



HAL
open science

La PolD, une ADN polymérase atypique d'arches et un nouvel outil prometteur pour les biotechnologies

Inès Marie Pauline Hugonneau Beaufet

► To cite this version:

Inès Marie Pauline Hugonneau Beaufet. La PolD, une ADN polymérase atypique d'arches et un nouvel outil prometteur pour les biotechnologies. Sciences du Vivant [q-bio]. 2019. dumas-02301874

HAL Id: dumas-02301874

<https://dumas.ccsd.cnrs.fr/dumas-02301874v1>

Submitted on 30 Sep 2019

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.

U.F.R DES SCIENCES PHARMACEUTIQUES

Année 2019

Thèse n°46

THÈSE POUR L'OBTENTION DU
DIPLÔME D'ÉTAT de DOCTEUR EN PHARMACIE

Présentée et soutenue publiquement

Le 12 juin 2019 à 17h

Par **Inès Marie Pauline HUGONNEAU BEAUFET**

Née le 5 mai 1994 à Bruges

**LA POLD, UNE ADN POLYMÉRASE ATYPIQUE
D'ARCHÉES ET UN NOUVEL OUTIL PROMETTEUR POUR
LES BIOTECHNOLOGIES**

Sous la direction du Professeur Pascale Dufourcq

Membres du jury :

Professeur Isabelle Bestel

Présidente

Docteur Ludovic Sauguet

Membre du jury

U.F.R DES SCIENCES PHARMACEUTIQUES

Année 2019

Thèse n°46

THÈSE POUR L'OBTENTION DU
DIPLÔME D'ÉTAT de DOCTEUR EN PHARMACIE

Présentée et soutenue publiquement

Le 12 juin 2019 à 17h

Par **Inès Marie Pauline HUGONNEAU BEAUFET**

Née le 5 mai 1994 à Bruges

**LA POLD, UNE ADN POLYMÉRASE ATYPIQUE
D'ARCHÉES ET UN NOUVEL OUTIL PROMETTEUR POUR
LES BIOTECHNOLOGIES**

Sous la direction du Professeur Pascale Dufourcq

Membres du jury :

Professeur Isabelle Bestel

Présidente

Docteur Ludovic Sauguet

Membre du jury

REMERCIEMENTS

À **Madame Pascale Dufoureq**, pour m'avoir fait l'honneur de diriger cette thèse. Vos cours m'ont toujours passionnée et ont beaucoup influencé mes choix tout au long de mon cursus . Merci pour votre disponibilité et pour toute l'aide que vous m'avez apportée pendant la rédaction de ce manuscrit.

À **Madame Isabelle Bestel**, pour avoir accepté la présidence de ce jury de thèse et pour l'intérêt que vous avez porté à mon travail. Merci pour les connaissances et les conseils que vous apportez à vos étudiants pendant ces années de pharmacie.

À **Monsieur Ludovic Sauguet**, pour avoir accepté de faire partie du jury et pour un million d'autres choses. Merci pour ta confiance, ta bienveillance, ta disponibilité sans faille et tout ce que tu m'apprends chaque jour. J'ai énormément évolué à tes côtés et j'espère pouvoir continuer à le faire pendant quelques années encore.

À **Monsieur Marc Delarue**, pour m'avoir accueillie au sein de son équipe. Merci pour vos conseils avisés, soyez assuré de mon profond respect.

À **Monsieur Patrick England** pour votre aide indispensable à la mise en place des tests d'interaction et pour m'avoir guidée tout au long de ces expériences avec beaucoup de patience.

À **Monsieur Ahmed Haouz** pour votre aide précieuse avant chaque run Synchrotron, pour votre disponibilité et tous vos conseils avisés.

À **Pierre**, pour tout ce que tu m'as appris pendant mon stage, pour avoir su prendre le temps de m'épauler, pour ta très grande bienveillance, je suis beaucoup plus cosy grâce à toi.

À **Camille**, pour tout ce que tu m'apportes depuis que tu es arrivée au labo. Cette aventure ne serait pas la même sans toi, et de jus de tomate en jus de tomate ma formation intensive en répatie commence à payer.

À **Clément**, pour ta bonne humeur, pour me faire mourir de rire tous les jours et pour ton soutien de tous les instants. Merci pour ton immense aide au quotidien et pendant la rédaction de cette thèse, pour toutes tes relectures et tes conseils (rond de doigt).

À tous les membres de l'équipe, Sandrine, Dariusz, et aussi Margaux et Victoire, pour la bonne ambiance qui règne au labo et qui rend ce lieu de travail si épanouissant. Margaux, Victoire, revenez vite nous voir !

À mes parents, mon frère et ma sœur, pour votre soutien indéfectible qui me porte tous les jours. Merci pour tout votre amour, vous êtes mes piliers sans qui cette réussite n'aurait pas été possible.

À ma grand-mère, pour m'avoir suivie dans toutes mes aventures avec autant de constance. Merci pour votre présence et votre soutien précieux. **À mon grand-père et à ma grand-mère,** que j'espère rendre fiers.

À Mélanie, Pauline, Charlène, Julie et Solenn, pour avoir toujours été présentes, et l'être encore aujourd'hui, même éparpillées aux quatre coins de la France (et depuis quelques mois, du monde !). Merci pour toutes ces années passées ensemble et toutes les prochaines à venir !

À Claire, pour être un véritable soutien tous les jours, bien qu'on soit rarement dans la même ville en même temps. Merci pour cette amitié qui compte énormément et pour tous ces souvenirs, même ceux qu'on ne connaît pas encore !

À Paul, pour cette première discussion plutôt mal entamée... et surtout pour le reste. Pour ce périple à l'autre bout du monde, pour être ma bulle d'air pendant les moments difficiles, et pour les millions d'autres choses que je pourrais écrire. Merci pour tout.

TABLE DES MATIÈRES

INTRODUCTION

INTRODUCTION GÉNÉRALE.....	11
I. LES ARCHEES : LE TROISIEME DOMAINE DU VIVANT	13
1. Découverte et classification des Archées	13
<i>Histoire de la découverte des Archées</i>	<i>13</i>
<i>Classification des Archées.....</i>	<i>14</i>
2. Caractéristiques comparées des Archées.....	16
<i>Biotopes et métabolismes.....</i>	<i>16</i>
<i>Morphologie cellulaire et biochimie des membranes.....</i>	<i>16</i>
<i>Génome et compaction de l'ADN.....</i>	<i>18</i>
3. Intérêts majeurs de l'étude des Archées	18
<i>Un changement radical de l'ordre établi</i>	<i>18</i>
<i>Un modèle d'étude simplifié.....</i>	<i>19</i>
<i>Un large panel de propriétés à exploiter.....</i>	<i>19</i>
4. Le réplisome des Archées.....	20
<i>Initiation de la réplication.....</i>	<i>22</i>
<i>Progression de la fourche de réplication</i>	<i>23</i>
II. LA POLD : UNE ADN POLYMERASE ATYPIQUE D'ARCHEE.....	24
1. Classification structurale des ADN polymérases.....	24
<i>Diversité des ADN polymérases réplicatives dans les trois domaines du vivant.....</i>	<i>24</i>
<i>Classification structurale des cœurs catalytiques des ADN polymérases</i>	<i>26</i>
<i>Classification structurale des domaines de relecture des ADN polymérases.....</i>	<i>27</i>
2. L'ADN polymérase de la famille D (Pold).....	27
3. Étude structurale de la Pold de <i>Pyrococcus abyssi</i>	28
4. Scénario de l'histoire évolutive des ADN polymérases réplicatives eucaryotes	30
5. Potentiel biotechnologique de la Pold	31
<i>Les ADN polymérases thermostables sont couramment utilisées en PCR</i>	<i>31</i>
<i>La Pold possède un grand potentiel biotechnologique.....</i>	<i>32</i>
III. LE PCNA : UN ACTEUR CLE DE LA FOURCHE DE REPLICATION	33
1. Fonctions biologiques et diversité structurale des facteurs de processivité.....	33
2. État de l'art concernant les interactions entre la Pold et le PCNA.....	34

MATÉRIEL ET MÉTHODES

I. EXPRESSION DES PROTEINES RECOMBINANTES	37
1. Constructions plasmidiques.....	37
2. Souches bactériennes d'expression.....	37
3. Transformation bactérienne par choc thermique	37
4. Induction de l'expression des protéines	38
II. PURIFICATION DES PROTEINES.....	38
1. Lyse cellulaire.....	39
2. Chromatographies	39
<i>Chromatographie de pseudo-affinité sur colonne de nickel</i>	<i>39</i>
<i>Chromatographie d'affinité sur colonne d'héparine</i>	<i>40</i>
<i>Chromatographie d'exclusion de taille</i>	<i>40</i>
III. CRISTALLOGRAPHIE AUX RAYONS X.....	41
1. Cristallogenèse.....	41
<i>Techniques utilisées.....</i>	<i>41</i>
<i>Essais de cristallogenèse sur le PCNA.....</i>	<i>42</i>
<i>Obtention des cristaux de PCNA.....</i>	<i>42</i>
2. Diffraction des rayons X et traitement des données.....	42
<i>Organisation d'un cristal</i>	<i>42</i>
<i>Diffraction des rayons X par un cristal.....</i>	<i>43</i>
<i>Traitement des données</i>	<i>44</i>
3. Phasage et construction du modèle	45
IV. TESTS D'INTERACTION PAR INTERFEROMETRIE A BIO-COUCHES	45
1. Principe général de la technique	45
2. Design des expériences	46

RÉSULTATS

I. PRODUCTION DU PCNA DE PYROCOCCUS ABYSSI	48
1. Expression recombinante.....	48
2. Purification	48
II. CRISTALLOGRAPHIE ET ETUDES STRUCTURALES DU PCNA DE PYROCOCCUS ABYSSI.....	51
1. Cristallogenèse et collecte des données de diffraction aux rayons X.....	51
2. Résolution de la structure cristallographique par remplacement moléculaire	51
3. Analyse structurale du PCNA de Pyrococcus abyssi	53

III. CARTOGRAPHIE DE L'INTERACTION ENTRE LE PCNA ET LA POLD DE <i>PYROCOCCUS ABYSSI</i>	56
1. Production des différentes constructions utilisées	56
2. Tests d'interaction par interférométrie à bio-couches	57
<i>Mise au point des conditions de l'expérience</i>	57
<i>Étude comparative des interactions entre le PCNA et les différentes versions tronquées de la PolD</i>	59
IV. PROPOSITION D'UN MODELE D'INTERACTION ENTRE LE PCNA ET LA POLD	60

DISCUSSION

I. DEVELOPPEMENT DU POTENTIEL BIOTECHNOLOGIQUE DE LA POLD	63
1. Le rôle clé des ADN polymérase.....	63
2. Les ADN polymérase thermostable utilisées en PCR.....	63
3. Une famille d'ADN polymérase au potentiel biotechnologique inexploité.....	64
4. Intérêts du couplage de la PolD avec le PCNA	65
5. Le PCNA, véritable maestro de la fourche de réplication.....	67
II. PERSPECTIVES	67
1. Sur le complexe d'initiation	67
2. Sur le complexe d'élongation	69

CONCLUSION

BIBLIOGRAPHIE

ABBREVIATIONS

<i>E. coli</i>	<i>Escherichia. coli</i>
<i>P. abyssi</i>	<i>Pyrococcus abyssi</i>
<i>P. furiosus</i>	<i>Pyrococcus furiosus</i>
<i>P. horikoshii</i>	<i>Pyrococcus horikoshii</i>
A	Adénine
ADN	Acide DésoxyriboNucléique
ARN	Acide RiboNucléique
BSA	Albumine de Sérum Bovin
C	Cytosine
Cdc45	<i>Cell Division Control protein 45</i>
CMG	Cdc45-MCM-GINS
DPANN	Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaeota
EDTA	Acide Ethylène Diamine Tétracétique
G	Guanine
GINS	Go-Ichi-Nii-San
HEPES	Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
IPTG	Iso-propylthiogalactopyranoside
LB	Milieu Luria Berani
MCM	<i>Minichromosome Maintenance protein Complex</i>
NTA	Acide nitrotriacétique
ORB	<i>Origin Recognition Box</i>
ORC	<i>Origin Recognition Complex</i>
PCNA	Proliferating Cell Nuclear Antigen
PDB	Protein Data Bank
PIP	<i>PCNA-Interacting Protein</i>
Pol-I	ADN polymérase I
Pol-III	ADN polymérase III
Polα	ADN polymérase α
Polδ	ADN polymérase δ
Polϵ	ADN polymérase ϵ

Poly	ADN polymérase γ
PolB	ADN Polymérase B
PolD	ADN Polymérase D
RT	<i>Reverse Transcriptase</i>
SDS	Dodécylsulfate de sodium
SDS-PAGE	Electrophorèse sur gel de polyacrylamide en condition dénaturante
SPR	Résonance plasmonique de surface
SSB	<i>Single-Stranded DNA Binding protein</i>
Sso	ADN polymérase PolY de <i>Sulfolobus solfataricus</i>
T	Thymine
TACK	Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota
Taq	ADN polymérase Pol-I de <i>Thermus aquaticus</i>
TdT	<i>Terminal Deoxynucleotidyl Transferase</i>
TRIS	Trishydroxyméthylaminométhane
U	Uracile

TABLE DES FIGURES

- Figure 1.** Arbre de la vie découpé en trois domaines du vivant
- Figure 2.** Arbre phylogénétique des trois domaines du vivant
- Figure 3.** Aperçu de la diversité morphologique des Archées
- Figure 4.** Modèle d'architecture de la fourche de réplication chez les Archées
- Figure 5.** Distribution des ADN polymérase répliquatives dans les trois domaines du vivant
- Figure 6.** Classification structurale mise à jour des cœurs catalytiques des ADN polymérase
- Figure 7.** Classification structurale mise à jour des sites exonucléases associés aux ADN polymérase répliquatives
- Figure 8.** Structure des sous-unités DP1 et DP2 de la PolD de *Pyrococcus abyssi* obtenue par cristallographie aux rayons X
- Figure 9.** Structure de la PolD entière de *Pyrococcus abyssi* en complexe avec l'ADN obtenue par cryo-microscopie électronique
- Figure 10.** Relation évolutive hypothétique entre les PolB et PolD des Archées et les ADN polymérase répliquatives eucaryotes
- Figure 11.** Structure des facteurs de processivité dans les trois domaines du vivant
- Figure 12.** Structure des ADN polymérase répliquatives Pol-III et PolB3 en complexe avec leurs facteurs de processivité respectifs, $\beta 2$ et PCNA
- Figure 13.** Cliché de diffraction
- Figure 14.** Principe de la technique d'interférométrie à bio-couches
- Figure 15.** Suivi des étapes de purification du PCNA via des profils électrophorétiques et chromatographiques
- Figure 16.** Obtention des cristaux de PCNA
- Figure 17.** Structure en cours d'affinement dans le logiciel COOT
- Figure 18.** Représentation tridimensionnelle du PCNA sous différentes formes oligomériques
- Figure 19.** Répartition du potentiel électrostatique de surface sur les deux faces du PCNA
- Figure 20.** Répartition des régions conservées à la surface du PCNA
- Figure 21.** Représentation schématique des différentes constructions de la PolD utilisées lors des tests d'interaction et concentrations choisies pour chacune d'elles
- Figure 22.** Choix de la protéine contrôle à utiliser
- Figure 23.** Résultats des tests d'interaction entre les différentes constructions de la PolD et le PCNA interférométrie en bio-couches
- Figure 24.** Représentations tridimensionnelles du modèle d'interaction proposé entre la PolD de *Pyrococcus abyssi* en complexe avec l'ADN et le PCNA de *Pyrococcus abyssi*
- Figure 25.** Cœurs catalytiques des ADN polymérase thermostables utilisées en PCR

Figure 26. Structure du complexe PolD-ADN-PCNA de *Pyrococcus abyssi* en mode élongation en cryo-microscopie électronique (en cours de résolution au laboratoire)

Figure 27. Structure de PriS en complexe avec PriL sans son domaine C-terminal obtenue par cristallographie aux rayons X à 1,6 Å de résolution

Figure 28. Résolution de structure de la primase entière de *Pyrococcus abyssi* en cours

TABLE DES TABLEAUX

Tableau 1. Comparaison des principales caractéristiques des Archées, Bactéries et Eucaryotes

Tableau 2. Comparaison des réplisomes dans les trois domaines du vivant

Tableau 3. Classification des ADN polymérase en sept familles distinctes

Tableau 4. État de l'art concernant la mise en évidence et la caractérisation des interactions entre le PCNA et la PolD

Tableau 5. Caractéristiques des constructions plasmidiques utilisées

Tableau 6. Tampons utilisés au cours des purifications

Tableau 7. Caractéristiques des jeux de donnée collectés et statistiques d'affinement de la structure

INTRODUCTION

INTRODUCTION GÉNÉRALE

Le monde du vivant est partagé en trois grands domaines : les Bactéries, les Archées et les Eucaryotes (1,2). Si tous ces organismes se distinguent d'un point de vue morphologique et physiologique, leur viabilité dépend de la même molécule : l'ADN. Support de leur information génétique, celle-ci assure le développement, le fonctionnement et la reproduction de tous les êtres vivants. Le dogme central de la biologie moléculaire, discipline consacrée à l'étude des molécules porteuses du message héréditaire, a été formulé en 1958 et décrit un flux d'information génétique allant principalement dans le sens ADN, ARN, protéine (3,4). L'ADN génomique est transcrit en ARN messager par les ARN polymérases, puis la molécule d'ARN est traduite en protéine par le ribosome.

Au cœur de ce processus, la synthèse de l'ADN est essentielle à la multiplication des cellules en orchestrant la transmission de l'information génétique de la cellule mère aux cellules filles. La molécule d'ADN de la cellule mère est copiée de manière fidèle et intégrale : on parle de duplication ou de réplication. La réplication est un processus très régulé dans l'espace et dans le temps, faisant intervenir des réseaux de protéines complexes, appelé le réplisome.

Les ADN polymérases sont des enzymes clés de ce réplisome, impliquées dans la réplication de l'ADN au cours du cycle cellulaire, mais aussi dans des processus de réparation et de recombinaison de l'ADN. En se basant sur leurs séquences protéiques et sur des analyses structurales, elles ont pu être classées en sept familles : PolA, PolB, PolC, PolD, PolX, PolY et les reverse transcriptases (RT) (5-7). Les ADN polymérases répliquatives impliquées dans la réplication du génome appartiennent aux familles A, B, C et D. Contrairement aux ARN polymérases et aux ribosomes, qui ont évolué par accréation de complexité moléculaire autour d'un cœur catalytique universel (8,9), l'évolution des ADN polymérases répliquatives est bien plus compliquée et aucune d'elles n'est retrouvée dans les trois domaines du vivant (10). En effet, le génome des Bactéries est principalement répliqué par une ADN polymérase de famille C, alors que le génome des Eucaryotes, des Archées et de nombreux virus est répliqué par une ADN polymérase de famille B.

