

## Ce qui est vrai pour l'Archaea *Pyrococcus* est (presque) vrai pour l'éléphant !

*La génomique comparée a confirmé de façon spectaculaire la ressemblance entre les mécanismes informationnels des Archaea et ceux des eucaryotes. Cela est particulièrement vrai pour la réplication. Ainsi, Pyrococcus abyssi réplique son chromosome selon un mode bactérien (origine unique) mais en utilisant des protéines de type eucaryote, dont la plupart n'ont pas d'homologues chez les bactéries. Ces protéines semblent jouer des rôles identiques, de*

*Pyrococcus à l'homme. Par analogie, on peut penser que de nombreuses protéines humaines sans fonction connue, qui ont des homologues chez les Archaea, sont des protéines informationnelles encore inconnues. Pour encourager leur étude, nous avons établi une première liste de PACE (proteins of archaea conserved in eukaryotes) dont certaines semblent impliquées dans la cancérogenèse.*

Cette citation de Jacques Monod « Ce qui est vrai pour *Escherichia coli* est vrai pour l'éléphant » était caractéristique du sentiment des pionniers de la biologie moléculaire qui venaient de mettre en évidence l'universalité du code génétique et des mécanismes de base de la réplication, de la transcription et de la traduction. Elle a servi, pendant des décennies, à justifier l'utilisation des bactéries, en particulier celle de *E. coli*, comme organismes modèles simples, permettant d'aborder l'étude des mécanismes moléculaires informationnels (traduction, transcription, réplication, réparation et recombinaison de l'ADN), avec comme objectif l'extrapolation des résultats obtenus aux eucaryotes. Cette approche s'est avérée fructueuse dans de nombreux cas, par exemple dans l'étude du système de correction des mésappariements chez *E. coli* dans la compréhension des mécanismes de cancérogenèse [1]. Toutefois, dès les années 1970, de nombreuses différences ont été trouvées entre la biologie moléculaire des eucaryotes et celle des bactéries, et l'extrapolation s'est souvent révélée impossible.

*A priori*, l'existence de ces différences pouvait être interprétée comme reflé-

tant la plus grande complexité des eucaryotes. Ainsi, les cellules humaines possèdent trois ARN polymérases distinctes, composées chacune par une douzaine de polypeptides différents, alors que l'unique ARN polymérase des bactéries n'en comprend que trois.

Nous savons aujourd'hui que cette façon de voir était en partie erronée. Si notre biologie moléculaire s'éloigne autant de celle des bactéries, ce n'est pas parce que nous sommes plus complexes, mais parce que nous n'appartenons pas au même monde de l'information. Pour conserver toute sa valeur à la citation de J. Monod, il suffit, toutefois de la modifier légèrement et de proclamer, « ce qui est vrai pour *Pyrococcus* est vrai pour l'éléphant ». Où est la différence ? *E. coli* est une bactérie, *Pyrococcus* une Archaea !

### L'étude des Archaea remet en question notre façon d'interpréter les différences entre bactéries et eucaryotes

La plupart des biologistes ont aujourd'hui entendu parler du domaine des Archaea [2], un groupe de procaryotes atypiques dont certains se sont fait remarquer par leur

capacité à vivre dans des conditions dites extrêmes (pH égal à 0 ou température de 110 °C, par exemple). En fait, il existe beaucoup d'Archaea qui vivent dans des conditions plus « normales » tout autour de nous (les Archaea non cultivables du sol, des lacs et de l'océan), et même en nous (les méthanogènes de notre intestin) [3]. Ce qui distingue toutefois les Archaea de tous les autres organismes, c'est leur biologie moléculaire et leur biochimie, différente à la fois de celle des bactéries et des eucaryotes, en particulier la structure de leurs lipides. Ce qui rend surtout les Archaea fascinantes, c'est que leurs mécanismes informationnels ressemblent beaucoup plus aux nôtres que ceux des bactéries [4, 5]. Ainsi, pour reprendre l'exemple cité plus haut, leurs ARN polymérases sont presque aussi complexes que les nôtres, avec une dizaine de sous-unités [6]. Les cellules des Archaea étant aussi « simples » que celles des bactéries, la sophistication de nos ARN polymérases ne peut donc plus s'expliquer par notre statut autoproclamé d'organisme supérieur ! Les mécanismes informationnels communs aux Archaea et aux eucaryotes, quelle que soit leur complexité, devaient bien être déjà présents dans leur

