été identifié dans le génome de *P. furiosus* (Uemori, Sato et al. 1997). Un gène possédant des homologies de séquence avec les protéines Eucaryotes cdc6 et Orc1 a été identifié, adjacent à l'ADN Polymérase D. L'identification de la protéine cdc6/orc1 a ensuite permis de déterminer expérimentalement l'origine de réplication Archaea en identifiant *in vitro* le site de fixation à l'ADN. Chez *P. abyssi*, cdc6/Orc1 est très probablement impliquée dans l'initiation de la réplication puisque la protéine se lie spécifiquement à la séquence oriC *in vivo* (Matsunaga, Forterre et al. 2001). Cette origine est composée d'une séquence riche en A-T entourée de deux séquences conservées de 13 paires de bases (pb). Ces séquences sont appelées respectivement ORB (Origine Recognition Box) et miniORB. La résolution de la structure tridimensionnelle de plusieurs homologues Archaea de la protéine cdc6/orc1 a révélé la présence de 3 domaines (Liu, Smith et al. 2000; Singleton, Dillingham et al. 2004). Les domaines I et II sont caractéristiques des protéines de la famille des ATPase AAA+ (ATPases Associated with diverse cellular Activities) tandis que le domaine III est un domaine en hélice ailée (« winged helix », WH) que l'on retrouve dans de nombreuses protéines se liant à l'ADN. La résolution de la structure de cdc6/orc1 de *P. furiosus* en présence d'ADN a révélé que les 3 domaines sont impliqués dans la liaison à l'ADN (Dueber, Corn et al. 2007; Gaudier, Schuwirth et al. 2007). Cette liaison induit une torsion de l'ADN qui provoque un déroulement local de la double hélice permettant l'initiation de la réplication. Chez *Halobacterium sp.* NRC-1, Orc10 et Orc2 sont des protéines essentielles (Berquist, DasSarma et al. 2007).

Lorsque l'ADN est déroulé à l'origine, les composants du replisome peuvent s'y assembler pour former le macro-complexe en charge de la réplication du génome appelé fourche de réplication. Les mécanismes impliqués dans l'assemblage de la fourche ne sont pas identifiés à l'heure actuelle.

2.1.1.2 OUVERTURE DE LA DOUBLE HELICE : L'HELICASE REPLICATIVE

Afin de répliquer un brin d'ADN, il est indispensable de le séparer de son brin apparié. Les hélicases forment une famille de molécules spécialisées dans l'ouverture catalytique de l'ADN. Les hélicases en charge du déroulement de l'ADN préalablement à la réplication sont appelées hélicases réplcatives. Les études biochimiques d'hélicases réplcatives bactériennes Eucaryotes et Archaea suggèrent que ces enzymes utilisent l'énergie dérivée de l'hydrolyse de nucléotides tri phosphates (NTPs) pour séparer le duplex d'ADN en amont de la fourche de réplication (Patel and Picha 2000).

L'hélicase bactérienne DnaB est généralement organisée en hexamère, cependant elle a été récemment décrite en dodécamère (Stelter, Gutsche et al. 2012), ayant la propriété de se lier et de glisser sur l'ADN simple comme double brin. Elle possède une activité ADN hélicase 5'→3' et une activité ATPase ADN-dépendante (Bochman and Schwacha 2009). L'hélicase réplcative Eucaryote est le complexe mini chromosome maintenance (MCM). C'est un hétéro hexamère constitué de six sous unités codées par six gènes (Mcm2-7). Des études biochimiques ont montré que le type d'assemblage du MCM eucaryote influençait son activité *in vitro*. Un complexe dimérique de l'hétérotrimère MCM4, 6, 7 possède une activité ADN hélicase 3'→5'; une activité de liaison à l'ADN simple brin et une activité ATPase ADN dépendante. *A contrario*, les complexes Mcm2 et Mcm3, 5 inhibent l'activité hélicase *in vitro*. Chacun de ces 6 gènes Mcm2-7 sont essentiels (Lei and Tye 2001; Forsburg 2004).

Il est proposé que le complexe MCM Archaea fonctionne similairement à son homologue Eucaryote. Le MCM Archaea possède une activité hélicase 3'→5' ATP dépendante, une activité ATPase ADN dépendante et est capable de se lier, de glisser sur l'ADN (Kelman and Kelman 2003; Kelman and White 2005). La plupart des espèces Archaea étudiées ne codent que pour un seul homologue de MCM. Dans le cas où l'on retrouve plusieurs homologues, il est possible que le complexe MCM soit un hétérocomplexe. La structures quaternaire de la protéine *in vivo* reste pour l'instant indéfinie puisque le complexe a été observé en hexamère, heptamère, dodécamère et également en filaments (Kelman and White 2005). La structure du domaine N-terminal en dodécamère a été résolue à 3Å (Fletcher, Bishop et al. 2003) (figure 10). Chaque sous unité contribue à la formation d'un site de liaison à l'ADN en épingle à cheveux (β -hairpin) au centre du cercle. Des études biochimiques chez *Sulfolobus* ont révélé que ce domaine contribue à la processivité de l'activité hélicase (Barry, McGeoch et al. 2007). En C- terminal du domaine de processivité se trouve un domaine ATPase AAA+ moteur de l'hélicase. Ce domaine possède à lui seul une activité hélicase *in vitro* (Barry and Bell 2006). Il est composé de deux motifs clés : un β -hairpin et un module β - α - β . Le domaine le plus C-terminal est un domaine WH. Sa délétion provoque une augmentation de l'activité hélicase suggérant un rôle régulateur (Jenkinson and Chong 2006). Les propriétés biochimiques et structurales des MCM Archaea ont permis de construire plusieurs modèles généraux, proposés pour décrire l'activité de MCM, *in vivo* (Sakakibara, Kelman et al. 2009) (Figure 9).

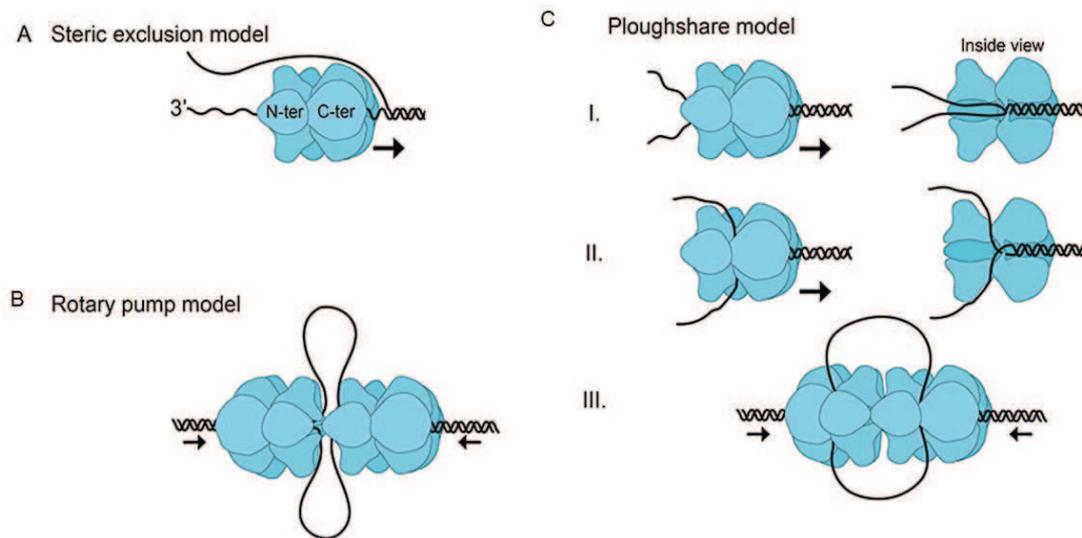


Figure 9: Mécanismes possibles pour l'activité hélicase de MCM *in vivo*.

A) Modèle d'exclusion stérique. Le MCM glisse de 3' en 5' sur un simple brin et déplace le brin opposé B) Modèle de pompe rotative. L'ADNsb est déroulé par deux complexes MCM tournants en sens opposé C) Modèle soc de charrue ("ploughshare"). 3 versions du modèle ploughshare sont proposées. Adapté de Sakakibara 2009.

L'interaction de *cdc6/Orc1* avec MCM a été démontrée *in vitro* chez *Methanothermobacter thermautotrophicus* (Shin, Grabowski et al. 2003; Kasiviswanathan, Shin et al. 2005). Le recrutement de l'hélicase répliquative à l'origine nécessite la présence de *cdc6/orc1*. Cette observation a alimenté le modèle où le MCM est recruté directement à l'origine via l'interaction à *cdc6/orc1* (Shin, Jiang et al. 2003). L'hélicase répliquative MCM est évidemment une protéine essentielle chez *Halobacterium sp.* NRC-1 (Berquist, DasSarma et al. 2007).

L'interaction entre l'hélicase répliquative et le facteur de liaison à l'origine n'est pas la seule interaction impliquée dans l'initiation de la réplication, les protéines associées aux hélicases répliquatives sont présentées ci-après.

2.1.1.3 LE COMPLEXE GINS

Le complexe GINS constitue l'un des complexes protéiques de la réplication identifié le plus récemment. Initialement découvert chez *Saccharomyces cerevisiae* (Takayama, Kamimura et al. 2003), ce complexe protéique est un facteur essentiel à l'initiation de la réplication et qui interagit avec le MCM.

Le complexe GINS Eucaryotes est constitué de 4 sous-unités Sld5, Psf1, Psf2 et Psf3 (GINs est un acronyme tiré du japonais « go-ichi-ni-san » signifiant 5-1-2-3, chiffres associés aux 4 sous-unités). Chacune des sous-unités possède un haut niveau de conservation suggérant une relation paralogue ancestrale. La découverte de ce complexe chez la levure a permis l'identification d'homologues Archaea. Le complexe GINS Archaea est composé de deux sous-unités GINS51 et GINS23. Elles possèdent des similarités de séquences avec Sld5 et Psf1, et Psf2 et Psf3, respectivement. Une interaction entre les protéines GINS23 et MCM de *Sulfolobus solfataricus* a été observée grâce à la technologie de « Yeast 2 Hybrid » (Y2H). La protéine GINS51 de *S. solfataricus* a été co-

purifiée lors d'une immunoprécipitation de GINS23 à partir d'un extrait cellulaire (Marinsek, Barry et al. 2006). SsoGINS possède une structure tetramérique avec un ratio équimolaire 2:2. SsoGINS interagit avec MCM et la primase, cependant aucun effet sur l'activité hélicase ou primase n'a été observée *in vitro* (Marinsek, Barry et al. 2006). En revanche, l'activité hélicase de MCM de *P. furiosus* est stimulée *in vitro* par l'ajout du complexe GINS (Yoshimochi, Fujikane et al. 2008) suggérant le complexe GINS comme un facteur accessoire de la réplication Archaea, possédant un rôle régulateur.

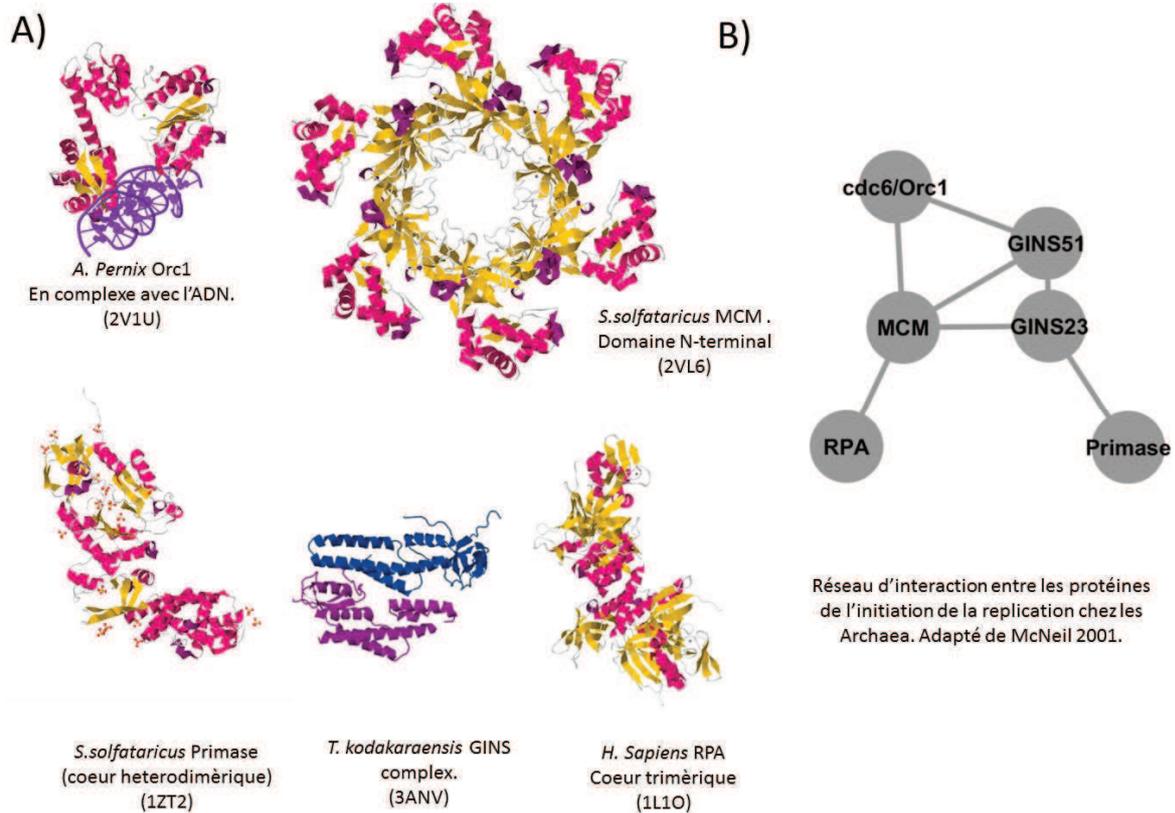


Figure 10: Protéines et complexes à l'initiation

A) Structure cristalline des protéines cdc6/orc1, de MCM, GINS, Primase et RPA. B) Interactions au sein de protéines de l'initiation de la réplication

Chez les Eucaryote l'hélicase MCM est associée au complexe GINS et à la protéine Cdc45. Ce complexe, CMG, est désormais identifié comme le complexe protéique en charge de l'activité hélicase réplivative Eucaryote (Fig. 11) (Makarova, Koonin et al. 2012). Chez les Archaea, une protéine de la famille Rec J appelé GAN (GINS Associated Nuclease) a été co-précipitée avec le complexe GINS chez *T. kodakarensis* (Li, Santangelo et al. 2010). Cette protéine constituerait l'homologue fonctionnel de cdc45 chez les Archaea. L'activité nucléase de la GAN est stimulée en présence du complexe GINS (Li, Pan et al. 2011). Les complexes CMG chez les Eucaryotes et les Archaea sont représentés schématiquement en figure 11.

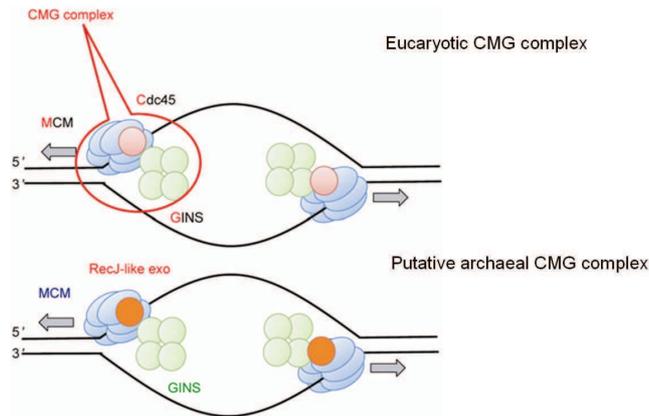


Figure 11: Complexe CMG Eucaryote et Archaea

Il est raisonnable de penser que d'autres facteurs protéiques jusque-là non identifiés puissent interagir avec le CMG et intervenir dans cette étape cruciale de l'initiation de la réplication. Le complexe CMG est en charge de l'initiation d'une étape cruciale de la vie de la cellule : l'entrée effective en réplication. L'initiation de la réplication de l'ADN constitue ainsi le point d'entrée dans un processus dont la complétion est fondamentale. L'identification de nouveaux acteurs des complexes protéiques en charge de cette étape constituerait une découverte d'un intérêt certain, d'autant qu'ils pourraient être conservés.

2.1.1.4 SYNTHÈSE DE L'AMORCE : L'ADN PRIMASE

La polymérisation du brin complémentaire à un ADN simple brin nécessite la synthèse préalable d'un court fragment nucléotidique appelé amorce (« primer »). Cette étape nécessaire à l'initiation de la polymérisation est prise en charge par une Primase. Elle synthétise un court fragment nucléotidique qui constituera le site d'initiation de la polymérisation du brin complémentaire.

La primase bactérienne est appelée DnaG (Mansfield and Dixon 2012). Il s'agit d'un monomère possédant trois domaines : un domaine N-terminal de fixation du Zn^{2+} jouant un rôle dans la reconnaissance de l'ADNsb. Un domaine central avec une activité ARN polymérase qui est le site permettant la synthèse du primer. Et un domaine C-terminal impliqué dans la liaison à DnaB. Chez les eucaryotes la primase est constituée d'un hétérodimère p48-p58 (sous unités catalytiques et régulatrices, respectivement), de la sous unité B et de l'ADN polymérase α formant le complexe alpha-primase DNA polymérase (Arezi and Kuchta 2000). La primase archaea est un hétérodimère composé uniquement de p41 et p46 (homologues respectifs de p48 et p58). La sous unité p46 forme un complexe stable avec p41 (Le Breton, Henneke et al. 2007). La sous unité p41 est la sous unité catalytique alors que la sous unité p46 régule l'activité du complexe. Les sous unités catalytiques (qui synthétisent le primer RNA) des Eucaryotes et des Archaea sont apparentées et possèdent des homologies avec des polymérases de la famille X (Lao-Sirieix, Pellegrini et al. 2005). Chez les Euryarchaea et les Crenarchaea, les ADN primases synthétisent des courts primers ARN (Matsunaga, Norais et al. 2003). Les ADN primases Archaea possèdent également, *in vitro*, des activités ; ADN polymérase (Bocquier, Liu et al. 2001; Liu, Komori et al. 2001; Lao-Sirieix and Bell 2004), deoxynucleotidyl terminal transférase (De Falco, Fusco et al.

2004) et de polymérisation des matrices discontinues (« polymerization across discontinuous templates », PADT)(Hu, Guo et al. 2012). Ces données biochimiques et structurales suggèrent l'implication des ADN primases Archaea dans la réparation de l'ADN, mais aucun mécanisme n'a été précisément décrit à ce jour.

Des homologues de la protéine DnaG ont été identifiés dans les génomes Archaea. En revanche, il semble que cette protéine intervient principalement dans la dégradation des ARN au sein de l'exosome (Evguenieva-Hackenberg, Walter et al. 2003). On peut tout de même signaler que SsoDnaG est capable de synthétiser de courts primers ARN de 13 pb (Zuo, Rodgers et al. 2010).

L'ADN primase de *Sulfolobus* interagit avec MCM via Gins23 *in vitro* (Marinsek, Barry et al. 2006). Cette interaction pourrait jouer chez les Archaea un rôle dans le couplage entre le déroulement de l'ADN et la synthèse des primers sur le brin retardé.

2.1.1.5 LES PROTEINES DE LIAISON A L'ADN SIMPLE BRIN: SSB ET RPA

Les protéines SSB chez les Bactéries et RPA (Replication Protein A) chez les Eucaryotes sont des protéines spécialisées dans la liaison de l'ADN simple brin pour sa protection contre les nucléases ou pour contrer les modifications chimiques pouvant survenir lors de l'ouverture de la double hélice ADN. La liaison des SSBs à l'ADN simple brin est notamment importante pendant la réplication. Les protéines SSBs tapissent l'ADN simple brin à la fourche de réplication maintenant la séparation des deux brins. En dépit de leurs faibles similarités de séquences, les protéines SSB et RPA possèdent le même domaine Oligonucleotide / Oligosaccharide - Binding (« OB-fold »), impliqué dans la liaison de la protéine à l'ADN. Le domaine OB fold consiste en un feuillet β (« five-stranded β -sheet ») replié en β -barrel fermé, souvent coiffé d'une hélice α (Pestryakov and Lavrik 2008, {Flynn, 2012 #509}). Bien que le nombre et la distribution des OB fold varient, la conservation de la structure entre SSB et RPA suggère une conservation du mécanisme de liaison à l'ADN simple brin, via l'OB-fold, dans l'ensemble du vivant.

Chez *E. coli*, SSB est un homotetramère d'un peptide de 20kDa contenant un seul OB-fold (Figure 12). Les SSBs des bactéries extrêmophiles *Deinococcus radiodurans* et *Thermus aquaticus* sont quant à eux constitués d'un homodimère d'un peptide contenant 2 OB-fold. Le SSB est décrit comme une protéine ubiquitaire des complexes protéiques de la maintenance génomique procaryote qui interagit via un domaine spécialisé à au moins 14 protéines du métabolisme de l'ADN (Shereda, Kozlov et al. 2008). Le RPA Eucaryote est, quant à lui, un hétérotrimère constitué de 3 sous unités de 70, 32 et 14 kDa. RPA70 possède 4 OB fold, RPA 32 un OB-fold et un domaine WH et RPA14 un OB-fold (Figure 12). Le RPA Eucaryote possède également un nombre élevé de partenaires impliqués dans la réparation, la réplication et la recombinaison de l'ADN (Tableau 1) (Fanning, Klimovich et al. 2006; Zou, Liu et al. 2006; Pike, Burgers et al. 2009; Hass, Lam et al. 2012). Le RPA Eucaryote est également capable de réguler l'activité de l'ARN polymérase en charge de la transcription (Sikorski, Ficarro et al. 2011).

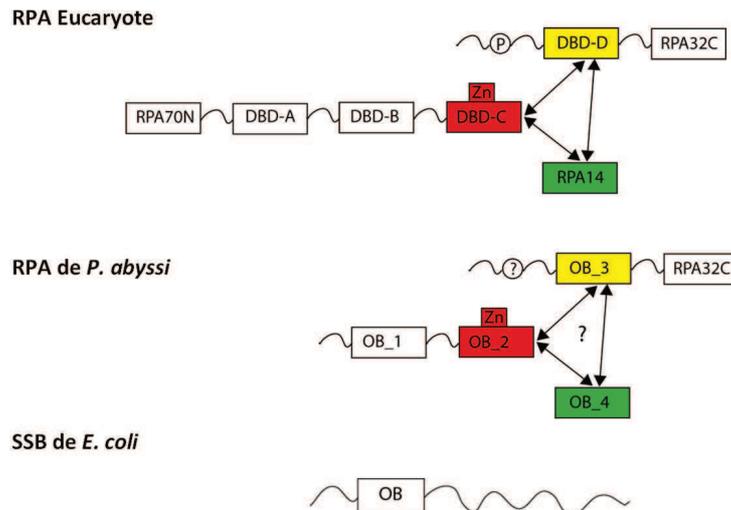


Figure 12: Schéma de l'organisation des protéines SSB et RPA.

Les domaines OB fold sont nommés DBD ou OB. Les domaines OB fold connus pour former le cœur de trimérisation du complexe eucaryote sont colorés

Tableau 1: Protéines de la maintenance génomique interagissant avec le RPA Eucaryote.

Protéine partenaire	Site d'interaction sur RPA	Site d'interaction sur le partenaire	Méthode
Cytidine Deaminase	RPA32	Indéterminé	e
Uracile-DNA glycosylase	RPA32163–217	29–75, N-terminus	cd
Rad52	RPA70169–326	221–280	ac
	RPA32224–271		b
SV40 T antigen	RPA70181–327	Origin binding domain 131–249	ab
	RPA32C		c
XPA	RPA70183–296	20–46	be
	RPA32C		a
p53	RPA70N1–120	38–58	ace
ATRIP	RPA70N	1–107, other?	e
FACT	RPA70N	Pob3237–477	ade
Rad51	RPA70181–291	1–93	ac
Werner syndrome helicase	RPA70168–308	N-terminal acidic region, C-terminus	abd
Bloom syndrome helicase	RPA70168–308	N-terminal acidic region, C-terminus	abd
Papillomavirus E1	RPA70181–291	Indéterminé	ab
Parvovirus NS1	RPA70, RPA32	Indéterminé	b
Pol-prim	RPA701–327	Primase p48/58	ab
	RPA32/14?		
RF-C	RPA70	p140, p40, p28	b
Rad9	RPA70, RPA32	Indéterminé	a
Rad17	RPA70, RPA32/14	Indéterminé	ae
p53BP1	RPA70, RPA32	Indéterminé	a
Nucleolin	RPA14	645–707	ae
BRCA2	Indéterminé	N-terminus	ae
XPG	Indéterminé	Indéterminé	ae
XPF–ERCC1	Indéterminé	Indéterminé	ae
Mre11–Rad50–Nbs1	Indéterminé	Indéterminé	e

Méthode : a, Purification d'affinité; b, ELISA; c, RMN; d, Y2H; e, Autre. (Fanning, Klimovich et al. 2006).

Chez les Archaea, on observe une forte hétérogénéité dans les structures des protéines SSBs. En généralisant, les Euryarchaea possèdent des protéines homologues au RPA tandis que l'organisation générale des SSB Crenarchaea est plus proche de celle identifiée chez les bactéries (un seul domaine OB fold et une longue extension C-terminale). Les caractéristiques biochimiques de la réaction de fixation sur l'ADNsb par le domaine OB fold des SSB Crenarchaea sont en revanche plus proches de celles décrites pour les sous unités du RPA Eucaryotes (Kerr, Wadsworth et al. 2003). Les premiers homologues de protéines de liaison à l'ADNsb ont été identifiés chez les Euryarchaea *M. jannaschii* et *M. thermoautotrophicus* (ChÃ©din, Seitz et al. 1998) (Kelly, Simanek et al. 1998). Ces protéines possèdent des homologies de séquences avec le RPA Eucaryote. Les méthanogènes *Methanosarcina acetivorans* codent trois protéines RPA (MacRPA1–3), chacune capable de s'assembler en homomultimère (Robbins, Murphy et al. 2004; Lin, Robbins et al. 2005; Robbins, McKinney et al. 2005). La mutation des résidus cystéine et histidine de MacRPA3 réduisent son affinité à l'ADNsb (Lin, Robbins et al. 2005). L'existence d'OB-Fold et de domaines Zinc Finger CX₂CX₈CX₂H est observée chez les RPA des Archaea méthanogènes (Robbins, McKinney et al. 2005) (Kelly, Simanek et al. 1998) ou non (Pugh, Lin et al. 2008). Chez *M. acetivorans* RPA est capable de stimuler la synthèse d'ADN par la polymérase B, *in vitro* (Robbins, Murphy et al. 2004).

Chez *H. volcanii*, trois protéines de liaison à l'ADN simple brin ont été identifiées (rpaA, rpaB et rpaC). Le gène rpaC est essentiel tout comme le produit des gènes rpaA et rpaB (qui peuvent en revanche être mutés individuellement (Skowrya and MacNeill 2012)). La délétion de RpaC peut cependant être complétée par une surexpression de RpaB indiquant une redondance fonctionnelle au sein des SSB halobactériens. La surexpression de rpaC induit une résistance accrue aux dommages de l'ADN suggérant une implication de rpaC dans les mécanismes de réparation de l'ADN (Skowrya and MacNeill 2012). RPA semble directement impliquée dans la recombinaison, en effet PfuRPA stimule la réaction d'échange de brin promue par RadA, *in vitro* (Komori and Ishino 2001).

En contraste de la complexité observée chez les protéines RPA Euryarchaea, le SSB Crenarchaea, bien caractérisé chez le thermophile *Sulfolobus solfataricus*, ne contient qu'un seul OB fold et un domaine C-terminal flexible. Le domaine C-ter, n'est pas impliqué dans la liaison à l'ADNsb (Kerr, Wadsworth et al. 2003). Il est probablement impliqué dans l'interaction avec d'autres protéines, au même titre que ce qui est observé chez les Bactéries. Cubeddu et White ont démontré que la protéine SSB de *Sulfolobus solfataricus* peut reconnaître spécifiquement les bases d'un ADN contenant un mésappariement ou une lésion, *in vitro* (Cubeddu and White 2005). Cette activité suggère une implication des SSB Crenarchaea dans la détection des dommages sur l'ADN.

La spécificité des protéines SSB et RPA à la forme simple brin de l'ADN, qui correspond à la conformation « active » (répliqué, transcrite, réparé, recombéné), en fait un acteur ubiquitaire des complexes protéines impliqués dans les transactions sur l'ADN. Les nombreux partenaires des SSBs des trois domaines du vivant reflètent leur caractère multifonctionnel.

2.1.2 SYNTHÈSE DU BRIN COMPLÉMENTAIRE

La synthèse du brin complémentaire, est effectuée par une famille de protéines spécialisées ; les ADN polymérase. *In vivo*, la réplication de l'ADN par les polymérase nécessite de nombreux facteurs protéiques associés, essentiels à la régulation de la polymérisation et à la maturation des brins d'ADN. Je présenterai ci-après les ADN polymérase puis leurs facteurs associés, PCNA, et RF-C.

2.1.2.1 ADN POLYMERASES

Les ADN polymérase sont les protéines en charge de l'ajout séquentiel des désoxynucléotides complémentaires au cours de la réplication d'un brin d'ADN. Décrire la somme des connaissances réunies au cours de dizaines d'années de recherche concernant cette famille d'enzymes dépasserait le cadre de ce manuscrit. Dans les paragraphes suivants je décrirai succinctement les propriétés générales des polymérase, et je présenterai ensuite les deux polymérase répliquatives Archaea de *P. abyssi* un peu plus en détails.

Les ADN polymérase catalysent la formation d'une liaison phosphodiester entre l'extrémité 3'OH terminale et le phosphate alpha du dNTP inséré en 3' du brin complémentaire. Toutes les polymérase possèdent cette activité ($ADN_n + dNTP \Rightarrow ADN_{n+1} + P_{pi}$) qui fait intervenir un site catalytique renfermant deux ions métalliques divalents (généralement Mg^{2+}) cofacteurs indispensables à la formation de la liaison covalente entre les nucléotides. La plupart des ADN polymérase possède une organisation structurale commune, comparable à une main avec des domaines doigts (« fingers »), paume (« palm ») et pouce (« thumb »). Les polymérase sont classées en 7 grandes familles, A, B, C, D, E, X, et Y (Burgers, Koonin et al. 2001)(tableau 2). Ce classement regroupe les polymérase selon :

- La processivité : nombre de molécules incorporées par cycle d'attachement (de 1 à 10 000).
- La fidélité : nombre de nucléotides correctement incorporés. Elle dépend de la sélectivité de la polymérase et de la plasticité de son site actif (Kool, 2002). La fidélité est améliorée en présence d'une fonction 3'→5' exonucléasique.
- L'affinité pour le substrat.
- Le taux d'extension de la chaîne d'ADN : nombre de nucléotides insérés par unité de temps.

Certaines polymérase possèdent des fonctions additionnelles (Pavlov et al 2006):

- L'activité exonucléase 3'→5' de la polymérase δ (« proof reading ») améliorant la fidélité.
- L'activité de déplacement de simple brin 5'→3' de la polymérase I d'*E. coli* (« nick translation »).
- L'activité 5' désoxyribose phosphate lyase (dRP lyase) chez la polymérase β .
- Activité ATPasique chez la polymérase θ .

La première ADN polymérase (ADN pol I) a été identifiée chez *E. coli* (Lehman, Bessman et al. 1958). Aujourd'hui, il est décrit 5 polymérase bactériennes; PolI, PolII et PolIII qui appartiennent aux familles A, B et C

respectivement, et PolIV et PolV qui sont des ADN polymérases translésionnelles de la famille Y. Une ADN polymérase translésionnelle possède la propriété de pouvoir poursuivre la réplication d'un brin d'ADN dont les bases sont endommagées, ce dont les ADN polymérases des autres familles sont incapables.

Tableau 2: Distribution, Fonction et fidélité des différentes familles d'ADN Polymérases.

Famille	Domaine	ADN Polymérase	Fonction	Fidélité
A	Bacteria	Pol I	Réplication - Réparation	10^{-5} à 10^{-6}
	Eukarya	Pol θ	Translésion	
		Pol ν	Réparation	
		Pol γ	Replication (mitochondriale)	
B	Bacteria	Pol II	Réplication	10^{-5} à 10^{-6}
	Archaea	Pol B I	Réplication – Réparation ?	
		Pol B II		
	Eukarya	Pol ϵ	Réplication	
		Pol δ	Réplication	
		Pol ζ	Translésion	
Pol α		Initiation de la réplication		
C	Bacteria	Pol III	Réparation	10^{-5} à 10^{-6}
D	Archaea	Pol D	Réplication-Réparation ?	10^{-5} à 10^{-6}
X	Eukarya	Pol β	Réparation	10^{-4} à 10^{-5}
		Pol λ	Réparation	
		Pol μ	Réparation	
		Pol σ	Réparation	
		Tdt	Réparation	
Archaea	PolX ??	Réparation ?		
Y	Bacteria	Pol IV	Translésion	10^{-2} à 10^{-4}
		Pol V	Translésion	
	Archaea	Dph	Translésion	
		Dpo4	Translésion	
		PolY1	Translésion	
	Eukarya	Pol η	Translésion	
		Pol ι	Translésion	
		Pol κ	Translésion	
Pol REV1		Translésion		

L'ADN polymérase répliquative est la Pol III. Composée de dix sous unités, elle est également appelée holoenzyme Pol III. Chez les Eucaryotes, les polymérases répliquatives sont la Pol α , Pol ϵ et Pol δ de la famille B et les polymérases translésionnelles (Pol η , Pol ι , Pol κ et PolREV1) appartiennent à la famille Y. L'ADN polymérase δ serait en charge de la synthèse du brin avancé tandis que l'ADN polymérase ϵ interviendrait dans la synthèse du brin retardé (Fukui, Yamauchi et al. 2004; Garg and Burgers 2005).

Dans les génomes Archaea 4 familles de polymérases, B,D, Y et X ont été identifiées. Les polymérases répliquatives *bona fide* sont les Pol B et D, leurs distributions varient selon le phylum considéré (Tableau 2). Elles nécessitent un court fragment complémentaire à la séquence à répliquer (« primer ») pour initier la synthèse de l'ADN. En raison de leur distribution, de leur essentialité et de leurs propriétés biochimiques, il est proposé

que la polymérase B Crenarchaea soit impliquée dans l'élongation des brins avancés et retardés (Böhlke, Pisani et al. 2002). Chez les Euryarchaea, Pol B réaliserait la synthèse continue et Pol D la synthèse discontinue (Henneke, Flament et al. 2005).

2.1.2.1.1 POLYMERASE B

L'organisation tri dimensionnelle des polymérases B Archaea est similaire à celle de la polymérase virale du bactériophage T4 RB69 avec un domaine polymérase C-terminal et un domaine exonucléase en N-terminal (Savino, Federici et al. 2004) (Figure 13). Les séquences d'acides aminés de ces enzymes présentent toutes des résidus conservés caractéristiques de la famille B (Hübscher, Maga et al. 2002).

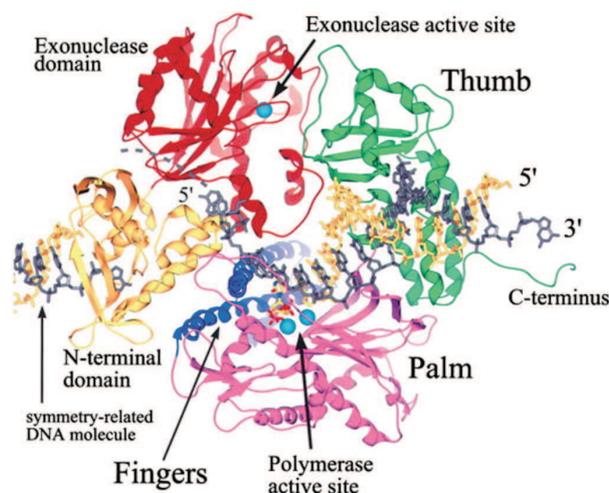


Figure 13: Structure de l'ADN Polymérase B du bactériophage RB69 en complexe avec l'ADN.

Les sous domaines « fingers », « palm » et « thumb » sont colorés en bleu, magenta et vert respectivement. Le domaine exonucléase est représenté en rouge, la matrice ADN en gris. D'après Franklin et al. 2001.

In vitro, la polymérase B de *P. abyssi* est capable d'initier la synthèse de l'ADN à partir d'un primer ADN mais pas ARN. *In vitro*, Elle ne réalise pas de déplacement de brin (Henneke, Flament et al. 2005). La polymérase B interagit avec le PCNA via un PIP-box motif et stimule son chargement (Rouillon, Henneke et al. 2007). L'interaction de PCNA à la polymérase B augmente considérablement sa processivité et lui permet de déplacer un brin d'ADN mais pas ARN (Henneke, Flament et al. 2005).

Lors de la réplication, les polymérases B des Archaea ont la capacité de reconnaître et de lier une base désaminée. La polymérase s'arrête 4 nucléotides en amont du dommage (Fogg, Pearl et al. 2002). La reconnaissance de l'uracile est basée sur un mécanisme de pré-lecture ("read-ahead") (Greagg, Fogg et al. 1999). Cette propriété empêche l'incorporation d'une adénine ou d'une cytosine en face de l'uracile ou de l'hypoxantine respectivement, réduisant le potentiel mutagène de ces lésions (Lindahl 1993; Wardle, Burgers et al. 2008; Russell, Richardson et al. 2009). Lorsque la polymérase est arrêtée sur un dommage, l'élongation de la réplication est stoppée et l'activité exonucléase est stimulée par un changement de conformation (mode

édition) (Russell, Richardson et al. 2009). La structure de la polymérase B de *P. abyssi* en mode édition sur une matrice d'ADN endommagée a été résolue récemment (Gouge, Ralec et al. 2012) (Figure 14).

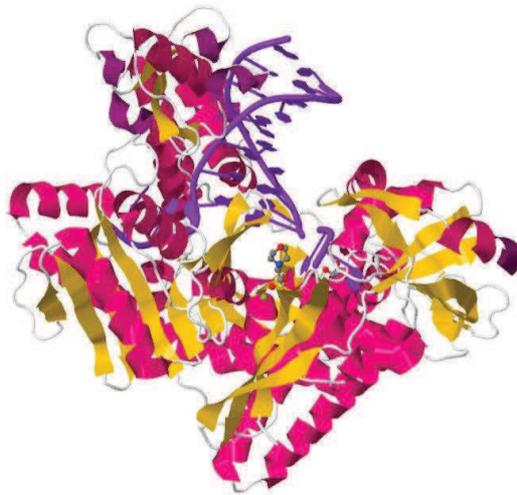


Figure 14: ADN polymérase B de *Pyrococcus abyssi* liée à un ADN double brin.

(mode édition). N°PDB : 4FLT (Gouge, Ralec et al. 2012).

Les polymérases B Crenarchaea sont également associées aux mécanismes de tolérance aux dommages. Les polymérases BII et BIII de *Pyrobaculum aerophilum* semblent être impliquées dans la voie de réparation par excision de bases (Base Excision Repair, BER) (Sartori and Jiricny 2003). Egalement, l'ADN Polymérase BI de *S. solfataricus* est bloqué par les sites AP et 8-oxo-dG. L'arrêt de la polymérisation après une incorporation en face d'un dommage déclenche le recrutement d'une ADN polymérase translésionnelle de la famille Y pour poursuivre la réplication, ce mécanisme est appelé « polymerase switch » (Grüz, Shimizu et al. 2003).

On peut également signaler que les polymérases B des Archaea thermophiles possèdent des ponts disulfures et des clusters d'acides aminés supplémentaires, chargés positivement, afin de favoriser les interactions électrostatiques de la polymérase à l'ADN et d'augmenter la stabilité de l'enzyme (Savino, Federici et al. 2004). Leurs propriétés font des Polymérases B des hyperthermophiles des outils précieux pour la PCR.

2.1.2.1.2 POLYMERASE D

Les polymérases D sont composées d'une petite sous unité (DP1), qui possède un domaine PHP que l'on retrouve notamment chez les polymérases Eucaryotes α , δ et ϵ , (Aravind and Koonin 1998) et une grande sous unité (DP2) qui possèdent l'activité de polymérisation (Shen, Musti et al. 2001). Des homologues de la Pol D ont été identifiés initialement chez les Euryarchaea et plus récemment Korarcheota, Thaumarchaeota et Aigarchaeota. L'hétérodimère Pol D possède une activité de polymérisation efficace et une activité 3'→5' exonucléase impliquée dans la correction des erreurs (proofreading) (Tang, Shen et al. 2004). La polymérase D de *P. abyssi* est un hétéromère (Gueguen, Rolland et al. 2001), les deux sous unités sont codées en tandem dans le génome et localisées au niveau de l'îlot de réplication contenant l'oriC et plusieurs protéines essentielles de la réplication, la réparation et la recombinaison (Cohen, Barbe et al. 2003). *In vitro*, *pabPolD* est

capable de réaliser un déplacement de brin lorsque qu'elle rencontre une séquence hybridé ARN mais pas ADN (Henneke, Flament et al. 2005). En présence de PCNA, la polymérase D déplace un brin hybridé qu'il soit ARN ou ADN.

Au regard des propriétés de la Pol B, il est probable que la polymérase D réalise la synthèse discontinue tandis que la polymérase B serait en charge de la polymérisation continue, sur le brin avancé (Henneke, Flament et al. 2005). L'absence de déplacement de brin et d'initiation de la réplication par un primer ARN de Pol B *in vitro* est en accord avec ce modèle. Les deux ADN polymérases pourraient interagir séquentiellement avec le PCNA lors de la réplication chez les Euryarchaea (Rouillon, Henneke et al. 2007). Un premier cycle d'élongation par la Pol D serait ainsi nécessaire pour créer le primer ADN indispensable à la Pol B.

2.1.2.1.3 POLYMERASES Y

Les polymérases translésionnelles de la famille Y sont des polymérases spécialisées dans la tolérance aux dommages. Les protéines Dbh et Dpo4 ont été les deux premières ADN polymérases de la famille Y identifiées chez les Archaea (Kulaeva, Koonin et al. 1996; Boudsocq, Iwai et al. 2001; Grüz, Shimizu et al. 2003). Ces polymérases ont la particularité de pouvoir répliquer un ADN endommagé. Chez les Crenarchaea, Dpo4 est par exemple capable de franchir un site AP, un dimère de thymine ou encore une lésion cis-platine. Dbh franchit efficacement l'hypoxanthine et l'uracile mais est bloqué par un site AP (Grüz, Shimizu et al. 2003). En présence d'ADN endommagé, la polymérase translésionnelle remplace la polymérase répllicative afin de passer le dommage (Friedberg, Lehmann et al. 2005). Ces polymérases sont, en revanche, sensiblement moins fidèles et peuvent induire des mutations. Chez *S. solfataricus*, le gène codant pour dbh n'est pas essentiel mais induit une augmentation des substitutions des bases G:C vers une paire T:A (Sakofsky, Foster et al. 2012). Aucune polymérase de la famille Y n'a été identifiée chez les Euryarchaea (Boudsocq, Iwai et al. 2001; Grüz, Shimizu et al. 2003).

2.1.2.2 PCNA, MAESTRO DE LA REPLICATION

Chez les Bactéries, les Eucaryotes et les Archaea les polymérases répliquatives sont associées à une protéine en forme d'anneau appelée « sliding clamp », que l'on traduit par pince coulissante ou facteur de glissement. Le sliding clamp ne possède pas d'activité enzymatique, sa fonction première est d'encercler l'ADN simple brin, matrice de la polymérase, et de l'y maintenir, augmentant ainsi considérablement la processivité de réaction de la polymérisation.

Chez les Bactéries, c'est le β -clamp qui forme une structure circulaire entourant l'ADN et facilitant la progression de la polymérase (Vivona and Kelman 2003). Le complexe gamma qui fait partie de l'holoenzyme est responsable du chargement et du décrochement de la sous unité beta dans une réaction ATP dépendante. Chez les Eucaryotes, c'est la protéine Proliferating Cell Nuclear Antigen, (PCNA) qui remplit le rôle essentiel de clamp (Vivona and Kelman 2003). La forme en anneau du sliding clamp est conservée pour l'ensemble du vivant, bien que PCNA et le β -clamp ne possèdent pas d'homologie de séquence (Figure 15). PCNA (« Proliferating Cell Nuclear Antigen ») a été identifié indépendamment par deux groupes de chercheurs

quasiment simultanément : Miyachi et ses collaborateurs (Miyachi, Fritzler et al. 1978) identifièrent initialement un auto-antigène chez des patients souffrant de lupus erythematosus systémique. Ils nommèrent cette protéine PCNA en raison de son observation dans le noyau de cellules en division. Au même moment Bravo et Celis (Bravo and Celis 1980) identifièrent une protéine de la phase S du cycle cellulaire qu'ils appelèrent cycline. Il s'agissait en réalité de la même protéine (Mathews, Bernstein et al. 1984).

La protéine PCNA est décrite comme «dancer with many partners » ou *maestro* de la fourche de réplication. En plus d'être un facteur de processivité, c'est une plateforme de chargement de nombreux facteurs à l'ADN. PCNA interagit avec de nombreuses protéines, dites « clientes », de la réplication et de la réparation (Vivona and Kelman 2003; Moldovan, Pfander et al. 2007).

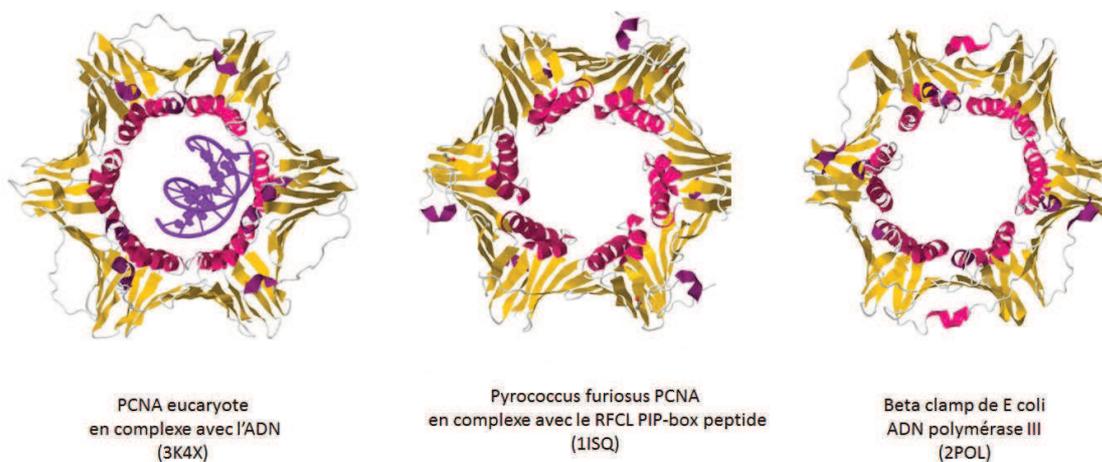


Figure 15: Structure cristalline des « sliding clamp » des trois domaines du vivant

De gauche à droite, le PCNA de *S. cerevisiae*, le PCNA de *P. furiosus* et le β clamp de *E. coli*.

Des homologues de la protéine PCNA ont été décrits chez toutes les Archaea. Les Crenarchaea possèdent plusieurs homologues (PCNA 1, 2, et 3) tandis que les Euryarchaea n'en possèdent généralement qu'un seul. *T. kodakarensis* possède deux homologues de PCNA (PCNA1 et 2), mais un seul, PCNA1, est essentiel (Kuba, Ishino et al. 2012). Le monomère PCNA est constitué de deux domaines reliés par une boucle IDCL (Inter Domain Connector Loop). PCNA adopte une structure tétramérique triédrique exposant les boucles IDCL à l'extérieur de l'anneau. Le trou central inhérent à la forme en « doughnut » du PCNA sert à accueillir l'ADN (Figure 15). Les boucles IDCL sont impliquées dans les interactions entre PCNA et ses partenaires via le motif PIP (« PCNA interacting Peptide ») ou PIP-box). La séquence consensus du PIP-box est « Qxxhxxaa », où « h » représente un résidu aromatique modérément hydrophobe (L, I, M) et « a » un résidu hydrophobe, généralement aromatique (F, Y), les acides aminés « x » peuvent être de tous types. Ce peptide a été identifié dans les protéines Archaea ADN polymérase B, RF-C, Fen-1, l'ADN ligase, (Meslet-Cladière, Norais et al. 2007) ou encore RNaseH2 et NucS chez *Pyrococcus abyssi* (Creze, Ligabue et al. 2012). Chacune de ces protéines interagissent avec le PCNA Archaea et forment des complexes dont la composition peut varier en fonction de la voie métabolique dans laquelle les partenaires du PCNA sont impliqués. Ainsi, un complexe fonctionnel comprenant le PCNA et trois

de ses partenaires : l'ADN polymérase B, l'ADN ligase et l'endonucléase Fen1 a été récemment décrit comme impliqué dans la maturation des fragments d'Okazaki chez *Sulfolobus solfataricus* (Beattie and Bell 2012).

Tableau 3: Protéines interagissant avec le PCNA chez les Archaea.

D'après Vivona and Kelman 2003 et Meslet cladière et al. 2007.

Protéine cliente	Effets de l'interaction avec PCNA	Fonction
DNA polymerase B (Pol B)	Stimulation de l'activité polymérase	Elongation de la réplication
DNA polymerase D (Pol D)	Stimulation de l'activité polymérase	Elongation de la replication
Replication factor C (RF-C)	Chargement de PCNA sur l'ADN	Initiation de la replication
DNA ligase	Stimulation de l'activité ligase	OFF
Flap endonuclease (Fen-1)	Stimulation de l'activité nucléase	OFF
RNase HII	Régulation de l'activité	OFF/Réparation
DNA polymerase Y1 (Pol Y1)	Stimulation de l'activité polymérase	Réparation
Radiation-sensitive mutant 2 (RAD2)		Réparation
Uracil DNA glycosylase (UDG)		Réparation
XPF	Stimulation de l'activité nucléase	Réparation

Le PCNA Eucaryote interagit également avec de nombreux facteurs (Figure 16) intervenant dans les mécanismes de réparation par excision de nucléotides (NER) (Riedl, Hanaoka et al. 2003), par mésappariement de base (MMR) (Umar, Buermeyer et al. 1996), et également par excision de base (BER) (Matsumoto and Kivie Moldave 2001). Lors de ce processus, le PCNA interagit notamment avec les ADN polymérases Eucaryotes β et λ , qui appartiennent à la famille X (Kedar, Kim et al. 2002; Maga, Villani et al. 2002).

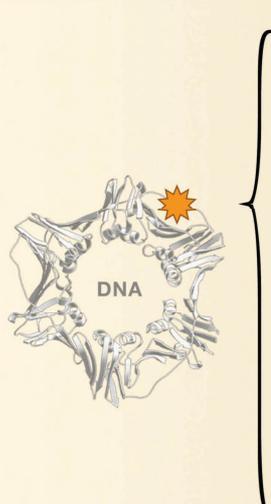
	Activities	Proteins
	DNA polymerases	Pol δ , Pol ϵ , Pol η , Pol ι , Pol κ , Pol ζ , Pol λ , Pol β , Rev1
	Clamp loader	Rfc1, Rfc3, Rfc4
	Flap-endonuclease	FEN-1
	DNA ligase	DNA Ligase 1
	Topoisomerase	Topo II α
	Replication licensing factor	Cdt1
	E3 ubiquitin ligases	Rad18, Rad5
	E2 SUMO-conjugating enzyme	Ubc9
	Helicases, ATPases	Srs2, Rrm3, Mgs1, WRN, RECQ5
	Mismatch repair enzymes	Msh3, Msh6, Mlh1, EXO1
	Base excision repair enzymes	UNG2, MPG, NTH1, hMYH, APE1, APE2, XRCC1
	Nucleotide excision repair enzyme	XPG
	Poly (ADP ribose) polymerase	PARP-1
	Histone chaperone	CAF-1
	Chromatin remodeling factor	WSTF
	Histone acetyltransferase	p300
	Histone deacetyltransferase	HDAC1
	DNA methyltransferase	DNMT1
	Sister-chromatid cohesion factors	Eco1, Chl1, Ctf18
	Protein kinases	CDK2, EGF Receptor
Cell-cycle regulators	p21, p57, Cyclin D1	
Apoptotic factors	Gadd45, ING1b, p53	

Figure 16: Partenaires du PCNA Eucaryote.

2.1.2.3 FACTEUR DE CHARGEMENT DU CLAMP

La structure en anneau des clamps nécessite d'être modifiée, *i.e.* ouverte, pour qu'ils puissent être chargés sur l'ADN. Une famille de protéines est spécialisée dans l'ouverture et le chargement du clamp, les « clamp-loader ». Chez les bactéries c'est le γ -complexe qui réalise cette fonction. Chez les Eucaryotes, le PCNA est chargé sur l'ADN par le RF-C (Replication Factor C). La structure du complexe clamp – clamp loader, ouvert d'un bactériophage est présentée en figure 17.

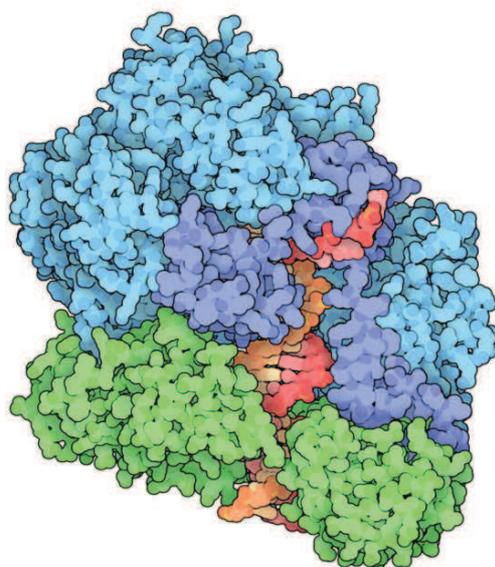


Figure 17: Structure tridimensionnelle du complexe clamp-clamp loader du bactériophage T4.

Ce cristal correspond à l'étape de jonction à la matrice amorcée (« primer template junction»). L'hybride ADN/ARN est en rouge et orange et le clamp loader en bleu. Le clamp, en vert, est en conformation ouverte. Adapté par www.pdb.org (3u5z) de Kelch, Makino et al. 2012.

Des protéines, homologues au RF-C, sont identifiées chez toutes les Archaea. Les protéines RF-C, tout comme PCNA sont très conservées entre les Archaea et les Eucaryotes, tant au niveau de la structure que de la fonction. Le niveau de conservation est tel qu'il est possible de former des complexes hétérologues Archaea/Eucaryotes fonctionnels (Henneke, Raffin et al. 2000). Chez les Archaea, le RF-C possède deux sous unités (grande RF-C L et petite RF-C S). Plusieurs conformations quaternaires du complexe RF-C ont été observées chez les Archaea. En effet, chez *S. solfataricus*, le RF-C formerait un hétéropentamère avec une grande sous-unité et quatre petites (Pisani, De Felice et al. 2000), tandis que chez *P. furiosus*, le RF-C consisterait soit en un hétéropentamère avec une grande sous-unité et quatre petites soit à en un hétérohexamère avec deux grandes sous-unités et quatre petites (Cann, Ishino et al. 2001). Un modèle du mécanisme du chargement ATP-dépendant du PCNA par le RF-C est présenté en figure 18.

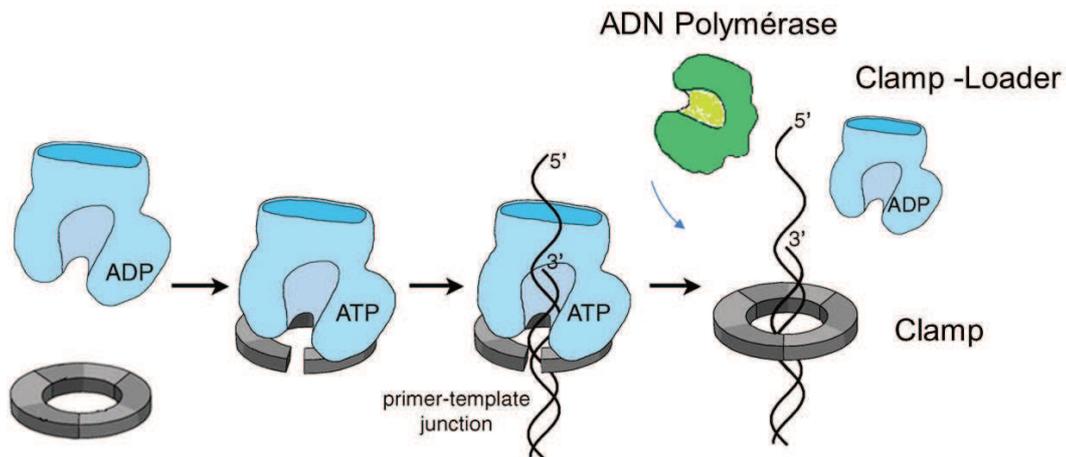


Figure 18: Modèle du chargement du clamp par son facteur de chargement.

Adapté de (Kelch, Makino et al. 2012).

La formation du complexe PCNA - RF-C va induire un changement de conformation de PCNA en ouvrant l'anneau trimerique qui permet le chargement du PCNA à l'ADN amorcé. Le RF-C possède une faible affinité pour l'ADN et le PCNA en absence d'ATP. Après liaison d'une molécule d'ATP, le RF-C se lie au PCNA et induit son ouverture. La liaison du complexe protéique à la matrice amorcée engendre l'hydrolyse de l'ATP, la fermeture du PCNA et l'éjection du RF-C. L'ADN polymérase peut alors être chargée pour la synthèse du brin complémentaire. Chez *P. abyssi*, la stimulation du chargement de PCNA par RF-C ne nécessite pas l'hydrolyse de l'ATP et il a été observé que le PCNA était capable de se charger sur l'ADN en absence du RF-C (Cann, Ishino et al. 1999). Ce chargement spontané peut s'expliquer par le nombre plus faible d'interactions non-covalentes entre les sous unités du PCNA Archaea, comparé aux PCNA Eucaryotes (Matsumiya, Ishino et al. 2001). *In vitro*, le chargement du PCNA sur l'ADN est stimulé par le RF-C mais également par l'ADN polymérase B (Rouillon, Henneke et al. 2007). Le PCNA de *P. abyssi* stimule également l'activité de l'ADN polymérase sur un ADN circulaire en absence de RF-C (Henneke, Raffin et al. 2000).

Le chargement du PCNA sur un ADN hybridé avec une amorce va permettre le recrutement de l'ADN polymérase. La réplication des brins avancés et retardés va générer des formes intermédiaires d'ADN. Les hybrides ADN/ARN, les fragments simples brins flottants (« flap ») et les discontinuités dans la liaison phosphodiester doivent être pris en charge pour obtenir une double hélice d'ADN intègre. Cette étape de la réplication est appelée maturation. Le processus de maturation principal a été identifié par Okazaki, il porte depuis son nom (« Okazaki Fragments Processing », OFP). Lors de la maturation, la polymérase engendre un déplacement de brin créant un simple brin 5' flottant (« flap »), ce brin est clivé par une nucléase et l'ADN ligase comble la discontinuité du squelette phosphodiester exposé. Chez les Eucaryotes et les Archaea, la coordination des activités des enzymes de la maturation est assurée par le PCNA.

2.1.3 MATURATION

La maturation fait intervenir des nucléases spécifiques des structures branchées dans l'ADN. Les nucléases de l'OFP sont RNase H2 pour les Bactéries et Fen-1 pour les Eucaryotes. La liaison phosphodiester entre le dernier nucléotide inséré et le premier du fragment adjacent est réalisée par l'ADN Ligase.

Très récemment, Bell et ses collaborateurs ont présenté un modèle de maturation du fragment d'Okazaki Archaea (Beattie and Bell 2012) (Figure 19). Ils démontrent que la maturation d'Okazaki est orchestrée par le PCNA. *In vitro*, ils réalisent de manière couplée une polymérisation, une élimination du primer ARN et une ligation. Six protéines sont nécessaires ; la polymérase B1, Fen-1, l'ADN ligase 1 et les trois sous unités PCNA (1-3) possédant chacune un site de fixation spécifique sur lesquelles viennent se fixer simultanément les trois acteurs de la synthèse et de la maturation du brin retardé. Ils démontrent que la réaction est plus efficace en présence d'un PCNA hétérotrimérique organisant un complexe avec les 3 autres protéines. Le modèle suggère que chacune des 3 protéines partenaires du PCNA se fixe spécifiquement à une sous unité. Dans le cas où le PCNA est un homotrimer, comme chez *P. abyssi* par exemple, le mécanisme de régulation de la stœchiométrie de complexe conservé est inconnu.

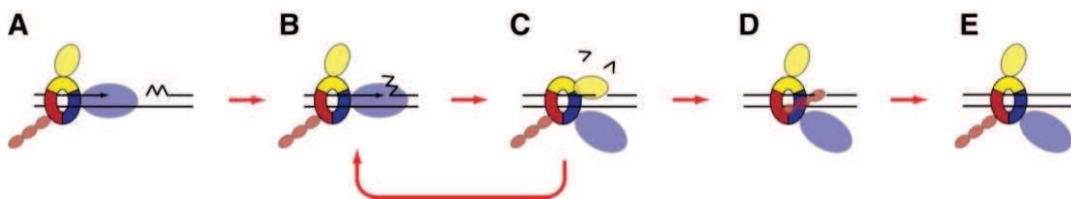


Figure 19 : Modèle de la maturation des fragments d'Okazaki chez les Archaea.

A) PolB1 (bleu), Fen1 (Jaune) et Ligase (rouge) chacune liées à une sous unité PCNA1-3 (Bleu foncé, Jaune foncé et rouge foncé). PolB1 synthétise l'ADN et PCNA transporte la ligase et Fen-1. B) Lorsque le complexe rencontre un fragment d'Okazaki, la polymérase réalise un déplacement de brin. C) Fen1 clive l'ADNsb générant un espace (« nick ») dans la séquence. Ce processus est répété jusqu'à l'élimination totale des fragments d'Okazaki. (D) La ligase se positionne au niveau du nick et catalyse la ligation. E) Le couplage de l'action du complexe s'assure de la conservation de l'intégrité du brin retardé naissant. (Beattie and Bell 2012).

2.1.3.1 FLAP ENDONUCLEASE 1

Fen-1 est une 5'-endonucléase qui reconnaît spécifiquement une structure d'ADN double brin avec un brin 5' non hybridé. L'élimination de ce brin déplacé flottant laissera une encoche (« nick ») dans l'ADN (Figure 19C). PCNA et Fen1 interagissent chez les Eucaryotes et les Archaea par une interaction PIP-box-IDCL (Meslet-Cladière, Norais et al. 2007). L'interaction de FEN-1 à PCNA va induire un changement de conformation de la nucléase stimulant son activité nucléase (Tom, Henricksen et al. 2000). En complexe avec PCNA, FEN-1 adopte trois conformations différentes (Nakane, Nakagawa et al. 2012). La résolution de structures et des études biochimiques de Fen1 Archaea (Matsui, Kawasaki et al. 1999; Matsui, Abe et al. 2004) ont contribué à la compréhension des mécanismes moléculaires conservés, à l'origine du clivage du flap. La protéine Fen-1 Archaea semble également être impliquée dans la réparation puisque la délétion du gène *fen1* de *H. volcanii* induit, en plus d'un ralentissement du cycle cellulaire, une sensibilité aux UV (Meslet-Cladière, Norais et al. 2007).

2.1.3.2 RNASE H II

Une deuxième nucléase pourrait également être impliquée dans la maturation de la réplication chez les Archaea, la Ribonucléase HII. Cette protéine est l'homologue de la nucléase impliquée dans la maturation chez les bactéries. La Ribonucléase H Eucaryote est, quand à elle, impliquée dans la réparation de l'ADN. Elle en reconnaît et élimine les ribonucléotides uniques insérés dans l'ADN (Cerritelli and Crouch 2009), en collaboration avec PCNA (Bubeck, Reijns et al. 2011). La délétion de RNase H2 chez les mammifères résulte en une augmentation du nombre de ribonucleotides insérées dans l'ADN génomique, une augmentation du taux de cassures de l'ADN, et en l'activation de la réponse cellulaire par les voies de réparation des dommages de l'ADN (Hiller, Achleitner et al. 2012).

La seule protéine homologue à une RNase H chez *P. abyssi* est la RNaseH II. Elle présente un PIP-Box (Meslet-Cladière, Norais et al. 2007) et possède une spécificité pour les substrats hybrides ADN/ARN (Le Laz, Le Goaziou et al. 2010). La formation d'un site de ligation par Fen-1 sur un substrat branché, est favorisée par la présence de RNase HII, *in vitro* (Henneke 2012). Ces données biochimiques suggèrent son implication dans la maturation des fragments d'Okazaki et de l'existence des deux voies, dépendantes ou non de RNase HII, de maturation chez les Archaea. Cependant, une étude génétique avait préalablement montré que la délétion du gène *rnaseHII* chez l'Archaea *H. volcanii* n'a pas d'effet sur le temps de doublement cellulaire mais induit par contre une légère sensibilité aux UV (Meslet-Cladière, Norais et al. 2007). La délétion simultanée des gènes *rnaseHII* et *fen1* provoque un phénotype encore plus sensible que le mutant unique *fen1*. Ces données génétiques penchent en faveur d'une implication de la RNase HII dans la réparation des ribonucléotides uniques comme chez les Eucaryotes. Le rôle biologique de la RNase HII chez les Archaea est donc, à ce jour, une question toujours ouverte.

2.1.3.3 ADN LIGASE

Les ADN ligases sont des molécules capables de lier des brins d'ADN par leurs extrémités en formant une liaison phosphodiester (Doherty and Suh 2000). Les ligases interviennent dans les mécanismes de réplication, de réparation et recombinaison. Lors de la réplication, l'ADN ligase catalyse la formation de la dernière liaison entre le 5'P et le 3'OH de chaque fragment d'Okazaki. Cette enzyme catalyse la réaction de ligature des encoches via trois réactions de transfert de nucleotidyl (Wilkinson, Day et al. 2001; Tomkinson, Vijayakumar et al. 2006). *Primo*, l'ADN ligase est activé par la liaison à un intermédiaire AMP en utilisant un ATP ou NAD⁺ en cofacteur de la réaction. *Secundo*, l'enzyme reconnaît son substrat ADN et l'AMP est transférée au phosphate 5' de l'ADN formant un intermédiaire ADN adenylé (AppADN). *Tertio*, le 5'AppDNA est sujet à une attaque nucléophile du groupement 3'hydroxyl, formant ainsi la dernière liaison phosphodiester. Ainsi, les ADN Ligases effectuent la dernière étape de la maturation correspondant d'une certaine façon à la terminaison de la réplication. Un schéma général de la réplication chez *P. abyssi* est présenté en figure 20.

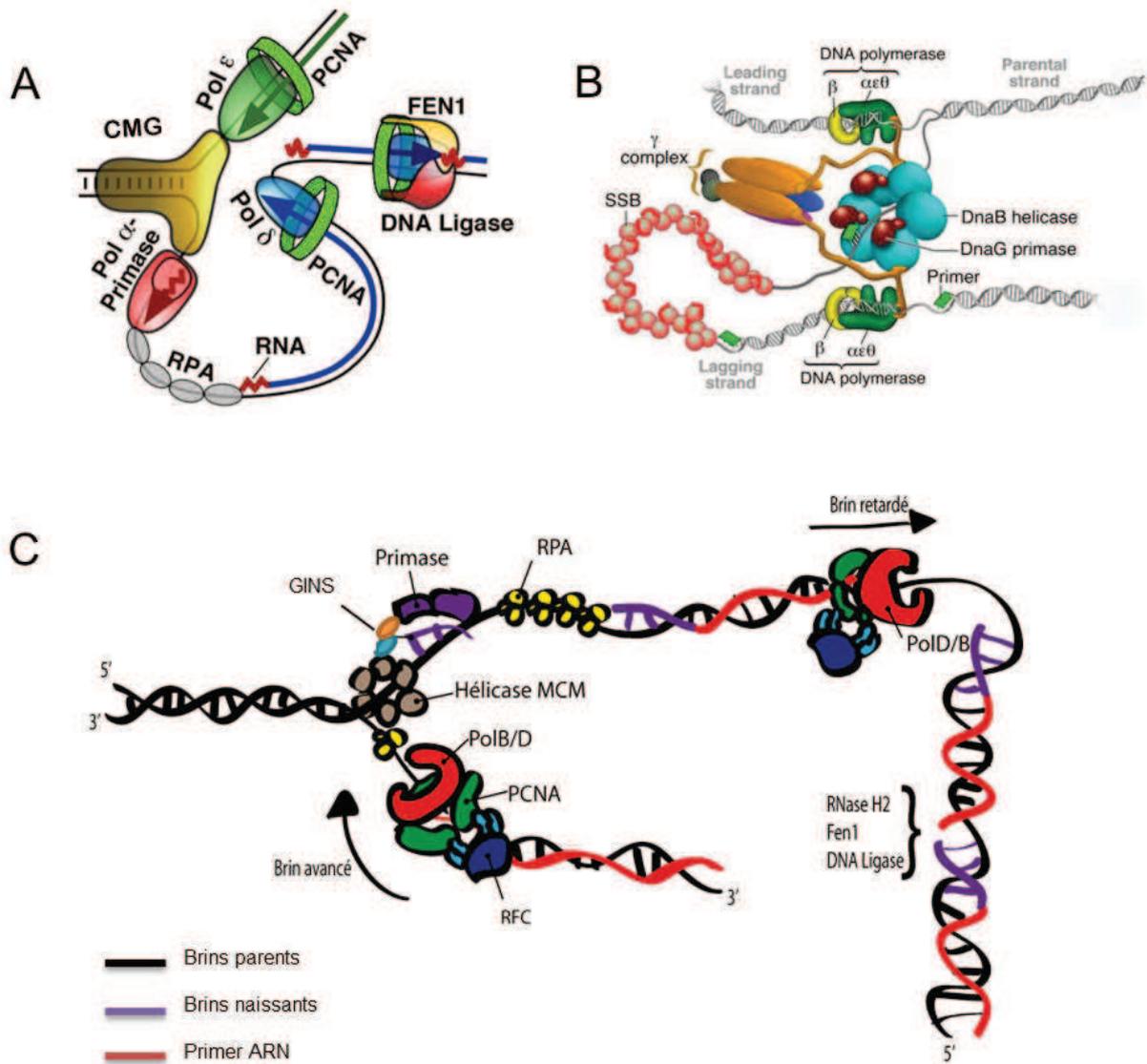


Figure 20 : Fourches de répliation.

A) Eucaryote B) Bactérienne C) Archaea

Après l'exposé des données bibliographiques concernant le mécanisme de répliation chez les Archaea, il s'avère que ce processus est relativement bien décrit, même si des données génétiques doivent encore confirmer certaines hypothèses basées sur les caractérisations biochimiques et structurales des acteurs du réplisome. Par exemple, aucune donnée fonctionnelle, *in vivo*, ne permet à ce jour de confirmer l'implication des Pol B et D des Euryarchaea dans le mécanisme de répliation. A contrario, les données sur les voies de réparation à l'œuvre chez les Archaea sont beaucoup plus parcellaires et ne nous permettent pas encore de déterminer quels sont les types de lésions apparaissant fréquemment dans le chromosome des Archaea et spécialement des archaea hyperthermophiles et quels sont les mécanismes moléculaires au sein des principales voies de réparations de ces dommages? La répliation peut être perturbée par des dommages sur le substrat ADN qui peuvent engendrer des arrêts de la fourche de répliation ou des mutations dont l'accumulation sera

délétère à court ou à long terme pour l'organisme. Dans les 3 domaines du vivant il existe plusieurs systèmes protéiques spécialisés dans la détection et la réparation des différents types de dommages recensés.

2.2 REPARATION DE L'ADN

La réparation est le mécanisme en charge de la détection et de la correction des erreurs de réplication ou de dommages sur la molécule d'ADN. La réparation de l'ADN est un processus essentiel à la conservation de l'intégrité de la structure de l'ADN et de l'information qu'elle contient. Les agressions chimiques ou physiques peuvent briser la liaison phosphodiester ou endommager les bases (Figure 21).

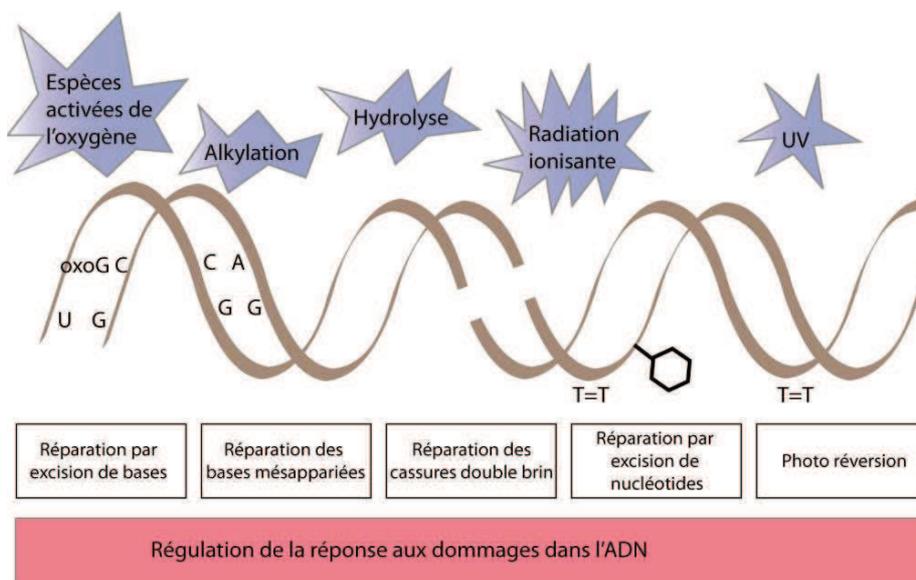


Figure 21: Les grandes voies de réparation de l'ADN et types de lésions associés

Tous les organismes vivants possèdent des systèmes de réparation de l'ADN. Il existe 7 voies principales de la réparation de l'ADN, deux de ces voies permettent la réversion du dommage (photo-réversion, transfert de groupement alkyl), trois autres impliquent l'élimination de la lésion avant réparation par synthèse d'ADN (réparation par excision de bases et de nucléotides et réparation des bases non appariées), les deux dernières permettent la réparation des cassures double brin (recombinaison homologue et liaison des extrémités non homologues ; NHEJ) tandis que l'action d'ADN polymérase translésionnelles permet à la fourche de réplication de passer les dommages sans provoquer son effondrement.