Les Archées expriment également une ADN polymérase répliquative atypique appartenant à la famille D, appelée PolD, et radicalement distincte des autres ADN polymérases (11,12). Mon laboratoire d'accueil, l'unité Dynamique Structurale des Macromolécules à l'Institut Pasteur de Paris, a déterminé la structure de la PolD par cristallographie aux rayons X et cryo-microscopie électronique. Ces données structurales leur ont permis de montrer que la PolD possédait un cœur catalytique totalement différent de celui des autres ADN polymérases mais similaire à celui des ARN polymérases, suggérant une origine évolutive commune entre la réplication et la transcription de l'ADN (13-15).

Ces travaux ouvrent la voie à une ingénierie enzymatique rationnelle, sur laquelle je me suis appuyée afin d'étudier l'interaction entre la PolD et un autre acteur clé de la fourche de réplication : le PCNA (*Proliferating Cell Nuclear Antigen*). Il s'agit d'une protéine de la famille des *sliding clamps* qui forme un anneau autour de l'ADN, participe à recruter d'autres partenaires de la fourche de réplication et stimule l'activité des ADN polymérases réplicatives chez les eucaryotes et chez les archées. Au cours de mes travaux, j'ai résolu la structure du PCNA de l'espèce *Pyrococcus abyssi* par cristallographie aux rayons X et j'ai étudié ses interactions avec la PolD. Ces données nous ont permis de construire un modèle tridimensionnel du complexe ternaire entre la PolD, l'ADN et le PCNA.

Une meilleure compréhension du fonctionnement du PCNA au sein de la fourche de réplication permettrait d'optimiser son couplage avec la PolD afin de développer le potentiel biotechnologique encore largement inexploré de cette dernière. Atypique et peu caractérisée avant les travaux réalisés ces dernières années dans mon laboratoire d'accueil, nombre de ses propriétés en font un nouvel outil très intéressant avec des applications en diagnostic clinique et en police scientifique. De plus, de nombreuses maladies héréditaires ou certains cancers sont dus à des mutations sur des protéines humaines qui possèdent des homologues très proches chez les Archées. Ces dernières constituent un modèle d'étude simplifié chez qui la compréhension de ces mécanismes peut s'avérer bien plus facile.

L'introduction de ce manuscrit de thèse présentera dans un premier temps les Archées, leur découverte et leur place dans la classification, leurs caractéristiques et l'intérêt de les étudier, pour enfin se concentrer sur le fonctionnement de leur réplisome. Dans un deuxième temps, nous nous focaliserons sur l'ADN polymérase D, son caractère atypique et sa place dans la classification, ainsi que son potentiel biotechnologique encore inexploité. Dans un troisième temps, nous nous intéresserons au PCNA afin de discuter de sa fonction biologique, de sa structure et de l'état de l'art concernant son interaction avec la PolD.

Ces travaux ont pour objectif d'établir une cartographie des interactions entre la PolD et le PCNA et de proposer un modèle du complexe ternaire qu'elles forment avec l'ADN. Comprendre la façon dont ce complexe fonctionne chez les Archées paverait la voie à une meilleure compréhension des mécanismes eucaryotes homologues. De plus, l'étude du réplisome des Archées hyperthermophiles présente un grand intérêt sur le plan biotechnologique. En effet, les propriétés des enzymes thermostables peuvent être utilisées pour améliorer les performances de réplication de l'ADN in vitro dans les techniques de PCR.

I. Les Archées : le troisième domaine du vivant

1. Découverte et classification des Archées

La très grande biodiversité présente à la surface de notre planète fait depuis longtemps l'objet d'un effort constant de classification aboutissant très souvent à un système dichotomique : animal/végétal, puis multicellulaire/unicellulaire avec l'apparition du microscope optique au XVII^{ème} siècle, et enfin eucaryote/procaryote suite au développement du microscope électronique au début du XX^{ème} siècle. Mais à partir des années 50, la meilleure compréhension de l'information génétique et de son codage, ainsi que la nécessité d'établir une classification homogène du vivant ont poussé les chercheurs à développer de nouvelles techniques de phylogénie moléculaire.

Histoire de la découverte des Archées

Les Archées ont été mises en lumière pour la première fois à la fin des années 70 par les travaux de Carl Woese et Georges Fox (16). Pour la première fois, le marqueur évolutif ne repose plus sur des critères morphologiques mais sur marqueur moléculaire universel : l'ARN ribosomique 16S. Composant de la petite sous-unité des ribosomes, cet acide nucléique est présent en grande quantités dans toutes les cellules vivantes et très conservé entre les différentes espèces. Il constitue donc un élément de comparaison qui a permis à Carl Woese d'établir des liens de parenté et de reconstituer un arbre phylogénétique universel (**Figure 1**).

Dans un premier temps, ce dernier a étudié plusieurs espèces « modèles » de bactéries, levures, plantes et animaux. En comparant les séquences nucléotidiques leurs ARN ribosomiques 16S, il a constaté que les espèces eucaryotes présentaient des motifs très similaires, et étaient donc proches du point de vue phylogénique. Les mêmes observations ont été faites pour deux espèces de bactéries. En revanche, ces motifs étaient très différents entre bactéries et eucaryotes, renforçant l'idée d'une profonde distinction eucaryote/procaryote.

C'est en étudiant une culture de bactéries atypiques appelées méthanogènes que Carl Woese découvrit des motifs complètement différents de ceux des eucaryotes, mais aussi de ceux de toutes les autres bactéries. Ces microorganismes, anaérobies strictes et producteurs de méthane, étaient déjà connus pour leur métabolisme « exotique » et Woese en conclut qu'ils formaient un groupe à part. Il nomma ce dernier *Archaeobacteria*, du grec « arkhaia » pour ancien (16). Dans les années qui suivirent, de nombreuses espèces de bactéries comme les halophiles (17) ou les hyperthermophiles (18) furent affiliées à ce nouveau domaine finalement baptisé *Archaea*.

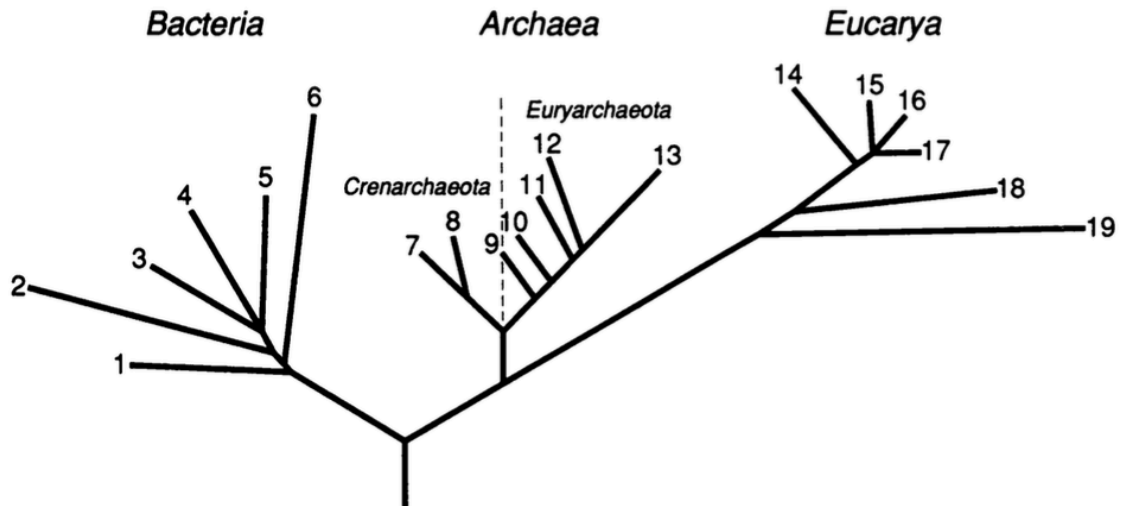


Figure 1. Arbre de la vie découpé en trois domaines du vivant (extraite de Woese *et al*, 1990) (1). L'ordre et la longueur des branches sont basées sur la comparaison des séquences d'ARNr 16S-18S. A la base de l'arbre, la racine représente LUCA (Last Universal Common Ancestor), le dernier ancêtre commun de tous les êtres vivants. La numérotation au bout des branches correspond aux groupes d'organismes suivants. **Dans le domaine des Bactéries :** 1, Thermogales ; 2, flavobactéries ; 3, cyanobactéries ; 4, Bactéries violettes ; 5, Bactéries Gram-positives ; 6, Bactéries vertes. **Dans le domaine des Archées :** le « règne » Crenarchaeota ; 7, Pyrodictium ; 8, Thermoproteus ; le « règne » Euryarchaeota ; 9, Thermococcales ; 10, Methanococcales ; 11, Methanobactériales ; 12, Methanomicrobiales ; 13, halophiles extrêmes. Dans le domaine des Eucaryotes : 14, animaux ; 15, ciliés ; 16, plantes vertes ; 17, champignons ; 18, flagellés ; 19, microsporidia.

Classification des Archées

À partir des années 1980, il fut admis qu'il existait non plus deux, mais trois domaines du vivant nommés *Eucarya*, *Bacteria* et *Archaea* et que le terme « procaryote » rassemblait artificiellement deux groupes aussi éloignés l'un de l'autre qu'ils pouvaient l'être des eucaryotes.

Depuis la mise en place des trois domaines du vivant, la diversité phylogénétique des Archées n'a cessé d'augmenter. À l'origine, Carl Woese et ses collègues avaient identifié deux phyla distincts au sein des Archées : les Crenarchaeota et les Euryarchaeota (1). Ceux-ci demeurent les plus étudiés aujourd'hui, notamment parce qu'ils sont cultivables en laboratoire. Les analyses phylogénétiques récentes reconnaissent à présent quatre superphyla (sous-domaines) au sein des archées : les Euryarchaeota, les TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota et Korarchaeota), les DPANN (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota et Nanohaloarchaeota) et les Asgard (Odinarchaeota, Thorarchaeota,

Heimdallarchaeota et Lokiarahaeta, d'après le nom du royaume des dieux de la mythologie nordique) (19,20) (Figure 2).

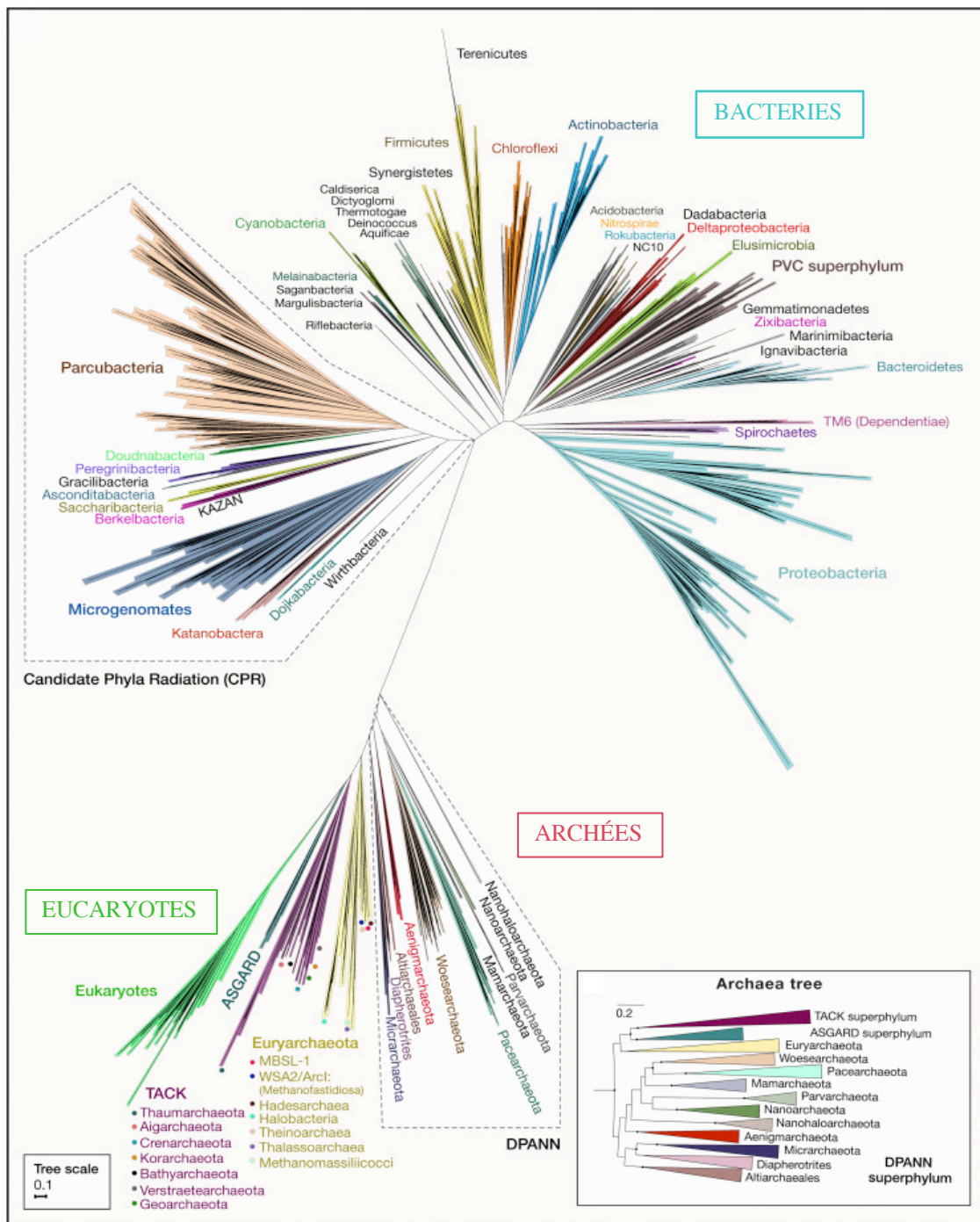


Figure 2. Arbre phylogénétique des trois domaines du vivant (extraite de Castelle and Banfield, 2018) (20). Dans ce nouveau modèle, le domaine des archées est composé de quatre superphylums : les Euryarchaeota, les ASGARD, les TACK et les DPANN. La branche qui regroupe les eucaryotes et les archées est placée entre un groupe de bactéries bien connues (comme les Probacteria, Actinobacteria, Firmicutes et Cyanobacteria) et un nouveau groupe monophylétique, Candidate Phyla Radiation. Les Euryarchaeota sont placées entre les ASGARD et les TACK, plus proches des eucaryotes, et les DPANN, plus éloignées, bien qu'il soit envisageable qu'ils forment un même superphylum avec les DPANN.

2. Caractéristiques comparées des Archées

Biotopes et métabolismes

Les Archées colonisent tous les biotopes et représenteraient 20% de la biomasse totale de notre planète (21). Elles sont ubiquitaires et peuvent vivre dans des habitats extrêmes comme des geysers ou des fumeurs noirs, où la température de l'eau est supérieure à 100°C (22–24), des environnements océaniques froids tels que les mers polaires (25), des environnements salins tels que la Mer Morte et les lacs salés (17), des environnements alcalins et acides pouvant atteindre un pH de 0 (26), mais également des habitats plus répandus comme les océans et leurs sédiments, les eaux usées, les marais et les sols de jardins et forêts (27).

Notons que les Archées ont également un caractère symbiotique. C'est notamment le cas des méthanogènes présentes dans le tube digestif des mammifères (28) et des Archées appartenant au superphylum DPANN, qui dépendraient principalement d'autres microorganismes (Bactéries, Archées ou Eucaryotes) pour satisfaire leurs besoins en nucléotides, acides aminés et lipides.

Morphologie cellulaire et biochimie des membranes

Au microscope, on distingue difficilement une Archée d'une Bactérie, et c'est d'ailleurs la raison de leur découverte tardive. De tailles similaires (de 0,1 à 15 µm), les Archées présentent souvent des flagelles protéiques permettant de se déplacer et adoptent, selon l'espèce, des formes de coques, de grappes, de bâtonnets, de filaments (**Figure 3**). En plus de se diviser par scission binaire, elles peuvent créer des biofilms, ont un unique chromosome circulaire, et peuvent posséder un ou plusieurs éléments extra-chromosomiques, les plasmides.

Il s'agit donc en apparence de véritables sosies des Bactéries. Cependant les chercheurs ont constaté que les Archées possédaient des caractéristiques uniques qui en font bel et bien un groupe distinct de celui des eucaryotes et des bactéries (**Tableau 1**). À l'inverse des Bactéries, la majorité des Archées ne possède qu'une seule membrane enveloppée de glycoprotéines qui forment une couche semi-cristalline à la surface de la cellule (la « *S-layer* ») (29). De plus, les bicouches lipidiques des Archées sont constituées de lipides atypiques qui diffèrent complètement des lipides membranaires bactériens et eucaryotes. Cette particularité de l'enveloppe cellulaire des Archées serait le résultat d'une adaptation aux environnements extrêmes.