**Tableau I.** Protéines présomptives de réplication de l'ADN chez *Pyrococcus abyssi* et leurs homologues humains.

Fonctions possibles chez <i>Pyrococcus</i>	Gène chez <i>Pyrococcus</i>	Nom chez <i>Pyrococcus</i>	Nom chez l'homme	Homologues bactériens (analogues fonctionnels)
Reconnaissance de l'origine chargement de l'hélicase MCM	PAB2265	Protéine Cdc6	Cdc6 (également homologue des protéines Orc1 et Orc4)	Absent (DnaA)
Ouverture de l'origine, hélicase	PAB2373	MCM	MCM 2, 3, 4, 5, 6, 7	Absent (DnaB)
Formation des amorces pour les fragments d'Okazaki	PAB2235	Primase, grande sous-unité	Primase, grande sous-unité, p58 (b)	Absent
	PAB2236	Primase, petite sous-unité (catalytique)	Primase, p49 Petite sous-unité	Absent (DnaG)
?	PAB0316	Primase DnaG-like	Absent	Domaine commun avec les primases bactériennes DnaG
Réplacase et/ou maturation des fragments d'Okazaki	PAB2404	ADN polymérase II, sous-unité catalytique	Absent	Absent (Pol III)
Protéine accessoire de la réplacase	PAB2266	ADN polymérase II, petite sous-unité	Petite sous-unité de l'ADN polymérase $\gamma$	Absent
Réplacase et/ou maturation des fragments d'Okazaki	pPAB1128	ADN polymérase I	ADN polymérases $\alpha$ , $\delta$ et $\epsilon$	ADN polymérase II (Pol I)
Facteur de « processivité » de la polymérase	PAB1465	PCNA	PCNA (proliferating cell nuclear antigen)	Absent (facteur de « processivité » $\beta$ )
Chargement du facteur de « processivité »	PAB0068	RFC	RFC (replication factor C)	Sous-unités $\gamma$ et $\tau$ de l'ADN polymérase III
	PAB0069	RFC, grande sous-unité	RFC	idem
Protection de l'ADN simple-chaîne	PAB2163	RPA	RPA-1 (RPAp 70) (replication protein A)	Absent (ssb)
	PAB2164	RPA-3	RPA-3 (RPAp 14)	Absent (ssb)
	PAB2165	RPA-2	RPA-2 (RPAp 34)	Absent (ssb)
Maturation des fragments d'Okazaki, ligation	PAB2002	Ligase (dépendante de l'ATP)	Ligase (dépendante de l'ATP)	Absent (ligase dépendante du NAD)
Maturation des fragments d'Okazaki, élimination de l'ARN	PAB1877	FEN1/Rad2	Flap-endonucléase I, DNase IV	Absent (ADN polymérase I)

ancêtre commun, il y a plus de trois milliards d'années !

**L'annotation des génomes à mis en lumière le caractère *a priori* eucaryote du mécanisme de réplication de l'ADN chez les *Archaea***

De tous les mécanismes informationnels, c'est la réplication de l'ADN qui

présente le plus de ressemblances entre *Archaea* et eucaryotes. La première observation remonte à 1984: il est apparu cette année-là que la synthèse de l'ADN chez certaines *Archaea* était bloquée *in vivo* par l'aphidicoline, un inhibiteur spécifique des ADN réplacases eucaryotes [7]. C'est toutefois l'analyse, en 1996, du premier génome d'*Archaea* complètement séquencé, *Methanococcus*

*jannashii*, qui a véritablement révélé l'étroite parenté entre notre mécanisme de réplication et celui des *Archaea* [8]. Le séquençage d'une dizaine de nouveaux génomes d'*Archaea* dans les années qui ont suivi a confirmé cette observation. Le *Tableau I* présente la liste des protéines de réplication présomptives que nous avons identifiées au cours de l'annotation du génome de *Pyro-*

*coccus abyssi*, une *Archaea* hyperthermophile dont le séquençage a été réalisé par le Genoscope [9]. Toutes ces protéines, à deux exceptions près, sont de type eucaryote. La plupart d'entre elles n'ont même pas d'homologue chez les bactéries et, dans les quelques cas où une protéine de réplication est présente dans les trois domaines, la protéine des *Archaea* présente toujours de plus fortes similitudes de séquence avec son homologue eucaryote qu'avec son homologue bactérien. Les deux protéines qui dérogent à la règle sont la sous-unité catalytique de l'ADN polymérase II, qui correspond à une nouvelle famille d'ADN polymérases uniquement présente chez les *Archaea* [10], et un homologue de la primase bactérienne, qui n'existe pas chez les eucaryotes. Curieusement, les *Archaea* possèdent à la fois un homologue de la primase eucaryote et un homologue de la primase bactérienne (Tableau I).