La cellule possède des protéines spécialisées dans la détection des dommages. A chaque type de dommage est associé une voie de réparation spécifique. C'est la détection d'un type de dommage particulier qui va déclencher l'entrée dans une voie de réparation spécialisée. L'étude des voies de réparation de l'ADN est un des champs les plus actifs de la biologie actuelle de par son intérêt fondamental et appliqué. Trois grands systèmes de réparation ont été décrits : la réversion du dommage, la réparation des cassures doubles brins et la réparation par excision de bases. Ils ont été identifiés chez les Eucaryotes et les Bactéries. La connaissance accumulée sur la réplication Archaea en fait un mécanisme désormais quasiment aussi bien décrit que chez les Bactéries et les Eucaryotes. Cependant, les systèmes protéiques en charge de la réparation de l'ADN chez les

Archaea sont toujours assez mal compris. La recherche d'homologues de protéines de la réparation dans les génomes Archaea a permis d'identifier plusieurs protéines de la réparation de type Eucaryote. En revanche de nombreux acteurs essentiels n'ont pas été identifiés et semblent absents des génomes Archaea.

2.2.1 REPARATION PAR REVERSION DU DOMMAGE

La réparation par réversion du dommage permet la réparation d'une base sans modification du squelette phosphodiester de l'ADN. La réparation par réversion du dommage fut la première réaction de réparation de l'ADN identifiée (Beukers and Berends 1960). La protéine CPD est une photolyase qui catalyse la monomérisation, dépendante de la lumière, des dimères de cyclobutyl-pyrimidine, formés entre base adjacente du même brin par les rayonnements ultraviolets. Une CPD de classe II a été caractérisée chez le méthanogène *Methanosarcina mazei* (Kiontke, Geisselbrecht et al. 2011). Un autre exemple de réversion des dommages est la réparation des O⁶-alkyl guanine par transfert irréversible du groupe alkyl de l'ADN à un résidu cystéine de la protéine. La protéine est ensuite dégradée. Ces alkyl transférases sont conservés dans l'ensemble du vivant et sont appelés AGT, OGT, ou MGMT (EC: 2.1.1.63). Le processus d'élimination des dommages par MGMT est présenté en figure 22. L'élimination des O⁶-alkylguanine par les alkyl transférases est un sujet de recherche très vaste (Pegg 2000; Perugino, Vettone et al. 2012). Pour les thermophiles, la réparation des lésions alkyl est cruciale car à haute température les bases alkylées sont rapidement converties en sites abasiques ou en cassure de brin. Une étude de l'activité de l'ADN alkyltransférase de *S. solfataricus* (SsOGT) a démontré sa grande efficacité à réparer les bases alkylées (Perugino, Vettone et al. 2012).

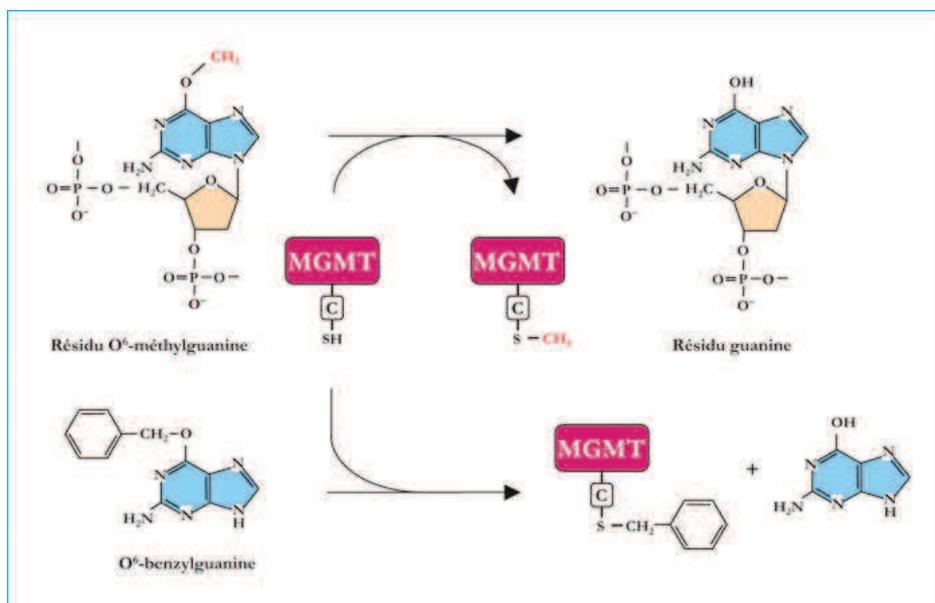


Figure 22 : Réparation des résidus O6-méthylguanine par la protéine MGMT

2.2.2 REPARATION PAR EXCISION DE NUCLEOTIDE NER

La réparation par excision de nucléotide (NER) est un système de réparation spécialisé dans la détection et le remplacement des nucléotides endommagés par des agents mutagènes, des ultraviolets ou des rayonnements ionisants. Il a été identifié chez les Eucaryotes et les Bactéries. Le NER consiste en une excision du fragment simple brin autour du dommage, suivi d'une synthèse par une ADN polymérase. L'importance de ce mécanisme de réparation est particulièrement illustrée chez l'homme où des mutations de protéines du NER provoquent des pathologies sévères telles que le Xeroderma pigmentosum et le syndrome de Cockayne. Elles sont causées par une hypersensibilité aux dommages de l'ADN.

Chez les Bactéries le NER est pris en charge par le complexe UvrABC. Ce complexe hélicase/endonucléase est en charge de la détection et de l'excision du fragment d'ADN endommagé. Chez les Eucaryotes, la machinerie NER est beaucoup plus complexe dans la mesure où plusieurs dizaines de protéines sont impliquées (Riedl, Hanaoka et al. 2003). Egalement, deux formes de NER ont été identifiées : le Global Genome Repair (GGR) et le Transcription Coupled Repair (TCR) (Fousteri and Mullenders 2008; Hanawalt and Spivak 2008). Ils se différencient par la nature de la détection du dommage. Dans le GGR, c'est la détection d'une distorsion dans la structure de l'ADN causée par un dommage qui va déclencher la reconnaissance du dommage et l'entrée dans la voie de réparation. Tandis que dans le TCR, le dommage est détecté au cours de la synthèse des transcrits par l'ARN polymérase. Il est suggéré que les Archaea utilisent une version simplifiée du NER Eucaryotes et que les homologues des protéines UvrABC identifiées chez certaines Archaea mésophiles aient été héritées d'un transfert latéral de gènes avec les Bactéries (Kelman and White 2005). Dans la majorité des cas, et notamment pour les hyperthermophiles, il n'existe pas d'homologues du système NER bactérien. En revanche, plusieurs homologues des protéines du NER Eucaryotes ont été identifiés dans les génomes Archaea (Rouillon and White 2010) (Figure 23). On retrouve généralement les nucléases XPF et Fen1 (XPG) et les hélicases XPB et XPD. Ces deux dernières sont également impliquées, chez les Eucaryotes, dans le complexe d'initiation de la transcription Eucaryote (TFIIH). A ce jour, il n'existe pas de preuves expérimentales de l'existence d'un NER spécifique aux Archaea. Les homologues des protéines impliquées dans la reconnaissance du dommage (XPA et XPC), chez les Eucaryotes, sont de plus absents des génomes Archaea. Cette absence suggère l'existence d'un mécanisme alternatif de détection des dommages. Les protéines SSB ou RPA sont parmi les candidats les plus sérieux pour assurer cette fonction de reconnaissance des dommages. *In vivo*, une surexpression de RPA de *Halobacterium* NRC-1 est corrélée à une résistance accrue aux radiations ionisantes et aux UV (DeVaux et al 2007, McCready 2005). Les protéines Bax1 et PCNA sont probablement impliquées dans le NER Archaea (Rouillon and White 2010).

Chez les Crenarchaea, la protéine SsoSSB possède la capacité de reconnaître les bases ADN mésappariées, un photo-produit de l'ADN ou une distorsion de la double hélice provoquée par un dommage (Cubeddu and White 2005). Il est suggéré que la liaison de SSB à l'ADNdb est rendue possible par la déstabilisation locale de l'ADN due au dommage et que cette reconnaissance initierait le recrutement des premières protéines effectrices de la réparation NER.

L'absence des protéines de la détection des dommages du GGR peut suggérer l'existence chez les Archaea d'un TCR très efficace. Dans l'hypothèse d'un TCR Archaea c'est la RNA Polymérase (RNAP) qui constituerait la protéine en charge de la détection des dommages. Cependant il a été démontré que les brins transcrits de *S. solfataricus* ne sont pas plus efficacement réparés (Romano, Napoli et al. 2007). Egalement, aucun homologue des protéines essentielles au TCR bactérien et eucaryotes (Mfd et CSA/CSB, respectivement) n'a été identifié chez les Archaea pour l'instant (Dorazi, Götz et al. 2007). Il est possible que l'extrême efficacité du CGR Archaea, s'il existe, rende non indispensable l'existence d'un TCR.

Pour résumer, il n'y a donc à ce jour que très peu d'évidences expérimentales de l'existence d'un tel mécanisme chez les Archaea hyperthermophiles. Cependant, la plupart de ces dernières possèdent les gènes codant pour les protéines du NER de type eucaryote, ce qui suggère que les HA utiliseraient une version simplifiée du NER eucaryote.

Cette hypothèse a été formulée par l'équipe de Malcolm White, suite à de nombreux travaux sur les homologues archaea des protéines du NER. Lui et ses collaborateurs ont pu montrer que les activités hélicases de XPD et XPB permettent l'ouverture du double brin portant la lésion. Le brin endommagé est ensuite incisé par les nucléases 3' flap et 5' flap, XPF et Bax1. D'autre part, il a également suggéré les participations hypothétiques du SSB, équivalent du RPA chez les Crenarchaea, pour la détection des lésions, et de l'endonucléase NucS pour la coupure du 3'flap (Figure 23) (Roberts, Bell et al. 2003; Cubeddu and White 2005; Newman, Murray-Rust et al. 2005; Roberts and White 2005; Roberts and White 2005; Rudolf, Makrantonis et al. 2006; Hutton, Roberts et al. 2008; Liu, Rudolf et al. 2008; Richards, Cubeddu et al. 2008; White 2009; Hutton, Craggs et al. 2010; Rouillon and White 2010; Rudolf, Rouillon et al. 2010; Rouillon and White 2011).

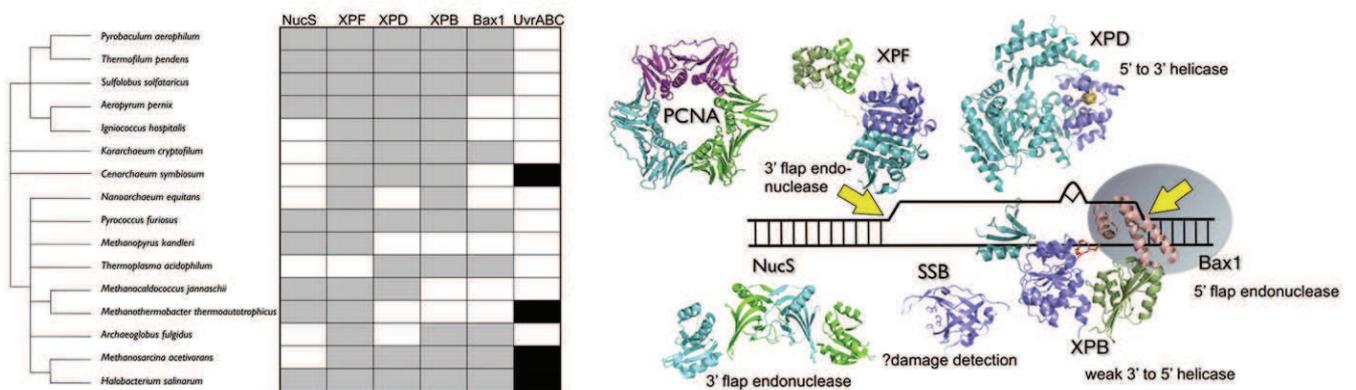


Figure 23 : Protéines du NER Archaea et leur distribution

(Rouillon and White 2010)

2.2.3 REPARATION PAR EXCISION DE BASE BER

Le système de réparation par excision de base (BER) est en charge de l'identification et de la réparation des bases modifiées ou endommagées par des processus chimiques tels que l'oxydation ou l'alkylation des bases, la méthylation ou la déamination des cytosines en uracile (Kelch, Makino et al. 2011). Ces lésions sont fréquentes chez tous les organismes et la fréquence de leur apparition augmente drastiquement avec la température. Le BER est un processus séquentiel initié par la détection de la base endommagée et son excision générant un site AP (apurinique ou apyrimidique). L'étape suivante est la prise en charge par une nucléase spécialisée dans la coupure du sucre en 5' du site AP. Il existe deux variantes de la suite de ce processus de réparation; soit « Single Nucleotide Repair » (SN-BER); soit « Long Patch » (LP-BER). Le nombre de nucléotides excisés et les protéines en jeu varient selon le processus. Dans le SN-BER Eucaryote, c'est la polymérase β qui remplace le nucléotide endommagé et qui excise le groupe 5'dRP grâce à son activité dRPase. L'extrémité 5' phosphate est ensuite ligaturé à l'extrémité 3' OH par l'ADN Ligase. Dans le LP-BER, la polymérase ajoute plusieurs nucléotides générant ainsi un flap par déplacement de brin. Le fragment d'ADN simple brin libre résultant est subséquemment clivé par Fen-1 et les fragments jointifs sont ligaturés par l'ADN ligase. Les processus des BER Eucaryote et Archaea sont présentées en figure 24.

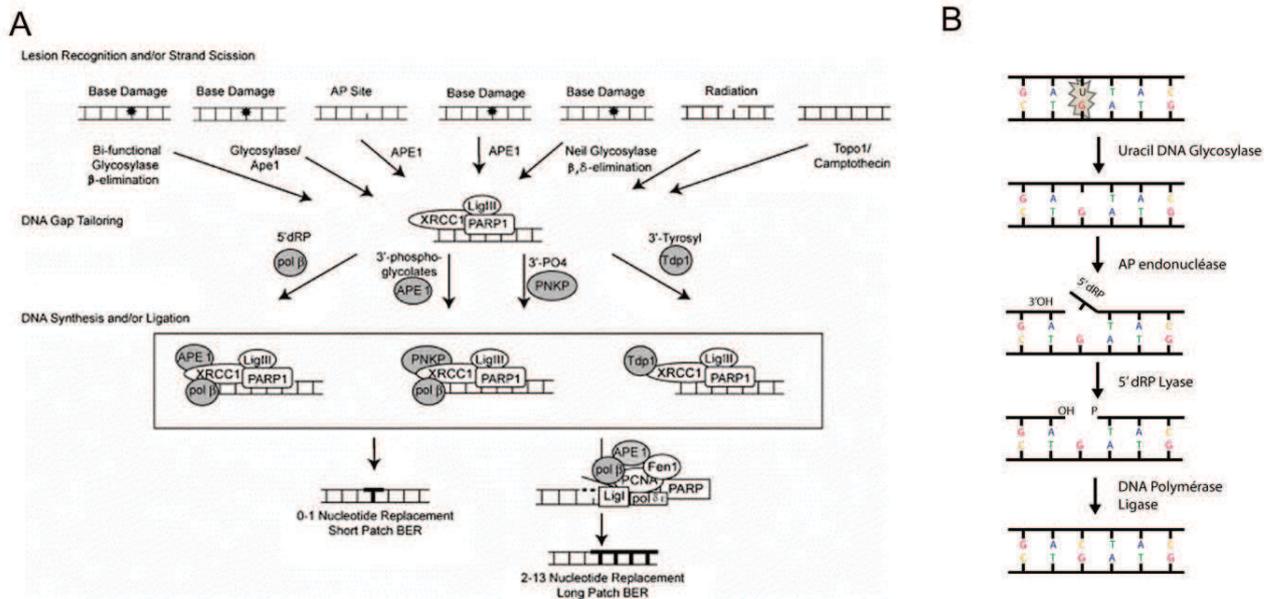


Figure 24: Base Excision Repair.

A) Modèle unifié du BER Eucaryote (Kelch, Makino et al. 2011) B) Modèle BER Archaea.(Sartori and Jiricny 2003)

La voie de réparation par excision de bases a été entièrement disséquée *in vitro*, sur le plan enzymologique, chez *Pyrobaculum aerophilum*, représentante du domaine des Crenarchaea. La voie complète fait intervenir une uracile-DNA glycosylase capable de reconnaître un mésappariement G-U dans l'ADN, formé par l'incorporation d'uracile au sein du brin néo-synthétisé sous forme de dUMP lors de la réplication. Cette enzyme reconnaît et élimine l'uracile, générant ainsi un siteapurinique(AP). Ce site AP est ensuite coupé du côté 5' par une AP endonucléase, l'extrémité 3' du brin incisé servant alors de primer pour une ADN polymérase de

réparation pour corriger le mésappariement, en l'occurrence une ADN polymérase B dont il existe 3 représentants chez *P. aerophilum*. Après élimination du résidu bloquant en 5' (dRP) par l'activité dRP lyase de la polymérase ou de l'ADN ligase, cette dernière raboute les deux extrémités adjacentes (Sartori and Jiricny 2003). Ces données suggèrent que le BER Archaea est conservé avec les Bactéries et les Eucaryotes.

2.2.4 REPARATION DES MESAPPARIEMENTS (MISMATCH REPAIR, MMR)

Le système de réparation des mésappariements intervient lorsque qu'une base non complémentaire à la matrice est insérée, notamment lors de la réplication par les ADN polymérases répliquatives ou translésionnelles. Un système de réparation des mésappariements a été identifié chez les Eucaryotes et les Bactéries. Les différentes voies du MMR chez les Eucaryotes et les Bactéries sont présentées en figure 25.

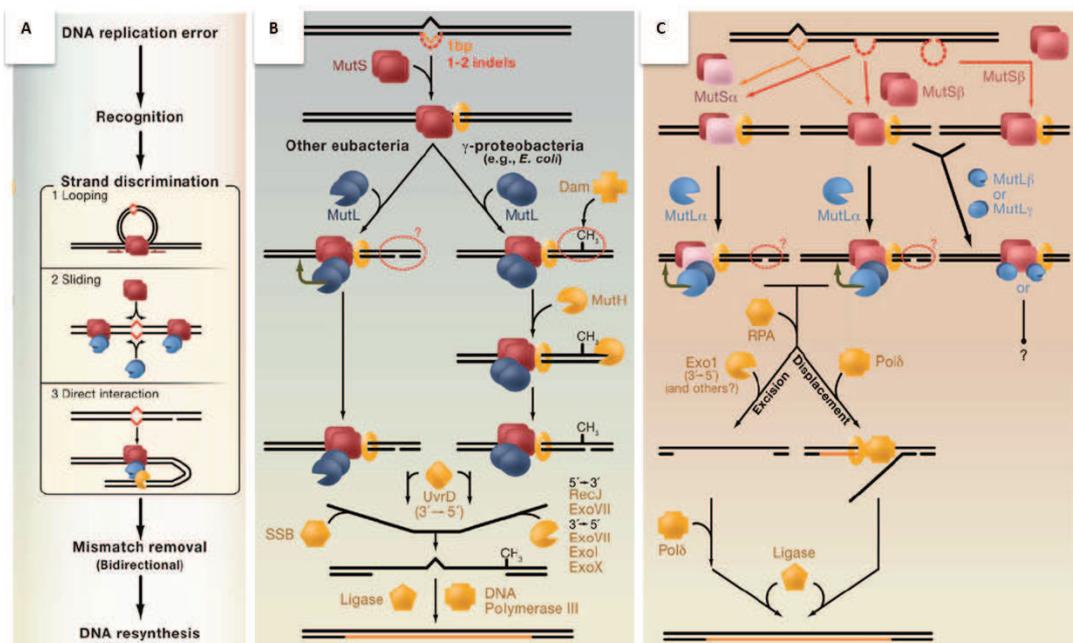


Figure 25 : MMR chez les bactéries et les eucaryotes.

A) Mécanisme général. B) MMR bactérien. Le MMR est initié lorsqu'un homodimère de MutS se lie à un ADN contenant un mis match. Un homodimère de MutL collabore avec MutS dans la reconnaissance du brin naissant contenant l'erreur. Chez *E. coli*, c'est l'absence de méthylation des adénines du brin fils qui permet la discrimination des brins. Après l'intervention de l'adénine méthyle-transferase Dam, le brin non méthylé est clivé par MutH. Le clivage va générer un nick qui permettra l'élimination du mis match par l'action couplée de UvrD, SSB et de nucléases spécifiques du simple brin (Endonucléases I, VIII, X et RecJ). L'ADN polymérase II synthétise à nouveau le brin qui contenait l'erreur et l'ADN ligase complète la réparation. C) MMR Eucaryote. Le MMR Eucaryote fonctionne similairement au système bactérien. Il implique cependant des complexes de protéines MutS et MutL différentes selon le type de mismatch détecté (cf tableau 4). C'est la protéine MutL qui est en charge de l'incision. L'incision formée par MutL permettra l'activité nucléase de l'Exo1, la resynthèse par pol δ , et la ligation.

Les protéines clés du MMR, MutS et MutL, sont très conservées chez les Bactéries et les Eucaryotes. MutS possède une activité ATPase et un motif de liaison à l'ADN (Acharya, Foster et al. 2003). MutL est un dimère en solution, elle fait partie des protéines de la famille Bergerat ATPase/kinase, elle intervient dans la formation du complexe en charge de l'incision (Bergerat, de Massy et al. 1997; McCreedy, Muller et al. 2005). Chez *E. coli*, le

MMR fait intervenir 12 protéines dont MutS, MutL et une 3^{ème} protéine spécifique au MMR : l'endonucléase MutH (Marti, Kunz et al. 2002) (Figure 25B). Chez les Bactéries, on estime le taux de mésappariement à un pour 10⁶ à 10⁷ nucléotides répliqués. Cette fréquence est multipliée par 100 lorsque les gènes MutS et MutL sont inactivés (Matic, Radman et al. 1997). Les Eucaryotes possèdent de multiples homologues de MutS et MutL formant des heterodimères au sein d'un système MMR plus complexe (Schofield and Hsieh 2003; Mackwan, Carver et al. 2007) (Figure 25B). Aucun homologue Eucaryote de la protéine MutH n'a été identifié.

Tableau 4 : Protéines du MMR et fonctions chez les Bactéries et les Eucaryotes

Protéine <i>E. coli</i>	Fonction	Homologues Eucaryotes	Fonction
MutS	Liaison au mésappariement	MSH2-MSH6 (MutS α)	Réparation des mésappariement simples et 1-2 bases
		MSH2-MSH3 (MutS β)	Réparation de certains mésappariement simples et supérieurs à 2 bases Partiellement redondant avec Msh2-Msh6
MutL	Coordination des complexes Mut	MLH1-PMS2 (γ PMS1) (MutL α)	Coordination des complexes
		MLH1-MLH2 (hPMS1) (MutL β)	Fonction inconnue
		MLH1-MLH3 (MutL γ)	Participe à la meiose
MutH	Incision du brin contenant l'erreur, en opposé d'un site hemiméthylé	-	
γ - δ Complex	Chargement du clamp	RFC complex	Charge le PCNA au dommage et regule la polarité de l'excision.
β -Clamp	Interagit avec MutS et pourrait être recruté à la fourche de réplication.	PCNA	Interagit avec MutS et MutL
	Augmente la processivité de DNA pol III		Recrutement du MMR au mesappariement Augmente la spécificité de la reconnaissance du dommage de Msh2-Msh6 Participe à l'excision et probablement à la signalisation du dommage Participe a la polymérisation
Hélicase II	Chargé à l'incision par MutS et MutL	-	
	Ouvre l'ADN préalablement à l'excision		
ExoI, ExoX	Excision 3'5' de l'ADNsb	EXOI (Rth1)	Excision de l'ADNdb
RecJ	Excision 5'3' de l'ADNsb	3'exo de pol δ	Excision de l'ADNsb
ExoVII		3'exo de pol η	Augmente la promotion des mutations générées par la délétion de l'ExoI
DNA pol III	Polymérisation fidèle	DNA pol δ	Polymérisation fidèle
SSB	Participe à l'excision et la synthèse de l'ADN	RPA	Participe à l'excision et la synthèse de l'ADN
DNA ligase	Ligation	DNA ligase	Ligation

Chez *Saccharomyces cerevisiae*, il existe 6 homologues mutS (MSH1-6), et 4 Homologues mutL (MLH1-3) et PMS1 (Prolla, Christie et al. 1994). Différentes combinaisons de ces gènes jouent un rôle dans la réparation des bases non-complémentaires insérées en face d'un dommage pendant la réplication (Earley and Crouse 1998) mais les protéines Mut eucaryotes ont un spectre d'activité plus large (Nakagawa, Datta et al. 1999). Le mécanisme de recrutement du MMR, au niveau des sites de mésappariement, est cependant assez mal décrit

chez les Eucaryotes., L'implication de protéines de la réplication, notamment le PCNA, suggère que le MMR pourrait être associé à la machinerie réplivative (Johnson, Kovvali et al. 1996; Umar, Buermeyer et al. 1996)..

Tableau 5 : Gènes Archaea codants pour des homologues des protéines Mut. (Grogan 2004).

	MutS	MutL
	COG0249	COG0323
ARCHAEA		
<i>Methanopyrus kandleri</i>	--	--
<i>Pyrococcus horikoshii</i>	--	--
<i>Pyrobaculum aerophilum</i>	--	--
<i>Aeropyrum pernix</i>	--	--
<i>Methanococcus jannashii</i>	--	--
<i>Archaeoglobus fulgidus</i>	--	--
<i>Sulfolobus solfataricus</i>	--	--
<i>Methanobacterium thermoautotrophicum</i>	--	--
<i>Thermoplasma acidophilum</i>	--	--
<i>Halobacterium salinarum</i>	VNG0172	VNG0159
<i>Methanosarcina mazei</i>	MM1683	MM1682

La majorité des Archaea ne possède pas d'homologue de protéines Mut. Ainsi le mécanisme en charge des mésappariements chez les Archaea n'est pas identifié (Tableau 5) (Grogan 2004). Les Archaea qui possèdent des homologues des protéines du MMR bactérien sont mésophiles, cela suggère un héritage des gènes Mut par transfert latéral tout comme pour XPD (Tableau 4). De plus, la délétion des homologues de type bactérien identifiés chez *H. salinarum*, MutS et MutL, ne provoque pas d'effet notable (Busch and DiRuggiero 2010) suggérant une fonction autre que dans la maintenance génomique chez les Archaea. Etant donné l'importance des gènes *mut* chez les Eucaryotes et les Bactéries (Matic, Radman et al. 1997), les Archaea doivent posséder des voies alternatives de réparation des mésappariements. L'une des explications possible à l'absence de MMR est que les Archaea possèdent un système de réparation alternatif qui prend en charge les dommages faisant intervenir MMR ou NER chez les autres organismes (Grogan 2004; Busch and DiRuggiero 2010) ;

2.2.5 REPARATION APRES UNE CASSURE DOUBLE BRIN

Les cassures doubles brin de l'ADN (« Double Strand Break », DSBs) sont des lésions hautement toxiques et qui génèrent une forte instabilité génomique. Pour préserver l'intégrité du génome, plusieurs systèmes de réparation des DSBs ont été développés. Les deux voies les plus représentatives sont la jonction des extrémités non-homologues (« Non Homologous End Joining », NHEJ) et la recombinaison homologue (RH) (« Homologous Recombination », HR) (Figure 24) (Chapman, Taylor et al. 2012). Le type de voie de réparation engagée dépend du type de dommage et de l'état énergétique de la cellule (Chapman, Taylor et al. 2012).

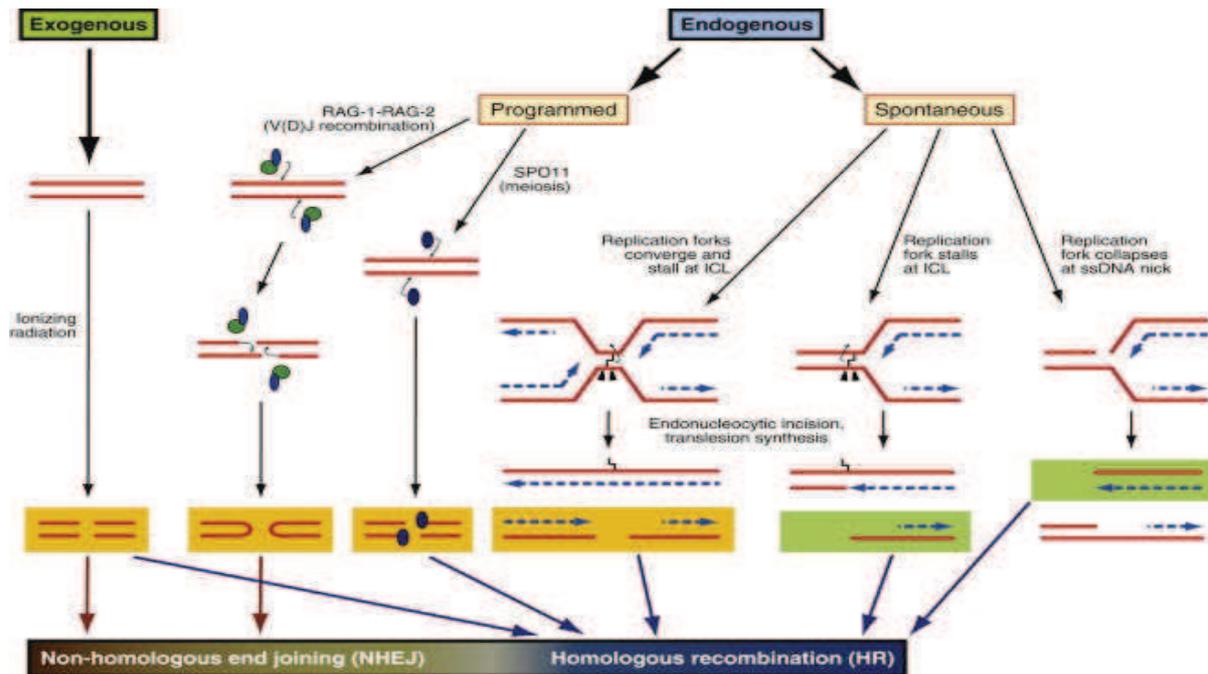


Figure 26: Dommages provoquant un DSB et voies de réparation

(Chapman 2012)

2.2.5.1 RECOMBINAISON HOMOLOGUE

La recombinaison homologue est un processus d'échange d'informations entre deux duplexes d'ADN. Elle est le processus fondamental à la base du brassage génomique qui étend la diversité génétique et également un processus essentiel de la réparation des DSBs. La description de la recombinaison sera limitée dans le manuscrit au contexte de la réparation. La RH classique commence par un DSB provenant d'une action enzymatique ou d'un dommage de l'ADN. Les extrémités sont dégradées pour produire une extrémité 3' sortante. Ces extrémités seront utilisées pour amorcer une invasion de brin (« stand-invasion ») qui permettra la formation d'une jonction de Holliday (« Holliday Junction », HJ). Les HJs sont des structures à 4 brins d'ADN hybridés, elles sont mobiles et peuvent migrer sur l'ADN. Les HJs sont résolues par une enzyme spécialisée la HJ-resolvase, cette nucléase libèrera les brins qui seront ensuite ligaturés par l'ADN ligase complétant ainsi le processus de recombinaison homologue. Les protéines qui catalysent le HR Archaea sont hautement conservées et quasiment ubiquitaires. Les protéines Mre11, Rad50 et RadA sont conservées avec les Eucaryotes. Les protéines Hjc, HerA et NurA sont spécifiques aux Archaea, suggérant l'existence de mécanismes alternatifs pour résoudre les DSBs.

2.2.5.1.1 RESECTION DES EXTREMITES

La première étape de la HR est la résection qui formera l'extrémité 3' sortante permettant l'invasion de brin. Chez *E. coli*, la reconnaissance et la résection sont catalysées par le complexe RecBCD ou par RecQ et RecJ ou encore par la voie AddAB (Niu, Raynard et al. 2009). Ce complexe possède une activité hélicase et nucléase

(Singleton, Dillingham et al. 2004). Cette étape est prise en charge par le complexe Mre11 – Rad50 (MR) chez les Eucaryotes (D'Amours and Jackson 2002). Ces protéines sont conservées chez les Archaea et sont souvent codées en opéron avec l'hélicase HerA (Constantinesco, Forterre et al. 2004) et la nucléase NurA (Constantinesco, Forterre et al. 2002). D'autre part, HerA interagit avec le complexe MR *in vitro* (Quaiser, Constantinesco et al. 2008),(Zhang, Wei et al. 2008). Le modèle de la résection Archaea est présenté en figure 27.

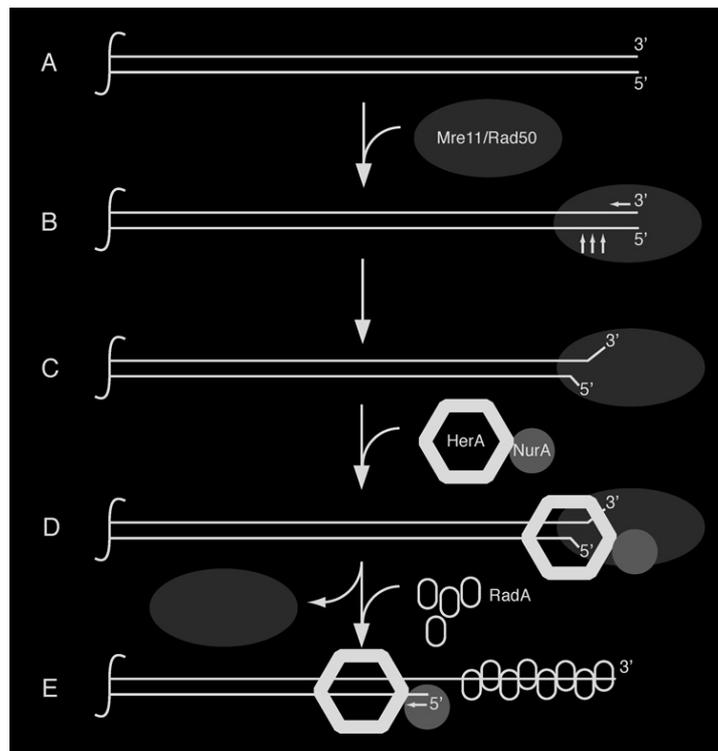


Figure 27 : Modèle de résection des DSB chez les Archaea.

Le complexe MR se lie aux extrémités 5' et 3' (A) et facilite l'ouverture des brins aidant au recrutement de HerA/NurA (B). La resection processive est catalysée par HerA/NurA (C et D) produisant le substrat simple brin de RadA (E)

Tous les génomes Archaea séquencés à ce jour possèdent donc des homologues des protéines eucaryotes Mre11 et Rad50 (MR). Les études biochimiques du complexe MR de *P. furiosus* indique que la fonction est conservée avec les complexes Eucaryotes et Bactériens. Le complexe MR de *P. furiosus* possède cependant une activité 3'-5' exonucléase dont l'orientation de l'activité est opposée à celle nécessaire pour l'initiation de la RH. Cela suggère la participation d'autres facteurs protéiques pour une résection efficace. Mre11 et Rad50 sont régulièrement observés en opéron avec les gènes *herA* et *nurA* Constantinesco, Forterre et al. 2002; Constantinesco, Forterre et al. 2004). Les 4 gènes sont co-transcrits chez *S. acidocaldarius*. HerA est une hélicase ATP-dépendent bidirectionnelle pouvant ouvrir l'ADN possédant une extrémité sortante 3' ou 5'. NurA définit une nouvelle famille de nucléases possédant à la fois une activité endonucléase simple brin et une activité 5'-3' exonucléase sur l'ADN simple et double brin (Constantinesco, Forterre et al. 2002; Constantinesco,

Forterre et al. 2004; Manzan, Pfeiffer et al. 2004). C'est la coopération des 4 protéines qui permet la résection 5' de l'ADN double brin (Hopkins and Paull 2008). Aucun homologue eucaryote de HerA et NurA n'a été identifié. La résection génère des extrémités simple brin, ainsi les SSBs sont susceptibles de s'y lier. Les SSBs doivent être éliminées pour autoriser la formation du filament RadA, essentiel à l'échange de brin. Le SSB de *S. Solfataricus* inhibe la formation d'une D-loop *in vitro* (McRobbie, Carter et al. 2009). Cette inhibition pourrait, comme chez les eucaryotes, être levée par des paralogues de RadA catalysant l'échange de brin (McRobbie, Carter et al. 2009), (Sheng, Zhu et al. 2008). L'échange de brin pourrait également dépendre d'une interaction directe entre le SSB et RPA. Cette interaction a été observée chez *S. Solfataricus* et *S. tokodaii* (Roflsmeier and Haseltine 2010) (Sheng, Zhu et al. 2008).

2.2.5.1.2 LE FILAMENT NUCLEOPROTEIQUE : RADA

La famille RecA inclut la protéine eucaryote Rad51, le paralogue Archaea RadA et la protéine RecA bactérienne. Le rôle principal des recombinases RecA est la formation d'un filament nucléoprotéique qui permettra la recherche d'une complémentarité dans un chromosome intact. La protéine RecA Archaea est plus proche des protéines Rad51 Eucaryotes. Elles partagent le site de liaison à l'ADNdb en N-terminal et ne possèdent pas le domaine spécifique des RecA bactériennes en C-terminal. La délétion de RadA chez *H. volcanii* perturbe la réparation de l'ADN, la recombinaison ainsi que la croissance cellulaire (Delmas 2009). La délétion de RecA chez les bactéries ou de Rad51 chez la levure est hautement délétère pour la cellule. Rad51 est essentielle chez les vertébrés (Sonoda, Sasaki et al. 1998). La plupart des génomes Archaea codent pour des paralogues de RadA (Haldenby 2009). Le premier identifié fut RadB. RadB ne possède pas de domaine NTD et ne catalyse pas l'échange de brin, *in vitro*. RadB est importante pour la recombinaison et la réparation de *H. volcanii* (Guy, Haldenby et al. 2006) mais la fonction régulatrice ou catalytique de la protéine n'est pas établie. Un autre gène paralogue a été identifié chez *S. solfataricus* (Sso2452), sa structure a été résolue par diffraction aux rayons X (McRobbie, Carter et al. 2009). Sso2452 inhibe la formation des filaments recA *in vitro* et pourrait être impliquée dans la régulation négative de la recombinaison homologue (McRobbie, Carter et al. 2009).

2.2.5.1.3 MIGRATION DES BRANCHES

Chez les bactéries le moteur de la migration est la protéine RuvAB qui, en se liant aux HJs, catalyse l'échange des brins (West 1996). L'homologue fonctionnel de RuvAB n'a été identifié ni chez les eucaryotes ni chez les Archaea. L'hélicase Hel308 est peut-être la protéine qui permet la migration des branches. *In vitro*, Hel308 est capable de dérouler des structures branchées et elle interagit avec une Hjc (« Hollyday junction cleavage ») Archaea (Li, Lu et al. 2008). Chez les Archaea une deuxième protéine HJ a été identifiée, Hje (« Hollyday junction endonuclease »), sa spécificité de substrat diffère de Hjc. Hjc est un homologue fonctionnel de RuvAB (Bolt, Lloyd et al. 2001). La protéine Hjc interagit physiquement avec PCNA via un PIP-box (Komori, Sakae et al. 1999). Cette interaction stimule l'activité de Hjc *in vitro* (Bond, Kvaratskhelia et al. 2001). Les mêmes auteurs proposent que cette interaction puisse permettre le chargement de Hjc à une fourche de réplication arrêtée pour qu'elle soit réparée par RH. La figure 28 résume l'état des connaissances sur les acteurs potentiellement

impliqués dans la réparation des cassures double brin dépendante de la recombinaison. Cependant, de manière surprenante, chez *H. volcanii* le complexe MR aurait une action inhibitrice vis-à-vis de la recombinaison homologue (Delmas, Shunburne et al. 2009). D'autre part, chez *Thermococcus kodakarensis*, les gènes *mre11*, *rad50*, *herA*, *nurA* and *radA* sont essentiels, ce qui est en complète opposition avec les résultats des analyses génétiques chez les Eucaryotes et les bactéries (Fujikane, Ishino et al. 2010). La fonction du complexe MR au sein de ce mécanisme est donc à ce jour toujours mystérieuse et intrigante chez les Archaea.

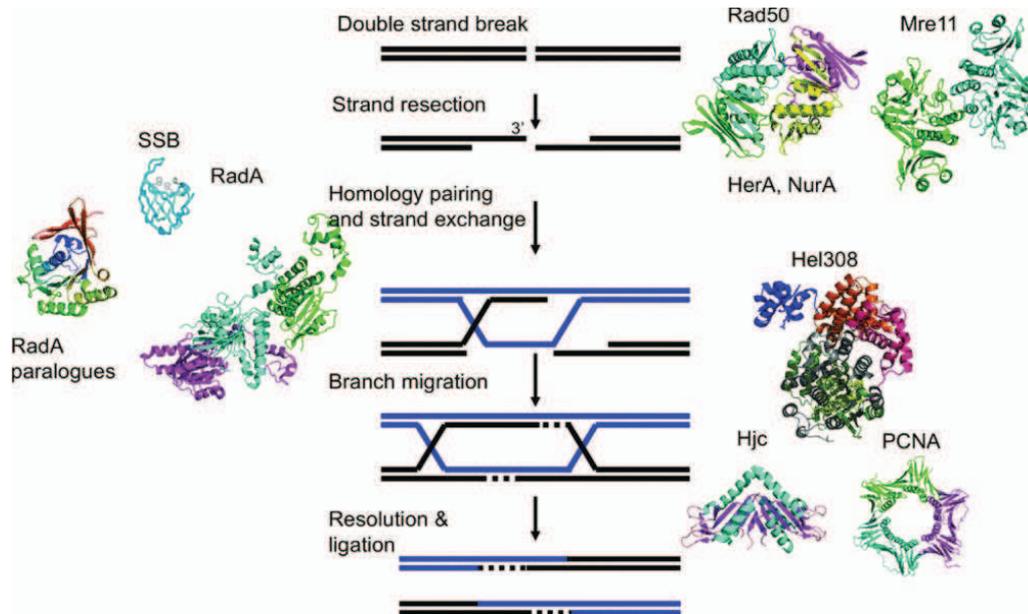


Figure 28: Protéines et voie de réparation par recombinaison homologue chez les Archaea.

La résection initiale est catalysée par le complexe Rad50–Mre11–HerA–NurA. Il génère une extrémité 3' sortante utilisée comme matrice pour une invasion de brin impliquant RadA, ses paralogues et les protéines SSB. L'invasion du brin va permettre une recherche d'homologie qui aboutira à la formation d'une jonction de Holliday. La migration de la jonction est probablement catalysée par Hel308, et Hjc. Hjc peut également former un complexe fonctionnel avec PCNA. Finalement l'ADN ligase peut lier les fragments d'ADN recombinés. (White 2011)

2.2.6 NON HOMOLOGOUS END JOINING (NHEJ)

Le NHEJ est une voie de réparation des cassures double brin par jonction des brins non homologues sans recherche d'homologies. Les extrémités des DSBs sont rapprochées par les protéines de liaison aux extrémités Ku et liées par une ADN ligase spécialisée. Lig4 chez les Eucaryotes et LigD chez les Bactéries (Shuman and Glickman 2007). Aucun homologue de protéines Ku et de ligase spécialisée n'a été identifié chez les Archaea à l'heure actuelle suggérant que ce mécanisme n'existe pas chez les Archaea ou bien qu'il est pris en charge par des acteurs, à ce jour encore inconnus.

2.3 REGULATION DE LA TOPOLOGIE DE L'ADN.

L'ADN est une structure contrainte où chaque déroulement provoque des tensions dans l'ADN. Les topoisomérases sont des protéines spécialisées dans la relaxation de l'ADN, notamment en amont des

hélicases (Koster, Crut et al. 2010). A l'inverse la Reverse Gyrase, spécifique des thermophiles, induit des super tours positifs afin de resserrer la double hélice ADN, par exemple après la réplication (Rodríguez 2002; Perugino, Valenti et al. 2009). Les protéines chromatiniennes jouent quant à elles un rôle primordial dans la compaction du génome et dans la régulation de l'expression des gènes. Plusieurs protéines chromatiniennes ont été identifiées chez les Archaea (White and Bell 2002). Les Archaea possèdent un système de compaction de l'ADN proche d'une version rudimentaire du système eucaryote. Beaucoup d'Euryarchaea possèdent des versions simplifiées des histones en charge de la compaction de l'ADN chromosomique. En absence d'histones, comme chez *S. solfataricus*, la fonction de compaction est probablement prise en charge par les protéines de la famille SulD7 (White and Bell 2002). La protéine Alba (« Acetylation Lowers Binding Affinity ») est une protéine chromatinienne identifiée chez les Archaea et quelques Eucaryotes (Forterre, Confalonieri et al. 1999; Bell, Botting et al. 2002). L'affinité de la protéine à l'ADN est régulée par acétylation / dé-acétylation. Ce mécanisme pourrait être l'un des processus clé de la modulation de l'accessibilité de l'ADN pour les machineries de réplication et de transcription Archaea.

Comment l'ensemble de ces protéines interagissent et collaborent pour conserver l'intégrité du génome est une question cruciale. Je détaillerai plus en détail dans la section "présentation de l'étude" mon apport pour tenter de répondre en partie à cette question. Toujours est-il qu'il est bien admis, chez les organismes modèles, que ces protéines s'associent au sein de complexes macromoléculaires pour assurer les différentes fonctions requises au sein des diverses voies métaboliques de la maintenance génomique. Comprendre ces mécanismes au niveau de la cellule nécessitera donc chez les Archaea de détailler les interactions physiques entre ces composants et, à terme, de mettre en œuvre des approches de biologie des systèmes pour en étudier les aspects dynamiques et fonctionnels. Dans la prochaine section, je présenterai la biologie des systèmes et son application aux systèmes protéiques.

3 BIOLOGIE DES SYSTEMES, LA CELLULE MICROMACHINE ET MICROORDINATEUR.

Depuis l'élaboration du dogme central de la biologie moléculaire, la compréhension de la fonction cellulaire et du génome a bénéficié de nombreuses découvertes radicales. La découverte des modifications post-transcriptionnelles et post-traductionnelles, de la rétro transcription, des ARN catalytiques ou non codants mais fonctionnels ou encore l'épigénétique sont parmi les découvertes qui ont considérablement modifiées la conception des processus cellulaires. La vision initiale du dogme de la biologie moléculaire où l'information est transmise linéairement d'une classe de molécules à une autre est désormais considérée comme singulièrement réductrice (Shapiro 2009). L'un des enseignements essentiels des découvertes faites depuis la découverte de l'ADN est que l'expression d'une fonction ne correspond pas à la synthèse d'une protéine à partir d'un gène mais émerge de l'exécution de processus interactifs et multimoléculaires.

3.1 BIOLOGIE DES SYSTEMES

Pour définir les processus moléculaires à l'origine de la morphogénèse, de la réponse globale à un stress ou encore l'homéostasie, il est indispensable d'appréhender la problématique à l'échelle cellulaire, unité du vivant. En 1865, Claude Bernard établit dans « Introduction à l'étude de la médecine expérimentale » que les molécules du « milieu intérieur » ne sont pas fondamentalement différentes du « milieu extérieur » mais plutôt que l'ensemble des molécules du milieu intérieur se comportent d'une manière contrôlée. Ce contrôle, il le définit en totale opposition au vitalisme en énonçant que la physique et la chimie des molécules organiques était la même que celle de la matière inorganique. Les « principes » de la biologie sont ainsi interprétables, au même titre que les lois de la physique, en termes mathématiques. Claude Bernard est reconnu par de nombreux auteurs comme le premier biologiste des systèmes (Noble 2008). Il est également reconnu pour ses travaux en physiologie et sur les principes de la démarche scientifique. John Von Neumann est un autre pionnier de la biologie des systèmes. Il définit en 1951, un automate capable de se répliquer et dont les données et le programme sont stockés dans le même espace physique (*Theory of Self Reproducing Automata*). La combinaison de systèmes logiques qu'il propose est réductible en termes d'acides nucléiques et de protéines (Figure 29) (Noireaux, Maeda et al. 2011).

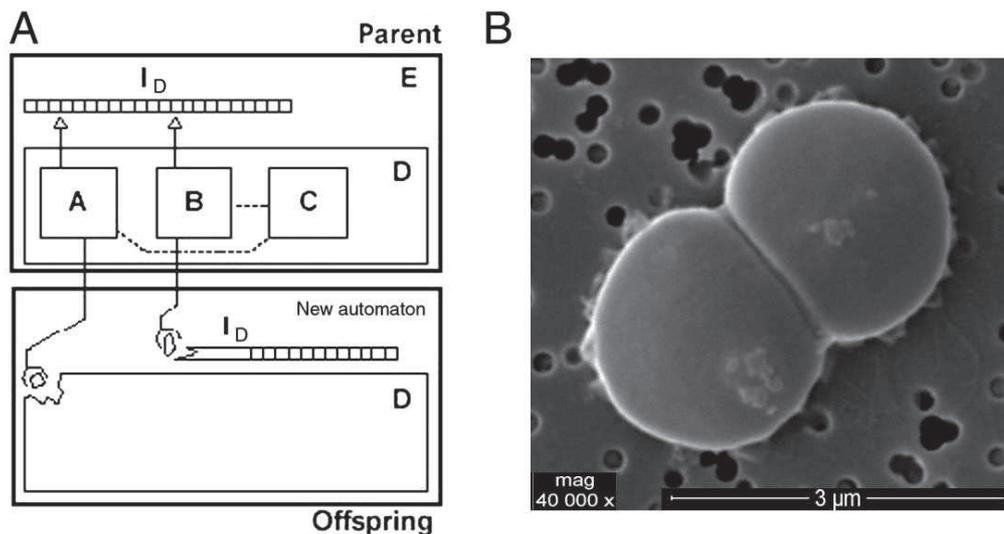


Figure 29 : Le système cellulaire

A) Automate de Von Neumann appliqué à la cellule. L'information est définie par la cassette I ("tape I"), qui constitue à la fois le code et le programme. La deuxième partie, D, de l'automate est composée de 3 parties A, B et C. A est un constructeur universel qui construit la partie D de la descendance en suivant l'information codée dans tape I quand le copieur B réplique les instructions de tape I. C est le contrôleur dirigeant A et B. Le nouvel automate suit le même schéma pour se reproduire. On peut comparer chacune des parties définies par une fonction cellulaire. Tape I est l'ADN génomique, A est la machinerie de transcription/traduction (protéines et RNA), B est la machinerie de réplication de l'ADN (protéines) et C le régulateur (ARN et protéines). B) *P. abyssi* observé en MEB Ifremer. Adapté de (Noireaux, Maeda et al. 2011)

Il est important de noter que la notion de cassette (« tape I ») (Figure 29) contenant à la fois le code et le programme est introduite par Von Neumann deux ans avant la résolution de la structure de l'ADN par Watson, Crick et Rose (Watson and Crick 1953). La découverte du code génétique et les premiers séquençages de génomes (Sanger, Nicklen et al. 1977), (Venter, Adams et al. 2001)) ont permis de mieux comprendre la nature de l'information de la « tape I ». En 2010, une équipe de chercheurs a obtenu la première cellule dont le génome a été entièrement synthétisé mettant en évidence l'aspect cassette du génome (Gibson, Glass et al. 2010). En transplantant une version minimale (où les gènes non essentiels ont été éliminés) du génome de *Mycoplasma mycoides* dans une souche de *Mycoplasma capricolum*, ils obtiennent finalement des cellules vivantes ne comprenant que le génome d'origine synthétique. Ces cellules dont le génome fait 10 kDa possèdent une croissance logarithmique et croissent plus rapidement que la souche naturelle. De nombreuses approches de biologie synthétique sont réalisées à partir de génomes naturels (approche « top-down »). Des approches nouvelles permettent aujourd'hui de reprogrammer des génomes entiers afin de modifier significativement les fonctions cellulaires (Wang, Isaacs et al. 2009). En revanche la création d'un système cellulaire synthétique « botum-up » *i.e.* la synthèse d'une cellule dont l'information est intégralement maîtrisée, est quant à elle encore plus complexe. En effet, la plupart des fonctions cellulaires ne peuvent pas être réduites en termes de présence ou d'absence de gènes mais le plus souvent en systèmes régulés et intégrés. Dans l'optique de créer une cellule *de novo* il est indispensable de comprendre l'« ordre créatif » du génome permettant l'intégration des nombreuses fonctions cellulaires. En confrontant la modélisation et l'expérimentation la biologie des systèmes tente d'appréhender les principes mathématiques fondamentaux qui régissent le fonctionnement d'une cellule, d'un organe (Bassingthwaighte 2000), d'un organisme (Buescher,

Liebermeister et al. 2012) ou même d'un écosystème (Flores, Wagner et al. 2012). La biologie des systèmes est aujourd'hui appliquée dans des domaines aussi variés que l'étude du métabolisme (Ni, Drengstig et al. 2009), la neurologie (Alivisatos, Chun et al. 2012), l'écologie microbienne (Flores, Wagner et al. 2012) ou encore le comportement des insectes sociaux (Blonder and Dornhaus 2011). Les progrès de la biologie des systèmes sont substantiellement liés à l'augmentation exponentielle de la puissance de calcul et des données disponibles permettant des modélisations et des simulations plus complexes et plus complètes.

Les approches de biologie des systèmes ont été d'un apport précieux pour étudier les protéines et leurs interactions. Je présenterai dans les paragraphes suivants le rôle fondamental des systèmes protéiques puis l'approche et l'apport de la biologie des systèmes pour l'étude des interactomes.

3.2 DES PROTEINES COMPLEXES ET DES COMPLEXES PROTEIQUES.

L'importance des protéines dans les organismes vivants revêt un double aspect structural et fonctionnel. Les protéines sont à la base de la morphogénèse et sont les acteurs essentiels des nombreuses réactions biochimiques indispensables au fonctionnement de la cellule. Une protéine constitue la traduction d'une information monodimensionnelle en une structure tridimensionnelle dans l'environnement aqueux du cytosol ou dans l'environnement multiphasique des membranes biologiques. Si l'on considère les acides nucléiques comme les molécules informationnelles, support de l'hérédité, alors les protéines sont les molécules fonctionnelles assurant la survie des organismes.

Les protéines assument une grande variété des fonctions biologiques et chaque protéine possède une ou plusieurs fonctions spécifiques. La fonction élémentaire d'une protéine est la fixation sélective de ligands ; soit une petite molécule, soit une macromolécule telle que les acides nucléiques ou d'autres protéines, *i.e.* la formation d'un complexe. La sélectivité ou spécificité d'une protéine dépend de sa composition en acides aminés et de leur organisation dans l'espace tridimensionnelle. Les protéines effectuent la catalyse chimique des réactions biologiques, le transport des molécules dans le cytosol ou à travers les membranes, la transduction des signaux chimiques ou lumineux, ou encore des actions mécaniques. Certaines protéines sont dites essentielles car leur suppression met en jeu la survie de l'organisme, ces protéines sont souvent conservées. L'une des propriétés intrinsèques des protéines est la capacité à s'assembler en complexe. Plusieurs protéines vont s'associer de manière non-covalente afin d'acquérir une fonction que les partenaires ne possèdent pas isolément. L'énergie d'activation des réactions non-covalentes est très faible, elles se produisent donc spontanément, et très rapidement, même à faible température et en l'absence de catalyseurs. Pour une protéine, former un complexe constitue la règle, la stabilité des différents complexes varie et est adaptée à la fonction. Une réaction enzyme-substrat doit pouvoir se faire et se défaire très rapidement tandis que des interactions proches de la covalence sont nécessaires aux processus longs ou aux complexes permanents. Les complexes protéiques sont des machines moléculaires composites impliquées dans l'ensemble des processus biologiques au sein d'une cellule. Ils se différencient des agrégats par leur spécificité et par une stœchiométrie définie. La spécificité de l'assemblage est garantie par l'affinité spécifique entre certains domaines sur la protéine. La régulation des complexes protéiques fait principalement intervenir des

modifications post traductionnelles ou des changements de conformations des protéines en complexe permettant, ou empêchant, l'assemblage de partenaires supplémentaires.

3.2.1 RESEAU D'INTERACTIONS PROTEINE - PROTEINE (PPIN)

L'interaction entre deux protéines est facilement modélisable en deux nœuds connectés par un lien discret. On peut représenter de la sorte l'ensemble des interactions d'un système protéique par un ensemble de nœuds interconnectés constituant un réseau. Les réseaux sont des objets mathématiques également appelés graphe (« graph »). Dans un réseau d'interaction protéine – protéine (« Protein - Protein Interaction Network », PPIN), les protéines sont des nœuds (« nodes ») et les interactions sont les liens (« edges »). Un organisme qui possède 2 000 protéines peut théoriquement former 199 000 paires de protéines uniques (difficiles à tester les une après les autres). Les technologies développées en biologie moléculaire et en protéomique au cours des dernières décennies ont rendu possible la cartographie à grande échelle des interactions physiques entre les protéines. Les premières études décrivant des PPIN ont été publiées au début des années 2000 (Uetz, Giot et al. 2000). Depuis, de nombreux autres PPIN ont été résolus chez la levure (Gavin, Bosche et al. 2002) (Ho, Gruhler et al. 2002) (Jeong, Mason et al. 2001) (Figure 30) et d'autres organismes modèles, y compris la Bactérie *Escherichia coli* (Butland, Peregrin-Alvarez et al. 2005), la drosophile *Drosophila Melanogaster* (Guruharsha, Rual et al. 2011), le ver *Caenorhabditis elegans* (Li, Armstrong et al. 2004) (Fig. 27), et l'homme (Rual, Venkatesan et al. 2005), (Stelzl, Worm et al. 2005)). Ils ont permis la cartographie à grande échelle des interactions entre les nombreuses protéines de ces organismes.

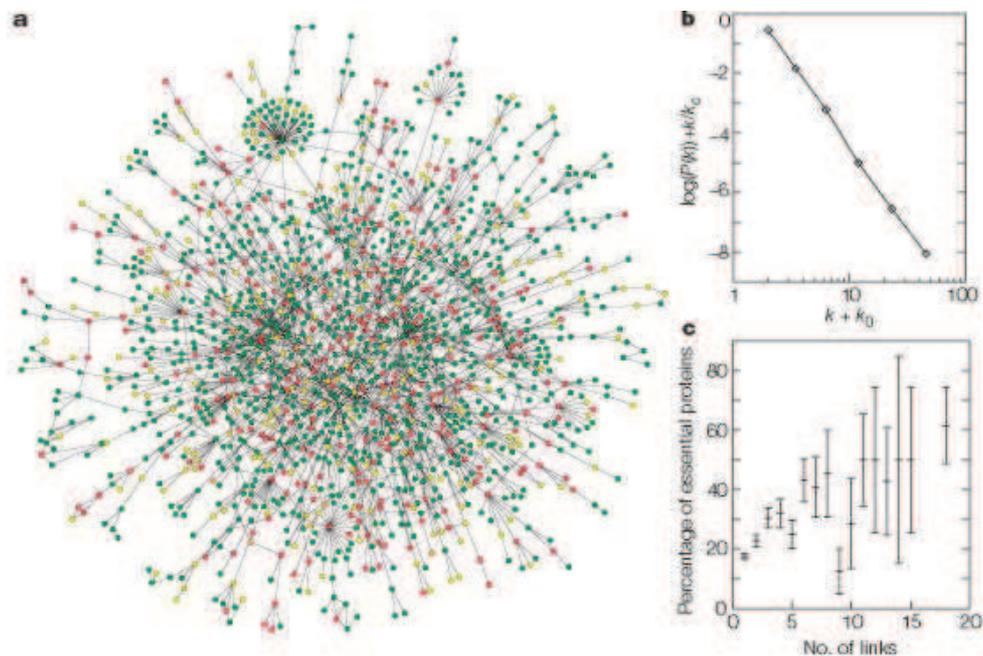


Figure 30 : Structure du PPINet et analyse de l'essentialité des protéines de *S. cerevisiae*.

a) PPIN de *S. Cerevisae*. La couleur des nœuds dépend de l'effet phénotypique de la suppression du gène codant pour la protéine représentée (rouge, létale; vert, létale; orange, croissance faible; jaune, inconnu). B) Distribution de la connectivité des nœuds du réseau (cf Section 3.2.1) C) Degré des nœuds en fonction de l'essentialité des protéines. Le nombre de liens augmente avec l'essentialité. (Jeong, Mason et al. 2001).

En identifiant les propriétés globales et locales d'un PPIN, il est possible d'obtenir des informations cruciales sur les composants et l'organisation d'un système protéique (Becker, Robisson et al. 2012) (Sarañs, Pancaldi et al. 2012). La taille et la complexité des cartes d'interactions nécessitent une post analyse informatique permettant d'identifier les propriétés essentielles du réseau. La théorie des réseaux complexes a fourni les outils mathématiques permettant de mesurer les réseaux et d'identifier leurs principes d'organisation.

3.2.2 PROPRIETES DES PPIN

Parmi les différents paramètres d'un réseau quelques-uns sont les plus significatifs. Le degré des nœuds k , leur distribution $P(k)$, la distance la plus courte entre deux nœuds l (« shortest path ») et le coefficient de clustering C (« clustering coefficient ») (Figure 31). A partir de ces mesures, il est possible d'identifier les propriétés principales d'un réseau.

3.2.2.1 ORIENTATION DU RESEAU

Tout d'abord, on considère deux types de réseaux ; orientés ou non-orientés. On dit qu'un réseau est orienté lorsque l'interaction entre deux nœuds A et B n'est pas équivalente que l'on considère A vers B ou B vers A. Les réseaux représentant une voie métabolique sont typiquement orientés puisqu'ils conduisent à la synthèse d'un produit à partir d'un substrat grâce à des réactions souvent irréversibles. *A contrario* les interactions au sein d'un PPIN sont non-orientées.

3.2.2.2 DEGRE, DISTRIBUTION DES DEGRES

Le paramètre le plus caractéristique d'un nœud dans un réseau est son degré (ou connectivité), k , qui indique le nombre de connections aux autres nœuds. Un réseau non orienté, de N nœuds et L liens, possède un degré moyen $\langle k \rangle = 2N/L$ (ou $\langle \rangle$ dénote la moyenne). La distribution des degrés, $P(k)$, donne la probabilité qu'un nœud possède exactement k liens.

3.2.2.3 PATH LENGTH

Les distances dans un réseau sont mesurées par le nombre de liens séparant deux nœuds (« path-length»). Dès lors qu'il existe plusieurs chemins entre deux nœuds, le chemin le plus court (« shortest path length », SPL) et sa moyenne (« average path-length », APL), jouent des rôles particuliers. Ils offrent un aperçu de la navigabilité globale au sein du réseau.

3.2.2.4 CLUSTERING COEFFICIENT

Dans beaucoup de réseaux, si un nœud A est connecté à B et que B est connecté à C, alors il est probable que A est également connecté à C. Ce phénomène peut être quantifié en utilisant le coefficient de clustering

(« clustering coefficient ») $C_i = 2n_{ii} / (k_i(k_i - 1))$, ou n_{ii} est le nombre de liens connectant les k_i voisins du nœud i (Watts and Strogatz 1998). Le coefficient de clustering moyen $\langle C \rangle$ caractérise la tendance globale des nœuds à former des clusters. Une mesure importante du réseau est la fonction $C(k)$ qui définit le coefficient de clustering moyen de tous les nœuds avec k liens.

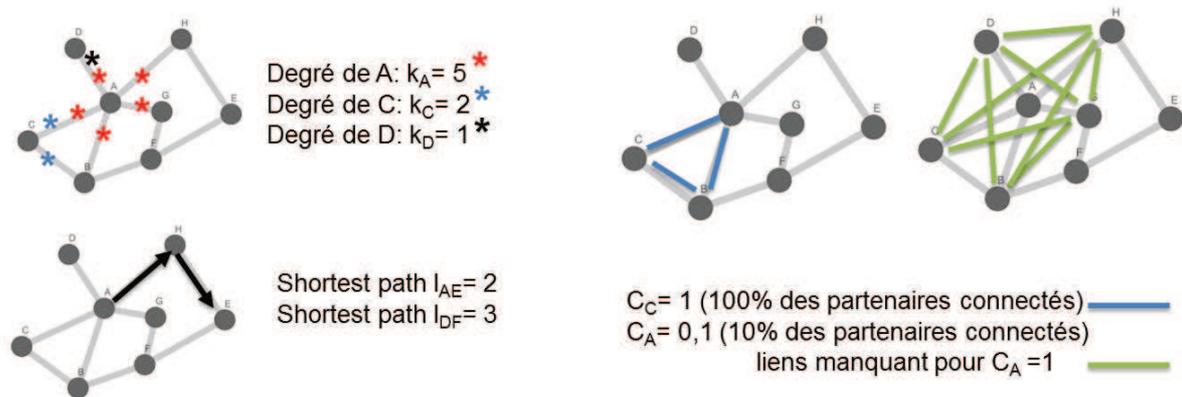


Figure 31 : Mesures élémentaires d'un graph.

Les paramètres présentés sont; le degré « k », le Shortest path « I » et le Clustering Coefficient « C ».

3.2.3 TOPOLOGIE D'UN RESEAU

A partir des paramètres que j'ai présenté, trois grand types de réseaux sont décrits : Les réseaux aléatoires (« random network »), les réseaux sans échelle (« scale-free ») et les réseaux hiérarchiques (« hierachical network »).

3.2.3.1.1 RESEAUX ALEATOIRES

Les réseaux aléatoires (Figure 32 A) ont été décrits selon le modèle Erdos-Rényi (ER) (Erdos 1960). Le degré des nœuds suit une distribution de Poisson indiquant que chacun des nœuds possèdent approximativement le même nombre de liens. Il existe un degré caractéristique du réseau et il n'y a pas dans les réseaux de nœuds significativement plus connectés. Dans un réseau aléatoire, le coefficient de clustering d'un nœud est indépendant de son degré, $C(k)$ apparaît donc comme une ligne horizontale si elle est tracée en fonction de k (figure 32 Ab). Egalement l'APL est proportionnel au logarithme de la taille du réseau, caractéristique des réseaux en petit monde (« small world »).

3.2.3.1.2 RESEAUX SCALE-FREE

Dans la plupart des réseaux réels décrits, les nœuds ne possèdent pas de connectivité typique. La distribution des degrés ne suit pas une loi de Poisson mais une loi de puissance (Figure 32 B). A l'échelle du réseau, cela signifie qu'une minorité de nœuds de hauts degrés (« hubs ») regroupent des ensembles de nœuds moins connectés. Ainsi la probabilité qu'un nœud soit très connecté est beaucoup plus élevée que dans un réseau aléatoire. Lorsque la distribution des degrés approxime une loi de puissance, le réseau est dit sans

échelle (« scale-free »). La notion de réseau scale-free a été introduite en 1999 par Albert et Barabasi (Barabasi and Albert 1999) après la cartographie et l'analyse de la topologie du réseau Internet. Les réseaux scale-free sont des réseaux « ultra small », ou l'APL est significativement plus court que les réseaux aléatoires en petit monde (Chung and Lu 2002) (Cohen and Havlin 2003).

3.2.3.1.3 RESEAUX HIERARCHIQUES

Les réseaux hiérarchiques sont des réseaux possédant une topologie scale-free et un coefficient de clusterisation élevé (Barabasi and Oltvai 2004) (Figure 32 C). Dans les réseaux hiérarchiques le tracé de $C(k)$ suit une droite dont la pente est de -1 sur une échelle logarithmique. Les zones hautement connectées constituent des modules connectés par un nombre limité de hub.

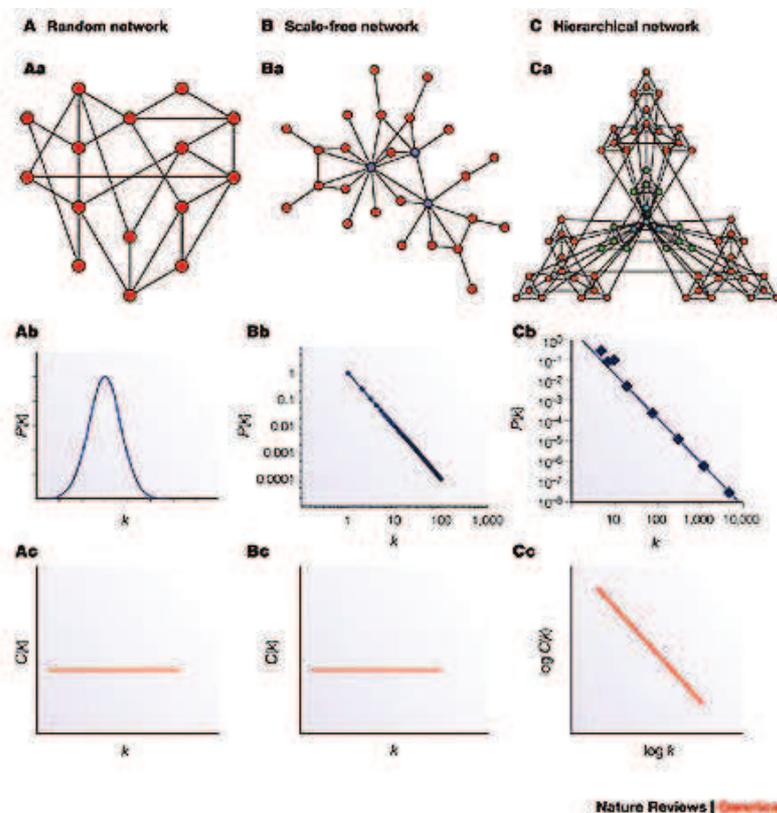


Figure 32 : Modèles de réseaux.

A) Réseau aléatoire ; la distribution des degrés suit une distribution de Poisson. B) Réseau Scale free la distribution des degrés suit une loi de puissance. Le clustering est indépendant du degré. C) Réseau hiérarchique ; la distribution suit une loi de puissance tout comme la distribution des coefficients de clustering (Barabasi and Oltvai 2004)

3.2.4 DE LA TOPOLOGIE A LA FONCTION

La topologie d'un PPIN reflète l'organisation du système protéique, elle possède une grande implication biologique (Becker, Robisson et al. 2012; Saraç, Pancaldi et al. 2012). La position d'une protéine dans un PPIN est liée à son essentialité (Batada, Reguly et al. 2006) (Han, Bertin et al. 2004) (Jeong, Mason et al. 2001). Il est suggéré que les PPIN sont scale-free et que la centralité des protéines « hub » reflète la topologie naturelle des PPIN (Maslov and Sneppen 2002). Les structures, généralement sans échelle (« scale-free ») régulièrement

observées dans les PPIN impliqueraient une résistance aux attaques aléatoires (« random component failure ») (Albert, Jeong et al. 2000) mais une forte sensibilité aux attaques ciblées, notamment sur les hubs (He and Zhang 2006).

Cependant, il est important de signaler que la topologie des PPIN est fortement influencée par les méthodes d'échantillonnage (Han, Dupuy et al. 2005). Cela suggère que l'échelle scale-free souvent observée n'est pas forcément représentative de l'échelle réelle des réseaux d'interactions biologiques. L'un des aspects les plus essentiels révélés par la biologie des systèmes est que les fonctions cellulaires sont réalisées de manière modulaire (Hartwell, Hopfield et al. 1999). La modularité, qui décrit l'achèvement de fonctions distinctes par un groupe de molécules physiquement ou fonctionnellement liées, est une propriété courante dans les réseaux biologiques (Alberts 1998). La protéine PCNA que nous avons décrite dans le chapitre précédent est un parfait exemple de protéine modulaire dans la mesure où elle est impliquée dans des assemblages protéiques ayant des fonctions distinctes. Les modules sont traduits dans un réseau sous la forme de clusters où le niveau d'interactions entre un groupe de nœud est élevé. Chaque module peut être représenté par un jeu de triangles. Une zone de forte densité en motifs triangulaires sera traduite par un coefficient de clustering élevé. Des signatures de modularité hiérarchique ont été identifiées dans de nombreux réseaux biologiques (Girvan and Newman 2002; Holme, Huss et al. 2003).

Afin de déterminer l'échelle réelle d'un PPIN il est nécessaire de le décrire dans son intégralité et d'intégrer l'aspect dynamique des interactions. La dynamique des réseaux est un sujet de recherche et de discussion très actif (Batada, Reguly et al. 2007) (Han, Bertin et al. 2004) (Batada, Reguly et al. 2007) (Steinacher and Soyer 2012). L'analyse de la structure des réseaux et de leur construction a également été utilisée dans le contexte d'études des mécanismes d'évolution à l'origine de la complexité et de la diversité des systèmes protéiques (Berg, Lassig et al. 2004) (Salathé, May et al. 2005) (Bersini, Lenaerts et al. 2006) (Batada, Hurst et al. 2006) (Becker, Robisson et al. 2012).

3.2.5 TECHNIQUES

Deux méthodes fournissent l'essentiel des données d'interactions protéine - protéine à grande échelle, le double hybride (Yeast two Hybrid, Y2H) et la purification d'affinité couplée à la spectrométrie de masse (AP-MS). Le Y2H repose sur la structure modulaire des facteurs de transcription dans les organismes Eucaryotes qui comportent un domaine de fixation à l'ADN (BD pour Binding Domain) et un domaine d'activation de la transcription (AD, pour Activating Domain). Pour tester l'interaction entre deux protéines X et Y on exprime ces protéines par construction génétique sous forme de protéines de fusion chez la levure. La protéine X est fusionnée avec le domaine AD de la protéine gal4 (régulateur du gène de la B-galactosidase), la protéine Y est fusionnée avec le domaine de liaison BD. L'interaction entre la protéine X et Y permettra l'assemblage d'un facteur de transcription fonctionnel et déclenchera l'expression du gène rapporteur. Le Y2H s'intéresse donc à des interactions binaires et permet de détecter des interactions labiles. La sensibilité de cette technique l'expose aussi à un taux de faux positif élevé.

L'AP-MS a pour objectif l'isolation et l'identification de complexes protéiques. Cette technique utilise la propriété d'affinité à un support solide de peptides ou de petites protéines appelées étiquettes (« tag »). Les étiquettes HIS (Histidine), GST (Glutathion S Transferase) ou MBP (Maltose Binding Protein) en sont quelques exemples. La protéine d'intérêt, fusionnée avec une de ces étiquettes, est incubée dans un extrait cellulaire. La composition des complexes protéiques isolés est ensuite déterminée par spectrométrie de masse. Les conditions expérimentales de l'AP-MS nécessitent le plus souvent des étapes de rinçages qui favorisent la détection de complexes protéiques stables à la défaveur d'interactions transitoires. Les interactions sont détectées au sein de complexes, cela implique qu'une interaction détectée uniquement en AP-MS ne peut pas être forcément considérée comme directe. On peut également citer le développement d'une technique dérivée de la purification d'affinité classique. Dans cette méthode appelée purification d'affinité en tandem (TAP-MS) on remplace la séquence codante *in vivo* par une version tagguée de la protéine. On purifie ensuite les complexes formés *in vivo* autour de la protéine d'intérêt sur colonne d'affinité. Cette technique a pour avantage de conserver les conditions d'expression du gène dans son contexte *in vivo* mais induit souvent à travailler avec de faibles quantités de protéines rendant l'identification du signal spécifique au sein des résultats de spectrométrie de masse plus difficile.