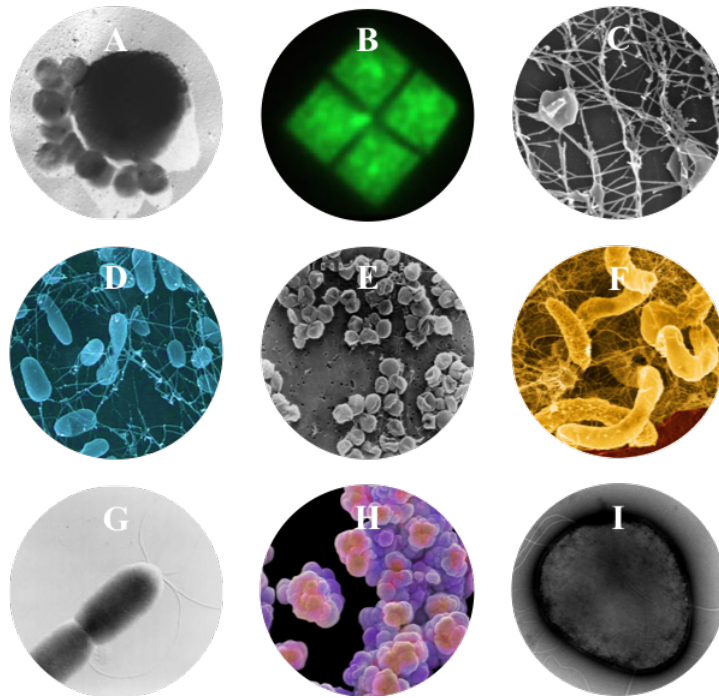


Figure 3. Aperçu de la diversité morphologique des Archées. (A) *Ignococcus* et son symbiote, *Nanoarcheum equitans* ; (B) *Haloquadratum* ; (C) *Pyrodictium abyssi* ; (D) *Pyrolobus fumarii* ; (E) *Pyrococcus abyssi* ; (F) *Desulfovibrio desulfuricans* ; (G) *Aquifex aeolicus* ; (H) *Methanosarcina sp.* ; (I) *Sulfolobus sp.*

		EUCARYOTES	ARCHÉES	BACTÉRIES
Taille		> 5 μm	1-4 μm	1-4 μm
Matériel génétique		Plusieurs chromosomes linéaires	Chromosome circulaire	Chromosome circulaire
Organelles		Oui	Non	Non
Noyau		Oui	Non	Non
Enveloppe cellulaire	Liaison des lipides	Ester	Ester	Ether
	Chaînes carbonées	Acides gras	Isoprènes	Acides gras
	Tête phosphate	D-glycérol-3-phosphate	L-glycérol-1-phosphate	D-glycérol-3-phosphate

Tableau 1. Comparaison des principales caractéristiques des Archées, Bactéries et Eucaryotes. D'un point de vue morphologique, les cellules d'Archées sont des cellules procaryotes typiques : elles ne possèdent ni noyau ni organites. La comparaison entre les trois domaines du vivant permet de mettre en évidence le caractère mosaïque de l'organisation moléculaire des cellules d'Archées qui « empruntent » certaines fonctions au domaine des Bactéries et d'autres à celui des Eucaryotes.

Génome et compaction de l'ADN

En termes de taille et d'organisation, le génome des Archées ne se distingue pas significativement du génome bactérien. Le chromosome des Archées est circulaire, comme celui des Bactéries, et sa taille est comprise entre ~0,5 Mpb (*Nanoarchaeum equitans*) et ~5,8 Mpb (*Methanosarcina mazei*). La majorité des protéines des Archées et des Bactéries sont codés dans des ORFs (*Open Reading Frame*) ininterrompus et organisés en opérons. A l'inverse, le génome des Eucaryotes est relativement plus large (2 850 Mpb chez l'humain) et contient des gènes codants interrompus par des introns et séparés par de longues régions inter-géniques (30).

Les Archées expriment une grande diversité de protéines de compaction génomique, ainsi l'organisation spatiale de leur génome varie selon les espèces. Le chromosome des Euryarchaeota est majoritairement condensé par des histones qui courbent ou enveloppent l'ADN, et par la protéine Alba qui forme des structures en boucle reliant deux duplexes d'ADN. En revanche, le chromosome des Crenarchaeota est majoritairement condensé par des petites protéines similaires aux protéines bactériennes (par exemple, Cren7 et Sul7 chez *Sulfolobus*), et par la protéine Alba (31).

3. Intérêts majeurs de l'étude des Archées

Un changement radical de l'ordre établi

La découverte des Archées a totalement remis en question l'organisation de l'arbre de la vie et l'hypothèse selon laquelle les Bactéries ont précédé et donné naissance aux Eucaryotes au cours de l'évolution. La comparaison des trois grands règnes permet d'établir un portrait-robot du dernier ancêtre commun universel LUCA (*Last Universal Common Ancestor*) à l'origine de l'ensemble des êtres vivants. Cependant, l'organisation des embranchements de l'arbre phylogénétique est encore vivement discuté au sein de la communauté scientifique.

La découverte du superphylum des Asgard est notamment à l'origine d'un débat sur la relation phylogénétique entre les Archées et les Eucaryotes. L'incorporation des Asgard dans les analyses phylogénétiques aurait pour conséquence de fusionner le domaine Archée et Eucaryote au sein d'un même domaine (32). Ces travaux sont néanmoins remis en question et, selon le marqueur évolutif utilisé, deux types d'arbres sont obtenus : un arbre dans lequel Archées et Eucaryotes ne forment qu'un seul domaine et un arbre dans lequel ils forment deux domaines distincts. Cette controverse permet de relancer un débat que certains ont pu croire clos sur l'origine des Eucaryotes.

Un modèle d'étude simplifié

Les Archées forment un groupe d'organismes distincts des Eucaryotes et des Bactéries, et sont donc d'excellents modèles pour comprendre les origines et les mécanismes d'évolution du vivant. En effet, l'étude de leurs génomes et de leurs protéines permet de distinguer les mécanismes communs à tous les types cellulaires de ceux caractéristiques d'une adaptation à des modes de vie particuliers, parfois extrêmes.

Chez les Archées, les processus moléculaires fondamentaux de la division cellulaire, de la maintenance du génome, de la synthèse des protéines, présentent de grandes similarités avec ceux des Eucaryotes. Il n'est pas toujours facile d'étudier les grandes machineries cellulaires présentes dans les cellules humaines tant leur purification et leurs activités sont difficiles à maîtriser au laboratoire. Les systèmes homologues que l'on trouve chez les Archées représentent des versions simplifiées beaucoup plus stables et faciles à produire. Celles-ci se révèlent donc être d'excellents modèles pour la biologie intégrative combinant études *in vivo*, biochimie, biophysique et biologie structurale.

Un large panel de propriétés à exploiter

La très grande diversité des milieux colonisés par les Archées permet de disposer d'un registre d'espèces couvrant des gammes de température, de pH, de salinité, de concentrations en métaux lourds et d'exposition aux rayonnements ionisants extrêmement étendu. Leurs constituants cellulaires possèdent de ce fait des propriétés structurales, de stabilité et fonctionnelles tout à fait exceptionnelles dont seule une partie est aujourd'hui exploitée en biotechnologies. Cette situation évolue progressivement sous la pression croissante de la demande en procédés industriels moins polluants, favorisant l'émergence d'alternatives combinant chimie et biocatalyse. Les enzymes d'Archées, en particulier d'extrêmophiles, constituent un atout majeur pour avancer dans cette direction.

L'étude des Archées met également en évidence leur importance au niveau de la biosphère. La découverte d'Archées effectuant le processus de nitrification laisse entrevoir leur rôle central dans le cycle de l'azote. De même, les méthanogènes pourraient jouer un rôle non négligeable dans les processus d'effet de serre, et dans le cycle du carbone.

Il est intéressant de noter qu'aucune Archée pathogène n'a été découverte pour le moment. Il semble toutefois que certaines pathologies bucco-dentaire et intestinales puissent être liées à des déséquilibres dans le rapport entre certaines Bactéries et Archées du microbiote.

4. Le réplisome des Archées

Le réplisome est un complexe très dynamique et hautement régulé à la fois dans l'espace et dans le temps. Sa composition varie au cours des trois étapes majeures de la réplication de l'ADN : l'initiation, l'élongation et la terminaison.

Pendant le processus de réplication, les deux brins d'ADN génomique, formant la bulle de réplication, servent de matrice pour la synthèse du nouvel ADN. La progression des fourches de réplication est homogène sur les deux brins d'ADN génomique, ce qui implique une duplication simultanée de ces deux brins. Or, les deux brins matrice sont orientés dans des directions opposées : le brin 5'-3' complémentaire du brin 3'-5'. Les ADN polymérases sont, quant à elles, seulement capables de synthétiser l'ADN dans le sens 5'-3' à partir d'une amorce d'ARN étirée par une primase. Par conséquent, seul l'un des deux brins d'ADN est synthétisé de manière continue (le brin « avancé »), alors que l'autre brin (le brin « retardé ») est synthétisé de manière discontinue, par petits fragments. La synthèse du brin avancé lors de la réplication de l'ADN nécessite peu d'évènements d'amorçage par la primase, alors que la synthèse du brin retardé en nécessite plusieurs, générant ainsi plusieurs fragments indépendants : les fragments d'Okazaki.

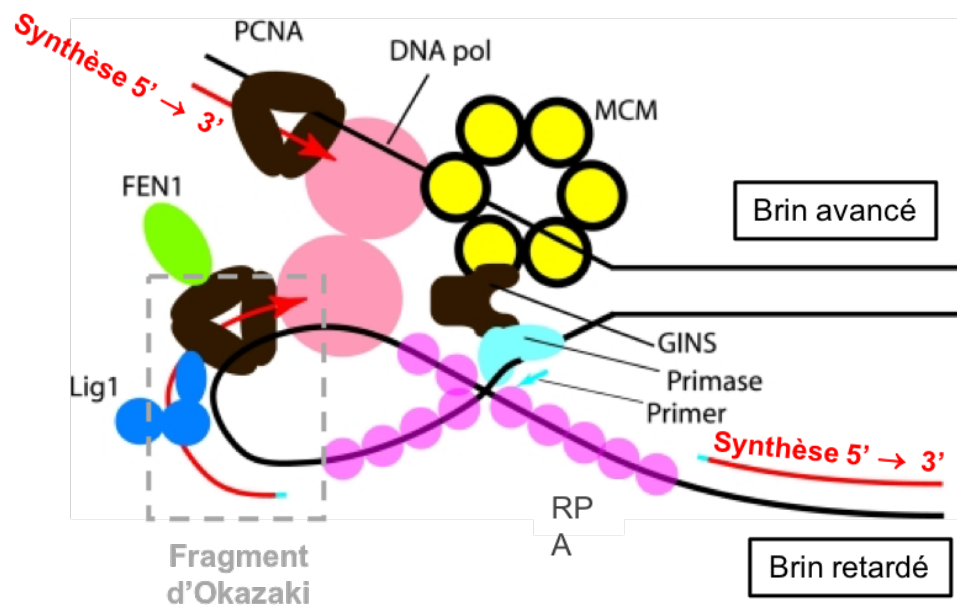


Figure 4. Modèle d'architecture de la fourche de réplication chez les Archées (adaptée de Barry and Bell, 2006) (33). Le brin d'ADN parental est indiqué en noir et le brin nouvellement synthétisé est indiqué en rouge. Les amorces ARN synthétisées par la primase sont indiquées en bleu. L'ADN simple brin est lié aux protéines RPA. L'hélicase MCM interagit avec le complexe de protéines GINS, interagissant lui-même avec la primase. L'ADN polymérase allonge l'amorce et interagit avec le PCNA trimérique. Le PCNA sert de plateforme permettant l'assemblage d'autres acteurs comme Fen1 ou Lig1. Le RFC, facteur de chargement du PCNA, n'est pas représenté ici.

L'organisation de la fourche de réplication chez les Archées est schématisée sur la **Figure 4** tandis que le **Tableau 2** présente une comparaison des réplisomes dans les trois domaines du vivant.

	BACTÉRIES	EUCARYOTES	ARCHÉES	
			EURYARCHAEOTA	CRENARCHAEOTA
Complexe de reconnaissance des origines de réplication				
Protéines de reconnaissance	DnaA	Orc	Orc1/Cdc6	Orc1/Cdc6 et WhiP
Hélicases	DnaB	MCM	MCM	MCM
Chargeurs d'hélicase	DnaC	Cdc6 et Cdt1	Orc1/Cdc6	Orc1/Cdc6
Complexe d'initiation				
Stimulateurs d'activité hélicase	-	Cdc45	Cdc45 ou GAN	Cdc45
	-	GINS	GINS	GINS
Protéines de fixation à l'ADN simple brin	SSB	RPA	RPA	SSB
Complexe d'élongation				
Primases	DnaG	Polα/PriS/PriL	PriS/PriL	PriS/PriL/PriX
Sliding clamps	β et β'	PCNA	PCNA	PCNA
Chargeurs de « pince »	Complexe τ	RFC	RFC	RFC
ADN polymérases répliquatives des brins sens et anti-sens	DnaE et PolC	Pol ϵ et Pol δ (ou Poly mitochondriale)	PolB et PolD	PolB1, PolB2 et PolB3
Maturation des fragments d'Okazaki				
Nucléases	Pol-I	Fen1	Fen1 ou GAN ou RNase H	Fen1 ou autre
ADN polymérases	Pol-I	Pol δ	PolB ou PolD	PolB1
ADN ligases	NAD ⁺ dépendante	ATP dépendante	NAD ⁺ ou ATP dépendante	ATP dépendante

Tableau 2. Comparaison des réplisomes dans les trois domaines du vivant (adapté de Kelman and Kelman, 2014) (34). Les homologies entre les réplisomes d'Eucaryotes et d'Archées sont indiquées en bleu, tandis que les homologies entre les réplisomes de Bactéries et d'Archées sont indiquées en jaune.

Initiation de la réplication

L'étape d'initiation est l'étape durant laquelle les amorces ARN, nécessaires pour amorcer l'activité des ADN polymérase répliquatives, sont synthétisées au niveau des origines de réplication.

- **Origines de réplication**

Comme pour les Bactéries, les origines de réplication de l'ADN chez les Archées ont une structure génomique clairement définie. Elles comportent une région riche en paires de bases AT (la région où la double hélice d'ADN sera dissociée), flanquée de part et d'autre par des éléments de reconnaissance, appelées ORB (*Origin Recognition Box*) sur lesquels se fixent les premières protéines du réplisome (35).

- **Reconnaissance des origines de réplication par Orc1/Cdc6**

Chez la plupart des Archées, les origines de réplication de l'ADN sont reconnues par la protéine Orc1/Cdc6, un homologue de deux protéines eucaryotes : Orc1 et Cdc6.

- **Recrutement de l'hélicase MCM**

Pour que la double hélice d'ADN génomique soit dupliquée par les ADN polymérase répliquatives, les deux brins doivent être dissociés l'un de l'autre, au moins localement. Cette séparation permet à chacun des brins de devenir accessible aux ADN polymérase et de leur servir de matrice. L'ouverture de la double hélice d'ADN génomique est localisée au niveau des origines de réplication et catalysée par l'hélicase MCM (*MiniChromosome Maintenance*) chez les Archées et les Eucaryotes. Chez la plupart des Archées, MCM forme un complexe homohexamérique codé par un seul gène *mcm* (35).

- **Ouverture de la double hélice d'ADN par le complexe CMG**

Le mécanisme moléculaire exact de la séparation des brins d'ADN par l'hélicase MCM au niveau des origines de réplication demeure inconnu. Il est cependant établi que MCM requiert le recrutement des protéines Cdc45 (*Cell Division Control protein 45*) et GINS (*Go-Ichi-Nii-San*) pour former un complexe hélicase actif, appelé CMG (Cdc45-MCM-GINS). Ce complexe CMG est capable d'ouvrir la double hélice d'ADN et de conduire la progression de la fourche de réplication le long du génome (36,37).

- **Protection de l'ADN matrice simple brin par les protéines SSB**

Les protéines SSB (*Single-Stranded DNA Binding protein*) jouent un rôle essentiel dans le réplisome en recouvrant l'ADN simple brin immédiatement après que le complexe CMG ait déroulé la double hélice d'ADN. Les protéines SSB empêchent ainsi la ré-hybridation de la double hélice et protègent l'ADN des attaques nucléasiques. De plus, il a récemment été montré que les protéines SSB stimulaient l'activité des ADN polymérase répliquatives PolB et PolD en empêchant la formation de structures secondaires dans l'ADN qui ralentissent les polymérase (36).

- **Initiation de la synthèse d'ADN par les primases**

Les ADN polymérase répliquatives PolB et PolD sont incapables d'initier la synthèse d'ADN *de novo* et requièrent une amorce ARN synthétisée par les primases. Celles-ci sont recrutées sur les deux brins d'ADN matrice au niveau de l'origine de réplication et synthétisent une amorce ARN de 9 à 16 ribonucléotides sur le brin avancé et plusieurs amorces sur le brin retardé (une pour chaque fragment d'Okazaki) (34).

Progression de la fourche de réplication

A l'issue de l'étape d'initiation, le complexe hélicase conduit la propagation de la fourche de réplication le long du génome. Durant cette étape d'élongation, les amorces ARN sont allongées par les ADN polymérase répliquatives (PolB et PolD) qui dupliquent ainsi l'ADN génomique.

- **Chargement du facteur de processivité PCNA par le complexe RFC**

Le facteur de processivité PCNA (*Proliferating Cell Nuclear Antigen*) joue un rôle central dans de nombreux mécanismes cellulaires et particulièrement la réplication de l'ADN. Il constitue une véritable plateforme d'interaction avec un grand nombre d'acteurs protéiques de la fourche de réplication. La fonction principale du PCNA est de recruter les ADN polymérase répliquatives PolB et PolD au niveau des fourches de réplication et d'augmenter leur capacité à allonger l'ADN sans se dissocier du brin matrice à l'aide de sa structure en anneau qui encercle l'ADN.

Dans la majorité des systèmes étudiés, le PCNA est incapable de s'assembler de manière efficace autour de l'ADN sans l'aide du facteur de chargement RFC. Il a été montré que l'activité de polymérisation des ADN polymérase PolB et PolD était stimulée en présence de PCNA et d'avantage stimulée en présence du complexe RFC-PCNA (38).

- **Maturation des fragments d'Okazaki**

Les amorces ARN qui initient la synthèse de chaque fragment d'Okazaki doivent être éliminées afin de former un brin retardé continu, composé exclusivement d'ADN. Ce processus nécessite l'action conjointe de plusieurs enzymes du réplisome : les ADN polymérases répliquatives PolB et PolD, les nucléases FEN1 et RNaseH et l'ADN ligase LIG1 (34).

Chez la plupart des Archées, le mécanisme moléculaire de la maturation des fragments d'Okazaki n'est pas compris. Pour cause, la fonction biologique des ADN polymérases répliquatives PolB3 et PolD au sein de la fourche de réplication n'est pas claire et les caractéristiques fonctionnelles de la PolD varient selon les espèces considérées.

II. La PolD : une ADN polymérase atypique d'archée

1. Classification structurale des ADN polymérases

Les ADN polymérases sont des enzymes clés impliquées dans la réplication de l'ADN au cours du cycle cellulaire, mais aussi dans des processus de réparation et de recombinaison de l'ADN.

Diversité des ADN polymérases répliquatives dans les trois domaines du vivant

En se basant sur leurs séquences protéiques et sur des analyses structurales, les ADN polymérases ont pu être classées en sept familles : PolA, PolB, PolC, PolD, PolX, PolY et RT (*Reverse Transcriptase*) (5–7) (**Tableau 3**).