Les protéines répliquatives identifiées chez *P. abyssi* sont toutes également présentes chez les *Archaea* au même royaume (les *Euryarchaeota* [2]) dont les génomes ont été séquencés; en revanche, trois d'entre elles n'ont pas d'homologues chez les *Archaea* de l'autre royaume, celui des *Crenarchaeota* [11]. Il est amusant de constater que les *Pyrococcus*, qui vivent à des températures comprises entre 95 et 105 °C, dans des sources marines [12], semblent bien, pour le moment, être les *Archaea* les plus proches de l'homme! En effet, nous avons pu identifier chez les *Pyrococcus* un homologue de la petite sous-unité de la protéine RPA humaine, qui n'est pas présent chez les autres *Archaea*.

### Les *Archaea* répliquent leur ADN à partir d'une origine unique, comme les bactéries

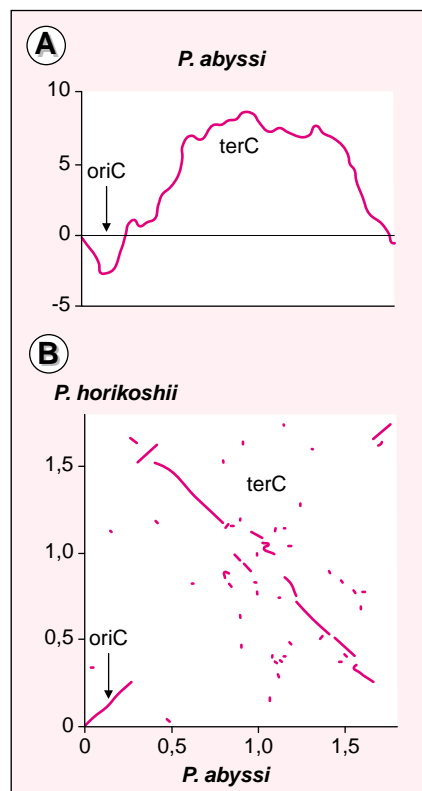
La similitude poussée entre les mécanismes de la réplication des eucaryotes et des *Archaea* semble bien témoigner d'une véritable parenté évolutive. En tout état de cause, elle ne peut pas être due à une structure similaire de leur chromatine ou à un mode de réplication identique et différent de celui des bactéries. En

effet, si *Pyrococcus* et d'autres *Euryarchaeota* possèdent bien des histones apparentées à celle des eucaryotes [13], les *Crenarchaeota*, qui ne possèdent pas ces histones, ont, elles aussi, des enzymes de réplication de type eucaryote [11]. Par ailleurs, nous avons pu montrer dans notre laboratoire, en collaboration avec l'équipe d'Hervé Philippe, que plusieurs *Archaea*, dont *P. abyssi*, répliquent leur ADN non pas comme les eucaryotes, mais au contraire selon un mode bactérien [14, 15].

Chez les bactéries, deux complexes de réplication (réplisomes) sont formés au niveau d'une origine unique et vont répliquer le chromosome de façon bidirectionnelle à très grande vitesse ( $\approx 1\ 000$  nucléotides par seconde). Au contraire, les eucaryotes répliquent leurs chromosomes à partir de multiples origines, et les

fourches ne parcourent que de courtes distances à faible vitesse (la taille moyenne d'un réplicon eucaryote varie de 10 à 100 kb, et la vitesse des fourches est de l'ordre de 100 nucléotides par seconde). Dans le cas de *P. abyssi*, nous avons pu montrer, en combinant des approches *in silico* et *in vivo*, que la réplication est bidirectionnelle et débute au niveau d'une origine unique, comme chez les bactéries (figure 1) [15]. Compte tenu de la taille du chromosome de *P. abyssi* (1,76 Mb) et de son temps de génération, qui est de 50 minutes à 95 °C, la vitesse des fourches de réplication chez les *Archaea* est donc comparable à celle observée chez *E. coli*.

La similitude entre le mode de réplication des bactéries et celui des *Archaea* semble également s'appliquer à la relation entre réplication



◀ **Figure 1. Origine de réplication de *P. abyssi*.** A. L'asymétrie dans la composition des deux brins d'ADN du génome de *P. abyssi* permet d'identifier l'origine de réplication. Chez les bactéries, pour une raison encore mal comprise, le brin répliqué de façon continue est souvent plus riche en G ou en certains mots, que le brin répliqué de façon discontinue, ce qui a permis de détecter l'origine et le terminus de réplication chez de nombreux génomes dont on connaissait la séquence complète. Dans certains cas, ces biais sont faibles et ils ne peuvent être mis en évidence qu'en étudiant (par intégration) leur effet cumulé tout au long de la séquence. Ici, le biais cumulatif de l'oligonucléotide GGGT par rapport à son complémentaire ACCC, calculé sur des fenêtres de 50 kbases le long de la séquence d'ADN complète du génome, présente un point singulier (flèche) qui permet d'identifier la région où se situe *oriC*, l'origine de réplication du chromosome de *P. abyssi* [15]. B. Les séquences nucléotidiques de deux souches très proches de Pyrococcales, *P. abyssi* et