3.3 PPIN DE LA MAINTENANCE GENOMIQUE

La maintenance génomique est un processus multimoléculaire (Budd, Tong et al. 2005; Hamdan and van Oijen 2010). Le système protéique en charge de la maintenance du génome a été étudié dans les trois domaines du vivant. Chez les Eucaryotes la définition du PPIN en charge de la maintenance génomique est un sujet actif de recherche dès lors qu'il permet la compréhension de certains mécanismes impliqués dans les cancers et ouvre la voie à la découverte de nouvelles cibles pharmacologiques potentielles (Reinhardt and Schumacher). Le réseau d'interaction des protéines de la maintenance génomique a été sujet à de nombreuses études chez les Eucaryotes et les Bactéries. Un exemple de réseau chez chacun de ces organismes est présenté en figures 33 et 34.

Y2H of DNA Damage Response Genes

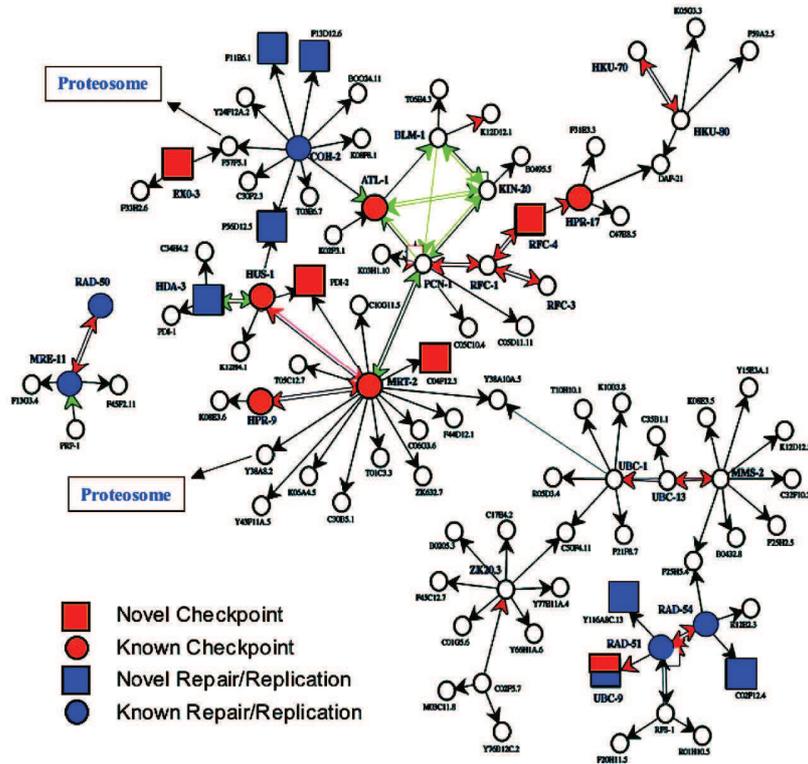


Figure 33 : PPINde la réponse au dommages de l'ADN chez les Eucaryotes.

(Boulton, Gartner et al. 2002)

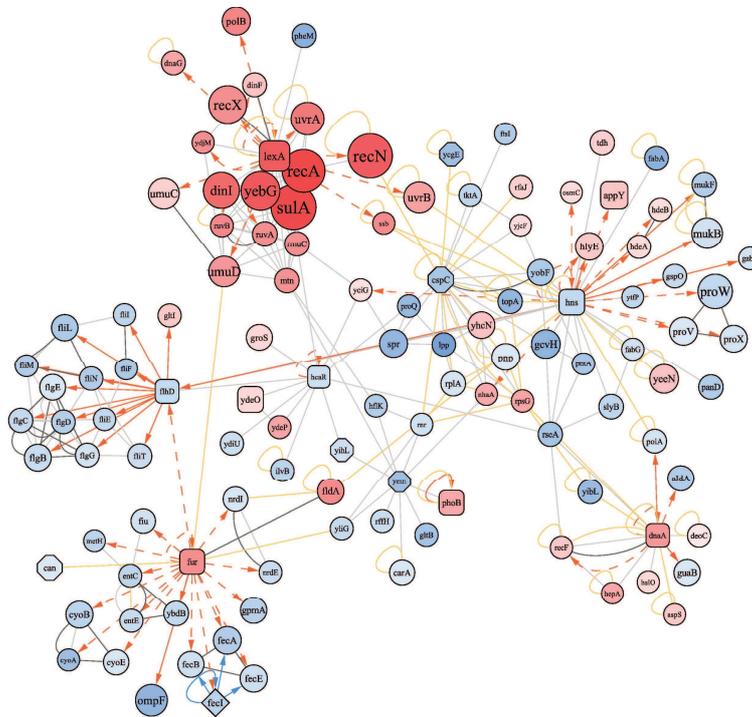


Figure 34 : PPIN des protéines de la réponse aux dommages de l'ADN chez les Bactéries.

Ce sous réseau des protéines LexA, Hns, Fur, CspC, FlhD, HcaR, DnaA et YeiE constitue le PPIN des protéines communes au système de réponse au dommage de l'ADN chez *E. coli*. (Hong, Ahn et al. 2009)

Li et ses collaborateurs en 2010 ont notamment identifié une nouvelle protéine (TK1252p) associée aux protéines GINS. Des études subséquentes ont identifié cette protéine comme une nucléase dont l'activité est régulée par le complexe GINS (Li, Pan et al. 2011). Cette protéine nommée GAN (« Gins Associated Nucléase) est probablement associée au complexe GINS – MCM lors de l'initiation de la réplication. Dans ces deux études du réseau d'interactions de la réplication Archaea, aucune analyse topologique du réseau d'interactions n'a été proposée. La taille limitée des jeux de données que les études ont fournis est probablement la cause principale.



Les Archaea constituent, au sein du domaine procaryote, une troisième branche du vivant indépendante de celle des Bactéries et des Eucaryotes. L'analyse des génomes Archaea a révélé une mosaïque génétique où les systèmes informationnels sont principalement conservés avec les Eucaryotes. L'un des systèmes le plus conservé est le réplisome, qui constitue en fait une version simplifiée de la machinerie de réplication Eucaryote. Cependant il est probable que de nouveaux acteurs du réplisome et notamment de la régulation de la réplication, restent encore à découvrir

Les Archaea Hyperthermophiles sont des extrémophiles dont la température optimale de croissance atteint souvent 100°C. Dans ces conditions, le niveau de dégradation spontanée de l'ADN est extrêmement élevé. Or, en dépit de la température élevée, les HA présentent des taux de mutations comparables à celui des organismes mésophiles. Cette propriété suggère l'existence chez les HA de mécanismes de réparation extrêmement efficaces. L'efficacité du système de réparation est illustrée par l'exceptionnelle capacité des HA à reconstruire un génome fragmenté par les radiations ionisantes. Les acteurs et les processus moléculaires autorisant cette réparation hors du commun ne sont que très partiellement décrits. De plus, les homologues de nombreux gènes essentiels à la réparation chez les Bactéries et les Eucaryotes sont absents des génomes Archaea. L'absence apparente de certains facteurs essentiels à la réparation suggère l'existence de mécanismes alternatifs de réparation intégrant des protéines spécifiques ou dont la divergence est telle qu'il est impossible de les identifier sur la base des homologies de séquence. En effet de l'état des connaissances sur la réparation de l'ADN chez les Archaea émane plusieurs questions qui restent en suspens :

Existe-t-il un NER fonctionnel ? Si oui, quelles sont les protéines impliquées dans la reconnaissance du dommage ?

Existe-t-il un mécanisme alternatif au MMR ?

Quel est le rôle du complexe Mre11/Rad50 ?

Quelles sont les protéines en charge de la régulation et du couplage des mécanismes de maintenance génomique ?

Pour tenter de répondre à ces questions, des expériences de purifications d'affinité ont été initiées par J. Briffotiaux au cours de sa thèse intitulée « Maintenance génomique chez l'Archaea hyperthermophile *Pyrococcus abyssi* : découverte de nouvelles interactions physiques et caractérisation fonctionnelle ». Au cours de ses travaux, il avait notamment réalisé des purifications d'affinités couplées à une analyse en spectrométrie de masse permettant d'identifier plusieurs protéines interagissant avec PCNA ou avec le complexe Mre11 Rad50 (Figure 37, les interactions en jaune ont été validées *in vitro*). Le projet de thèse que j'ai mené s'inscrit dans la continuité de ce travail avec l'objectif d'étendre ce réseau préliminaire impliqué dans la maintenance génomique de *P. abyssi*.

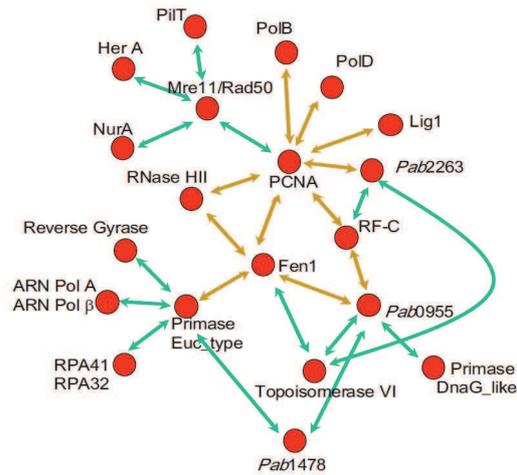


Figure 37 : PPIN de la maintenance génomique de *P. abyssi* proposé par J. Briffotiaux en 2008.

Pour ce faire, j'ai bénéficié d'une approche de production haut-débit de protéines recombinantes réalisée en collaboration avec l'équipe de Mirjam Czjzek (CNRS, UMR 7139) qui nous a permis d'obtenir les versions recombinantes de nombreuses protéines impliquées dans les mécanismes de réparation, réplication et recombinaison de l'ADN chez *P. abyssi* (Groissillier et al. 2010). Au cours de ma thèse j'ai validé la purification de près de 40 protéines recombinantes dont 23 ont finalement été utilisées comme « appâts » dans un protocole de purification d'affinité couplé à une analyse en spectrométrie de masse. L'identification des complexes protéiques impliqués dans la maintenance génomique a permis de dessiner le PPIN de la maintenance génomique de *P. abyssi* et d'identifier de nouvelles interactions et de nouvelles protéines impliquées dans la réparation, la réplication et la recombinaison chez les Archaea. Une attention toute particulière est portée aux protéines et systèmes, conservés avec les Eucaryotes

L'identification et la caractérisation des principaux acteurs de la maintenance génomique chez ce type d'organisme ouvrent des perspectives pour la découverte de nouvelles protéines ou de nouveaux mécanismes de la réparation de l'ADN. La découverte de nouvelles enzymes de l'ADN revêt également un aspect appliqué dès lors que ces enzymes, thermostables, sont des outils précieux en biotechnologie (ADN polymérases et facteurs associées, nouvelles nucléases, ligases etc.). À terme, mes travaux visent à améliorer les connaissances sur les bases moléculaires de la maintenance génomique chez les Archaea et les Eucaryotes.

5.1 PRODUCTION ET PURIFICATION DES PROTEINES APPATS

La mise au point du protocole de couplage des purifications par affinité de complexes avec l'identification des partenaires par spectrométrie de masse, ainsi que l'analyse des résultats MS et leur intégration au sein d'un réseau, a constitué une part importante de mon travail de thèse. Je pense que l'exposé de ces résultats et leurs discussions pourra intéresser ceux qui à l'avenir voudront mettre en place ce type d'approche méthodologique, c'est pourquoi j'ai décidé de la détailler dans ce chapitre. Par soucis de clarté, je reprends certains éléments du chapitre « Matériels et Méthodes ».

Nous avons comme objectif de purifier à partir d'un extrait cellulaire total, dont la préparation sera discutée ultérieurement, les protéines partenaires d'une protéine utilisée comme appât. Pour cela, il est nécessaire initialement de produire chaque protéine sous une version recombinante.

Une version modifiée du gène codant pour chaque protéine d'intérêt est insérée via un plasmide dans différentes souches d'*E. coli*. Les protéines Archaea qui seront produites dans les différents clones bactériens seront fusionnées avec une étiquette moléculaire (« Tag »). Cette étiquette est un polypeptide inséré au sein de la séquence à l'extrémité N-terminale, dans notre cas. Les étiquettes que nous utilisons sont constituées soit de 6 acides aminés Histidines, soit d'une Glutathion-S-transférase. Ces HIS-Tag et GST-Tag ont la propriété d'être chacun affins à une matrice solide spécifique permettant leurs purifications. Les clones dont j'ai disposé proviennent de la collection du laboratoire ou du protocole MARINE-EXPRESS. Les résultats de ce projet de production de protéines recombinantes haut débit ont été publiés dans la revue *Microbial Cell Factories* (Groisillier, Herve et al. 2010) (Annexe 1).

Au cours de ma thèse plus de 50 souches d'*E. coli* ont été utilisées comme « usine moléculaire » pour produire la version recombinante de, finalement, 28 protéines, complexes ou domaines protéiques solubles. Le protocole expérimental (Figure 38) a consisté à purifier, sur billes magnétiques couvertes d'une matrice fixant les ions de Cobalt, et à utiliser directement ces complexes billes-appâts dans une expérience de Pull-down au sein de l'extrait cellulaire de *P. abyssi*.

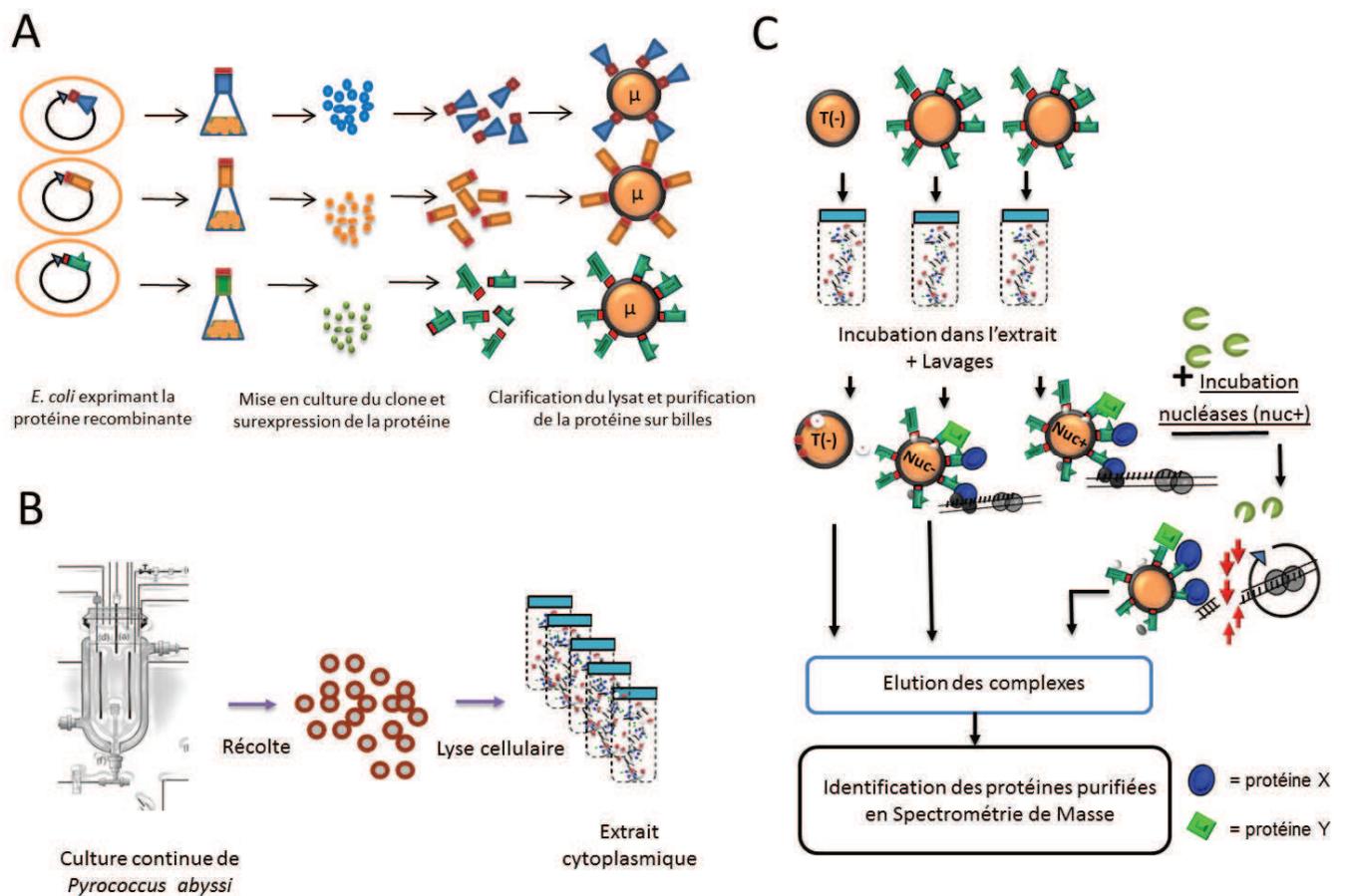


Figure 38 : Schéma général du protocole expérimental.

A) Production des protéines recombinantes et purification. B) Production de la biomasse de *P. abyssi* et extraction du contenu cellulaire. C) Protocole d'AP-MS des protéines de la maintenance génomique. La recherche de partenaires protéiques est effectuée par deux expériences AP-MS, avec ou sans étape d'incubation des complexes purifiés avec des nucléases. Des expériences témoins, où les billes nues sont incubées avec l'extrait cellulaire, sont également réalisées. Les protéines purifiées seront identifiées ultérieurement par spectrométrie de masse en tandem.

La surproduction de la protéine recombinante est déclenchée dans une culture bactérienne puis les cellules sont récoltées et lysées. J'ai ensuite incubé, dans le lysat bactérien clarifié, une quantité de billes permettant de fixer théoriquement 30 µg d'une protéine recombinante de 35 kDa (selon les recommandations du fournisseur, Invitrogen). Après une série de lavages, la qualité de la purification et la quantité de protéines purifiées sur billes sont vérifiées sur gel SDS-PAGE (Figure 39). Dans le cas où la quantité de protéines purifiées apparaît trop faible, on double, triple, etc., la quantité de lysat bactérien utilisé, ceci jusqu'à obtenir la quantité satisfaisante de protéines appâts sur billes. En gardant une même quantité de billes, j'ai pu fixer et purifier entre 10 et 30 µg de protéines sur les billes magnétiques, en fonction des appâts considérés. Certaines protéines se sont montrées cependant réfractaires, soit à l'étape de fixation et de purification sur billes, soit à l'étape initiale de surexpression. L'aspect « haut débit » et exploratoire du projet nous a incité à abandonner temporairement ces appâts dont la purification, difficile, aurait nécessité un temps d'investissement important avec une perspective de résultats incertaine. Vingt-sept protéines, complexes ou domaines protéiques ont été finalement purifiés de manière satisfaisante. Je n'ai pas cherché à effectuer une purification à l'homogénéité

de tous les appâts utilisés. D'une part, parce que l'entreprise se serait avérée trop gourmande en temps et que d'autre part, les éventuels contaminants bactériens ne gêneront pas l'analyse en spectrométrie de masse puisque l'identification des protéines partenaires se fera en interrogeant uniquement la banque de protéines de *P. abyssi*. Cependant je suis conscient que l'on pourra opposer à cette option la critique selon laquelle des contaminants de *E. coli* présents, même en faible quantité, sur les billes magnétiques pourraient également interagir de façon hétérologue avec des protéines de l'extrait de *P. abyssi* et ainsi induire la présence de faux positifs dans les résultats. Cependant, dans chaque purification effectuée, la protéine de *P. abyssi* constituait la protéine majoritaire à plus de 80%.

Avec ces 28 appâts, nous couvrons un spectre fonctionnel étendu sur les 3 fonctions: réplication, recombinaison et réparation. Nous possédons également plusieurs protéines dont la fonction est indéterminée. On pourra cependant regretter l'absence de certaines protéines pour lesquelles nous avons un fort intérêt telles que les hélicases XPF et XPB ou encore les histones A et B.

Les clones sur-exprimant les protéines recombinantes des gènes *pab1530*, *pab0164*, *pab2270*, *pab0067*, *pab0560*, *pab1695*, *pab1466*, *pab0956*, *pab0034*, *pab0128*, *pab0801*, *pab7094*, *pab0418*, *pab2423*, *pab0063* proviennent du programme MARINE-EXPRESS (Grosillier, Herve et al. 2010). Les protéines codées par les gènes *pab2266*, *pab2404*, (Palud, Villani et al. 2008), *pab0352*, *pab1465*, *pab2002*, *pab2373*, (Meslet-Cladière, Norais et al. 2007), *pab2263*, (Ren, Kuhn et al. 2007), *pab2236*, *pab2235* (Le Breton, Henneke et al. 2007), *pab2163*, *pab2164*, *pab2165*, *pab0811*, *pab0812* et *pab1128* ont été clonées au laboratoire. La liste des protéines appâts et leurs propriétés sont synthétisées dans le tableau 6. Les résultats des protéines Alba, CopG, SBDS et CCA adding enzyme n'ont pas été considérés dans le réseau. En effet CopG, SBDS et CCA-adding enzyme ne font pas partie de la maintenance génomique tandis qu'une contamination lors des purifications d'affinité avec la protéine Alba est suspectée. La liste des protéines utilisées en AP-MS et leurs propriétés sont représenté dans le Tableau 6. Des exemples de purification des protéines recombinantes sont présentés en figure 39.

Tableau 6 : Protéines de fusion utilisées en appâts dans les expériences d'AP-MS.

Nom	Uniprot ID	ORF	Fonction	Propriétés de l'app Tag	
Alka	Q9UZ73	pab1530	Réparation des bases alkylées	FL	HIS
DNA ligase	POCL75	pab2002	Ligation	FL	HIS
DNA2 homolog	Q9V2G5	pab0067	Helicase/Nu clease	C-ter 1070K-1308A	HIS
Fen1	Q9V0P9	pab1877	Endonucléase de l'OFP	FL	HIS
Gins 23	Q9UYR8	pab0956	Coordination de l'architecture de la fourche de replication	FL	HIS
Gins 51	Q9UYX9	pab1466	Coordination de l'architecture de la fourche de replication	FL	HIS
MCM	Q9UYR7	pab2373	Hélicase replicative	FL	HIS
Mre11, Rad50	Q9UZY0	pab0811	Réparation des cassures double brin	Complex	HIS
	Q9UZY8	pab0812	Réparation des cassures double brin	Complex	HIS
Protein deacetylase	Q9UZE7	pab0801	Inconnue	FL	HIS
OGG1	Q9UZY0	pab1695	Réparation (BER)	FL	GST
NucS	Q9V2E8	pab2263	Endonucléase (NER?)	FL	HIS
Primase ss P41	Q9V292	pab2236	Petite sous unité de la Primase	FL	HIS
Primase ls P46	Q9V291	pab2235	Grande sous unité de la Primase	FL	HIS
PCNA	Q9UYX8	pab1465	Sliding Clamp	FL	HIS
PHP domain protein	Q9V0F9	pab0560	Inconnue	FL	HIS
Polymerase B	POCL76	pab1128	ADN polymerase B	FL	HIS
Polymerase D	Q9V2F3	pab2266	ADN polymerase D petite sous unité	Complex	HIS
	Q9V2F4	pab2404	DNA polymerase D grande sous unité	Complex	HIS
Rad25-like	Q9V278	pab0128	Inconnue	FL	HIS
RadA	Q9V233	pab0164	Recombinaison	FL	HIS
RadB	Q9V2F6	pab2270	Recombinaison	FL	HIS
Rec-J like ss exonuclease	Q9V2K8	pab0034	Nucléase associé au GINS (GAN)	FL	HIS
Reverse gyrase dom1	Q9UZ86	pab2423	Super-enroulement de l'ADN	N-ter	HIS
Ribonuclease HII	Q9V1A9	pab0352	Nucléase spécifiques des hybrides ADN/ARN	FL	HIS
RPA (14, 32, 41)	Q9V1Z0	pab2163	SSB homologue à RPA14	Complex	HIS
	Q9V1Z1	pab2164	SSB homologue à RPA32	Complex	HIS
	Q9V1Y9	pab2165	SSB homologue à RPA41	Complex	HIS
Alba	Q9V1N3	pab7094	Histone-like	FL	HIS
CCA adding enzyme	Q9V2H1	pab0063	Inconnue	FL	HIS
SBDS protein	Q9V121	pab0418	Assemblage du ribosome	FL	GST

FL: protéine entière, Complex: complexe protéique. N-ter/C-ter : domaine protéique.

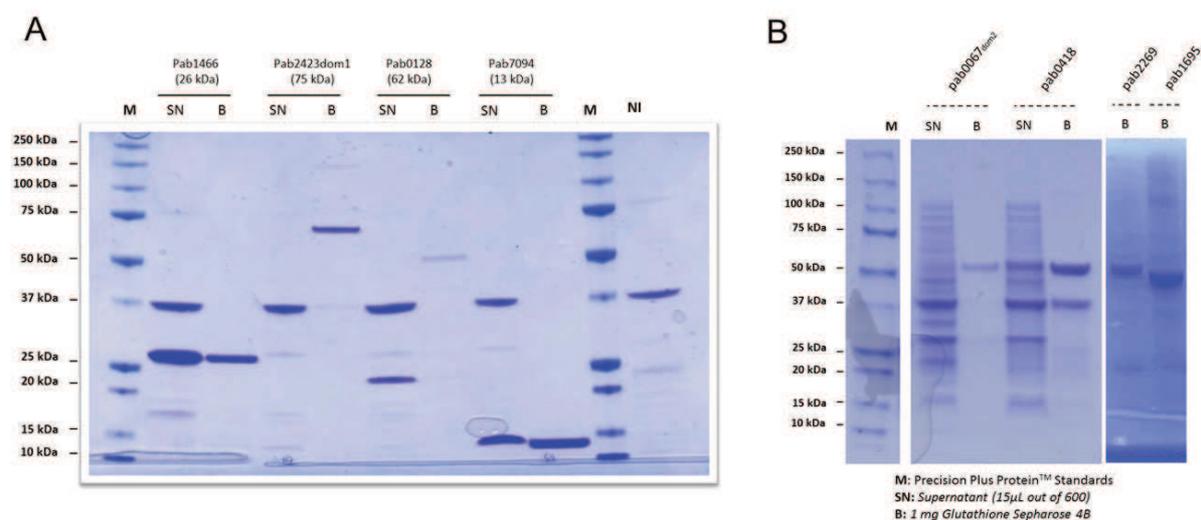


Figure 39 : Exemples de purifications des protéines appâts.

Etapes finales de la purification des protéines His-tag (A) et GST-tag (B). M : Marqueur de taille. SN : Surnageant après chocs thermiques. B : protéines purifiées sur billes. NI surnageant *E. coli* non induit après choc thermique.

5.2 MISE AU POINT DE LA PURIFICATION D’AFFINITE

Les conditions expérimentales permettant la formation d’interactions spécifiques sont déterminées en utilisant comme témoin l’interaction, bien décrite, entre PCNA et Fen1. L’interaction se réalise grâce à l’affinité de la boucle IDCL (interdomain connecting loop) du PCNA pour le PIP motif de Fen1. L’objectif de cette expérience est de déterminer les conditions permettant la formation de cette interaction *in vitro*. Pour cette expérience j’ai utilisé deux protéines recombinantes purifiées ; la protéine PCNA fusionnée à un tag histidine et l’endonucléase Fen1 qui en est démunie. 5 µg de la protéine PCNA purifiée sont fixés sur 1 mg de billes et incubés avec une solution contenant 10 µg de la protéine Fen1 purifiée. L’incubation est réalisée en milieu PBS pendant 1 h à 4°C. Les billes sont ensuite lavées par 3 x 1 ml de PBS, puis les protéines fixées sont éluées et déposées sur gel SDS-PAGE. J’ai pu ainsi montrer que la fixation de la protéine étiquetée est maintenue tout au long de l’expérience, que la protéine Fen1 non étiquetée ne se fixe pas sur les billes et, enfin, que la protéine Fen1 est bien co-purifiée par les billes chargées avec le PCNA (figure 40). Cela indique que l’interaction entre les deux protéines recombinantes se forme dans les conditions expérimentales utilisées et que le complexe formé est suffisamment stable pour résister aux séries de lavages. Le tampon PBS semble convenir à la formation de complexes spécifiques. Plusieurs travaux sur la formation de complexes protéiques ont été réalisés dans des tampons salins comparables. Le pouvoir tamponnant à pH neutre est également important dans la mesure où un saut de pH vers des valeurs trop acides provoquerait une déstabilisation de l’interaction entre l’étiquette de la protéine appât et la matrice.

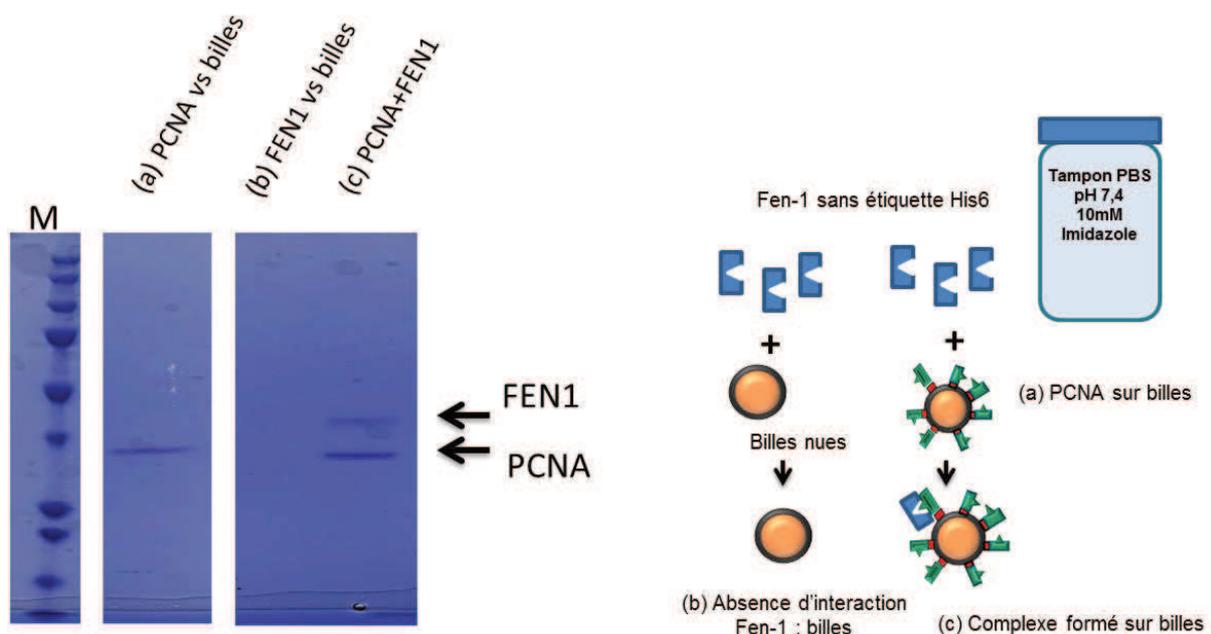


Figure 40 : Mise au point des conditions de tampon.

A) gauche : Interaction entre PCNA et Fen1 recombinants. A) droite : Schematisation de l’expérience.

5.3 PREPARATION DE L'EXTRAIT CELLULAIRE

La biomasse de *P. abyssi* produite en bioréacteur est utilisée pour préparer un extrait cellulaire duquel sera purifié les complexes protéiques via les protéines appâts fixées sur billes, présentées dans le paragraphe précédent. Afin de réaliser les expériences de purification d'affinité il est nécessaire de produire un extrait cellulaire dont les concentrations en sels et en protéines seront adaptées à la formation de complexes protéiques (Figure 38B).

Afin de préserver au maximum l'intégrité des protéines, l'ensemble des expériences est réalisé sur glace. Les cellules sont décongelées dans le milieu tampon salin standard PBS (Phosphate Buffered Saline) dont je me suis assuré de la compatibilité avec la formation d'interactions protéine-protéine spécifiques (voir chapitre précédent). J'effectue ensuite une lyse cellulaire par sonication directement dans le tampon PBS préalablement additionné d'inhibiteurs de protéases. Après cette étape qui permet la libération du contenu cytoplasmique des cellules de *P. abyssi* dans le tampon, les débris cellulaires sont éliminés par centrifugation. Le surnageant isolé est additionné de 10mM d'imidazole et le titre ajusté à 8 mg.ml⁻¹ de protéines totales. Le pH est finalement corrigé à 7.4 (4°C). La concentration en protéines que nous utilisons peut paraître élevée mais elle reste inférieure aux concentrations intracellulaires dont les valeurs dépassent largement les dizaines de milligrammes par millilitre et probablement au-delà localement. Une concentration protéique élevée favorise le repliement correct des protéines et la formation de complexes, ceci tant qu'elle ne favorise pas la formation d'agrégats, quant à eux, destructurants, désordonnés et insolubles (Minton 2000).

La sonication, utilisée pour lyser les cellules de *P. abyssi*, va déstabiliser et fragmenter les membranes libérant le contenu cytoplasmique. L'énergie dégagée par la source d'ultrason va agir sur l'ensemble des éléments cellulaires, y compris les acides nucléiques. La sonication aura deux effets principaux sur l'ADN, substrat des protéines sujettes à l'étude. Le premier est un dommage mécanique direct, le second un effet chimique indirect. Les ultrasons provoquent deux phénomènes physiques principaux en milieu aqueux ; des effets de cavitation et une accélération des molécules du solvant. La cavitation provoque une augmentation extrême et localisée de la température (« bulk heating »). L'énergie dégagée est suffisante pour, non seulement, briser la liaison phosphodiester mais également libérer des radicaux libres du solvant qui pourront attaquer les constituants de l'ADN. Un autre effet, indépendant de la cavitation, dégrade également l'ADN. L'agitation des molécules « compresse » l'ADN contre son solvant provoquant une dégradation comparable à celle provoquée par un effet thermique (Fuciarelli, Sisk et al. 1995).

Ces effets se traduisent par une fragmentation hétérogène et une dégradation chimique de l'ADN au sein de l'extrait pendant l'extraction des composants cellulaires solubles. En gel d'agarose non dénaturant, l'ADN de l'extrait cellulaire présente effectivement un profil de dégradation correspondant à une fragmentation de l'ADN (Figure 41). Dans l'extrait cellulaire, l'ADN de *P. abyssi* est donc fortement endommagé et les protéines probablement intègres, puisqu'extrêmement stables. En SDS-PAGE, on observe un profil de migration net composé de nombreuses bandes discrètes suggérant la conservation de l'intégrité de la structure primaire des protéines de l'extrait.

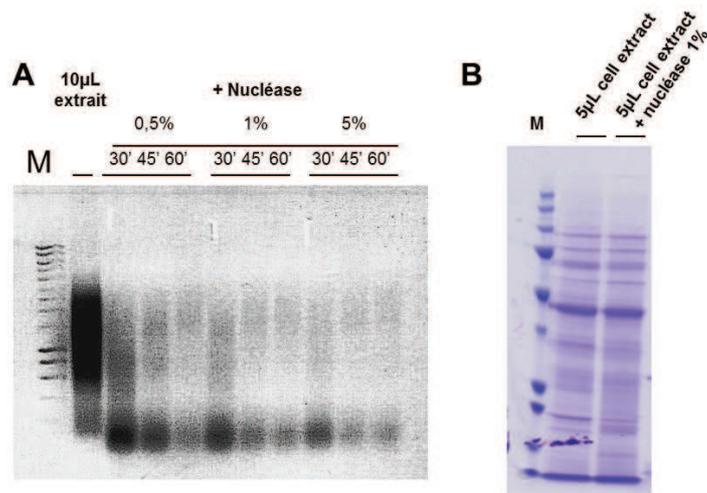


Figure 41 : Effet des nucléases sur l'extrait cellulaire de *P. abyssi*.

A) Extrait cellulaire de *P. abyssi* gel d'agarose (coloration BET). Effet des nucléase à différents temps et différentes concentrations. B) Profil de migration des protéines de l'extrait en SDS- PAGE, avec ou sans incubation pendant 45' en présence de 1% de nucléases.

5.4 MISE AU POINT DU PROTOCOLE DE PURIFICATION D'AFFINITE DES COMPLEXES.

Comme précédemment, c'est avec la protéine PCNA que nous avons déterminé les conditions permettant la formation et la purification des complexes spécifiques, cette fois ci au sein de l'extrait cellulaire de *P. abyssi*. Environ 30 µg de PCNA sont fixées sur les billes magnétiques et incubées dans 4 ml de l'extrait cellulaire de *P. abyssi*. Après incubation, pendant 3 h, à 4 °C, sous une agitation douce, les billes magnétiques sont séparées par aimantation de l'extrait et lavées extensivement par le tampon PBS. Cette étape permet d'éliminer les dernières traces de l'extrait et un maximum de partenaires non spécifiques. Les complexes, lavés et purifiés, sont ensuite élués préalablement à l'analyse MS (Figure 38 C).

Les protéines de la maintenance génomique ont, par nature, la propension à se lier à l'ADN ou à former des complexes dont l'ADN est un des composants. Afin de ne préserver que ces complexes spécifiques, que leurs formations soit dépendantes ou non de l'ADN, et de se débarrasser des partenaires non-spécifiques dont la présence est due à leurs affinités pour l'ADN, j'ai réalisé une seconde expérience d'AP-MS (Figure 38). Cette fois, les complexes purifiés sur billes ont été incubés en présence de nucléases, dégradant les acides nucléiques. Dans cette deuxième expérience (nuc+) les interactions faisant intervenir de longs fragments d'ADN sont éliminés. J'ai déterminé la concentration et le temps d'incubation nécessaires pour observer la dégradation de l'ADN génomique de *P. abyssi* dans l'extrait (Figure 41). La condition retenue est une incubation de 45' en présence de 1% (vol/vol) d'un mélange de nucléases ADN/ARN (Invitrogen). Les purifications d'affinité, en présence ou absence de nucléases, réalisées avec la protéine PCNA, ainsi qu'un témoin négatif sont présentées en figure 42. Les témoins négatifs consistent à incuber l'extrait de *P. abyssi* avec des billes « nues » et à éluer les protéines qui s'y sont fixées.

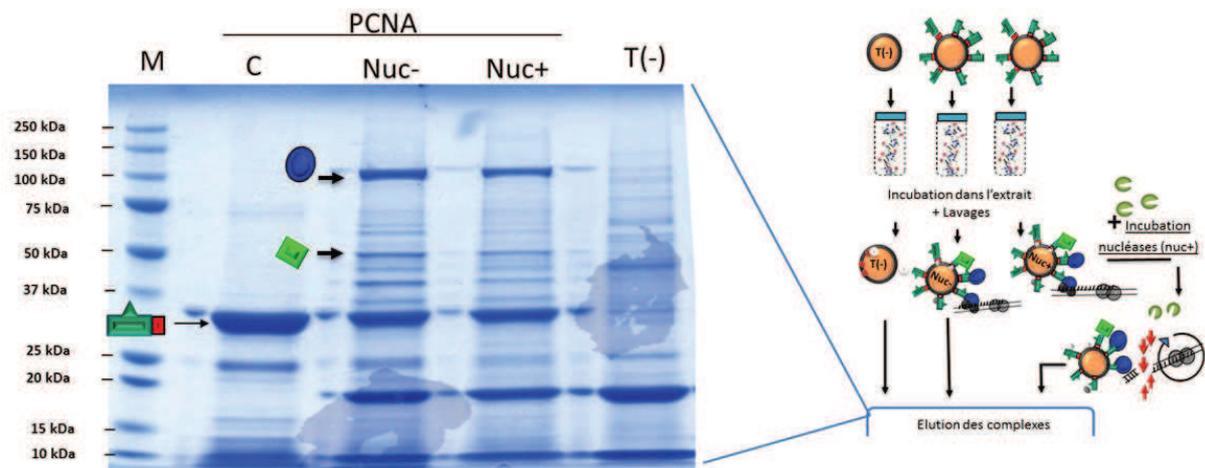


Figure 42 : Purifications d'affinité des complexes, l'exemple PCNA.

Les deux purifications nuc- et nuc + sont migrées en parallèle d'un témoin de charge de la protéine sur les billes (C) et d'un témoin négatif (T-). La position du PCNA est marquée par une flèche.

Nous observons une stabilité certaine de la proie sur les billes, même si la quantité de l'appât finalement présent diminue. Cette diminution du signal peut s'expliquer par le décrochage possible d'une partie des protéines appât mais également par la perte d'une partie des billes lors des étapes de lavages. Les billes ont également tendance à coller aux plastiques de certains récipients ce qui provoque une légère diminution de la quantité finale de billes. Cela met en exergue l'enjeu crucial que constitue la purification d'une quantité importante de protéines.

Nous observons en SDS-PAGE la présence de nombreuses bandes surnuméraires en présence de l'appât PCNA. Ces bandes sont d'intensités variables et de masses apparentes variées. On observe par exemple une bande à 100 kDa dont l'intensité est quasiment équivalente à celle de l'appât. Dans la mesure où une bande surnuméraire, absente à la fois dans le témoin de charge (piste C, Figure 42) et le témoin négatif (piste T(-), Figure 42), constitue une protéine potentiellement partenaire de l'appât, la quantité purifiée est largement à la portée d'une identification par MS. Il est difficile d'estimer l'élimination d'espèces protéiques à partir d'observation du gel, l'effet de l'incubation avec les nucléases sera plus probablement identifiable dans les données de MS.

On comprend ici également la difficulté d'une extraction manuelle des bandes d'autant que le PCNA constitue un des cas les plus nets en termes d'apparitions de bandes surnuméraires sur gel. L'imprécision de la découpe et la diffusion des protéines impliqueront nécessairement l'identification d'une pluralité d'espèces dans une seule bande du gel rendant l'identification de l'espèce protéique d'intérêt difficile.

La quantité de protéines se fixant de façon non spécifique sur les billes reste significative malgré les lavages intensifs (piste T(-), Figure 42). Afin de diminuer le nombre de partenaires aspécifiques, plusieurs adaptations du protocole expérimental ont été entreprises. J'ai notamment réalisé une pré-incubation de l'extrait avec une importante quantité de billes en espérant éliminer de l'extrait les espèces interagissant systématiquement avec les billes. Dans 3 lysats de volume et de concentration équivalents à ceux d'une expérience de pull down, j'ai

incubé 2, 8, et 16 mg de billes nues pendant 1 h. Ensuite les billes sont éliminées, et les extraits déplétés incubés avec des billes chargées en PCNA, pour une purification d'affinité. Je présente ci-après les profils de migration en SDS-PAGE des protéines éliminées lors de la pré-incubation (Figure 43A) et des purifications d'affinités dans les extraits cellulaires déplétés (Figure 43B). Le principe expérimental est présenté en figure 43 C. On observe que la déplétion n'a que peu d'effets malgré l'utilisation d'une quantité de billes jusqu'à 8 fois supérieure à la quantité utilisée lors d'une AP standard.

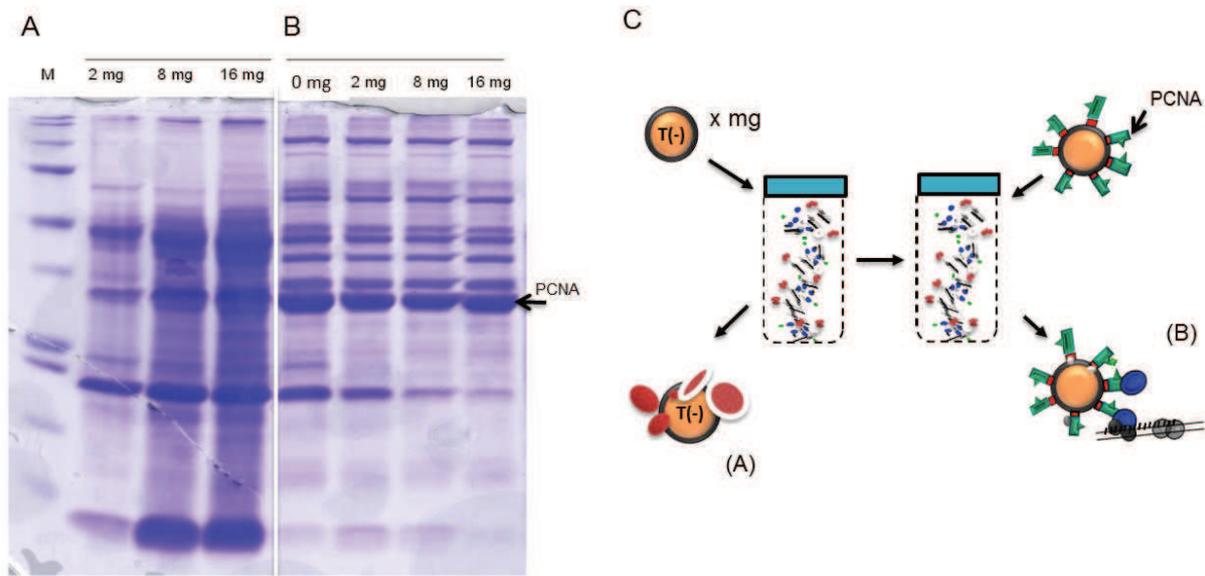


Figure 43: Effet de la déplétion sur un extrait cellulaire utilisé en AP.

A) Profil d'éluion des protéines se fixant sur 2, 8 et 16 mg de billes. B) Profil d'éluion de purification d'affinité dans un extrait pré incubé avec 0, 2, 8 et 16 mg s billes C) Schématisation de l'expérience

La concentration élevée de l'extrait cellulaire apparaît donc comme un obstacle important à une déplétion réellement efficace. Au final, je n'ai donc pas retenu cette étape de déplétion dans le protocole expérimental. Par contre, nous avons à ce stade décidé de caractériser de manière exhaustive la liste des protéines de *P. abyssi* se liant de façon non spécifique sur les matrices de Cobalt, au sein des témoins négatifs. Cette connaissance s'avère en effet indispensable pour la consistance et la robustesse de cette approche. Je détaillerai un peu plus loin la démarche de filtrage des données de spectrométrie de masse, prenant en compte ce listing. J'ai effectué 5 témoins négatifs indépendants et pas moins de 33 témoins négatifs, relatifs à chaque pulldown. Ceci nous a permis d'obtenir une somme de données considérable sur les sources potentielles de résultats non-spécifiques, ce qui est un point clé de cette approche méthodologique. On peut également signaler qu'en faisant varier la concentration d'imidazole aucun effet notable n'a été observé. La variation de ce paramètre fut limitée aux concentrations ne risquant pas d'entraîner le décrochage de la protéine appât. De même l'ajout d'un détergent aurait pu limiter les interactions hydrophobes des protéines de l'extrait aux billes. En revanche, une élimination imparfaite du détergent pourra diminuer notre sensibilité en MS. Il existe des détergents compatibles à la spectrométrie de masse, en revanche leur prix (35€ par milligramme) confine leur utilisation en faible volume.

Au regard de la concentration de l'extrait et de la quantité de billes utilisées, l'intensité du signal non spécifique est maîtrisée et témoigne d'une certaine stabilité. Il est probable que deux facteurs essentiels influencent sa composition, le premier est l'affinité à la matrice ou à l'ion présenté et le second l'abondance des protéines dans l'extrait (Tableau 7). Dans la mesure où la préparation de l'extrait est constante on peut raisonnablement espérer conserver cette constance dans la population de protéines se fixant sur les billes dans les différentes purifications d'affinité. Afin de déterminer le mode d'élution nous offrant le signal le plus net en MS, nous avons comparé la performance de l'analyse MS en utilisant différentes conditions.

5.5 MISE AU POINT DES CONDITIONS D'ELUTIONS

Afin d'être identifiés en spectrométrie de masse, les complexes doivent être séparés des billes dans les conditions qui permettent une analyse en MS dans des conditions optimales *i.e.* une concentration minimale en sels, détergents ou autres molécules diminuant drastiquement la performance de l'analyse. Toute molécule ayant la capacité de se charger lors de la désorption du solvant sera dirigée vers le détecteur et son signal en masse sur charge créera un bruit de fond qui peut se révéler rédhibitoire pour l'analyse.

En utilisant comme preuve de concept, une purification d'affinité en présence de l'appât PCNA, j'ai comparé plusieurs conditions d'élutions des complexes. Nous avons choisi initialement de ne pas séparer les complexes sur gel avant l'analyse MS, mais de les éluer lors de l'étape finale de l'expérience pull-down pour ensuite analyser en LC-MS/MS l'intégralité des protéines présentes dans l'échantillon. Il fallait donc à ce stade déterminer la composition idéale du tampon d'élution, capable de décrocher les protéines et complexes formés à la périphérie des billes de Co et qui soit compatible avec la digestion trypsique et l'analyse MS. D'emblée nous avons écarté l'utilisation d'Imidazole et d'EDTA qui posent problèmes pour la MS. Nous avons testé trois conditions différentes couramment utilisées pour ce type d'analyses (Padliya and Wood 2004; Huang, Nichols et al. 2009; Austin, Smidansky et al. 2012).

(1) élution pendant 30', 95°C dans un tampon Tris Hcl 40 mM pH 8

(2) élution 30' à 95°C dans du Rapigest 0,1%, détergent doux compatible avec la MS et utilisé pour solubiliser et dénaturer les protéines sans réduire l'activité protéolytique de la trypsine.

(3) élution par précipitation des protéines en présence d'acetonitrile 40% et TFA 0,1 %. Après dessiccation, les protéines sont reprises dans une solution Rapigest 1%.

Après élution dans ces trois conditions, les échantillons ont été envoyés à la plateforme protéomique de BioGenOuest (Université de Rennes I) pour des analyses LC-MS/MS et une fraction a été déposée sur gel SDS-PAGE pour vérifier l'efficacité de l'élution. La figure 44 montre que l'élution des complexes dans les conditions 1 et 2 se fait de manière comparable et efficace. Par contre l'élution en condition 3, par précipitation des protéines, est incomplète, et d'autre part la quantité de protéines récupérées après resuspension dans une solution de Rapigest 1% est assez faible. La figure 44B décrit le nombre de protéines identifiées, avec un score

Mascott supérieur à 100 (voir section analyse de spectrométrie de masse), pour chaque condition testée. On voit clairement que la condition d'élution 2, en présence uniquement de Rapigest, permet d'obtenir un plus grand nombre d'identifications de partenaires potentiels. D'autre part nous retrouvons dans la liste de ces protéines, les interactants bien connus du PCNA, à savoir : Fen1 et RF-C, ainsi que de nouveaux partenaires potentiels, par exemple mre11 et Rad50 qui confirmaient les résultats précédemment obtenus par J. Briffotiaux en utilisant une approche différente (voir présentation de l'étude). La comparaison des résultats entre les conditions 1 et 2 tend à démontrer que l'étape de trypsinisation des protéines thermostables constitue probablement l'étape limitante pour l'identification des protéines en MS. A ce stade nous pensions donc avoir trouvé des conditions nous permettant de démarrer l'étude à plus grande échelle.

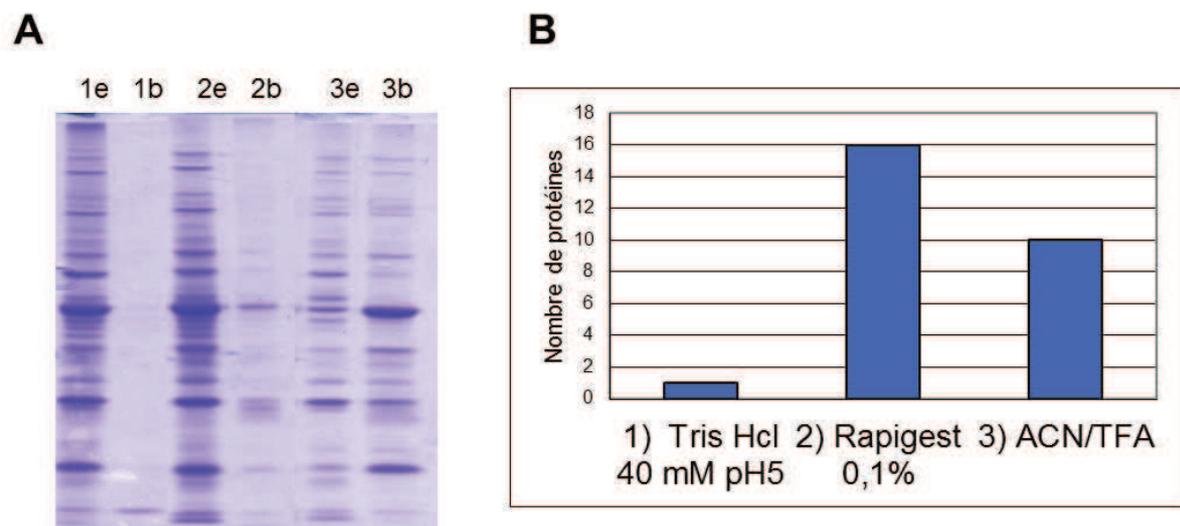


Figure 44 : Conditions d'élution des complexes et intensité du signal en MS.

A) Comparaisons des conditions d'élutions. Les complexes purifiés avec PCNA sont élués dans un Tampon Tris Hcl 40 mM pH5 (1e); Rapigest® 0,1% (2e) ; H2O après précipitation Acetonitrile 40% ; TFA 0,1 % (3e). Les protéines non éluées sont migrées en SDS-PAGE 1b ,2b, 3b. B) Comparaison du niveau de détection en MS pour l'analyse des éluats liquides avec les tampons 1, 2 et 3.

Seulement, si on replace ces évènements dans l'ordre chronologique, il s'est passé un délai de 5 mois entre l'envoi des trois échantillons de test pour les tampons d'élution et le retour des résultats de l'analyse par la plate forme protéomique. J'ai voulu mettre à profit ce laps de temps en tentant d'améliorer la dernière étape de pulldown par un lavage à l'éthanol 30% avant élution. Le lavage à l'éthanol est couramment utilisé pour débarrasser les échantillons de traces de sel avant analyse en MS. J'ai donc utilisé cette étape supplémentaire, toujours en utilisant le PCNA comme appât, dans trois pulldowns reprenant les conditions d'élution, décrites plus haut. Après migration sur gel SDS-PAGE des protéines éluées, j'obtenais des profils identiques aux essais précédents ne contenant pas cette nouvelle étape de lavage à l'alcool. Je pensais donc au final avoir déterminé la condition d'élution idéale en tampon Rapigest 0,1% et en plus avoir sensiblement amélioré le niveau de sensibilité de détection en MS par cette étape supplémentaire de lavage (même si je ne l'avais pas testé jusqu'à l'analyse MS). J'ai donc à ce stade réalisé en série plus de 15 pulldowns qui constituaient le cœur de

mon travail expérimental, en incluant cette étape de lavage, et j'ai envoyé les échantillons à la plateforme protéomique. Le mieux étant parfois l'ennemi du bien, lorsque j'ai reçu les premiers résultats, que je pensais définitifs, nous nous sommes aperçus que les résultats n'étaient absolument pas en rapport avec ceux des tests préliminaires. A titre d'exemple, en utilisant le PCNA comme appât et en incluant l'étape de lavage à l'éthanol, je n'identifiais plus qu'une seule protéine avec un score Mascott spécifique, l'endonucléase Fen1. Après discussion avec les ingénieurs de la plate forme, nous nous sommes rendu compte que nos échantillons contenaient une quantité importante de Polyéthylène Glycols (PEG) qui était responsable de la mauvaise qualité des résultats. Une pollution aux PEGs provient le plus souvent d'une solution impure ou de la dissolution de plastiques par certains solvants, notamment l'éthanol. A ce stade, l'intégralité des échantillons n'avait pas encore été traités en analyse MS, et nous avons tenté de nous débarrasser des PEGs en faisant migrer les échantillons restant dans un gel SDS-PAGE, sur 2 cm, puis de découper l'intégralité du profil protéique en 3 bandes de largeur équivalente. Nous espérions que les PEGs, qui sont de gros polymères branchés, ne migrent pas dans le gel. Les protéines contenues dans ces trois bandes ont été ensuite analysées en LC-MS/MS après extraction du gel et digestion trypsique. La figure 45 montre la comparaison des résultats de deux pulldown PCNA obtenus, lors des tests préliminaires et après migration des échantillons contaminés aux PEGs dans un gel SDS-PAGE. On voit clairement que l'étape de migration sur gel, non seulement permet d'éliminer l'effet négatif des PEGs mais en plus améliore énormément la sensibilité de l'analyse MS. Cette péripétie dans la mise au point de ce protocole nous aura finalement permis d'améliorer la qualité générale des résultats obtenus.

En plus de l'élimination des PEGs deux autres facteurs peuvent expliquer une qualité de la détection plus élevée pour l'élution SDS-PAGE. Tout d'abord, il a été démontré que les protéines d'organismes hyperthermophiles étaient particulièrement réfractaires à l'étape de digestion trypsique en conditions natives (Lee, Sevinsky et al. 2009). La compacité et la rigidité des protéines hyper thermostables rend inaccessible à la trypsine de nombreux sites de digestion diminuant ainsi drastiquement le nombre de peptides détectables. Ensuite, les échantillons migrés en SDS-PAGE sont fractionnés en trois morceaux augmentant la sensibilité de l'analyse. En effet le fractionnement de l'analyse permet d'utiliser la gamme dynamique du détecteur sur l'intégralité de l'échantillon, mais avec une définition beaucoup plus élevée. Ce fractionnement limite également le recouvrement de spectre.

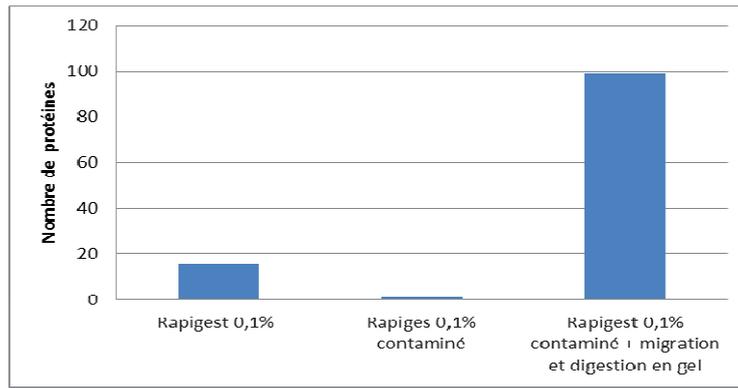


Figure 45 : Nombre d'espèces protéiques identifiés selon le mode d'éluéon.

5.6 ANALYSE EN SPECTROMETRIE DE MASSE

Les protéines partenaires des protéines appâts utilisées ont été identifiées en utilisant l'approche dite LC-MS/MS, qui consiste à coupler la séparation de peptides en chromatographie liquide (LC) avec leur identification par une étape de spectrométrie de masse en tandem (MS/MS).

La spectrométrie de masse permet d'identifier le rapport masse sur charge de molécules, y compris les polypeptides, en identifiant leurs comportements dans le vide sous l'influence d'un champ magnétique. Dans nos expériences, les protéines digérées sont injectées dans une colonne de nano chromatographie liquide avant d'être vaporisées et ionisées dans le vide de la chambre du spectromètre. Sous l'influence du champ magnétique les ions sont dirigés vers le détecteur. Grâce à un second champ magnétique (perpendiculaire au premier) les ions seront mis en orbite dans la chambre du détecteur. On calcule à partir de leurs fréquences de rotation le rapport masse sur charge de l'ion considéré, dans notre cas du polypeptide. De ce rapport, on déduit la masse et la composition possible de ce peptide en acides aminés. Cette information est comparée au protéome de *P. abyssi* traduit et digéré *in silico* (Figure 46). Ce type d'analyse de mélanges protéiques complexes est appelé approche « shot gun ».

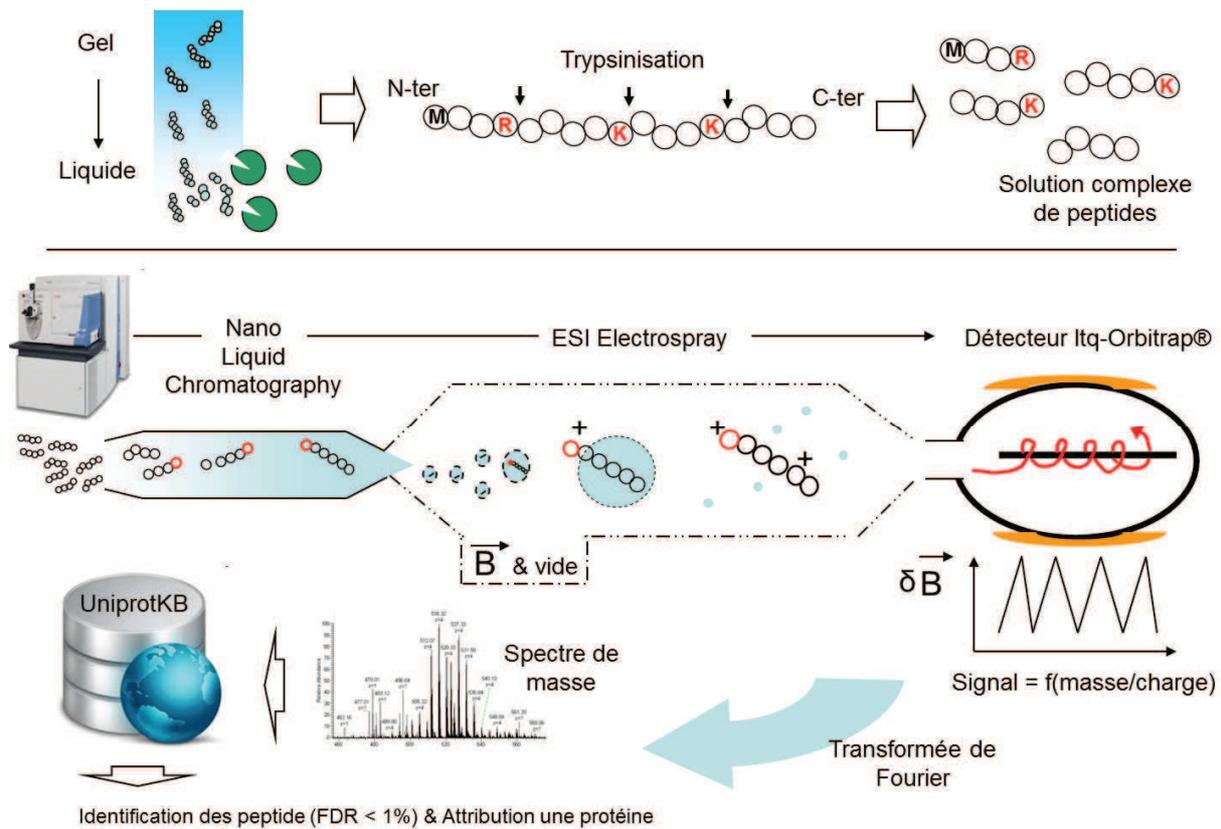


Figure 46 : Schématisation du processus d'identification des protéines en MS.

Les protéines purifiées sont extraites du gel et digérées par la trypsine. Les peptides sont ensuite injectés dans le spectromètre de masse. A la sortie de la nano chromatographie liquide, le solvant des peptides est évaporé, les ions se chargent et sont dirigés ainsi par un champ magnétique en direction du détecteur. Dans les détecteurs que nous utilisons, les peptides sont mis en orbite. La fréquence de rotation du peptide dans le détecteur permettra d'identifier son rapport masse sur charge à une approximation inférieure à 1 ppm.

Chaque peptide identifié est attribué au protéome de *P. abyssi* traduit et digéré *in silico* dans la base de données Uniprot (www.uniprot.org). Les paramètres de détection sont optimisés pour ne conserver que les détections non ambiguës, *i.e.* ne possédant qu'une seule correspondance. Une protéine est donc détectée avec un nombre de correspondance (« match ») peptidique couvrant un pourcentage déterminé de la séquence. Le processus d'attribution des peptides est basé sur l'algorithme probabiliste nommé MASCOT. L'algorithme MASCOT calcule la probabilité P qu'une correspondance à une entrée de la base de données soit erronée, *i.e.* une correspondance aléatoire. Pour lui donner une apparence plus intuitive le score est présenté sous la forme $S = -10 \cdot \log(P)$ de sorte que la valeur soit positive. Plus le score MASCOT est élevé, meilleure est l'identification.

La spectrométrie de masse n'est cependant pas une méthode strictement quantitative lorsque l'analyse porte sur des mélanges complexes de protéines non marquées. En effet, la dépendance de la détection à la composition du polypeptide analysé produit une décorrélation entre la quantité des espèces protéiques et leurs niveaux respectifs de détection. En effet, à nombre de molécules égal, une grosse protéine de 1000 acides aminés a beaucoup plus de chance d'être détectée qu'une protéine n'en possédant que 100. Au biais de la taille s'ajoute celui de l'hétérogénéité de charge des peptides. Les peptides chargés négativement ne volent pas bien, voire pas du tout, n'atteignant jamais le détecteur du spectromètre. Ces peptides ont donc moins de

chance d'être détectés même s'ils sont présents dans l'échantillon. Aussi des modifications post-traductionnelles telles que les phosphorylations seront défavorables au vol du peptide modifié. L'injection des peptides dans le spectromètre par nano-chromatographie liquide couplé à une injection par nébulisation ou « electrospray ionization mass spectrometry » (ESI-MS) résout en partie ce problème. La vaporisation d'un solvant liquide permet la formation d'ions multichargés pouvant contrecarrer la charge négative de certains peptides.

Ces différentes propriétés font de notre analyse en spectrométrie de masse une technique semi-quantitative. Le score MASCOT peut cependant être considéré comme une bonne approximation de la quantité relative de la même espèce protéique dans des analyses différentes. Le nombre de peptides détectés et le taux de couverture de séquence d'une même protéine augmenteront conjointement à sa quantité. Cela se traduira par une augmentation du score MASCOT. Cette propriété établit le score MASCOT comme un outil robuste dès lors que l'on compare une série d'analyse à des expériences témoins. En comparant la composition protéique d'une purification d'affinité à des expériences témoins, nous pourrions détecter un enrichissement en certaines espèces, spécifiques à une protéine appât. Nous postulons que cet enrichissement aura été causé par l'interaction spécifique entre les protéines fixées sur les billes et les protéines de l'extrait cellulaire. C'est grâce à cet effet de co-purification que nous détectons les partenaires spécifiques d'une protéine. Nous avons également démontré précédemment que ceci n'est valide que dans le cas où les paramètres de digestions sont constants.

5.7 FILTRE DES DONNEES

L'analyse en spectrométrie de masse a identifié plus de 100 protéines dans chaque essai PCNA_{nuc-} et PCNA_{nuc+}. Parmi les nombreuses protéines identifiées, seule une fraction correspond à des partenaires spécifiques tandis que les autres « matchs » correspondent à des interactions non-spécifiques avec les billes magnétiques. Cela implique que les données brutes doivent être filtrées afin d'éliminer ces données non informatives. D'où l'intérêt de posséder des données exhaustives sur les protéines se fixant de manière non-spécifiques sur les billes.

De nombreuses méthodes ont été développées pour l'identification du signal spécifique en spectrométrie de masse (Brymora, Valova et al. 2001; Yu, Ivanic et al. 2009). Nous avons pris le parti de développer notre propre outil avec une méthode simple basée sur la reproduction expérimentale. En effet, d'une certaine manière, les essais nuc+ et nuc- correspondent à un duplicat de l'expérience, basée également sur une connaissance exhaustive des protéines de *P. abyssi* se liant de manière non-spécifiques sur les billes de Co, plutôt que sur un outil statistique sophistiqué dont la maîtrise est périlleuse sans une bonne compréhension des concepts mathématiques complexes sur lesquels ils reposent.

Nous avons observé que la population des espèces protéiques se fixant de manière non-spécifiques sur les billes est globalement constante en l'observant en SDS-PAGE. La stabilité apparente de la population fixée directement sur les billes rend envisageable l'identification de l'ensemble de ses composants. Nous avons donc

mis en place un protocole simple et adapté qui consiste à comparer les scores représentant le niveau d'apparition d'une protéine dans un essai, à son niveau maximal observé dans l'ensemble des 5 témoins négatifs et dans le témoin relatif (réalisé en même temps que l'expérience pulldown avec le même extrait cellulaire). Si nous considérons le score MASCOT comme la mesure de l'intensité du signal d'une protéine dans une purification alors il constitue, dans un témoin, la mesure du bruit de fond. Le profil de détection des protéines détectées au moins une fois avec un score mascot supérieur à 20 dans les témoins négatifs indépendants est présenté en figure 47.

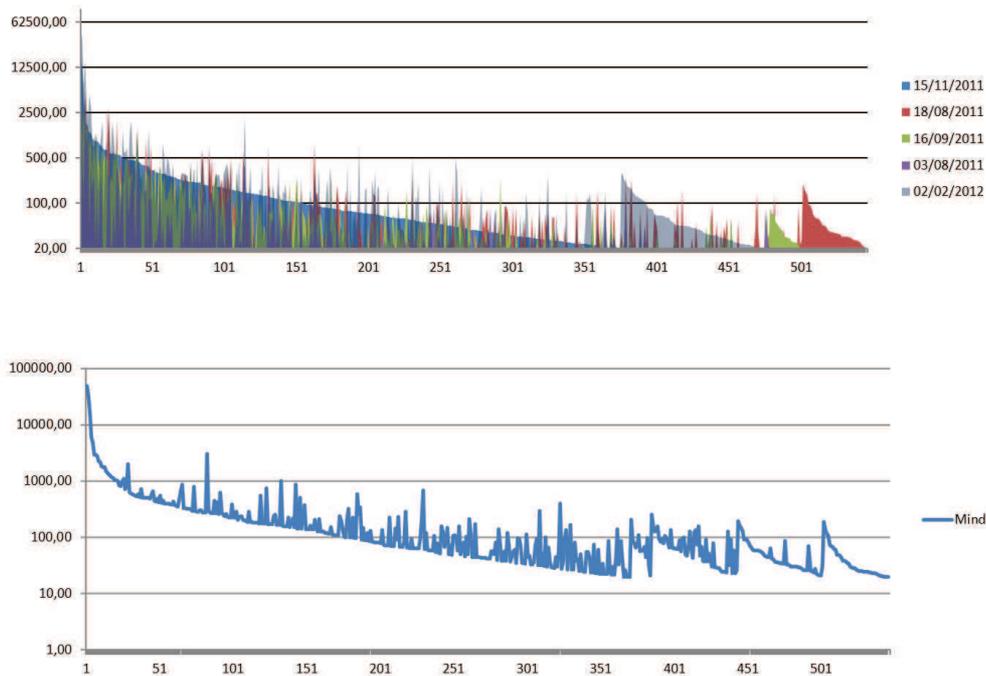


Figure 47 : Profil d'identification des protéines des témoins négatifs indépendants.

Les scores Mascot des protéines de chaque témoin indépendant sont compilés et classés par ordre successif. Seule la valeur maximale obtenue (Mind) est représentée dans le deuxième graphique. Chaque unité de l'axe X correspond à une protéine, l'axe Y correspond à son score Mascot.

Pour identifier les interactions spécifiques à la présence d'une protéine appât, nous calculons le rapport SR entre le score MASCOT maximal des témoins (Mneg) et le score MASCOT des purifications sans et avec traitement aux nucléases (Mnuc- et Mnuc+ respectivement). Ce ratio, SR- ou SR+, correspond à un rapport "bruit sur signal" qui tendra vers zéro lorsque l'écart entre l'essai et le témoin augmentera. Ainsi, un SR de 0,1 correspond à la détection d'une protéine dans une purification avec un score MASCOT 10 fois plus élevé que son score MASCOT le plus important au sein des témoins. Le SR n'est pas calculé pour les cas où le contrôle Mneg est supérieur aux essais Mnuc+/- puisque qu'il correspond aux protéines non spécifiques.

Le score MASCOT de chaque protéine détectée est comparé à celui obtenu dans 6 témoins négatifs, lorsque la protéine considérée est bien sur présente dans au moins un des témoins négatifs. Dans le cas où une protéine n'est détectée dans aucun des témoins, on lui attribuera arbitrairement une valeur Mneg de 20. Pour une protéine qui n'a jamais été détectée dans les témoins négatifs (Mneg=20) elle obtiendra un SR de 0,2 ou moins

dès lors qu'elle franchira la valeur 100 en score MASCOT. Cinq témoins dont l'analyse a été réalisée indépendamment des séries d'expériences de purification d'affinité constituent une base de données de témoins indépendants pour l'ensemble des analyses. Les 20 protéines détectées avec les scores les plus élevés et le rang des protéines de la maintenance génomique est présentée dans le tableau 7A. Il y est également présenté les résultats d'analyse en masse de l'extrait total (Tableau 7B ; le score présenté correspond au score le plus élevé de deux analyses de 20µl d'extrait déposé sur gel). Parmi les protéines les plus détectées dans les témoins négatifs sur billes nues, plusieurs ont une affinité pour les ions divalents et possèdent certainement une affinité pour les ions sur les billes. Certaines protéines sont fortement détectées à des niveaux élevés dans l'extrait brut et dans les témoins. Les protéines chromatinienne Histone A et Alba sont parmi les protéines les plus détectées dans les témoins et l'extrait brut. Comme elles apparaissent systématiquement, l'identification de ces deux protéines en tant que partenaire spécifique sera assez difficile. Cette information croisée, sur l'abondance cellulaire et la tendance à se fixer de façon non-spécifique sur les billes, nous a donc amené à traiter de façon prudente les interactions faisant intervenir ce type de protéines. On observe cependant que les protéines de la maintenance génomique, si elles ont été détectées, le sont principalement avec des scores faibles, me donnant bon espoir de pouvoir identifier le signal spécifique d'un enrichissement en AP parmi le bruit de fond inhérent à l'analyse. Le témoin de l'extrait brut est présenté à titre indicatif, il n'est pas inclus dans le calcul du SR que je décris ci-après dans le texte.

Tableau 7 : Protéines détectées dans les témoins négatifs et dans l'extrait brut.