	BACTÉRIES	ARCHÉES	EUCARYOTES	VIRUS
A	Pol-I		Pol γ , Pol θ , Pol ω	T3, T5, T7
B	Pol-II	PolB1, B2, B3	Pol δ , Pol ϵ , Pol ζ	T4, T6, ϕ 29, RB69
C	Pol-III			PolC
D		PolD		
X	PolX		Pol λ , Pol β , Pol μ , TdT	ASFV
Y	Pol IV, V	Dpo4, ...		
RT			téломérase	<i>reverse transcriptase</i>

Tableau 3. Classification des ADN polymérases en sept familles distinctes. La TdT est une *terminal deoxynucleotidyl transferase*. Dpo4 est un exemple de polymérase de la famille Y appartenant à l'Archée *Sulfolobus solfataricus*.

Bien que le principe de la réplication de l'ADN soit conservé dans tout le vivant, aucune ADN polymérase n'est universelle. Quatre familles différentes d'ADN polymérases ont évolué pour répliquer l'information génétique dans les trois domaines du vivant : les familles A, B, C et D. Chez les Bactéries, le génome est répliqué par des ADN polymérases appartenant aux familles A et C. Chez les Eucaryotes, le génome nucléaire est répliqué par trois ADN polymérases répliquatives distinctes, toutes appartenant à la famille B, alors que le génome mitochondrial est répliqué principalement par une ADN polymérase de famille A. Enfin, chez les Archées, la situation est plus complexe et les ADN polymérases répliquatives ne sont pas caractérisées avec autant de certitudes que les deux autres domaines du vivant. Les Archées contiennent à la fois des ADN polymérases B, homologues à celles des Eucaryotes, et une ADN polymérase répliquative appartenant à la famille D appelée PolD (**Figure 5**).

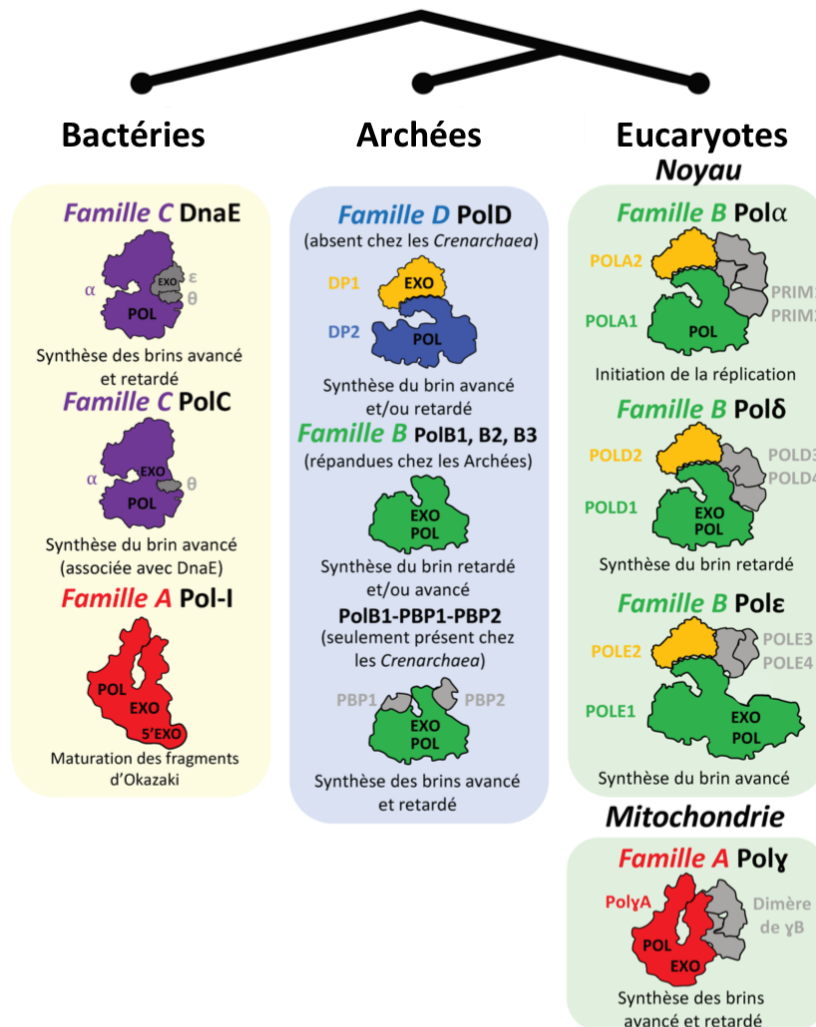


Figure 5. Distribution des ADN polymérases répliquatives dans les trois domaines du vivant (extraite de Raia *et al*, 2019b) (15). Représentation schématisée de la composition en sous-unités des polymérases répliquatives dans les trois domaines du vivant. Les sites actifs polymérase et exonucléase sont respectivement indiqués par « POL » et « EXO ». Les sous-unités catalytiques eucaryotes sont représentées en vert, les sous-unités régulatrices en jaune et les sous-unités accessoires en gris.

En déterminant la structure de cette PolD, mon laboratoire d'accueil a mis en évidence une polymérase atypique qui étend le répertoire des repliements connus pour leur implication dans la réplication de l'ADN. Ces travaux ont permis de mettre à jour la classification structurale des ADN polymérases en faisant un état des lieux de la diversité rencontrée dans le monde du vivant pour répliquer le génome (13–15).

Classification structurale des cœurs catalytiques des ADN polymérases

Avant la détermination de la structure de la PolD, les ADN polymérases étaient divisées en deux super-groupes basés sur le repliement structural de leur cœur catalytique (39) (Figure 6). Le premier groupe contient les ADN polymérases de type Klenow (40), similaires aux ADN polymérases répliquatives Pol-I d'Escherichia coli, regroupant les familles de polymérase A, B, Y et RT. Le second groupe contient les ADN polymérases de type Pol β , incluant les familles de polymérase C et X (41). La résolution de la structure de la PolD a mis en évidence une troisième classe de cœurs catalytiques capables de polymériser des désoxyribonucléotides (Figure 6). Le cœur catalytique de la PolD adopte un repliement en deux tonneaux- β , qui se distingue des cœurs catalytiques de type Klenow et Pol β , retrouvés, sans exception, chez toutes les autres familles d'ADN polymérases.

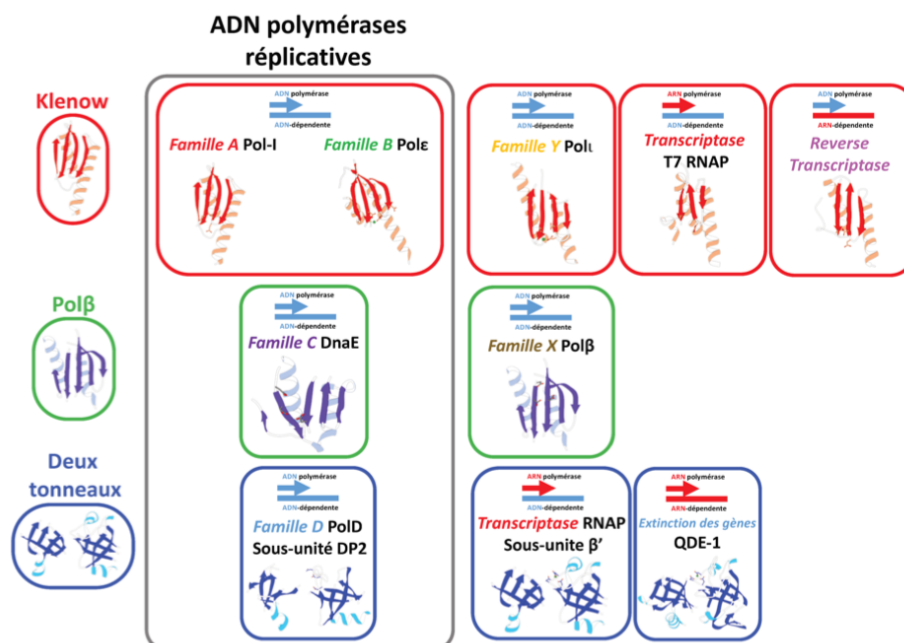


Figure 6. Classification structurale mise à jour des cœurs catalytiques des ADN polymérases (extraite de Raia *et al.*, 2019b) (15).

Classification structurale des domaines de relecture des ADN polymérase

Il existe deux cas de figure au sein des ADN polymérase répliquatives : le domaine exonucléase est inséré au sein de la sous-unité catalytique, c'est notamment le cas des ADN polymérase de famille A (Pol-I) et de famille B (Pol α , Pol δ , Pol ϵ , PolB1, PolB2 et PolB3), ou le domaine exonucléase est une enzyme indépendante et c'est le cas des ADN polymérase de famille C (Pol-II) et de famille D (PolD).

Les ADN polymérase répliquatives ont fait évoluer trois classes structurales distinctes de domaine de relecture exonucléase : le repliement exonucléase DnaQ, qui est largement exprimé dans toutes les formes de vie et associé aux ADN polymérase répliquatives de famille A, B et C, le repliement exonucléase PHP (*Polymerase and Histidinol Phosphatase*), qui est exclusivement associé aux ADN polymérase de type Pol β , et le repliement exonucléase PDE (*Phosphodiesterase*), qui est exclusivement associé à la PolD dans sa forme active (**Figure 7**). Les domaines PHP et PDE sont deux exemples de repliements exonucléase ancestraux retrouvés sous forme inactive chez certaines ADN polymérase répliquatives contemporaines.

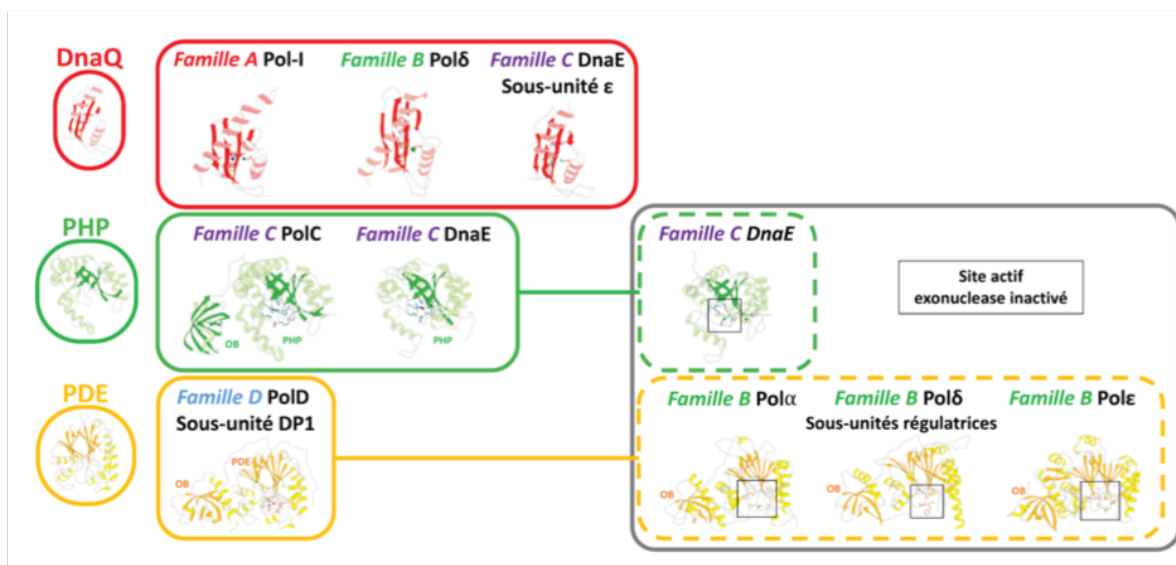


Figure 7. Classification structurale mise à jour des sites exonucléase associés aux ADN polymérase répliquatives (extraite de Raia *et al*, 2019b) (15).

2. L'ADN polymérase de la famille D (PolD)

La PolD a été isolée il y a une vingtaine d'année par le groupe du Professeur Yoshizumi Ishino au Japon à partir d'archées extrémophiles du genre *Pyrococcus*. Il s'agit d'un enzyme hétérodimérique composée de deux sous-unité appelées DP1 et DP2 (12). Une caractérisation biochimique a montré que la petite sous-unité DP1 porte la fonction correctrice d'erreurs 3'-5'

exonucléase tandis que la grande sous-unité DP2 est responsable de l'activité polymérase 5'-3' (42–45). L'interaction entre DP1 et DP2 est nécessaire au fonctionnement optimal de la PolD, stimulant à la fois les activités polymérase et exonucléase (44).

Après la découverte de la PolD, deux décennies de caractérisation biochimique, d'analyses fonctionnelles *in vitro* et d'études génétiques et structurales ont contribué à montrer que la PolD joue un rôle essentiel dans la réplication de l'ADN chez les Archées et constitue une classe atypique d'ADN polymérases. On la retrouve chez toutes les archées, sauf dans le phylum des Crénarchées.

Avant les travaux réalisés au laboratoire durant ces dernières années, les données structurales sur la PolD se limitaient à une structure cristallographique des 200 premiers résidus de la sous-unité DP2 (46) et une structure RMN des 70 premiers résidus de la sous-unité DP1 de *Pyrococcus horikoshii* (47). Cependant, aucune donnée n'était disponible sur les domaines les plus intéressants : le domaine polymérase et le domaine exonucléase.

3. Étude structurale de la PolD de *Pyrococcus abyssi*

Afin de mieux comprendre les origines évolutives de la PolD, mon laboratoire d'accueil a ouvert un axe de recherche sur la détermination de sa structure. Ils ont choisi comme modèle d'étude la PolD de *Pyrococcus abyssi* pour deux raisons. En premier lieu pour son potentiel biotechnologique largement inexploré mais aussi parce qu'elle fait l'objet d'une collaboration historique avec l'équipe du Docteur Ghislaine Henneke à l'IFREMER de Brest.

Les structures des deux sous-unités DP1 et DP2 ont pu être déterminées séparément à des résolutions respectives de 2,5 Å et 2,2 Å (13) (Figure 8). La construction de DP1 utilisée inclut les acides aminés 144 à 622 de la petite sous-unité, et ne prend pas en compte une région N-terminale flexible non conservée et non nécessaire pour l'activité exonucléase. La construction de DP2 utilisée inclut les acides aminés 1 à 1050 de la grande sous-unité, et ne prend pas en compte un domaine C-terminal (1051-1270) responsable de l'interaction avec DP1.

Plus récemment, la structure de la PolD entière en complexe avec l'ADN a été déterminée en cryo-microscopie électronique au laboratoire à une résolution de 7,1 Å. Ce travail a permis d'aller encore plus loin dans la caractérisation et la classification de la PolD (Figure 9) (14,15).

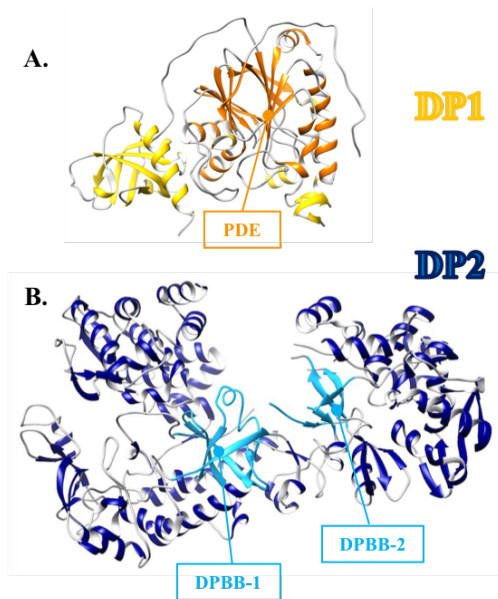


Figure 8. Structure des sous-unités DP1 (PDBid : 5IHE) et DP2 (PDBid : 5IJL) de la PolD de *Pyrococcus abyssi* obtenue par cristallographie aux rayons X (13). Les proportions et la position des deux sous-unités l'une par rapport à l'autre ne sont pas respectées. (A) Petite sous-unité DP1 de la PolD. Le domaine phosphodiesterase (PDE) responsable de l'activité exonucléase est représenté en orange. (B) Grande sous-unité DP2 de la PolD portant l'activité polymérase. Les deux tonneaux β , appelés DPBB1 (*double-psi β -barrel 1*) et DPBB2 (*double-psi β -barrel 2*) sont représentés en bleu clair.

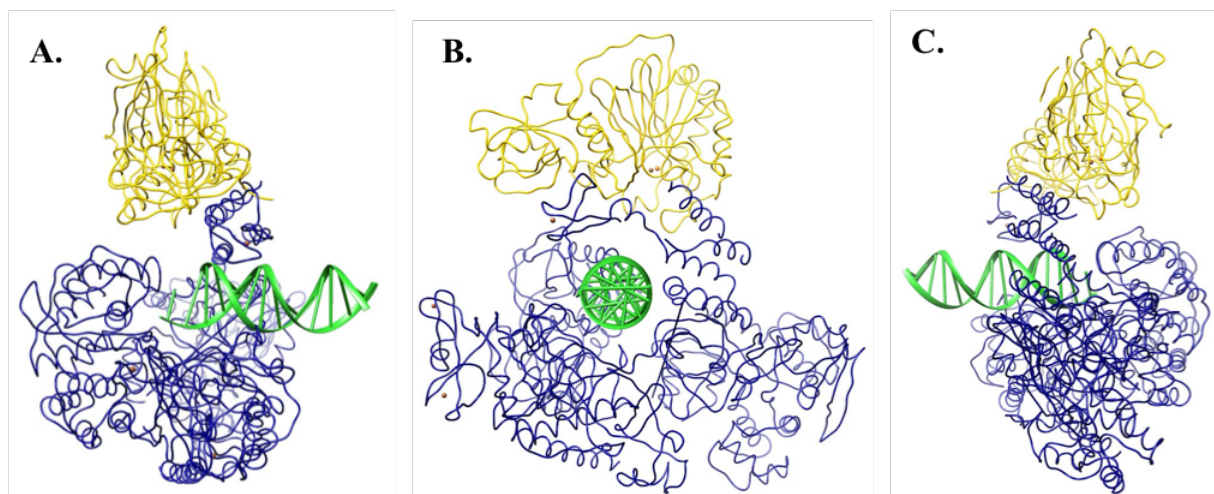


Figure 9. Structure de la PolD entière de *Pyrococcus abyssi* en complexe avec l'ADN obtenue par cryo-microscopie électronique (PDBid : 6HMS) (14). (A) Vue latérale du complexe. La petite sous-unité à l'activité exonucléase DP1 de la PolD est colorée en jaune et la grande sous-unité à l'activité polymérase DP2 de la PolD est colorée en bleu. L'ADN est coloré en vert. (B) Vue de face du même complexe. (C) Vue latérale du même complexe.