*P. horikoshii*, ont été alignées l'une par rapport à l'autre au moyen du programme BLAST. Les segments de très grande similitude sont représentés en fonction des coordonnées (en Mégabases) de chacun des génomes. Les réarrangements des chromosomes des deux souches sont identifiés par les différences par rapport à la diagonale qui représente une colinéarité complète des séquences. La flèche indique l'emplacement de l'origine (*oriC*), la région correspondant au terminus de réplication (*TerC*) est également indiquée.

et organisation des génomes. Chez les bactéries, des études d'imagerie *in situ*, menées en particulier sur *Bacillus subtilis*, ont montré que ces réplisomes restent associés physiquement entre eux pendant la réplication [16]. Ceci pourrait expliquer l'existence d'événements de recombinaison qui se produisent de façon symétrique par rapport à l'axe reliant l'origine et le terminus [17]. Ces recombinaisons auraient lieu entre les régions d'ADN répliquées au même moment par les deux réplisomes. Or, la comparaison des génomes de *P. abyssi* et de *P. horikoshii*, une espèce sœur séquencée au Japon, a mis en évidence une grande inversion de l'un des deux génomes, symétrique par rapport à l'origine, en accord avec l'hypothèse d'une association des réplisomes [18]. Enfin, cette analyse a montré que ces deux génomes divergeaient principalement à partir de leur région de terminaison [15]. Le terminus du chromosome est donc un point chaud de recombinaison, chez les *Archaea* comme chez les bactéries (figure 1).

### Les *Archaea* utilisent les mêmes protéines que nous pour répliquer leur ADN

Comme il n'existe pas encore d'outil génétique permettant l'inactivation des gènes ou la production de mutants chez *Pyrococcus*, nous ne savons pas avec certitude si les enzymes de réplication de type eucaryote détectées chez cette *Archaea* sont bien celles qui répliquent son ADN. Au vu de son mode de réplication de type bactérien, on pourrait légitimement avoir des doutes. Il semble toutefois très probable que les protéines en question soient les bonnes. D'une part, la plupart d'entre elles ou leurs homologues chez d'autres *Archaea* ont maintenant été purifiées, et les activités détectées *in vitro* ont toujours été celles prédites par l'annotation *in silico*, correspondant bien à ce que l'on attend de protéines répliquatives (pour une revue récente, voir [19]). D'autre part, nos études *in vivo* et *in silico* ont montré que l'origine de réplication de *P. abyssi* est localisée dans une région qui regroupe des gènes

codant précisément pour plusieurs de ces protéines (figure 2) [15]. De façon significative, les origines de réplication chez les *Pyrococcus* et une autre *Archaea*, *Methanobacterium thermoautotrophicum* [14, 15] correspondent à une région intergénique localisée en amont du gène codant pour un homologue de la protéine eucaryote Cdc6. Cette région comprend une zone riche en AT, qui est probablement requise pour l'ouverture de la double hélice, et plusieurs séquences répétées, conservées entre ces différentes *Archaea*, qui pourraient être reconnues par la protéine Cdc6. Cette dernière joue précisément un rôle déterminant dans le déclenchement de la réplication chez les eucaryotes, en recrutant l'hélicase MCM (*mini chromosome maintenance*) au niveau des complexes d'initiation, les ORC (*origin of replication complexes*) [20]. Nous avons pu montrer récemment que la protéine Cdc6 de *P. abyssi* est spécifiquement associée à l'origine de réplication *in vivo*, ce qui confirme l'identification de Cdc6 comme étant la protéine initiateur de la réplication chez les *Archaea*