A Témoins indépendants				B Extrait brut			
Rang	Accession	Protéine	Mneg	Rang	Accession	Description	Mascot
1	Q9UYI0	MoaB-2 molybdenum cofactor biosynthesis pro	49219,3	1	Q47950	Glutamate dehydrogenase OS=Pyrococcus abys	45033,2
2	Q9V0N3	Putative uncharacterized protein	32297,2	2	Q9V0N3	Putative uncharacterized protein OS=Pyrococcu	36409,9
3	Q9V0Y8	Putative snRNP Sm-like protein	16609,8	3	Q9V0Z4	EndO 1,4-beta-glucanase (Cellulase M) (EGM) OS	12010,2
4	Q9UXV7	Putative uncharacterized protein	6117,5	4	Q9V1B2	Serine hydroxymethyltransferase OS=Pyrococcu	10344,3
5	Q9V0P3	ABC-type periplasmic dipeptide transport prote	4788,4	5	Q9UZE9	For formaldehyde:ferredoxin oxidoreductase O	9500,6
6	Q9UZH3	Glucose-1-phosphate thymidyltransferase	3032,4	6	Q9UYZ4	Pvruvate synthase subunit porA OS=Pyrococcus	7996,9
7	Q9UZC1	Mannosyl-3-phosphoglycerate synthase	3008,5	7	Q9V0V7	Elongation factor 1-alpha OS=Pyrococcus abys	7963,7
8	Q9V2H7	Probable phosphoenolpyruvate synthase	2970,1	8	Q9UY58	Putative uncharacterized protein OS=Pyrococcu	7756,7
9	Q9V1N3	DNA/RNA-binding protein Alba	2828,0	9	Q9V2H7	Probable phosphoenolpyruvate synthase OS=Py	7063,8
10	Q9V1P9	Predicted RNA-binding protein	2285,5	10	Q9V2Q7	Thermosome subunit OS=Pyrococcus abys	6780,4
11	Q9UZ85	Putative uncharacterized protein	2205,8	11	Q9V0Q7	Pyridoxal phosphate-dependent aminotransfer	6624,6
12	Q9V0T6	MoaA-like intein containing molybdenum cofac	2049,5	12	Q9UY99	Pvruvate/ketoisovalerate oxidoreductases com	6260,1
13	Q9UYX4	Riboflavin kinase	1866,2	13	Q9UY77	Acetyl coA C-acetyltransferase OS=Pyrococcus	4145,3
14	Q9V0L0	Putative uncharacterized protein	1807,7	14	Q9V0F0	SgaA-like serine-glyoxylate aminotransferase re	4110,1
15	Q9V1B7	Iron (III) ABC transporter,	1766,3	15	P61881	Archaeal histone A OS=Pyrococcus abys	3574,8
16	Q9UZ78	Putative tRNA	1523,3	16	Q9V1Z8	Elongation factor 2 OS=Pyrococcus abys	3473,3
17	Q9UYR4	Dehydrogenase	1420,5	17	Q9V1Q0	Aminopeptidase from family M42 OS=Pyrococc	3383,0
18	Q9UYV0	6-pyruvoyl-tetrahydropterin synthase	1310,1	18	Q9V0M8	SerA D-3-phosphoglycerate dehydrogenase OS=	3310,1
19	Q9UZE9	For formaldehyde:ferredoxin oxidoreductase	1261,6	19	Q9UXZ0	Enolase OS=Pyrococcus abys	3266,6
20	Q9V1G0	Translation initiation factor 2 subunit gamma	1189,3	20	Q9UZ57	Indolepyruvate oxidoreductase subunit iorA OS	2979,8
48	Q9V2G4	Replication factor C small subunit	559,5	38	Q9V1N3	DNA/RNA-binding protein Alba OS=Pyrococcus	1284,5
58	P61881	Archaeal histone A	503,2	61	Q9V0B3	Predicted DNA methylase OS=Pyrococcus abys	909,0
95	Q9V2S0	Cell division protein FtsZ 1	294,2	76	Q9UZ86	Reverse gyrase OS=Pyrococcus abys	664,2
106	Q9V1T1	Metal-dependent phosphohydrolase, HD super	272,9	85	Q9UYJ2	ParB-like nuclease, putative OS=Pyrococcus aby	611,1
107	Q9V233	DNA repair and recombination protein radA	272,6	86	Q9V2S0	Cell division protein ftsZ homolog 1 OS=Pyrococ	588,5
108	Q9V135	Type 2 DNA topoisomerase 6 subunit B	268,8	91	Q9V1N2	Cdc48 cell division control protein 48, AAA fami	572,3
133	Q9UZC8	DNA double-strand break repair Rad50 ATPase	215,0	100	Q9V147	ATPase involved in chromosome partitioning, m	495,6
137	Q9UZ86	Reverse gyrase [Includes: Helicase	205,1	110	Q9UYX8	DNA polymerase sliding clamp OS=Pyrococcus	415,2
162	Q9V0B3	Adenine-specific DNA methylase	166,6	117	Q9V0U9	Predicted endonuclease OS=Pyrococcus abys	384,9
171	Q9V2F6	DNA repair and recombination protein radB	160,4	120	Q9UY85	RecJ-like phosphoesterase, containing RNA binc	350,4
195	Q9V2G3	Replication factor C large subunit	139,1	135	Q9V2A5	RecA superfamily ATPase implicated in signal tr	298,8
197	Q9V2H1	CCA-adding enzyme	137,4	143	Q9V1X8	HsdM type I restriction modification enzyme, su	286,2
198	Q9V278	DNA repair protein RAD25	136,9	157	Q9V2G5	DNA replication helicase, dna2 homolog OS=Py	267,5
211	Q9V134	Type II DNA topoisomerase VI subunit A	123,9	190	Q9V2G4	Replication factor C small subunit OS=Pyrococc	214,8
221	Q9V2K8	RecJ-like single-stranded exonuclease recJ hom	114,7	220	Q9UZ61	Cell division protein ftsZ homolog 2 OS=Pyrococ	174,8
306	Q9V2F4	DNA polymerase II large subunit	107,0	230	Q9UZC8	DNA double-strand break repair rad50 ATPase (164,7
236	Q9V1Y9	RPA41 subunit of the hetero-oligomeric comple	102,9	232	Q9V135	Type 2 DNA topoisomerase 6 subunit B OS=Pyrr	162,8
263	Q9V0P9	Flap endonuclease 1	90,2	241	Q9V185	DNA ligase OS=Pyrococcus abys	155,9
268	Q9V2F4	DNA polymerase II large subunit	86,3	257	Q9V1F5	Archaeal histone B OS=Pyrococcus abys	137,5
287	Q9UZC9	DNA double-strand break repair protein Mre11	79,0	270	Q9V1Y9	RPA41 subunit of the hetero-oligomeric comple	131,4
305	Q9V1Y1	HsdR type I restriction-modification enzyme, R	70,6	301	Q9V233	DNA repair and recombination protein radA OS	109,2
306	Q9UZT3	TatD related DNase	70,1	306	Q9V2F4	DNA polymerase II large subunit OS=Pyrococcus	107,0
307	Q9UYW2	N6-adenine-specific DNA methylase	69,7				
320	Q9UYX8	DNA polymerase sliding clamp	66,1				

Les protéines sont classées selon leur score. Les protéines associées à la maintenance génomique sont surlignées en bleu.

On associe également un témoin relatif supplémentaire à chaque AP. Il est réalisé en parallèle de chaque pull-down et analysé simultanément. Bien qu'il puisse constituer un témoin de la variation relative à l'expérience, il ne constituera finalement qu'une vérification de l'absence de contamination croisée des échantillons lors de l'analyse. En effet, il a été démontré que des peptides rémanents d'une expérience précédente pouvaient contaminer plusieurs analyses LC-MS/MS (Bern, Phinney et al. 2009). C'est l'étape de nano-LC qui est souvent mise en cause. Certains peptides s'avèrent extrêmement affins à la matrice de la colonne et ne sont libérés de la colonne que par compétition avec de nouveaux peptides, provenant d'une seconde analyse effectuée avec un appât différent, et injectés sur la colonne contaminée. Ce biais pouvant donc potentiellement entraîner l'attribution erronée d'un partenaire à un appât quelconque, alors qu'en fait il interagissait avec l'appât de l'expérience AP/MS précédente.

Pour nous assurer de détecter ce type d'évènements de contaminations croisées, les échantillons sont systématiquement analysés dans l'ordre suivant en spectrométrie de masse :

1 Témoin négatif,

2 Essai traité à la nucléase,

3 Essai sans traitement à la nucléase.

Ainsi, chaque contamination potentielle, due à l'analyse précédente, pourra être détectée dans le témoin négatif relatif, injecté en premier. Les paramètres de filtres et de seuils de significativité ont été déterminés en utilisant la protéine PCNA comme référence. Les quantités de protéines partenaires purifiées semblent relativement importantes sur SDS-PAGE, on peut donc s'attendre à observer une variation significative du score MASCOT de certaines espèces.

Lorsqu'on analyse les résultats de l'AP-MS de PCNA avec un score SR limite de 0,2, nous identifions d'emblée 4 protéines dont les interactions étaient déjà décrites dans la littérature ; l'ADN polymérase 1, NucS, RF-C, et FEN1 (Figure 48). Ces interactions constituent des complexes intervenant à des étapes clé de la maintenance génomique notamment la réplication. Les résultats obtenus avec PCNA nous fournissent une confirmation expérimentale de la validité du concept de notre approche. Ils témoignent de notre capacité à former *in vitro* des complexes spécifiques et d'en identifier immédiatement les composants protéiques. Nous avons également observé plusieurs interactions décrites dans la littérature dont la détection s'effectuait au delà des seuils établis. Certaines protéines ne sont plus détectées (ou avec un SR au-dessus du seuil arbitraire) dans l'essai traité avec la nucléase quand d'autres présentent le comportement inverse, en ayant un score plus élevé dans l'essai non traité à la nucléase. Nous observons par exemple, que la petite sous unité du RF-C n'est détectée à un seuil satisfaisant que dans l'essai nuc+, contrairement à la grande sous unité (Figures 49 et 50). Obtenir un filtrage optimal avec un seuil de détection fixé empiriquement nécessiterait une précision de mesure beaucoup plus élevée avec idéalement une relation directe entre le score utilisé et la quantité de protéines en présence. Le caractère semi quantitatif intrinsèque à la spectrométrie de masse et les conditions de stringence de notre filtre nous expose à un taux de faux négatifs élevé. Ce seuil permet néanmoins d'identifier nettement plusieurs des partenaires décrits pour la protéine PCNA. Nos conditions de seuil SR 0.2 sont probablement imparfaites et un peu trop restrictives, mais l'exclusion de certaines interactions des données haute confiance s'explique aussi certainement par les effets de l'incubation avec des nucléases.

Tableau 8 : Partenaires de PCNA identifiées en AP-MS.

Protéines détectées					Nuc-			Nuc+			Témoins			Scores	
ID uniprot	# AAs	MW [kDa]	pI	Description	Σcov	Σ# Pep	Mnuc-	Σcov	Σ# Pep	Mnuc+	MInd	MRel	Mneg	SR-	SR+
Q9V0P9	343	39	8,9	FEN1	67,6	490	2614,3	69,1	440	4503,6	90,2	84,2	90,2	0,03	0,02
Q9V1Z5	752	86	9,5	ERCC4/XPF	29,5	71	434,1	18,2	17	242,5	25,8	1,0	25,8	0,06	0,11
Q9UYE0	516	60	6,3	Metallo-beta-lactamase pab1035	26,4	23	268,8	14,9	11	194,4	1,0	1,0	20,0	0,07	0,10
P0CL76	771	89	8,4	DNA polymerase 1	33,7	37	253,1	33,2	39	347,0	1,0	1,0	20,0	0,08	0,06
Q9V104	417	48	7,6	Q9V104	41,0	44	236,0	33,6	50	386,0	4,3	1,0	20,0	0,08	0,05
Q9V2E8	251	29	8,0	NucS	57,8	27	227,9	48,2	26	298,1	1,0	1,0	20,0	0,09	0,07
Q9V2G3	479	55	9,2	RFC1s	59,5	134	1489,0	37,2	96	894,6	139,1	144,2	144,2	0,10	0,16
Q9UZC8	880	104	7,0	Rad50	50,0	229	2020,9	57,1	875	12606,3	215,0	1,0	215,0	0,11	0,02
Q9UZC9	423	49	6,3	Mre11	56,3	117	712,4	62,9	322	2353,8	79,0	1,0	79,0	0,11	0,03
Q9V2F2	419	48	6,8	Cdc6	20,5	8	116,7	10,5	3	32,4	1,0	23,9	23,9	0,20	#N/A
Q9V1X9	427	49	8,8	HsdS	16,4	6	93,5	19,7	10	103,8	1,0	1,0	20,0	0,21	0,19
Q9V1X8	623	73	5,7	HsdM	27,9	24	412,1	16,2	17	273,9	90,9	49,3	90,9	0,22	0,33
Q9UYB0	631	73	9,1	ApeH acylamino-acid-releasing enzyme	11,6	13	155,9	#N/A	#N/A	#N/A	48,1	1,0	48,1	0,31	#N/A
Q9V2G4	1437	165	8,8	RFCss	15,7	298	1784,9	16,0	377	4055,2	559,5	378,7	559,5	0,31	0,14
Q9V1Y9	375	43	5,0	RPA41	30,1	37	295,5	#N/A	#N/A	#N/A	102,9	1,0	102,9	0,35	#N/A
Q9UYX8	249	28	4,6	PCNA	84,3	2165	15946,0	87,2	2469	24488,1	66,1	5983,7	5983,7	0,38	0,24
Q9UYB7	570	65	5,6	MutS	15,3	8	120,9	3,2	2	48,7	48,4	1,0	48,4	0,40	#N/A
Q9V135	564	64	9,1	Type II DNA topoisomerase VI subunit B	38,7	63	668,5	35,1	57	611,5	268,8	102,8	268,8	0,40	0,44
Q9V2B5	464	52	9,5	Site specific DNA-methyltransferase	37,3	32	270,7	29,1	18	180,0	145,1	1,0	145,1	0,54	0,81
Q9UY97	537	60	9,5	Q9UY97	5,8	4	113,8	#N/A	#N/A	#N/A	61,5	1,0	61,5	0,54	#N/A
Q9V297	453	51	4,9	Maltotriose-binding protein	19,2	15	204,4	8,2	4	104,8	113,3	1,0	113,3	0,55	#N/A
Q9UYW5	480	54	8,9	Q9UYW5	11,5	9	129,0	#N/A	#N/A	#N/A	82,7	1,0	82,7	0,64	#N/A
Q9V134	382	44	8,3	Type II DNA topoisomerase VI subunit A	21,2	12	179,6	28,3	20	196,7	123,9	106,7	123,9	0,69	0,63
Q9V0H3	694	78	8,9	Q9V0H3	38,0	44	576,8	7,1	5	56,4	404,0	38,9	404,0	0,70	#N/A
Q9UZG3	479	56	9,2	Sulfatase	10,2	8	90,7	#N/A	#N/A	#N/A	69,0	1,0	69,0	0,76	#N/A
Q9UYC6	998	112	6,3	Archaeal Lon protease	7,6	9	110,1	#N/A	#N/A	#N/A	87,4	1,0	87,4	0,79	#N/A
Q9UZD0	548	61	5,3	Predicted ATPase	38,9	44	736,3	44,9	50	891,2	554,4	590,0	590,0	0,80	0,66
P61881	67	7	9,1	Archaeal histone A	47,8	98	609,2	47,8	62	1042,5	503,2	165,0	503,2	0,83	0,48
Q9UXU7	1017	114	6,9	V-type ATP synthase alpha chain	22,2	46	649,0	6,2	9	140,1	591,6	215,0	591,6	0,91	#N/A

Les protéines détectées sont identifiées par leur référence UniprotKB. Pour chaque protéine, la couverture (Σcov), le nombre de peptides identifiés (Σpep), et le score Mascot (Mnuc-/+) des expériences nuc- et nuc+ sont présentés. Les scores de chaque protéine dans les témoins indépendants (MInd), relatifs (MRel) et les scores résultants SR sont également présentés. Le SR de PCNA est, dans cette expérience, artificiellement diminué par la détection d'une contamination dans le témoin relatif. Cette contamination provient d'une analyse précédente.

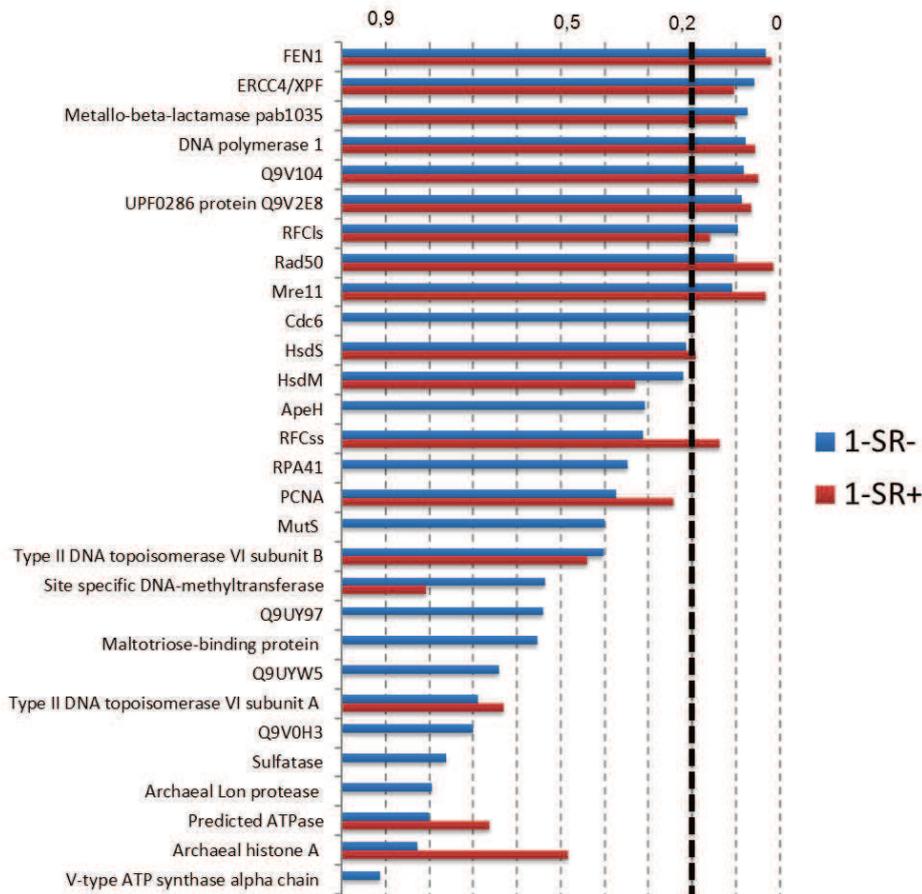


Figure 48 : Scores Mascot de l'AP-MS PCNA.

Les valeurs présentées correspondent aux scores SR de chaque protéine. Plus les barres sont proches de zéro, plus le signal est fort. La barre pointillée en gras représente le seuil qui permet de considérer une interaction comme étant de haute confiance.

Les complexes de la maintenance génomique entretiennent une relation privilégiée avec les acides nucléiques. Dans nos expériences d'AP-MS, les complexes que nous purifions de l'extrait cellulaire sont au moins partiellement complexés avec des ADN et des ARN. En purifiant un complexe assemblé sur un fragment d'ADN nous purifions également les protéines interagissant directement avec celui-ci. Cette propriété peut augmenter considérablement la complexité des données obtenues en spectrométrie de masse. L'ajout d'une étape d'incubation avec des nucléases dans une deuxième AP-MS réalisée pour chaque protéine appât permet d'appréhender ce problème en identifiant les interactions indirectes. En présence des nucléases, les fragments d'acides nucléiques exposés seront dégradés et les protéines les liants éliminées. Ainsi, chaque interaction de haute confiance a résisté à une dégradation enzymatique des acides nucléiques, ce qui nous permet de penser que la majorité est directe, ou bien qu'ils sont fixés de manière très stable sur l'ADN/ARN, empêchant ainsi une action efficace des nucléases.

En revanche, il existe des complexes dont l'assemblage et la stabilité est dépendante de la présence d'un substrat nucléique. Ainsi on peut penser que certaines interactions biologiquement significatives sont abolies par l'effet des nucléases.

5.8 PARTENAIRES SPECIFIQUES DE 23 PROTEINES DE LA MAINTENANCE GENOMIQUE

Les conditions expérimentales d'AP-MS déterminées avec la protéine PCNA ont été appliquées aux autres protéines dont nous possédons une version recombinante tagguée. Les expériences de recherche de partenaires ont été entreprises pour un total de 29 appâts. En effet, en plus des 27 protéines que j'ai purifiées moi-même, j'ai également réalisé, dans le cadre de collaborations et donc annexes à mon sujet de thèse, des purifications d'affinité pour une ARN ligase et une protéine de la famille des B Casp en collaboration avec Beatrice Clouet d'Orval (Laboratoire de Microbiologie et Génétique Moléculaires) et Hubert Becker (Laboratoire d'Optique et Bioscience). Le contenu des résultats obtenus ne sera pas discuté dans ce manuscrit puisque les travaux sont toujours en cours, mais on peut dire que les données obtenues se sont révélées informatives et cohérentes.

Pour identifier les interactions les plus significatives, les résultats des purifications ont été analysés et filtrés en conservant le seuil de SR_{nuc-} et SR_{nuc+} à 0,2. Les résultats sont présentés en Tableau 9 ci-après. Pour chaque protéine appât sont présentées les protéines proies ayant des scores $SR \leq 0,2$.

Tableau 9 : Résultats des AP-MS, interactions haute confiance.

Protéine appât	Proie	Description	# AAs	MW	pI	SR-	SR+
DNA ligase	POCL75	DNA ligase	559	64	7,5	0,00	0,00
	Q9UZ86	Reverse gyrase	1214	140	8,3	0,09	0,12
	Q9V114	RNAP rpoA1	907	103	7,7	0,12	0,15
DNA polymerase B	POCL76	DNA polymerase B	771	89	8,4	0,00	0,01
MCM	Q9UYR7	MCM	1112	126	5,8	0,01	0,01
	Q9UZ12	ERCC2/XPD/rad3 DNA repair helicase	637	73	8,0	0,04	0,05
GINS23	Q9UYR8	GINS23	168	20	8,7	0,04	0,04
	Q9V0Z9	Atp-binding transport protein	244	27	8,7	0,20	0,12
	Q9UYR2	Q9UYR2	264	30	5,8	0,19	0,16
PCNA	Q9V0P9	FEN1	343	39	8,9	0,03	0,02

	Q9V1Z5	ERCC4/XPF	752	86	9,5	0,06	0,11
	Q9UYE0	Metallo-beta-lactamase pab1035	516	60	6,3	0,07	0,10
	POCL76	DNA polymerase 1	771	89	8,4	0,08	0,06
	Q9V104	Q9V104	417	48	7,6	0,08	0,05
	Q9V2E8	UPF0286 protein Q9V2E8	251	29	8,0	0,09	0,07
	Q9V2G3	RF-Cls	479	55	9,2	0,10	0,16
	Q9UZC8	Rad50	880	104	7,0	0,11	0,02
	Q9UZC9	Mre11	423	49	6,3	0,11	0,03
GINS51	Q9UYX9	GINS51	185	21	5,1	0,00	0,00
	Q9V2F3	DNA polymerase II ss	619	69	5,0	0,07	0,11
Alka	Q9UZ73	Alka 3-methyladenine DNA glycosylase	276	32	9,1	0,00	0,00
	Q9V113	RNAP rpoA2	397	45	6,0	0,01	0,01
	Q9UZM4	Lhr-2 large helicase-related protein	867	100	8,9	0,02	0,01
	Q9V1X9	HsdS	427	49	8,8	0,02	0,01
	Q9V1Y1	HsdR	1104	131	7,7	0,02	0,04
	Q9V2I6	Fructose-bisphosphate aldolase class 1	281	31	7,1	0,02	0,03
	Q9V1Z1	RPA32	272	32	5,2	0,02	0,07
	Q9V051	Putative ATPase of the AAA superfamily	436	50	9,2	0,03	0,03
	Q9V0V0	Glycerol-1-phosphate dehydrogenase	346	37	9,5	0,05	0,05

Q9V2F9	RNA-binding protein AU-1	469	54	6,8	0,05	0,04
Q9V2B6	Q9V2B6	226	26	8,7	0,05	0,05
Q9UZ86	Reverse gyrase	1214	140	8,3	0,05	0,05
Q9V114	RNAP rpoA1	907	103	7,7	0,05	0,05
Q9V076	Metallo-beta-lactamase pab1751	451	50	6,6	0,06	0,07
Q9V2K6	Q9V2K6	148	17	9,6	0,06	0,07
Q9UZB6	DNA helicase, putative	653	75	9,0	0,06	0,06
Q9UYS8	DNA topoisomerase 1	685	78	9,6	0,07	0,05
Q9UY25	Endonuclease IV	281	32	8,4	0,08	0,11
Q9V1Y9	RPA41	375	43	5,0	0,09	0,13
Q9V115	RNAP rpoB	1117	127	6,8	0,09	0,08
Q9UY23	RNAP rpoE1	188	21	8,2	0,10	0,13
Q9UZP0	RNAP rpoF	121	14	5,2	0,10	0,11
Q9UZ60	minD/MRP superfamily ATPase	257	29	5,4	0,10	0,14
Q9UXY5	Histone acetyltransferase Elp3 homolog	589	69	9,2	0,10	0,09
Q9V2J5	MoaB-1	199	23	9,0	0,10	0,07
Q9V0W6	Endonuclease III	222	26	9,2	0,11	0,12
Q9V1X8	HsdM	623	73	5,7	0,12	0,10
Q9V291	DNA primase P46	393	45	7,2	0,13	0,12

	Q9UY16	Q9UY16	150	17	9,2	0,13	0,14
	Q9V2D7	Q9V2D7	158	18	9,5	0,14	0,10
	Q9UZD5	L-tyrosine decarboxylase	384	43	8,6	0,17	0,18
Reverse gyrase	Q9UZ86	Reverse gyrase	1214	140	8,3	0,01	0,01
	Q9V113	RNAP rpoA2	397	45	6,0	0,07	0,08
Mre11/Rad50 cplx	Q9UZC8	Rad50	880	104	7,0	0,01	0,01
	Q9UYX8	PCNA	249	28	4,6	0,02	0,10
	Q9UZC9	Mre11	423	49	6,3	0,04	0,06
NAD-dpd deacetylase	Q9UZE7	NAD-dpd deacetylase	250	28	8,1	0,02	0,02
	Q9UYR4	Dehydrogenase	392	44	6,4	0,10	0,12
OGG1 homolog	Q9V1Z0	RPA14	117	14	4,9	0,00	0,00
	Q9UY07	Nucleotidyltransferase	419	46	6,1	0,00	0,00
	Q9UZY0	N-glycosylase/DNA lyase/OGG1	239	28	9,4	0,01	0,01
	Q9UZM4	Lhr-2 large helicase-related protein	867	100	8,9	0,01	0,01
	Q9V1Z1	RPA32	272	32	5,2	0,01	0,01
	Q9UYS8	DNA topoisomerase 1	685	78	9,6	0,01	0,01
	Q9V0H2	Lhr-1 large helicase-related protein	912	105	9,2	0,02	0,02
	Q9V1Y9	RPA41	375	43	5,0	0,02	0,02
	Q9UZ86	Reverse gyrase	1214	140	8,3	0,02	0,02

	Q9V113	RNAP rpoA2	397	45	6,0	0,03	0,03
	Q9V0W6	Endonuclease III	222	26	9,2	0,07	0,10
	Q9V2J5	MoaB-1	199	23	9,0	0,07	0,08
	Q9V292	DNA primase P41	345	40	8,7	0,07	0,08
	Q9V227	GptA purine phosphoribosyltransferase	215	25	5,2	0,08	0,07
	Q9UZP0	RNAP rpoF	121	14	5,2	0,08	0,10
	Q9UY23	RNAP rpoE1	188	21	8,2	0,09	0,10
	Q9V1N7	Methylated-DNA--protein-cysteine methyltransferase	172	20	8,9	0,09	0,11
	Q9UZX1	Q9UZX1	256	29	9,4	0,10	0,10
	Q9V0F2	Phosphoesterase, putative	229	26	6,7	0,10	0,10
	Q9V2M4	RNAP rpoL	95	11	6,8	0,11	0,12
	Q9V0B1	Q9V0B1	153	18	9,8	0,12	0,10
	Q9V114	RNAP rpoA1	907	103	7,7	0,13	0,10
	Q9V116	RNAP rpoH	82	9	9,2	0,14	0,13
	Q9UZW9	Q9UZW9	125	15	6,2	0,16	0,21
	Q9V115	RNAP rpoB	1117	127	6,8	0,16	0,13
	Q9V2B6	Q9V2B6	226	26	8,7	0,17	0,16
PHP domain protein	Q9V0F9	Putative phosphotransferase, php family	216	25	5,3	0,00	0,00
	Q9UZT4	Q9UZT4	259	28	7,4	0,20	0,20

FEN1	Q9V0P9	FEN1	343	39	8,9	0,00	0,00
	Q9UYE0	Metallo-beta-lactamase pab1035	516	60	6,3	0,09	0,06
	Q9UYX8	PCNA	249	28	4,6	0,11	0,11
RNaseH2	Q9V1A9	Ribonuclease HII	224	25	9,0	0,00	0,00
	Q9V2M1	ATP-dependent RNA helicase, putative	855	98	9,0	0,16	0,18
RPA	Q9UY07	Nucleotidyltransferase	419	46	6,1	0,00	0,00
	Q9V1Z1	RPA32	272	32	5,2	0,00	0,00
	Q9V1Y9	RPA41	375	43	5,0	0,01	0,00
	Q9V1Z0	RPA14	117	14	4,9	0,01	0,00
	Q9V1Z5	ERCC4/XPF	752	86	9,5	0,04	0,08
	Q9V113	RNAP rpoA2	397	45	6,0	0,05	0,08
	Q9UYS8	DNA topoisomerase 1	685	78	9,6	0,06	0,07
	Q9UZM4	Lhr-2 large helicase-related protein	867	100	8,9	0,06	0,05
	Q9UZ86	Reverse gyrase	1214	140	8,3	0,06	0,14
	Q9V114	RNAP rpoA1	907	103	7,7	0,06	0,10
	Q9V0H2	Lhr-1 large helicase-related protein	912	105	9,2	0,06	0,09
	Q9V2G5	DNA2	1308	152	6,8	0,07	0,10
	Q9UYC4	ERCC3/XPB	453	52	8,9	0,09	0,13
	Q9V1X8	HsdM	623	73	5,7	0,10	0,11

RadA	Q9V233	radA	356	39	7,0	0,01	0,01
	Q9V076	Metallo-beta-lactamase pab1751	451	50	6,6	0,09	0,14
	POCL75	DNA ligase	559	64	7,5	0,12	0,09
Rad25like	Q9V278	DNA repair protein RAD25	446	51	9,3	0,00	0,00
	Q9V113	RNAP rpoA2	397	45	6,0	0,02	0,01
	Q9V1Z1	RPA32	272	32	5,2	0,05	0,13
	Q9V2F9	RNA-binding protein AU-1	469	54	6,8	0,06	0,08
	Q9V114	RNAP rpoA1	907	103	7,7	0,07	0,07
	Q9UZP0	RNAP rpoF	121	14	5,2	0,10	0,16
	Q9V2B6	Q9V2B6	226	26	8,7	0,10	0,09
	Q9UY96	Flavoprotein monooxygenase, putative	370	41	9,5	0,12	0,14
	Q9UZ86	Reverse gyrase	1214	140	8,3	0,16	0,12
	Q9V115	RNAP rpoB	1117	127	6,8	0,16	0,17
	Q9UY23	RNAP rpoE1	188	21	8,2	0,16	0,20
	Q9UZM4	Lhr-2 large helicase-related protein	867	100	8,9	0,17	0,04
P46	Q9V291	DNA primase P46	393	45	7,2	0,00	0,00
	Q9V233	radA	356	39	7,0	0,03	0,03
P41	Q9V233	radA	356	39	7,0	0,01	0,01
	Q9V292	DNA primase P41	345	40	8,7	0,04	0,02

	P0CL75	DNA ligase	559	64	7,5	0,10	0,08
NucS	Q9V2E8	NucS	251	29	8,0	0,00	0,00
	Q9UYX8	PCNA	249	28	4,6	0,00	0,00
	Q9V1Z5	ERCC4/XPF	752	86	9,5	0,03	0,04
	Q9UYE0	Metallo-beta-lactamase pab1035	516	60	6,3	0,03	0,03
	Q9UZ60	minD/MRP superfamily ATPase	257	29	5,4	0,04	0,15
	Q9V076	Metallo-beta-lactamase pab1751	451	50	6,6	0,05	0,08
	Q9V2F8	Q9V2F8	128	15	9,5	0,06	0,11
	Q9V113	RNAP rpoA2	397	45	6,0	0,07	0,12
	Q9V2G3	RF-Cls	479	55	9,2	0,07	0,12
	Q9UYC4	ERCC3/XPB	453	52	8,9	0,08	0,15
	Q9V1X9	HsdS	427	49	8,8	0,09	0,13
	Q9UXY5	Histone acetyltransferase Elp3 homolog	589	69	9,2	0,09	0,11
	Q9V1F5	Archaeal histone B	67	7	9,8	0,09	0,08
	Q9V0B3	Adenine-specific DNA methylase	1010	116	9,1	0,10	0,19
	Q9V2G4	RF-Css	1437	165	8,8	0,11	0,11
	Q9UZY0	N-glycosylase/DNA lyase/OGG1	239	28	9,4	0,12	0,19
	Q9UYC3	BAX1/Uncharacterized ArCR	379	44	9,3	0,13	0,13
	Q9UYB0	ApeH acylamino-acid-releasing enzyme	631	73	9,1	0,13	0,12

	Q9V135	Type II DNA topoisomerase VI subunit B	564	64	9,1	0,14	0,17
	Q9UZT4	Q9UZT4	259	28	7,4	0,15	0,18
	P61881	Archaeal histone A	67	7	9,1	0,15	0,15
	Q9V2K6	Q9V2K6	148	17	9,6	0,18	0,12
DNA polymerase D	Q9V2F4	DNA polymerase II ls	1455	166	6,4	0,00	0,00
	Q9V2F3	DNA polymerase II ss	619	69	5,0	0,00	0,00
	Q9V1Z1	RPA32	272	32	5,2	0,04	0,09
	Q9V291	DNA primase P46	393	45	7,2	0,06	0,13
	Q9UY07	Nucleotidyltransferase	419	46	6,1	0,07	0,16
	Q9V1Y9	RPA41	375	43	5,0	0,10	0,19
RadB	Q9V2F6	radB	239	27	9,4	0,01	0,01
	Q9UYR4	Dehydrogenase	392	44	6,4	0,13	0,20
	Q9UZT4	Q9UZT4	259	28	7,4	0,14	0,12
	P0CL75	DNA ligase	559	64	7,5	0,20	0,17
DNA2dom2	Q9V2G5	DNA2	1308	152	6,8	0,01	0,00
	Q9UY85	RecJ-like phosphoesterase	740	83	6,4	0,04	0,02
	Q9V1Z1	RPA32	272	32	5,2	0,04	0,02
	Q9UYI3	GTP-binding protein	397	44	6,4	0,04	0,03
	Q9V0E5	RNA-binding protein, containing THUMP domain	393	45	8,9	0,08	0,17

	Q9UZB6	DNA helicase, putative	653	75	9,0	0,09	0,10
	Q9V2F9	RNA-binding protein AU-1	469	54	6,8	0,10	0,09
	Q9V1V1	50S ribosomal protein L6P	184	21	9,3	0,11	0,18
	Q9UZ86	Reverse gyrase	1214	140	8,3	0,14	0,12
	Q8J2X9	Q8J2X9	234	28	5,6	0,14	0,06
RecJ-like ss exo	Q9V2K8	RecJ-like single-stranded exonuclease	481	54	6,1	0,00	0,00
	Q9V256	Q9V256	269	31	4,7	0,04	0,20

Après filtrage des données, ce set de protéines constitue l'ensemble des associations que nous qualifions de "haute confiance". Nous n'avons retenu que le score SR pour constituer cet ensemble au contraire des études précédentes qui généralement effectuent un tri des associations en prenant également en compte la catégorie fonctionnelle des protéines potentiellement associées (Li, Santangelo et al. 2010). En effet ces auteurs considèrent que si deux protéines sont associées dans un réseau et appartiennent à des catégories fonctionnelles trop distinctes, cette association est probablement non-spécifique. Pour notre part, nous n'avons pas voulu utiliser ce critère qui nous semble trop subjectif et qui potentiellement peut conduire à l'élimination de l'analyse d'associations réelles dont la signification biologique nous échappe, dans l'état actuelle des connaissances. Nous avons pu nous permettre de ne pas utiliser ce critère de catégorisation fonctionnelle puisque posséder une liste exhaustive des protéines non-spécifiques et avoir effectué deux analyses LC-MS/MS par appâts nous a conduits à pouvoir éliminer en grande partie les signaux non-spécifiques des résultats finaux. Bien sur, je suis conscient qu'il reste encore des interactions non-spécifiques dans le set de haute confiance, cette présence de bruit de fond est inhérente à l'approche AP-MS utilisée (von Mering, Krause et al. 2002). Cependant je pense que nous avons considérablement réduit la part de ce signal non informatif par rapport aux interactions spécifiques. Comme le montre le tableau 8, la somme d'informations obtenues par cette approche est très importante (23 protéines appâts, 150 interactions). Pour les analyser sous la forme d'un réseau d'interactions, nous avons établi une collaboration avec Jacques Nicolas (IRISA, Univ. Rennes I) afin de construire le réseau, d'en analyser les caractères topologiques principaux et de tenter de corrélérer la modularité de ce réseau avec les connaissances actuelles sur les voies métaboliques de l'ADN chez les Archaea. Jacques Nicolas a développé plusieurs algorithmes capables de déterminer un grand nombre de paramètres du réseau. J'ai ensuite effectué deux séjours dans son laboratoire afin d'utiliser les algorithmes d'analyse topologique, sous sa direction.

5.9 ANALYSE TOPOLOGIQUE DU RESEAU D'INTERACTION

La quantité et la complexité des données recueillies, cela même avec un seuil de significativité élevé est difficilement interprétable en l'état (Tableau 7). Afin d'avoir une meilleure compréhension des données, les interactions identifiées avec le plus haut degré de confiance ont été compilées dans un réseau d'interactions (Figure 50A). Dans ce réseau les protéines sont représentées par les nœuds et les interactions par des arrêtes reliant chaque nœud suivant un modèle de spoke. Dans ce modèle on considère que chaque partenaire détecté significativement interagit directement avec l'appât. Le réseau représente ainsi l'ensemble des complexes protéiques purifiés. Le réseau est composé de 93 nœuds, dont environ 1/4 sont des protéines ou des complexes appâts, et de 136 interactions (au sein du réseau, les sous unités de complexes sont rassemblées en une seule entité, ce qui diminue le nombre d'interactions totales, Figure 50A). Le réseau présente un taux de clusterisation de 10,3% corrigé à 11,1% lorsque l'on prend en compte le biais de notre modèle où les protéines uniquement proies ne peuvent pas posséder d'interactions entre elles. Le réseau est constitué d'une entité principale contenant 81 nœuds et de 4 autres structures plus petites (1 fois 3 et 3 fois 2 protéines) qui ne lui sont pas connectées. Les protéines de la maintenance génomique représentent 43% des protéines alors que 19% n'ont pas de fonctions identifiées. Nous observons que 22% sont associées à la transcription, la maturation et la traduction des ARN, 7% à des fonctions de modification/restriction. Seulement 9% des protéines possèdent une ontologie associée à d'autres métabolismes. Ainsi, 93% des protéines du réseau sont liées au métabolismes des acides nucléiques ou ne sont pas identifiées.

La distribution des degrés suit une loi de puissance à l'asymptote (Figure 50B). Aussi, la distribution des coefficients de clusterisation $C(k)$ semble tendre vers une loi de puissance signifiant un certain degré de hiérarchisation dans le réseau. Notre réseau est donc de type scale free avec, à priori, un certain degré de hiérarchisation. Ces propriétés sont typiques d'une organisation modulaire, ainsi dans notre réseau, il existe probablement des modules de nœuds identifiables.

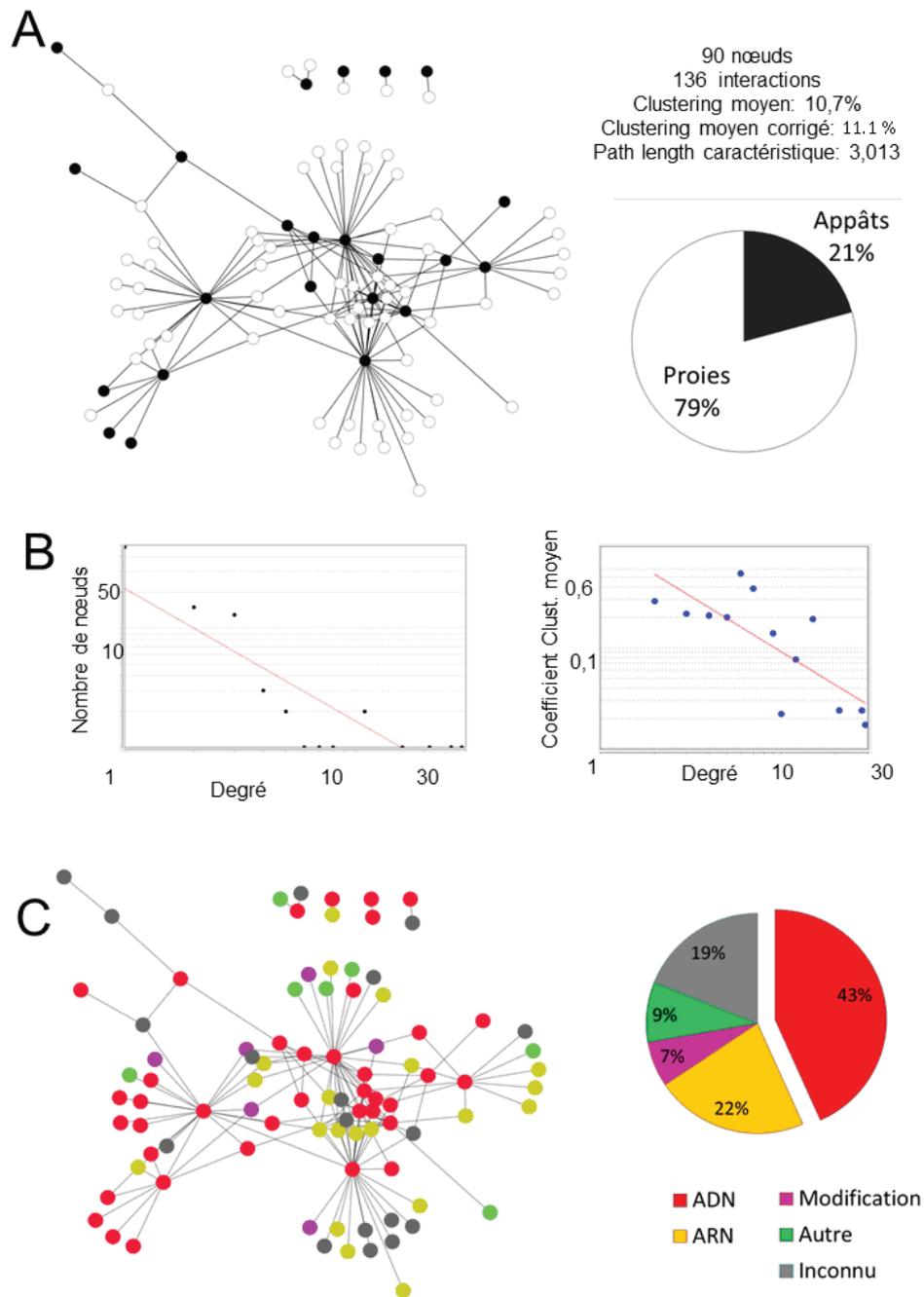


Figure 49 : Paramètres généraux du réseau.

A) Structure et paramètres simples du réseau. Les protéines appâts sont signalées en noir. B) Distribution des degrés et des coefficients de clusterisations. Les axes x et y sont représentés en échelle logarithmique. Une loi de puissance est appliquée chacune des représentations. C) Répartition en termes de métabolismes dans le réseau. 43% des protéines sont affiliées à la maintenance génomique.

Nous avons réalisé une analyse du réseau afin d'en comprendre la structure et d'en identifier les éventuels modules pouvant correspondre à des fonctions biologiques. En étudiant la topologie du réseau nous avons identifié ses propriétés essentielles. L'analyse topologique du réseau a consisté à en déterminer les zones de fort recoupement (« clusters »), les protéines ayant un nombre de partenaires communs « élevés » (« groups »)

ainsi que le squelette du réseau (« backbone ») basé sur le calcul de l'arbre des "longueurs de chemins" (path-length) optimaux. Cette analyse nous a permis d'identifier, en son sein, les complexes protéiques les plus significatifs. J'ai tout d'abord identifié au sein du réseau les clusters optimaux d'au moins 3 protéines possédant un coefficient de clusterisation supérieur à 10% (Figure 50). J'ai identifié 4 clusters autour des protéines PCNA, RPA, l'ADN primase p41 et la sous unité A2 de la RNA polymérase. Ces structures possèdent un coefficient de clusterisation élevé ce qui indique un recouplement des informations obtenues en AP-MS. Les protéines ayant un degré supérieur à 1 et un coefficient de clusterisation supérieur à 5% sont présentées en figure 50B. Au sein des clusters optimaux identifiés, près de 80 % des protéines appartiennent à la maintenance génomique (« ADN » figure 50C). On observe que le cluster RNAP rpoA2 est, en fait, majoritairement dominé par le cluster RPA. Aussi la présence de la sous unité A1 de la RNA polymérase dans le cluster RPA suggère que ces deux clusters ne seraient en fait qu'une seule structure partagée entre la RNA polymérase et le PCNA.

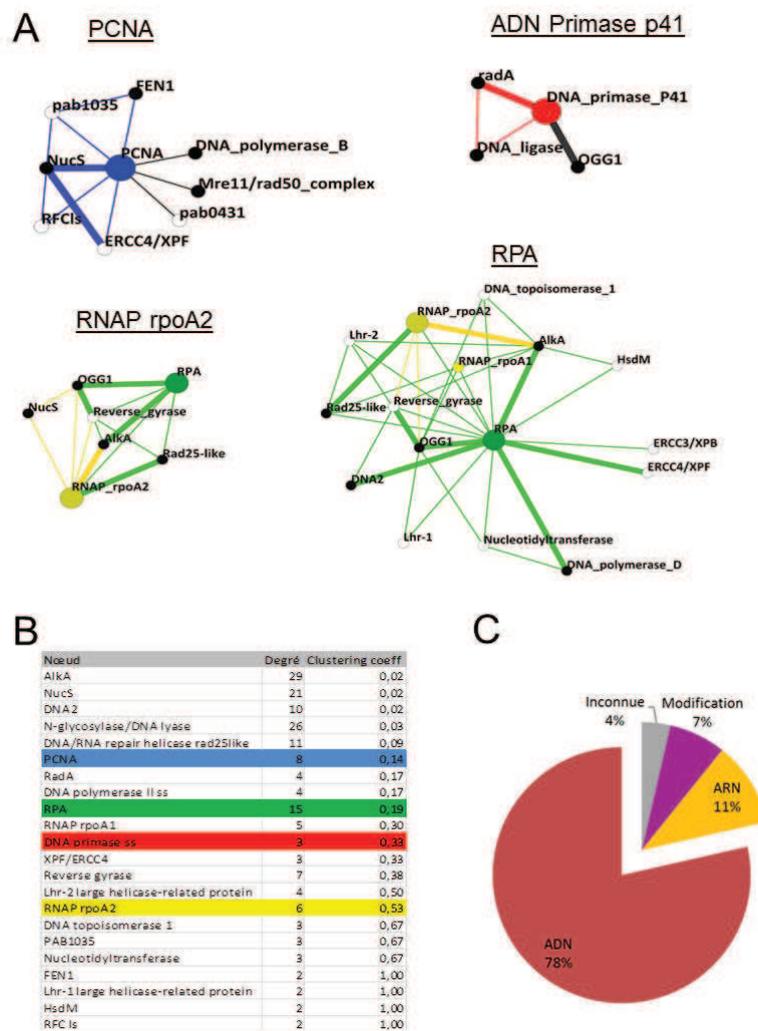


Figure 50: Clusters au sein du réseau haute confiance.

A) Clusters optimaux identifiés dans le réseau. B) Degré et Coefficient de clusterisation des nœuds du réseau. C) Répartition en termes de métabolismes au sein des clusters. Les protéines attribuées à la maintenance génomique représentent 78% des nœuds. Les clusters sont enrichis en protéines de la maintenance génomique.

Au lieu de considérer le voisinage d'un nœud (protéine), on peut considérer l'environnement des arêtes (interactions). L'indice de « clustering coefficient mutuel » dans un graphe est un coefficient de clusterisation sur des paires de nœuds. Il calcule la probabilité (p-value) d'observer autant d'interactions entre deux protéines étant donné le nombre total d'interactions et le nombre d'interactions de chacune. L'interaction possible entre les deux protéines observées n'est pas prise en compte. Ce coefficient peut être utilisé comme une statistique de test de la probabilité d'une interaction entre deux protéines. Il peut donc être utilisé aussi bien pour la validation des interactions observées ainsi que pour la sélection de protéines non observées en interaction, mais probablement liées. J'ai utilisé l'algorithme « mutualcoeff » pour identifier les groupes optimaux de 2 à 3 protéines ayant le coefficient mutuel le plus élevé (Figure 51). Cette recherche correspond à la recherche de groupes de protéines ayant une forte cooccurrence. Les groupes identifiés sont les suivants : RPA et Rad25-like ; Fen1 et NucS ; RNAP rpoa1 et RNAP rpoa2 ; Alka et OGG1 ; RadA, RadB et DNA2 ; DNA primase p46 et DNA primase p41 ; MoaB1 avec RNAP rpoF et Q9V2B6 ; PCNA avec HsdS. Plusieurs de ces groupes correspondent à des sous unités de complexes multimériques (DNA Primase, RNA polymérase), à des protéines fonctionnellement associées (RadA/RadB, ou Alka /OGG1), ou encore, ont un partenaire commun connu (NucS/Fen1 avec PCNA). Les groupes de protéines identifiées sont différents des clusters mais les groupes identifiés sont cohérents. Les paires Alka : OGG1, RpoA1 :RpoA2, RPA rad25-like et Fen1 :NucS appartiennent au même cluster. Aussi, les paires RadA :RadB et p41 :p46 se chevauchent avec le cluster p41. Ces données suggèrent que ces 4 protéines constituent un module .

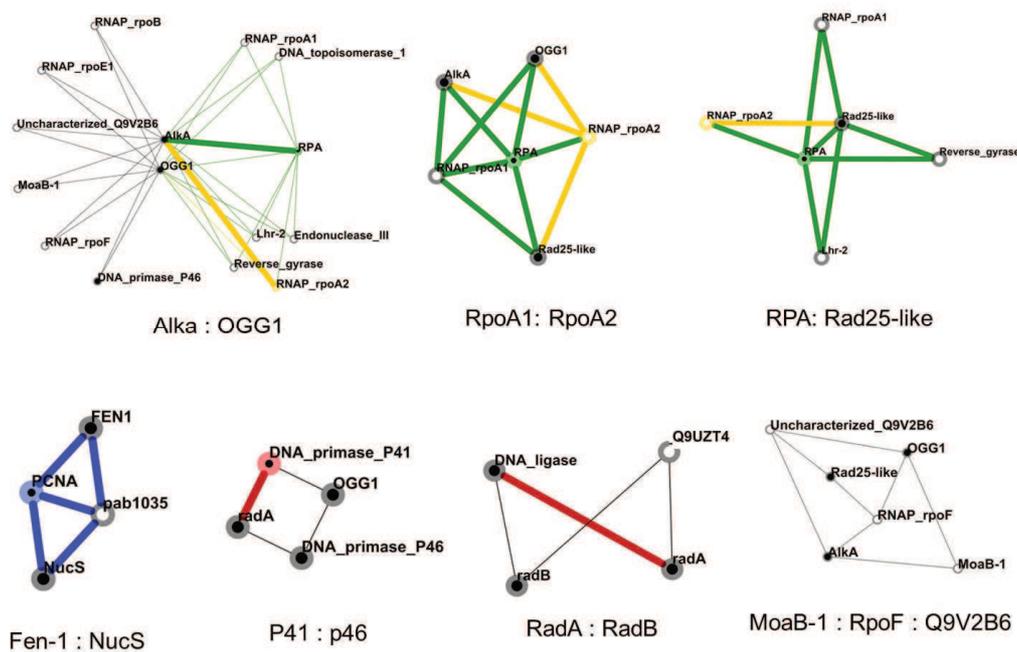


Figure 51 : Groupes de protéines optimaux sur la base du coefficient mutuel.

Les différents groupes de protéines sont présentés avec leurs partenaires (protéine 1 : protéine 2 : protéine 3). Les couleurs des interactions correspondent aux clusters présentés précédemment.

Au lieu d'identifier des sous graphes qui constitueraient des modules ou leurs motifs, on peut réduire l'ensemble du réseau à un jeu minimal d'interactions et ainsi identifier les relations essentielles dans le réseau. Pour ce faire, j'ai identifié l'arbre de recouvrement minimal du jeu d'interactions dominantes grâce à l'algorithme « backbone ».

Un jeu dominant est un ensemble de nœuds connectés (il existe un chemin d'interactions entre n'importe quelle paire de protéines) de telle sorte que chaque autre nœud du réseau soit en interaction directe avec au moins un de ses éléments. Ainsi, en identifiant le jeu dominant du réseau, on réalise une bipartition des nœuds selon leur appartenance, ou pas, au cœur du réseau en lieu et place de la bipartition appâts / proies inhérente à l'expérience d'AP-MS. C'est au sein de ce jeu dominant que j'ai identifié l'arbre minimum de recouvrement. Un arbre de recouvrement minimum (Minimum Spanning Tree) est l'ensemble minimal d'interactions connectant l'intégralité des nœuds. Cet arbre est construit sur la base du chemin le plus court (« shortest path length »), il ne contient pas de relation transitive, ainsi, il n'existe qu'un seul chemin pour aller d'un nœud à un autre. Dans chaque réseau, il peut exister une pluralité d'arbres de recouvrement. En revanche, il est possible de déterminer un arbre optimal selon les paramètres des interactions. En utilisant l'algorithme « backbone », j'ai identifié l'arbre minimal de recouvrement du réseau en utilisant le score SR comme paramètre d'optimisation. Ainsi, j'ai identifié les interactions essentielles qui constituent l'épine dorsale du réseau d'interactions (Figure 52). Cette épine dorsale est constituée de 16 protéines et 15 interactions dont la somme des scores SR- est de 0,73. Selon ces critères, ce backbone peut être interprété comme les interactions les plus profondes et les plus fiables du réseau. C'est au sein de cet arbre que la probabilité d'une interaction directe est la plus forte.

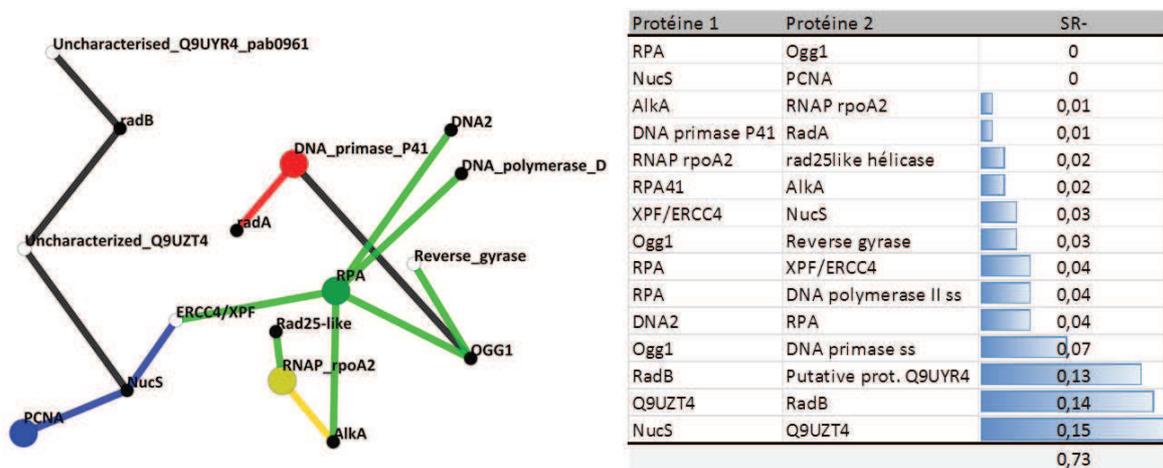


Figure 52: Epine dorsale du réseau haute confiance.

A gauche, l'arbre de recouvrement minimal du réseau. Les couleurs correspondent aux clusters décrits précédemment. A droite, le tableau récapitule les valeurs du SR- des interactions de l'arbre.

La présentation des résultats des analyses AP/MS au sein d'un réseau et les interprétations des fonctions biologiques de certaines associations détectées sont décrites sous forme d'un article dans le prochain chapitre.

5.10 PRESENTATION DE L'ARTICLE

Bien que n'ayant pas produit le nombre d'articles suffisants pour présenter un manuscrit de thèse sur articles, nous avons décidé de passer plus de temps sur une mise en forme de ces résultats qui soit directement publiable, c'est pourquoi je l'inclue, tel quel, dans la partie Résultats et Discussions plutôt qu'en Annexe.

Au début du dernier semestre de mon contrat de thèse, j'ai participé à un congrès international sur la biologie moléculaire des Archaea (Marburg, 2-4 Juin 2012), pendant lequel j'ai présenté une grande partie de mes résultats. A cette occasion, nous avons appris qu'une des interactions que j'ai détecté dans le réseau et qui concerne RadB et une nouvelle protéine, Pab0961, potentiellement impliquée dans la régulation de la réplication, a également été détectée par l'équipe Génétique et Génomique du laboratoire de Thorsten Allers (Université de Nottingham). Ces derniers ont, de plus, effectué des expériences génétiques sur l'homologue de Pab0961 chez *Haloferax volcanii*, et les résultats sont cohérents avec l'identification des partenaires de cette protéine chez *P. abyssi*. Nos deux études sont donc complémentaires et se soutiennent l'une l'autre. Cependant ces deux études divergent trop dans les approches et les attendues, ce qui rend impossible une publication commune. Nous avons donc décidé de déposer nos deux manuscrits en même temps, pour proposer une parution conjointe dans un même journal. Thorsten Allers met actuellement en forme les résultats de son analyse et nous avons convenu que les papiers seront soumis fin Janvier à la revue Nucleic Acids Research.

Les points principaux décrits dans cet article concernent la première description d'un réseau d'interactions impliqué dans la maintenance de l'intégrité génomique d'une Archaea hyperthermophile. Nous démontrons que ce réseau est consistant avec ce qui est déjà connu dans la littérature et qu'il est donc à même de révéler de nouveaux aspects de la maintenance génomique. Certaines interactions ont été confirmées par des approches fonctionnelles ou d'interactions physiques complémentaires. D'autre part, ce réseau révèle également des interactions non suspectées entre des homologues de protéines Eucaryotes, ce qui renforce l'idée que les Archaea constituent un bon modèle pour l'étude des mécanismes informationnels. D'un point de vue plus général, les données de cet article suggèrent qu'il existe chez les Archaea une relation étroite et contrôlée entre les composants du réplisome et les protéines de la recombinaison de l'ADN au niveau de la fourche de réplication. Il est possible que la coordination de ces activités soit une caractéristique des Archaea se développant dans des conditions extrêmes de température.

[Depuis la rédaction de l'article, certaines protéines ont été exclues du réseau, quelques ajustements d'attribution d'ontologie ont été réalisés ainsi certaines données peuvent légèrement différer de celles du manuscrit, sans conséquences majeures pour l'analyse que nous présentons.]

An extended network of genomic maintenance in the archaeon *Pyrococcus abyssi* highlights unexpected associations between eucaryotic homologs

Pierre François Pluchon^{1,2,3}, Thomas Fouqueau⁴, Christophe Crezé^{1,2,3}, Sébastien Laurent^{1,2,3}, Julien Briffotiaux^{1,2,3,§}, Gaëlle Hogrel^{1,2,3}, Adeline Palud^{1,2,3}, Winfried Hausner⁴, Michael Thomm⁴, Jacques Nicolas⁵ and Didier Flament^{1,2,3,*}

¹Ifremer, UMR6197, Laboratoire de Microbiologie des Environnements Extrêmes, 29280 Plouzané, France.

²Université de Bretagne Occidentale, UMR6197, Laboratoire de Microbiologie des Environnements Extrêmes, 29280 Plouzané, France.

³CNRS, UMR6197, Laboratoire de Microbiologie des Environnements Extrêmes, 29280 Plouzané, France.

⁴Lehrstuhl für Mikrobiologie, Universität Regensburg, Universitätsstrasse 31, 93053 Regensburg, Germany.

⁵IRISA-INRIA, Campus de Beaulieu, 35042 Rennes cedex, France

* To whom correspondence should be addressed. Tel: 0033 (0)2 98 22 45 27 ; Email: dflament@ifremer.fr

[§]Present Address: Julien Briffotiaux, Laboratoire d'Optique et Biosciences, CNRS, Ecole Polytechnique, 91128 PALAISEAU, France.

ABSTRACT

A protein-interaction network centred on informational processes of *Pyrococcus abyssi* was generated by affinity purification coupled with mass spectrometry. The network consists of 147 interactions linking 93 proteins. These interactions give insights into the connections of DNA replication with recombination and repair. They also lead to the discovery of new archaeal components and of associations between eucaryotic homologs. Thus, our interaction map provides a valuable tool to reveal aspects of genome integrity both in Archaea and Eucaryotes.

INTRODUCTION

The control of genome integrity is of pivotal importance for the cell. Failures in maintaining genome stability result in the accumulation of mutations, genome rearrangements and cell death. Molecular interactions involved in DNA repair and recombination and their coordination with DNA synthesis are required for genome maintenance. These genome transactions and their proper

coordination require assembling multi-protein machines at specific sites at the chromosome or chromatin level. The large variety of lesions that can arise in DNA and the biological importance of coping with these are reflected by the diversity of biochemical strategies for DNA repair found in model micro-organisms. In this context, hyperthermophilic Archaea (HA) are of special interest as DNA in hyperthermophiles is exposed to temperatures that increase the rate of its decomposition reactions (Lindahl 1993). However, in the hyperthermophilic euryarchaeon *P. abyssi*, the level of apurinic/apyrimidinic sites was shown to be only 10-fold higher than in *E. coli* and the level of spontaneous mutation in HA has been shown to be close to the standard rate for DNA-based microbes (Grogan, Carver et al. 2001), suggesting active and efficient repair pathways (Palud, Villani et al. 2008). In addition to this adaptation at extreme temperature HA have a deep evolutionary divergence from well studied micro-organisms. How they maintain genomic integrity is thus an interesting question from both physiological and phylogenetic perspectives (Grogan 2004). However, very little is known concerning pathways and proteins that perform DNA repair in HA. Genomic sequence analyses suggest that most of the archaeal proteins predicted to participate in DNA replication, repair and recombination are more closely related to eucaryotic than bacterial proteins (Kelman and White 2005; Yutin, Makarova et al. 2008).

So far, proteins of base excision repair, alkyl transfer, damage reversion and trans-lesion pathways have been detected in HA (Boudsocq, Iwai et al. 2001; Sartori and Jiricny 2003; Lucas-Lledo and Lynch 2009; Perugino, Vettone et al. 2012). In contrast, genomic content analysis show that mismatch repair (MMR) machinery is absent in most archaea species. The presence of the *mutS/mutL* orthologs genes of MMR pathway is only restricted to halophiles and methanogen species and have been shown to be dispensable for the maintenance of a low mutation rate in *Halobacterium salinarum* NRC-1 (Busch and DiRuggiero 2010). In this context, the low mutation rate observed could imply that Archaea have evolved mechanisms to improve the correct insertion of nucleotide and proofreading activities during DNA replication or that a distinct MMR, which has escaped genomic analysis, operates in Archaea. Furthermore, the existence of nucleotide excision repair (NER) pathway has not been demonstrated, yet it is suggested by the presence and properties of NER nucleases and helicases in HA. However, none of the proteins responsible for the lesion recognition step has been found in archaeal genomes (Kelman and White 2005). In this context, the implication of new components or new DNA repair pathways can not be excluded. Another conspicuous feature

highlighted by genomic analysis is the relative paucity of DNA polymerases dedicated to repair functions in archaeal genomes which seems inconsistent with the lifestyle of these micro-organisms at high temperatures and raises the hypothesis that archaeal replicative DNA polymerases could have additional repair functions alone or associated with other proteins. Many aspects of DNA repair and more generally regulation of genomic maintenance thus remain enigmatic. To gain insight into these fundamental processes, genetic systems in Archaea have been developed and a wide range of genes involved in DNA replication and repair has been disrupted in order to understand their physiological functions in HA (Atomi, Imanaka et al. 2012). However, to date, very few reports applied this functional approach to the analysis of DNA repair in anaerobic hyperthermophilic archaea (Fujikane, Ishino et al. 2010; Li, Santangelo et al. 2010).

Identification of protein-protein interaction networks in organisms can complement information obtained from these functional genomics approaches. Indeed, to fulfill their biological activities in the cell, most proteins function in association with protein partners, or as part of large molecular assemblies. Hence, the knowledge of the interaction context of a protein is crucial to understand its cellular functions. Two previous studies reported the elucidation of protein-protein networks of archaeal DNA replication, using yeast two-hybrid screens or purification of stable complexes assembled *in vivo*, in *Archaeoglobus fulgidus* and *Thermococcus Kodakarensis*, respectively (Motz, Kober et al. 2002; Li, Santangelo et al. 2010). The latter study led to the important discovery and characterization of a novel GINS-associated nuclease, GAN, which might be the archaeal homologue of Cdc45 in the CMG complex (Li, Pan et al. 2011; Makarova, Koonin et al. 2012).

In this report, we delineated a protein-protein interaction network of genomic maintenance in *P. abyssi* using affinity purification *in vitro* coupled with mass spectrometry identification of protein partners (AP/MS). Compared to the purification of complex *in vivo*, this approach allows the purification of proteins that are normally not expressed under laboratory conditions. In contrast the stoichiometry of interacting proteins is not maintained and the high concentration of bait proteins used can lead to false positive identifications. To minimize this latter negative effect, we duplicated the AP/MS assay, per bait, and compared the results against five independent negative controls. In addition, we used a simple algorithm to provide clear cut results of the mass spectrometry identification. 23 HIS or GST-tagged proteins presumably implicated in DNA replication, repair and recombination were used to co-isolate protein complexes (Groisillier, Herve et al. 2010), allowing us to

propose an extended network of genome stability at high temperature in *P. abyssi*. For some of the interactions detected, we demonstrated physical associations, using surface plasmon resonance and pulldown-immunoprecipitation assays or a functional association using an enzymatic assay, as exemplified by the activation of transcription in the presence of a RPA complex. Interestingly, using this approach we revealed and confirmed unsuspected interaction between eucaryotic homologs.

MATERIAL AND METHODS

***Pyrococcus abyssi* protein extract**

P. abyssi was grown in continuous culture in a 5 L Gazlift bioreactor in MES medium under anaerobic conditions at 95°C and pH 6.4 (Godfroy, Postec et al. 2006). Growth was monitored by density measurements with a Thoma cell counting chamber (0.02 mm depth). Cells were maintained in exponential growth phase between 2 and 4 x 10⁸ cells.ml⁻¹. Culture medium was harvested on ice. Cells were pelleted at 6000 g, 45 min at 4°C and dry cells pellets were stored at -80°C. *P. abyssi* cells pellets were resuspended in two volume (w/v) of PBS buffer supplemented with imidazole; 140 mM NaCl, 2,7 mM KCl, 10 mM Na₂HPO₄, 1,76 mM KH₂PO₄, 10 mM imidazole, pH 7.4, with a mix of anti protease EDTA free (Roche). Cell lysis was completed by sonication using a Vibracell sonifier (BioBlock Scientific, 2 x 10 min of 0.5 second on/off pulses at 375 W, 40% amplitude, on ice). *P. abyssi* cells debris were eliminated by centrifugation at 15000 g, 45 min at 4°C. Soluble fractions concentrations were measured by Dc Protein Assay (Bio-Rad) with Bovine Serum Albumine as the standard and adjusted at 8 mg.ml⁻¹ and pH 7.4.

Production and purification

The production and purification of the bait proteins were either already reported (see table 1 for a listing of the protein baits used and the associated references) or is described in the supplemental materials. Depending on the clone used, expression vectors might differ, however principle was the same. Briefly; plasmid pGEX-4T-1 (GE Healthcare Life Science) was used to express protein in fusion with GST while pFO4, pET15b, pE26b+, pET30 (Novagen) and pQE80 (QiaGen) expression vectors were used to express protein in fusion with a HIS tag. The sequences of the inserts were checked and the recombinant clones were used to transform *E. coli* RosettaTM strain. *E. coli* clones were grown in 300 mL Lysogeny broth (LB) medium (Beckton, Dickinson and Co). Protein production under LacZ

promoter was induced in exponential growth phase at $OD_{600nm} = 0.6$ by 1 mM IPTG, 3h at 37°C. Cell lysis was performed by 3 x 10 min 80°C heat shocks for HIS tagged protein and by sonication for GST tagged protein, using a Vibracell sonifier (10 min of 0.5 second on/off pulses at 375 W, 40% amplitude). Cell lysates were cleared by centrifugation at 10000 g, 10 min at 4°C. Recombinant proteins were purified from soluble fraction on magnetic beads (Dynabeads® His-Tag Isolation & Pulldown or Promega MagneGST™ Pull-Down System). Two milligrams of beads, equilibrated in PBS buffer were incubated 15 min at room temperature under rotation with 4 ml of cell lysate. Purified proteins were washed 4 times with 1mL of PBS buffer. Quality of the purification and apparent size of the protein were checked on SDS-PAGE.

Affinity purification of *P. abyssi* protein complexes

Purified HIS and GST tagged proteins were used as baits in the *P. abyssi* protein extract. Depending on the tagged proteins sizes and binding properties, between 10 and 30 µg of bait proteins were immobilized on 2 mg of either cobalt or glutathione coated magnetic beads. The complex baits-beads were further incubated with 4 ml of *P. abyssi* proteins extract under rotation for 3 hours at 4°C . Protein complexes formed *in vitro* were separated on a magnet and washed extensively by 4 x 4 ml and 2 x 1 ml of PBS buffer.

A second assay was performed with an additional step of nucleotides degradation to eliminate non-specific protein interactions *via* DNA or RNA fragments. To this aim, purified complexes were incubated with 80/1.2 units of a DNase/RNase mix (GE healthcare) 45 minutes in PBS buffer at room temperature. Those two affinity purification experiments will be referred as nuc- and nuc+, respectively. Finally, a control assay was also performed, that consisted in carrying out of the same experiment except that Glutathione- and cobalt-coated magnetic beads were used instead of the complexes baits-beads. Five such independent control assays were done to identify the exhaustive listing of proteins engaged in non-specific interactions with the beads. Purified proteins complexes were eluted in 100 µL of Laemmli buffer at 95°C for 15 min and separated on SDS-PAGE during 20 min at 120V in 12% Criterion XT Precast Gels (Bio-Rad). Proteins were extracted from the gels by cutting each full protein tracks in 3 slices of the same dimension. Gel slices were then separately analysed in shotgun proteomics analysis. The procedure that consisted in separating the proteins on denaturing electrophoresis gel turned out to greatly improve the efficiency of the downstream

trypsinisation step as it has already been shown that some hyperthermostable proteins may resist proteolytic digestion at moderate temperatures (Lee, Sevinsky et al. 2009).

Protein prefractionation and digestion

The fragments of gels were first treated with 50 mM NH_4HCO_3 in acetonitrile/water 1:1 (v/v), dehydrated with 100 % acetonitrile and rehydrated in 100 mM NH_4HCO_3 . Next they were washed again with 50 mM NH_4HCO_3 in acetonitrile /water, 1:1 (v/v) and dehydrated with 100 % acetonitrile. The slices were then treated with 65 mM DTT for 15 minutes at 37°C, and with 135 mM iodoacetamide in the dark at room temperature. Finally, the samples were washed with 100 mM NH_4HCO_3 in acetonitrile/water, 1:1 (v/v), and dehydrated with 100 % acetonitrile before being rehydrated in 100 mM NH_4HCO_3 , washed with 100 mM NH_4HCO_3 in acetonitrile/water, 1:1 (v/v) and then dehydrated again with 100% acetonitrile. Proteins were digested overnight at 37°C with 4 ng/ μl of modified trypsin (Promega) in 50 mM NH_4HCO_3 . Peptides were extracted by first incubating the slices in 80 μl of acetonitrile/water/formic acid (70/30/0.1; v/v/v) for 20 minutes, and then in 40 μl of 100% acetonitrile for 5 minutes and finally in 40 μl of acetonitrile/water/ formic acid (70/30/0.1; v/v/v) for 15 minutes. Supernatants were transferred into fresh tubes and concentrated in a SpeedVac (Thermo Scientific) for 15 minutes to a final volume of 40 μl .

Mass spectrometry analysis

The MS measurements were done with a nanoflow high-performance liquid chromatography (HPLC) system (Dionex, LC Packings Ultimate 3000) connected to a hybrid LTQ-Orbitrap XL (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source (New Objective). The HPLC system consists of a solvent degasser nanoflow pump, a thermostated column oven kept at 30°C, and a thermostated autosampler kept at 8 °C to reduce sample evaporation. Mobile A (99.9 % MilliQ water and 0.1% formic acid (v:v)) and B (99.9% acetonitrile and 0.1% formic acid (v:v)) phases for HPLC were delivered by the Ultimate 3000 nanoflow LC system (Dionex, LC Packings). To prevent cross-contaminations between different affinity purification sets, for each baits analyzed the control assay were first applied on the HPLC column followed by the nuc+ and the nuc- assays. 10 μl of prepared peptide mixture was loaded on a trapping precolumn (5 mm \times 300 μm i.d., 300 Å pore size, Pepmap C18, 5 μm) for 3 minutes in 2% buffer B at a flow rate of 25 $\mu\text{l}/\text{minute}$. This step was followed by reverse-phase separations at a flow rate of 0.250 $\mu\text{l}/\text{minute}$ using an analytical column (15 cm \times 300

μm i.d., 300 Å pore size, Pepmap C18, 5 μm , Dionex, LC Packings). We ran a gradient ranging from 2% to 35% buffer B for the first 60 minutes, 35% to 60% buffer B from 60 to 85 minutes, and 60% to 90% buffer B from 85 to 105 minutes. Finally, the column was washed with 90% buffer B for 16 minutes, and with 2% buffer B for 19 minutes prior to loading of the next sample. The peptides were detected by directly eluting them from the HPLC column into the electrospray ion source of the mass spectrometer. An ESI voltage of 1.5 kV was applied to the HPLC buffer using the liquid junction provided by the nanoelectrospray ion source and the ion transfer tube temperature was set to 200°C. The MS instrument was operated in its data-dependent mode by automatically switching between full survey scan MS and consecutive MS/MS acquisition. Survey full scan MS spectra (mass range 400 – 2000) were acquired in the Orbitrap section of the instrument with a resolution of $R = 60,000$ at m/z 400; ion injection times were calculated for each spectrum to allow for accumulation of 106 ions in the Orbitrap. The seven most intense peptide ions in each survey scan with an intensity above 2000 counts (to avoid triggering fragmentation too early during the peptide elution profile) and a charge state ≥ 2 were sequentially isolated at a target value of 10,000 and fragmented in the linear ion trap by collision induced dissociation (CID). Normalized collision energy was set to 35 % with an activation time of 30 milliseconds. Peaks selected for fragmentation were automatically put on a dynamic exclusion list for 120 seconds with a mass tolerance of ± 10 ppm to avoid selecting the same ion for fragmentation more than once. The following parameters were used: the repeat count was set to 1, the exclusion list size limit was 500, singly charged precursors were rejected, and a maximum injection time was set at 500 ms and 300 ms for full MS and MS/MS scan events, respectively. For an optimal duty cycle the fragment ion spectra were recorded in the LTQ mass spectrometer in parallel with the Orbitrap full scan detection. For Orbitrap measurements, an external calibration was used before each injection series ensuring an overall error mass accuracy below 5 ppm for the detected peptides. MS data were saved in RAW file format (Thermo Fisher Scientific) using XCalibur 2.0.7 with tune 2.4.