4. Scénario de l’histoire évolutive des ADN polymérases répliquatives eucaryotes

Aucune famille d’ADN polymérase répliquative ne s’est imposée de manière universelle dans les trois domaines du vivant. Il apparaît désormais plus clairement que la machinerie de répllication de l’ADN a évolué indépendamment chez les Bactéries d’une part et chez les Archées/Eucaryotes d’autre part (48). Ainsi, les ADN polymérases répliquatives d’Archées (PolB et PolD) partagent une histoire évolutive commune avec les ADN polymérases répliquatives des Eucaryotes (Pol α , Pol δ et Pol ϵ de famille B) (49–51). La structure de la PolD résolue au laboratoire pendant la thèse du Docteur Raia a permis de proposer un scénario évolutif expliquant le lien entre les ADN polymérases répliquatives d’Archées et les polymérases répliquatives eucaryotes.

Les Pol α , Pol δ et Pol ϵ sont composées d’une sous-unité catalytique, appartenant à la famille B des ADN polymérases, et une sous-unité régulatrice, appartenant à la famille des phosphodiesterase (PDE). L’étude structurale menée sur la PolD a montré que DP1 était l’ancêtre des sous-unités régulatrices eucaryotes et que les régions C-terminales des sous-unités catalytiques des polymérases eucaryotes Pol α , Pol δ et Pol ϵ , et la région C-terminale de la sous-unité DP2 de la PolD adoptent un repliement similaire qui leur permet respectivement d’interagir avec les sous-unités régulatrices B et leur sous-unité DP1 (**Figure 10 A**).

Cette analyse structurale a permis de proposer le scénario évolutif suivant : la sous-unité catalytique DP2 de la PolD, composée d’un cœur catalytique à deux tonneaux β , aurait été remplacée au cours de l’évolution par une sous-unité catalytique appartenant à la famille B des ADN polymérases, composée d’un domaine polymérase de type Klenow et fusionnée à un domaine exonucléase de type DnaQ (**Figure 10 B**). Par la suite, le complexe entre le domaine C-terminal et la sous-unité régulatrice chez les Eucaryotes aurait perdu la fonction de relecture exonucléase 3’-5’ et la capacité de fixer l’ADN. Chez les Pol δ et Pol ϵ , la fonction de relecture est assurée par le domaine exonucléase DnaQ, fusionné au domaine polymérase de la sous-unité catalytique de famille B. La sous-unité régulatrice a cependant été conservée au cours de l’évolution, pour stabiliser la structure de la polymérase et interagir avec les partenaires de la fourche de répllication (14).

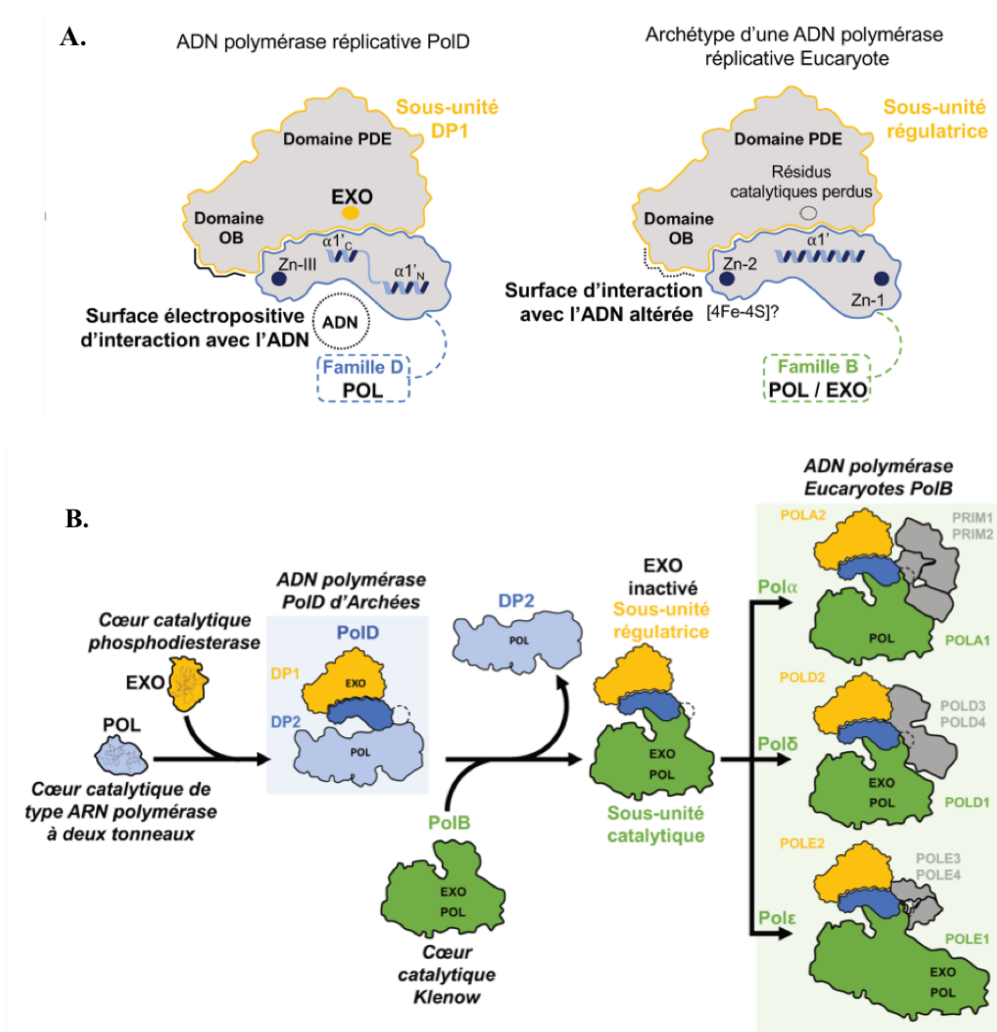


Figure 10. Relation évolutive hypothétique entre les PolB et PolD des Archées et les ADN polymérases répliquatives eucaryotes (extraite de Raia *et al*, 2019a) (14). (A) Représentation schématique de l'interface DP1/DP2 de la PolD et de l'interface entre les sous-unités catalytique et régulatrice des ADN polymérases répliquatives eucaryotes, mettant en évidence les différences entre les deux structures. EXO : site actif exonucléase ; OB : *Oligonucleotide Binding domain* ; PDE : domaine phosphodiesterase. (B) Relation évolutive complexe entre la PolD et les PolB eucaryotes, impliquant des substitutions et inactivations de domaines.

5. Potentiel biotechnologique de la PolD

Les ADN polymérases thermostables sont couramment utilisées en PCR

Au-delà de leurs fonctions biologiques essentielles, les ADN polymérases sont couramment utilisées en biologie moléculaire. Une des techniques les plus connues et les plus anciennes est la PCR (*Polymerase Chain Reaction*). Celle-ci permet de dupliquer en grand nombre une séquence d'ADN ou d'ARN connue, à partir d'une faible quantité d'acide nucléique

et d'amorces spécifiques constituées d'oligonucléotides de synthèse. En PCR, l'amplification de l'ADN est catalysée par des ADN polymérase thermostables capables de résister aux variations de températures induites par les cycles PCR. Il existe un large catalogue commercial de polymérase thermostables différentes qui possèdent toutes des caractéristiques distinctes de thermostabilité, de processivité, de fidélité, de spécificité, de sélectivité des nucléotides, et de résistance aux contaminants ou aux inhibiteurs (52). Il n'existe cependant aucune ADN polymérase ultime, le choix de l'enzyme adéquate dépend de l'application souhaitée. En effet, il existe un grand nombre de PCR différentes, avec des buts distincts, par exemple : des PCR à haut rendement, à haute-fidélité, sur colonies bactériennes, avec des nucléotides modifiés, ou sur des génomes entiers.

Aujourd'hui, toutes les polymérase thermostables commercialisées appartiennent invariablement aux familles d'ADN polymérase A, B et Y. Les polymérase de famille A proviennent principalement de Bactéries thermophiles ou hyperthermophiles, telle que la polymérase répliquatives Pol-I de *Thermus aquaticus* (Taq). Les polymérase de famille B proviennent d'Archées hyperthermophiles, telle que la polymérase répliquatives PolB3 de *Pyrococcus furiosus* (Pfu). Enfin, les polymérase de famille Y proviennent d'Archées hyperthermophiles, telle que la PolY de *Sulfolobus solfataricus* (Sso) (53).

Pour répondre aux besoins de plus en plus spécifiques des nouvelles techniques de PCR, les formes natives des ADN polymérase thermostables ne suffisent pas. Par exemple, le diagnostic de maladies génétiques ou infectieuses nécessite des ADN polymérase capables d'amplifier directement l'ADN contenu dans des échantillons cliniques tels que le sang ou les fluides corporels. De nouvelles polymérase sont alors continuellement modifiées par ingénierie enzymatique pour répondre à ces besoins.

La PolD possède un grand potentiel biotechnologique

La PolD possède des propriétés de processivité, de spécificité et de thermostabilité qui lui confèrent un potentiel biotechnologique important, exploitable en PCR (Killelea et al. 2014). La PolD possède également des propriétés biochimiques qui la distingue des polymérase conventionnelles commercialisées à ce jour. Contrairement aux polymérase « Taq » bactériennes, la PolD est capable d'étendre une amorce contenant jusqu'à deux mésappariements à l'extrémité 3'. Elle est également davantage résistante aux cycles successifs de dénaturation à haute température induite par les cycles successifs de PCR et montre une résistance supérieure aux inhibiteurs classiques (comme les ions NaCl et CaCl₂ ou le détergent SDS) et une résistance totale à l'hémoglobine. La PolD de *Pyrococcus abyssi* présente la plus haute tolérance aux ions calcium Ca²⁺ par rapport aux autres ADN polymérase thermostables

commercialisées, plaçant ainsi la PolD comme une enzyme de choix dans l'amplification d'ADN présent dans des échantillons riches en Ca^{2+} , tels que les produits alimentaires (lait, fromage, ...) et des échantillons humains (dents, os, ...) (54).

III. Le PCNA : un acteur clé de la fourche de réplication

1. Fonctions biologiques et diversité structurale des facteurs de processivité

Chez les Bactéries, les Eucaryotes et les Archées, l'efficacité de la synthèse de l'ADN est intimement liée à la présence de protéines appelées facteurs de processivité, qui stimulent considérablement la processivité des ADN polymérases répliquatives et sont indispensables pour dupliquer de grands génomes. Ce sont généralement des protéines oligomériques en forme d'anneau, capables d'encercler une double hélice d'ADN. Une fois chargées sur l'ADN, elles peuvent recruter une ADN polymérase et se déplacer le long du génome sans se dissocier comme des « pinces coulissantes » (*sliding clamps*).

Dans les trois domaines du vivant, il existe deux principales familles de facteurs de processivité associés aux ADN polymérases répliquatives : le clamp $\beta 2$, associé aux Pol-III et Pol-I bactériennes et le PCNA (*Proliferating Cell Nuclear Antigen*), associé aux Pol δ et Pol ϵ eucaryotes et aux PolD et PolB des Archées. Les facteurs de processivité possèdent tous une structure circulaire avec une cavité centrale suffisamment large pour accueillir une double hélice d'ADN. La plupart des facteurs de processivité sont des homomultimères : le PCNA est un homotrimère et le clamp $\beta 2$ est un homodimère (55).

Le clamp $\beta 2$ a la forme d'un anneau hexagonal composé de deux sous-unités identiques β . Chaque sous-unité β est composée d'une répétition de trois domaines similaires, formant ainsi un anneau avec une pseudo-symétrie d'ordre six.

Chez les Eucaryotes et les Archées, la structure du PCNA est très similaire à celle du clamp $\beta 2$ bactérien. La principale différence entre le PCNA et le clamp $\beta 2$ réside dans le fait que chaque sous-unité du PCNA est composée de seulement deux domaines, au lieu de trois, et que le PCNA trimérise pour former un anneau avec une pseudo-symétrie d'ordre six, semblable à celle du clamp $\beta 2$ (56). Chez certaines espèces de Crenarchées, le PCNA est constitué de trois sous-unités distinctes : PCNA1, PCNA2 et PCNA3, chacune interagissant avec des protéines différentes de la fourche de réplication. L'architecture hétérotrimérique du PCNA est spécifique des Crenarchées (57) (**Figure 11**).

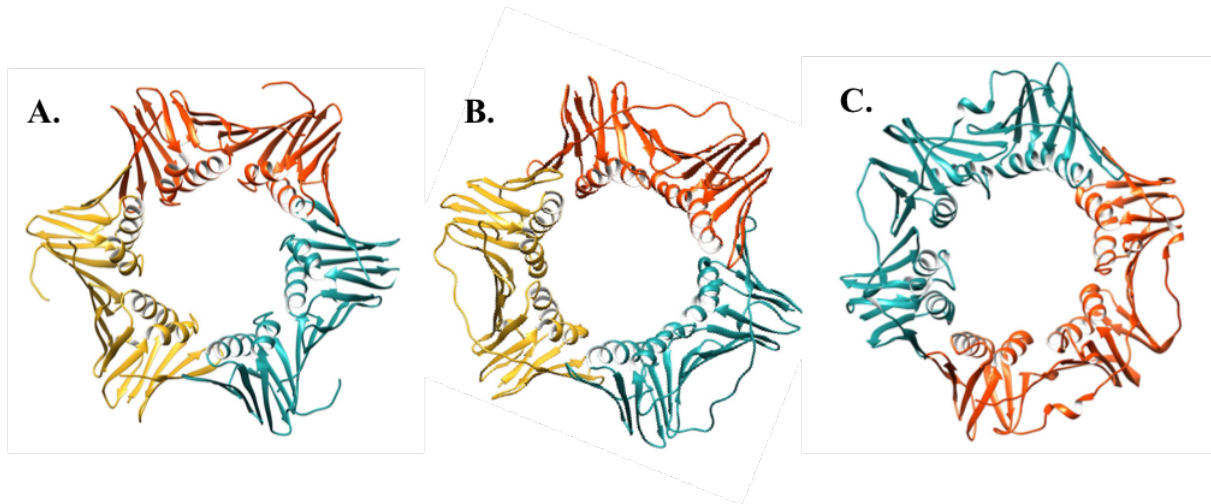


Figure 11. Structure des facteurs de processivité dans les trois domaines du vivant. (A) Structure cristallographique du PCNA trimérique de *Pyrococcus furiosus* (PDBid : 1GE8) (58) ; (B) Structure cristallographique du PCNA trimérique de *Sacharomyces cerevisiae* (PDBid : 1PLR) (56) ; (C) Structure cristallographique de l’analogue bactérien du PCNA, la sous-unité β dimérique de *Escherichia coli* (PDBid : 2POL) (59).

2. État de l’art concernant les interactions entre la PolD et le PCNA

Le plus souvent, les ADN polymérases (ainsi que d’autres partenaires de la fourche de réplication) interagissent avec le facteur de processivité via un motif conservé localisé à l’extrémité N- ou C-terminale des sous-unités catalytiques : le motif QL[S/D]LF interagit avec le clamp $\beta 2$ et le motif Qxx[I/L/M/V]xx[F/Y][F/Y] (aussi appelé motif PIP) interagit avec le PCNA, où « x » représente tout type d’acide aminé (55,60). Cependant, de nouveaux sites d’interaction ont été mis en évidence au-delà du motif PIP (61).

Des études biochimiques ont montré que le PCNA augmentait considérablement les performances de la PolD de *P. abyssi in vitro* (62,63).

Des expériences en immunoprécipitation ont montré que le PCNA de *Pyrococcus furiosus* semblait interagir principalement avec la grande sous-unité DP2 de la PolD de *P. furiosus* et de façon très faible avec la petite sous-unité DP1 (64), résultats appuyés quelques années plus tard par les travaux d’une autre équipe en résonance plasmonique de surface (SPR) (65). De nouvelles données, obtenues avec la même technique, avait entre-temps montré que le PCNA de *Pyrococcus horikoshii* semblait au contraire interagir exclusivement avec la petite sous-unité DP1 de la PolD de *P. horikoshii*, mais en aucun cas avec la grande sous-unité DP2 (66).

En revanche, aucune interaction directe n'a été observée en SPR entre le PCNA de *P. abyssi* et la PolD de *P. abyssi* (62), suggérant que l'interaction PCNA-PolD était ADN-dépendante entre le PCNA et la PolD. Un résumé des études concernant l'interaction entre la PolD et le PCNA est présenté dans le **Tableau 4**.

À ce jour, il n'existe aucune structure à haute résolution d'un complexe entre une ADN polymérase répliquative et son facteur de processivité. Il existe cependant un modèle de la Pol-III d'*Escherichia coli* en complexe avec le clamp $\beta 2$ et l'ADN, à 7,3 Å de résolution, déterminé par cryo-microscopie électronique (67) ainsi qu'un modèle de la PolB3 de *Pyrococcus furiosus* en complexe avec le PCNA et l'ADN, à 19 Å de résolution, construit par microscopie à coloration négative (68) (**Figure 12**). Au cours de mes travaux, j'ai étudié l'interaction entre la PolD et le PCNA de *Pyrococcus abyssi* afin d'identifier les régions importantes et d'être en mesure de réaliser une construction rationalisée du complexe que ces deux protéines forment avec l'ADN. Cette étude a pour objectif principal d'ouvrir la voie à l'ingénierie enzymatique rationnelle visant à améliorer les performances de ce complexe, notamment en conditions de PCR.