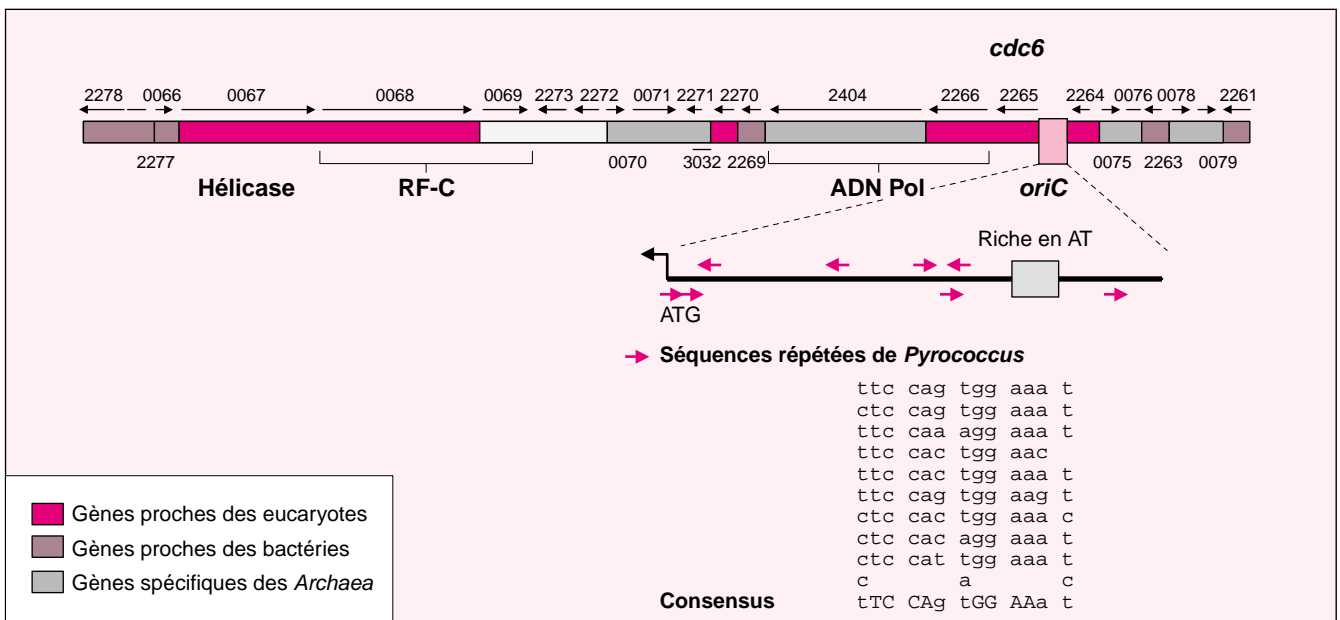


Figure 2. Vue élargie de la région située à l'origine de réplication du chromosome de *P. abyssi*. Plusieurs gènes impliqués directement dans la réplication de l'ADN (RF-C, DNA Pol et *cdc6*) sont situés dans la proximité immédiate d'*oriC*, l'origine de réplication proprement dite. *oriC* contient d'autre part des segments répétés très conservés (flèches rouges), ainsi qu'une région riche en nucléotides AT, favorable à l'ouverture des deux brins d'ADN lors de l'étape d'initiation de la réplication. Les nombres correspondent aux ORF annotés dans la séquence du génome de *P. abyssi*. Les flèches indiquent le sens de transcription.

[21]. De même, nous venons de montrer que les fragments d'Okazaki avaient exactement la même taille chez *P. abyssi* que chez la levure ( $\approx 150$  pb, contre de 1 à 2 kb chez les bactéries), ce qui est en accord avec l'idée selon laquelle l'étape d'élongation utilise les mêmes protéines dans les deux domaines.

### Le mécanisme de réplication des *Archaea* est un système eucaryote simplifié

Bien que le mécanisme de réplication des *Archaea* soit typiquement eucaryote, il s'agit d'un système simplifié quant au nombre de protéines et de polypeptides impliqués. Par exemple, l'hélicase MCM des *Archaea* est un hexamère composé de six sous-unités identiques tandis que son homologue eucaryote est formée de six sous-unités différentes [19]. De même la protéine RFC des *Archaea* possède deux sous-unités contre cinq pour son homologue eucaryote [19]. Notons également que les *Archaea* ne possèdent pas l'équivalent de l'ADN polymérase  $\alpha$ , qui intervient chez les eucaryotes pour ajouter un fragment d'ADN aux amorces ARN produites par la primase. Chez les eucaryotes, ces amorces ARN/ADN seront ensuite utilisées par les ADN polymérases  $\delta$  ou  $\epsilon$  pour initier la synthèse des fragments d'Okazaki. Chez les *Archaea*, la primase semble suffisante pour fabriquer l'amorce et passer directement le relais à la réplicase. Finalement, le génome de *P. abyssi* ne codant pas pour des homologues des composants de l'ORC, la protéine Cdc6 de *P. abyssi* doit être capable à la fois de reconnaître l'origine et de recruter l'hélicase MCM. En résumé, *P. abyssi* réplique son ADN plus vite que nous, et avec un plus petit nombre de protéines semblables aux nôtres, c'est à se demander qui est le plus archaïque des deux !