Data processing and identification of peptides and proteins

The data analysis was performed with the Proteome Discoverer 1.2 software supported by Mascot (Matrixscience) and SEQUEST database search engines for peptide and protein identification. MS/MS spectra were search against the SwissProt Database filtered with the *Pyrococcus abyssi* taxonomy (release-2010_10, 1812 sequences, 545621 residues), for protein identification using the Proteome

Discoverer software (version 1.2.0.208). Mass tolerance for MS and MS/MS was set at 10 ppm and 0.5 Dalton, respectively. The enzyme selectivity was set to full trypsin with one miss cleavage allowed. Identified peptides were filtered based on Xcorr values and the Mascot score to obtain a false discovery rate of 1 %.

Identification of specific interactions

The MS data were processed manually in order to identify bait-prey specific interaction signals. To rationalize this process, we considered that the identification of a protein was significant when, at least, 5 independent peptides covering at least 15% of the protein sequence were identified and that the resulting mascot score (M) was over 100. We then defined a noise/signal significance ratio (SR) for each identified proteins. SR was calculated by dividing the independent control assay maximal Mascot scores (M_{neg}) by the pull-down Mascot scores M_{nuc-} or M_{nuc+} . Proteins that never appeared in the control assays were arbitrarily given a M_{neg} value of 1. We set the significance threshold at a SR score of 0.2 ($M_{nuc+/-} \geq 5 \times M_{neg}$); an interaction was considered to be of reasonable confidence when both SR_{nuc-} and SR_{nuc+} were below 0.2 and moderate when at least one of the two SR score was between 0.5 and 0.2. Any SRs over 0.5 classified the related proteins as part of the non-specific signal.

Protein Network and Bioinformatics analyses

The network:

The protein-protein interaction network is enriched by associating to each node GO terms and Interpro domains extracted from available protein annotations. Furthermore, we used the Gene Ontology hierarchy (v 15/11/2011) for the characterization of protein clusters and the Domine database v2.0 (Yellaboina, Tasneem et al. 2011) for the prediction of domain interactions with an associated confidence level. We used a synonym table to translate Interpro domains into Pfam domains and confidence values of interactions were translated into discrete levels according to the following coding scheme: NA=1, LC=2, MC=4, HC=5.

Topology of the network:

The clustering coefficient of a node (protein) is computed as the proportion of links observed between its direct neighbors (proteins interacting with it) with respect to the maximum possible number of links they could share. It takes value 1 for a clique (complete set of interactions) and 0 for a star (protein connected to a set of independent proteins). In case of bipartite graphs like PPI networks (only a subset of real interactions can be observed, those between a bait and a prey), the ratio is made with respect to the cross product of baits and preys. Main clusters of the network have been extracted according to the following procedure filtering pseudo-cliques in the network: all proteins whose clustering coefficient exceeds a stringent threshold (20% in the network) are first extracted and form the basis of a cluster with the subset of their interacting partners. In a second step, the proteins that are supported by only one interaction in a cluster are discarded, together with all clusters that are included in a larger cluster.

Another important aspect of the network structure analysis is to try to distinguish direct interactions from indirect ones. The aim is to obtain a scaffold, a spine or a backbone of the network that provides an abstract view of the distances between proteins in terms of communication between subparts of the network. Minimal Spanning Trees (MST) have been recognized as a good means to unravel the intricate structure emerging from affinity purification data (Friedel and Zimmer 2009). A spanning tree is a tree connecting all vertices in a graph. If the edges are weighted, it is possible to compute a MST, that is, a spanning tree with a minimal total weight. In an interaction network however, all proteins are not equal and considering the whole set of proteins may hinder the search for its core spine. The proteins having a topologically central place have been shown to play a more important biological role than others on the human PPI network (Milenkovic, Memisevic et al. 2011), using the notion of Minimum Connected Dominating Set (MCDS). A connected dominating set in a graph is a subset of its nodes such that they are connected and there exists an edge between each remaining node and a node in this subset. A MCDS has furthermore a minimal size. We have combined the two concepts by looking for a minimal spanning tree on the subgraphs of minimum connected dominating sets (i.e, a MST with respect to all possible MCDS). Thus, the produced backbone is a minimal weighted tree of interacting proteins such that all other proteins have a direct interaction with one of its proteins. In order to get the best support in the data, we use both the interaction network and the network of known associated interacting domains. Three criteria are cumulated in the weight: the SR (highest priority), the opposite of the associated domain confidence, and the domain number (lowest priority).

Finally, we have used a power graph analysis (Royer, Reimann et al. 2008) a compression of the interaction graph that help to organize and visualize in a more abstract way the whole set of interactions. A power interaction graph is defined like ordinary graphs by a set of power nodes (subsets of proteins) and a set of power edges. If a power edge connect two power nodes A and B, then each node in A is connected to all nodes in B. Two conditions are further required: power nodes have to be either disjoint or embedded sets of proteins and there is no loss of information (all interactions have to be represented in the power graph).

Pull-down immunoprecipitation assays

For the interaction PCNA:Pab0431 : the Pull-down experiments were performed as described earlier ((Meslet-Cladiere, Norais et al. 2007; Ren, Kuhn et al. 2009)) with the exception that 5µg of Pab0431 was incubated with 1 mg of Co²⁺ coated magnetic microbeads (Invitrogen) and that the beads-baits complexes formed were then incubated with 16.5 mg of *P. abyssi* cell extract for 2 h at 4 °C.

For the interaction PCNA:Mre11/Rad50 : 10 µg of Mre11:Rad50 complex was immobilized on 240 µg of magnetic microbeads (Invitrogen) in 200 µL of binding buffer (TrisHCl 20 mM pH 8, NaCl 150 mM, DTT 1 mM) in the presence of 20 µg of BSA. When indicated, DNA was degraded by the addition of 5 U.I. of DnaseI (Sigma) in 200 µL of TrisHCl 50 mM pH 8, MgCl₂ 4 mM during 1 h at room temperature. After 3 washing steps in binding buffer, 0.5 µg of PCNA was added for 1 h at room temperature in 200 µL of binding buffer in the presence of 30 µg of BSA. After 3 final washing steps, the proteins were eluted from the beads in Laemli buffer 5 minutes at 95°C and separated on SDS PAGE (4-20%, Pierce). Proteins were transferred onto a PVDF membrane (Thermo Scientific). PCNA and Mre11 were probed, using a polyclonal anti-PCNA and monoclonal anti-HisTag antibodies (Invitrogen), respectively. Dylight fluorophore coupled secondary antibodies (Thermo Scientific) were used to reveal the proteins. Fluorescence was detected using Typhoon (Amersham Bioscience) imager and analysed with ImageQuant software.

***In Vitro* Transcription Assay**

In vitro transcription reactions were performed as described previously (Spitalny and Thomm 2003). For basal transcription, reaction mixtures were assembled in 25 µL (final volume) of transcription buffer (40 mM HEPES-Na pH 7.3, 250 mM NaCl, 2.5 mM MgCl₂, 0.1 mM ZnSO₄, 0.1 mM EDTA, 0.1 mg/ml

BSA, 5% Glycerol). 10 nM gdh-C20 template DNA was combined with 100 nM TBP, 35 nM TFB, and 10 nM RNAP. Transcription was initiated by the addition of ATP, CTP, and GTP to a final concentration of 40 μ M and UTP to a final concentration of 2.7 μ M, including 2 μ Ci of [α - 32 P]UTP (3000 Ci/mmol). After incubation at 70 $^{\circ}$ C for 15 min, the transcripts were analyzed by denaturing PAGE.

Surface Plasmon resonance

Data were acquired at 25 $^{\circ}$ C using Reichert SR7000DC spectrometer instrument (Reichert Inc., Buffalo, USA). Running buffer was PBS Tween 0.5% and the flow rate was 20 μ L.min $^{-1}$. Streptavidin (Sigma, France) was immobilized on a mixed SAM (1 C $_{11}$ -(OEG) $_6$ -COOH :10 C $_{11}$ -(OEG) $_3$ -OH) surface *via* classical amine coupling chemistry. Then, 66 μ RIU of the 5' biotinylated ssDNA 5'-TGCCAAGCTTGCATGCCTGCAGGTCGACTCTA-3' was immobilized on the biotin surface. Each measurement cycle consisted first in a saturation of the ssDNA chip by RPA (10 min at 100 nM) followed by an injection of one of the 3 DNA polymerases from *P.abysssi* at the following concentrations: 16.67 nM, 50 nM, 150 nM and 450 nM. Regeneration of the sensor was achieved with 3 injections of 0.5 % SDS for 60 sec. Each curve was referenced against ssDNA/RPA dissociation curves. Data were then fitted using a global analysis method with Scrubber 2.0a software (Biologic Software, Australia).

RESULTS AND DISCUSSION

P. abyssi cells were grown in a continuous non-synchronized culture and maintained under an exponential growth phase for 6 weeks. A total amount of 350 g of cell pellet was harvested. From the cell pellet, we prepared a concentrated cytoplasmic extract, buffered in a saline solution. Twenty three recombinant proteins (Table1), protein complexes and truncated domains of proteins, were used as molecular baits to look for potential interacting partners in the cell extracts of *P. abyssi*, using affinity purification coupled to mass spectrometry (AP/MS) identification analysis. These proteins were either documented or predicted to be implicated in genome maintenance processes in *Archaea*. The recombinant forms of these baits, either HIS $_6$ - or GST-tagged, were expressed in *E. coli* and affinity purified on magnetic microbeads from the soluble bacterial cell extract (Groisillier, Herve et al. 2010). The purified proteins were then incubated in the *P. abyssi* cellular extracts to identify their potential interacting partners using AP/MS experiments. To avoid non-specific interactions mediated by

nucleotidic substrates, we performed two distinct pull-down experiments per bait, in the presence or absence of RNA/DNA nucleases. The proteins identified by mass spectrometry, that were present in both AP/MS experiments and for which the Mascot score was more than 5 fold higher in the assays compared to the negative controls, were regarded as being able to specifically interact with the bait under consideration. This corresponded to significance ratio (SR) values below the threshold of 0.2 (see Materials and methods section). We reasoned that this exploratory analysis should unravel new and unsuspected relationships between known factors and novel components involved in nucleic acid processes and thus we only used the objective criterion of SR score to classify the proteins identified in mass spectrometry, rather than excluding proteins based on functional annotations. As a result, some non-specific interactions might remain in the network depicted but we believe that this non informative part had been reduced to the minimum thanks to the analysis of two independent AP/MS experiments per bait. In support of this, the results include 72% of proteins described or predicted to act in nucleic acid pathways and contain already documented interactions between established

Table 1: List of bait proteins used in affinity purification experiments

Bait protein	functions	Tag	Property	References
Alka	DNA damage detection	HIS N-ter	FL	Groisillier <i>et al.</i> , 2010
DNA ligase	Ligase involved in	HIS N-ter	FL	Meslet-Cladiere <i>et al.</i> , 2007
DNA2 homolog	Helicase/nuclease	HIS N-ter	C-ter 1070K-1308A	Groisillier <i>et al.</i> , 2010
Fen1	Okazaki fragment	HIS N-ter	FL	Meslet-Cladiere <i>et al.</i> , 2007
Gins 23	Coordinates replication	HIS N-ter	FL	Groisillier <i>et al.</i> , 2010
Gins 51	Coordinates replication	HIS N-ter	FL	Groisillier <i>et al.</i> , 2010
MCM	Replicative helicase	HIS N-ter	FL	Supplementary data S1
Mre11, Rad50	DNA double-strand	HIS N-ter	FL complex	Supplementary data S1
	-	-	FL complex	-
NAD-dependantprotein	Unknown	HIS N-ter	FL	Groisillier <i>et al.</i> , 2010
N-glycosylase/DNA	DNA repair enzyme that	GST N-ter	FL	Groisillier <i>et al.</i> , 2010
NucS	Endonuclease	HIS N-ter	FL	Ren <i>et al.</i> , 2007
Primase small subunit	Primase small subunit	HIS N-ter	FL	Le Breton <i>et al.</i> , 2007
Primase large subunit	Primase large subunit	HIS N-ter	FL	Le Breton <i>et al.</i> , 2007
PCNA	DNA polymerase sliding	HIS N-ter	FL	Meslet-Cladiere <i>et al.</i> , 2007
php domain containing	Unknown	HIS N-ter	FL	Groisillier <i>et al.</i> , 2010
Polymerase B	DNA polymerase B	HIS N-ter	FL	Gouje <i>et al.</i> , 2012

Polymerase D	DNA polymerase D	HIS N-ter	FL complex	Palud <i>et al.</i> , 2008
	DNA polymerase D	-	FL complex	-
Rad-25 like	DNA/RNA helicase of	HIS N-ter	FL	Groisillier <i>et al.</i> , 2010
RadA	DNA repair and	HIS N-ter	FL	Groisillier <i>et al.</i> , 2010
RadB	DNA repair and	HIS N-ter	FL	Groisillier <i>et al.</i> , 2010
Rec-J like ss	Gins Associated	HIS N-ter	FL	Groisillier <i>et al.</i> , 2010
Ribonuclease HII	RNA of RNA-DNA	HIS N-ter	FL	Le Laz <i>et al.</i> , 2010
RPA (14, 32, 41)	Single-stranded DNA-	-	FL complex	Supplementary data S1
	Single-stranded DNA-	-	FL complex	-
	Single-stranded DNA-	HIS N-ter	FL complex	-

FL: full length, N-ter: N-terminus domain, C-ter: C-terminus domain

protein partners. In addition, we also identified interactions that have already been described in Eucaryotes but as yet undiscovered in Archaea (see following sections). The putative interactions detected with the highest degree of confidence with and without nuclease treatment, are presented in Table 2 (voir Tableau 9 dans le paragraphe précédent du manuscrit de thèse).

Global Network Description

In order to understand the data more thoroughly, the high confidence data set was computed in a network. Graph analysis helped us to identify the main features of the network such as clusters, protein groups and network backbone (Figure 1A). Proteins were represented as nodes and putative interactions as edges, following a spoke model (Bader and Hogue 2002). The network was composed of 93 nodes and 147 edges. Baits represented 21% of the total number of nodes. 65% of the nodes had an established or predicted role in nucleic acid metabolisms and 19% corresponded to uncharacterized functions. Modification enzymes and other metabolisms represented respectively 7% and 9% of the proteins. In addition, 38% of the interactions linked proteins of the DNA replication/repair/recombination functional categories, which also indicated that the fraction of false positives has been strongly reduced after filtering of the AP/MS data (von Mering, Krause et al. 2002). The network presented an average clustering coefficient of 10,3 % corrected to 11,1 % when the bias of prey-prey interactions blindness was taken into account. The most isolated protein is DNA2 (clust. coef. 0.8%) and the proteins that are linked to a highly connected cluster and thus most likely interact inside a protein complex (clust. coef. $\geq 50\%$, cluster size ≥ 4) are the DNA topoisomerase 1 and the

Lhr-2 protein which are both related to the Alka-OGG1-RPA set of proteins. Using graph analysis, we identified highly connected areas in the network *i.e.* the optimal cliques of at least 3 proteins in the network (Figure 1B).

The clustering centers were Proliferating Cell Nuclear Antigen homotrimer (PCNA), Replication Protein A heterotrimer (RPA), the catalytic subunit of DNA primase p41 and RNA polymerase subunit A2. In addition, we also computed the minimum spanning tree (MST) of the dominating set of the network, using the SR score as the sole constraint. This sub-graph (edges highlighted in thick gray lines cf in Figure 1) corresponded to a backbone network that contained only the essential topological information to represent the protein network, based on the consistency of the AP/MS analysis. It represented the interactions that are more likely direct within the network. This sub-tree presented a marked root corresponding to the RPA complex. Altogether, the clustering and MST analysis emphasized the fact that PCNA and RPA complexes are essential proteins in Archaea, involved in distinct and numerous DNA transactions, spanning DNA replication, DNA repair and DNA recombination, and that both factors are involved in interactions with many different partners (Berquist, DasSarma et al. 2007; Skowyra and MacNeill 2012) (Kelman and Hurwitz 2000; Komori and Ishino 2001; Robbins, Murphy et al. 2004; Lin, Guzman et al. 2006) (Warbrick 2000; MacNeill 2011; Creze, Ligabue et al. 2012) (Pan, Kelman et al. 2011).

In addition, complementary to the clustering analysis, we performed a co-occurrence analysis that allowed us to determine the optimal groups of 2 or 3 proteins sharing the largest numbers of common partners (Figure 1C). Although these proteins were not linked within the network, the co-occurrence of similar preys indicated a possible relationship between the baits concerned. The protein groups we identified were RPA and Rad25-like helicase ; the single strand endonucleases Fen1 and NucS ; the subunits rpoa1 and rpoa2 of the RNA polymerase (RNAP) ; Alka and OGG1 ; RadA, RadB and DNA2 ; DNA primase p46 and DNA primase p41 ; MoaB1 with the subunit rpoF of RNAP and the hypothetical protein PAB0105 ; PCNA with HsdS subunit of DNA restriction/modification system. As expected, these groups corresponded to either subunits of multimeric proteins (RNAP, DNA primase) or to functionally related proteins known to interact (RadA/RadB, (Komori, Miyata et al. 2000)), hence validating the relevance of the co-occurrence analysis and thus highlighting the putative inter-relationship between the proteins present in these groups. In this light, the interaction between

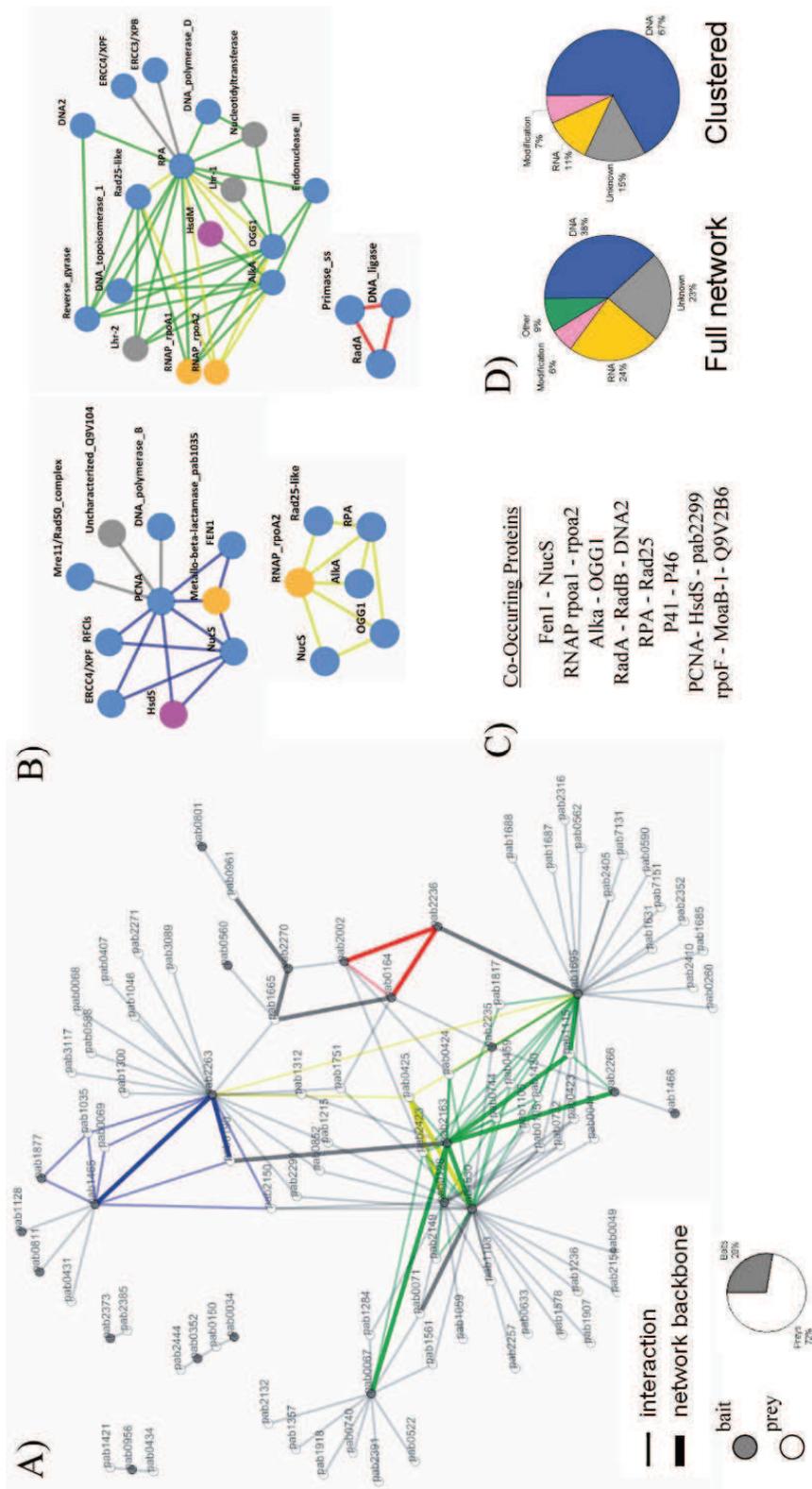


Figure 1: Statistic and topological analysis of the protein interaction network. A, protein interaction map. Baits are indicated in grey dots and preys in open circles. Network backbone is represented in thick lines. Protein clusters are highlighted in blue (PCNA), Green (RPA) Yellow (RNAP rpoa2) and red (DNA primase ss). B, optimum clusters. Nodes colour follow attributed function of represented protein ; DNA replication/recombination/repair in blue, RNA metabolisms in yellow, Methylation/modification in pink and unknown function in grey. C, optimum groups of co-occurring proteins. D, Repartition of metabolisms in full network and in clusters.

Rad25-like, belonging to the superfamily 2 of helicases and thus generating single strand nucleic acids, and RPA that has a high affinity with single strand DNA, seems to be consistent.

Are the interactions detected specific and biologically relevant ?

Others have already attempted to reveal the interaction network involved in Archaeal DNA replication using either Y2H and AP/MS *in vivo* approaches, in *Archaeoglobus fulgidus* and *Thermococcus kodakaraensis*, respectively (Motz, Kober et al. 2002; Li, Santangelo et al. 2010). We compared our data to these studies and found few common interactions (Table 2), representing 6,6 and 8 % of overlap when we only considered the interactions detected by the same baits, used by Motz and Li, respectively. These three studies used different methodological approaches on different species to determine potential physical associations, which probably accounts for the differences in coverage observed (von Mering, Krause et al. 2002). However, we took advantage of these two datasets to analyze our own raw data in greater depth. We were thus able to identify two additional potential associations (Table 2), also found by Moltz and co-workers that were originally considered of moderate confidence in our dataset, based on the SR score only ; these results will be discussed later (see following sections). Only two associations were common to the three studies, PCNA-Fen1 and PCNA-RF-C, likely reflecting the high affinities between these components and their high abundance in the cells.

Table 2. Common interactions in protein-protein interaction network of DNA pathways in HA

Interactions	Datasets *
PolD su1-PolD su2	1,2,3
PCNA - Fen-1	1,2,3
PCNA - RF-C	1,2,3
PCNA - PolD	1,2
RPA - Q9V1R8	1,2
RPA - Nth endonuclease III	1,2
RPA - Reverse gyrase	2,3
GINS51 - PolD	2,3

* 1, Motz *et al.* 2002 (Y2H) ; 2, Li *et al.* 2010 (*in vivo* AP/MS) ; 3, this study (*in vitro* AP/MS).

Then, we compared our data with the previously established or reported interactions between components of genomic maintenance in Archaea. We found that the PCNA replication clamp interacted with the small subunit of the loading factor RF-C, the replicative DNA polymerase B and the

flap endonuclease Fen1. These corresponding interactions have been shown to take place within the chromosome replication machinery (Meslet-Cladiere, Norais et al. 2007; Rouillon, Henneke et al. 2007). In addition, we also found the PCNA associated with the recently described ssDNA nuclease NucS, as well as an interaction between NucS and Hef/Xpf proteins. We already reported these associations, using different technical approaches, and proposed that they could be implicated in branched-DNA repair pathway (Ren, Kuhn et al. 2009; Creze, Ligabue et al. 2012). We found that the subunit GINS51 of the GINS complex interacts with the exonuclease subunit of the DNA polymerase D (PoID), the interaction between the polymerase subunit of PoID and GINS51 was also detected albeit displaying higher SR scores (SR⁻:0,22 ; SR⁺:0,41). Although this interaction has not been characterized to date, it has been reported in the study of Li *et al.* that GINS51 interacts with both subunits of PoID (Table 2). These congruent data suggest that the interaction between GINS51 and PoID is specific and reinforce the hypothesis that “GINS functions as the center of the replisome, linking the polymerase, helicase, and primase components” (Li, Santangelo et al. 2010).

Subunits RPA32 and RPA41 of the RPA complex were co-purified with the HIS-tagged Polymerase D complex (PoID). The interaction between RPA41 and PoID was also reported in *Pyrococcus furiosus* using immunoprecipitation experiments (Komori and Ishino 2001), suggesting that a specific interaction between RPA and PoID is likely to occur at the replication fork in *P. abyssi*. In this latter report, Komori *et al.* also reported interactions between RPA41 and both DNA primase and DNA polymerase B (PoIB), which were not detected in our high confidence dataset, although His-tagged RPA complex co-isolated both subunits of the DNA primase but with moderate SR scores (SR⁻/SR⁺: 0,09/0,26 for PriL and SR⁻/SR⁺:0,36/0,51 for PriS). To clarify this situation, we sought to characterize the physical interactions between the RPA complex and the DNA polymerases from *P. abyssi*. To this aim, we designed a SPR experiment using RPA complex immobilized on a 32mer ssDNA oligonucleotide to monitor the interaction with the polymerases injected over the flow cell. This experimental set up not only allowed us to orient the RPA molecules on the chip, which is known to improve the analysis of the SPR signals, but also had the potential advantage of approaching the physiological conditions required for these interactions to take place. Figure 2 clearly demonstrated that both PoID and the small catalytic subunit of DNA primase interacted with RPA, bound to ssDNA, with similar affinities, whereas PoIB did not bind to RPA complex. According to the dissociation constant measured (Figure 2), the RPA/PoID complex is roughly twice as more stable than the

RPA/catalytic subunit of DNA primase complex, which could explain the differences observed between the SR scores of PolD and DNA primase in our AP/MS experiments. These SPR results are thus in good agreement with the results of the AP/MS experiments. On the other hand, it is in contradiction with the data of Komori *et al.* that showed an interaction between RPA and PolB, however, it is not unreasonable to suggest that the very weak signal obtained using immunoprecipitation for PolB in the latter study, when compared to those for DNA primase and PoD, could be an artefact. It has been proposed, based on biochemical properties, that both PolD and PolB from *P. abyssi* could be present at the replication fork and that Pol B might be the candidate to replicate the leading strand while Pol D could synthesize the lagging strand (Henneke, Flament et al. 2005). The differences in RPA binding could thus reflect separate mechanisms for the initiation and elongation steps of PolD and DNA primase on lagging strand and PolB on leading strand.

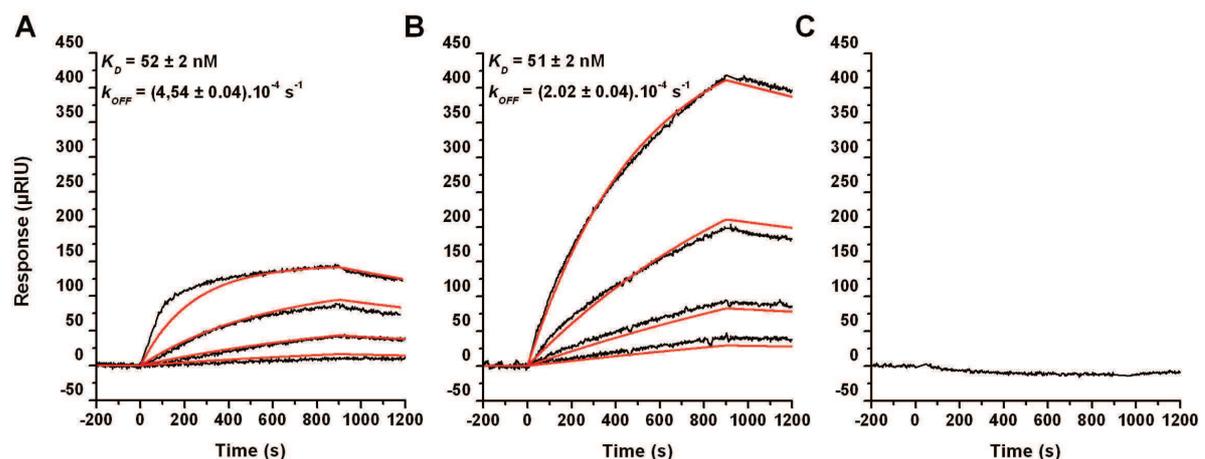


Figure 2: Sensorgrams of the interaction between RPA bound to ssDNA and the 3 DNA polymerases from *Pyrococcus abyssi*. The catalytic subunit of the DNA primase (A), PolD (B) and PolB (C) were injected at different concentrations (16.67 nM, 50 nM, 150 nM and 450 nM) on a chip coated by the assembly of a 32mer ssDNA bound by RPA. Signals reported are already subtracted from a RPA dissociation curve and data were fitted using a 1:1 binding reaction model.

Altogether, with the limitation that not all known complexes could be detected, possibly because they may not assemble under our growth conditions or because the tag interferes with complex assembly, these data indicate that the filtering process of our data is consistent and that the threshold used to sort out the mass spectroscopy data is selective enough to ensure that many biologically relevant interactions involved in DNA transactions can be revealed through this approach.

Besides these already reported interactions, the network encompassed also, either unknown interactions between documented components of the archaeal genomic maintenance, or potential new components. In the next sections we will describe these associations with a special emphasis on the main PCNA, RPA and DNA primase clusters detected in the network.

PCNA: the fellowship of the ring enlarges

Mre11/Rad50 Complex

PCNA, used as bait, co-isolated both Mre11 and Rad50. The reverse experiment, performed with the tagged complex Mre11/Rad50 also co-isolated the PCNA, strongly suggesting that PCNA could form a stable complex with Mre11/Rad50 assembly in *P. abyssi*. PCNA, Mre11 and Rad50 from *P. abyssi* are close homologs to their eucaryal counterparts and as the interaction between these components has not been reported as yet in both domains, we sought to confirm this finding. With that in mind, we carried out co-immunoprecipitation *in vitro* with the purified PCNA and Mre11/Rad50 complexes. Figure 3 clearly showed that the complex Mre11/Rad50 co-immunoprecipitated the PCNA either in presence or absence of nuclease treatment, thereby validating the result of the AP/MS experiment.

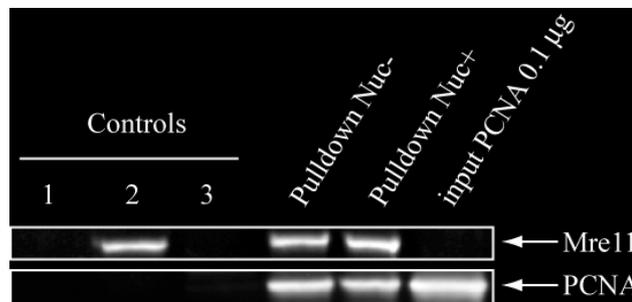


Figure 3: PCNA interacts with Mre11/Rad50 complex. 10 µg of HIS fusion Mre11/Rad50 immobilized on Co²⁺ magnetic beads was incubated with 0.5 µg of PCNA in absence or presence of DNase I (Pulldown Nuc- or Pulldown Nuc+). Control lanes consisted in the Co²⁺ beads only (1) or incubated with either the complex Mre11/Rad50 (2) or the PCNA (3). Bait and prey proteins were analysed by Western blot, and probed with PCNA or HIS antibodies.

The eucaryotic Mre11 complex is composed of the meiotic recombination 11 (Mre11), Rad50 and Nbs1 (Nijmegen breakage syndrome 1). The Mre11 complex is a sensor of double-strand breaks

(DSBs) that also controls the DNA damage response by governing the activation of the central transducing kinase ATR/ATM. In addition, the Mre11 complex regulates DSB repair, through the homologous recombination repair (HR) and non-homologous end-joining (NHEJ) (Stracker and Petrini 2011). However, a number of recent reports are consistent with the eucaryal Mre11 complex having a primary role in the maintenance of the replication fork during DNA replication and in the HR-dependent resolution of DNA replication-associated DSBs (Stracker and Petrini 2011). All the archaeal genomes sequenced to date contain clear homologs of eucaryal Mre11 and Rad50 and the initial biochemical characterization of the *P. furiosus* homologs indicate that the archaeal Mre11 complex is functionally similar to those from Bacteria and Eukarya. However, Nbs1 additional component has not been found in Bacteria and Archaea. In contrast, the HerA helicase and NurA nuclease, specific to the Archaea and frequently found in thermophilic archaeon, have been found to be physically and functionally associated with the archaeal Mre11 complex (Constantinesco, Forterre et al. 2002; Constantinesco, Forterre et al. 2004; Hopkins and Paull 2008; Quaiser, Constantinesco et al. 2008). Indeed, these four proteins co-operate for the 5' strand resection at DSB prior to HR (Hopkins and Paull 2008). However, the physiological functions of the Mre11 complex remain unsolved in Archaea, due to the conflicting data of the genetic studies performed so far. (Delmas, Shunburne et al. 2009; Fujikane, Ishino et al. 2010).

To date, the physical interaction between PCNA and Mre11 complex has never been reported either in Eucarya or in Archaea. However, it has been shown that the eucaryotic Mre11 complex colocalizes with the PCNA at replication foci during S phase and that the eucaryal Mre11 complex plays a primary role in the maintenance of the replication fork during DNA replication (Huertas, Cortes-Ledesma et al. 2008). If this were also true for the Archaea, it would be tempting to speculate that the archaeal PCNA could be involved in the recruitment and dynamics of the Mre11 complex at the replication fork, either to stabilize the components of the replisome or to repair stalled replication fork intermediates by HR. Another non exclusive hypothesis could be that PCNA might also, in some way, enhances the mechanism of end resection carried out by the Mre11 complex, NurA and HerA, prior to strand invasion at DSB sites.

Pab0431, a new nuclease associated with PCNA ?

Pab0431 co-precipitated with the PCNA bait (Figure 1) and we confirmed, using pull-down immunoprecipitation (Fig. 4A) that this interaction was specific. As numerous client proteins of the PCNA interact with the replication clamp via a PIP box (Beattie and Bell 2011), we looked for the occurrence of such motif in the sequence of Pab0431. The alignment of *P. abyssi* proteins, known to interact with the PCNA via the PIP motif (Henneke, Gueguen et al. 2002; Meslet-Cladiere, Norais et al. 2007; Castrec, Rouillon et al. 2009; Creze, Ligabue et al. 2012), with the C-terminal end of Pab0431 indicates that a well conserved consensus sequence is present in this portion of the protein (Fig 3B).

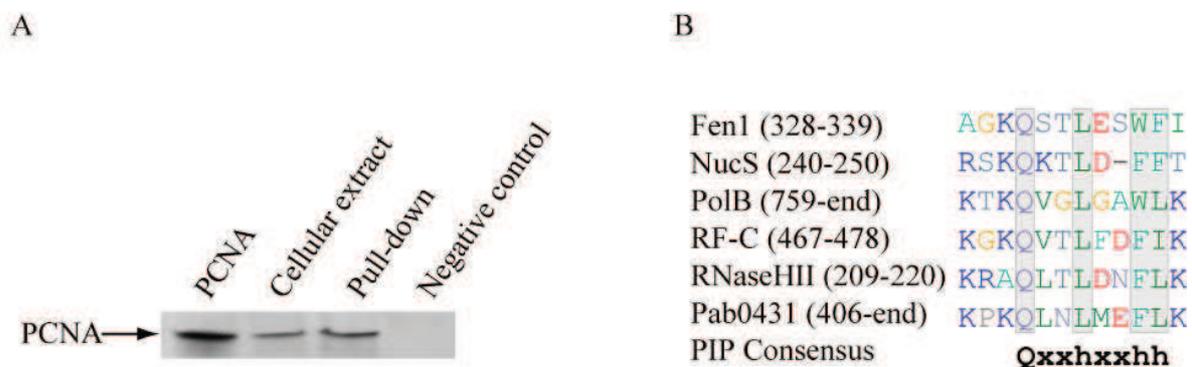


Figure 4: Pab0431 interacts with PCNA and displays a putative C-terminal PIP motif. A, 5 µg of histidine-tagged Pab0431 bound to magnetic beads (lane pull-down), and a negative control that consists of the beads only, were incubated with 16 mg of *P. abyssi* cellular extract. Control lanes contained 0.33 µg of cell free extract and 8.75 ng of PCNA. The presence of PCNA on the blot was revealed using anti-PCNA antibody. B, Pab0431 has a canonical C-terminal PIP box motif when compared to the consensus sequence (x and h respectively refer to any amino acids and hydrophobic residues) and various sequences of proteins from *P. abyssi* known to interact with the PCNA through the PIP motif. The position of the PIP box within the sequences is indicated

Pab0431 possesses a N-terminal PHP (p_olymerase and h_istidinol p_hosphatase) domain (5-226) and a Zn ribbon motif (248-270) ; it belongs to COG1379, which members do not have a clear function. However, the PHP domain is a putative phosphoesterase domain found, isolated or associated, in proteins with diverse functions, including bacterial DNA polymerase III, eucaryotic DNA polymerase, X-family of DNA polymerases and histidinol phosphatases (Aravind and Koonin 1998). This domain consists of four conserved core regions containing invariant histidine and aspartate residues implicated in metal ion coordination. These four motifs were found conserved in the sequence of Pab0431 (data not shown). The proteins of the COG1379 family are found in Bacteria

and Archaea ; within this latter phylum, they are restricted to Euryarchaea with the exception of halobacteriales, thermoplasmatales and Archaeoglobales orders.

A HHpred search with the pab0431 sequence indicated significant similarity (E-value=2.2e-61) with the N-terminal sequence (7-436) of DNA polymerase III alpha subunit (PolIII α) from *Thermus aquaticus*. This region contains a PHP domain which is present in all PolIII α sequences and was found to exhibit zinc ion-dependent 3'-5' exonuclease Activity (Stano, Chen et al. 2006). Pab0431 also displays similarity to a stand-alone PHP domain of the YcdX Protein from *E. coli* that is closely related to the PolIII α PHP domain. The structure of YcdX revealed a Zn²⁺ trinuclear center with characteristics similar to several phosphoesterases (Teplyakov, Obmolova et al. 2003). Interestingly, YcdX is significantly induced in response to DNA damage caused by mitomycin C in *E. coli* (Khil and Camerini-Otero 2002).

We demonstrated that Pab0431 is a new partner of the PCNA and is thus likely to be involved in a DNA metabolism process. Pab0431 displays a canonical PIP motif at its C-terminal end and although mutation analyses are needed to assess the role of this motif regarding the interaction detected, it is tempting to speculate that this sequence could mediate the interaction with the replication clamp. The presence of a PHP domain, containing the four conserved motifs responsible for metal coordination, as well as a Zn-ribbon motif tips the balance in favour of potential metallo-dependent hydrolase and nucleic acid binding activities. In addition, Pab0431 has significant similarities to the PHP domain of bacterial PolIII α , recently described to have a zinc dependent proofreading activity, and with the YcdX protein of *E. coli* that contains a stand-alone PHP domain and also may be related to DNA repair. Taken together, these data suggest that Pab0431 may have a nuclease activity related to DNA replication, as a separate subunit providing an additional proofreading activity for the replicative DNA polymerases, or could be involved in a DNA repair process that remain to be determined. This 'hypothetical protein' thus appears to be an exciting target for further functional characterization. The putative PIP motif seems to be conserved only in the sequence of the representatives of the Thermococcales order. For this criterion to be validated, prioritizing the functional characterization of this phylogenetic group is necessary.

Pab1035, a putative novel archaeal β -CASP DNA nuclease.

In the cluster PCNA, Pab1035 was found associated with PCNA, RF-C, NucS and Fen1 (Figure 1). Pab1035 belongs to COG0595 and arCog00547 (Makarova, Sorokin et al. 2007), which encompass predicted hydrolases of the metallo-beta-lactamase super-family. Sequence analysis indeed indicated that Pab1035 belongs to the β -CASP subfamily of zinc-dependent metallo- β -lactamases (named for metallo- β -lactamase, CPSF, Artemis, Snml, Pso2 (Callebaut, Moshous et al. 2002)). In contrast to canonical metallo- β -lactamases, members of the β -CASP family contain an additional inserted domain (β -CASP) and this subfamily comprises of numerous proteins acting on nucleic acid substrates, involved in DNA repair and RNA processing (for a review, c.f. (Dominski 2007)). This β -CASP domain contains three conserved motifs (motifs A, B and C) and it has been suggested that the nature of a particular amino acid included in the motif C could discriminate between RNA and DNA processing enzymes (Callebaut, Moshous et al. 2002). On this basis Pab1035 is annotated as a predicted exonuclease of the beta-lactamase fold involved in RNA processing. However more recent functional characterizations of RNases of the β -CASP family indicate a lack of discrimination between RNA and DNA which may be common among the members of this family (Clouet-d'Orval, Rinaldi et al. 2010). Besides, studies on Artemis, a eucaryotic β -CASP member involved in V(D)J Recombination and DNA double-strand break repair by NHEJ, demonstrated that despite having a valine in motif C instead of the canonical histidine in RNA processing enzymes, it can act as a 5' exo-nuclease on both single stranded DNA and RNA substrates (Ma, Pannicke et al. 2002). In addition, replacement of this residue with alanine, histidine or serine did not rule out the nuclease activity of Artemis, suggesting that this residue is not essential (Poinsignon, Moshous et al. 2004), and thus should not be used any more as a criterion for functional prediction of the substrate preference of the member of the β -CASP family. *P. abyssi* genome encodes three members of this family. A recent phylogenomic analysis of archaeal β -CASP orthologs (Clouet-d'orval, personal communication) suggested that Pab1751 is a orthologue of bacterial RNase J with 5' exo-nuclease activity *in vitro* (Clouet-d'Orval, Rinaldi et al. 2010), Pab1868 is related to the Eucaryotic CPSF-73 (cleavage/polyadenylation specificity factor) and displays endo- and exo-ribonucleolytic activity reminiscent to that of the eucaryotic counterpart. Both of these RNases could be involved in the processing of mRNA precursors. In contrast, Pab1035 belongs to a highly divergent cluster and its function is still yet unknown. However, the position of Pab1035 in the network,

potentially associated with two ssDNA nucleases NucS and Fen1 and with the replication clamp, and the evolutionary divergence from the canonical RNA nucleases of the same family, suggest that the function of this enzyme might not be restricted exclusively to RNA metabolism. This enzyme could in fact process DNA substrates and be involved in the DNA repair mechanisms. As the other members of the β -CASP family, acting on DNA substrates, are mainly involved in the repair of ICLs and double-strand break repair by NHEJ (Dominski 2007), it is tempting to speculate that Pab1035 could act as a 5' exo-nuclease in one or both of these pathways.

RPA serves as a mediator for genome maintenance in *P. abyssi*

RPA activates transcription in vitro

The interaction observed between the large subunit RPA41 of the heterotrimeric RPA complex and the two A1 and A2 subunits of the RNA polymerase (RNAP) suggested the possibility that RPA could play a role in transcription (Figure 1). As the purified RNAP from *P. abyssi* was not available, we tested this hypothesis using the basal transcription machinery from *P. furiosus* and the RPA complex from *P. abyssi* in a promoter-specific transcription assay using the *gdh* promoter as a template ((Spitalny and Thomm 2003). Both RPA and RNAP complexes from these two *Pyrococcus* strains display high sequence identities and the numbers and domain organisation of the subunits are similar (Komori and Ishino 2001; Kusser, Bertero et al. 2008; Waegel, Schmid et al. 2010). RPA complex was incubated with DNA, Transcription Factor B (TFB) and TATA-box Binding Protein (TBP). Reactions were initiated by addition of nucleoside triphosphates and the corresponding run-off transcript was formed (Figure 5). Interestingly, with the addition of increasing amounts of RPA the total amount of synthesized transcripts also accumulated up to almost three fold (Figure 5, compare lanes 1-6). Control experiments with buffer or in the absence of transcription factors confirmed that the activation of transcription depends on the presence of RPA. The same transcription profile was observed with the circular template (*gdh* template cloned in pUC19 ; data not shown).

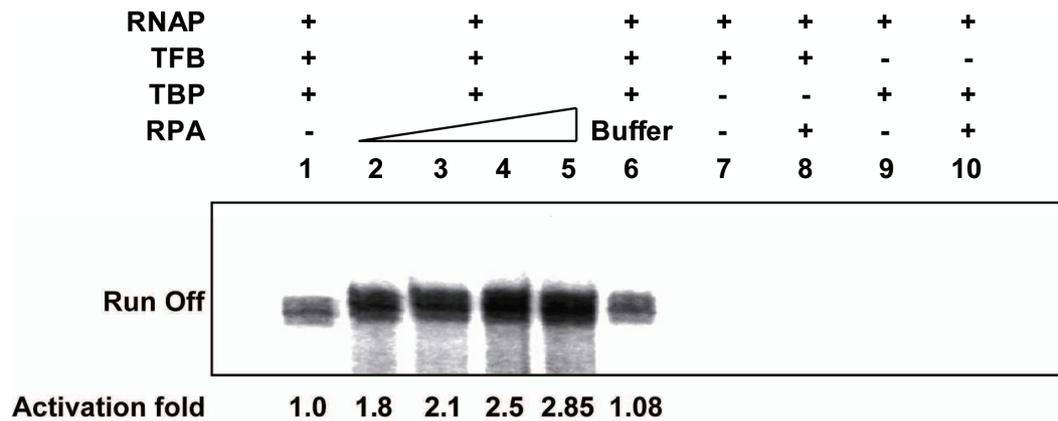


Figure 5: RPA stimulates *in vitro* transcription. The effect of RPA on transcription was analyzed by *in vitro* transcription experiments either in the absence or in the presence of increasing amounts of RPA (10 nM, lane 2; 20 nM, lane 3; 30 nM, lane 4; 50 nM, lane 5). The presence or absence of TBP or TFB was indicated on top of the lanes. The effect of RPA on transcription activity was quantified by phosphorimaging (Fujifilm FLA-5000) and indicated as activation fold at the bottom of each lane.

Archaea have been found to possess multiple RPA and SSB homologs and among the *EURYARCHAEA*, RPAs display an unusual diversity and distribution (Kernchen and Lipps 2006; Lin, Lin et al. 2008; Skowrya and MacNeill 2012). The RPAs from *P. furiosus* and *P. abyssi* exist as a stable hetero-oligomeric complex consisting of three subunits, RPA41, RPA14, and RPA32 (Komori and Ishino 2001). Comparisons of amino acid sequences and protein domain organisations strongly suggest that the euryarchaeal RPA proteins are more closely related to the eucaryal one than to the crenarchaeal and eubacterial SSBs. Studies of archaeal RPAs and SSBs have provided many insights into the role these single strand DNA binding proteins serve in DNA processing pathways, including DNA replication, recombination and repair. In particular, the crenarchaeal SSB of *Sulfolobus Solfataricus* was shown to enhance transcription in TBP limiting conditions (Richard, Bell et al. 2004), but this is the first study demonstrating that a euryarchaeal heterotrimeric RPA, reminiscent to the organisation of eucaryotic RPA, increases transcription *in vitro*.

Archaea and eukarya share a common basal transcription apparatus that consists of a multisubunit RNAP and three transcription factors, TBP, TFB, and transcription factor E. Richard et al. reported a specific physical interaction between the SSB and RNAP from *S. solfataricus* and demonstrated that SSB supports transcription in the absence of TBP and stimulates the polymerase at the level of recruitment and/or initiation (Richard, Bell et al. 2004). They suggested a model in which SSB in transcription could have a role in disruption of chromatin structure at the promoter and

recruitment of RNA polymerase to form the pre-initiation complex. In sharp contrast with this model, using abortive transcription experiments, we demonstrated that archaeal RPA did not stimulate initiation of transcription and could not substitute for a basic transcription factor in initiation (full analysis to be presented elsewhere). These properties were rather indicative of a mechanism of stimulation of transcription similar to the one demonstrated recently for the eucaryal RPA heterotrimer. Indeed, quantitative proteomic analyses highlighted the yeast RPA complex as an interaction partner of the transcription machinery (Sikorski, Ficarro et al. 2011). Further genetic and biochemical experiments led to the suggestion that RPA interacts with the non-template strand of RNA polymerase II complexes during elongation. However, the function of eucaryal RPA associated to the elongating-transcription complex is still unknown. In that regard, further works are needed to determine whether the archaeal RPA may also interact with elongation complexes and bind to non transcribed templates. Further understanding of the functional and physical interaction found between archaeal RPA and RNAP could provide exciting advances on transcription mechanisms both in Archaea and Eucarya.

We then looked at all the subunits of RNAP present in the network and extracted a sub-graph representing their first direct neighbours (Figure 6). Interestingly, three proteins implicated in base excision repair in eucarya or bacteria, namely 8-oxoguanine DNA glycosylase OGG1 (Pab1695), 3-methyl-adenine DNA glycosylase AlkA (Pab1530) and DNA ligase (Pab2002), associated with different subunits of RNAP. Although transcription-coupled repair has not been demonstrated yet in archaea (Dorazi, Gotz et al. 2007; Romano, Napoli et al. 2007), this finding might suggest the existence of transcription-dependent repair of oxidative DNA damages, at least, in hyperthermophilic archaea. Such a transcription-dependent repair of oxidative DNA damages was recently highlighted in mammal cells by the interaction between DNA glycosylase NEIL2 and RNA polymerase II (Banerjee, Mandal et al. 2011).

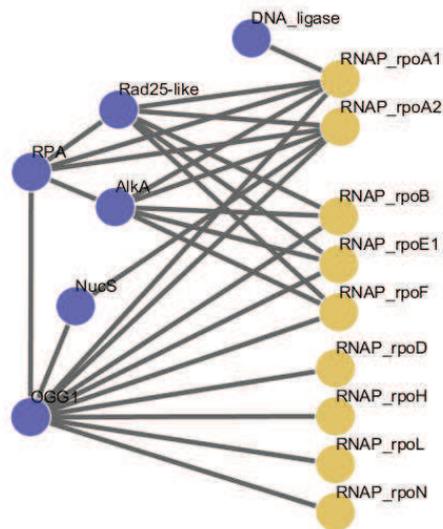


Figure 6: RNAP subunits interacting partners. Sub-graph representing the direct neighbours of all the RNAP subunits detected in the network.

The Rad25-like helicase (Pab0128) was also found associated with many subunits of RNAP (Figure 6). Rad25-like is a superfamily II helicase, closely related to archaeal XPB and its function is still unknown. The putative helicase is the representative of a new family of DNA/ RNA helicases present in Archaea, Bacteria and viruses from the three domains (Cohen, Barbe et al. 2003). The eucaryal XPB is a component of the TFIIH complex and is essential for both NER and transcription initiation (Coin, Oksenysh et al. 2007; Oksenysh, Bernardes de Jesus et al. 2009). Although the Rad25-like protein is less related to the eucaryal XPB than the genuine XPB homolog encoded in the genome, it is tempting to speculate that this helicase could have a role in the initiation of transcription, ancestral to the function of the eucaryal XPB in transcription as a component of the TFIIH complex.

Is there a Dna2 dependent pathway for Okazaki fragments maturation in Archaea ?

An association was found between the subunit RPA32 and the protein Dna2 (Pab0067) using either the RPA complex or Dna2 as bait, strongly suggesting that this interaction is specific. As the full Dna2 protein from *P. abyssi* was poorly expressed and essentially in the insoluble fraction, we truncated the protein Pab0067 in two domains, the N-terminal portion encompassing the predicted nuclease domain of RecB family (1-600) and the C-terminal halves comprising the superfamily I DNA/RNA helicase domain (601-1308). We only succeeded to express and purify the C-terminal helicase domain and used this truncated domain as bait in our AP/MS assay, thereby suggesting that this portion of Dna2 potentially interacts with RPA. Pab0067 encodes a distantly related homologue of the eucaryal Dna2

protein, which is formed by the fusion of a helicase and an endonuclease domain. Dna2 is postulated to cleave long DNA flaps that escape Fen1 activity during Okazaki fragment (OF) maturation in yeast (Ayyagari, Gomes et al. 2003). In contrast, human Dna2 seems to play a role in DNA replication that is distinct from Fen1 and OF maturation (Duxin, Moore et al. 2012). It has already been demonstrated that OF have similar sizes in *P. abyssi* and Eucarya (around 150 nucleotides) (Matsunaga, Norais et al. 2003). A recent biochemical study has suggested that the coordinated activities of RNaseHIII, Fen1, DNA polymerases B and D and DNA ligase1 are involved in RNA primer removal in *P. abyssi* (Henneke 2012). This study was performed using short flap substrates (max 16 nt), thereby not precluding the possibility that an other pathway might also exist to cleave long DNA flaps that escape Fen1 activity. In this regard, the interaction between RPA and Dna2 could be indicative of an alternative RPA/Dna2/Fen1 pathway for maturation of OF in Archaea.

The DNA Primase clusters with RadA and DNA ligase in the network.

Interestingly, the cluster primase consisted of the catalytic subunit of DNA primase, DNA ligase and RadA, suggesting that DNA primase and DNA ligase I could be implicated in recombinational repair of DNA double-strand breaks. Given the relative paucity of DNA polymerases dedicated to repair functions in archaeal genomes and the architectural connection between Pol X DNA polymerases and the small subunit of archaeal primases (Lao-Sirieix, Pellegrini et al. 2005), it is not unreasonable to speculate that the archaeal DNA primase could also have DNA repair functions. In this respect, the primase complex from *S. solfataricus* was shown to be capable of template-dependent polymerization across discontinuous DNA templates (Hu, Guo et al. 2012). Based on this new biochemical property, it was suggested that DNA primase may be involved in double-stranded break repair in Archaea, which is consistent with our data.

Independently of the clusters detected, other associations in the network attracted our attention and, we believe, provide firm basis for further investigations.

Is XPD an accessory helicase at the replisome ?

Strikingly, the helicase XPD was found in interaction with the supposed replicative helicase MCM, independently of the main network (Figure 1). The presence of XPD only in the MCM pull-down and its

total absence in the five independent negative controls, strongly suggests that we are observing a true interaction, rather than a non-specific effect.

XPD is a superfamily-2 (SF2) 5'-3' helicase containing an iron-sulphur cluster (Rudolf, Makrantonis et al. 2006). As a component of transcription factor II H (TFIIH), the eucaryal XPD is involved in DNA unwinding during nucleotide excision repair (NER) (Coin, Oksenyich et al. 2007). Archaeal XPD is closely related in sequence to the eucaryal enzyme but the biological functions of XPD in archaeal cells still remain unclear. Biochemical properties are consistent with an implication in DNA excision repair pathway (Liu, Rudolf et al. 2008; Rudolf, Rouillon et al. 2010) but a recent genetic analysis of DNA repair in *Thermococcus kodakarensis*, suggested that XPD does not seem to play a major role in *T. kodakarensis* NER pathway, given the high level of resistance to UV, methyl methanesulfonate and mytomycin C exhibited by the Δxpd mutant (Fujikane, Ishino et al. 2010). Moreover, apart from our study, no physical interactions between archaeal XPD and other proteins have been demonstrated that could give clues to the physiological functions of this helicase or enforce the hypothesis of a role in archaeal NER.

A recent genetic study demonstrating that DinG is essential for efficient replication across highly transcribed regions in *E. Coli* (Boubakri, de Septenville et al. 2010) may shed light on the interaction found between MCM and XPD. Indeed, DinG is a bacterial representative of the SF2 family of helicases and translocates in the 5'-3' direction on ssDNA. The bacterial enzyme behaves *in vitro* similarly to the archaeal XPD (White 2009) and knockouts do not render *E. coli* sensitive to DNA damaging agents (ref). Using construct strains carrying chromosomal inversions, Boubakri and co-workers, demonstrated that DinG is recruited to the replication fork to allow replication across oppositely oriented transcribed ribosomal operons, removing R-loops and participating in RNA polymerase removal (Boubakri, de Septenville et al. 2010).

In archaea, the enzymes implicated in clearing away the path of replication fork from the RNA polymerase are not known, however, it has been shown *in vitro* that the replicative helicase MCM of *Methanothermobacter thermautotrophicus* could not unwind DNA in the presence of stalled transcription elongation complex, which raised the hypothesis that, at least *in vitro*, additional proteins may facilitate the removal of the RNA polymerase complex (Shin, Santangelo et al. 2007). In this light, and without ruling out a possible role of XPD in NER pathway, it is tempting to speculate that this

helicase could also be implicated in RNA polymerase removal. In this respect, after fork arrest due to collisions between replication and transcription complexes, archaeal XPD could be recruited and then act either alone or in a concerted manner with MCM to dislodge the RNA polymerase and remove the R-loops. Another possibility could be that XPD travels with the fork and may act in preventing fork arrest and separation of the replisome components upon collisions. The activities of archaeal XPD against R-loops and DNA bound by the RNA polymerase has not yet been tested, further genetic and biochemical experiments are thus required to confirm this hypothesis.

Pab0961 potentially couple DNA replication and DNA recombination at the replication fork.

Pab0961 was found associated with RadB and a NAD-dependent protein deacylase of Sir2 family (Pab0801) in the high confidence dataset of proteins.

Independently, the interaction between the homologous proteins of Pab0961 and RadB from *Haloflexax volcanii* was also detected by wardell *et al.* in the laboratory of Thorsten Allers (To be submitted along with this report in the same journal). In a report published in the same issue, they demonstrate the physical interaction between the homologs of Pab0961 (Hvo_2384) and both RadA and RadB. In addition, phenotypic analysis of the Δ HVO_2384 mutant shows that it has a normal growth rate and is not sensitive to UV light. However, it is less sensitive to mitomycin C and more efficient at homologous recombination than the wild-type. They proposed that in the absence of HVO_2384, replication forks that are stalled at interstrand crosslinks, caused by mitomycin C, are restarted more efficiently because homologous recombination is uninhibited. These findings indicate that HVO_2384 could prevent recombination, at the replication fork, presumably by inhibiting RadB. This data are consistent with our finding of Pab0961 physically associated with RadB, thereby indicating that this interaction, found in two independent studies is likely specific and suggesting that this protein could have similar functions in *P. abyssi*.

Pab0961 is a member of the CBS-domain-containing membrane protein COG3448. It contains two pairs of Cystathionine β -synthase (CBS) domains. These domains are often found in proteins with completely different functions, which contain other domains that are usually enzymatic, membrane transporters or DNA-binding domains. It has been shown that CBS domains regulate enzyme activity based on the concentration of AMP/ATP or other adenosine derivatives, but they may also bind metallic ions such as Mg^{2+} (Scott, Hawley *et al.* 2004; Ishitani, Sugita *et al.* 2008). Pab0961 do not

contain additional domains and is thus likely devoid of enzymatic activity. Standalone CBS domain proteins, found particularly in prokaryotes, might form complexes upon binding to other proteins to which they interact with and regulate.

Interestingly, when we weighted the SR ratio of the MS results with operons and genomic neighbourhood data, GINS23, which is encoded in close vicinity and pulled down Pab0961 with a moderate SR score (SR- :0.23 ; SR+ :0.27), was reclassified as a putative interacting partner of this protein. This interaction between GINS23 and TK1186, the homologue of Pab0961 in *Thermococcus Kodakarensis*, was also observed by Li and co-workers (Li, Santangelo et al. 2010), indicating that this interaction is likely specific. In addition the recently described GINS associated nuclease (GAN) of the RecJ family, used as bait in our AP/MS assay, also co-purified Pab0961, but only in the nuclease free assay. GINS, GAN and MCM form a complex that appears to be a key component of the DNA replication machinery in all Archaea (Makarova, Koonin et al. 2012). This complex is homologous and most likely functionally analogous to the eucaryotic CMG complex.

The fact that this standalone CBS domain protein co-purified with GINS23 subunit in two independent studies and at least in the nuclease free assay with the GAN nuclease raises the hypothesis that Pab0961 might be involved in the regulation of the activity of the archaeal CMG complex. This hypothesis is still in accordance with the genetic analysis performed in *H. volcanii*, since HVO_2384 seems to target replication fork sites where the ternary CMG complex is supposed to serve as the replicative helicase. Therefore, both studies strongly support an implication of Pab0961 and HVO_2384 in the coupling of DNA replication and DNA recombination at replication forks; we thus refer to these proteins as RcrA (recombination coupled to replication protein A).

It has been proposed that prokaryotic stand-alone CBS domain proteins interact with and regulate the activities of effectors proteins by binding adenine nucleotides depending on the energy charge of cells ((King, Lee et al. 2008). Given the central roles of CBS domains in eucaryotic metabolism, it is thus likely that they also play important roles in the cellular physiology of prokaryotes.

A major contribution of the present study is that we reveal for the first time the identity of such potential effectors proteins presumably linking DNA replication and DNA recombination at the replication fork, the activity of which could be regulated in response to the energy charge of cells.

Further genetic and biochemical experiments are now required to better understand the function of this new regulation protein.

CONCLUSIONS

In conclusion, in this report we describe the first protein interaction network of genomic maintenance in Archaea and provide experimental confirmation for some of the associations detected.

In particular, we observed interactions between replisome and DNA recombination components, PCNA-Mre11/Rad50, DNA primase-RadA. These results seem to indicate a close coordination of DNA replication and recombination activities in the maintenance of genome stability of hyperthermophilic Archaea. HA might have evolved a high-fidelity recombination-dependent replication mechanism that can accurately repair the damage, which could thereby explain the lack of translesional polymerases observed in euryarchaea. This interplay between the recombination and replication machineries likely interface with regulatory elements involved in the control of the DNA damage response, as exemplified by the identification of a new factor, RcrA, presumably involved in the coupling of DNA recombination and DNA synthesis at the replication fork.

Further studies are fundamental to unravel the complexities of these interactions and hopefully in turn will solve the mystery of how HA preserve genome integrity in such a harsh environment.

We hope that this promising initial study will provide the corner stone for a deeper understanding of the cellular processes of genomic maintenance are coordinated in Archaea, with the potential to highlight the molecular mechanisms implicated in similar processes in Eucaryotes.

REFERENCES

- Aravind, L. and E. V. Koonin (1998). "Phosphoesterase domains associated with DNA polymerases of diverse origins." *Nucleic Acids Res* **26**(16): 3746-52.
- Atomi, H., T. Imanaka, et al. (2012). "Overview of the genetic tools in the Archaea." *Front Microbiol* **3**: 337.
- Ayyagari, R., X. V. Gomes, et al. (2003). "Okazaki fragment maturation in yeast. I. Distribution of functions between FEN1 AND DNA2." *J Biol Chem* **278**(3): 1618-25.
- Bader, G. D. and C. W. Hogue (2002). "Analyzing yeast protein-protein interaction data obtained from different sources." *Nat Biotechnol* **20**(10): 991-7.
- Banerjee, D., S. M. Mandal, et al. (2011). "Preferential repair of oxidized base damage in the transcribed genes of mammalian cells." *J Biol Chem* **286**(8): 6006-16.

- Beattie, T. R. and S. D. Bell (2011). "Molecular machines in archaeal DNA replication." Curr Opin Chem Biol **15**(5): 614-9.
- Berquist, B. R., P. DasSarma, et al. (2007). "Essential and non-essential DNA replication genes in the model halophilic Archaeon, Halobacterium sp. NRC-1." BMC Genet **8**: 31.
- Boubakri, H., A. L. de Septenville, et al. (2010). "The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo." EMBO J **29**(1): 145-57.
- Boudsocq, F., S. Iwai, et al. (2001). "Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta." Nucleic Acids Res **29**(22): 4607-16.
- Busch, C. R. and J. DiRuggiero (2010). "MutS and MutL are dispensable for maintenance of the genomic mutation rate in the halophilic archaeon Halobacterium salinarum NRC-1." PLoS One **5**(2): e9045.
- Callebaut, I., D. Moshous, et al. (2002). "Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family." Nucleic Acids Res **30**(16): 3592-601.
- Castrec, B., C. Rouillon, et al. (2009). "Binding to PCNA in Euryarchaeal DNA Replication requires two PIP motifs for DNA polymerase D and one PIP motif for DNA polymerase B." J Mol Biol **394**(2): 209-18.
- Clouet-d'Orval, B., D. Rinaldi, et al. (2010). "Euryarchaeal beta-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity." J Biol Chem **285**(23): 17574-83.
- Cohen, G. N., V. Barbe, et al. (2003). "An integrated analysis of the genome of the hyperthermophilic archaeon Pyrococcus abyssi." Mol Microbiol **47**(6): 1495-512.
- Coin, F., V. Oksenysh, et al. (2007). "Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair." Mol Cell **26**(2): 245-56.
- Constantinesco, F., P. Forterre, et al. (2002). "NurA, a novel 5'-3' nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea." EMBO Rep **3**(6): 537-42.
- Constantinesco, F., P. Forterre, et al. (2004). "A bipolar DNA helicase gene, herA, clusters with rad50, mre11 and nurA genes in thermophilic archaea." Nucleic Acids Res **32**(4): 1439-47.
- Creze, C., A. Ligabue, et al. (2012). "Modulation of the Pyrococcus abyssi NucS endonuclease activity by the replication clamp PCNA at functional and structural levels." J Biol Chem.
- Delmas, S., L. Shunburne, et al. (2009). "Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon Haloferax volcanii by restraining homologous recombination." PLoS Genet **5**(7): e1000552.
- Dominski, Z. (2007). "Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism." Crit Rev Biochem Mol Biol **42**(2): 67-93.
- Dorazi, R., D. Gotz, et al. (2007). "Equal rates of repair of DNA photoproducts in transcribed and non-transcribed strands in Sulfolobus solfataricus." Mol Microbiol **63**(2): 521-9.
- Duxin, J. P., H. R. Moore, et al. (2012). "Okazaki fragment processing-independent role for human Dna2 enzyme during DNA replication." J Biol Chem **287**(26): 21980-91.
- Friedel, C. C. and R. Zimmer (2009). "Identifying the topology of protein complexes from affinity purification assays." Bioinformatics **25**(16): 2140-6.
- Fujikane, R., S. Ishino, et al. (2010). "Genetic analysis of DNA repair in the hyperthermophilic archaeon, Thermococcus kodakaraensis." Genes Genet Syst **85**(4): 243-57.
- Godfroy, A., A. Postec, et al. (2006). 4 Growth of Hyperthermophilic Microorganisms for Physiological and Nutritional Studies. Methods in Microbiology, Academic Press. **Volume 35**: 93-108.

- Grogan, D. W. (2004). "Stability and repair of DNA in hyperthermophilic Archaea." Curr Issues Mol Biol **6**(2): 137-44.
- Grogan, D. W., G. T. Carver, et al. (2001). "Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*." Proc Natl Acad Sci U S A **98**(14): 7928-33.
- Groisillier, A., C. Herve, et al. (2010). "MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms." Microb Cell Fact **9**: 45.
- Henneke, G. (2012). "In vitro reconstitution of RNA primer removal in Archaea reveals the existence of two pathways." Biochem J **447**(2): 271-80.
- Henneke, G., D. Flament, et al. (2005). "The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication." J Mol Biol **350**(1): 53-64.
- Henneke, G., Y. Gueguen, et al. (2002). "Replication factor C from the hyperthermophilic archaeon *Pyrococcus abyssi* does not need ATP hydrolysis for clamp-loading and contains a functionally conserved RF-C PCNA-binding domain." J Mol Biol **323**(5): 795-810.
- Hopkins, B. B. and T. T. Paull (2008). "The *P. furiosus* mre11/rad50 complex promotes 5' strand resection at a DNA double-strand break." Cell **135**(2): 250-60.
- Hu, J., L. Guo, et al. (2012). "Template-dependent polymerization across discontinuous templates by the heterodimeric primase from the hyperthermophilic archaeon *Sulfolobus solfataricus*." Nucleic Acids Res **40**(8): 3470-83.
- Ishitani, R., Y. Sugita, et al. (2008). "Mg²⁺-sensing mechanism of Mg²⁺ transporter MgtE probed by molecular dynamics study." Proc Natl Acad Sci U S A **105**(40): 15393-8.
- Kelman, Z. and J. Hurwitz (2000). "A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon *Methanobacterium thermoautotrophicum deltaH*." J Biol Chem **275**(10): 7327-36.
- Kelman, Z. and M. F. White (2005). "Archaeal DNA replication and repair." Curr Opin Microbiol **8**(6): 669-76.
- Kernchen, U. and G. Lipps (2006). "Thermodynamic analysis of the single-stranded DNA binding activity of the archaeal replication protein A (RPA) from *Sulfolobus solfataricus*." Biochemistry **45**(2): 594-603.
- Khil, P. P. and R. D. Camerini-Otero (2002). "Over 1000 genes are involved in the DNA damage response of *Escherichia coli*." Mol Microbiol **44**(1): 89-105.
- King, N. P., T. M. Lee, et al. (2008). "Structures and functional implications of an AMP-binding cystathionine beta-synthase domain protein from a hyperthermophilic archaeon." J Mol Biol **380**(1): 181-92.
- Komori, K. and Y. Ishino (2001). "Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination." J Biol Chem **276**(28): 25654-60.
- Komori, K., T. Miyata, et al. (2000). "Both RadA and RadB are involved in homologous recombination in *Pyrococcus furiosus*." J Biol Chem **275**(43): 33782-90.
- Kusser, A. G., M. G. Bertero, et al. (2008). "Structure of an archaeal RNA polymerase." J Mol Biol **376**(2): 303-7.
- Lao-Sirieix, S. H., L. Pellegrini, et al. (2005). "The promiscuous primase." Trends Genet **21**(10): 568-72.
- Lee, A. M., J. R. Sevinsky, et al. (2009). "Proteomics of *Pyrococcus furiosus*, a hyperthermophilic archaeon refractory to traditional methods." J Proteome Res **8**(8): 3844-51.
- Li, Z., M. Pan, et al. (2011). "A novel DNA nuclease is stimulated by association with the GINS complex." Nucleic Acids Res **39**(14): 6114-23.

- Li, Z., T. J. Santangelo, et al. (2010). "Affinity purification of an archaeal DNA replication protein network." *MBio* **1**(5).
- Lin, Y., C. E. Guzman, et al. (2006). "Methanosarcina acetivorans flap endonuclease 1 activity is inhibited by a cognate single-stranded-DNA-binding protein." *J Bacteriol* **188**(17): 6153-67.
- Lin, Y., L. J. Lin, et al. (2008). "Engineering of functional replication protein a homologs based on insights into the evolution of oligonucleotide/oligosaccharide-binding folds." *J Bacteriol* **190**(17): 5766-80.
- Lindahl, T. (1993). "Instability and decay of the primary structure of DNA." *Nature* **362**(6422): 709-15.
- Liu, H., J. Rudolf, et al. (2008). "Structure of the DNA repair helicase XPD." *Cell* **133**(5): 801-12.
- Lucas-Lledo, J. I. and M. Lynch (2009). "Evolution of mutation rates: phylogenomic analysis of the photolyase/cryptochrome family." *Mol Biol Evol* **26**(5): 1143-53.
- Ma, Y., U. Pannicke, et al. (2002). "Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination." *Cell* **108**(6): 781-94.
- MacNeill, S. A. (2011). "Protein-protein interactions in the archaeal core replisome." *Biochem Soc Trans* **39**(1): 163-8.
- Makarova, K. S., E. V. Koonin, et al. (2012). "The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes." *Biol Direct* **7**: 7.
- Makarova, K. S., A. V. Sorokin, et al. (2007). "Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea." *Biol Direct* **2**: 33.
- Matsunaga, F., C. Norais, et al. (2003). "Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin." *EMBO Rep* **4**(2): 154-8.
- Meslet-Cladiere, L., C. Norais, et al. (2007). "A novel proteomic approach identifies new interaction partners for proliferating cell nuclear antigen." *J Mol Biol* **372**(5): 1137-48.
- Milenkovic, T., V. Memisevic, et al. (2011). "Dominating biological networks." *PLoS One* **6**(8): e23016.
- Motz, M., I. Kober, et al. (2002). "Elucidation of an archaeal replication protein network to generate enhanced PCR enzymes." *J Biol Chem* **277**(18): 16179-88.
- Oksenyich, V., B. Bernardes de Jesus, et al. (2009). "Molecular insights into the recruitment of TFIIH to sites of DNA damage." *EMBO J* **28**(19): 2971-80.
- Palud, A., G. Villani, et al. (2008). "Intrinsic properties of the two replicative DNA polymerases of *Pyrococcus abyssi* in replicating abasic sites: possible role in DNA damage tolerance?" *Mol Microbiol* **70**(3): 746-61.
- Pan, M., L. M. Kelman, et al. (2011). "The archaeal PCNA proteins." *Biochem Soc Trans* **39**(1): 20-4.
- Perugino, G., A. Vettone, et al. (2012). "Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage." *J Biol Chem* **287**(6): 4222-31.
- Poinsignon, C., D. Moshous, et al. (2004). "The metallo-beta-lactamase/beta-CASP domain of Artemis constitutes the catalytic core for V(D)J recombination." *J Exp Med* **199**(3): 315-21.
- Quaiser, A., F. Constantinesco, et al. (2008). "The Mre11 protein interacts with both Rad50 and the HerA bipolar helicase and is recruited to DNA following gamma irradiation in the archaeon *Sulfolobus acidocaldarius*." *BMC Mol Biol* **9**: 25.
- Ren, B., J. Kuhn, et al. (2009). "Structure and function of a novel endonuclease acting on branched DNA substrates." *EMBO J* **28**(16): 2479-89.