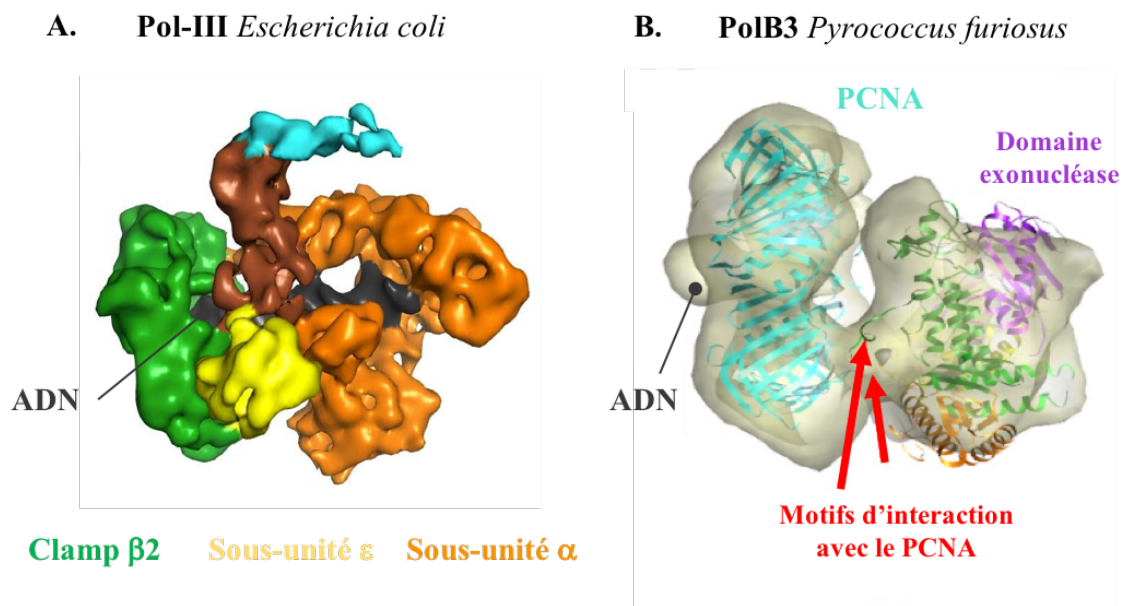


Figure 12. Structure des ADN polymérases répliquatives Pol-III et PolB3 en complexe avec leurs facteurs de processivité respectifs, $\beta 2$ et PCNA (extraites de Fernandez-Leiro *et al*, 2015 et Mayanagi *et al*, 2011) (67,68). (A) Carte de densité de microscopie électronique de l'ADN polymérase bactérienne Pol-III en complexe avec l'ADN et le clamp $\beta 2$. (B) Carte de densité de microscopie électronique en coloration négative de l'ADN polymérase d'Archée PolB3 en complexe avec l'ADN et le PCNA.

Date	Résultats	Organismes	Références
1995 - 1998	Découverte d'une nouvelle famille d'ADN polymérase répliquatives chez les archées : PoID	<i>P. furiosus</i>	Imamura <i>et al</i> , 1995 ; Uemori <i>et al</i> , 1997 ; Ishino <i>et al</i> , 1998 ; Cann <i>et al</i> , 1998
1999	Le PCNA stimule l'activité de la PoID in vitro . Etudes d'immunoprécipitation mettent en évidence une interaction entre la PoID et le PCNA impliquant DP2 mais pas DP1.	<i>P. furiosus</i>	Cann <i>et al</i> , 1999
2004	Double-hybride : Interaction PCNA-DP1 abolie par la délétion des 200 résidus N-terminaux de DP1. SPR avec immobilisation des 200 résidus N-terminaux de DP1 sur une puce senseur CM5 : Interaction entre le PCNA et la région N-terminale de DP1	<i>P. horikoshii</i>	Tang <i>et al</i> , 2004
2005	Caractérisation de la PoID de <i>Pyrococcus abyssi</i> . PCNA stimule son activité <i>in vitro</i> . SPR : Interaction PoID-PCNA dépendante de la présence d'ADN .	<i>P. abyssi</i>	Henneke <i>et al</i> , 2005
2007	Tests d'activité <i>in vitro</i> et SPR : identification d'un motif d'interaction (PIP-motif) à l'extrémité C-terminale de DP2 (avec PCNA immobilisé sur une puce senseur CM5, KD = 76,4 nM). Délétion des 11 derniers résidus C-terminaux de DP2 altère l'interaction PoID-PCNA et la stimulation de l'activité de la PoID par le PCNA.	<i>P. furiosus</i>	Tori <i>et al</i> , 2007
2009	SPR avec immobilisation de l'ADN sur une puce senseur SA : Délétion du PIP-motif C-terminal de DP2 ne suffit pas à abolir l'interaction → multiples sites d'interaction entre le PCNA et la PoID ?	<i>P. abyssi</i>	Castrec <i>et al</i> , 2009

Tableau 4. État de l'art concernant la mise en évidence et la caractérisation des interactions entre le PCNA et la PoID.

MATÉRIEL ET MÉTHODES

I. Expression des protéines recombinantes

1. Constructions plasmidiques

Les caractéristiques des différentes constructions plasmidiques utilisées au cours de ces travaux sont détaillées dans le **Tableau 5**.

Enzymes	Dénomination des constructions	Poids moléculaire (kDa)	Étiquette poly-histidines	Sites de clivage de l'étiquette	Vecteur	Sites de clonage utilisés
PCNA	PCNA	28,4	6-His, N-ter	X	pQE80L	BamH1-Not1
PoID	PoID	213,9	14-His, N-ter	TEV	pRSF-duet	BamH1-Not1 & Nde1-Xho1
	DP1	69,2	14-His, N-ter	TEV	pRSF-duet	BamH1-Not1
	DP1-delNTD	54,3	14-His, N-ter	TEV	pRSF-duet	BamH1-Not1
	DP1-NTD	9,3	14-His, N-ter	TEV	pRSF-duet	BamH1-Not1
	DP2 catalytique	120,6	14-His, N-ter	TEV	pRSF-duet	BamH1-Not1
	DP1-delNTD + DP2-CTD long	69,4	14-His, N-ter	TEV	pRSF-duet	BamH1-Not1 & Nde1-Xho1
	DP1 + DP2-CTD court	84,6	14-His, N-ter	TEV	pRSF-duet	BamH1-Not1 & Nde1-Xho1

Tableau 5. Caractéristiques des constructions plasmidiques utilisées.

2. Souches bactériennes d'expression

La souche bactérienne d'expression utilisée est la souche *Escherichia coli* BL21-AI qui contient le plasmide RIPL. Ce plasmide surexprime les ARN de transfert correspondant aux codons rares de *E. coli*, et qui sont régulièrement présents dans les gènes de *Pyrococcus abyssi*.

3. Transformation bactérienne par choc thermique

Cette étape permet d'introduire les gènes d'intérêt chez la bactérie. N'étant cependant pas complètement efficace, il faut sélectionner les bactéries qui ont été transformées en utilisant un gène de résistance antibiotique porté exclusivement par nos plasmides. Ici, les bactéries

ayant intégré nos constructions plasmidiques deviennent résistantes à la kanamycine et au chloramphénicol.

Ainsi, les bactéries *E. coli* BL21-AI ont été transformées séparément avec chacune des constructions. La transformation est réalisée en mélangeant 1 µL de plasmide et 100 µL de bactéries compétentes. Après incubation dans la glace 20 minutes, le mélange est soumis à un choc thermique de 30 secondes à 42°C suivi de 2 minutes dans la glace. Les cellules sont ensuite incubées 1h à 37°C en milieu SOC avant d'être étalées sur des boîtes LBKC (un milieu Luria-Bertani supplémenté avec 50 µg/mL de kanamycine et de chloramphénicol) et incubées 16 heures à 37°C.

4. Induction de l'expression des protéines

La pré-culture est réalisée en milieu liquide LBKC à partir d'une colonie prélevée dans la boîte obtenue au point précédent et incubée 16h à 37°C jusqu'à obtenir une densité optique de 3. Un volume de pré-culture est ensuite inoculé dans 1,8 L de LBKC de façon à ce que la densité optique à 600 nm soit de 0,05 et incubé à 37°C. L'induction de l'expression des protéines est réalisée en ajoutant 0,5 mM d'iso-propylthiogalapyranoside (IPTG) et 1mM d'arabinose 20% lorsque la densité optique de la culture atteint des valeurs comprises entre 0,7 et 0,8. La durée de l'induction est de 20h à 20°C.

L'IPTG est un analogue de l'allolactose, un métabolite du lactose qui active la transcription de l'opéron lactose et en particulier du gène de la β-galactosidase, lacZ. Comme l'allolactose ou le lactose, l'IPTG se lie au répresseur de l'opéron lactose, LacI, ce qui empêche sa liaison à l'opérateur et induit donc la transcription du gène de la β-galactosidase. L'IPTG peut être utilisé dans le cas présent pour induire l'expression de nos protéines recombinantes placées sous le contrôle du promoteur lac dans *E. coli*.

II. Purification des protéines

Les différentes constructions ont pu être isolées en suivant un protocole de purification qui consiste en quatre étapes chromatographiques et une digestion par une protéase. Les protocoles de purification utilisés sont identiques pour toutes les constructions et les tampons utilisés sont présentés dans le **Tableau 6**.

Tampon	Composition
A	300 mM NaCl ; 10 mM imidazole ; 50 mM HEPES pH 8
B	300 mM NaCl ; 500 mM imidazole ; 50 mM HEPES pH 8
C	50 mM HEPES pH 8
D	50 mM NaCl ; 50 mM HEPES pH 8
E	2 M NaCl ; 50 mM HEPES pH 8
F	200 mM NaCl ; 20 mM HEPES pH 8

Tableau 6. Tampons utilisés au cours des purifications. Les tampons A et B sont utilisés lors de l'étape de chromatographie de pseudo-affinité sur colonne de nickel pendant laquelle l'élution est assurée par un gradient d'imidazole. Le tampon C est utilisé afin de diluer les sels présents dans le tampon à l'issue de l'élution sur la colonne de nickel. Les tampons D et E sont utilisés lors de l'élution sur colonne héparine. Cette dernière est une échangeuse de cations, l'élution se fait par gradient salin. Le tampon F est utilisé lors de l'étape sur colonne de gel filtration.

1. Lyse cellulaire

La première étape est la lyse des cellules bactériennes, qui permet de récupérer les protéines solubles exprimées dans le cytoplasme. La lyse est réalisée sous 1,35 kbar pression à l'aide d'un disrupteur de cellules (TS series). Un cocktail d'inhibiteurs de protéases est ajouté et l'ADN est dégradé par l'ajout de 2 μ l de benzonase. Les constructions utilisées étant thermostables, chauffer l'extrait brut 10 minutes à 60°C permet de précipiter sélectivement les protéines endogènes. Une centrifugation à 35 000 xg pendant 30 minutes permet de séparer le matériel soluble du matériel insoluble. Le surnageant de lyse est ensuite filtré (0,8 μ m) afin de ne pas colmater les colonnes de chromatographie.

2. Chromatographies

Chromatographie de pseudo-affinité sur colonne de nickel

Le principe consiste à fixer de manière covalente un ligand sur un support solide. Lorsqu'on place un mélange de macromolécules en présence de ce support fonctionnalisé, celles reconnaissant le ligand se fixent sélectivement à sa surface, ce qui permet de les séparer des autres après lavage par élution avec un tampon adéquat. Ici, tout le principe repose sur les interactions ioniques qui se forment entre les ions nickel et plusieurs noyaux imidazoles de résidus histidines. L'élution de la protéine est ensuite réalisée par un mécanisme compétitif en ajoutant des quantités croissantes d'imidazole. Cette technique de chromatographie ne permet

pas de retenir la molécule d'intérêt, mais les composants que la résine de nickel reconnaît, c'est pourquoi elle est qualifiée de pseudo affinité.

Les protéines étudiées possèdent une étiquette constituée d'histidines qui permet d'avoir une affinité accrue pour le Ni^{2+} . Les colonnes HisTrap HP (GE Healthcare) sont composées d'une matrice NTA (acide nitriloacétique) chélateur du Ni^{2+} . La colonne est préalablement équilibrée avec du tampon A, puis l'élution est assurée par une concentration d'imidazole contenue dans le tampon B.

Chromatographie d'affinité sur colonne d'héparine

Ces colonnes permettent notamment de retenir les protéines liant l'ADN, ce qui est le cas de nos protéines d'intérêt. La colonne Hitrap (GE Healthcare) utilisée est pré-équilibrée dans le tampon D faiblement concentré en sel puis élué avec un gradient de NaCl allant de 50 mM à 2M grâce à l'ajout de tampon E.

L'étiquette histidine est ensuite clivée. Pour cela les protéines sont digérées 16h à température ambiante en présence de 200 ng de protéase TEV. Le mélange réactionnel est ensuite chargé sur une colonne de nickel afin de retenir le produit non-clivé et l'étiquette histidine clivée.

Chromatographie d'exclusion de taille

Aussi appelée filtration sur gel, cette technique de chromatographie permet de séparer des molécules en fonction de leur taille. On utilise pour cela des granules de gel poreux. Les grosses molécules (dont le diamètre est supérieur à celui des pores) sont exclues et sont donc éluées les premières, tandis que les petites et moyennes molécules sont éluées plus tardivement car incluses dans le gel, leur migration est freinée. Les solutés sont donc élués dans l'ordre inverse des masses moléculaires.

La protéine purifiée est injectée sur une colonne de filtration sur gel pré-équilibrée avec du tampon F. La colonne utilisée est une Superdex 200 Increase 10/300 (GE Healthcare). Toutes les étapes de la purification sont réalisées sur un système HPLC Äkta Purifier (GE Healthcare).

Les protéines purifiées sont concentrées par centrifugation à 6000 xg en utilisant des filtres Amicon Ultracel-50 (seuil de coupure 30 kDa). La concentration est obtenue en mesurant l'absorbance des noyaux aromatiques de la protéine à 280 nm au Nanodrop 2000 (Thermo

Scientific). Les coefficients d'absorption molaire ont été calculés de façon théorique en utilisant Protparam (<http://web.expasy.org/protparam>). L'évolution de la pureté des protéines est contrôlée à chacune des étapes en déposant des échantillons des différentes fractions sur un gel de polyacrylamide en conditions dénaturantes (SDS-PAGE) 4-12%. Le gel est révélé au bleu de Coomassie.

III. Cristallographie aux rayons X

La cristallographie aux rayons X est une science qui se consacre à l'étude des cristaux en partant de l'échelle atomique. Au carrefour de disciplines aussi diverses que la physique, la chimie, les mathématiques et même la biophysique, la biologie et la médecine, elle a permis de montrer que les propriétés physico-chimiques d'un cristal sont étroitement liées à l'arrangement spatial des atomes dans la matière. La découverte des rayons X lui a fait accomplir des bonds de géant au cours du XX^{ème} siècle. En effet, ces rayons ont des longueurs d'ondes suffisamment petites pour être sensibles aux détails de la matière à l'échelle des atomes et des molécules. Ces derniers vont transformer la propagation rectiligne des faisceaux de lumière traversant un échantillon de matière et former sur un écran une figure de diffraction. Les figures obtenues ainsi par diffraction des rayons X contiennent de nombreuses informations sur les cristaux et plus généralement des assemblages structurés d'atomes et de molécules.

1. Cristallogenèse

Une étape préalable est indispensable à l'analyse cristallographique : la cristallogenèse. Cette technique consiste à faire des cristaux de protéine à partir de protéine soluble. Le cristal de protéine est un arrangement d'ordre moléculaire. L'empilement des protéines, ordonné de façon périodique et régulier dans les trois directions de l'espace, donne naissance au cristal. Pour obtenir ces cristaux de protéine, il faut trouver les conditions physico-chimiques qui font passer la protéine de l'état soluble à un état solide cristallin, sans faire précipiter.

Techniques utilisées

La technique utilisée ici est appelée diffusion de vapeur. Une goutte contenant l'échantillon protéique est mélangée à la solution de cristallisation dans un environnement hermétique en présence d'un réservoir contenant une grande quantité de la solution de cristallisation. Dans ces conditions, l'évaporation et la diffusion de l'eau et de solvants volatiles entre la goutte et le réservoir permettent d'équilibrer les concentrations entre le réservoir le

mélange protéine/agent précipitant. Il en résulte une augmentation progressive de la protéine et de l'agent précipitant vers une zone de nucléation favorable à l'apparition de cristaux. Au fur et à mesure de la croissance des cristaux, la concentration de la protéine diminue pour atteindre la courbe de solubilité qui se caractérise par un niveau de saturation juste suffisant pour maintenir l'intégrité sur cristal. Deux dispositifs existent selon que la goutte contenant le mélange protéine/précipitant soit assise ou suspendue.

Essais de cristallogenèse sur le PCNA

Le PCNA purifié et concentré à 30 mg/ml a été soumis à des tests de cristallogenèse. Les criblages de cristallisation ont été réalisés en utilisant les kits Crystal Screen 1 et 2 et Index de Hampton Research, ainsi que les kits JBS de 1 à 8 de Jena Bioscience. Ces conditions de cristallisation ont été testées par diffusion vapeur en goutte assise avec un robot Mosquito à 18°C en utilisant des microplaques GREINER (CrystalQuick 96 réservoirs 3 puits à fond concave en polystyrène) avec des volumes de réservoir de 200 µL, et en mélangeant 200 nL de solution protéique et 200 nL de solution de cristallisation.

Obtention des cristaux de PCNA

Les premiers cristaux de PCNA ont été observés à J1 dans certains puits des kits Crystal Screen et Index de Hampton Research. La qualité de ces cristaux était suffisante pour collecter un jeu de données exploitables. Les meilleurs cristaux ont été obtenus à partir de 30 mg/mL de PCNA dans une solution 0,1 M de BIS-TRIS pH 5,5 et 45% (poids/volume) de 2-méthyl-2,4-pentanediol (MPD). Il est nécessaire de plonger les cristaux pêchés dans une solution cryoprotectrice avant de les congeler afin d'éviter la formation de glace organisée susceptible d'interférer avec la diffraction du cristal. Cette étape n'a cependant pas été nécessaire pour les cristaux obtenus dans 45% de MPD puisque celui-ci est un agent cryoprotectant.

2. Diffraction des rayons X et traitement des données

Organisation d'un cristal

Un cristal est un solide dans lequel les molécules sont ordonnées de façon périodique dans les trois dimensions de l'espace. L'unité de base du cristal est la maille élémentaire, dont la répétition par des éléments de translation forme le système cristallin. Cette maille est délimitée par un triplet de vecteurs $\{a, b, c\}$, qui définissent les paramètres de mailles : trois

longueurs a , b , c , et trois angles α , β , γ . En fonction des valeurs que peuvent prendre ces six paramètres, on compte sept systèmes cristallins différents (triclinique, monoclinique, orthorhombique, tétragonal, trigonal, hexagonal, cubique). À l'intérieur de la maille, on trouve un motif (composé d'une ou plusieurs molécules), répété par des éléments de symétrie. Ces symétries s'organisent autour de points particuliers, appelés nœuds, généralement situés aux sommets de la maille. En fonction des positions possibles de ces nœuds, on compte quatre types de mailles élémentaires différentes (primitive, bases centrées, corps centré, faces centrées). En combinant les 7 systèmes cristallins, les quatre types de mailles élémentaires, et les différentes opérations de symétries retrouvées en cristallographie des protéines, on obtient 65 groupes d'espaces possibles. Les premières étapes du traitement des données de cristallographie aux rayons X visent à déterminer le système cristallin et le groupe d'espace du cristal.