En tout cas, cette simplification du système de réplication chez les *Archaea* devrait faciliter son étude, et permettre de répondre plus facilement aux nombreuses questions qui se posent encore sur les mécanismes de la réplication chez l'homme. Ainsi, l'étude des protéines MCM

chez les *Archaea* a facilité la mise en évidence de son activité hélicase, ce qui s'était avéré difficile avec l'enzyme eucaryote. De même, notre observation sur la taille des fragments d'Okazaki chez les *Archaea* permet d'exclure l'hypothèse selon laquelle la vitesse réduite des fourches de réplication chez les eucaryotes serait due à la petite taille de leurs fragments d'Okazaki. De façon non exclusive, la différence observée entre la vitesse des fourches de réplication chez les *Archaea* et les eucaryotes pourrait s'expliquer, soit par la structure différente de la chromatine dans les deux domaines, soit par la simplification du système chez les *Archaea*, soit encore par une plus grande efficacité des protéines répliquatives des *Archaea* par rapport à leurs homologues eucaryotes.

### Pourquoi les bactéries font-elles bande à part ?

L'absence d'homologie entre, d'un côté, les protéines de réplication des *Archaea* et des eucaryotes et, de l'autre, celles des bactéries, ne semble pas pouvoir résulter d'une accélération drastique de la vitesse d'évolution de ces protéines dans l'un ou l'autre des trois domaines, ce qui aurait rendu impossible la détection de leur homologie. En effet, dans plusieurs cas, il est clair que les protéines de la réplication bactériennes se sont formées indépendamment de leurs homologues fonctionnels chez les *Archaea* et les eucaryotes. Ainsi, l'hélicase bactérienne DnaB est beaucoup plus apparentée à RecA qu'à l'hélicase MCM des *Archaea* et des eucaryotes [22]. De même, la primase des *Archaea* et des eucaryotes appartient à la superfamille des nucléotidyl-transférases de la classe X, qui ne comprend pas DnaG [23].

Deux hypothèses ont été avancées pour expliquer l'originalité du système de réplication bactérien. La première envisage l'invention indépendante du mécanisme de réplication, d'une part chez les bactéries, et de l'autre dans une lignée commune aux *Archaea* et aux eucaryotes [24]. Dans ce cas, le dernier ancêtre com-

mun aux trois domaines, LUCA (*the last universal cellular ancestor* [6]) possédait encore un génome à ARN. La seconde hypothèse envisage le remplacement des protéines de réplication initialement présentes chez LUCA par des analogues fonctionnels d'origine virale [25, 26]. De nombreuses protéines de réplication codées par des virus ne présentent en effet que très peu de similarité de séquence avec leurs homologues cellulaires. Le remplacement des protéines répliquatives de LUCA par des protéines répliquatives virales chez les bactéries ou chez les *Archaea*/eucaryotes pourrait ainsi expliquer l'absence d'homologie observée aujourd'hui entre protéines répliquatives de différents domaines.

Dans les deux cas, il reste à expliquer la ressemblance des autres mécanismes informationnels (transcription, traduction) entre les *Archaea* et les eucaryotes. On peut imaginer, soit que les *Archaea* et les eucaryotes partagent une histoire commune, distincte de celle des bactéries [27], soit que ces mécanismes étaient déjà présents chez LUCA et ont été fortement modifiés dans la lignée conduisant aux bactéries [28].

### Comment utiliser *Pyrococcus* pour identifier de nouvelles fonctions chez l'homme ? le cas des PACE

Quel que soit le scénario évolutif qui s'imposera dans le futur pour expliquer l'étroite parenté entre les *Archaea* et nous, cette constatation devrait conduire à des applications pratiques. Jusqu'à présent, les découvertes de nouvelles fonctions chez les eucaryotes grâce aux *Archaea* se sont souvent faites par hasard. Ainsi, il y a quelques années, la mise en évidence d'une nouvelle famille d'ADN topoisomérase de type II chez les *Archaea* a contribué de façon inattendue à l'identification de la protéine responsable de la cassure des chromosomes, la protéine SPO11, au cours de la recombinaison méiotique [29]. Aujourd'hui, les données de la génomique comparative devraient permettre d'orienter la recherche. Ainsi, nous avons récemment établi un catalogue de 32 protéines d'*Archaea*

de fonction inconnue qui sont présentes chez eucaryotes et absentes chez les bactéries [30]. Vu leur ancienneté et leur degré de conservation, ces protéines, que nous avons appelées PACE (pour *Proteins from Archaea Conserved in Eucarya*), doivent jouer un rôle majeur dans le fonctionnement de nos cellules, et leur étude pourrait être importante pour la recherche médicale. En effet, sur treize PACE pour lesquelles il existe des données génétiques ou biochimiques préliminaires chez l'homme ou d'autres eucaryotes, huit semblent impliquées dans la cancérogenèse [30]. Nous avons donc sans doute à portée de main plusieurs nouveaux oncogènes ou anti-oncogènes parmi la vingtaine de PACE humaines pour lesquelles nous ne savons encore absolument rien !