- Richard, D. J., S. D. Bell, et al. (2004). "Physical and functional interaction of the archaeal single-stranded DNA-binding protein SSB with RNA polymerase." Nucleic Acids Res **32**(3): 1065-74.
- Robbins, J. B., M. C. Murphy, et al. (2004). "Functional analysis of multiple single-stranded DNA-binding proteins from *Methanosarcina acetivorans* and their effects on DNA synthesis by DNA polymerase BI." J Biol Chem **279**(8): 6315-26.
- Romano, V., A. Napoli, et al. (2007). "Lack of strand-specific repair of UV-induced DNA lesions in three genes of the archaeon *Sulfolobus solfataricus*." J Mol Biol **365**(4): 921-9.
- Rouillon, C., G. Henneke, et al. (2007). "DNA polymerase switching on homotrimeric PCNA at the replication fork of the euryarchaea *Pyrococcus abyssi*." J Mol Biol **369**(2): 343-55.
- Royer, L., M. Reimann, et al. (2008). "Unraveling protein networks with power graph analysis." PLoS Comput Biol **4**(7): e1000108.
- Rudolf, J., V. Makrantonis, et al. (2006). "The DNA repair helicases XPD and FancJ have essential iron-sulfur domains." Mol Cell **23**(6): 801-8.
- Rudolf, J., C. Rouillon, et al. (2010). "The helicase XPD unwinds bubble structures and is not stalled by DNA lesions removed by the nucleotide excision repair pathway." Nucleic Acids Res **38**(3): 931-41.
- Sartori, A. A. and J. Jiricny (2003). "Enzymology of base excision repair in the hyperthermophilic archaeon *Pyrobaculum aerophilum*." J Biol Chem **278**(27): 24563-76.
- Scott, J. W., S. A. Hawley, et al. (2004). "CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations." J Clin Invest **113**(2): 274-84.
- Shin, J. H., T. J. Santangelo, et al. (2007). "Archaeal minichromosome maintenance (MCM) helicase can unwind DNA bound by archaeal histones and transcription factors." J Biol Chem **282**(7): 4908-15.
- Sikorski, T. W., S. B. Ficarro, et al. (2011). "Sub1 and RPA associate with RNA polymerase II at different stages of transcription." Mol Cell **44**(3): 397-409.
- Skowrya, A. and S. A. MacNeill (2012). "Identification of essential and non-essential single-stranded DNA-binding proteins in a model archaeal organism." Nucleic Acids Res **40**(3): 1077-90.
- Spitalny, P. and M. Thomm (2003). "Analysis of the open region and of DNA-protein contacts of archaeal RNA polymerase transcription complexes during transition from initiation to elongation." J Biol Chem **278**(33): 30497-505.
- Stano, N. M., J. Chen, et al. (2006). "A coproofreading Zn(2+)-dependent exonuclease within a bacterial replicase." Nat Struct Mol Biol **13**(5): 458-9.
- Stracker, T. H. and J. H. Petrini (2011). "The MRE11 complex: starting from the ends." Nat Rev Mol Cell Biol **12**(2): 90-103.
- Teplyakov, A., G. Obmolova, et al. (2003). "Crystal structure of the *Escherichia coli* YcdX protein reveals a trinuclear zinc active site." Proteins **51**(2): 315-8.
- von Mering, C., R. Krause, et al. (2002). "Comparative assessment of large-scale data sets of protein-protein interactions." Nature **417**(6887): 399-403.
- Waage, I., G. Schmid, et al. (2010). "Shuttle vector-based transformation system for *Pyrococcus furiosus*." Appl Environ Microbiol **76**(10): 3308-13.
- Warbrick, E. (2000). "The puzzle of PCNA's many partners." Bioessays **22**(11): 997-1006.
- White, M. F. (2009). "Structure, function and evolution of the XPD family of iron-sulfur-containing 5'→3' DNA helicases." Biochem Soc Trans **37**(Pt 3): 547-51.

Yellaboina, S., A. Tasneem, et al. (2011). "DOMINE: a comprehensive collection of known and predicted domain-domain interactions." Nucleic Acids Res **39**(Database issue): D730-5.

Yutin, N., K. S. Makarova, et al. (2008). "The deep archaeal roots of eukaryotes." Mol Biol Evol **25**(8): 1619-30.

5.11 RESULTATS COMPLEMENTAIRES SUR L'INTERACTION RPA-RNAP

Dans cet article, nous avons présenté un réseau d'interaction de la maintenance génomique de *Pyrococcus abyssi*. Au sein de ce réseau, nous avons identifié plusieurs complexes protéiques de la maintenance génomique déjà établis et plusieurs nouveaux complexes dont certains ont été confirmés indépendamment. L'interaction entre l'ARN polymérase et RPA est observée en AP-MS et est appuyée dans le réseau par un recouvrement des clusters défini par ces deux protéines. Afin de déterminer quelle est la fonction biologique de cette interaction nous avons étudié l'effet de RPA sur la transcription, *in vitro*.

Pour cela, j'ai bénéficié de la mise en place d'une collaboration entre notre laboratoire et celui de Michael Thomm à l'Université de Regensburg, pour aborder cette question. En utilisant le RPA de *P. abyssi* produit au laboratoire, ces derniers ont pu montrer que RPA stimulait l'activité transcriptionnelle de la RNA polymérase. Cette activation est décrite dans l'article que nous allons soumettre. Cependant, pour essayer de mieux cerner le rôle du RPA deux expériences supplémentaires ont été réalisées qui permettent de confirmer que le RPA stimule la transcription mais qu'il n'a d'effets ni sur l'initiation ni sur la terminaison. Ces données suggèrent un rôle régulateur de la phase d'élongation de synthèse des ARN. Les résultats de ces deux expériences supplémentaires sont présentés en figures 53 et 54.

5.12 RPA N'INFLUENCE PAS L'INITIATION DE LA TRANSCRIPTION, IN VITRO

En utilisant un ADN contenant le promoteur *gdh* comme matrice, un dinucléotide 5'-GpU-3' comme primer, et des ribonucleotides radio-marqués, nous avons testé la capacité de l'ARN polymérase à incorporer un résidu [α -³²P]-UTP au niveau du primer dinucléotide GU, en absence ou présence de RPA. Cette expérience de transcription avortée permet de déterminer l'influence de facteurs sur l'initiation de la transcription. La figure 53 montre que le RPA ne modifie pas le niveau d'incorporation des nucléotides marqués (Figure 53 pistes 1, 2 et 3). Nous en concluons que RPA est incapable de stimuler l'initiation de la transcription, *in vitro*.

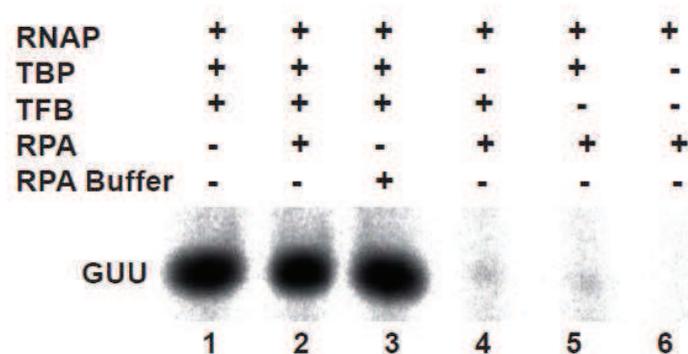


Figure 53 : Effet de RPA sur l'initiation de la transcription *in vitro*

Synthèse d'un transcrit 3 nt abortif dans une réaction de transcription *in vitro*. Les produits ARN sont séparés en gel d'acrylamide 28%. RNAP : ARN polymérase. TBP : TATAbox Binding Protéine. TFB : Transcription factor B.

Nous montrons aussi que le RPA ne peut pas se substituer aux facteurs d'initiation TBP et TFB (Figure 53, pistes 4, 5 et 6), contrairement à ce qui est observé avec le SSB *Crenarchaea* (Richard D.J., et al. 2004). Comme RPA ne possède pas d'effet sur l'initiation, nous avons étudié l'effet de RPA sur la terminaison de la transcription.

5.13 RPA N'INFLUENCE PAS LA TERMINAISON DE LA TRANSCRIPTION, IN VITRO

Afin de tester cette hypothèse, nous avons étudié le profil de terminaison de transcription du gène de l'histone H1 en fonction de la présence de RPA. Le mécanisme de terminaison pour ce substrat est apparenté à celui de l'ARN polymérase II eucaryote et chez *P. furiosus*, cette terminaison est dépendante de la température (Spitalny and Thomm 2008). A 80°C, les terminaisons intermédiaires sont quasiment toutes abolies et la majorité des ARN sont transcrits en entier. A 90°C, la plupart des transcrits sont terminés à la première des séquences de terminaison (T1). La figure 54 montre que, en présence de RPA, la quantité de transcription augmente globalement mais qu'il n'influence pas les profils de terminaisons, et cela pour les deux températures testées. De plus cette expérience permet de démontrer que l'effet de stimulation observé n'est pas particulier au promoteur *gdh* puisque nous avons utilisé une matrice différente dans cette expérience ; le promoteur de l'histone A1.

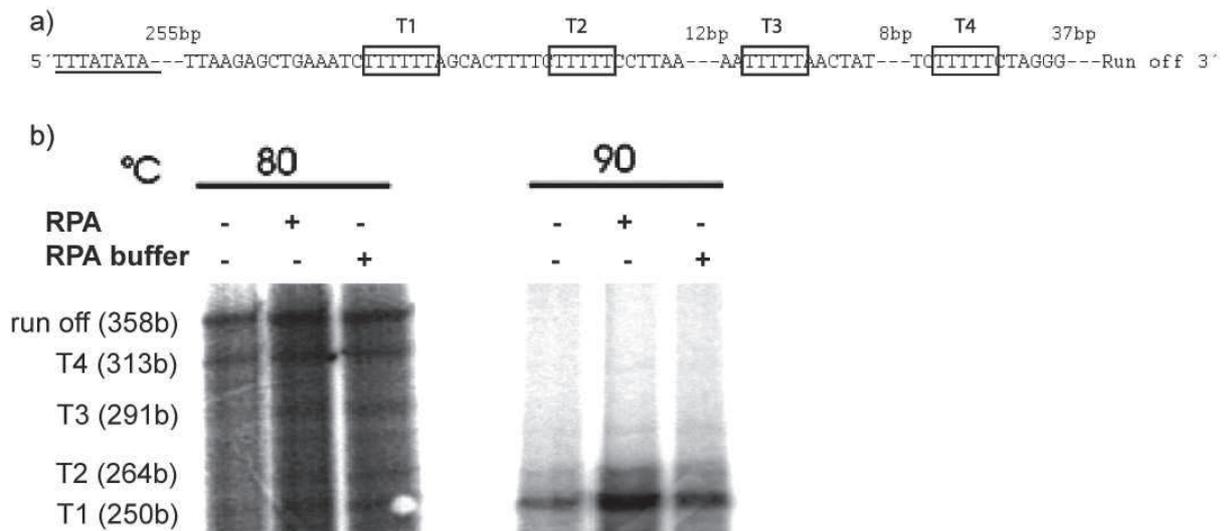


Figure 54 : Effet de RPA sur la terminaison de la transcription.

a) Séquence de la matrice ADN_{hpyA1}. En 5' le site d'initiation est souligné, les 4 sites de terminaison 3'T1 à T4 sont représentés par un rectangle. b) Transcription, *in vitro*, de la matrice hpyA1 à 80 et 90 °C. La longueur des transcrits T1-4 obtenus est indiquée à gauche.

La description de l'interaction entre la RNAP et RPA est inédite chez les Archaea. En revanche, cette interaction a été identifiée récemment chez les Eucaryotes (Sikorski, Ficarro et al. 2011). Les auteurs de cette étude observent notamment une co-précipitation de RPA avec la RNAP et une localisation de RPA aux régions transcrites de l'ADN. Ils proposent finalement que le RPA intervienne au niveau des bulles de transcription, pendant la phase d'élongation, en tapissant le brin opposé au brin transcrit.

Dans le modèle proposé par ces auteurs, après la transition entre la phase d'initiation et la phase d'élongation, le RPA pourrait s'associer avec le brin non transcrit de la bulle de transcription créée par le complexe d'élongation (Figure 55).

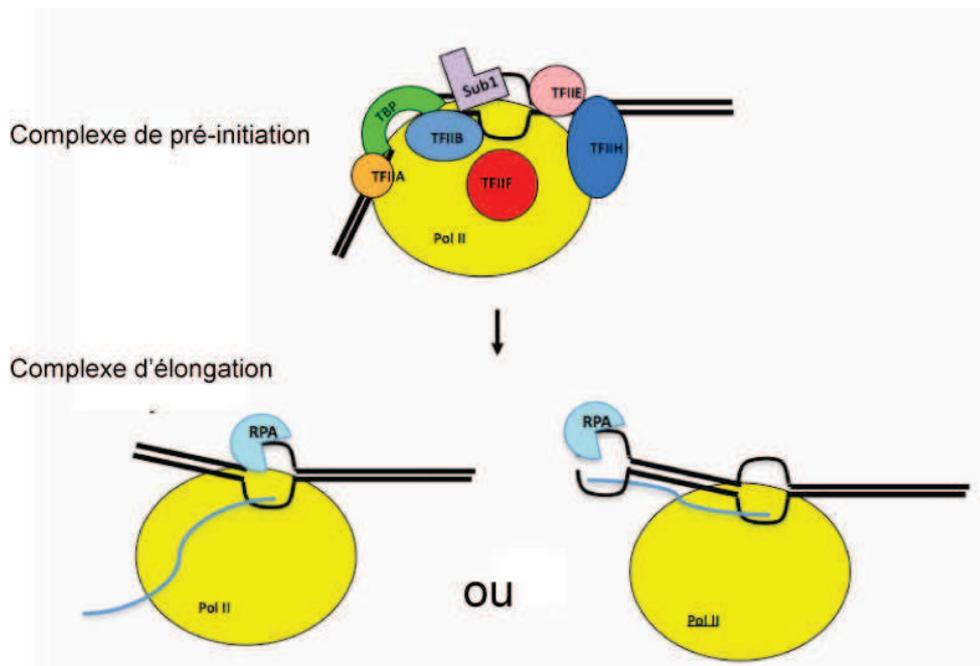


Figure 55: modèle de l'interaction entre RPA et la RNA polymérase

L'hypothèse principale est que la présence de RPA sur le brin non-transcrit au niveau de la bulle de transcription protège l'ADN exposé (environ 5bp) et améliore l'ouverture de la bulle de transcription en amont. Une autre hypothèse pouvant expliquer la présence du RPA au niveau du complexe d'élongation est que cette association pourrait atténuer ou empêcher des réactions de recombinaison couplées à la transcription. En effet, la transcription peut entraîner de tels événements, probablement par la formation de R-loop (structure hybride DNA/RNA, voir modèle ci-dessus) entre le transcrit RNA et le brin transcrit DNA, en aval du complexe d'élongation, c'est-à-dire de manière indépendante de la bulle de transcription. Le brin non-transcrit est dans ce cas probablement l'agent recombino-génique, et la fixation de RPA pourrait empêcher l'étape d'invasion de brins.

Nos données suggèrent que le RPA Archaea agit plutôt pendant la phase d'élongation. Si cela est vrai, il est probable que le mécanisme d'action est différent de celui proposé chez les Eucaryotes car nous avons démontré au laboratoire que la fixation du RPA nécessite une taille d'ADN simple brin d'au moins douze nucléotides et donc, si la taille du brin non-transcrit est bien similaire à celui exposé dans le complexe d'élongation eucaryote, le RPA ne pourrait s'y fixer.

Afin de caractériser plus précisément la fonction du RPA dans la régulation de la transcription, d'autres expériences sont donc nécessaires. La caractérisation fonctionnelle du complexe chez les Archaea pourrait

fournir d'importants éléments de compréhension sur le rôle biologique des protéines de liaison à l'ADN simple brin au cours de la transcription et de son couplage avec la maintenance génomique.



6.1.1 PROBLEMATIQUE

Nous avons observé, et ce dès la mise au point avec le PCNA, que le seuil de significativité strict que nous avons dû établir est probablement un peu élevé et certainement imparfait. Ainsi il est probable que plusieurs interactions biologiquement significatives aient été éliminées par ce seuil établi empiriquement. Afin d'explorer plus en profondeur les données d'AP-MS, une version étendue du réseau a été construite. Ce réseau est construit sur une base SR moins stricte ; une interaction est intégrée au réseau dès lors qu'elle apparaît au moins une fois avec un SR- inférieur à 0,3.

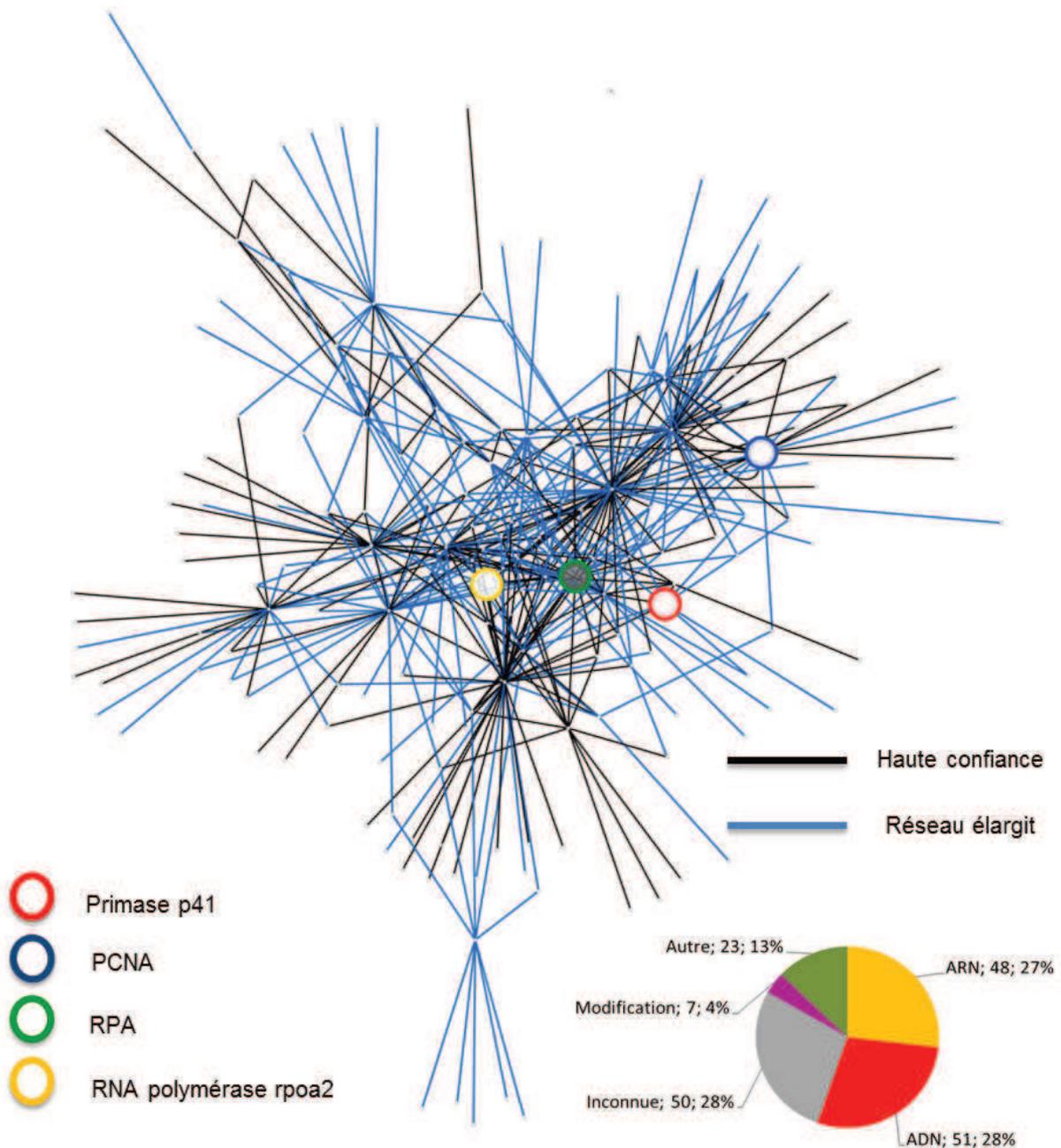


Figure 56: Réseau étendu et position des clusters haute confiance.

Les interactions du réseau présentées dans l'article sont en noir, les interactions supplémentaires identifiées avec une confiance moindre sont représentés en bleu. Les centres des clusters hautes confiances sont représentés en gros et en couleur. Ces 4 clusters sont toujours des centres de clusterisation dans ce réseau (données non présentées).

Ce seuil permet d'inclure les interactants dont le comportement a été modifié par les nucléases ainsi que les protéines dont les scores affleurent le seuil de significativité établi arbitrairement. Les paramètres de couverture ne sont plus éliminatoires et le nombre de peptides est abaissé à une tolérance de 3 peptides uniques. Un score MASCOT supérieur à 100 reste un paramètre essentiel. Dans sa version étendue le réseau contient 179 nœuds et 403 interactions (Figure 55). Seul le couple MCM – XPD reste isolé, le reste des interactions formant une seule et même structure. Afin de ne pas abuser de complexité, je ne présenterai du réseau étendu que quelques exemples de sous graphes au sein desquels la validité des interactions est supportée par l'ontologie des partenaires protéiques. En utilisant l'algorithme « opéron », conçu pour cette exploration du réseau étendu, j'ai identifié les protéines dont la position dans le génome est corrélée aux interactions observées. Les paires de protéines codées à moins de 5000 bases l'une de l'autre sont présentées dans le tableau 10 en fonction de leur distance, du brin sur lequel elles sont codées et de leurs connections dans le réseau. Ce faisant, l'algorithme « opéron » met en évidence la relation entre l'organisation du génome et l'organisation de certains complexes protéiques. En effet, il est bien établi que l'expression des fonctions biologiques est intrinsèquement liée à l'organisation du génome. Par exemple, les opérons procaryotes permettent un certain niveau de co-régulation des éléments codés en son sein. Même si cela n'est pas systématique, on observe que les sous unités de complexes protéiques sont régulièrement codés dans le même opéron. L'assemblage d'un complexe protéique lambda sera facilité si ses éléments sont synthétisés au même endroit et au même moment. L'algorithme « opéron » met en évidence plusieurs corrélations de proximité génétique entre des partenaires de la maintenance génomique dans le réseau (Tableau 10).

Tableau 10 : Résultats de l'algorithme « operon » :

Protéine 1	Protéine 2	Brin gene 1	Brin gene2	Distance (pb)	Partenaire commun
pab1035	pab1036	+	-	75	NucS
pab0961	pab1421	+	-	994	non
XPB/ERCC3	Bax1	-	+	1037	RPA : NucS
MCM	pab0961	+	+	1568	non
pab0431	NatA-likeABC - pab0434	+	+	2080	non
ADN polymerase II	NucS	-	-	4073	non
RFClS	Q9V2F8 pab2271	+	-	4118	NucS
RNAPrhoE1	UncharacterizedHTH PYRAB16920	+	-	4722	AlkA
RFClS	RadB	+	-	4848	non
Gins23	pab0961	+	+	4899	non

Les paires de protéines identifiées sont présentées en fonction de la distance en paire de bases de leur séquences codantes respectives. Les couleurs surlignées correspondent aux complexes présentés en figures 57, 58 et 59.

NucS interagit avec les deux protéines du NER, Bax1 et XPB, qui sont codées de façon juxtaposées et dans le sens opposé. Le RF-C semble quant à lui entretenir une relation privilégiée avec NucS et pab 2271 (Q9V2F8). D'autre part, nous avons décrit dans l'article la protéine RcrA (Pab0961 dans le tableau ci-dessus), qui semble interagir avec RadB, GINS23 et la nucléase GAN, or, le gène codant cette protéine est présent dans un même

locus comprenant également les gènes GINS23, MCM et Q9UYR2 (Pab1421 dans le tableau 10). Cette dernière protéine, de fonction inconnue, interagit également avec GINS23 dans le réseau étendu. Il est possible que cette nouvelle protéine révélée par l'analyse du réseau étendu et des informations de proximité génomique puisse être un nouvel acteur de la régulation de la réplication de l'ADN associé à la protéine RcrA. Les corrélations les plus évidentes sont observées entre des protéines de la réparation codées sur des locus génétiques spécifiques à la maintenance génomique. Ainsi, ces protéines pourraient constituer des éléments important de complexes de la réplication et/ou de réparation de l'ADN. Les sous graphes identifiés grâce à l'algorithme « opéron » sont détaillés dans les figures 57 à 58. Ils permettent notamment de décrire un système NER potentiel (Figures 56) ainsi que de nouveaux partenaires de la protéine RcrA (Figure 59).

6.1.2 NUCLEOTIDE EXCISION REPAIR

Une des principales questions, toujours sans réponse, chez les archaea hyperthermophiles, est l'existence de la voie de réparation par excision de nucléotides (NER) qui permet d'éliminer les adduits sur l'ADN. Aucune des protéines responsables de la reconnaissance des lésions pour cette voie de réparation n'a été trouvée à ce jour chez les Archaea (Rouillon and White, 2011). Le système NER n'est pas encore complètement compris, cependant la proximité phylogénétique des protéines Archaea avec leurs homologues Eucaryotes est telle que des analyses biochimiques et structurales pourraient avoir un impact direct sur la compréhension des phénotypes de mutations chez l'homme.

Si ce métabolisme est bien conservé chez les Eucaryotes, les Bactéries et les archaea mésophiles, il n'y a à ce jour que très peu d'évidences expérimentales de l'existence d'un tel mécanisme chez les archaea hyperthermophiles (HA). Cependant, la plupart de ces dernières possèdent les gènes codant pour les protéines du NER de type eucaryote, ce qui suggère que les HA utiliseraient une version simplifiée du NER eucaryote.

Cette hypothèse a été formulée par l'équipe de Malcolm White, suite à de nombreux travaux sur les homologues archaea des protéines du NER. Lui et ses collaborateurs ont pu montrer que les activités hélicases de XPD et XPB permettent l'ouverture du double brin portant la lésion, le brin endommagé est ensuite incisé par les nucléases 3' flap et 5' flap, XPF et Bax1. D'autre part, il a également suggéré les participations hypothétiques du SSB, équivalent du RPA chez les Crenarchaea, pour la détection des lésions, et de l'endonucléase NucS pour la coupure du 3' flap (Figure 23) (Roberts, Bell et al. 2003; Cubeddu and White 2005; Newman, Murray-Rust et al. 2005; Roberts and White 2005; Roberts and White 2005; Rudolf, Makrantonis et al. 2006; Hutton, Roberts et al. 2008; Liu, Rudolf et al. 2008; Richards, Cubeddu et al. 2008; White 2009; Hutton, Craggs et al. 2010; Rouillon and White 2010; Rudolf, Rouillon et al. 2010; Rouillon and White 2011).

L'interaction de NucS et de RPA avec la quasi totalité des acteurs du NER, dans le réseau étendu, argumente en faveur de la participation de ces deux composants dans cette voie métabolique de la réparation de l'ADN (Figure 54A). En accord avec cette hypothèse, une revue récente de Christophe Rouillon propose que NucS soit un nouvel acteur du NER, étant donné qu'elle présente une spécificité de substrat similaire à celle de Hef/XPF (dénommée ERCC4/XPF dans le réseau ci-dessous) et clive l'ADN à la jonction entre le simple et le double brin

(Rouillon and White 2011). Le RPA semble également interagir avec les protéines du NER ; dans ce contexte, son rôle pourrait être la détection des lésions dans l'ADN. Plusieurs études montrent que le RPA, ainsi que le SSB (Single Strand Binding protein), l'équivalent du RPA chez les Crenarchaea, interviennent probablement dans la réponse aux dommages (McCready, Muller et al. 2005; DeVeaux, Muller et al. 2007). De plus, le SSB de *Sulfolobus solfataricus* est capable de reconnaître spécifiquement une lésion au sein d'un duplex d'ADN (Cubeddu and White 2005). Cette fonction n'a pas encore été démontrée pour le RPA des Euryarchaea, plus proche du RPA eucaryote que le SSB. Si cette hypothèse s'avère juste, cette fonction de détection de lésions pourrait expliquer l'absence de gènes codant pour la détection de lésions de type NER dans le génome des Archaea.

Ce tableau général des acteurs potentiels du NER chez les HA est donc cohérent avec le graphe d'interaction de la figure 54. Les deux activités 5'→3' et 3'→5' ADN hélicases, portées respectivement par la protéine Hef/XPF et le complexe XPB/Bax1, sont associées au facteur RPA, ce qui est cohérent avec l'hypothèse que le RPA pourrait détecter la lésion et initier une ouverture locale de la double hélice, facilitant ainsi le positionnement des hélicases, pour continuer à ouvrir le double brin jusqu'à une taille critique déclenchant l'excision du brin endommagé. Les nucléases Bax1 et NucS, ainsi que Hef/XPF et NucS peuvent prendre en charge les substrats 5' et 3' flap générés de part et d'autre de la lésion, respectivement.

En plus de ces données sur l'interaction physique entre ces composants, nous avons également obtenu des données sur leurs environnements génétiques. L'algorithme opéron a permis de montrer que les séquences codantes pour XPB et Bax1 sont adjacentes dans le génome (Figure 55C). De plus, les homologues de RPA et XPF sont codés à proximité. Même s'il est impossible de conclure directement sur le sens biologique de cette observation, la corrélation entre la position génomique et les complexes observés est frappante. Il est raisonnable de penser que la co-localisation est une signature fonctionnelle et qu'elle permet une co-régulation, par exemple transcriptionnelle. Il est important de noter que *xpb* et *bax1* ne sont codés, ni consécutivement, ni sur le même brin. Ils n'appartiennent pas au même transcrit. Cette orientation pourrait permettre à la cellule de synthétiser les transcrits XPB et Bax1 avec une ouverture minimale de l'hélice d'ADN.

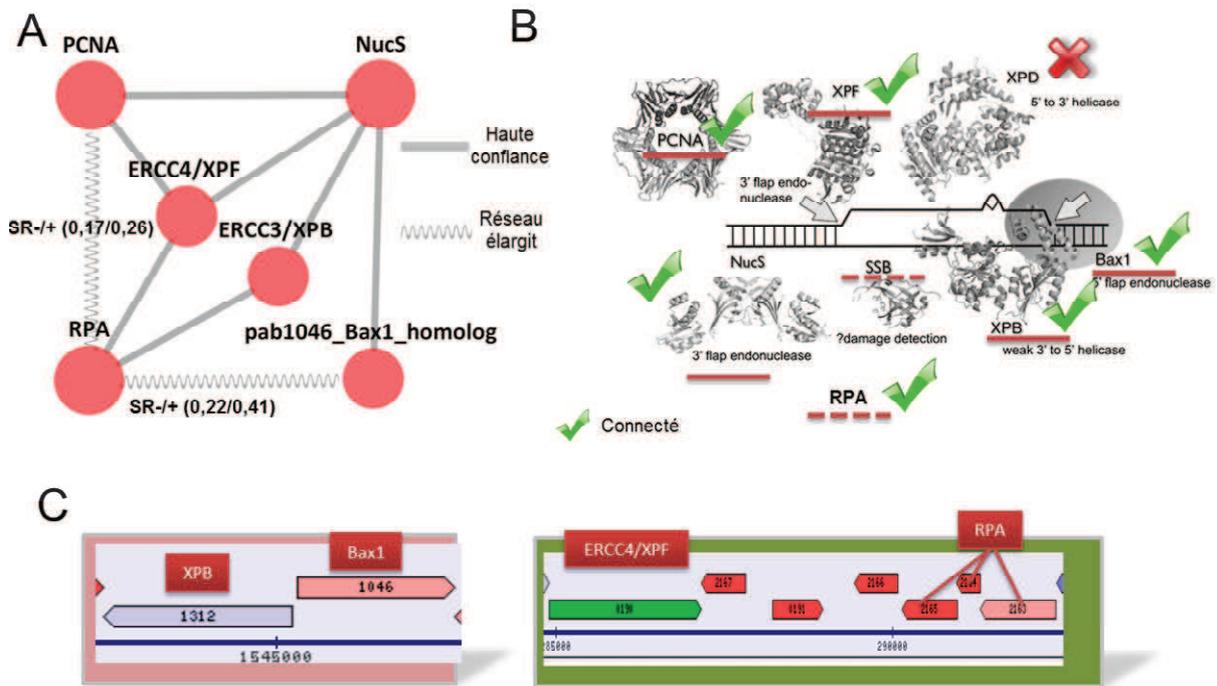


Figure 57: Acteurs établis du NER Archaea dans le réseau.

A) Protéines du NER et leurs interactions au sein du réseau élargi. Les interactions appartenant au réseau élargi uniquement sont signalées par des traits sinués. B) Modèle du NER Archaea (Rouillon et White 2010). Les protéines connectées dans le sous graphe sont signalées par un signe de validation vert. C) Position des gènes xpb, bax1, xpf et rpa14, 32,41 dans le génome de *P. abyssi* (<http://www-archbac.u-psud.fr/>).

NucS est très nettement observé dans un contexte de réparation. Il est donc possible que d'autres protéines de la réparation soient présentes dans l'environnement de NucS. L'algorithme « opéron » identifie une protéine potentiellement impliquée dans la réparation de l'ADN ; Q9UYD9 (pab1036) (Figure 55). Pab1036, dont les scores SR avec NucS sont tous les deux au-dessus de 0,3, est codée en opéron avec pab1035, laquelle serait impliquée dans la réparation de l'ADN (cf l'article). Même si aucune homologie ou signature de domaine n'a été identifiée, nos données suggèrent que pab1036 pourrait intervenir dans la réparation de l'ADN, en collaboration avec pab1035.

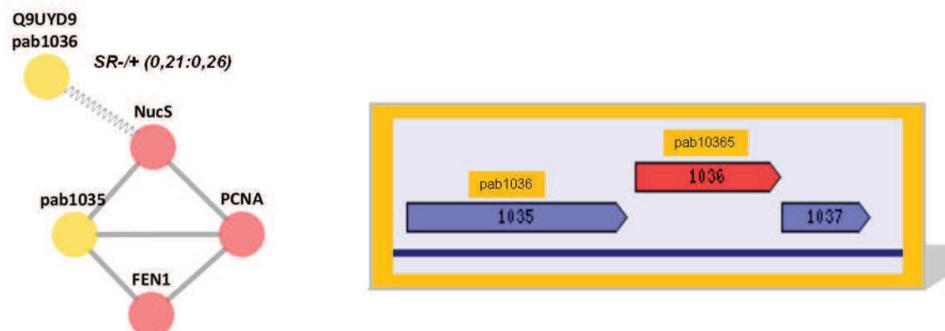
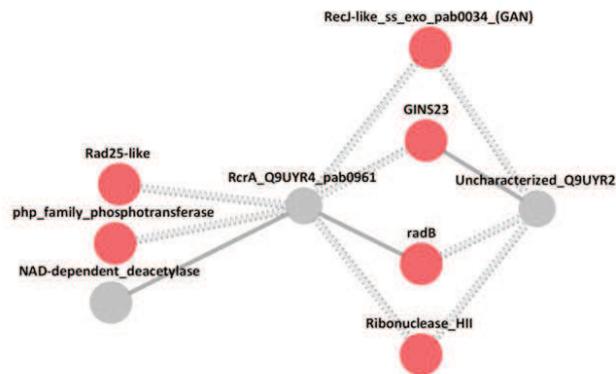


Figure 58: Pab1036 nouvelle nucléase associée à la réparation ?

6.1.3 RCRA ET SES COMPLEXES.

La protéine Q9UYR4 (pab0961) est certainement l'une des découvertes les plus intéressantes de mes recherches, au regard des interactions physiques observées lors de nos travaux, de ceux de l'équipe de Z. Kelman (Li, Santangelo et al. 2010) et des données génétiques obtenus chez *H. volcanii* par le laboratoire de T. Allers. La protéine codée par pab0901 et HVO2384 a été nommée RcrA (recombination coupled to replication protein A), signifiant son implication dans la recombinaison associée à la réplication. Cette protéine possède 2 paires de domaines CBS formant deux domaines Bateman. La structure cristalline de l'homologue de RcrA chez *M. Jannashii* a été résolue et s'est révélée être un homologue très conservé de l'AMPK- γ Eucaryote (Gomez-Garcia, Oyenarte et al.). Le rôle de cette protéine a été discuté dans l'article. Au sein des données haute confiance cette protéine interagit avec RadB et la deacetylase pab0801. Cependant, lorsque l'on observe cette protéine dans son contexte étendu, on observe que ; *primo*, RcrA possède 4 partenaires communs à Q9UYR2 (pab1421) ; RadB (pab2270), Gins23 (pab0956), GAN (pab0034) et la RNAsell (pab0352) (Figure 57A), *secundo*, RcrA est codé à proximité de pab1421, de Gins23 et de MCM (Figure 57B). Ces données renforcent les hypothèses émises sur RcrA. Son implication dans la régulation de la réplication pourrait faire intervenir le complexe GINS, sa nucléase associée GAN et l'hélicase répllicative, cette fois sur la base du contexte génomique. La figure 56A montre bien à quel point la protéine Pab1421 (Q9UYR2) a des partenaires communs avec RcrA. Cette protéine pourrait donc également s'associer à RcrA au sein d'un complexe d'ordre supérieur contenant les partenaires décrits plus haut. La fonction de cette protéine est inconnue, cependant elle présente des homologies avec une protéine bactérienne de la famille recX impliquée dans la régulation de la recombinaison, ainsi qu'un domaine de liaison à l'ADN putatif en C-terminal, de type OB fold, identifié par l'algorithme HHpred via une recherche d'homologie de séquence secondaire. Sur cette base, nous pensons donc que Pab1421 est également une cible très intéressante, qu'il faudra inclure dans les études futures qui seront menées pour élucider la fonction de RcrA.

A



B

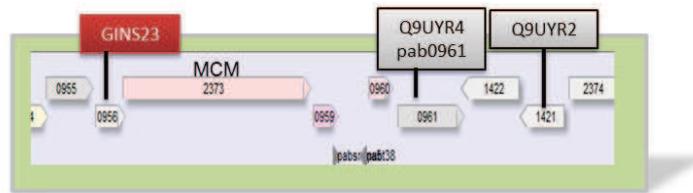


Figure 59 : RcrA et ses complexes.

A) Partenaires des protéines RcrA et Q9UYR2 (pab1421). 4 protéines ont ces deux protéines en partenaires commun. Gins23, la nucléase associée au GINS, :GAN, radB et la RNase H2. Ces interactions étaient pour la plupart sensibles aux nucléases B) Position des gènes gins23, pab0961 et pab1421. A

Ce rapide aperçu du réseau élargi est extrêmement encourageant. Il indique que les données recueillies en MS sont d'une grande qualité et qu'elles recèlent encore certainement d'autres complexes protéiques essentiels de la maintenance génomique de *P. abyssi*.

7 CONCLUSIONS ET PERSPECTIVES

A l'issue de ce travail, j'ai pu mettre au point une méthode robuste pour la recherche de partenaires de protéines hyperthermostables. La simplicité de la méthode permet de purifier et d'identifier les partenaires de n'importe quelle protéine recombinante possédant un tag HIS ou GST. Ceci nous a permis de mettre en évidence un réseau d'interaction des protéines de la maintenance génomique chez l'Archaea hyperthermophile *Pyrococcus abyssi*. Ce réseau, même s'il est incomplet, intègre les fonctions de réplication, de recombinaison, de réparation et de transcription de l'ADN. L'analyse topologique a permis d'identifier les paramètres topologiques essentiels et d'en déduire ses interactions les plus significatives (Figure 60). Nous n'avons pas observé de cloisonnement des composants en fonction de voies métaboliques précises, par contre plusieurs acteurs de la réplication de l'ADN s'associent avec des protéines de la réparation, ce qui semble indiquer une forte coopération entre les fonctions de réplication et de réparation. Ceci renforce l'idée que la machinerie moléculaire en charge de la réplication de l'ADN joue un rôle crucial dans le maintien de la stabilité génomique. D'autre part, nous avons également observé des interactions entre les protéines de la maintenance génomique et celles de la transcription, principalement l'ARN polymérase, posant la question de l'existence d'un mécanisme de réparation couplé à la transcription chez les Archaea.

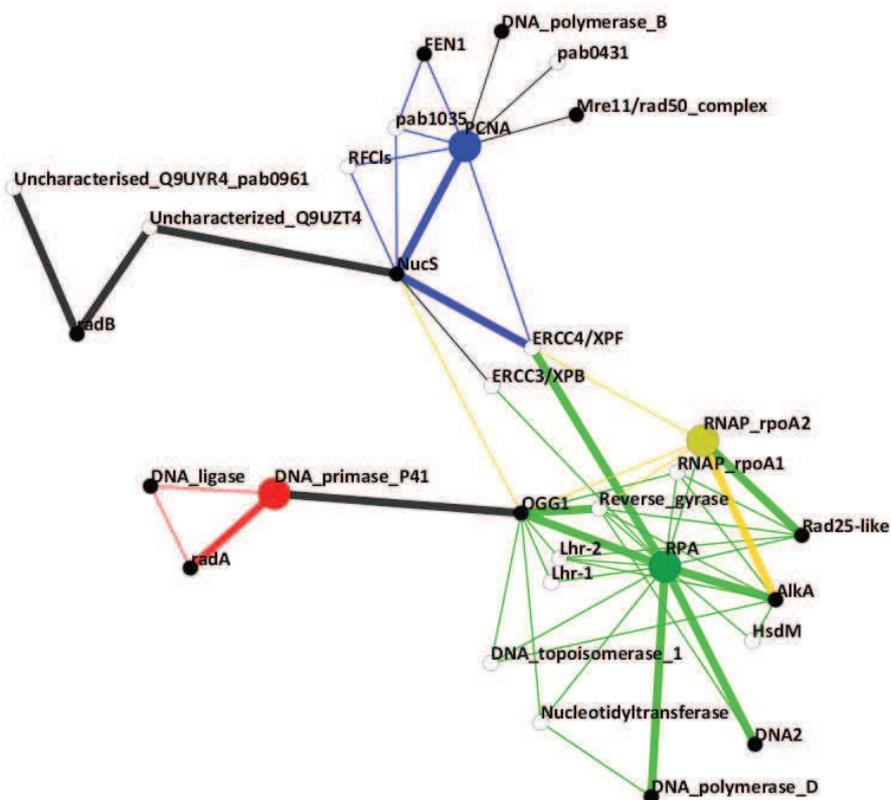


Figure 60: Union des clusters et de l'épine dorsale du réseau

Contrairement au programme initial de la thèse, je n'ai pas pu réaliser de caractérisation fonctionnelle d'un ou plusieurs des complexes identifiés par cette approche. La raison à cela est un manque de temps dû aux délais

d'obtention des résultats de spectrométrie de masse (12 mois cumulés), cependant j'ai pu mettre à profit ces temps d'attente pour me former à la manipulation et au concept des graphes pour l'analyse des réseaux d'interactions. L'étude topologique du réseau a permis de mettre en évidence la forte connexion des protéines RPA et PCNA avec des protéines de la maintenance génomique suggérant leurs implications dans de nombreux complexes essentiels. La recherche des co-occurrences a également permis d'aller plus loin dans l'analyse du réseau pour proposer des interactions entre composants qui ne sont pas directement reliés. Enfin, l'utilisation de l'algorithme permettant l'intégration au réseau des données de proximité génomique a permis d'aller rechercher dans le jeu de données de moindre confiance des interactions dont la signification biologique semble cohérente.

Au final ce travail est donc resté très exploratoire, néanmoins il ouvre de nombreuses perspectives de recherche pour étudier, sur le plan fonctionnel, la machinerie moléculaire en charge de la conservation de l'intégrité du génome.

Nouveaux complexes de la réplication, réparation et recombinaison de l'ADN.

La protéine PCNA est décrite comme le *maestro* de la fourche de réplication. Dans notre réseau, PCNA constitue effectivement un centre de clusterisation hautement connecté avec des protéines de la réplication, du NER et de la recombinaison.

Le complexe Mre11 : Rad50 interagit directement avec le PCNA. Ce complexe pourrait être impliqué dans la réparation des cassures double brin générées par un effondrement de la fourche de réplication. Ces événements peuvent se produire lorsque la progression de la fourche est interrompue par la présence d'une lésion simple brin en amont de la fourche ou de protéines sur l'ADN. En interagissant avec le PCNA, le complexe Mre11 : Rad50 pourrait être recruté à ces cassures cassures double brins et initierait leur réparation par recombinaison homologue. Il est également possible que le complexe MR soit constitutivement présent à la fourche de réplication et qu'il la stabilise afin de prévenir l'effondrement de la fourche lorsque ces mêmes cassures double brins sont générées par le passage du replisome. Cette dernière hypothèse est supportée par le fait qu'il existe une colocalisation de Mre11 et du PCNA au foci de réplication pendant la phase S chez les Eucaryotes (Mirzoeva and Petrini 2003). L'observation de ce complexe peut également signifier une implication du PCNA lors de la résection des fragments d'ADN lors de la recombinaison homologue. En défaveur de cette dernière hypothèse, il faut noter que les protéines HerA et NurA, connues pour participer au mécanisme de résection de brin en présence du complexe Mre11 : Rad50, n'ont pas été trouvées dans nos essais AP/MS quelque soit l'appât utilisé, le complexe Mre11 : Rad50 ou le PCNA.

Les perspectives concernant la caractérisation de cette interaction sont assez directes pour l'instant. Nous envisageons de tester l'action du PCNA dans des tests nucléases *in vitro* en présence du complexe Mre11 : Rad50. Il faudra ainsi tester une modulation possible des activités endo- et exonucléase du complexe par le PCNA sur des substrats ADN simple et double brin. Dans ce but, les activités nucléases du complexe Mre11 : Rad50 de *P. abyssi* ont été étudiées, malheureusement sans succès, nous suspectons que la présence du tag His

en N-terminal de la sous unité Mre11 en est la cause. Pour remédier à cela, nous venons de recevoir de la part de Tanya Paull (Université d'Austin, Texas) le clone permettant de produire le complexe Mre11 : Rad50 de *P. furiosus* dont le tag Histidine est placé en C-terminal de la sous unité Mre11 et dont l'activité nucléase est bien décrite (Hopfner, Karcher et al. 2000; Hopfner, Craig et al. 2002).

La deuxième connection entre réplication et recombinaison est observée au sein du cluster formé par p41, RadA et l'ADN ligase. P41 est la sous unité catalytique de la primase, protéine en charge de la polymérisation du primer lors de la réplication et RadA une protéine de la famille Rec associée à la recombinaison homologue. Le rôle de l'ADN primase Archaea dans la réparation a été suggéré *in vivo* (Le breton et al. 2007) et *in vitro* (Hu, Guo et al. 2012). En revanche, aucun mécanisme n'est précisément décrit. Les données que je présente suggèrent que la primase, radA et l'ADN ligase pourrait collaborer lors d'un mécanisme de réparation dépendant de la recombinaison. RadA serait en charge de la recherche d'homologies, grâce à son activité d'invasion de brins, l'ADN primase pourrait ensuite synthétiser des matrices discontinues et l'ADN ligase lierait les fragments à la fin d'un tel processus (Figure 61). Dans un tel mécanisme la sous unité catalytique aurait un rôle analogue à celui de la polymérase δ Eucaryote impliquée dans la synthèse de l'ADN lors des événements de recombinaison (Maloisel, Fabre et al. 2008). En soutien à cette hypothèse, seule la sous unité catalytique de l'ADN primase est présente dans ce sous-réseau, hors nous avons déjà montré au laboratoire que cette sous unité, en absence de la sous unité régulatrice, n'est capable d'incorporer que des résidus dNTP (Le Breton, Henneke et al. 2007).

Dans ce contexte, le recrutement de la primase pourrait être médié par une interaction avec RadA. En support de cette hypothèse, la protéine RadB, protéine médiatrice de la recombinaison homologue, est également connectée à l'ADN ligase. RadB serait impliquée dans la régulation de la polymérisation du filament RadA, mais sa fonction reste mal définie.

On ne peut pas cependant écarter l'hypothèse que ce cluster d'interactions pourrait suggérer à nouveau, comme dans le cas de l'interaction PCNA : Mre11/Rad50, une forte implication des événements de recombinaison pour réparer les dommages rencontrés par la fourche de réplication ou pour recruter les éléments du réplisome pour le redémarrage de la réplication.

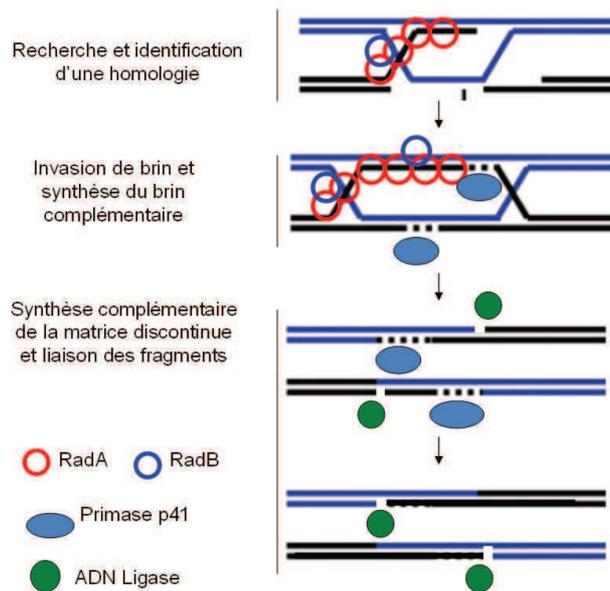


Figure 61: Modèle de la synthèse des brins complémentaires par l'ADN primase lors de la recombinaison.

Lorsqu'une séquence homologue a été identifiée, la sous unité p41 de la Primase pourrait être recrutée et effectuer la polymérisation de la matrice discontinue. Ce modèle n'inclut pas l'intégralité des protéines de la RH Archeae qui sont présentées dans l'introduction

Dans un premier temps, il faudra valider ces interactions physiques par une autre approche. Si ces interactions sont bien valides, il faudra alors étudier l'effet de la primase sur l'activité de recombinaison de RadA et inversement, l'effet de RadA sur l'activité d'extension d'amorces de la sous unité catalytique de la primase.

La 3^{ème} relation entre recombinaison et réplication que nous avons identifiée est centralisée sur la protéine pab0961, désormais nommée RcrA. Cette protéine de fonction inconnue, codée dans le même locus que l'hélicase répliquative MCM et GINS23, est observée en complexe principalement avec GINS23, la nucléase associée GAN et la protéine radB. Au regard de l'interaction avec radB et deux des composants du complexe CMG (GINS23 et nucléase GAN), acteur clé de la réplication de l'ADN, il est raisonnable de penser que RcrA est certainement également impliquée dans le couplage des activités de réplication et de recombinaison. En accord avec cette hypothèse, le phénotype observé chez le mutant RcrA de *H. volcanii* suggère un rôle dans la régulation des phénomènes de recombinaison lors d'un arrêt de la fourche de réplication. L'identification de RcrA constitue la découverte d'une nouvelle famille de protéines de la maintenance génomique Archaea. La structure de la RcrA de *M. jannaschii* a révélé des homologies structurales avec l'AMPK – γ Eucaryote, kinase majeure de la régulation du cycle cellulaire. Cette connexion est intrigante, puisque l'AMPK eucaryote est un puissant régulateur des fonctions cellulaires, en fonction de la charge énergétique. On peut imaginer que RcrA pourrait jouer un rôle similaire mais cette fois-ci uniquement ciblé au niveau de la réplication de l'ADN. D'un point de vue évolutif, RcrA constituerait le module de base, capable de détecter le niveau de charge énergétique dans la cellule, auquel se serait ajouté les sous-unités catalytiques présentes dans le complexe AMPK eucaryote.

Au regard de sa co-occurrence avec RcrA et de sa position génétique, il est tentant de proposer que pab1421 interagisse avec RcrA et qu'elle puisse ainsi constituer la deuxième protéine Rcr, RcrB.

Ces résultats semblent indiquer qu'il existe une coordination étroite entre les activités de réplication et de recombinaison de l'ADN et qu'un mécanisme de contrôle assurerait le lien entre les machineries de réplication et de recombinaison. Soit pour répondre à l'apparition de dommages détectés par la fourche de réplication, soit pour réguler l'activité de réplication de l'ADN, en fonction de l'énergie disponible dans la cellule.

Quoiqu'il en soit, la compréhension des relations entre réplication et recombinaison chez les Archaea pourrait nous permettre de mieux comprendre comment ces organismes assurent la stabilité de leurs génomes au sein d'un environnement propice à l'apparition de lésions dans l'ADN.

Nouveaux complexes associés à la machinerie de transcription.

Ce travail, en collaboration avec le laboratoire de Michael Thomm, a permis de découvrir l'interaction physique et fonctionnelle entre le RPA et l'ARN polymérase. L'étude de cette interaction pourra avoir des conséquences pour la compréhension de ce même complexe eucaryote, puisque l'unique ARN polymérase Archaea est l'homologue de l'ARN polymérase II et que le RPA de *P. abyssi* est très proche de la version eucaryote.

Nos travaux ont permis de montrer l'interaction physique entre ces deux composants et la caractérisation fonctionnelle suggère que le RPA améliore le taux de transcrit lors de l'étape d'élongation. La caractérisation de cette interaction est toujours en cours. Nos collègues Allemands sont en train de produire le RPA de *P. furiosus* pour tester, dans un contexte homologue, l'hypothèse d'une implication du RPA au niveau de la bulle de transcription du complexe d'élongation. Pour notre part, nous allons caractériser par SPR l'interaction directe entre le RPA et l'ARN polymérase en présence ou absence d'ADN, ainsi que le comportement de fixation du RPA sur des substrats de type bulle. L'ensemble de ces données devrait pouvoir éclaircir le rôle du RPA dans l'activation de la transcription.

Mes travaux ont également permis d'identifier des relations entre les protéines de la réparation et de la transcription. En effet, plusieurs sous unités de la RNAP interagissent avec les protéines de la réparation de l'ADN ; l'ADN glycosylase OGG1 (Pab1695), la 3-méthyl-adenine DNA glycosylase AlkA (Pab1530), l'ADN ligase (Pab2002), ou encore l'hélicase de fonction inconnue Rad25 (pab0128). L'interaction de ces protéines avec l'ARN polymérase pourrait entrer en jeu dans la réparation des dommages oxydatifs couplée à la transcription. L'ADN glycosylase OGG1 pourrait même interagir directement avec l'ARN polymérase au même titre que ce qui est observé pour l'ADN glycosylase Eucaryote NEIL2. L'ARN polymérase est présenté dans son contexte élargi en figure 61. Il est très probable que l'on puisse identifier de nouveaux partenaires de l'ARN polymérase dans des régions du réseau encore peu explorées en détails. Certains de ces facteurs pourraient être impliqués dans le couplage entre la transcription et la réparation.

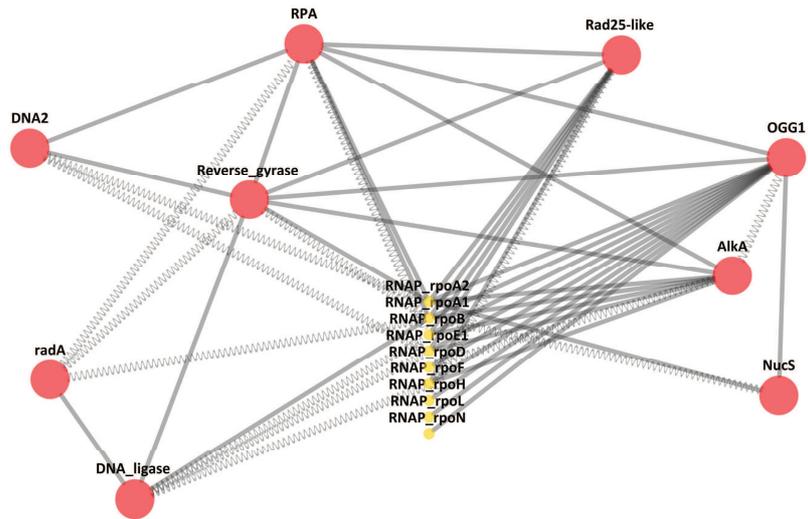


Figure 62: L'ARN polymérase dans le réseau étendu.

Les partenaires directs des sous unités de l'ARN polymérase identifiés dans le réseau étendu. Les traits sinueux sont des interactions du réseau élargi. L'ARN polymérase ne possède que deux protéines partenaire supplémentaires dans le réseau élargi: DNA2 qui interagit avec RPA et RadA dont l'interaction est abolie dans l'essai nucléase.

XPB partenaire inédit de l'hélicase répllicative

L'une des interactions les plus intrigantes est celle du MCM avec XPB. XPB est une hélicase 5'-3' de la famille SF2. XPB fait partie du complexe de transcription eucaryote TF2H, et participe activement à la réparation NER. La protéine XPB Archaea possède de fortes homologies de séquences avec la protéine des Eucaryotes. Cependant, sa délétion chez *T. kodakarensis* n'augmente pas la sensibilité aux dommages de l'ADN. Ainsi la fonction du XPB Archaea *in vivo*, reste mystérieuse. Dans le réseau d'interaction les composants du NER que nous avons identifiés ne sont pas connectés à XPB, qui reste isolée avec MCM. A l'heure actuelle, aucune interaction entre XPB et l'hélicase MCM n'a été identifiée. XPB et MCM diffèrent par leur polarité, 5'-3 et 3'-5', respectivement. Aussi XPB est capable d'ouvrir un brin d'ADN sur lequel est fixé une ARN polymérase, ce dont est incapable le MCM. Ainsi XPB pourrait venir en aide à MCM lorsqu'il s'agit de répliquer des régions encombrées de l'ADN, notamment par le système transcriptionnel. L'expérience que j'ai réalisée ne permet pas de déterminer si XPB est associée à l'hélicase pour la réplication en général ou seulement pour répliquer les matrices encombrées (Figure 63).

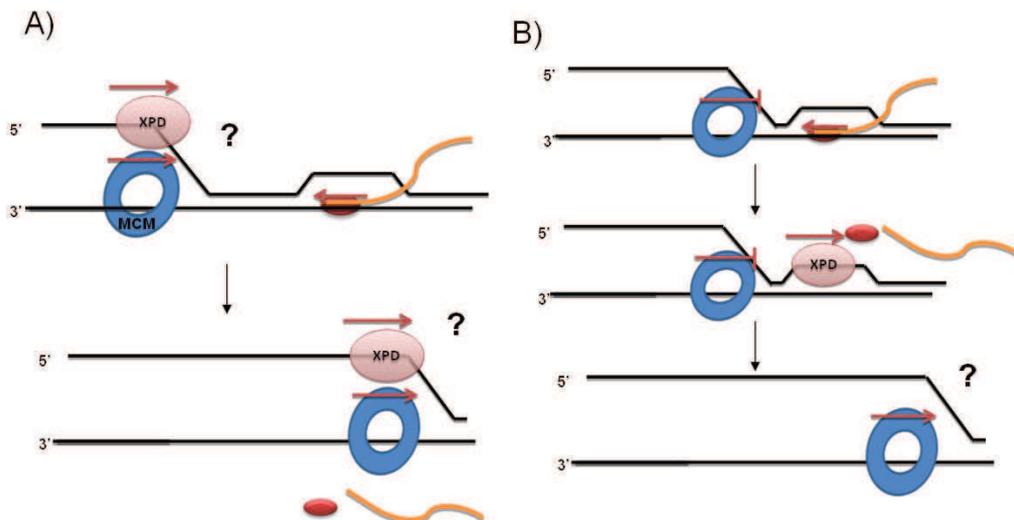


Figure 63: Schématisation de la collaboration entre MCM et XPB.

A) XPB interagît en permanence avec MCM lors de l'ouverture de l'ADN. B) XPB est recruté lors d'un arrêt de MCM en amont d'un fragment d'ADN transcrit par l'ARN polymérase.

Gaëlle Hogrel, une nouvelle thésarde dans le laboratoire, qui travaillera sur la caractérisation de plusieurs complexes du réseau, est d'ores et déjà en train de tester l'hypothèse selon laquelle XPB pourrait améliorer l'activité hélicase de MCM sur des substrats différents, contenant ou non des structures de type R-loop. Pour l'instant nous ne disposons que de la protéine XPB de *Sulfolobus solfataricus*, si les premiers tests fonctionnels s'avèrent probants, il faudrait produire et purifier l'hélicase XPB de *P. abyssi* pour étudier de manière homologue les interactions physiques et fonctionnelles entre ces deux protéines.

Implications biotechnologiques

Nous avons également identifié la protéine pab0431 comme un nouveau partenaire de PCNA. Cette protéine présente un pip-box motif et un domaine N-terminal homologue à l'extrémité N-terminal de l'ADN pol III de *Thermus aquaticus*. Pab0431 pourrait intervenir dans la réplication en constituant une sous unité d'un complexe PCNA : Polymérase : pab0431. Cette dernière pourrait alors conférer une activité de relecture supplémentaire lors de la réplication et pourrait ainsi améliorer la fidélité de la polymérase par un mécanisme qui reste à déterminer. Outre l'aspect fondamental, cette protéine est également intéressante, d'un point de vue biotechnologique, car elle pourrait permettre d'améliorer la fidélité des ADN polymérase dans une réaction PCR.

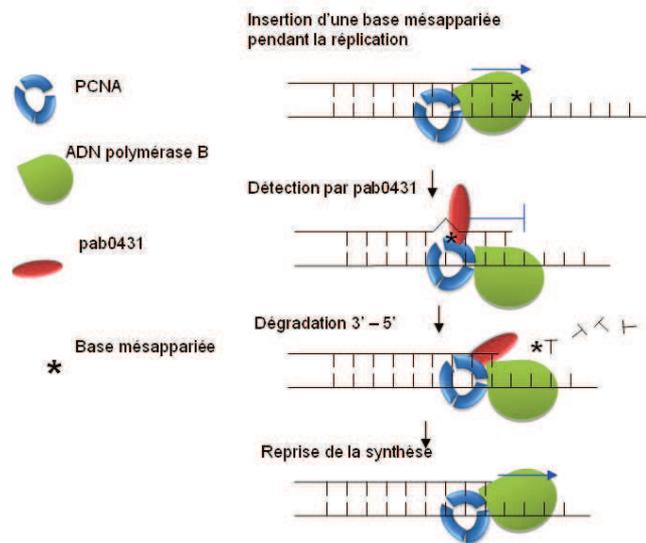


Figure 64: Modèle de l'action de pab0431

Pab0431 permettrait l'amélioration de la fidélité de l'ADN polymérase. Pab0431 voyage sur l'ADN lié au PCNA ou est recruté lors de la détection d'un mis-match. En tant que module isolé pab0431 serait capable d'initier une activité exo 3'-5' et de promouvoir ainsi l'élimination de la base mésappariée préalablement à une re-synthèse fidèle.

J'ai également mis en évidence la présence de plusieurs complexes spécifiques dans les données de confiance plus faible, ainsi il est probable que l'on puisse décrire d'autres interactions biologiquement significatives à partir des données de plus faible confiance. Les candidats les plus intéressants sont identifiables en orientant la recherche des critères ontologiques, comme je l'ai fait pour les protéines RcrA et B.

De nombreux autres aspects peuvent être intégrés au réseau pour mieux définir les interactions. Pour ce faire les bases de données biologiques mettent à disposition une quantité très importante d'informations qui peuvent être intégrées au réseau comme paramètre des nœuds ou des arêtes. Il est par exemple possible d'intégrer au réseau les domaines fonctionnels présents sur chaque protéine afin d'identifier l'interaction domaine à domaine entre deux protéines.

Afin de mieux comprendre la fonction des différentes interactions que nous avons décrites, il est bien évidemment prioritaire d'identifier les fonctions des complexes *in vitro*. En s'appuyant sur l'ontologie des protéines et leur environnement dans le réseau, il est possible de proposer une hypothèse sur un mécanisme pris en charge par un groupe de protéine et de tester ensuite *in vitro* son activité vis-à-vis d'un substrat nucléique déterminé. Ainsi, le réseau est un puissant générateur d'hypothèses lorsqu'il s'agit de déterminer les conditions expérimentales qui permettront d'identifier la fonction d'une protéine ou d'un complexe. Il est cependant difficile de proposer une fonction biologique aux associations détectées sans données fonctionnelles supplémentaires. Cependant nous avons maintenant des pistes plus précises pour essayer de répondre aux questions posées au début de l'étude.

En effet, l'absence de MMR chez les Archaea, incohérente avec le faible taux de mutations observées, pourrait s'expliquer par le fait que la recombinaison pourrait prendre en charge de manière très efficace les dommages rencontrés ou générés par l'activité de réplication, comme le suggère les interactions entre composants du réplisome et de la machinerie de recombinaison homologue. D'autre part il pourrait exister un mécanisme de régulation qui permettrait effectivement le couplage de ces activités au niveau de la fourche de réplication.

D'autre part, nos données semblent bien confirmer la présence d'une voie de réparation NER chez les Archaea, vu la clusterisation des acteurs potentiels de ce mécanisme. D'autre part, les interactions entre le RPA et les hélicases impliquées dans ce processus, suggèrent que cette protéine de fixation à l'ADN simple brin pourrait être impliquée dans la détection de la lésion sur l'ADN, elle constituerait alors le chaînon manquant vers un NER complet et fonctionnel. Il sera donc important de vérifier cette hypothèse en analysant le comportement de fixation du RPA sur des substrats ADN double brin contenant des lésions encombrantes, déstabilisant la double hélice, connues pour déclencher chez les Eucaryotes une réponse de type NER.

Mise à part la concentration des protéines hypothétiques du NER nous n'avons pas pu identifier des modules fonctionnels, bien définis dans le réseau, qui seraient en charge d'une voie métabolique précise. Cependant le réseau est très certainement une base solide pour les identifier en utilisant d'autres approches de biologie des systèmes

Approche systémique, vers l'intégration des données

Il apparaît ainsi vraiment indispensable de coupler ces résultats avec des travaux de génétiques pour à la fois confirmer *in vivo* les interactions détectées, mais aussi pour mieux cerner les voies métaboliques dans lesquels ces nouveaux complexes pourraient être impliqués. Le laboratoire est actuellement en train de développer des outils génétiques pour les thermococcales, proche de *P. abyssi*, ainsi que de mettre au point les méthodes de transformation et de sélection. Une fois au point, la génération de mutants pour les gènes de ce réseau couplée à l'observation phénotypique à la suite de différents traitements mutagéniques, devrait permettre de comprendre la fonction biologique de ces interactions. La résolution des structures tridimensionnelles des nouvelles protéines découvertes et de leur complexes serait également d'un intérêt certain.

Après avoir décrit la structure et l'organisation du réseau, si l'on veut décrire le comportement réel du système en charge de la maintenance génomique, il est primordial d'y intégrer son aspect dynamique. La dynamique d'un réseau peut être intégrée localement en déterminant les constantes d'affinité de chaque complexe où appréhendée au niveau du système en déterminant les régulations transcriptionnelles, traductionnelles ou post traductionnelles de chacune des protéines. On peut ainsi envisager de comparer les profils de transcription entre deux cultures de *P. abyssi* dont l'une aura été soumise à un agent physique ou chimique provoquant des dommages sur l'ADN. Dans la même expérience, on peut également envisager d'identifier les profils des polysomes qui nous permettrait de savoir si un groupe de protéines est traduit spécifiquement en réponse à un type de dommages donné. Grâce à cette approche, on pourra déterminer si les groupes de protéines identifiés dans le réseau constituent des modules fonctionnel co-régulés ce qui pourrait potentiellement conduire à l'identification de nouvelles voies de réparation ou à l'identification de nouveaux candidats par recherche de promoteurs.



8.1 CULTURE CONTINUE DE *P. ABYSSI*

La première étape des travaux réalisés a consisté à produire la biomasse suffisante pour réaliser l'ensemble des expériences subséquentes. Pour cela, j'ai réalisé une culture continue et anaérobie de l'Archaea *P. abyssi* en bioréacteur gaz lift 5 litres (Figures 65 et 66). Ce bioréacteur conçu spécialement pour la culture des microorganismes hyperthermophiles nous a permis de réaliser une culture à grande échelle de *P. abyssi*. Le bioréacteur est paramétré afin d'obtenir les conditions optimales de croissance de *P. abyssi*. *P. abyssi* est cultivée dans un milieu SME modifié (Godfroy, Postec et al. 2006). Le milieu est stérilisé par filtration à l'aide de filtres 0,22 μ m (Millipore). L'état d'oxydation du milieu est contrôlé grâce à la présence de résazurine dont la couleur rose indique un faible taux d'oxygène dans le milieu de culture. Le milieu est dépourvu de Pipes et de Na₂S. L'apport de soufre, utilisé comme accepteur final de la chaîne d'électron, est assuré par la présence de L-cystéine. La température, le pH et l'agitation au sein du bioréacteur sont maintenus automatiquement. Un thermostat maintient la température à 96°C (Godfroy, Postec et al. 2006) tandis que 2 pompes contrôlent un influx de NaOH 1M ou de HCl 1M afin de maintenir la culture à un pH 6,4. La culture est maintenue en anaérobiose par la présence d'un flux d'azote. La surpression créée par l'ascension du gaz dans le « draft tube » permet également l'agitation du milieu de culture par convection. Un influx continu de milieu frais est appliqué et l'efflux est récolté dans un récipient fermé maintenu dans un bain de glace. Le flux continu de milieu est réglé à un taux de dilution de 0.1 L.h⁻¹. La concentration cellulaire au sein de la culture est stabilisée entre 1x10⁸ et 5x10⁸ cellules par millilitres pendant 6 semaines. La croissance a été suivie par des mesures de densité sur une chambre de comptage cellule de Thoma (0,02 mm de profondeur). La concentration de *P. abyssi* dans le bioréacteur est présentée en fonction du nombre de jours de culture en figure 66. La culture de *P. abyssi* est extraite par un capillaire du réacteur et récoltée dans un récipient fermé plongé dans un bain de glace. L'état des cellules récoltées est vérifié chaque jour sous microscope optique à contraste de phase. Après récolte, les cellules sont séparées du milieu de culture par deux étapes de centrifugation à 6000 g, pendant 45 min à 4 °C. Les culots, formés par les cellules Archaea, sont aliquotés et conservés à -80°C. En 6 semaines de culture continue, nous avons obtenu 350 g de culot cellulaire. Les cellules de *Pyrococcus abyssi* issues de la culture ont également été observées sous microscope électronique à balayage (Figure 7), le spécimen présenté est probablement en fin de division cellulaire.

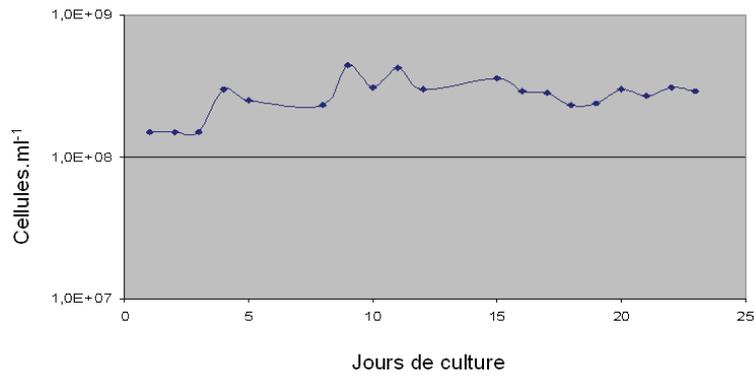


Figure 65 : Suivi de la concentration cellulaire lors de la culture continue

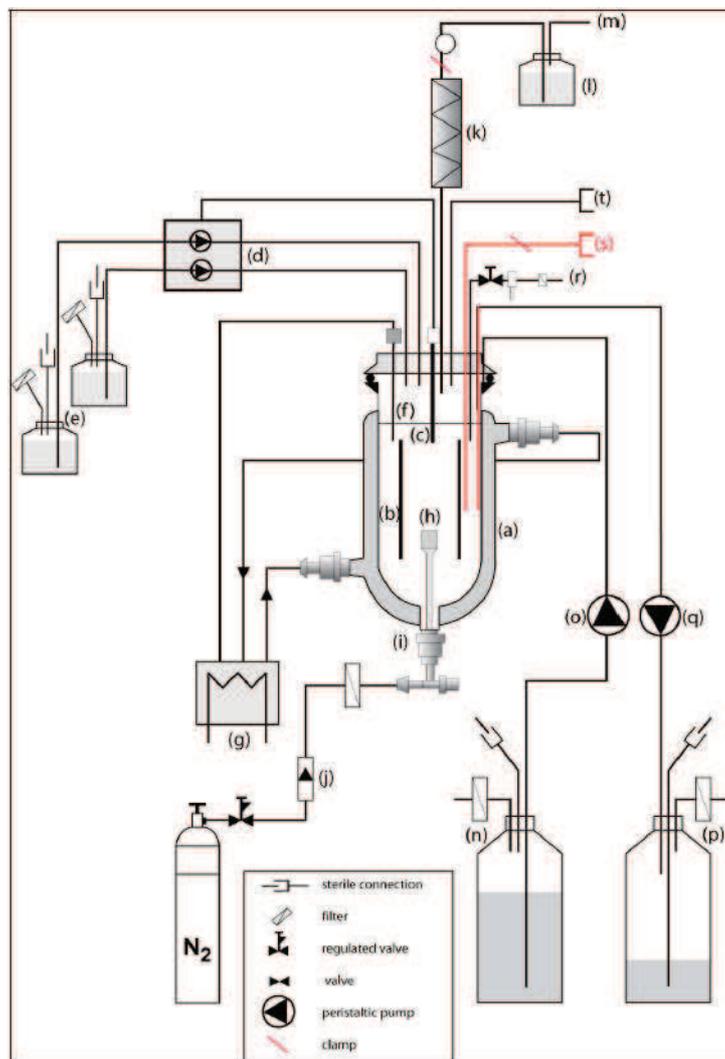


Figure 66 : Représentation schématique du système expérimental de culture en bioréacteur gazlift.

(a) cuve en verre à double enveloppe ; (b) « draft-tube » ; (c) sonde de pH ; (d) contrôleur de pH ; (e) flacons d'acide (HCl 1 N) et de base (NaOH 1 N) ; (f) sonde de température ; (g) bain thermostaté ; (h) bulleur ; (i) robinet du bulleur ; (j) débitmètre à bille pour gaz ; (k) condenseur ; (l) trappe de condensation ; (m) effluent gazeux ; (n) bidon d'entrée (milieu de culture) ; (o) pompe d'entrée ; (p) bidon de sortie (récolte) ; (q) pompe de sortie ; (r) rampe aérobie de prélèvement d'échantillons ; (t) système d'inoculation. (Godfroy et al., 2006).

8.2 PREPARATION DE L'EXTRAIT CELLULAIRE

Les culots de *P. abyssi* sont directement resuspendus dans le tampon utilisé en purification d'affinité. Le tampon utilisé est un tampon salin standard PBS (140 mM de NaCl, 2,7 mM de KCl, 10 mM de Na₂HPO₄, 1,76 mM de KH₂PO₄, pH 7,4) dans lequel on ajoute 10 mM d'imidazole et un mélange d'anti protéase sans EDTA (Roche). La lyse des cellules a été réalisée par sonication en utilisant un sonicateur Vibracell (Bioblock Scientific, 2 x 10 min de 0,5 seconde marche / arrêt des impulsions à 375 W, amplitude de 40%, sur glace). Les débris cellulaires ont été éliminés par centrifugation à 15 000 g, 45 min à 4 ° C. La concentration de la fraction soluble a été mesurée par Protein Assay Dc (Bio-Rad) avec l'albumine sérique bovine comme standard. La concentration est finalement ajustée à 8 mg.ml⁻¹ et le pH à 7,4.