Diffraction des rayons X par un cristal

Lorsque qu'un cristal de protéines est traversé par un faisceau de rayons X, le rayon incident est diffracté dans des directions particulières de l'espace. Ces rayons diffractés sont mesurés sur un détecteur, et on observe des taches de diffraction (**Figure 13**).

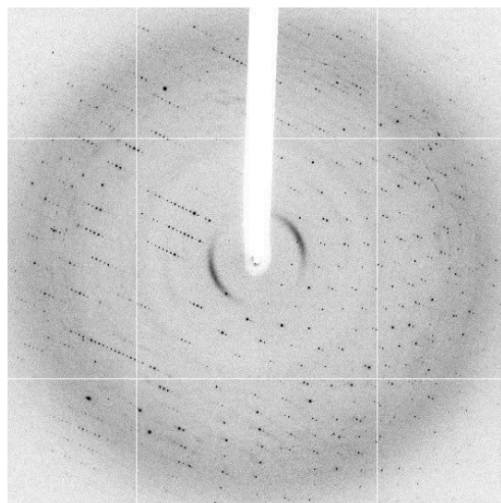


Figure 13. Cliché de diffraction

L'onde diffractée par le cristal dans une direction (h,k,l) donnée de l'espace est définie mathématiquement par le facteur de structure. Ce nombre complexe est caractérisé par son amplitude et sa phase. Cependant, la mesure lors de l'expérience de diffraction nous donne accès uniquement à l'intensité des taches sur le détecteur, qui est proportionnelle au carré de l'amplitude de ces facteurs de structure. L'information de phase est perdue. Pour retrouver cette information, il existe plusieurs méthodes dites de phasage.

Traitement des données

La mesure de la position des tâches sur le détecteur permet de déterminer les paramètres de mailles et le système cristallin. La mesure de l'intensité des tâches permet de caractériser le motif à l'intérieur du cristal ainsi que le groupe d'espace.

- **Détermination de la maille et du groupe d'espace**

Lors de cette étape, appelée indexation, un programme détermine plusieurs possibilités de systèmes cristallins et de paramètres de maille compatibles avec la position des tâches de diffraction. Ces solutions sont classées en fonction d'une pénalité qui tient compte de la déformation que doit subir la maille prédite pour satisfaire les contraintes géométriques du système cristallin.

- **Intégration des données**

En fonction du groupe d'espace choisi précédemment, les positions des tâches de diffraction sont prédites, l'intensité des tâches est mesurée, et un triplet de coordonnées h , k et l leur est attribué.

- **Mise à l'échelle, moyennation et conversion des intensités**

Au fur et à mesure de l'exposition aux rayons X, les dommages dus à l'irradiation causent une diminution du pouvoir diffractant du cristal. Ainsi, l'intensité des réflexions mesurées décroît au cours du temps. Il est donc important de remettre à l'échelle ces intensités pour obtenir un jeu de données complet. À cette étape, les intensités équivalentes par symétrie qui ont été mesurées plusieurs fois sont moyennées. Le facteur RSYM permet d'évaluer la différence entre ces réflexions équivalentes (plus ce facteur est élevé, plus l'écart d'intensité entre les tâches équivalentes est grand). C'est également durant cette étape que le jeu de données est « coupé » à une résolution donnée. Cela peut être fait en fonction de différents critères, comme la valeur du rapport I/σ , ou celle du coefficient de corrélation $CC1/2$.

- **Contenu de l'unité asymétrique**

L'unité asymétrique est le plus petit volume à l'intérieur de la maille à partir duquel l'ensemble du cristal peut être reconstitué en appliquant les éléments de symétrie du groupe

d'espace. Elle peut contenir plusieurs molécules, qui ne sont donc pas reliées par des opérations de symétrie cristallographique. Il est possible d'en estimer le nombre par le calcul du coefficient de Matthews (V_M). Ce coefficient doit être compris entre 1.6 et 4 Å³/Da, ce qui correspond à des fractions de solvant comprises entre 30 et 75% du volume de la maille.

3. Phasage et construction du modèle

Le problème principal de la résolution d'une structure cristallographique vient de l'incapacité à mesurer expérimentalement la phase des ondes diffractées par le cristal. En effet, les détecteurs mesurent uniquement les intensités des faisceaux diffractés, et les phases sont perdues lors de la mesure. Or, pour recalculer la densité électronique diffractante, il est indispensable de connaître ces deux informations. Au cours de mes travaux, j'ai utilisé une technique de phasage mathématique, appelée remplacement moléculaire, qui utilise les phases d'un modèle homologue dont la structure est connue.

L'approche utilisée pour déterminer la structure cristallographique du PCNA de *Pyrococcus abyssi* est détaillée dans la partie résultats. Les statistiques des données collectées et du modèle affiné sont résumées dans le **Tableau 7** (p. 52).

IV. Tests d'interaction par interférométrie à bio-couches

1. Principe général de la technique

La technique utilisée ici pour analyser les interactions intermoléculaires est appelée interférométrie à bio-couches. Ce travail a été réalisé sur un appareil Octet RED384 (Pall ForteBio) qui permet de mesurer les interactions macromoléculaires en analysant les motifs d'interférence de la lumière blanche réfléchi depuis la surface de la pointe d'un biocapteur. La lumière blanche provenant d'une LED est dirigée par la fibre optique vers le bas du biocapteur où se trouvent deux interfaces séparées par une mince couche : une couche biocompatible (ligand/analyte) et une couche de référence interne. Chacune de ces couches réfléchit la lumière, donnant lieu à des interférences constructives ou destructives à différentes longueurs d'onde qui sont enregistrées grâce à un capteur. Lorsque les longueurs d'onde identiques de chaque faisceau sont analysées ensemble, une figure d'interférences apparaît. En représentant les amplitudes observées en fonction des longueurs d'ondes, un profil interférométrique peut être obtenu. Lorsque les molécules se lient à la surface du biocapteur, la longueur du trajet de la réflexion entre la surface et la solution augmente tandis que celle de l'autre réflexion de

référence reste inchangée. Ce phénomène modifie les modèles d'interférences pour toutes les longueurs d'onde et le nouveau profil interférométrique se trouve décalé vers la droite du profil d'origine. Ce dispositif permet de suivre en temps réel la fixation de partenaires sur la protéine immobilisée.

L'extrémité de chaque biocapteur est revêtue d'une matrice biocompatible à base de silice qui minimise la liaison non spécifique à la surface. La construction de PCNA utilisée ayant été purifiée avec son étiquette 6-His, nous avons décidé de nous servir des biocapteurs NTA - Ni-NTA sur lesquelles du nickel est fixé sur de l'acide nitrotriacétique (NTA) et qui présentent par conséquent une forte affinité pour les protéines marquée par des histidines. Le principe de cette méthode est schématisé dans la **Figure 14**.

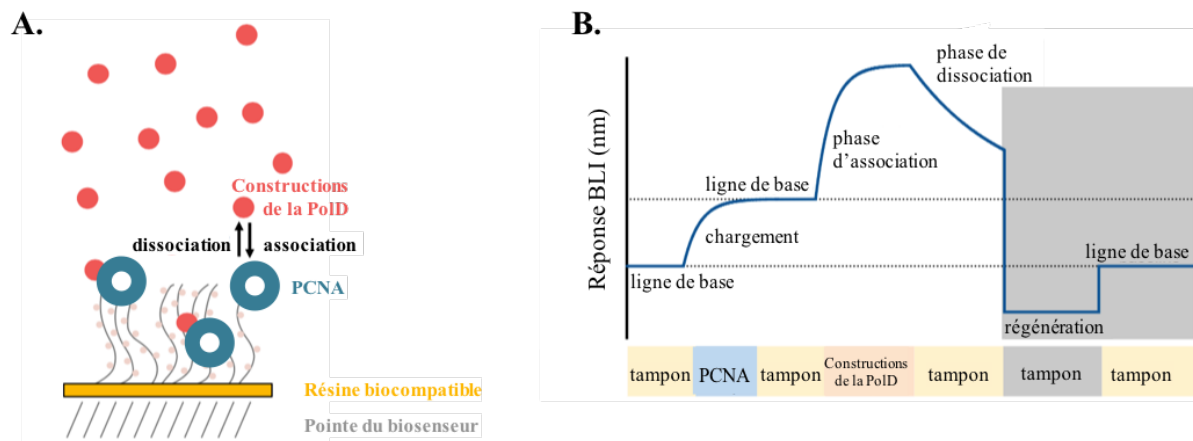


Figure 14. Principe de la technique d'interférométrie à bio-couches. (A) Le PCNA est immobilisé à la pointe d'un biosenseur revêtue d'une résine biocompatible et sa liaison aux différentes constructions de la PoID est mesurée. La variation du nombre de molécules liées à la pointe entraîne un motif d'interférence mesuré en temps réel (nm). (B) Courbe type obtenue lors des différentes étapes de l'expérience.

2. Design des expériences

Les tests d'interaction par interférométrie à bio-couches ont été conduits sur un Octet RED384 (Pall ForteBio) en utilisant des plaques opaques 96 puits. Le volume d'échantillon ou de tampon déposé dans chaque puits est compris entre 230 et 300 μL et les travaux sont réalisés à température ambiante. Le tampon utilisé est du tampon F (200 mM NaCl ; 20 mM HEPES pH 8) contenant de l'albumine de sérum bovin (BSA) à une concentration de 20 mg/ml et utilisée dans le but de limiter les interactions non spécifiques.

Avant chaque expérience, chaque pointe Ni-NTA est équilibrée dans 250 μL de tampon pendant 60 secondes. Ces dernières sont ensuite chargées avec du PCNA dilué à la

concentration de 120 nM pendant 300 secondes. Une étape d'équilibration dans du tampon permet d'éliminer toutes les molécules de PCNA non chargées. S'en suit l'étape d'association avec les différentes constructions pendant 600 secondes puis l'étape de dissociation par trempage des pointes dans du tampon pendant 300 secondes. Chaque pointe est ensuite régénérée en passant successivement dans du tampon, de la glycine à pH 2,1, du tampon, de l'EDTA 0,5 M, du tampon et enfin du chlorure de nickel (NiCl₂) 1 mM.

Le cycle est alors entièrement répété en remplaçant cette fois le PCNA par notre contrôle négatif à 267 nm (la protéine soluble his-taguée provenant du domaine N-terminal d'un récepteur membranaire qui ne partage aucun lien avec la PolD ou le replisome). Puisque la PolD n'interagit pas avec cette dernière, ceci permet en effet de voir son taux de fixation non-spécifique sur le PCNA. La concentration de cette protéine « contrôle » a été choisie pour garantir un même taux de chargement sur les pointes qu'avec le PCNA.

A la fin des expérimentations, les données sont exportées à partir du logiciel Octet, et retravaillées dans un premier temps sur le logiciel Scrubber (BioLogic Software). Lors de cette étape, chaque série de données peut être isolée et alignée sur une même ligne de base. Le traitement final s'effectue sur le logiciel BIAevaluation (Biacore). Les courbes obtenues peuvent y être retraitées et faire l'objet de différents calculs afin d'obtenir une idée plus précise du modèle d'interaction mis en jeu.

Ainsi, pour chaque construction, plusieurs superpositions de modèles ont été testés et nous ont permis d'obtenir des valeurs approximatives des constantes d'association (k_{on}), de dissociation (k_{off}) et de la constante de dissociation à l'équilibre (K_d). Le modèle « Langmuir binding » est le modèle d'interaction 1:1 le plus simple entre un analyte et un ligand immobilisé, et suppose que l'analyte se lie de la même façon sur chaque ligand immobilisé. Le modèle « biphasic dissociation » n'analyse que la phase de dissociation et permet de mieux comprendre combien de phases la compose, et si celles-ci sont lentes ou rapides. Le modèle « heterogenous ligand (parallel reactions) » décrit une interaction entre un analyte et deux ligands indépendants. La courbe d'association obtenue est la somme des deux réactions indépendantes. Cette situation est fréquemment retrouvée lorsque le ligand est immobilisé de façon hétérogène, quand il possède de multiples points d'attache, ou encore lorsque l'échantillon de ligand utilisé est lui-même hétérogène. Le modèle « two-state (conformational change) » décrit une association 1:1 de l'analyte au ligand immobilisé suivi d'un changement de conformation du complexe. Pour simplifier le modèle, on suppose que le complexe dont la conformation a été changée ne peut se dissocier que par le retour au complexe sans changement conformationnel, ainsi : $A + B \leftrightarrow AB \leftrightarrow AB^*$.

RÉSULTATS

I. Production du PCNA de *Pyrococcus abyssi*

Les études structurales sur le PCNA et les tests d'interaction entre le PCNA et la PolD réalisées dans le cadre de mon stage nécessitent des milligrammes de protéine purifiée. Une construction permettant l'expression recombinante chez *Escherichia coli* du PCNA de *P. abyssi* (voir la partie Matériel et méthodes) a été réalisée par l'équipe du Docteur Ghislaine Henneke de l'IFREMER de Brest avant mon arrivée au laboratoire. Cette construction porte une étiquette poly-histidines en position N-terminale. Au cours de mon stage, j'ai mis au point un protocole de purification élaboré permettant d'obtenir des quantités de protéines suffisantes avec un degré de purification élevé.

1. Expression recombinante

La construction a été exprimée dans la souche d'expression *E. coli* BL21-CodonPlus (DE3)-RIPL. La surexpression de la protéine a été déclenchée par ajout d'IPTG et d'arabinose dans la culture. Après induction, la culture est incubée à 20°C afin de ralentir l'expression du PCNA et d'améliorer ainsi l'efficacité de son repliement. Les profils d'expression et de solubilité du PCNA recombinant ont été analysés sur gel SDS-PAGE (**Figure 15**). La protéine exprimée se retrouve majoritairement sous forme insoluble, cependant la fraction soluble contient une quantité de protéine suffisante pour entreprendre une purification de l'enzyme.

2. Purification

Le protocole de purification mis au point au laboratoire est composé de trois étapes chromatographiques. La première étape consiste en une chromatographie de pseudo-affinité sur colonne de nickel. Le profil d'élution présente un pic très étalé qui semble multiple (**Figure 15 A**). La réalisation d'un gel SDS-PAGE permet de confirmer la présence d'un contaminant élué en début de gradient, soit aux plus faibles concentrations d'imidazole (**Figure 15 B**).

Les fractions contenant le PCNA sont rassemblées puis dialysées afin de diminuer la concentration en NaCl qui pourrait interférer avec l'étape chromatographique suivante, la fixation du PCNA sur colonne héparine. Dans notre cas, cette étape est indispensable car le pic d'absorbance mesuré après passage sur colonne de nickel se situe principalement à 260 nm, signe d'une contamination nucléique (**Figure 15 E**). A l'instar de nombreuses protéines qui interagissent avec l'ADN, le PCNA présente une forte affinité pour la colonne héparine qui mime le squelette phosphodiester de l'ADN. Ainsi, le PCNA est retenu sur la colonne tandis que certains contaminants (dont l'ADN) sont éliminés. La séquence du PCNA de *P. abyssi*

comprend plusieurs tyrosines, mais aucun tryptophane. Son maximum d'absorbance dans l'UV se situe donc à 274 nm et non à 280 nm. Après passage sur la colonne héparine le pic d'absorbance se situe à 274 nm, ce qui confirme l'élimination de la contamination nucléaire (**Figure 15 E**).

Enfin, une étape de polissage est réalisée sur une colonne de gel filtration. Plusieurs pics sont observés sur le chromatogramme. La réalisation d'un gel SDS-PAGE permet de confirmer la présence d'un contaminant dans le dernier pic (**Figure 15 G**). En revanche, l'analyse sur gel indique que le PCNA est présent majoritairement dans les fractions correspondant à tous les autres pics, suggérant que plusieurs formes oligomériques de cette protéine coexistent en solution. Ce résultat n'est pas surprenant puisque les monomères de PCNA s'organisent en trimères en forme d'anneaux qui s'associent et se dissocient autour de l'ADN. Le pic d'élution majoritaire a été réinjecté, et deux pics ont été obtenus sur le chromatogramme (**Figure 15 F**). Le premier, d'un poids moléculaire inférieur, correspondrait à la forme monomérique du PCNA tandis que le second correspondrait à sa forme trimérique. Des études d'analyses biophysiques comme une ultracentrifugation analytique permettrait de déterminer avec précision la stœchiométrie des différents pics de gel filtration. Néanmoins, nos résultats de chromatographie d'exclusion démontrent que l'association du complexe est réversible et concentration-dépendante.

Deux purifications ont été réalisées avec un rendement final de 3 mg de protéine par litre de culture. La PCNA produit à l'issue de la première purification a permis de mener les tests d'interaction avec les différentes constructions de la PolD par interférométrie à bio-couches, tandis que la seconde purification, a permis de conduire une étude structurale par cristallographie aux rayons X. Il est à noter que l'étiquette poly-histidines n'est pas clivée au cours de la purification ; elle sera d'ailleurs nécessaire afin de lier le PCNA aux senseurs lors des tests d'interaction.