Il est déjà possible d'obtenir des informations au sujet de certaines PACE en étudiant l'environnement de leurs gènes chez les *Archaea*. Dans les génomes procaryotes, le regroupement des gènes codant pour des protéines qui interagissent physiquement ou fonctionnellement les uns avec les autres est en effet fréquent. Ainsi, l'environnement génomique chez les *Archaea* des PACE homologues de plusieurs oncogènes, répresseurs d'oncogènes, ou inducteurs d'apoptose de cellules tumorales, suggère que ces protéines

appartiennent ou interagissent avec le système de traduction [30]. Premier exemple, les gènes codant pour les PACE 4 et 5 sont situés côte à côte dans quatre génomes d'*Archaea* et sont, à chaque fois, proches du gène codant pour le facteur d'initiation de la traduction eIF1 $\alpha$  (figure 3). Cela suggère la possibilité d'une interaction chez l'homme entre les homologues des PACE 4 et 5 et eIF1 $\alpha$ . Cette hypothèse mériterait d'être testée vu l'importance potentielle de la PACE 5 humaine, appelée Rip-1, qui est connue d'une part pour interagir avec la protéine rev2 du virus du SIDA et, de l'autre, pour être impliquée dans le cancer des bronches [31]. Deuxième exemple, le gène codant pour l'homologue chez les *Archaea* de la protéine humaine XAB1, connue pour interagir avec la protéine de réparation XPA [32], est situé chez les trois *Pyrococcus* entre le gène codant pour MCM et celui codant pour une protéine apparentée à MinD (contrôle de la division cellulaire chez les bactéries). Il est tentant de penser que, en plus de son rôle dans la réparation, XAB1 intervient aussi chez l'homme dans le couplage entre l'initiation de la réplication et la division cellulaire. Signalons enfin que la plupart des PACE ne semblent pas présenter de repliements connus, ce qui devrait rendre leur analyse d'autant plus

intéressante pour la génomique structurale. L'étude des PACE chez l'homme et dans les organismes modèles devrait donc être un axe de recherche prometteur dans les années qui viennent. Plus généralement, nous allons certainement beaucoup apprendre de l'étude approfondie des mécanismes informationnels chez les *Archaea*. Il est temps de réaliser que nous avons enfin réussi à trouver à quoi ressemble un éléphant ! ■

## RÉFÉRENCES

1. Harfe BD, Jinks-Robertson S. DNA mismatch repair and genetic instability. *Annu Rev Genet* 2000; 34: 359-99.
2. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; 87: 4576-9.
3. Schleper C. Les Archaeobactéries sont parmi nous. *La Recherche* 1999; 317: 30-3.
4. Olsen GJ, Woese CR. Archaeal genomics: an overview. *Cell* 1997; 89: 991-4.
5. Forterre P. Archaea: what can we learn from their sequences? *Curr Opin Genet Dev* 1997; 7: 764-70.
6. Langer D, Hain J, Thuriaux P, Zillig W. Transcription in *Archaea*: similarity to that in eucarya. *Proc Natl Acad Sci USA* 1995; 92: 5768-72.
7. Forterre P, Elie C, Kohiyama M. Aphidicolin inhibits growth and DNA synthesis in halophilic archaeobacteria. *J Bacteriol* 1984; 159: 800-2.
8. Edgell DR, Doolittle WF. *Archaea* and the origin(s) of DNA replication proteins. *Cell* 1997; 89: 995-8.
9. <http://www.genoscope.cnrs.fr/Pab>
10. Cann IK, Komori K, Toh H, Kanai S, Ishino Y. A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc Natl Acad Sci USA* 1998; 95: 14250-5.
11. Myllykallio H, Forterre P. Mapping of a chromosome replication origin in an archaeon: response. *Trends Microbiol* 2000; 8: 537-9.
12. Erauso G, Reysenbach AL, Godfroy A, et al. *Pyrococcus abyssi*, sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch Microbiol* 1993; 160: 338-49.