8.3 CLONAGE DES GENES CIBLES

Les protéines recombinantes des gènes pab1530, pab0164, pab2270, pab0067, pab0560, pab1695, pab1466, pab0956, pab0034, pab0128, pab0801, pab7094, pab0418, pab2423, pab0063 proviennent du programme MARINE-EXPRESS (Groisillier, Herve et al. 2010). Les protéines, pab2266, pab2404, (Palud, Villani et al. 2008), pab0352 (Le Laz, Le Goaziou et al. 2010), pab1465, pab1877, pab2002, pab2373, (Meslet-Cladière, Norais et al. 2007), pab2263, (Ren, Kuhn et al. 2007), pab2236, pab2235 (Le Breton, Henneke et al. 2007), pab2163, pab2164, pab2165, pab0811, pab0812 et pab1128 ont été individuellement clonés au laboratoire. La liste de protéines appâts et leurs propriétés sont résumées dans le tableau 1. La procédure de clonage des protéines non référencées est détaillée dans le tableau S1 de l'article.

Selon l'origine du clone, le vecteur d'expression utilisé peut différer (données supplémentaires S1) mais le principe de purification est cependant le même. Brièvement, la séquence d'ADN ciblée a été amplifiée par PCR avec des amorces spécifiques visant à ajouter en 5' et 3' des sites de restriction. Les séquences, amplifiées et digérées, ont été clonées dans des vecteurs de multiplication puis d'expression en phase avec une étiquette de poly-histidine (HIS6) ou une glutathion S-transférase (GST). Le plasmide PGEX-4T-1 (GE Healthcare Life Sciences) a été utilisé pour exprimer la protéine en fusion avec une GST, alors que pFO4 (Groisillier, Hervé et al. 2010), pQE80, pE26b+, pET28a+, et pET30 (Novagen) ont été utilisés pour exprimer les protéines en fusion avec une étiquette histidine. La transcription du trans-gène est sous le contrôle du promoteur inductible LacZ. Les plasmides contenant les séquences codantes pour les protéines Archaea ont été utilisés pour transformer une souche modifiée d'*E. coli* (RosettaTM Novagen). Ces Bactéries modifiées ont la particularité de coder les ARN de transfert pour les codons rares dont l'absence peut entraîner des arrêts de la traduction et diminuer drastiquement la production d'une protéine hétérologue.

8.4 PRODUCTION ET PURIFICATION DES PROTEINES APPATS

Les clones bactériens sont cultivés dans du milieu LB (Beckton, Dickinson and Co). En Erlenmeyer 500 ml on incube une pré-culture de 30 mL (en phase stationnaire) dans 250 ml de LB. Le milieu est additionné d'antibiotiques (ampicilline 25mg. ML^{-1} et kanamycine ou chloramphénicol) permettant la conservation des plasmides dans la culture. La production de protéines a été induite en phase de croissance exponentielle à $\text{DO}_{600\text{nm}} = 0,6$ par 1 mM d'IPTG, 3h à 37 ° C. La transcription du gène codant pour la protéine de fusion est sous le contrôle du promoteur LacZ. Ainsi, la production de la protéine est inductible. En effet, la régulation négative de l'opéron Lac est levée par l'ajout de 100mM d'Isopropyl β -D-1-thiogalactopyranoside (IPTG), analogue non métabolisable du lactose. La levée d'inhibition de l'expression du gène cloné va *in fine* conduire à une surproduction de la protéine de fusion par l'ensemble des Bactéries de la culture. Après 3h d'induction à l'IPTG les cellules bactériennes sont récoltées par centrifugation. Dans le cas où la protéine produite est HIS-tag la protéine est purifiée directement de la fraction soluble et thermostable du lysat bactérien (figure 39). Un protocole de purification plus doux, avec une lyse cellulaire par sonication à froid est appliqué aux protéines GST-tag (Figure 39).

Les cellules ayant exprimées la protéine recombinantes sont centrifugées à 10 000 g, 20 minutes à 4°C et le surnageant éliminé. Les culots sont suspendus dans du PBS à un dixième du volume de culture et on réalise une lyse cellulaire, soit par 3 chocs thermiques de 10 min à 80 °C pour les protéinées tagguées HIS soit par sonication (10 min de 0,5 seconde marche / arrêt des impulsions à 375 W, 40% d'amplitude) pour les protéines munies d'un tag GST, non thermostable. La clarification des lysats cellulaires a été réalisée par centrifugation à 10000 g, 10 min à 4°C. Les protéines recombinantes mono et multimeriques ont été purifiés directement à partir de la fraction soluble sur des billes magnétiques (Dynabeads® Isolation His-Tag & Pull-Down ou Promega MagneGST™ Pull-Down System). Deux milligrammes de billes, équilibrée dans le tampon PBS ont été incubées, sur roue, 15 min, à température ambiante, avec 1.5 ml de lysat bactérien. Les protéines fixées sur les billes sont purifiées à l'aide d'un aimant et lavées 4 fois avec 1 ml du tampon de purification d'affinité. La qualité de la purification et la taille apparente de la protéine ont été vérifiées sur SDS-PAGE.

8.5 PURIFICATION D'AFFINITE DES COMPLEXES PROTEIQUES DE P. ABYSSI

Les protéines purifiées ont été utilisées comme appât contre l'extrait cellulaire de *P. abyssii*. En fonction de la taille protéine et de ses propriétés de liaison, entre 10 et 30 μg de protéines appâts ont été immobilisées sur 2 mg de billes magnétiques. Les protéines appâts, fixées sur les billes magnétiques ont été incubées avec 4 ml d'extrait cellulaire dans des tubes à hémolyse d'une capacité de 5 ml ceci sous rotation, pendant 3 heures, à 4 ° C. Les complexes protéiques formés *in vitro* ont été séparés sur un aimant et lavés abondamment par 4 fois 4 ml et 2 fois 1 ml de tampon PBS. Pour chaque protéine appât, on réalise en parallèle un essai supplémentaire auquel on ajoute une étape d'incubation avec un mélange de nucléases dégradant les ADN et ARN. Les complexes purifiés et lavés ont été incubés 45 minutes à 25°C avec un mélange de nucléase (DNAse et RNAse) (GE Healthcare). Pour chaque protéine appât nous procédons donc à deux expériences de purification par affinité, l'une sans et l'une avec un traitement nucléase, ces essais seront dénommés nuc- et nuc+,

respectivement. Les complexes protéiques purifiés ont été élués directement dans 100 µl de tampon de Laemmli à 95 °C pendant 15 min. Les protéines dénaturées sont ensuite séparées par SDS-PAGE pendant 20 min à 120V~ dans un gel Criterion XT 12% (Bio-Rad). Les protéines ont été extraites à partir des gels en découpant l'intégralité de chaque piste en 3 tranches de la même dimension. Les tranches de gel ont ensuite été analysées séparément en spectrométrie de masse. La séparation des protéines sur gel dénaturant électrophorèse s'est avérée améliorer considérablement l'efficacité de l'étape trypsinisation. Comme il a déjà été démontré, certaines protéines peuvent résister à la digestion hyperthermostable protéolytique à des températures modérées (Lee, Sevinsky et al. 2009). Un témoin négatif, où les billes sont incubées nues dans l'extrait, est également réalisé pour chaque essai. En outre, cinq témoins négatifs indépendants ont été réalisés afin d'identifier le pool de protéines se liant non spécifiquement aux billes.

8.6 SPECTROMETRIE DE MASSE

Les paramètres de l'analyse en spectrométrie de masse (MS) sont détaillés, en anglais, dans l'article.

8.7 IDENTIFICATION DES INTERACTIONS SPECIFIQUES

Les données MS ont été traitées manuellement afin d'identifier le signal spécifique du bruit de fond. Les résultats des purifications d'affinité sont comparés à une base de données témoin. L'intensité de la détection de la protéine est exprimée par un ratio signal sur bruit. Le signal correspond au score MASCOT de l'essai, le bruit de fond au score MASCOT des témoins. On considère de prime abord l'identification d'une protéine significative lorsqu'au moins 4 peptides indépendants couvrant au moins 10% de la séquence protéique ont été identifiés et que le score MASCOT résultant est supérieur à 100. Cette rationalisation nous permet de travailler uniquement avec des données MS de haute qualité. Afin d'identifier la population de protéines de *P. abyssi* qui se fixent non spécifiquement aux billes, nous avons réalisé cinq témoins indépendants et un témoin relatif à chaque expérience où les billes ont été incubées en l'absence de protéines appâts. Ces essais sont compilés et constituent une base de données de témoins négatifs donnant une valeur référence M_{neg} pour chaque protéine dans chaque expérience. Cette valeur M_{neg} est attribuée à l'ensemble des protéines codées dans le génome de *P. abyssi* mais peut parfois varier d'une expérience à une autre car elle prendra la valeur la plus élevée des témoins indépendants M_{ind} ou du témoin relatif M_{rel} . Ce dernier témoin relatif à l'expérience nous permet non seulement un degré de confiance encore plus élevé mais nous assure surtout qu'aucune contamination croisée en MS ne pollue nos résultats. Dans le cas où une protéine n'est détectée dans aucun des témoins, on lui attribuera arbitrairement une valeur M_{neg} de 20. L'intensité du signal d'une protéine est exprimée par le ratio signal / bruit (SR) calculé pour chaque protéine partenaire potentiellement identifiée. Les scores $SR+$ et $SR-$ ont été calculés pour chaque interaction putative en divisant le M_{neg} par les scores $nuc+$ et $nuc-$, respectivement. Nous avons fixé le seuil de signification à un score de SR de 0,2 correspondant à un signal cinq fois plus important que dans les témoins ($M_{nuc+} / - \geq 5 \times M_{neg}$). Pour une protéine qui n'a jamais été

détectée dans les témoins négatifs ($M_{neg}=20$) elle obtiendra un SR de 0,2 ou moins dès lors qu'elle franchira la valeur 100 en score MASCOT. Ce seuil SR à 0,2 correspond également à une détection cinq fois plus forte dans l'essai que dans l'ensemble des témoins.

La confiance dans la spécificité de la détection d'une interaction est considérée comme élevée lorsque les deux SR_{nuc} et SR_{nuc+} étaient en dessous de 0,2.

8.8 CONSTRUCTION DU RESEAU ET ANALYSE TOPOLOGIQUE.

Les interactions détectées en MS sont compilées en réseau. Ce réseau constitue une représentation graphique de l'ensemble de nos données. Les protéines sont représentées par les sommets (ou nœuds) connectés par des arêtes représentant les interactions en suivant le modèle de Spoke (ou « en étoile »). Ce modèle considère chaque interaction comme étant directe ; chaque proie est reliée à ses appât par une arête. Les complexes formés par les 3 sous unités du RPA, les deux sous unités de la Polymérase 2 utilisées en complexes comme appât sont réunis sous un seul nœud, RPA et Polymérase D respectivement. De la même manière le complexe Mre11/Rad50 est représenté par un seul nœud. Dans une première approche nous utilisons uniquement les données de haute confiance. Un algorithme « pseudoclique » nous permet d'identifier les zones de recoupement des données en identifiant les clusters. Un algorithme « mutcoeff » nous permet d'identifier les protéines qui possèdent le plus grand nombre de partenaires communs en identifiant les groupes minimum de 2 ou 3 protéines du réseau. Un algorithme « backbone » nous permet d'identifier le squelette du réseau en calculant l'arbre minimal de recouvrement (minimum spanning tree) du jeu dominant (dominant set). L'algorithme opéron permet d'identifier les protéines codées à proximité dans le génome. Les réseaux sont représentés en utilisant le logiciel libre Cytoscape© (Shannon, Markiel et al. 2003). Les représentations utilisées sont adaptées d'une représentation « Edge weighted Force directed » qui utilise le nombre d'interactions communes pour pondérer la taille de l'interaction et la proximité des nœuds dans la représentation bidimensionnelle du réseau. L'indice « Edge betweenness » est un bon indicateur de l'interaction entre deux protéines dans un réseau d'interaction (Dunn, Dudbridge et al. 2005). Cette représentation automatisée est ensuite adaptée manuellement pour une lecture optimale.

BIBLIOGRAPHIE

- Acharya, S., P. L. Foster, et al. (2003). "The Coordinated Functions of the E. coli MutS and MutL Proteins in Mismatch Repair." *Molecular Cell* **12**(1): 233-246.
- Albers, S.-V. and B. H. Meyer (2011). "The archaeal cell envelope." *Nat Rev Micro* **9**(6): 414-426.
- Albers, S.-V. and M. Pohlschröder (2009). "Diversity of archaeal type IV pilin-like structures." *Extremophiles* **13**(3): 403-410.
- Albert, R., H. Jeong, et al. (2000). "Error and attack tolerance of complex networks." *Nature* **406**(6794): 378-382.
- Alberts, B. (1998). "The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists." *Cell* **92**(3): 291-294.
- Alivisatos, A. Â P., M. Chun, et al. (2012). "The Brain Activity Map Project and the Challenge of Functional Connectomics." *Neuron* **74**(6): 970-974.
- Aravind, L. and E. V. Koonin (1998). "Phosphoesterase domains associated with DNA polymerases of diverse origins." *Nucleic Acids Res* **26**(16): 3746-52.
- Aravind, L. and E. V. Koonin (1998). "Phosphoesterase domains associated with DNA polymerases of diverse origins." *Nucleic Acids Research* **26**(16): 3746-3752.
- Arbing, M. A., S. Chan, et al. (2012). "Structure of the surface layer of the methanogenic archaean *Methanosarcina acetivorans*." *Proceedings of the National Academy of Sciences* **109**(29): 11812-11817.
- Arezi, B. and R. D. Kuchta (2000). "Eukaryotic DNA primase." *Trends in Biochemical Sciences* **25**(11): 572-576.
- Atomi, H., T. Imanaka, et al. (2012). "Overview of the genetic tools in the Archaea." *Front Microbiol* **3**: 337.
- Austin, R. J., H. M. Smidansky, et al. (2012). "Proteomic analysis of the androgen receptor via MS-compatible purification of biotinylated protein on streptavidin resin." *PROTEOMICS* **12**(1): 43-53.
- Ayyagari, R., X. V. Gomes, et al. (2003). "Okazaki fragment maturation in yeast. I. Distribution of functions between FEN1 AND DNA2." *J Biol Chem* **278**(3): 1618-25.
- Bader, G. D. and C. W. Hogue (2002). "Analyzing yeast protein-protein interaction data obtained from different sources." *Nat Biotechnol* **20**(10): 991-7.
- Banerjee, D., S. M. Mandal, et al. (2011). "Preferential repair of oxidized base damage in the transcribed genes of mammalian cells." *J Biol Chem* **286**(8): 6006-16.
- Barabasi, A.-L. and Z. N. Oltvai (2004). "Network biology: understanding the cell's functional organization." *Nat Rev Genet* **5**(2): 101-113.
- Barabasi, A. L. and R. Albert (1999). "Emergence of Scaling in Random Networks." *Science* **286**(5439): 509-512.
- Barry, E. R. and S. D. Bell (2006). "DNA Replication in the Archaea." *Microbiol. Mol. Biol. Rev.* **70**(4): 876-887.
- Barry, E. R., A. T. McGeoch, et al. (2007). "Archaeal MCM has separable processivity, substrate choice and helicase domains." *Nucleic Acids Research* **35**(3): 988-998.
- Bassingthwaighte, J. (2000). "Strategies for the Physiome Project." *Annals of Biomedical Engineering* **28**(8): 1043-1058.
- Batada, N. N., L. D. Hurst, et al. (2006). "Evolutionary and Physiological Importance of Hub Proteins." *PLoS Comput Biol* **2**(7): e88.

- Batada, N. N., T. Reguly, et al. (2006). "Stratus Not Altocumulus: A New View of the Yeast Protein Interaction Network." *PLoS Biol* **4**(10): e317.
- Batada, N. N., T. Reguly, et al. (2007). "Still Stratus Not Altocumulus: Further Evidence against the Date/Party Hub Distinction." *PLoS Biol* **5**(6): e154.
- Beattie, T. R. and S. D. Bell (2011). "Molecular machines in archaeal DNA replication." *Curr Opin Chem Biol* **15**(5): 614-9.
- Beattie, T. R. and S. D. Bell (2012). "Coordination of multiple enzyme activities by a single PCNA in archaeal Okazaki fragment maturation." *Embo J* **31**(6): 1556-1567.
- Becker, E., B. t. Robisson, et al. (2012). "Multifunctional proteins revealed by overlapping clustering in protein interaction network." *Bioinformatics* **28**(1): 84-90.
- Bell, S. D., C. H. Botting, et al. (2002). "The Interaction of Alba, a Conserved Archaeal Chromatin Protein, with Sir2 and Its Regulation by Acetylation." *Science* **296**(5565): 148-151.
- Ben Hania, W., R. Ghodbane, et al. (2011). "Cultivation of the first mesophilic representative (â€œmesotogaâ€) within the order Thermotogales." *Systematic and Applied Microbiology* **34**(8): 581-585.
- Berg, J., M. Lassig, et al. (2004). "Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications." *BMC Evolutionary Biology* **4**(1): 51.
- Bergerat, A., B. de Massy, et al. (1997). "An atypical topoisomerase II from archaea with implications for meiotic recombination." *Nature* **386**(6623): 414-417.
- Bern, M., B. S. Phinney, et al. (2009). "Reanalysis of Tyrannosaurus rex Mass Spectra." *Journal of Proteome Research* **8**(9): 4328-4332.
- Berquist, B., P. DasSarma, et al. (2007). "Essential and non-essential DNA replication genes in the model halophilic Archaeon, Halobacterium sp. NRC-1." *BMC Genetics* **8**(1): 31.
- Berquist, B. R., P. DasSarma, et al. (2007). "Essential and non-essential DNA replication genes in the model halophilic Archaeon, Halobacterium sp. NRC-1." *BMC Genet* **8**: 31.
- Bersini, H., T. Lenaerts, et al. (2006). "Growing biological networks: Beyond the gene-duplication model." *Journal of Theoretical Biology* **241**(3): 488-505.
- Beukers, R. and W. Berends (1960). "The effect of paramagnetic substances on the conversion of some pyrimidines by ultraviolet radiation." *Biochim Biophys Acta* **38**: 573-5.
- Birrien, J.-L., X. Zeng, et al. (2011). "Pyrococcus yayanosii sp. nov., an obligate piezophilic hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent." *International Journal of Systematic and Evolutionary Microbiology* **61**(12): 2827-2881.
- Blonder, B. and A. Dornhaus (2011). "Time-Ordered Networks Reveal Limitations to Information Flow in Ant Colonies." *PLoS ONE* **6**(5): e20298.
- Bochman, M. L. and A. Schwacha (2009). "The Mcm Complex: Unwinding the Mechanism of a Replicative Helicase." *Microbiology and Molecular Biology Reviews* **73**(4): 652-683.
- Bocquier, A. A., L. Liu, et al. (2001). "Archaeal primase: bridging the gap between RNA and DNA polymerases." *Current Biology* **11**(6): 452-456.
- Böhlke, K., F. Pisani, et al. (2002). "Archaeal DNA replication: spotlight on a rapidly moving field." *Extremophiles* **6**(1): 1-14.

- Bolt, E. L., R. G. Lloyd, et al. (2001). "Genetic analysis of an archaeal Holliday junction resolvase in *Escherichia coli*." Journal of Molecular Biology **310**(3): 577-589.
- Bond, C. S., M. Kvaratskhelia, et al. (2001). "Structure of Hjc, a Holliday junction resolvase, from *Sulfolobus solfataricus*." Proceedings of the National Academy of Sciences **98**(10): 5509-5514.
- Boubakri, H., A. L. de Septenville, et al. (2010). "The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo." EMBO J **29**(1): 145-57.
- Boudsocq, F., S. Iwai, et al. (2001). "Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta." Nucleic Acids Res **29**(22): 4607-16.
- Boudsocq, F. o., S. Iwai, et al. (2001). "Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic polĪ." Nucleic Acids Research **29**(22): 4607-4616.
- Boulton, S. J., A. Gartner, et al. (2002). "Combined Functional Genomic Maps of the *C. elegans* DNA Damage Response." Science **295**(5552): 127-131.
- Bravo, R. and J. E. Celis (1980). "A search for differential polypeptide synthesis throughout the cell cycle of HeLa cells." The Journal of Cell Biology **84**(3): 795-802.
- Brochier-Armanet, C., P. Forterre, et al. (2011). "Phylogeny and evolution of the Archaea: one hundred genomes later." Current Opinion in Microbiology **14**(3): 274-281.
- Brock, T., K. Brock, et al. (1972). "Sulfolobus: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature." Archiv fÄ¼r Mikrobiologie **84**(1): 54-68.
- Brymora, A., V. A. Valova, et al. (2001). Protein-Protein Interactions Identified by Pull-Down Experiments and Mass Spectrometry. Current Protocols in Cell Biology, John Wiley & Sons, Inc.
- Bubeck, D., M. A. M. Reijns, et al. (2011). "PCNA directs type 2 RNase H activity on DNA replication and repair substrates." Nucleic Acids Research **39**(9): 3652-3666.
- Budd, M. E., A. H. Y. Tong, et al. (2005). "A Network of Multi-Tasking Proteins at the DNA Replication Fork Preserves Genome Stability." PLoS Genet **1**(6): e61.
- Buescher, J. M., W. Liebermeister, et al. (2012). "Global Network Reorganization During Dynamic Adaptations of *Bacillus subtilis* Metabolism." Science **335**(6072): 1099-1103.
- Burgers, P. M. J., E. V. Koonin, et al. (2001). "Eukaryotic DNA Polymerases: Proposal for a Revised Nomenclature." Journal of Biological Chemistry **276**(47): 43487-43490.
- Busch, C. R. and J. DiRuggiero (2010). "MutS and MutL Are Dispensable for Maintenance of the Genomic Mutation Rate in the Halophilic Archaeon *Halobacterium salinarum* NRC-1." PLoS ONE **5**(2): e9045.
- Busch, C. R. and J. DiRuggiero (2010). "MutS and MutL are dispensable for maintenance of the genomic mutation rate in the halophilic archaeon *Halobacterium salinarum* NRC-1." PLoS One **5**(2): e9045.
- Butland, G., J. M. Peregrin-Alvarez, et al. (2005). "Interaction network containing conserved and essential protein complexes in *Escherichia coli*." Nature **433**(7025): 531-537.
- Callebaut, I., D. Moshous, et al. (2002). "Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family." Nucleic Acids Res **30**(16): 3592-601.
- Cann, I., S. Ishino, et al. (1999). "Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea." J Bacteriol **181**: 6591 - 6599.

- Cann, I. K. O., S. Ishino, et al. (2001). "Biochemical Analysis of Replication Factor C from the Hyperthermophilic Archaeon *Pyrococcus furiosus*." Journal of Bacteriology **183**(8): 2614-2623.
- Castrec, B., C. Rouillon, et al. (2009). "Binding to PCNA in Euryarchaeal DNA Replication requires two PIP motifs for DNA polymerase D and one PIP motif for DNA polymerase B." J Mol Biol **394**(2): 209-18.
- Castrec, B. t., S. b. Laurent, et al. (2010). "The Glycine-Rich Motif of *Pyrococcus abyssi* DNA Polymerase D Is Critical for Protein Stability." Journal of Molecular Biology **396**(4): 840-848.
- Castrec, B. t., C. Rouillon, et al. (2009). "Binding to PCNA in Euryarchaeal DNA Replication Requires Two PIP Motifs for DNA Polymerase D and One PIP Motif for DNA Polymerase B." Journal of Molecular Biology **394**(2): 209-218.
- Cavicchioli, R. (2006). "Cold-adapted archaea." Nat Rev Micro **4**(5): 331-343.
- Cerritelli, S. M. and R. J. Crouch (2009). "Ribonuclease H: the enzymes in eukaryotes." FEBS Journal **276**(6): 1494-1505.
- ChÃ©din, F. d. r., E. M. Seitz, et al. (1998). "Novel homologs of replication protein A in Archaea: implications for the evolution of ssDNA-binding proteins." Trends in Biochemical Sciences **23**(8): 273-277.
- Chapman, J. Å R., MartinÅ R. G. Taylor, et al. (2012). "Playing the End Game: DNA Double-Strand Break Repair Pathway Choice." Molecular Cell **47**(4): 497-510.
- Charmasson, S., P.-M. Sarradin, et al. (2009). "High levels of natural radioactivity in biota from deep-sea hydrothermal vents: a preliminary communication." Journal of Environmental Radioactivity **100**(6): 522-526.
- Chung, F. and L. Lu (2002). "The average distances in random graphs with given expected degrees." Proceedings of the National Academy of Sciences **99**(25): 15879-15882.
- Clouet-d'Orval, B., D. Rinaldi, et al. (2010). "Euryarchaeal beta-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity." J Biol Chem **285**(23): 17574-83.
- Cohen, G. N., V. Barbe, et al. (2003). "An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*." Molecular Microbiology **47**(6): 1495-1512.
- Cohen, G. N., V. Barbe, et al. (2003). "An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*." Mol Microbiol **47**(6): 1495-512.
- Cohen, R. and S. Havlin (2003). "Scale-Free Networks Are Ultrasmall." Physical Review Letters **90**(5): 058701.
- Coin, F., V. Oksenysh, et al. (2007). "Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair." Mol Cell **26**(2): 245-56.
- Constantinesco, F., P. Forterre, et al. (2002). "NurA, a novel 5'-3' nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea." EMBO Rep **3**(6): 537-42.
- Constantinesco, F., P. Forterre, et al. (2002). "NurA, a novel 5[prime]-3[prime] nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea." EMBO Rep **3**(6): 537-542.
- Constantinesco, F., P. Forterre, et al. (2004). "A bipolar DNA helicase gene, herA, clusters with rad50, mre11 and nurA genes in thermophilic archaea." Nucleic Acids Research **32**(4): 1439-1447.
- Constantinesco, F., P. Forterre, et al. (2004). "A bipolar DNA helicase gene, herA, clusters with rad50, mre11 and nurA genes in thermophilic archaea." Nucleic Acids Res **32**(4): 1439-47.
- Corliss, J. B., J. Dymond, et al. (1979). "Submarine Thermal Springs on the GalÃ¡pagos Rift." Science **203**(4385): 1073-1083.

- Creze, C., A. Ligabue, et al. (2012). "Modulation of the *Pyrococcus abyssi* NucS endonuclease activity by the replication clamp PCNA at functional and structural levels." J Biol Chem.
- Creze, C., A. Ligabue, et al. (2012). "Modulation of the *Pyrococcus abyssi* NucS Endonuclease Activity by Replication Clamp at Functional and Structural Levels." Journal of Biological Chemistry **287**(19): 15648-15660.
- Cubeddu, L. and M. F. White (2005). "DNA damage detection by an archaeal single-stranded DNA-binding protein." J Mol Biol **353**(3): 507-16.
- Cubeddu, L. and M. F. White (2005). "DNA Damage Detection by an Archaeal Single-stranded DNA-binding Protein." Journal of Molecular Biology **353**(3): 507-516.
- D'Amours, D. and S. P. Jackson (2002). "The MRE11 complex: at the crossroads of DNA repair and checkpoint signalling." Nat Rev Mol Cell Biol **3**(5): 317-327.
- De Falco, M., A. Fusco, et al. (2004). "The DNA primase of *Sulfolobus solfataricus* is activated by substrates containing a thymine-rich bubble and has a 3' terminal nucleotidyl-transferase activity." Nucleic Acids Research **32**(17): 5223-5230.
- Delmas, S., L. Shunburne, et al. (2009). "Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon *Haloferax volcanii* by restraining homologous recombination." PLoS Genet **5**(7): e1000552.
- Delmas, S. p., L. Shunburne, et al. (2009). "Mre11-Rad50 Promotes Rapid Repair of DNA Damage in the Polyploid Archaeon *Haloferax volcanii* by Restraining Homologous Recombination." PLoS Genet **5**(7): e1000552.
- DeLong, E. F. (1998). "Everything in moderation: Archaea as "non-extremophiles"." Current Opinion in Genetics & Development **8**(6): 649-654.
- Desmond, E., C. Brochier-Armanet, et al. (2011). "On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes." Research in Microbiology **162**(1): 53-70.
- DeVeaux, L. C., J. A. Muller, et al. (2007). "Extremely radiation-resistant mutants of a halophilic archaeon with increased single-stranded DNA-binding protein (RPA) gene expression." Radiat Res **168**(4): 507-14.
- Doherty, A. J. and S. W. Suh (2000). "Structural and mechanistic conservation in DNA ligases." Nucleic Acids Research **28**(21): 4051-4058.
- Dominski, Z. (2007). "Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism." Crit Rev Biochem Mol Biol **42**(2): 67-93.
- Dorazi, R., D. Gotz, et al. (2007). "Equal rates of repair of DNA photoproducts in transcribed and non-transcribed strands in *Sulfolobus solfataricus*." Mol Microbiol **63**(2): 521-9.
- Dorazi, R., D. Götz, et al. (2007). "Equal rates of repair of DNA photoproducts in transcribed and non-transcribed strands in *Sulfolobus solfataricus*." Molecular Microbiology **63**(2): 521-529.
- Dueber, E. L. C., J. E. Corn, et al. (2007). "Replication Origin Recognition and Deformation by a Heterodimeric Archaeal Orc1 Complex." Science **317**(5842): 1210-1213.
- Dunn, R., F. Dudbridge, et al. (2005). "The Use of Edge-Betweenness Clustering to Investigate Biological Function in Protein Interaction Networks." BMC Bioinformatics **6**(1): 39.
- Duxin, J. P., H. R. Moore, et al. (2012). "Okazaki fragment processing-independent role for human Dna2 enzyme during DNA replication." J Biol Chem **287**(26): 21980-91.

- Earley, M. C. and G. F. Crouse (1998). "The role of mismatch repair in the prevention of base pair mutations in *Saccharomyces cerevisiae*." Proceedings of the National Academy of Sciences of the United States of America **95**(26): 15487-15491.
- Erauso, G. I., A.-L. Reysenbach, et al. (1993). "Pyrococcus abyssi sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent." Archives of Microbiology **160**(5): 338-349.
- Erdos, P. (1960). "On the evolution of random graphs." Publications of the Mathematical Institute of the Hungarian Academy of Sciences **5**: 17-61.
- Evguenieva-Hackenberg, E., P. Walter, et al. (2003). "An exosome-like complex in *Sulfolobus solfataricus*." EMBO Rep **4**: 889 - 893.
- Fanning, E., V. Klimovich, et al. (2006). "A dynamic model for replication protein A (RPA) function in DNA processing pathways." Nucleic Acids Research **34**(15): 4126-4137.
- Fletcher, R. J., B. E. Bishop, et al. (2003). "The structure and function of MCM from archaeal *M. Thermoautotrophicum*." Nat Struct Mol Biol **10**(3): 160-167.
- Flores, G. E., I. Wagner, et al. (2012). "Distribution, abundance, and diversity patterns of the thermoacidophilic "Deep-sea Hydrothermal Vent Euryarchaeota 2" (DHVE2)." Frontiers in Microbiology **3**.
- Fogg, M. J., L. H. Pearl, et al. (2002). "Structural basis for uracil recognition by archaeal family B DNA polymerases." Nat Struct Mol Biol **9**(12): 922-927.
- Forsburg, S. L. (2004). "Eukaryotic MCM Proteins: Beyond Replication Initiation." Microbiology and Molecular Biology Reviews **68**(1): 109-131.
- Forterre, P., F. Confalonieri, et al. (1999). "Identification of the gene encoding archeal-specific DNA-binding proteins of the Sac10b family." Molecular Microbiology **32**(3): 669-670.
- Fousteri, M. and L. H. F. Mullenders (2008). "Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects." Cell Res **18**(1): 73-84.
- Friedberg, E. C., A. R. Lehmann, et al. (2005). "Trading Places: How Do DNA Polymerases Switch during Translesion DNA Synthesis?" Molecular Cell **18**(5): 499-505.
- Friedel, C. C. and R. Zimmer (2009). "Identifying the topology of protein complexes from affinity purification assays." Bioinformatics **25**(16): 2140-6.
- Fuciarelli, A. F., E. C. Sisk, et al. (1995). "Induction of base damage in dna solutions by ultrasonic cavitation." Free Radical Biology and Medicine **18**(2): 231-238.
- Fujikane, R., S. Ishino, et al. (2010). "Genetic analysis of DNA repair in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*." Genes & Genetic Systems **85**(4): 243-257.
- Fujikane, R., S. Ishino, et al. (2010). "Genetic analysis of DNA repair in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*." Genes Genet Syst **85**(4): 243-57.
- Fukui, T., K. Yamauchi, et al. (2004). "Distinct roles of DNA polymerases delta and epsilon at the replication fork in *Xenopus* egg extracts." Genes to Cells **9**(3): 179-191.
- G., E., CHARBONNIER, et al. (1992). Caractérisation préliminaire d'une archaebactérie hyperthermophile possédant un plasmide, isolée d'une source hydrothermale du bassin Nord Fidjien. Paris, FRANCE, Elsevier.
- Gao, J. and J. Wang (2012). "Re-Annotation of Two Hyperthermophilic Archaea &i>Pyrococcus abyssi&i>; GE5 and &i>Pyrococcus furiosus&i>; DSM 3638." Current Microbiology **64**(2): 118-129.

- Garg, P. and P. Burgers (2005). "DNA Polymerases that Propagate the Eukaryotic DNA Replication Fork." Critical Reviews in Biochemistry and Molecular Biology **40**(2): 115 - 128.
- Gaudier, M., B. S. Schuwirth, et al. (2007). "Structural Basis of DNA Replication Origin Recognition by an ORC Protein." Science **317**(5842): 1213-1216.
- Gavin, A.-C., M. Bosche, et al. (2002). "Functional organization of the yeast proteome by systematic analysis of protein complexes." Nature **415**(6868): 141-147.
- Gérard, E. G., E. J. Jolivet, et al. (2001). "DNA protection mechanisms are not involved in the radioresistance of the hyperthermophilic archaea *Pyrococcus abyssi* and *P. furiosus*." Molecular Genetics and Genomics **266**(1): 72-78.
- Gibson, D. G., J. I. Glass, et al. (2010). "Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome." Science **329**(5987): 52-56.
- Girvan, M. and M. E. J. Newman (2002). "Community structure in social and biological networks." Proceedings of the National Academy of Sciences **99**(12): 7821-7826.
- Godfroy, A., A. Postec, et al. (2006). 4 Growth of Hyperthermophilic Microorganisms for Physiological and Nutritional Studies. Methods in Microbiology, Academic Press. **Volume 35**: 93-108.
- Godfroy, A., N. D. Raven, et al. (2000). "Physiology and continuous culture of the hyperthermophilic deep-sea vent archaeon *Pyrococcus abyssi* ST549." FEMS Microbiol Lett **186**(1): 127-32.
- Gomez-Garcia, I., I. Oyenarte, et al. "The Crystal Structure of Protein MJ1225 from *Methanocaldococcus jannaschii* Shows Strong Conservation of Key Structural Features Seen in the Eukaryal γ -AMPK." Journal of Molecular Biology **399**(1): 53-70.
- Gouge, J. r. m., C. I. Ralec, et al. (2012). "Molecular Recognition of Canonical and Deaminated Bases by *P. abyssi* Family B DNA Polymerase." Journal of Molecular Biology **423**(3): 315-336.
- Graham, D. E., R. Overbeek, et al. (2000). "An archaeal genomic signature." Proceedings of the National Academy of Sciences **97**(7): 3304-3308.
- Greagg, M. A., M. J. Fogg, et al. (1999). "A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil." Proceedings of the National Academy of Sciences **96**(16): 9045-9050.
- Grogan, D. W. (2004). "Stability and Repair of DNA in Hyperthermophilic Archaea."
- Grogan, D. W. (2004). "Stability and repair of DNA in hyperthermophilic Archaea." Curr Issues Mol Biol **6**(2): 137-44.
- Grogan, D. W., G. T. Carver, et al. (2001). "Genetic fidelity under harsh conditions: Analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*." Proceedings of the National Academy of Sciences **98**(14): 7928-7933.
- Grogan, D. W., G. T. Carver, et al. (2001). "Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*." Proc Natl Acad Sci U S A **98**(14): 7928-33.
- Groisillier, A., C. Herve, et al. (2010). "MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms." Microbial Cell Factories **9**(1): 45.
- Groisillier, A., C. Herve, et al. (2010). "MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms." Microb Cell Fact **9**: 45.

- Grüz, P., M. Shimizu, et al. (2003). "Processing of DNA lesions by archaeal DNA polymerases from *Sulfolobus solfataricus*." Nucleic Acids Research **31**(14): 4024-4030.
- Gueguen, Y., J. Rolland, et al. (2001). "Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*." Eur J Biochem **268**: 5961 - 5969.
- Gunbin, K., D. Afonnikov, et al. (2009). "Molecular evolution of the hyperthermophilic archaea of the *Pyrococcus* genus: analysis of adaptation to different environmental conditions." BMC Genomics **10**(1): 639.
- Guruharsha, K. G., J.-F. o. Rual, et al. (2011). "A Protein Complex Network of *Drosophila melanogaster*." Cell **147**(3): 690-703.
- Guy, C. P., S. Haldenby, et al. (2006). "Interactions of RadB, a DNA Repair Protein in Archaea, with DNA and ATP." Journal of Molecular Biology **358**(1): 46-56.
- Hamdan, S. M. and A. M. van Oijen (2010). "Timing, Coordination, and Rhythm: Acrobatics at the DNA Replication Fork." Journal of Biological Chemistry **285**(25): 18979-18983.
- Han, J.-D. J., N. Bertin, et al. (2004). "Evidence for dynamically organized modularity in the yeast protein-protein interaction network." Nature **430**(6995): 88-93.
- Han, J.-D. J., D. Dupuy, et al. (2005). "Effect of sampling on topology predictions of protein-protein interaction networks." Nat Biotech **23**(7): 839-844.
- Hanawalt, P. C. and G. Spivak (2008). "Transcription-coupled DNA repair: two decades of progress and surprises." Nat Rev Mol Cell Biol **9**(12): 958-970.
- Hartman, A. L., C. d. Norais, et al. (2010). "The Complete Genome Sequence of *Haloferax volcanii* DS2, a Model Archaeon." PLoS ONE **5**(3): e9605.
- Hartwell, L. H., J. J. Hopfield, et al. (1999). "From molecular to modular cell biology." Nature.
- Hass, C. S., K. Lam, et al. (2012). "Repair-specific Functions of Replication Protein A." Journal of Biological Chemistry **287**(6): 3908-3918.
- He, X. and J. Zhang (2006). "Why Do Hubs Tend to Be Essential in Protein Networks?" PLoS Genet **2**(6): e88.
- Hekinian, R., M. Fevrier, et al. (1983). "East Pacific Rise Near 13{degrees}N: Geology of New Hydrothermal Fields." Science (New York, N.Y.) **219**(4590): 1321-1324.
- Henneke, G. (2012). "In vitro reconstitution of RNA primer removal in Archaea reveals the existence of two pathways." Biochem J **447**(2): 271-80.
- Henneke, G. (2012). "In vitro reconstitution of RNA primer removal in Archaea reveals the existence of two pathways." Biochemical Journal **447**(2): 271-280.
- Henneke, G., D. Flament, et al. (2005). "The Hyperthermophilic Euryarchaeota *Pyrococcus abyssi* Likely Requires the Two DNA Polymerases D and B for DNA Replication." Journal of Molecular Biology **350**(1): 53-64.
- Henneke, G., D. Flament, et al. (2005). "The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication." J Mol Biol **350**: 53 - 64.
- Henneke, G., D. Flament, et al. (2005). "The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication." J Mol Biol **350**(1): 53-64.
- Henneke, G., Y. Gueguen, et al. (2002). "Replication factor C from the hyperthermophilic archaeon *Pyrococcus abyssi* does not need ATP hydrolysis for clamp-loading and contains a functionally conserved RFC PCNA-binding domain." J Mol Biol **323**(5): 795-810.

- Henneke, G., J.-P. Raffin, et al. (2000). "The PCNA from *Thermococcus fumicolans* Functionally Interacts with DNA Polymerase β ." Biochemical and Biophysical Research Communications **276**(2): 600-606.
- Hiller, B., M. Achleitner, et al. (2012). "Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity." The Journal of Experimental Medicine **209**(8): 1419-1426.
- Ho, Y., A. Gruhler, et al. (2002). "Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry." Nature **415**(6868): 180-183.
- Holme, P., M. Huss, et al. (2003). "Subnetwork hierarchies of biochemical pathways." Bioinformatics **19**(4): 532-538.
- Hong, J., J.-M. Ahn, et al. (2009). "Construction of a functional network for common DNA damage responses in *Escherichia coli*." Genomics **93**(6): 514-524.
- Hopfner, K. P., L. Craig, et al. (2002). "The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair." Nature **418**(6897): 562-6.
- Hopfner, K. P., A. Karcher, et al. (2000). "Mre11 and Rad50 from *Pyrococcus furiosus*: cloning and biochemical characterization reveal an evolutionarily conserved multiprotein machine." J Bacteriol **182**(21): 6036-41.
- Hopkins, B. B. and T. T. Paull (2008). "The *P. furiosus* mre11/rad50 complex promotes 5' strand resection at a DNA double-strand break." Cell **135**(2): 250-60.
- Hu, J., L. Guo, et al. (2012). "Template-dependent polymerization across discontinuous templates by the heterodimeric primase from the hyperthermophilic archaeon *Sulfolobus solfataricus*." Nucleic Acids Res **40**(8): 3470-83.
- Hu, J., L. Guo, et al. (2012). "Template-dependent polymerization across discontinuous templates by the heterodimeric primase from the hyperthermophilic archaeon *Sulfolobus solfataricus*." Nucleic Acids Research **40**(8): 3470-3483.
- Huang, H. Z., A. Nichols, et al. (2009). "Direct Identification and Quantification of Aspartyl Succinimide in an IgG2 mAb by RapiGest Assisted Digestion." Analytical Chemistry **81**(4): 1686-1692.
- Hübscher, U., G. Maga, et al. (2002). "EUKARYOTIC DNA POLYMERASES." Annual Review of Biochemistry **71**(1): 133-163.
- Huertas, P., F. Cortes-Ledesma, et al. (2008). "CDK targets Sae2 to control DNA-end resection and homologous recombination." Nature **455**(7213): 689-92.
- Hutton, R. D., T. D. Craggs, et al. (2010). "PCNA and XPF cooperate to distort DNA substrates." Nucleic Acids Res **38**(5): 1664-75.
- Hutton, R. D., J. A. Roberts, et al. (2008). "PCNA stimulates catalysis by structure-specific nucleases using two distinct mechanisms: substrate targeting and catalytic step." Nucleic Acids Res **36**(21): 6720-7.
- Ishibaschi, #160, et al. (1994). Fluctuation of chemical compositions of the phase-separated hydrothermal fluid from the North Fiji Basin Ridge. Amsterdam, PAYS-BAS, Elsevier.
- Ishitani, R., Y. Sugita, et al. (2008). "Mg²⁺-sensing mechanism of Mg²⁺ transporter MgtE probed by molecular dynamics study." Proc Natl Acad Sci U S A **105**(40): 15393-8.
- Jenkinson, E. R. and J. P. J. Chong (2006). "Minichromosome maintenance helicase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert." Proceedings of the National Academy of Sciences **103**(20): 7613-7618.
- Jeong, H., S. P. Mason, et al. (2001). "Lethality and centrality in protein networks." Nature **411**(6833): 41-42.

- Johnson, R. E., G. K. Kovvali, et al. (1996). "Evidence for Involvement of Yeast Proliferating Cell Nuclear Antigen in DNA Mismatch Repair." Journal of Biological Chemistry **271**(45): 27987-27990.
- Joyce, G. F. (2012). "Bit by Bit: The Darwinian Basis of Life." PLoS Biol **10**(5): e1001323.
- Kandler, O. and H. König (1998). "Cell wall polymers in Archaea (Archaeobacteria)." Cellular and Molecular Life Sciences **54**(4): 305-308.
- Kashefi, K. and D. R. Lovley (2003). "Extending the Upper Temperature Limit for Life." Science **301**(5635): 934.
- Kasisviswanathan, R., J.-H. Shin, et al. (2005). "Interactions between the archaeal Cdc6 and MCM proteins modulate their biochemical properties." Nucleic Acids Research **33**(15): 4940-4950.
- Kedar, P. S., S.-J. Kim, et al. (2002). "Direct Interaction between Mammalian DNA Polymerase β and Proliferating Cell Nuclear Antigen." Journal of Biological Chemistry **277**(34): 31115-31123.
- Kelch, B. A., D. L. Makino, et al. (2011). "How a DNA Polymerase Clamp Loader Opens a Sliding Clamp." Science **334**(6063): 1675-1680.
- Kelly, T. J., P. Simanek, et al. (1998). "Identification and characterization of a single-stranded DNA-binding protein from the archaeon *Methanococcus jannaschii*." Proceedings of the National Academy of Sciences **95**(25): 14634-14639.
- Kelman, L. M. and Z. Kelman (2003). "Archaea: an archetype for replication initiation studies?" Molecular Microbiology **48**(3): 605-615.
- Kelman, Z. and J. Hurwitz (2000). "A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon *Methanobacterium thermoautotrophicum deltaH*." J Biol Chem **275**(10): 7327-36.
- Kelman, Z. and M. F. White (2005). "Archaeal DNA replication and repair." Current Opinion in Microbiology **8**(6): 669-676.
- Kelman, Z. and M. F. White (2005). "Archaeal DNA replication and repair." Curr Opin Microbiol **8**(6): 669-76.
- Kernchen, U. and G. Lipps (2006). "Thermodynamic analysis of the single-stranded DNA binding activity of the archaeal replication protein A (RPA) from *Sulfolobus solfataricus*." Biochemistry **45**(2): 594-603.
- Kerr, I. D., R. I. M. Wadsworth, et al. (2003). "Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein." Embo J **22**(11): 2561-2570.
- Khil, P. P. and R. D. Camerini-Otero (2002). "Over 1000 genes are involved in the DNA damage response of *Escherichia coli*." Mol Microbiol **44**(1): 89-105.
- King, N. P., T. M. Lee, et al. (2008). "Structures and functional implications of an AMP-binding cystathionine beta-synthase domain protein from a hyperthermophilic archaeon." J Mol Biol **380**(1): 181-92.
- Kiontke, S., Y. Geisselbrecht, et al. (2011). "Crystal structures of an archaeal class II DNA photolyase and its complex with UV-damaged duplex DNA." Embo J **30**(21): 4437-4449.
- Komori, K. and Y. Ishino (2001). "Replication Protein A in *Pyrococcus furiosus* Is Involved in Homologous DNA Recombination." Journal of Biological Chemistry **276**(28): 25654-25660.
- Komori, K. and Y. Ishino (2001). "Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination." J Biol Chem **276**(28): 25654-60.
- Komori, K., T. Miyata, et al. (2000). "Both RadA and RadB are involved in homologous recombination in *Pyrococcus furiosus*." J Biol Chem **275**(43): 33782-90.

- Komori, K., S. Sakae, et al. (1999). "A Holliday junction resolvase from *Pyrococcus furiosus*: Functional similarity to *Escherichia coli* RuvC provides evidence for conserved mechanism of homologous recombination in Bacteria, Eukarya, and Archaea." Proceedings of the National Academy of Sciences **96**(16): 8873-8878.
- Koonin, E. V., K. S. Makarova, et al. (2001). "HORIZONTAL GENE TRANSFER IN PROKARYOTES: Quantification and Classification1." Annual Review of Microbiology **55**(1): 709-742.
- Koster, D. A., A. Crut, et al. (2010). "Cellular Strategies for Regulating DNA Supercoiling: A Single-Molecule Perspective." Cell **142**(4): 519-530.
- Kuba, Y., S. Ishino, et al. (2012). "Comparative analyses of the two proliferating cell nuclear antigens from the hyperthermophilic archaeon, *Thermococcus kodakarensis*." Genes to Cells **17**(11): 923-937.
- Kulaeva, O. I., E. V. Koonin, et al. (1996). "Identification of a DinB/UmuC homolog in the archeon *Sulfolobus solfataricus*." Mutat Res **357**(1-2): 245-53.
- Kusser, A. G., M. G. Bertero, et al. (2008). "Structure of an archaeal RNA polymerase." J Mol Biol **376**(2): 303-7.
- Lao-Sirieix, S.-h. and S. D. Bell (2004). "The Heterodimeric Primase of the Hyperthermophilic Archaeon *Sulfolobus solfataricus* Possesses DNA and RNA Primase, Polymerase and 3'-terminal Nucleotidyl Transferase Activities." Journal of Molecular Biology **344**(5): 1251-1263.
- Lao-Sirieix, S.-h., L. Pellegrini, et al. (2005). "The promiscuous primase." Trends in Genetics **21**(10): 568-572.
- Lao-Sirieix, S. H., L. Pellegrini, et al. (2005). "The promiscuous primase." Trends Genet **21**(10): 568-72.
- Le Breton, M., G. Henneke, et al. (2007). "The Heterodimeric Primase from the Euryarchaeon *Pyrococcus abyssi*: A Multifunctional Enzyme for Initiation and Repair?" Journal of Molecular Biology **374**(5): 1172-1185.
- Le Laz, S. b., A. Le Goaziou, et al. (2010). "Structure-Specific Nuclease Activities of *Pyrococcus abyssi* RNase HIII." Journal of Bacteriology **192**(14): 3689-3698.
- Ledford, H. (2012). "Life-changing experiments: The biological Higgs." Nature **483**(7391): 528-530.
- Lee, A. M., J. R. Sevinsky, et al. (2009). "Proteomics of *Pyrococcus furiosus*, a Hyperthermophilic Archaeon Refractory to Traditional Methods." Journal of Proteome Research **8**(8): 3844-3851.
- Lee, A. M., J. R. Sevinsky, et al. (2009). "Proteomics of *Pyrococcus furiosus*, a hyperthermophilic archaeon refractory to traditional methods." J Proteome Res **8**(8): 3844-51.
- Lehman, I. R., M. J. Bessman, et al. (1958). "Enzymatic Synthesis of Deoxyribonucleic Acid: I. PREPARATION OF SUBSTRATES AND PARTIAL PURIFICATION OF AN ENZYME FROM *ESCHERICHIA COLI*." Journal of Biological Chemistry **233**(1): 163-170.
- Lei, M. and B. K. Tye (2001). "Initiating DNA synthesis: from recruiting to activating the MCM complex." Journal of Cell Science **114**(8): 1447-1454.
- Li, S., C. M. Armstrong, et al. (2004). "A Map of the Interactome Network of the Metazoan *C. elegans*." Science **303**(5657): 540-543.
- Li, Z., S. Lu, et al. (2008). "Hjm/Hel308A DNA Helicase from *Sulfolobus tokodaii* Promotes Replication Fork Regression and Interacts with Hjc Endonuclease In Vitro." Journal of Bacteriology **190**(8): 3006-3017.
- Li, Z., M. Pan, et al. (2011). "A novel DNA nuclease is stimulated by association with the GINS complex." Nucleic Acids Research **39**(14): 6114-6123.
- Li, Z., M. Pan, et al. (2011). "A novel DNA nuclease is stimulated by association with the GINS complex." Nucleic Acids Res **39**(14): 6114-23.

- Li, Z., T. J. Santangelo, et al. (2010). "Affinity Purification of an Archaeal DNA Replication Protein Network." mBio **1**(5).
- Li, Z., T. J. Santangelo, et al. (2010). "Affinity purification of an archaeal DNA replication protein network." MBio **1**(5).
- Lin, Y., C. E. Guzman, et al. (2006). "Methanosarcina acetivorans flap endonuclease 1 activity is inhibited by a cognate single-stranded-DNA-binding protein." J Bacteriol **188**(17): 6153-67.
- Lin, Y., L. J. Lin, et al. (2008). "Engineering of functional replication protein a homologs based on insights into the evolution of oligonucleotide/oligosaccharide-binding folds." J Bacteriol **190**(17): 5766-80.
- Lin, Y., J. B. Robbins, et al. (2005). "A CCCH Zinc Finger Conserved in a Replication Protein A Homolog Found in Diverse Euryarchaeotes." Journal of Bacteriology **187**(23): 7881-7889.
- Lindahl, T. (1993). "Instability and decay of the primary structure of DNA." Nature **362**(6422): 709-15.
- Lindahl, T. (1993). "Instability and decay of the primary structure of DNA." Nature **362**(6422): 709-715.
- Lindahl, T. and B. Nyberg (1972). "Rate of depurination of native deoxyribonucleic acid." Biochemistry **11**(19): 3610-3618.
- Liu, H., J. Rudolf, et al. (2008). "Structure of the DNA repair helicase XPD." Cell **133**(5): 801-12.
- Liu, J., C. L. Smith, et al. (2000). "Structure and Function of Cdc6/Cdc18: Implications for Origin Recognition and Checkpoint Control." Molecular Cell **6**(3): 637-648.
- Liu, L., K. Komori, et al. (2001). "The Archaeal DNA Primase." Journal of Biological Chemistry **276**(48): 45484-45490.
- Lucas-Lledo, J. I. and M. Lynch (2009). "Evolution of mutation rates: phylogenomic analysis of the photolyase/cryptochrome family." Mol Biol Evol **26**(5): 1143-53.
- Ma, Y., U. Pannicke, et al. (2002). "Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination." Cell **108**(6): 781-94.
- Mackwan, R. R., G. T. Carver, et al. (2007). "An unusual pattern of spontaneous mutations recovered in the halophilic archaeon *Haloferax volcanii*." Genetics **176**(1): 697-702.
- MacNeill, S. A. (2011). "Protein-protein interactions in the archaeal core replisome." Biochem Soc Trans **39**(1): 163-8.
- Maga, G., G. Villani, et al. (2002). "Human DNA Polymerase δ Functionally and Physically Interacts with Proliferating Cell Nuclear Antigen in Normal and Translesion DNA Synthesis." Journal of Biological Chemistry **277**(50): 48434-48440.
- Makarova, K., E. Koonin, et al. (2012). "The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes." Biology Direct **7**(1): 7.
- Makarova, K. S., E. V. Koonin, et al. (2012). "The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes." Biol Direct **7**: 7.
- Makarova, K. S., A. V. Sorokin, et al. (2007). "Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea." Biol Direct **2**: 33.
- Makarova, K. S., N. Yutin, et al. (2010). "Evolution of diverse cell division and vesicle formation systems in Archaea." Nat Rev Micro **8**(10): 731-741.
- Malandrin, L., H. Huber, et al. (1999). "Nucleoid Structure and Partition in *Methanococcus jannaschii*: An Archaeon With Multiple Copies of the Chromosome." Genetics **152**(4): 1315-1323.

- Maloisel, L., F. Fabre, et al. (2008). "DNA polymerase delta is preferentially recruited during homologous recombination to promote heteroduplex DNA extension." Mol Cell Biol **28**(4): 1373-82.
- Mansfield, Robyn E. and Nicholas E. Dixon (2012). "Priming the Engine of DNA Synthesis." Structure **20**(9): 1447-1448.
- Manzan, A., G. Pfeiffer, et al. (2004). "MlaA, a hexameric ATPase linked to the Mre11 complex in archaeal genomes." EMBO Rep **5**(1): 54-9.
- Marinsek, N., E. R. Barry, et al. (2006). "GINS, a central nexus in the archaeal DNA replication fork." EMBO Rep **7**(5): 539-545.
- Marti, T. M., C. Kunz, et al. (2002). "DNA mismatch repair and mutation avoidance pathways." Journal of Cellular Physiology **191**(1): 28-41.
- Maslov, S. and K. Sneppen (2002). "Specificity and Stability in Topology of Protein Networks." Science **296**(5569): 910-913.
- Mathews, M. B., R. M. Bernstein, et al. (1984). "Identity of the proliferating cell nuclear antigen and cyclin." Nature **309**(5966): 374-376.
- Matic, I., M. Radman, et al. (1997). "Highly Variable Mutation Rates in Commensal and Pathogenic Escherichia coli." Science **277**(5333): 1833-1834.
- Matsui, E., J. Abe, et al. (2004). "Aromatic Residues Located Close to the Active Center Are Essential for the Catalytic Reaction of Flap Endonuclease-1 from Hyperthermophilic Archaeon Pyrococcus horikoshii." Journal of Biological Chemistry **279**(16): 16687-16696.
- Matsui, E., S. Kawasaki, et al. (1999). "Thermostable Flap Endonuclease from the Archaeon, Pyrococcus horikoshii, Cleaves the Replication Fork-like Structure Endo/Exonucleolytically." Journal of Biological Chemistry **274**(26): 18297-18309.
- Matsumiya, S., Y. Ishino, et al. (2001). "Crystal structure of an archaeal DNA sliding clamp: Proliferating cell nuclear antigen from Pyrococcus furiosus." Protein Science **10**(1): 17-23.
- Matsumoto, Y. and S. M. A. K. M. R. S. L. S. H. W. Kivie Moldave (2001). Molecular mechanism of PCNA-dependent base excision repair. Progress in Nucleic Acid Research and Molecular Biology, Academic Press. **Volume 68**: 129-138.
- Matsunaga, F., P. Forterre, et al. (2001). "In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin." Proceedings of the National Academy of Sciences **98**(20): 11152-11157.
- Matsunaga, F., C. Norais, et al. (2003). "Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin." EMBO Rep **4**: 154 - 158.
- Matsunaga, F., C. Norais, et al. (2003). "Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin." EMBO Rep **4**(2): 154-8.
- McCready, S., J. A. Muller, et al. (2005). "UV irradiation induces homologous recombination genes in the model archaeon, Halobacterium sp. NRC-1." Saline Systems **1**: 3.
- McRobbie, A.-M., L. G. Carter, et al. (2009). "Structural and Functional Characterisation of a Conserved Archaeal RadA Paralog with Antirecombinase Activity." Journal of Molecular Biology **389**(4): 661-673.
- Meselson, M. and F. W. Stahl (1958). "The replication of DNA in Escherichia coli." Proceedings of the National Academy of Sciences **44**(7): 671-682.
- Meslet-Cladière, L., C. d. Norais, et al. (2007). "A Novel Proteomic Approach Identifies New Interaction Partners for Proliferating Cell Nuclear Antigen." Journal of Molecular Biology **372**(5): 1137-1148.

- Meslet-Cladiere, L., C. Norais, et al. (2007). "A novel proteomic approach identifies new interaction partners for proliferating cell nuclear antigen." J Mol Biol **372**(5): 1137-48.
- Meslet-Cladière, L., C. d. Norais, et al. (2007). "A Novel Proteomic Approach Identifies New Interaction Partners for Proliferating Cell Nuclear Antigen." Journal of Molecular Biology **372**(5): 1137-1148.
- Meslet-Cladière, L., C. d. Norais, et al. (2007). "A Novel Proteomic Approach Identifies New Interaction Partners for Proliferating Cell Nuclear Antigen." Journal of Molecular Biology **372**(5): 1137-1148.
- Milenkovic, T., V. Memisevic, et al. (2011). "Dominating biological networks." PLoS One **6**(8): e23016.
- Minton, A. P. (2000). "Implications of macromolecular crowding for protein assembly." Current Opinion in Structural Biology **10**(1): 34-39.
- Mirzoeva, O. K. and J. H. J. Petrini (2003). "DNA Replication-Dependent Nuclear Dynamics of the Mre11 Complex1 1 NIH GM56888 and the DOE ER62859." Molecular Cancer Research **1**(3): 207-218.
- Miyachi, K., M. J. Fritzler, et al. (1978). "Autoantibody to a Nuclear Antigen in Proliferating Cells." The Journal of Immunology **121**(6): 2228-2234.
- Moldovan, G.-L., B. Pfander, et al. (2007). "PCNA, the Maestro of the Replication Fork." Cell **129**(4): 665-679.
- Motz, M., I. Kober, et al. (2002). "Elucidation of an Archaeal Replication Protein Network to Generate Enhanced PCR Enzymes." Journal of Biological Chemistry **277**(18): 16179-16188.
- Motz, M., I. Kober, et al. (2002). "Elucidation of an archaeal replication protein network to generate enhanced PCR enzymes." J Biol Chem **277**(18): 16179-88.
- Nakagawa, T., A. Datta, et al. (1999). "Multiple functions of MutS- and MutL-related heterocomplexes." Proceedings of the National Academy of Sciences **96**(25): 14186-14188.
- Nakamura, T., Y. Zhao, et al. (2012). "Watching DNA polymerase [eegr] make a phosphodiester bond." Nature **487**(7406): 196-201.
- Nakane, S., N. Nakagawa, et al. (2012). "The role of the PHP domain associated with DNA polymerase X from *Thermus thermophilus* HB8 in base excision repair." DNA Repair **11**(11): 906-914.
- Newman, M., J. Murray-Rust, et al. (2005). "Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition." EMBO J **24**(5): 895-905.
- Ng, S. Y. M., B. Zolghadr, et al. (2008). "Cell Surface Structures of Archaea." Journal of Bacteriology **190**(18): 6039-6047.
- Ni, X. Y., T. Drenth, et al. (2009). "The Control of the Controller: Molecular Mechanisms for Robust Perfect Adaptation and Temperature Compensation." Biophysical Journal **97**(5): 1244-1253.
- Niu, H., S. Raynard, et al. (2009). "Multiplicity of DNA end resection machineries in chromosome break repair." Genes & development **23**(13): 1481-1486.
- Noble, D. (2008). "Claude Bernard, the first systems biologist, and the future of physiology." Experimental Physiology **93**(1): 16-26.
- Noireaux, V., Y. T. Maeda, et al. (2011). "Development of an artificial cell, from self-organization to computation and self-reproduction." Proceedings of the National Academy of Sciences **108**(9): 3473-3480.
- Oksenyich, V., B. Bernardes de Jesus, et al. (2009). "Molecular insights into the recruitment of TFIIH to sites of DNA damage." EMBO J **28**(19): 2971-80.

- Pace, N. R. (2009). "Mapping the Tree of Life: Progress and Prospects." Microbiology and Molecular Biology Reviews **73**(4): 565-576.
- Padliya, N. D. and T. D. Wood (2004). "A strategy to improve peptide mass fingerprinting matches through the optimization of matrix-assisted laser desorption/ionization matrix selection and formulation." PROTEOMICS **4**(2): 466-473.
- Palud, A., G. Villani, et al. (2008). "Intrinsic properties of the two replicative DNA polymerases of *Pyrococcus abyssi* in replicating abasic sites: possible role in DNA damage tolerance?" Molecular Microbiology **70**(3): 746-761.
- Palud, A., G. Villani, et al. (2008). "Intrinsic properties of the two replicative DNA polymerases of *Pyrococcus abyssi* in replicating abasic sites: possible role in DNA damage tolerance?" Mol Microbiol **70**(3): 746-61.
- Pan, M., L. M. Kelman, et al. (2011). "The archaeal PCNA proteins." Biochem Soc Trans **39**(1): 20-4.
- Patel, S. S. and K. M. Picha (2000). "STRUCTURE AND FUNCTION OF HEXAMERIC HELICASES1." Annual Review of Biochemistry **69**(1): 651-697.
- Pegg, A. E. (2000). "Repair of O6-alkylguanine by alkyltransferases." Mutation Research/Reviews in Mutation Research **462**(2&3): 83-100.
- Pelve, E. A., A.-C. Lindås, et al. (2012). "Four chromosome replication origins in the archaeon *Pyrobaculum calidifontis*." Molecular Microbiology **85**(5): 986-995.
- Perugino, G., A. Valenti, et al. (2009). "Reverse gyrase and genome stability in hyperthermophilic organisms." Biochemical Society Transactions **37**(Pt 1): 69-73.
- Perugino, G., A. Vettone, et al. (2012). "Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage." J Biol Chem **287**(6): 4222-31.
- Perugino, G., A. Vettone, et al. (2012). "Activity and Regulation of Archaeal DNA Alkyltransferase: CONSERVED PROTEIN INVOLVED IN REPAIR OF DNA ALKYLATION DAMAGE." Journal of Biological Chemistry **287**(6): 4222-4231.
- Pestryakov, P. E. and O. I. Lavrik (2008). "Mechanisms of Single-Stranded DNA-Binding Protein Functioning in Cellular DNA Metabolism." Biochemistry (Moscow).
- Pike, J. E., P. M. J. Burgers, et al. (2009). "Pif1 Helicase Lengthens Some Okazaki Fragment Flaps Necessitating Dna2 Nuclease/Helicase Action in the Two-nuclease Processing Pathway." Journal of Biological Chemistry **284**(37): 25170-25180.
- Pisani, F. M., M. De Felice, et al. (2000). "Biochemical characterization of a clamp-loader complex homologous to eukaryotic replication factor C from the hyperthermophilic archaeon *Sulfolobus solfataricus*." Journal of Molecular Biology **301**(1): 61-73.
- Poinsignon, C., D. Moshous, et al. (2004). "The metallo-beta-lactamase/beta-CASP domain of Artemis constitutes the catalytic core for V(D)J recombination." J Exp Med **199**(3): 315-21.
- Prolla, T. A., D. M. Christie, et al. (1994). "Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homologs of the bacterial mutL gene." Molecular and Cellular Biology **14**(1): 407-415.
- Pugh, R. A., Y. Lin, et al. (2008). "Ferroplasma acidarmanus RPA2 Facilitates Efficient Unwinding of Forked DNA Substrates by Monomers of FacXPD Helicase." Journal of Molecular Biology **383**(5): 982-998.
- Quaiser, A., F. Constantinesco, et al. (2008). "The Mre11 protein interacts with both Rad50 and the HerA bipolar helicase and is recruited to DNA following gamma irradiation in the archaeon *Sulfolobus acidocaldarius*." BMC Molecular Biology **9**(1): 25.

- Quaiser, A., F. Constantinesco, et al. (2008). "The Mre11 protein interacts with both Rad50 and the HerA bipolar helicase and is recruited to DNA following gamma irradiation in the archaeon *Sulfolobus acidocaldarius*." BMC Mol Biol **9**: 25.
- R., D., K. M., et al. (1998). Thermococcus acidaminovorans sp. nov., a new hyperthermophilic alkalophilic archaeon growing on amino acids. Heidelberg, ALLEMAGNE, Springer.
- Reinhardt, H. C. and B. r. Schumacher "The p53 network: cellular and systemic DNA damage responses in aging and cancer." Trends in Genetics **28**(3): 128-136.
- Ren, B., J. Kuhn, et al. (2009). "Structure and function of a novel endonuclease acting on branched DNA substrates." EMBO J **28**(16): 2479-89.
- Ren, B., J. Kuhn, et al. (2007). "Crystallization and preliminary X-ray analysis of a RecB-family nuclease from the archaeon *Pyrococcus abyssi*." Acta Crystallographica Section F **63**(5): 406-408.
- Reyes-Lamothe, R., C. Possoz, et al. (2008). "Independent Positioning and Action of *Escherichia coli* Replisomes in Live Cells." Cell **133**(1): 90-102.
- Richard, D. J., S. D. Bell, et al. (2004). "Physical and functional interaction of the archaeal single-stranded DNA-binding protein SSB with RNA polymerase." Nucleic Acids Res **32**(3): 1065-74.
- Richards, J. D., L. Cubeddu, et al. (2008). "The archaeal XPB protein is a ssDNA-dependent ATPase with a novel partner." J Mol Biol **376**(3): 634-44.
- Riedl, T., F. Hanaoka, et al. (2003). "The comings and goings of nucleotide excision repair factors on damaged DNA." Embo J **22**(19): 5293-5303.
- Robbins, J. B., M. C. McKinney, et al. (2005). "The Euryarchaeota, Nature's Medium for Engineering of Single-stranded DNA-binding Proteins." Journal of Biological Chemistry **280**(15): 15325-15339.
- Robbins, J. B., M. C. Murphy, et al. (2004). "Functional analysis of multiple single-stranded DNA-binding proteins from *Methanosarcina acetivorans* and their effects on DNA synthesis by DNA polymerase BI." J Biol Chem **279**(8): 6315-26.
- Robbins, J. B., M. C. Murphy, et al. (2004). "Functional Analysis of Multiple Single-stranded DNA-binding Proteins from *Methanosarcina acetivorans* and Their Effects on DNA Synthesis by DNA Polymerase BI." Journal of Biological Chemistry **279**(8): 6315-6326.
- Roberts, J. A., S. D. Bell, et al. (2003). "An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA." Mol Microbiol **48**(2): 361-71.
- Roberts, J. A. and M. F. White (2005). "An archaeal endonuclease displays key properties of both eukaryal XPF-ERCC1 and Mus81." J Biol Chem **280**(7): 5924-8.
- Roberts, J. A. and M. F. White (2005). "DNA end-directed and processive nuclease activities of the archaeal XPF enzyme." Nucleic Acids Res **33**(20): 6662-70.
- Rodríguez, A. C. (2002). "Studies of a Positive Supercoiling Machine." Journal of Biological Chemistry **277**(33): 29865-29873.
- Rolfsmeier, M. L. and C. A. Haseltine (2010). "The Single-Stranded DNA Binding Protein of *Sulfolobus solfataricus* Acts in the Presynaptic Step of Homologous Recombination." Journal of Molecular Biology **397**(1): 31-45.
- Romano, V., A. Napoli, et al. (2007). "Lack of Strand-specific Repair of UV-induced DNA Lesions in Three Genes of the Archaeon *Sulfolobus solfataricus*." Journal of Molecular Biology **365**(4): 921-929.
- Romano, V., A. Napoli, et al. (2007). "Lack of strand-specific repair of UV-induced DNA lesions in three genes of the archaeon *Sulfolobus solfataricus*." J Mol Biol **365**(4): 921-9.

- Rouillon, C., G. Henneke, et al. (2007). "DNA polymerase switching on homotrimeric PCNA at the replication fork of the euryarchaea *Pyrococcus abyssi*." J Mol Biol **369**(2): 343-55.
- Rouillon, C., G. Henneke, et al. (2007). "DNA Polymerase Switching on Homotrimeric PCNA at the Replication Fork of the Euryarchaea *Pyrococcus abyssi*." Journal of Molecular Biology **369**(2): 343-355.
- Rouillon, C. and M. F. White (2010). "The evolution and mechanisms of nucleotide excision repair proteins." Research in Microbiology **162**(1): 19-26.
- Rouillon, C. and M. F. White (2010). "The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: implications for eukaryal and archaeal nucleotide excision repair." J Biol Chem **285**(14): 11013-22.
- Rouillon, C. and M. F. White (2011). "The evolution and mechanisms of nucleotide excision repair proteins." Res Microbiol **162**(1): 19-26.
- Royer, L., M. Reimann, et al. (2008). "Unraveling protein networks with power graph analysis." PLoS Comput Biol **4**(7): e1000108.
- Rual, J.-F., K. Venkatesan, et al. (2005). "Towards a proteome-scale map of the human protein-protein interaction network." Nature **437**(7062): 1173-1178.
- Rudolf, J., V. Makrantonis, et al. (2006). "The DNA repair helicases XPD and FancJ have essential iron-sulfur domains." Mol Cell **23**(6): 801-8.
- Rudolf, J., C. Rouillon, et al. (2010). "The helicase XPD unwinds bubble structures and is not stalled by DNA lesions removed by the nucleotide excision repair pathway." Nucleic Acids Res **38**(3): 931-41.
- Russell, H. J., T. T. Richardson, et al. (2009). "The 3'→5' proofreading exonuclease of archaeal family-B DNA polymerase hinders the copying of template strand deaminated bases." Nucleic Acids Research **37**(22): 7603-7611.
- Sakakibara, N., L. M. Kelman, et al. (2009). "Unwinding the structure and function of the archaeal MCM helicase." Molecular Microbiology **72**(2): 286-296.
- Sakofsky, C. J., P. L. Foster, et al. (2012). "Roles of the Y-family DNA polymerase Dbh in accurate replication of the *Sulfolobus* genome at high temperature." DNA Repair **11**(4): 391-400.
- Salathé, M., R. M. May, et al. (2005). "The evolution of network topology by selective removal." Journal of The Royal Society Interface **2**(5): 533-536.
- Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proceedings of the National Academy of Sciences **74**(12): 5463-5467.
- Saraç, A. m. S., V. Pancaldi, et al. (2012). "Topology of functional networks predicts physical binding of proteins." Bioinformatics **28**(16): 2137-2145.
- Saraç, A. m. S., V. Pancaldi, et al. (2012). "Topology of functional networks predicts physical binding of proteins." Bioinformatics **28**(16): 2137-2145.
- Sartori, A. A. and J. Jiricny (2003). "Enzymology of Base Excision Repair in the Hyperthermophilic Archaeon *Pyrobaculum aerophilum*." Journal of Biological Chemistry **278**(27): 24563-24576.
- Sartori, A. A. and J. Jiricny (2003). "Enzymology of base excision repair in the hyperthermophilic archaeon *Pyrobaculum aerophilum*." J Biol Chem **278**(27): 24563-76.
- Savino, C., L. Federici, et al. (2004). "Insights into DNA Replication: The Crystal Structure of DNA Polymerase B1 from the Archaeon *Sulfolobus solfataricus*." Structure **12**(11): 2001-2008.
- Schofield, M. J. and P. Hsieh (2003). "DNA mismatch repair: molecular mechanisms and biological function." Annu Rev Microbiol **57**: 579-608.

- Scott, J. W., S. A. Hawley, et al. (2004). "CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations." J Clin Invest **113**(2): 274-84.
- Shannon, P., A. Markiel, et al. (2003). "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." Genome Research **13**(11): 2498-2504.
- Shapiro, J. A. (2009). "Revisiting the Central Dogma in the 21st Century." Annals of the New York Academy of Sciences **1178**(1): 6-28.
- Shen, Y., K. Musti, et al. (2001). "Invariant Asp-1122 and Asp-1124 Are Essential Residues for Polymerization Catalysis of Family D DNA Polymerase from *Pyrococcus horikoshii*." Journal of Biological Chemistry **276**(29): 27376-27383.
- Sheng, D., S. Zhu, et al. (2008). "The in vitro activity of a Rad55 homologue from *Sulfolobus tokodaii*, a candidate mediator in RadA-catalyzed homologous recombination." Extremophiles **12**(1): 147-157.
- Shereda, R. D., A. G. Kozlov, et al. (2008). "SSB as an Organizer/Mobilizer of Genome Maintenance Complexes." Critical Reviews in Biochemistry and Molecular Biology **43**(5): 289-318.
- Shin, J.-H., B. Grabowski, et al. (2003). "Regulation of Minichromosome Maintenance Helicase Activity by Cdc6." Journal of Biological Chemistry **278**(39): 38059-38067.
- Shin, J.-H., Y. Jiang, et al. (2003). "Substrate requirements for duplex DNA translocation by the eukaryal and archaeal MCM helicases." J. Biol. Chem.: M308599200.
- Shin, J. H., T. J. Santangelo, et al. (2007). "Archaeal minichromosome maintenance (MCM) helicase can unwind DNA bound by archaeal histones and transcription factors." J Biol Chem **282**(7): 4908-15.
- Shuman, S. and M. S. Glickman (2007). "Bacterial DNA repair by non-homologous end joining." Nat Rev Micro **5**(11): 852-861.
- Sikorski, Timothy W., Scott B. Ficarro, et al. (2011). "Sub1 and RPA Associate with RNA Polymerase II at Different Stages of Transcription." Molecular Cell **44**(3): 397-409.
- Sikorski, T. W., S. B. Ficarro, et al. (2011). "Sub1 and RPA associate with RNA polymerase II at different stages of transcription." Mol Cell **44**(3): 397-409.
- Singleton, M. R., M. S. Dillingham, et al. (2004). "Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks." Nature **432**(7014): 187-193.
- Skowrya, A. and S. A. MacNeill (2012). "Identification of essential and non-essential single-stranded DNA-binding proteins in a model archaeal organism." Nucleic Acids Research **40**(3): 1077-1090.
- Skowrya, A. and S. A. MacNeill (2012). "Identification of essential and non-essential single-stranded DNA-binding proteins in a model archaeal organism." Nucleic Acids Res **40**(3): 1077-90.
- Sonoda, E., M. S. Sasaki, et al. (1998). "Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death." Embo J **17**(2): 598-608.
- Spitalny, P. and M. Thomm (2003). "Analysis of the open region and of DNA-protein contacts of archaeal RNA polymerase transcription complexes during transition from initiation to elongation." J Biol Chem **278**(33): 30497-505.
- Spitalny, P. and M. Thomm (2008). "A polymerase III-like reinitiation mechanism is operating in regulation of histone expression in archaea." Molecular Microbiology **67**(5): 958-970.
- Stano, N. M., J. Chen, et al. (2006). "A coproofreading Zn(2+)-dependent exonuclease within a bacterial replicase." Nat Struct Mol Biol **13**(5): 458-9.

- Steinacher, A. and O. S. Soyer (2012). Evolutionary Principles Underlying Structure and Response Dynamics of Cellular Networks
Evolutionary Systems Biology. N. Back, I. R. Cohen, A. Lajtha, J. D. Lambris and R. Paoletti, Springer New York. **751**: 225-247.
- Stelter, M., I. Gutsche, et al. (2012). "Architecture of a Dodecameric Bacterial Replicative Helicase." Structure **20**(3): 554-564.
- Stelzl, U., U. Worm, et al. (2005). "A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome." Cell **122**(6): 957-968.
- Stetter, K. O. (1999). "Extremophiles and their adaptation to hot environments." FEBS Letters **452**(1&2): 22-25.
- Stracker, T. H. and J. H. Petrini (2011). "The MRE11 complex: starting from the ends." Nat Rev Mol Cell Biol **12**(2): 90-103.
- Takayama, Y., Y. Kamimura, et al. (2003). "GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast." Genes & development **17**(9): 1153.
- Tang, X.-F., Y. Shen, et al. (2004). "Domain Topology of the DNA Polymerase D Complex from a Hyperthermophilic Archaeon *Pyrococcus horikoshii*" Biochemistry **43**(37): 11818-11827.
- Teplyakov, A., G. Obmolova, et al. (2003). "Crystal structure of the Escherichia coli YcdX protein reveals a trinuclear zinc active site." Proteins **51**(2): 315-8.
- Tom, S., L. A. Henricksen, et al. (2000). "Mechanism Whereby Proliferating Cell Nuclear Antigen Stimulates Flap Endonuclease 1." Journal of Biological Chemistry **275**(14): 10498-10505.
- Tomkinson, A. E., S. Vijayakumar, et al. (2006). "DNA Ligases: Structure, Reaction Mechanism, and Function." Chemical Reviews **106**(2): 687-699.
- Uemori, T., Y. Sato, et al. (1997). "A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus* : gene cloning, expression, and characterization." Genes Cells **2**: 499 - 512.
- Uetz, P., L. Giot, et al. (2000). "A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*." Nature **403**(6770): 623-627.
- Umar, A., A. B. Buermeier, et al. (1996). "Requirement for PCNA in DNA Mismatch Repair at a Step Preceding DNA Resynthesis." Cell **87**(1): 65-73.
- Venter, J. C., M. D. Adams, et al. (2001). "The Sequence of the Human Genome." Science **291**(5507): 1304-1351.
- Vivona, J. B. and Z. Kelman (2003). "The diverse spectrum of sliding clamp interacting proteins." FEBS Letters **546**(2-3): 167-172.
- von Mering, C., R. Krause, et al. (2002). "Comparative assessment of large-scale data sets of protein-protein interactions." Nature **417**(6887): 399-403.
- Waage, I., G. Schmid, et al. (2010). "Shuttle vector-based transformation system for *Pyrococcus furiosus*." Appl Environ Microbiol **76**(10): 3308-13.
- Wang, H. H., F. J. Isaacs, et al. (2009). "Programming cells by multiplex genome engineering and accelerated evolution." Nature **460**(7257): 894-898.
- Warbrick, E. (2000). "The puzzle of PCNA's many partners." Bioessays **22**(11): 997-1006.

- Wardle, J., P. M. J. Burgers, et al. (2008). "Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya." Nucleic Acids Research **36**(3): 705-711.
- Watson, J. D. and F. H. C. Crick (1953). "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid." Nature **171**(4356): 737-738.
- Watts, D. J. and S. H. Strogatz (1998). "Collective dynamics of 'small-world' networks." Nature **393**(6684): 440-442.
- West, S. C. (1996). "The RuvABC proteins and Holliday junction processing in *Escherichia coli*." Journal of Bacteriology **178**(5): 1237-41.
- White, M. F. (2009). "Structure, function and evolution of the XPD family of iron-sulfur-containing 5'→3' DNA helicases." Biochem Soc Trans **37**(Pt 3): 547-51.
- White, M. F. and S. D. Bell (2002). "Holding it together: chromatin in the Archaea." Trends in Genetics **18**(12): 621-626.
- Wilkinson, A., J. Day, et al. (2001). "Bacterial DNA ligases." Molecular Microbiology **40**(6): 1241-1248.
- Woese, C. R. and G. E. Fox (1977). "Phylogenetic structure of the prokaryotic domain: The primary kingdoms." Proceedings of the National Academy of Sciences **74**(11): 5088-5090.
- Woese, C. R., O. Kandler, et al. (1990). "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya." Proceedings of the National Academy of Sciences **87**(12): 4576-4579.
- Yan, J., S. C. M. Haaijer, et al. (2012). "Mimicking the oxygen minimum zones: stimulating interaction of aerobic archaeal and anaerobic bacterial ammonia oxidizers in a laboratory-scale model system." Environmental Microbiology: n/a-n/a.
- Yellaboina, S., A. Tasneem, et al. (2011). "DOMINE: a comprehensive collection of known and predicted domain-domain interactions." Nucleic Acids Res **39**(Database issue): D730-5.
- Yoshimochi, T., R. Fujikane, et al. (2008). "The GINS Complex from *Pyrococcus furiosus* Stimulates the MCM Helicase Activity." Journal of Biological Chemistry **283**(3): 1601-1609.
- Yu, X., J. Ivanic, et al. (2009). "A Novel Scoring Approach for Protein Co-Purification Data Reveals High Interaction Specificity." PLoS Comput Biol **5**(9): e1000515.
- Yutin, N., K. S. Makarova, et al. (2008). "The deep archaeal roots of eukaryotes." Mol Biol Evol **25**(8): 1619-30.
- Zeng, X., J.-L. Birrien, et al. (2009). "Pyrococcus CH1, an obligate piezophilic hyperthermophile: extending the upper pressure-temperature limits for life." ISME J **3**(7): 873-876.
- Zhang, S., T. Wei, et al. (2008). "Archaeal DNA helicase HerA interacts with Mre11 homologue and unwinds blunt-ended double-stranded DNA and recombination intermediates." DNA Repair **7**(3): 380-391.
- Zou, Y., Y. Liu, et al. (2006). "Functions of human replication protein A (RPA): From DNA replication to DNA damage and stress responses." Journal of Cellular Physiology **208**(2): 267-273.
- Zuo, Z., C. J. Rodgers, et al. (2010). "Characterization of a Functional DnaG-Type Primase in Archaea: Implications for a Dual-Primase System." Journal of Molecular Biology **397**(3): 664-676.

Annexe 1 : Groisillier, A., C. Herve, et al. (2010). "MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms." *Microbial Cell Factories* 9(1): 45.

RESEARCH

Open Access

MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms

Agnès Groisillier^{1,2}, Cécile Hervé^{1,2}, Alexandra Jeudy^{1,2}, Etienne Rebuffet^{1,2}, Pierre F Pluchon³, Yann Chevotot⁴, Didier Flament³, Claire Geslin³, Isabel M Morgado⁵, Déborah Power⁵, Margherita Branno⁶, Hervé Moreau⁷, Gurvan Michel^{1,2}, Catherine Boyen^{1,2} and Mirjam Czjzek*^{1,2}

Abstract

Background: The production of stable and soluble proteins is one of the most important steps prior to structural and functional studies of biological importance. We investigated the parallel production in a medium throughput strategy of genes coding for proteins from various marine organisms, using protocols that involved recombinatorial cloning, protein expression screening and batch purification. This strategy was applied in order to respond to the need for post-genomic validation of the recent success of a large number of marine genomic projects. Indeed, the upcoming challenge is to go beyond the bioinformatic data, since the bias introduced through the genomes of the so called model organisms leads to numerous proteins of unknown function in the still unexplored world of the oceanic organisms.

Results: We present here the results of expression tests for 192 targets using a 96-well plate format. Genes were PCR amplified and cloned in parallel into expression vectors pFO4 and pGEX-4T-1, in order to express proteins N-terminally fused to a six-histidine-tag and to a GST-tag, respectively. Small-scale expression and purification permitted isolation of 84 soluble proteins and 34 insoluble proteins, which could also be used in refolding assays. Selected examples of proteins expressed and purified to a larger scale are presented.

Conclusions: The objective of this program was to get around the bottlenecks of soluble, active protein expression and crystallization for post-genomic validation of a number of proteins that come from various marine organisms. Multiplying the constructions, vectors and targets treated in parallel is important for the success of a medium throughput strategy and considerably increases the chances to get rapid access to pure and soluble protein samples, needed for the subsequent biochemical characterizations. Our set up of a medium throughput strategy applied to genes from marine organisms had a mean success rate of 44% soluble protein expression from marine bacteria, archaea as well as eukaryotic organisms. This success rate compares favorably with other protein screening projects, particularly for eukaryotic proteins. Several purified targets have already formed the base for experiments aimed at post-genomic validation.

Background

The marine environment is highly complex and contains the vast majority of known and unknown biodiversity. It is also the last frontier to understand the control of the global climate and hides a wealth of biological resources

still to be tapped for food, health and energy. Up until very recently, few genomic data were available for oceanic organisms, but this panorama is rapidly changing with a number of genomic projects now underway, which focus on marine organisms, ranging from microbes [1,2] to multicellular eukaryotes including vertebrates or macroalgae [3], as well as the generation of resources and access to genomes or EST libraries for various eukaryotic systems [4-7]. The wealth of sequence data arising from

* Correspondence: czjzek@sb-roscoff.fr

¹ UPMC Univ Paris 6, UMR 7139 Végétaux marins et Biomolécules, LIA DIAMS, Station Biologique, F 29682, Roscoff, France

† Contributed equally

Full list of author information is available at the end of the article

these projects, means that researchers are confronted with a huge number of putative genes, the function of which are, at best, so far only deduced from sequence comparisons (automatic annotation). The pressing question is how to analyze the genomic data with respect to original biological processes in diverse marine organisms (i.e. their development or stress response), their importance in adaptation to the particular habitat and how to identify new enzymes and/or metabolites of biotechnological interest. Thus, the availability of complete genome data has resulted in the development of transcriptomic and proteomic methods that can be used to study regulatory networks and interactions of thousands of genes in parallel, allowing an efficient global analysis of genomic information. However, there are a number of clear drawbacks with these methods in so far that they are strongly dependent on the quality of the genome annotation, which at present assumes conserved functions across often widely distant taxa. Furthermore, these techniques give at most only an indication of the regulation/metabolic pathway the corresponding gene product belongs to, and little or no information on the precise biochemical function of unknown genes.

To understand the precise biological function of a single gene, the biochemical and physiological characterization of its product is essential and this is often greatly aided by the availability of 3-D structural information. Although the 3D structure does not always reveal the natural substrate, it has been shown repeatedly that it helps at least find the class of compounds among which the substrate will be found [8,9]. Several bottlenecks exist in the analysis of individual proteins; generally the techniques utilized require systems for the efficient over-

expression of the target gene in order to produce sufficient recombinant protein. Furthermore, to constitute an assessment for potential biotechnological applications of the discovered proteins/enzymes, the effective recombinant expression of biologically active proteins is essential. When aiming at the 3D-structure of the protein of interest, a second bottleneck is encountered at the step of crystallization of these proteins. Recent developments in the field of structural genomics have demonstrated that medium/high throughput strategies are most adapted to the production of large numbers of soluble and active gene products and/or protein crystals at a time [10,11], since they allow simultaneous testing of numerous conditions with an optimized effort.

In this study, our objective was to set up and apply a medium throughput strategy for the production, expression and, at least for some candidates, the crystallization of selected proteins from marine organisms. With this approach, we targeted proteins from various marine organisms including archaea, bacteria and eukaryotes but also proteins from different families of interest as for example enzymes involved in the metabolism of carbohydrates, stress related proteins or cellular division proteins (Table 1). On the basis of its ease of genetic manipulation, its rapid growth rate, low cost and robustness *Escherichia coli* still remains the most popular expression system [12]. Moreover, the results of high-throughput programs (such as those funded under the NIH Protein structure Initiative (PSI) [13] available on the site PSI/targetDB <http://targetdb.pdb.org/> show that in term of full-length proteins, up to 50% of proteins from the Eubacteria or Archaea and 10% of proteins from Eukarya can be expressed in *E. coli* in soluble form [14]. We have there-

Table 1: Availability of genomic data to the Marine Express partners.

Organism	Taxonomy	Gene families of interest	State of genomic data	Partner
<i>Pyrococcus abyssi</i>	Euryarchaeota	Proteins from the DNA replication system	Genome published [1]	IFREMER Brest, France
<i>Zobellia galactanivorans</i>	Flavobacteria	Polysacchride metabolism, sulfatases	Genome complete	SBR Roscoff, France
<i>Ectocarpus siliculosus</i>	Brown alga	Stress related genes, carbohydrate active enzymes	14000 ESTs, Genome complete	SBR Roscoff, France
<i>Ostreococcus tauri</i>	Green microalga	Biological clock	Genome published [4]	Arago Banyuls, France
<i>Ciona intestinalis</i>	Urochordate	Hox-genes, Ci-msx, Ci-RX	Genome published [5]	SZ A. Dohrn, Napoli, Italia
<i>Sparus aurata</i>	Teleost fish	Hormones, calcium and musculo-skeletal development, stress related	30000 ESTs and 20 full length cDNAs	CCMAR Faro, Protugal
<i>Archaeal virus (PAV1)</i>	Not yet classified	Any	Genome published [28]	UBO, LM2E, Brest, France

fore chosen this expression system for our purpose and report the results of the parallel expression strategy in *E. coli* 192 targeted genes, in two compatible expression vectors, using a 96-well plate format. A selection of individual targets that were successfully overexpressed in this program, were subsequently further characterized in order to determine their functional partners, cellular localization, 3D structure and/or biochemical activity. As a conclusion, three selected examples of these studies are briefly discussed.

Results

Selection and bioinformatics analysis of the target genes

A subset of 200 target genes was selected by the different partners involved in the program to evaluate the potential for heterologous expression of marine organism genes in *E. coli*. The gene families of interest, defined by the partners and organized according to the organism of origin and from three domains of life, are shown in Table 2. The sequences of the genes and their corresponding predicted proteins were analyzed by bioinformatic tools to assess their suitability for expression in our bacterial medium throughput strategy. The nucleotide-sequences were screened for the absence of the selected restriction sites for the gene cloning strategy. Where possible, the potential signal peptides and transmembrane domains were removed in order to express soluble proteins. These were therefore identified using SignalP and TMHMM servers, respectively [15,16]. Another important feature that was analyzed in this step was the protein modularity. It is often assumed that single domains of proteins contain important biological functions and are more likely to be successfully expressed than the large, multi-modular proteins. Moreover, the modularity of a protein is usually an obstacle to its crystallization. Consequently, targeted proteins were, where possible, cloned as full length proteins,

but also in the form of their individual modules. The modular architecture of each target was examined using Blast queries against UniProt database, as well as domain searches with the InterPro server [17]. The crucial choices of the N- and C-terminal boundaries of each module were refined using Hydrophobic Cluster Analysis (HCA) [18]. From 200 starting sequences, eighteen were eliminated after restriction site analysis and eight because of the presence of transmembrane zones throughout the protein sequences.. Finally, 192 sequences encoding proteins ranging from 7 to 140 kDa were analyzed and classified according to different restriction site strategies as follows: *Bam*HI/*Eco*RI, *Bam*HI/*Mfe*I, *Bgl*II/*Eco*RI and *Bgl*II/*Mfe*I.

Amplification and cloning of DNA fragments

To optimize the amplification steps through PCR, specific primers were designed which had the same theoretical T_m value for all targets.

The starting genetic material was genomic DNA for prokaryotes and full length cDNA for eukaryotic organisms. Therefore, in parallel to the bioinformatic analysis, an important step for eukaryotic proteins was to retrieve full length cDNAs from the genetic material. This step was performed separately by each partner by 5'RACE PCR on available cDNA libraries. Selected target genes were amplified by PCR using a general set of gene-specific primers (cf Methods). Even non-amplified targets were included in subsequent steps and permitted recovery of two genes in the cloning step. Overall, 183 out of 192 PCR fragments were obtained (Figure 1). These recovered, non-amplified genes corresponded to cDNA (obtained in weak concentration) or genomic DNA with a size greater than 3500 base pairs.

In the present study, two expression vectors were used to test recombinant protein expression and solubility.

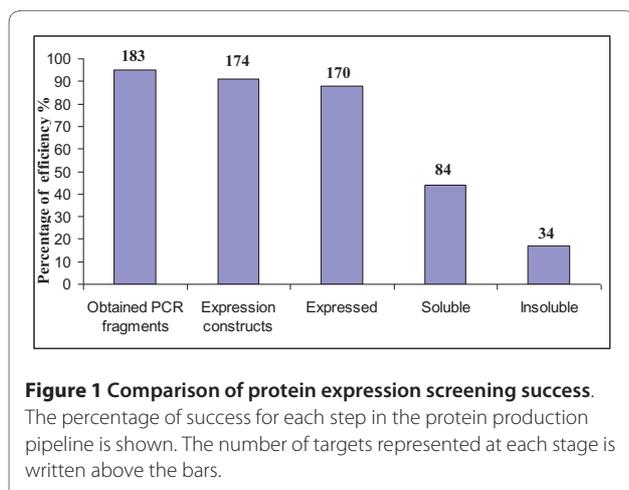
Table 2: Comparison of obtained results covering three domains of living organisms

	Eukarya		Bacteria		Archaea		Total		
	pFO4	pGEX	pFO4	pGEX	pFO4	pGEX	pFO4	pGEX	combined ^c
number of soluble proteins	14	13	13	25	18	26	45	64	84
number of insoluble proteins	10	19	4	5	12	24	26	48	34
no expression ^a	42	37	11	2	31	11	84	50	52
no clone ^b	8	5	4	0	9	12	37	30	22
percentage of soluble proteins	7%	7%	7%	13%	9%	14%	23%	33%	44%
percentage of soluble + insoluble proteins	13,00%	17%	9%	16%	16%	26%	37%	58%	61%

^a number of targets for which no expected bands were visualized on SDS-PAGE gels

^b number of target genes did not clone in expression vectors

^c number of different soluble proteins obtained in pGEX or in pFO4 vectors



One vector (pFO4) contains a His₆-tag at the N-terminal for affinity purification; the second (pGEX-4T-1) allows the production of a fusion protein with an N-terminal glutathion-S-transferase (GST). This parallel PCR cloning procedure could easily be performed in 96-well format. The approach relies on the use of a single PCR product for each gene that is compatible for ligation to both the expression vectors, pGEX-4T-1 and pFO4. Since the expression vectors were digested with *Bam*HI and *Eco*RI, the upstream and downstream PCR primers introduced *Bam*HI (or its isocaudomer *Bgl*II) and *Eco*RI (or its isocaudomer *Mfe*I) restriction sites, respectively, upon PCR amplification. After transformation in *E. coli* DH5 α strains, the plasmids were validated by PCR screening of colonies using primers specific for the expression vectors and flanking the cloning sites. In this way, we obtained 174 cloned target genes (Figure 1). The efficiency for direct cloning of these target genes from PCR products was 95% (174 out of 183).

Expression and purification of recombinant proteins in small-scale experiments

The validated plasmids were used to transform appropriate *E. coli* expression strains (Table 3). Fusion proteins in pFO4 vector are under-control of a T7 promoter and in *E. coli* cells containing a chromosomally located defective prophage DE3 must be used for transformation. For Seventy-seven genes that were cloned from archaea Rosetta or Rosetta (DE3) strains were transformed, which compensate for a number of rare codons in *E. coli*. For cloned genes that contain several cysteines in their sequences Origami or Origami (DE3) strains (14 genes) were transformed. Indeed, these cells carry mutations for both the thioredoxin reductase (*trx*B) and glutathione reductase (*gor*) genes, mutations which greatly enhance disulfide bond formation in the cytoplasm. For all other recombinant vectors BL21 or BL21 (DE3) strains were used to be transformed.

For 167 out of 183 cloned constructs that contained inserts of the expected size small-scale experiments for soluble protein expression were screened. A key step in the automation of the small-scale experimental setup is the development of auto-induction media (ZYP5052 medium used in the present study), which contain differentially metabolized carbon sources that promote growth to relatively high cell densities and then auto-induce by the utilization of lactose. These media remove the need to monitor cell densities or to add an inducer such as IPTG in T7-based expression systems [19]. To further optimize soluble protein expression, we used a culture temperature of 20°C for 3 days, since it has been established by the *Structural Proteomics In Europe* consortium [20] and others that lower temperatures tend to be more effective [21]. Moreover, in the present study 24-well plates instead of the usual 96-well plates [14] were used to obtain better aeration of the cells [22]. The final OD₆₀₀ values that were reached in the small-scale cultivation vessels ranged between 10 and 16.

The expression and the solubility of the recombinant proteins were tested by SDS-PAGE. Only for few target clones it was possible to visually detect the protein band corresponding to the fusion protein directly on SDS-PAGE gels, as illustrated for some examples in Figure 2a. Consequently, to validate the exact number of clones that expressed proteins in the soluble fractions, we used Microspin GST/HIS mini-columns (GE Healthcare Life Science, USA) to perform a mini-purification screen. Indeed, these mini-columns are used to purify a GST fusion protein or a HIS fusion protein by affinity chromatography. After incubation and washing, bound His-tagged or GST-tagged proteins were eluted with elution buffer containing imidazole or reducing-glutathion, respectively. The elution results were again analyzed on SDS-PAGE gels (Figure 2b). This evaluation method was much more sensitive and revealed that 118 proteins out of the 174 constructs (68%) were expressed, while 84 of these 118 (71%) were found to be soluble (Table 2 and Additional file 1, Table S1).

Scale-up expression and purification of targets from *Pyrococcus abyssi*

A relative high success rate (45% of solubly expressed targets) was obtained for targets originating from the hyperthermophilic archaee *Pyrococcus abyssi*. A total of 37 out of 77 expressed proteins were expressed soluble and easily purified in the small scale set up. To pursue the biochemical characterization of these proteins, to date the purification of 33 proteins has been performed at a larger scale. The up-scaling was performed using culture volumes of 200 ml and based on the protocols of the medium throughput strategy. Independent of size or protein family, reasonable yields in the range of several mg of

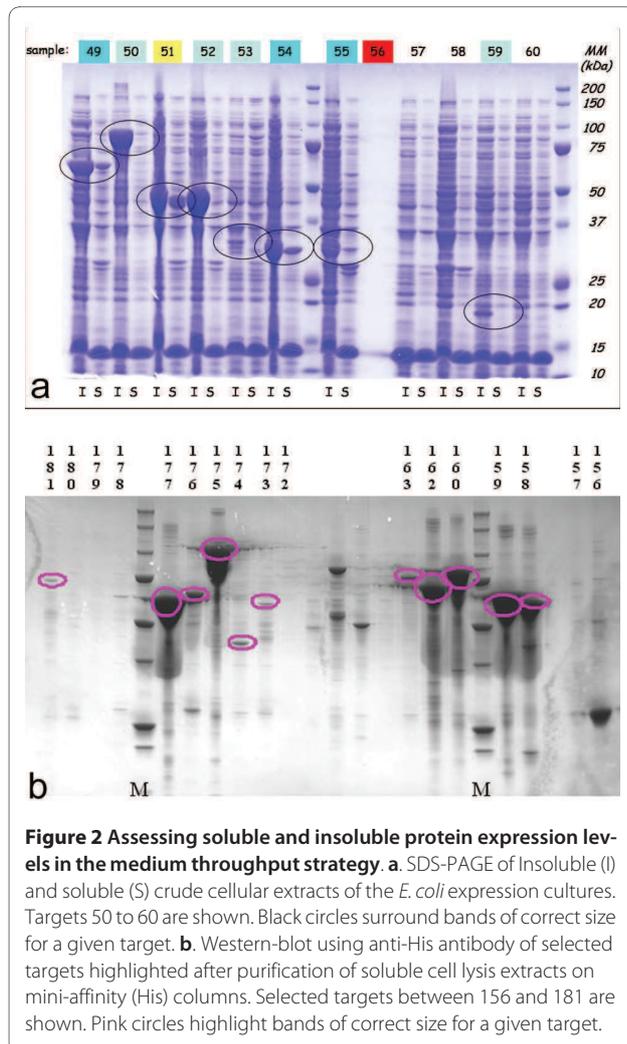


Figure 2 Assessing soluble and insoluble protein expression levels in the medium throughput strategy. a. SDS-PAGE of Insoluble (I) and soluble (S) crude cellular extracts of the *E. coli* expression cultures. Targets 50 to 60 are shown. Black circles surround bands of correct size for a given target. **b.** Western-blot using anti-His antibody of selected targets highlighted after purification of soluble cell lysis extracts on mini-affinity (His) columns. Selected targets between 156 and 181 are shown. Pink circles highlight bands of correct size for a given target.

pure protein were obtained for the scaled-up targets, as exemplified by four targets in Figure 3a. One of the aims of the follow-up study based on the genome data of *P. abyssi* is to identify novel partners in the archaeal DNA repair/replication system. The purified and tagged proteins are therefore subsequently used in pull-down/MS experiments as previously described [23]. One example of a successful pull-down experiment is illustrated in Figure 3b. The up-scaled, produced and purified target 92 (PAb0164, see Additional file 1, Table S1) was immobilized on the column. Interacting proteins were co-purified and subsequently identified using an Ultraflex MALDI-TOF/TOF instrument (Bruker Daltonics).

Up-scale, purification and crystallization of R-Z3597 from *Zobellia galactanivorans*

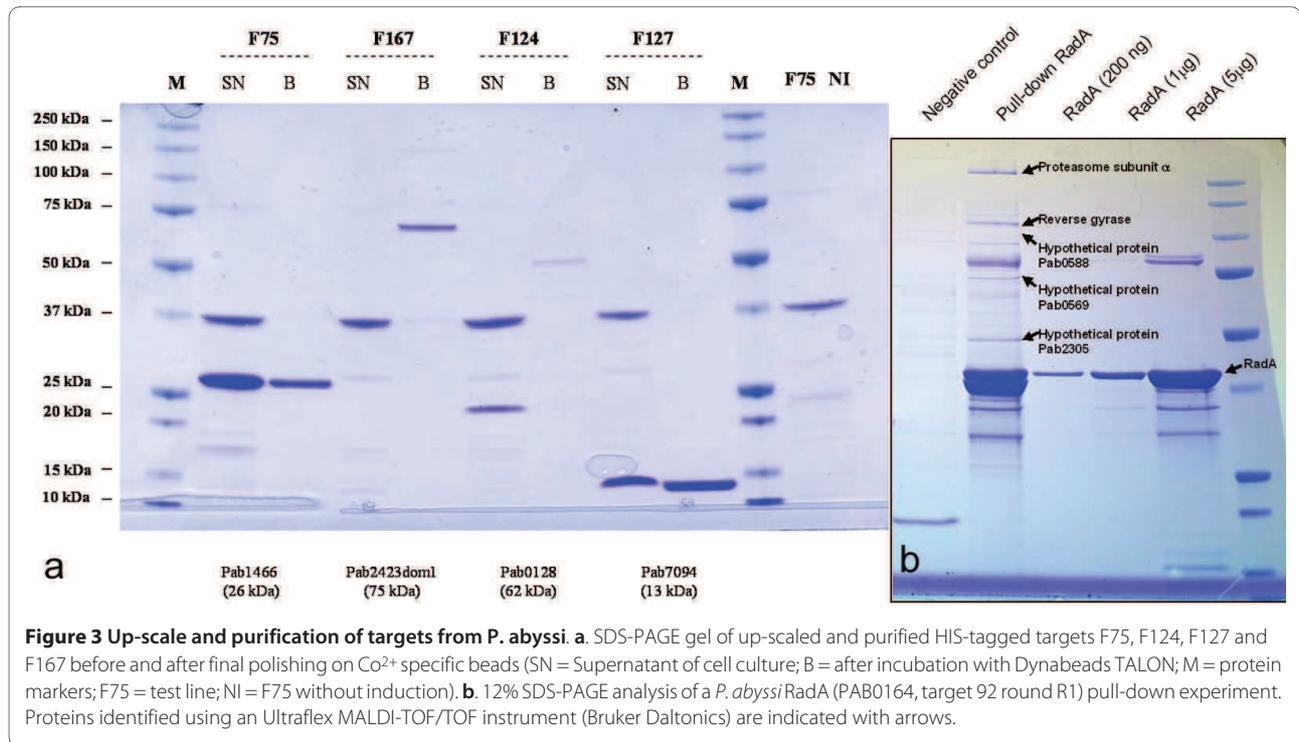
The genome of *Z. galactanivorans* contains approximately 50% of 'hypothetical proteins', as identified by bioinformatic analysis (Barbeyron et al., in preparation). A

selection of these proteins, chosen because of very distant similarity to glycoside hydrolases or because they were collinear and found in close genomic context to other glycoside hydrolases, were targets for the heterologous expression, to produce sufficient pure protein samples for a complete structural and biochemical characterization. In particular, R-Z3597 (target 43, Additional file 1, Table S1) was successfully over expressed in soluble form and in high yield. The scale-up of expression was performed using the protocol established in the medium throughput trial, but for 500 ml culture medium. The final yield after two steps of purification (Nickel affinity and size exclusion) was 15 mg (Figure 4a). A solubility screen, following the strategy of Collins et al. [24] indicated that the best results were obtained with a buffer composed of 50 mM CHES pH 9.5, 500 mM NaCl and 100 mM imidazole. The protein was concentrated in this buffer to a final concentration of 12 mg/ml. Crystallisation conditions were screened using three different commercial kits (PEGI, PACT and JCSG+). Orthorhombic single crystals grew in the optimized condition, containing 20% (w/v) PEG 3350 and 200 mM calcium acetate, within one or two weeks at 292 K (Figure 4b). The further biochemical characterization, especially the screening for activity on various marine oligo- and polysaccharides is currently in progress.

Scale-up, purification and crystallization of Stanniocalcin1A from *Sparus aurata*

Functional characterization of proteins from *Sparus aurata* involved in handling and regulation of calcium: several categories of fish cDNA were provided to the flagship project MARINE-EXPRESS and included hormones, hormone responsive transcription factors, chaperones and matrix proteins involved in hormone driven calcium turnover. Target constructs used for further studies included sea bream and *Tetraodon nigroviridis* Stanniocalcins. In particular the expression of stanniocalcin1 (STCA1) from *Tetraodon nigroviridis* was scaled-up following the protocol established in the medium throughput trial, but for 500 ml ZYP5052 culture medium. The final yield after two steps of purification (Nickel affinity and size exclusion) was about 12 mg (Figure 5a).

Crystallisation conditions were screened using three variants of commercial kits (PEGI, PACT and JCSG+). First thin, plate-like crystals grew in the condition, containing PEG 4000 25% 0.1 M di-sodium citrate pH 5.6 and 0.2 M ammonium sulphate, within one or two weeks at 292 K (Figure 5b). The further biochemical characterization and optimization of crystallization conditions to produce crystals suitable for X-ray analysis is currently under progress.



Discussion

Based on sequence analysis, a large proportion of genes from genomic data of marine organisms have unknown cellular and/or molecular functions. One major challenge is to assign biological function and to elucidate the mech-

anism of action of such genes. This challenge involves techniques to elucidate the structure and function of the gene products, interactions between proteins and/or global protein changes. For example, the three-dimensional structure of a protein can often provide functional clues, primarily by detecting structural similarity with a protein of known function even when sequence identity is low [9]. Purified protein is generally required in these

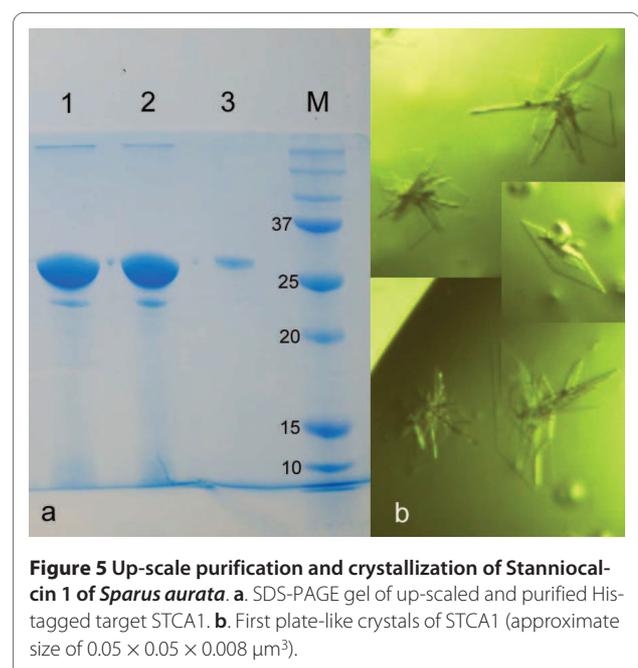
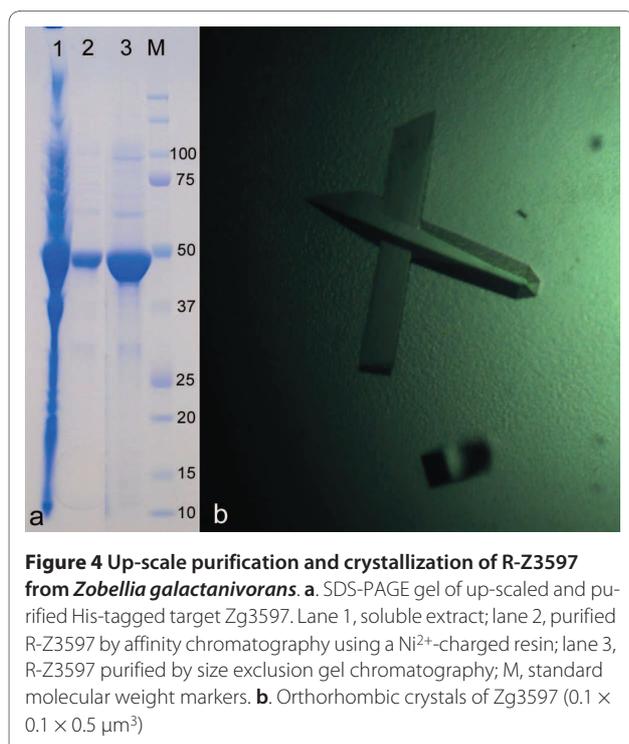
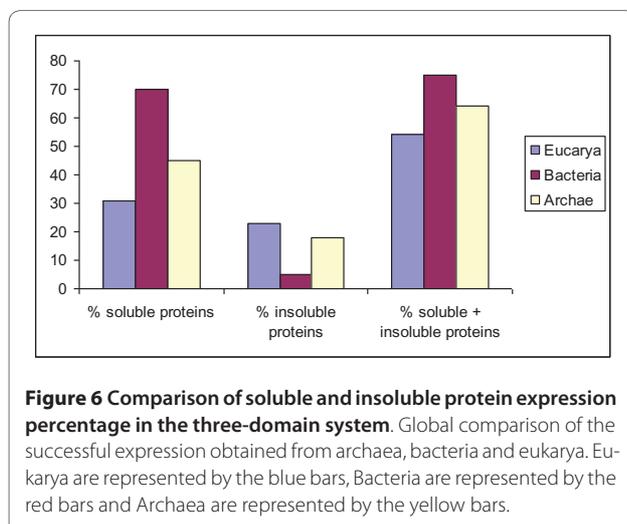


Table 3: Escherichia coli strains and plasmids used in this study.

Genotypes and relevant properties		Sources or references
Strains		
DH5α	F ⁻ <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17</i> (rK ⁻ mK ⁺), λ ⁻	Promega, USA
BL21	F ⁻ <i>ompT gal dcm lon hsdS_B(r_B⁻ m_B⁻)</i>	Novagen, USA
BL21 (DE3)	F ⁻ <i>ompT gal dcm lon hsdS_B(r_B⁻ m_B⁻) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])</i>	Novagen, USA
Rosetta	F ⁻ <i>ompT hsdS_B(R_B⁻ m_B⁻) gal dcm</i>	Novagen, USA
Rosetta (DE3)	F ⁻ <i>ompT hsdS_B(R_B⁻ m_B⁻) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])</i>	Novagen, USA
Origami	F ⁻ <i>ompT hsdS_B r_B⁻ m_B⁻) gal dcm lacY1 ahpC gor522::Tn10 (Tc^R) trxB::kan pAR5615 (Ap^R)</i>	Novagen, USA
Origami (DE3)	F ⁻ <i>ompT hsdS_B r_B⁻ m_B⁻) gal dcm lacY1 ahpC gor522::Tn10 (Tc^R) trxB::kan pAR5615 (Ap^R) λ(DE3[lacI lacUV5-T7 gene 1 ind1 sam7 nin5])</i>	Novagen, USA
Plasmids		
pGEX-4T1	Amp ^R , tac promoter, GST.Tag	GE HealthCare, USA
pFO4	Amp ^R , T7lac promoter, His.Tag	This Study

studies and is at the basis of the development of medium/high throughput strategies to produce a large number of soluble proteins [10,11]. A key feature to the success of medium/high throughput cloning strategies is the optimization of an identical treatment of all targets. Moreover, multiplying constructs, vectors and targets consequently increase the chances to obtain pure, soluble protein samples to pursue biochemical analyses. This has also been demonstrated more recently by other expression systems, such as the ligation-independent cloning (LIC) method of *Mycobacterium tuberculosis* gene sequences [25].



The increased number of genomic projects concerning marine organisms that are available, including prokaryotic organisms [1,2] or eukaryotic organisms [3-7], as well as projects still in progress (*Ectocarpus siliculosus*, *Zobellia galactanivorans*), require the application of medium/high throughput transcriptomic and proteomic methods.

Here, we show that a general scheme for bacterial expression of genes originating from marine organisms could be successfully implemented for the production of soluble proteins. Relatively few studies have been performed to assess medium/high throughput expression of soluble proteins from marine organisms. However, in general efforts have been concentrated on individual organisms. For an example, the Southeast Collaboratory for Structural Genomics has developed high throughput protein production and crystallization of genes originating from *Pyrococcus furiosus* [26]. More recently, a series of diatom expression vectors based on the Invitrogen Gateway technology for high throughput protein tagging and overexpression in *Phaeodactylum tricornutum* has been described [27].

In the present study a medium throughput approach was applied for the heterologous protein expression of diverse marine organisms (including prokaryotes, archae and eukaryotes) in the same system (Figure 6). This is in contrast with previous studies that generally focus on one organism of origin. The general concept of the MARINE-EXPRESS program can be divided into three main parts: i) the bioinformatics analysis of the target genes selected by different partners, ii) the cloning of targeted genes into

two different expression vectors and iii) the soluble expression of the targeted proteins. Table 1 shows the MARINE-EXPRESS consortium partners and the main organisms studied. Moreover, we have analyzed target genes from a lemon-shaped Virus-like Particle (PAV1) isolated from the hyperthermophilic euryarchaeote *Pyrococcus abyssi* [28] and some target genes from the brown alga *Laminaria digitata* and the red alga *Chondrus crispus* [6,7]. To enhance protein solubility and to facilitate affinity chromatography in the next purification steps, we used a gene fusion approach. Several tags have been developed commercially to facilitate rapid single-step purification [29,30]. The most popular are glutathione-S-transferase (GST) [31] and polyhistidine (His₆) [32] tags. The His₆-tag is popular because of its small size. GST fusions help in obtaining soluble products of the whole fusion protein simply by its presence. Each expression scenario requires a specific vector. Re-cloning genes into each of these vectors is extremely labor-intensive. Recombinatorial cloning methods provide an opportunity to minimize the effort required for alternate expression [33]. The present study includes two expression vectors with either a His-tag (pFO4) or GST-tag (pGEX-4T-1) at the N-terminal that allow the same restriction-enzyme-based cloning in *Bam*HI/*Eco*RI. Figure 1 shows an average cloning success superior to 80%. This success rate is clearly higher than those obtained with different organisms [34] or at different centers from the Protein Structure Initiative [13]. The use of auto-induction media for small-scale protein production was successful. This is measured by the high level of target protein expression (>80%) that has been achieved.

The next stage, analysis of protein expression can be carried out without affinity purification using SDS-PAGE analysis which is a very reliable method, but not very sensitive. Indeed, for direct detection of the His-tagged product in the soluble fraction, a dot-blot procedure with an anti-His antibody is often applied [35-37]. Dot-blot is a fast method to screen expression and solubility of recombinant proteins using a convenient 96-well format. However, the reliability of this method is limited due to lack of specificity of the detection method, and it does not give information about the size and the purity of detected protein. One solution is to couple dot-blot with techniques providing information about the actual size such as capillary electrophoresis, SDS-PAGE or Western blotting. In our study, the use of affinity mini-columns increased the percentage of detected soluble targets by 15% for His-tagged and 60% for GST-tagged targets. Moreover, this method permitted to the expression level of recombinant proteins to be estimated more precisely and confirmed their correct molecular weight. In fact, the number of obtained soluble proteins is generally under evaluated. Indeed, we have seen that for some targets that were

judged negative in small-scale experiments, culturing them in a larger volume of auto-inducible medium, such as 50 ml, in some cases allowed soluble expression of these 'negatively' judged targets (data not shown). The triage based on the small-scale results reduces the number of targets that progress to large-scale culture preparation [32], but in rare particular cases misses potential soluble expression.

Previous studies have indicated that approximately 50% of full-length proteins from the *Eubacteria* or *Archaea* and only 10-20% of proteins from *Eucarya* can be expressed in *E. coli* in soluble form [38,39]. This percentage has been significantly increased (nearly 50%) for human targets proteins using a multi-construct approach [40]. In the present study, 44% of soluble proteins were obtained. The best results are obtained for marine bacteria with 67% of soluble proteins, then archaea with 45% and as expected, *Eucarya* give the smallest percentage with 31%. These differences decrease if we take both insoluble and soluble proteins into account (Figure 6). In summary, we used a parallel production approach for bacterial expression in medium throughput to yield 84 soluble proteins from a total of 192 marine targets (44%).

While expression or crystallization strategies can be generalized to a common factor like their marine bacterial or eukaryotic origin, the setup for the functional screening is intimately bound to the gene family of interest to the different consortium members (Table 1). This latter step, performing the functional/biochemical characterization of the soluble expressed proteins, will therefore be conducted by each partner and will focus on the family of genes of their interest.

In conclusion, the present project provided purified proteins that are key reagents for numerous assays that address fundamental questions about their structure, function and regulation. For the first time, our medium throughput project allowed the expression of various proteins of marine origin in parallel, independent of organism. A rapid and cost-effective small-scale screening method for soluble expression of proteins from marine organisms in *E. coli* has been established, allowing the different partners to access large quantities of purified protein and to choose among targets of their interest for subsequent functional and/or structural analysis, which is currently underway.

Methods

Strains, plasmids and culture conditions

The sources and relevant genotypes of the bacterial strains as well as of the plasmids used are listed in Table 3. Plasmid pGEX-4T-1 (GE Healthcare Life Science, USA) is used to express protein in fusion with a glutathione-S-transferase (GST) at the N-terminal. Plasmid pFO4 (a vector modified from pET15b (Novagen, USA) to be

compatible with the *Bam*HI/*Eco*RI ligation strategy) generates a hexa-histidine tail at the N-terminal of recombinant protein. *Escherichia coli* DH5 α , used for standard cloning procedures, and *E. coli* strains used for gene expression experiments, were grown in Luria-Bertani (LB) liquid or on LB solid medium, as described by Sambrook et al. in 1989 [41]. For expression tests, an auto-inducible ZYP5052 medium [19] was used. All media were supplemented, when necessary, with 100 μ g ml⁻¹ ampicillin (sodium salt).

Bioinformatics analysis of the target sequences

The potential signal peptides and transmembrane domains have been predicted using SignalP and TMHMM, respectively [15,16]. The modularity of each target protein has been examined using Blast queries against UniProt database, as well as domain searches with the InterPro server [17]. The precise delineation of each module has been refined using Hydrophobic Cluster Analysis (HCA) [18]. For this study, 192 modules were chosen with predicted masses between 7 and 140 kDa and rearranged into two 96-well plates. All procedures were performed where possible in this 96-well format.

Primers design and cloning method

Expression vectors (pGEX-4T1 and pFO4) were digested by *Bam*HI and *Eco*RI. For each target sequence, we sought the restriction site recognized by *Bam*HI, *Eco*RI or their isocaudomers (respectively *Bgl*II and *Mfe*I) using BioEdit Sequence Alignment Editor (Ibis Biosciences Inc., USA). The target genes were classified into four compatible cloning strategies (*Bam*HI/*Eco*RI, *Bam*HI/*Mfe*I, *Bgl*II/*Eco*RI and *Bgl*II/*Mfe*I) in order to design the correct oligonucleotide primers, and assign targets in 96-wells plate). The standard scheme for primer design was defined as: for the forward primers, 5'-[hexa-G tail]-[*Bam*HI or *Bgl*II]-[Hybridization site]-3' and for the reverse primers, 5'-[hexa-C tail]-[*Eco*RI or *Mfe*I]-[stop anticodon]-[Hybridization site]-3'. Oligonucleotides for PCR were purchased in 96-well plates from Operon Biotechnologies GmbH (Cologne, Germany). PCR amplification was performed on a GeneAmp^R PCR System 2700 (Applied Biosystems, USA). The thermocycle utilized was: denaturation at 95°C for 5 min and thirty cycles of denaturing at 95°C for 30 s, annealing at 50°C for 30 s and polymerization at 72°C for 4 min. Template amplification was performed with Pfu polymerase (PROMEGA, USA) and used with the conditions recommended by the supplier. PCR reactions were analyzed on 1% agarose gels using standard procedures [41]. The resulting PCR products were purified using the QIAquickTM 96 PCR purification Kit (QIAGEN, USA), digested with appropriate restrictions enzymes and cloned in parallel into the pFO4 and pGEX-4T1 expression vectors using standard proce-

dures [41]. PCR-screening was performed directly on the DH5 α bacterial colonies to verify clones with inserts on expected size, using PCR primers which annealed upstream and downstream of the insertion site of pGEX-4T1 and pFO4. Target fragments were amplified using 10 μ l of PCR Master Mix (PROMEGA, USA) added to 0.2 μ l of each primer (100 μ M) with the same program described above. Plasmid extraction was performed using MiniPrep SV purification Kit (PROMEGA, USA) and recombinant plasmids were used to transform *E. coli* expression strains.

Screening for protein expression using 2 ml cultures

E. coli clones, for which the presence of the expression gene had been verified by colony PCR as described previously, were tested for the expression of the desired protein. Screening was done using 2 ml cultures in 24-deep well plates. Cultivation was performed in two phases. First, transformed colonies were grown at 37°C overnight in LB medium containing 100 μ g ml⁻¹ ampicillin. Then, cultures were diluted 1:100 with auto-inducible ZYP5052 medium [19] containing 100 μ g ml⁻¹ ampicillin and subjected to further incubation at 20°C until the desired density.

Lysis of cells and detection of proteins

For solubility assay, cell pellets from small-scale expression cultures were resuspended in 500 μ l of lysis buffer (Tris-HCl 50 mM, pH 7.5; NaCl 250 mM; EDTA 1 mM; lysosyme 1 mg ml⁻¹; DNase 0.1 mg ml⁻¹) and incubated at 18°C for 1 hour and the soluble and insoluble fractions separated by centrifugation (12000 g, 20 min, 4°C). Insoluble pellets were resuspended in 200 μ l of lysis buffer supplemented with urea 6 M. Samples from soluble and insoluble fractions were separated by 12% sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using 12% Criterion precast Bis-Tris gels with 26 wells. Targets were scored as positive for expression and solubility if a detectable fusion protein of the correct molecular weight was observed after Coomassie-staining. In parallel, soluble fractions were purified using His or GST Microspin columns (GE Healthcare Life Science, USA) according to the protocol recommended by the supplier. The results were also analyzed by 12% SDS-PAGE.

Additional material

Additional file 1 Table S1. Summary of all 192 targets, listing the partners providing the targets, the organism of gene origin, the accession number (where known), the primers used for amplification and eventually the follow-up experiments that have been performed.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

GM, CB and MC have set up and designed the study; AG and CH have set up and performed all experiments; GM, AG, CH and AJ have performed bioinformatic analysis; AJ has performed cloning and expression experiments in round 2; ER, PP and IM have performed up-scale expression, purification and crystallization experiments of selected targets; DF, CG, DP, MB, HM, CB and GM have provided genomic material, as well as access to genomic data and designed follow up experiments. AG, CH and MC have analyzed the data and written the manuscript; all authors discussed the results and commented the manuscript and all authors read and approved the final manuscript.

Acknowledgements

This work has been funded by the network of excellence (GOCE-CT-2004-505403) 'Marine Genomics Europe' through a Flagship program entitled 'Marine-express' (2006-2008). DF is supported by grant 4852-REPAR/CREATE from the Brittany Regional Council. This work was also supported by the ANR program ARCREP (contract number: ANR-07-BLAN-0371-01) to DF and MC.

Author Details

¹UPMC Univ Paris 6, UMR 7139 Végétaux marins et Biomolécules, LIA DIAMS, Station Biologique, F 29682, Roscoff, France, ²CNRS, UMR 7139 Végétaux marins et Biomolécules, LIA DIAMS, Station Biologique, F 29682, Roscoff, France, ³Fremer-CNRS-UBO, UMR6197, Microbiology of Extreme Environments, Technopole Brest-Iroise, BP70, 29280 Plouzané, France, ⁴Institut des nanotechnologies de Lyon (INL), UMR5270, Ecole Centrale de Lyon Bât. F 7, 36 Av Guy de Collongue, 69134 Ecully Cedex, France, ⁵Centro de Ciências do Mar, Universidade do Algarve, Comparative and molecular Endocrinology, 8005-139 Faro, Portugal, ⁶Cellular and Developmental Biology, Stazione Zoologica A. Dohrn, Villa Comunale, 80121 Napoli, Italia and ⁷Observatoire de Banyuls sur mer, Models in Cellular and Evolutionary Biology, UMR7628, CNRS-UPMC, Banyuls sur mer, France

Received: 9 March 2010 Accepted: 14 June 2010

Published: 14 June 2010

References

- Cohen G, Barbe V, Flament D, Galperin M, Heilig R, Lecompte O, Poch O, Prieur D, Quéréllou J, Ripp R, Thierry JC, Van der Oost J, Weissenbach J, Zivanovic Y, Forterre P: **An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi***. *Mol Microbiol* 2003, **47**:1495-1512.
- Glöckner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W, Gade D, Beck A, Borzym K, Heitmann K, Rabus R, Schlesner H, Amann R, Reinhardt R: **Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1**. *Proc Natl Acad Sci USA* 2003, **100**:8298-8303.
- Peters AF, Marie D, Scornet D, Kloareg B, Cock JM: **Proposal of *Ectocarpus siliculosus* (Ectocarpales, Phaeophyceae) as a model organism for brown algal genetics and genomics**. *J Phycol* 2004, **40**:1079-1088.
- Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, Robbins S, Partensky F, Degroevie S, Echeynie S, Cooke R, Saeys Y, Wuyts J, Jabbari K, Bowler C, Panaud O, Piegue B, Ball SG, Ral JP, Bouget FY, Piganeau G, De Baets B, Picard A, Delseny M, Demaille J, Van de Peer Y, Moreau H: **Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features**. *Proc Natl Acad Sci USA* 2006, **103**:11647-11652.
- Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KE, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang HG, Awazu S, Azumi K, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee BI, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoaka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Dettler C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS: **The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins**. *Science* 2002, **298**:2157-2167.
- Collén J, Roeder V, Rousvoal S, Collin O, Kloareg B, Boyen C: **An expresseq sequence tag analysis of thallus and regenerating protoplasts of *Chondrus crispus* (Gigartinales, Rhodophyceae)**. *J Phycol* 2006, **42**:104-112.
- Roeder V, Collén J, Rousvoal S, Corre E, Leblanc L, Boyen C: **Identification of stress gene transcripts in *Laminaria digitata* (Phaeophyceae) protoplast cultures by expressed sequence tag analysis**. *J Phycol* 2005, **41**:1227-1235.
- Thornton JM, Todd AE, Milburn D, Borkakoti N, Orengo CA: **From structure to function: Approaches and limitations**. *Nat Struct Biol Structural Genomics Supplement* 2000:991-994.
- Watson JD, Laskowski RA, Thornton JM: **Predicting protein function from sequence and structural data**. *Curr Opin Struct Biol* 2005, **15**:275-284.
- Sugar FJ, Jenney FE Jr, Poole FL, Brereton PS, Izumi M, Shah C, Adams MW: **Comparison of small- and large-scale expression of selected *Pyrococcus furiosus* genes as an aid to high-throughput protein production**. *J Struct Funct Genomics* 2005, **6**:149-158.
- Busso D, Poussin-Courmontagne P, Rose D, Ripp R, Litt A, Thierry JC, Moras D: **Structural genomics of eukaryotic targets at a laboratory scale**. *J Struct Funct Genomics* 2005, **6**:81-88.
- Baneyx F: **Recombinant protein expression in *Escherichia coli***. *Curr Opin Biotechnol* 1999, **10**:411-421.
- Fox BG, Goulding C, Malkowski MG, Stewart L, Deacon A: **Structural genomics: from genes to structures with valuable materials and many questions in between**. *Nat Methods* 2008, **5**(2):129-132.
- Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park HW, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim SH, Rao Z, Shi Y, Terwilliger TC, Kim CY, Hung LW, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyti T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schütz A, Heinemann U, Yokoyama S, Büssov K, Günsalus KC: **Protein production and purification**. *Nat Methods* 2008, **5**:135-146. Review
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0**. *J Mol Biol* 2004, **340**:783-795.
- Krogh A, Larsson B, von Heijne G, Sonnhammer ELL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes**. *J Mol Biol* 2001, **305**:567-580.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R: **InterProScan: protein domains identifier**. *Nucleic Acids Res* 2005, **33**:116-120.
- Lemesle-Varloot L, Henrissat B, Gaboriaud C, Bissery V, Morgat A, Mornon JP: **Hydrophobic cluster analysis: procedures to derive structural and functional information from 2-D-representation of protein sequences**. *Biochimie* 1990, **72**:555-574.
- Studier FW: **Protein production by auto-induction in high density shaking cultures**. *Protein Expr Purif* 2005, **41**:207-234.
- Berrow NS, Büssov K, Coutard B, Diprose J, Ekberg M, Folkers GE, Levy N, Lieu V, Owens RJ, Peleg Y, Pinaglia C, Quevillon-Cheruel S, Salim L, Scheich C, Vincentelli R, Busso D: **Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study**. *Acta Crystallogr D Biol Crystallogr* 2006, **62**:1218-1226.
- Schein CD: **Production of soluble recombinant proteins in bacteria**. *Biotechnol* 1989, **7**:1141-1148.
- Nguyen H, Martinez B, Oganessian N, Kim R: **An automated small-scale protein expression and purification screening provides beneficial information for protein production**. *J Struct Funct Genomics* 2004, **5**:23-27.
- Ren B, Kuhn J, Meslet-Cladiere L, Briffotoux J, Norais C, Lavigne R, Flament D, Ladenstein R, Myllykallio H: **Structure and function of a novel endonuclease acting on branched DNA substrates**. *Embo J* 2009, **28**(16):2479-2489.

24. Collins BK, Tomanicek SJ, Lyamicheva N, Kaiser MW, Mueser TC: A preliminary solubility screen used to improve crystallization trials: crystallization and preliminary X-ray structure determination of *Aeropyrum pernix* flap endonuclease-1. *Acta Cryst D-Biol Cryst* 2004, **60**:1674-1678.
25. Qin H, Hu J, Hua Y, Challa SV, Cross TA, Gao FP: Construction of a series of vectors for high throughput cloning and expression screening of membrane proteins from *Mycobacterium tuberculosis*. *BMC Biotechnol* 2008, **16**:8-51.
26. Wang BC, Adams MW, Dailey H, DeLucas L, Luo M, Rose J, Bunzel R, Dailey T, Habel J, Horanyi P, Jenney FE Jr, Kataeva I, Lee HS, Li S, Li T, Lin D, Liu ZJ, Luan CH, Mayer M, Nagy L, Newton MG, Ng J, Poole FL, Shah A, Shah C, Sugar FJ, Xu H: Protein production and crystallization at SECSG -- an overview. *J Struct Funct Genomics* 2005, **6**(23):233-243.
27. Siaut M, Heijde M, Mangogna M, Montsant A, Coesel S, Allen A, Manfredonia A, Falciatore A, Bowler C: Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* 2007, **406**(1-2):23-35.
28. Geslin C, Gaillard M, Flament D, Rouault K, Le Romancer M, Prieur D, Erauso G: Analysis of the first genome of a hyperthermophilic marine virus-like particle, PAV1, isolated from *Pyrococcus abyssi*. *J Bacteriol* 2007, **189**:4510-4519.
29. Stevens RC: Design of high-throughput methods of protein production for structural biology. *Structure* 2000, **8**:R177-185.
30. Terpe K: Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 2003, **60**:523-533.
31. Smith DB: Generating fusions to glutathione S-transferase for protein studies. *Methods Enzymol* 2000, **326**:54-70.
32. Hochuli E: Purification of recombinant proteins with metal chelate adsorbent. *Genet Eng (N Y)* 1990, **12**:87-98.
33. Lesley SA: High-throughput proteomics: protein expression and purification in the postgenomic world. *Prot Expr Purif* 2001, **22**:159-164.
34. Cowieson NP, Listwan P, Kurz M, Aagaard A, Ravasi T, Wells C, Huber T, Hume DA, Kobe B, Martin JL: Pilot studies on the parallel production of soluble mouse proteins in a bacterial expression system. *J Struct Funct Genomics* 2005, **6**:13-20.
35. Knaust RK, Nordlund P: Screening for soluble expression of recombinant proteins in a 96-well format. *Anal Biochem* 2001, **297**:79-85.
36. Busso D, Kim R, Kim SH: Expression of soluble recombinant proteins in a cell-free system using a 96-well format. *J Biochem Biophys Methods* 2003, **55**:233-240.
37. Vincentelli R, Canaan S, Offant J, Cambillau C, Bignon C: Automated expression and solubility screening of His-tagged proteins in 96-well format. *Anal Biochem* 2005, **346**:77-84.
38. Braun P, LaBaer J: High throughput protein production for functional proteomics. *Trends Biotechnol* 2003, **21**:383-388.
39. Büssov K, Scheich C, Sievert V, Harttig U, Schultz J, Simon B, Bork P, Lehrach H, Heinemann U: Structural genomics of human proteins--target selection and generation of a public catalogue of expression clones. *Microb Cell Fact* 2005, **5**:21.
40. Gräslund S, Sagemark J, Berglund H, Dahlgren LG, Flores A, Hammarström M, Johansson I, Kotenyova T, Nilsson M, Nordlund P, Weigelt J: The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expr Purif* 2008, **58**:210-221.
41. Sambrook J, Fritsch EF, Maniatis T: *Molecular cloning: A Laboratory Manual*. Cold Spring Harbor, Cold Spring Harbor Press; 1989.

doi: 10.1186/1475-2859-9-45

Cite this article as: Groisillier *et al.*, MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms *Microbial Cell Factories* 2010, **9**:45

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Annexe 2: **Supplementary Material and Methods**

Cloning, production and purification of the RPA complex.

The structural genes of *PabRPA41*, *PabRPA14* and *PabRPA32* were amplified using the genome of *Pyrococcus abyssi* as a template, the high fidelity Pfu DNA polymerase (Promega) and the following primers:

RPA41_For: 5'-CAAAGGTGGTTCATATGACGGTGTGACG-3'

RPA41_Rev: 5'-CTCTAACTTGAGGATCCATCTCACATCATC-5'

RPA14_For: 5'-GGGTGATGTCATATGGAAGTTCAAGTTAGG-3'

RPA14_Ter: 5'-CCTAGTCGCGGGGATCCTCTTCTTCATTTTC-3'

RPA32_For: 5'-GTGTGGTGATTCATATGAAGAAGAGGATGC-3'

RPA32_Rev: 5'-CCTTCCATGGGATCCACCTTACCCTCTATC-3'

The PCR products were digested by *NdeI* and *BamHI* and cloned into the pET 28a+ vector (Novagen) for RPA41 and into the pET 25b+ vector (Novagen) for RPA14 and RPA32. Thereafter, to improve the solubility of RPA32, the gene cloning for this subunit was inserted in frame with the gene of RPA14 and under the T7 promoter control. To this aim, the recombinant pET 25b+/RPA14 and pET 25b+/RPA32 plasmids were double digested by *BamHI/NotI* and *BgIII/NotI*, respectively. The *BgIII-NotI* fragment carrying the RPA32 and the T7 promoter as well as the ribosome binding site sequence was then ligated with the *BamHI-NotI* fragment, hence downstream and in frame with the RPA14 gene. The resulting plasmids, pET 28a+/RPA41 and pET 25b+/RPA14-RPA32, were used to transform *Escherichia coli* Rosetta (DE3) pLysS strain (Novagen) for the production of the three subunits, with RPA41 fused to a histidine tag at the N-terminus. The nucleotide sequence of each plasmid was checked by DNA sequencing. Although numerous clones have been tested, the insert corresponding to the gene of RPA41 was always found with at least one mutation in the 5' part. One of these mutated clones exhibited a transition leading to a leucine 4 to valine mutation of the amino acid deduced sequence. The attempts to correct this mutation failed and induced additional mutations; consequently, as this mutation was considered to be conservative enough, the clone was used for further analysis. Cells were grown in 8 litres of LB medium supplemented with chloramphenicol (34 µg/ml), ampicillin (50 µg/ml) and kanamycin (50 µg/ml) at 37°C to $A_{600nm} = 0.6$ and expression of RPA proteins were induced by the addition of 0.5 mM IPTG (final concentration). Four hours after induction, cells were harvested by centrifugation and stored at -80°C. The pellets were resuspended in 100 ml of lysis buffer (30 mM Tris-HCl, pH 7.5, NaCl 300 mM, ZnCl₂ 10 µM and imidazol 10 mM). Cells were heated at 80°C for 15 minutes and then frozen at -80°C for 30 minutes; this freeze/heat cycle was applied three times. Cell debris were removed by centrifugation (12 000 g for 30 minutes at 15°C). Most of the remaining *Escherichia coli* proteins were removed using another heat step: 30 minutes at 80°C, followed by a centrifugation step (12 000 g for 30 minutes at 15°C). Supernatant was filtered

(0,40 μm) and loaded onto a Ni^{2+} affinity column pre-equilibrated in buffer A (30 mM Tris-HCl, pH 7.5, NaCl 300 mM and imidazol 10 mM). Non-bounded fraction, containing the small subunit, was collected, filtered using a 30kDa-cutoff concentrator (Vivaspin, Sartorius), heated (80°C for 15 minutes) and centrifugated (12 000 g for 30 minutes at 15°C). A gradient of buffer A to buffer B (30 mM Tris-HCl, pH 7.5, NaCl 300 mM and imidazol 500 mM) was applied to the column, and the His tagged RPA complex was eluted at 100 mM imidazol. Fractions containing the heterotrimer RPA were pooled and concentrated to 10 mg/ml, and applied onto a size-exclusion column (Superdex 200 – GE Healthcare) equilibrated in buffer C (30 mM Tris-HCl, pH 7.5, NaCl 300 mM). Excess of medium subunit appeared to be adsorbed on the column, and was released using NaOH 200 mM. Fractions were titrated with 1 M Tris-HCl, pH 7.5 and protein was stored in buffer C. To eliminate DNA in each protein or complex, NaCl concentration was adjusted to 1 M and Poly-imine P was added to a final concentration of 0,5%. DNA was then removed by centrifugation (15 000 g for 1 hour at 15°C). Protein was precipitated with AmSO_4 to remove Poly-imine P, and solubilized in buffer D (30 mM Tris-HCl, pH 7.5, NaCl 500 mM). AmSO_4 was eliminated by dialysis against buffer D. The large subunit (RPA41) and the full complex were then separated on an size-exclusion column (Superdex 200 – GE Healthcare) equilibrated in the same buffer. Fractions containing subunits alone or the complex were separately pooled and concentrated to 10 mg/ml, and stored at 12°C in buffer D.

Cloning, production and purification of MCM.

The *mcm* gene, was shown to contain two intein coding regions, which are intervening sequences spliced out as a protein and not as a mRNA. To prevent a possible toxic effect of the inteins in *E. coli*, the genes encoding intein were deleted using the splicing by overlap extension PCR approach [Ho, 1990 #162], with the following primers:

MCM-*ini*: 5'-GGATGGGTGAAGCATATGGATAGAGAGGAG-3'

MCM-*int1R*: 5'-ATGTATCTGAGAAGTTGGCTGTTTTTCGCGACTCCCGGATCTC-3'

MCM-*int1F*: 5'-GAGATCCGGGAGTCGCGAAAAACAGCCAACTTCTCAGATACAT-3'

MCM-*int2F*: 5'-TCGTCCCTAACCACCGCGGCGAGCCCAGCGGCAGAACTGC-3'

MCM-*int2R*: 5'- GCAGTTCTGCCGCTGGGCTCGCCGCGGTGGTTAGGGACGA-3'

MCM-*end*: 5'-AAAGAGTAGTCGACTCAGACGGTTCTGTAATA-3'

The resulting PCR product contained the MCM gene minus the intein. This product was double digested by *NdeI/SalI* (restriction sites in bold in the sequences of primers), cloned into pQE-80 and sequenced. The constructed expression vector pQE-80L/*PabMCM* was introduced into host *E.coli* Rosetta (DE3) pLysS. The transformed cells were grown at 37°C in 6 L LB medium supplemented with 50 $\mu\text{g/ml}$ ampicillin and 34 $\mu\text{g/ml}$ chloramphenicol. When culture reached an A_{600} of 0.6 OD, protein expression was induced by addition of 1 mM IPTG. Induced cells were incubated at 37°C during 4

hours and harvested by centrifugation. The pellet was resuspended in lysis buffer (10 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM DTT, and Complete Mini-ethylenediaminetetraacetic acid [EDTA]-free anti protease (Roche)). Cells lysis was done using "One Shot" Cell Disrupter (Constant Systems LDT, Daventry, UK) at a pressure of 1.9 kbars. The resulting lysate was centrifuged at 8000 x g for 15 min at 4°C to eliminate cell debris. The supernatant was then subjected to heat treatment at 75°C for 10 min and the thermo-precipitated proteins were removed by centrifugation at 8000 x g for 15 min at 4°C. The clarified supernatant was applied to a His Prep FF 16/10 column (GE Healthcare Life Sciences) preequilibrated with buffer A (10 mM Tris HCl pH 7.5, 150 mM NaCl, 10 mM imidazole, 1 mM DTT). Elution was performed with 140 ml of 0.01 to 0.5 M imidazole linear gradient in buffer A. 1.0 ml fractions were collected and analyzed by SDS-PAGE to detect the MCM polypeptide. Fractions containing MCM were pooled and dialyzed overnight against buffer B (10 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM DTT). The dialysate was loaded onto a Superdex S200 10/300 GL column (GE Healthcare Life Sciences) pre-equilibrated with buffer B. The peak fractions were pooled, dialyzed against buffer B and concentrated using a Vivaspin Turbo 15 concentrator (Sartorius) at 2.7 mg/ml. The purified enzyme was then aliquoted and stored at -20 °C.

Cloning, production and purification of the Mre11/Rad50 complex

The genes encoding *Pab0810* (*PabMre11*) and *Pab0811* (*PabRad50*) have been amplified targeting *P. abyssi* genomic DNA with high fidelity Pfu DNA polymerase (Promega) and the following primers:

Mre11_Lic_For: 5'-GACGACGACAAGATGTGGACTTCTGAGGTGTTTCG-3'

Mre11_Lic_Rev: 5'-CGCGGGCGGCCGTCACCCTTTAACCCAGGCC -3'

Rad50_Lic_Pour: 5'-GCGGGCCCGGCCTGCATGAAGATAGAGGAGC-3'

Rad50_Lic_rev: 5'-GAGGAGAAGCCCGGTTTCATGACACCACCTCTACC -3'

Both PCR products were inserted in the same pET-30 Ek/LIC Vector (Novagen) that allowed multiple ORF cloning in the same plasmid. *PabMre11* was fused to a HIS-tag at the N-terminus. The recombinant plasmid was used to transform competent *E.coli* BL21(DE3). The transformed cells were grown at 37°C in 6 L LB medium supplemented with 50 µg/ml of kanamycin. When culture reached an A_{600} of 0.6 OD, protein expression was induced by addition of 1 mM IPTG at 15°C overnight. Cells were harvested at 4°C for 45 min at 5000 g. The pellet was resuspended in lysis buffer (20 mM Tris HCl pH 8, 20 mM Imidazole, 1 mM DTT, and Complete Mini-ethylenediaminetetraacetic acid [EDTA]-free anti protease (Roche)). Cells lysis was achieved after three runs in French pressure cell (40K cell, 1000 psi). The resulting lysate was centrifuged at 10000 x g for 15 min at 4°C to eliminate cell debris. The supernatant was then subjected to heat treatment at 75°C for 10 min and the thermo-precipitated proteins were removed by centrifugation at 14000 x g for 45 min at 4°C. The clarified supernatant was filtered (0,40 µm) and loaded onto a Ni²⁺ affinity column pre-equilibrated in buffer A (20 mM Tris-HCl, pH 8, 20 mM Imidazol, 1 mM DTT). A gradient of buffer A to buffer B (20 mM Tris-HCl, pH 8, 1 M Imidazol, 150mM NaCl, 1 mM DTT) was applied to the column, and the His-tagged Mre11/Rad50

complex was eluted at 400 mM imidazol. Fractions containing the heterodimer were pooled, dialyzed overnight against buffer C (20 mM Tris-HCl, pH 8, 150mM NaCl, 1 mM DTT) and concentrated using a Vivaspin Turbo 15 concentrator (Sartorius) at 0.3 mg/ml. The purified complex was then aliquoted and stored at $-20\text{ }^{\circ}\text{C}$.





Abstract

DNA replication, recombination and repair are central and essential mechanisms in all cells. Highly efficient high-fidelity chromosome replication is vital for maintaining the integrity of the genetic information and for the avoidance of genetic disease. Archaeal replisome is described as simplified version of the eukaryotic system. However, DNA repair is still enigmatic, as many essential repair proteins have not been identified in Archaeal genomes. The question of DNA repair is even more puzzling while many Archaea lives under extreme temperature that promotes DNA instability and catalyses nucleobase damages. Thus, Hyperthermophilic Archaea (HA) must have solved a molecular problem (spontaneous loss of native DNA structure) at a magnitude that mesophilic organisms do not face. A highly adapted DNA maintenance system must operate in order to maintain DNA integrity. Those mechanisms and their possible coordination with DNA replication are still unknown.

Here, I report the first protein-protein interaction network of genomic maintenance in HA. Using AP-MS approach we identified new protein complexes potentially implicated in DNA replication, recombination and repair of HA *P. abyssi*. Topological analysis of the network highlighted both known and unknown partners of essential and conserved protein of genomic maintenance. From the network emerges multifunctional clusters integrating both replication and recombination proteins and revealing new aspects of the transcription machinery. I also provide experimental confirmation of some of the interactions we detected.

I propose that the interactions we observe reflects the interplay between recombination and replication machineries that likely interfaces with regulatory elements involved in the control of the DNA damage response, as shown by the identification of a new factors, presumably involved in the coupling of DNA recombination and DNA synthesis at the replication fork.

Les organismes vivants doivent reproduire et transmettre *ne variatur* l'information contenue dans les chromosomes. Ainsi, la conservation de l'intégrité du génome est un processus biologique fondamental. La maintenance génomique constitue l'ensemble des processus biologiques impliqués dans la conservation, la duplication et la transmission de l'information génétique contenue dans les chromosomes. La machinerie répliquative des Archaea est décrite comme une version simplifiée de celle connue chez les Eucaryotes faisant des Archaea un excellent modèle d'étude de la réplication. Contrairement à la réplication, les processus Archaea de réparation de l'ADN sont encore mystérieux. En effet, plusieurs protéines essentielles de la réparation semblent absentes des génomes Archaea et ce même chez les espèces Hyperthermophiles (HA). A la température optimale de croissance, ces Archaea doivent posséder des capacités considérables de réparation des dommages de l'ADN, catalysés à haute température. Ainsi les Archaea hyperthermophiles sont probablement dotées d'un système de réparation alternatif extrêmement efficace. Ce système et sa coordination avec la réplication sont inconnus.

Un protocole de purification d'affinité couplée à la spectrométrie de masse des protéines permet d'identifier les complexes protéiques impliqués dans la maintenance génomique de l'Archaea hyperthermophile *P. abyssi*. Les complexes identifiés sont compilés dans un réseau d'interaction. Soumis à une étude topologique le réseau révèle notamment de nouvelles interactions entre des protéines essentielles de la maintenance génomique, conservées avec les Eucaryotes. Plusieurs interactions sont vérifiées indépendamment où caractérisées fonctionnellement *in vitro* ou *in vivo*.

Ces travaux mettent en lumière l'étroite collaboration entre la réplication et la recombinaison et révèlent de nouveaux aspects de la machinerie de transcription.