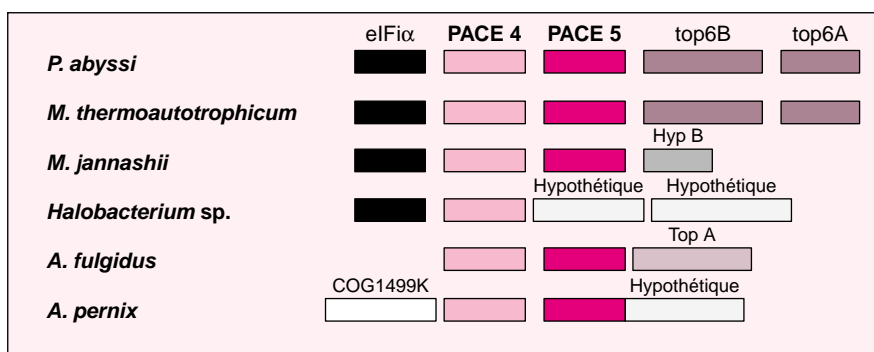


Figure 3. Environnement génomique conservé des PACE 4 et 5 chez différentes espèces d'*Archaea*. La PACE 4 est l'homologue de la protéine SudD du champignon *Neurospora crassa*; la PACE 5 est l'homologue de la protéine Rip-1 de l'homme. Une même couleur indique l'homologie des protéines. Top6A: topo-isomérase sous-unité A; Top6B: topo-isomérase sous-unité B; Hyp B: 3-hydroxybutyrate déshydrogénase présumptive; TopA: ADN topo-isomérase I; COG1499K: protéine de la transcription, intervenant dans la stabilité du ribosome et de l'ARN messager.

## RÉFÉRENCES

13. Higashibata H, Fujiwara S, Takagi M, Imanaka T. Analysis of DNA compaction profile and intracellular contents of archaeal histones from *Pyrococcus kodakaraensis* KOD1. *Biochem Biophys Res Commun* 1999; 258: 416-24.
14. Lopez P, Philippe H, Myllykallio H, Forterre P. Identification of putative chromosomal origins of replication in archaea. *Mol Microbiol* 1999; 32: 883-6.
15. Myllykallio H, Lopez P, Lopez-Garcia P, et al. Bacterial Mode of replication with an eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* 2000; 288: 2212-5.
16. Lemon KP, Grossman AD. Localization of bacterial DNA polymerase: evidence for a factory model of replication. *Science* 2000; 282: 1516-9.
17. Tillier ERM, Collins RA. Genome rearrangement by replication-directed recombination. *Nat Genet* 2000; 26: 195-8.
18. Makino S, Suzuki M, Zivanovic Y, Myllykallio H, Forterre P. Bacterial genomic reorganization upon DNA replication. *Science* 2001; 292: 803.
19. MacNeill SA. Understanding the enzymology of archaeal DNA replication: progress in form and function. *Mol Microbiol* 2001; 40: 520-9.
20. Kelly TJ, Brown GW. Regulation of chromosome replication. *Annu Rev Biochem* 2000; 69: 829-80.
21. Matsunaga F, Forterre P, Ishino Y, Myllykallio H. *In vivo* interactions of archaeal Cdc/Orcl and MCM proteins with the replication origin. *Proc Natl Acad Sci USA* 2001 (sous presse).
22. Leipe DD, Aravind L, Grishin NV, Koonin EV. The bacterial replicative helicase DnaB evolved from a RecA duplication. *Genome Res* 2000; 10: 5-16.
23. Kirk BW, Kuchta RD. Arg304 of human DNA primase is a key contributor to catalysis and NTP binding: primase and the family X polymerases share significant sequence homology. *Biochemistry* 1999; 38: 7727-36.
24. Leipe DD, Aravind L, Koonin EV. Did DNA replication evolve twice independently? *Nucleic Acids Res* 1999; 27: 3389-401.
25. Forterre P. Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins *Mol Microbiol* 1999; 33: 457-65.
26. Villarreal LP, DeFilippis VR. A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J Virol* 2000; 74: 7079-84.
27. Woese CR. Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci USA* 2000; 18: 8392-6.
28. Forterre P, Philippe H. Where is the root of the universal tree of life?. *Bioassays* 1999; 21: 871-9.
29. Bergerat A, De Massy B, Gabelle D, Varoutas PC, Nicolas A, Forterre, P. An atypical type II DNA topoisomerase from *Archaea* with implication for meiotic recombination. *Nature* 1997; 386: 414-7.
30. Matte-Taillez O, Forterre P, Zivanovic, Y. Mining archaeal proteomes for eukaryotic proteins with novel functions: the PACE case. *Trends Genet* 2000; 16: 533-6.
31. Gure AO, Altorki NK, Stockert E, Scallan MJ, Old LJ, Chen YT. Human lung cancer antigens recognized by autologous antibodies: definition of a novel cDNA derived from the tumor suppressor gene locus on chromosome 3p21.3. *Cancer Res* 1998; 58: 1034-41.
32. Nitta M, Saijo M, Kodo N, et al. A novel cytoplasmic GTPase XAB1 interacts with DNA repair protein, XPA *Nucleic Acids Res* 2000; 28: 4212-8.

**Patrick Forterre**  
**Yvan Zivanovic**  
**Oriane Matte-Taillez**  
**Fujihiko Matsunaga**  
**Hannu Myllykallio**

*Institut de génétique et de microbiologie, Bât 409, Cnrs, UMR 8621, Université Paris-Sud, 91405 Orsay cedex, France.*

## TIRÉS A PART

P. Forterre.