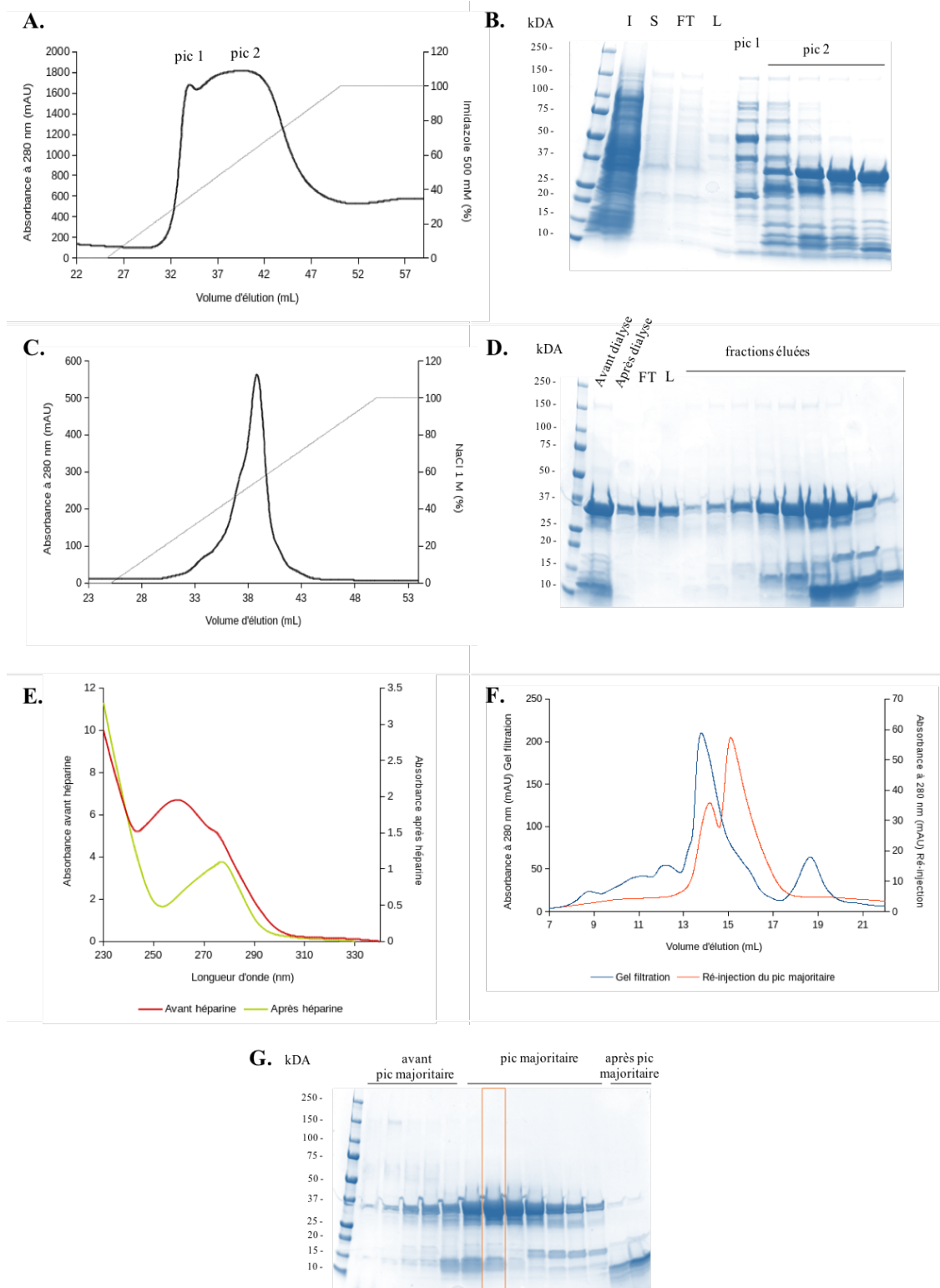


Figure 15. Suivi des étapes de purification du PCNA *via* des profils électrophorétiques et chromatographiques. (A) Chromatogramme d'élué du PCNA sur une colonne de nickel. (B) Gel SDS-PAGE des fractions insolubles *I*, solubles *S*, flow-through *FT*, de lavage *L*, et des fractions éluées. (C) Chromatogramme de l'élué du PCNA sur une colonne héparine. (D) Gel SDS-PAGE des fractions avant dialyse, après dialyse, flow-through *FT*, de lavage *L*, et des fractions éluées de la colonne d'héparine. (E) Mesure de l'absorbance avant et après passage sur colonne héparine. (F) Chromatogrammes de l'élué du PCNA et de l'élué du pic majoritaire réinjecté en gel filtration. (G) Gel SDS-PAGE des fractions éluées en gel filtration. La fraction réinjectée est encadrée en orange.

II. Cristallographie et études structurales du PCNA de *Pyrococcus abyssi*

1. Cristallogenèse et collecte des données de diffraction aux rayons X

Plusieurs cribles ont été réalisés avec un robot de cristallogenèse sur la plateforme de cristallographie de l'Institut Pasteur avec le PCNA purifié et concentré à 30 mg/ml. Les meilleurs cristaux de PCNA ont été observés après 24 heures dans deux solutions (**Figure 16**). Ces cristaux ont une forme hexagonale régulière et mesurent environ 60 μm x 60 μm x 30 μm .

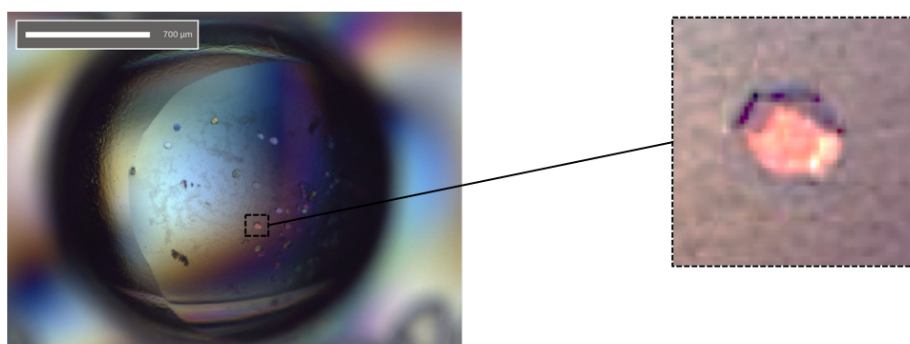


Figure 16. Obtention des cristaux de PCNA. Les cristaux ont été obtenus par criblage sur un robot de cristallogenèse à 18°C. Les meilleurs cristaux sont apparus après 24 heures, ont une forme hexagonale régulière. La condition présentée est composée de 0,1 M de BIS-TRIS à pH 5.5 et de 45% v/v de MPD.

2. Résolution de la structure cristallographique par remplacement moléculaire

Les dimensions réduites des cristaux du PCNA de *P. abyssi* (< 100 μm) ne sont pas compatibles avec la configuration de toutes les lignes de lumière. En revanche, la ligne de lumière micro-focus Proxima 2 (Synchrotron SOLEIL, Gif-sur-Yvette) possède une dimension de faisceau d'environ 5 μm x 3,5 μm , adapté à la taille de nos cristaux, et qui nous a permis de collecter des jeux de données exploitables avec ces cristaux diffractant à une résolution maximale de 2,3 Å. Les données collectées ont été intégrées et mises à l'échelle. Le coefficient de Matthews prédit la présence d'un seul monomère dans l'unité asymétrique avec 55 % de solvant. Le groupe d'espace le plus probable est le P6₃, le système réticulaire est donc hexagonal.

Nous disposons d'un bon modèle de départ avec le monomère de PCNA de *P. furiosus* (**58**) (code PDB : 1GE8) qui présente une identité de séquence de 89% avec le PCNA de *P. abyssi*. La structure cristallographique du PCNA de *P. abyssi* a donc pu être déterminée par la méthode du remplacement moléculaire avec le logiciel PHASER dans la suite du programme PHENIX, en imposant la recherche d'un monomère dans l'unité asymétrique et le groupe d'espace P6₃. Une solution unique contenant le monomère dans l'unité asymétrique est obtenue

avec les statistiques suivantes : LLG de 999,686 et TFZ de 26,9. Ces deux scores sont utilisés pour estimer la justesse du modèle. Le LLG (Log Likelihood Gain) mesure à quel point le résultat peut être mieux prédit avec notre modèle par rapport à une distribution aléatoire des mêmes atomes : le LLG doit être positif et le plus élevé possible. Le plus souvent, une valeur de TFZ supérieure à 8 indique qu'une solution correcte a bien été trouvée.

Le premier modèle obtenu par remplacement moléculaire a été affiné avec des valeurs de R-work et de R-free respectives de 29,24% et 35,32%. Ces statistiques encourageantes confirment que la solution trouvée est correcte et en accord avec les valeurs de LLG et de TFZ. Ce premier modèle été modifié manuellement dans COOT afin de muter les 10% de résidus non conservés chez *P. abyssi* et d'obtenir ainsi la bonne séquence en acides aminés. Des cycles d'affinement ont été lancés avec BUSTER, suivis de modifications manuelles dans COOT de la structure pour placer au mieux les acides aminés dans la densité électronique, en se servant également du modèle muté en cas de difficulté (**Figure 17**). Les valeurs respectives de R-work et de R-free finales sont de 20,40% et 24,30% et 56 molécules d'eau ont été modélisées. Seule la boucle connective qui relie les deux domaines du PCNA n'a pas pu être modélisée, cette région ne présentant qu'une très faible densité électronique. Les caractéristiques du jeu de données exploité ainsi que les statistiques d'affinement de la structure sont présentées dans le **Tableau 7**.

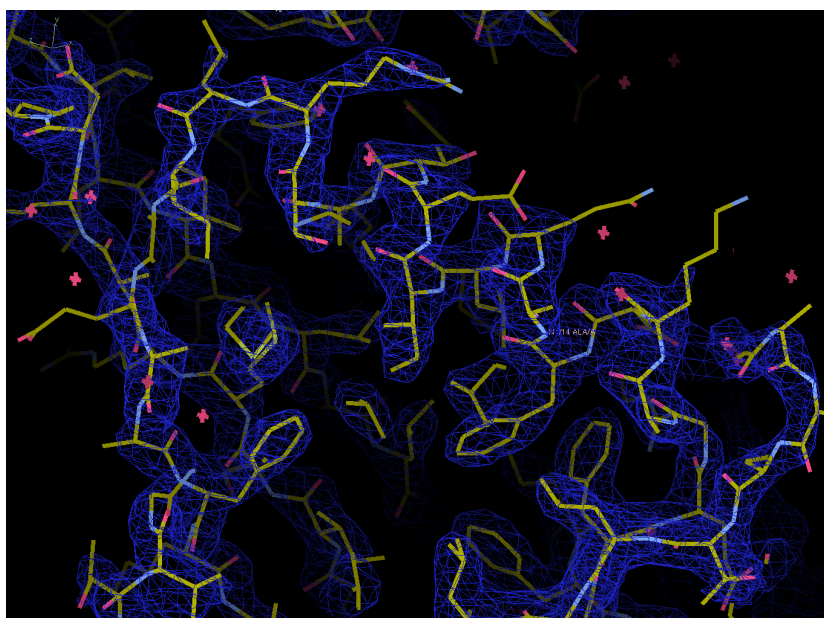


Figure 17. Structure en cours d'affinement dans le logiciel COOT. La densité électronique est colorée en bleu et les molécules d'eau sont indiquées par des croix roses.

PCNA <i>P. abyssi</i>	
Collecte des données	
Groupe d'espace	P6 ₃
Paramètres de maille	
<i>a, b, c</i> (Å)	91.90, 91.90, 64.19
α, β, γ (°)	90, 90, 120
Résolution (Å)	39.79 – 2.30 (2.38 – 2.30)*
<i>R</i> _{sym}	0.027 – (1.03)
CC(1/2)#	1 (0.416)
<i>I</i> / σI	18.3 (0.9)
Complétude (%)	99.9 (99.4)
Multiplicité	13.7 (10.7)
Facteur B de Wilson (Å ²)	73.53
Affinement	
Résolution (Å)	20 – 2.3
Nombre de réflexions	13820
<i>R</i> _{work} / <i>R</i> _{free}	0.204 / 0.243
Modèle	
Protéine	1877
Ligands/Ions	N/A
Molécules d'eau	56
Facteurs B (Å ²)	
Protéine	77.237
Ligands/Ions	N/A
Molécules d'eau	76.987
Déviations moyennes (r.m.s.d.)	
Longueur des liaisons (Å)	0.01
Angles entre les liaisons (°)	1.24
*Les valeurs entre parenthèses correspondent à la coquille de plus haute résolution	
# Mn (I) half-set correlation CC(1/2)	

Tableau 7. Caractéristiques des jeux de donnée collectés et statistiques d'affinement de la structure.

3. Analyse structurale du PCNA de *Pyrococcus abyssi*

Le PCNA de *P. abyssi* présente une architecture en forme d'anneau, avec un diamètre d'environ 90 Å, une ouverture centrale d'un diamètre de 32 Å, et une épaisseur de 30 Å (**Figure 18 B**). Le monomère du PCNA de *P. abyssi* possède deux sous-domaines, un sous-domaine N-terminal et un sous-domaine C-terminal, reliés entre eux par une boucle connective (**Figure 18 A**). Cette région ne présentant qu'une faible densité électronique, elle n'a cependant pas pu être modélisée. Chaque sous-domaine est composé de 2 feuillets β antiparallèles, l'un comprenant 4 brins β et l'autre 5 brins β , bordés par deux hélices α parallèles. Le PCNA s'associe en trimère par l'intermédiaire des liaisons hydrogène entre un brin β du sous-domaine N-terminal d'un monomère, et un brin β du sous-domaine C-terminal du monomère adjacent.

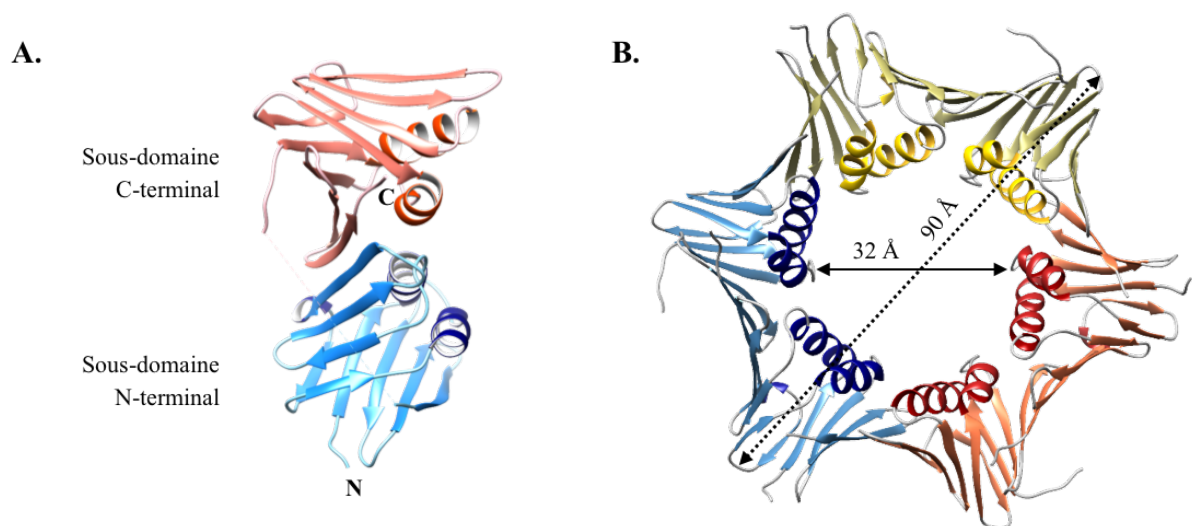


Figure 18. Représentation tridimensionnelle du PCNA sous différentes formes oligomériques. (A) Structure d'un monomère de PCNA composé d'un sous-domaine N-terminal et d'un sous-domaine C-terminal. Chaque sous-domaine est composé de 2 feuillets β antiparallèles bordés par deux hélices α parallèles. (B) Structure du trimère de PCNA formé par liaison entre un brins β du sous-domaine N-terminal d'un monomère et un brin β du sous-domaine C-terminal d'un autre monomère. L'anneau mesure environ 90 Å de large et sa cavité centrale a un diamètre d'environ 32 Å.

Le PCNA de *P. abyssi* a cristallisé dans un groupe d'espace $P6_3$ et présente un monomère dans son unité asymétrique. Le trimère peut être reconstitué par symétrie cristalline mais plusieurs arguments suggèrent que l'assemblage quaternaire observé n'est pas un artefact de cristallographie. Tout d'abord, cet agencement est retrouvé chez de nombreux autres PCNA procaryotes et eucaryotes dont les structures ont été résolues en cristallographie avec des groupes d'espace et des empilements cristallins différents. Chez les archées, c'est par exemple le cas pour *Haloferax volcanii* dont la structure a été résolue dans le groupe d'espace $C1\ 2\ 1$ avec deux trimères dans l'unité asymétrique (code PDB : 3HI8), ou de *Thermococcus gammatolerans* dont la structure a été résolue dans le groupe d'espace $P3$ avec comprend deux monomères dans l'unité asymétrique (code PDB : 5A6D). Les structures du PCNA de *P. abyssi* et des deux autres espèces sont superposables avec des valeurs de RMSD de 1,3 Å et 0,55 Å, respectivement calculées sur 672 et 681 C α . De plus, des structures du PCNA trimérique ont déjà été déterminées par microscopie électronique, notamment celle du PCNA de *P. furiosus* en complexe avec la PolB et l'ADN (68). Sans contrainte cristallographique, le PCNA se présente également sous forme d'anneau et les structure du PCNA de *P. abyssi* et de *P. furiosus* sont superposables avec une valeur de RMSD de 1,16 Å calculé sur la superposition de 644 C α .

Le potentiel électrostatique de surface du PCNA présente une distribution largement biaisée (**Figure 19**). La surface interne de l'anneau, qui interagit avec l'ADN, est chargée positivement avec de nombreux résidus lysine et arginine. Une analyse des régions conservées par le logiciel Consurf permet de distinguer deux faces chez le PCNA, l'une très conservée entre les différentes espèces et l'autre plus variable. La superposition de la structure du PCNA de *P. furiosus* lié à un peptide de 11 résidus contenant le motif PIP de la grande sous-unité du facteur de réplication C (RFC) (PDB ID : 1ISQ) (**69**) sur la structure du PCNA de *P. abyssi* montre que la région occupée par le motif PIP coïncide avec une région très conservée du PCNA (**Figure 20 A et B**). Les informations structurales apportées par la cristallographie nous ont permis de mieux comprendre l'architecture du PCNA et d'émettre de premières hypothèses quant aux zones concernées par l'interaction avec la PolD. Ces données ont pu être complétées par la suite en effectuant des tests d'interaction entre le PCNA et différentes constructions de la PolD par interférométrie à bio-couches.

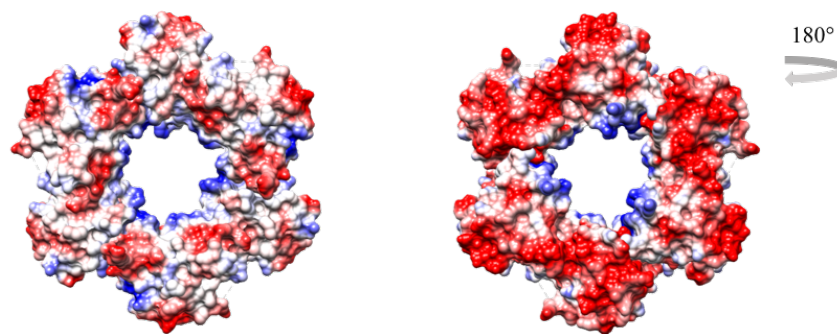


Figure 19. Répartition du potentiel électrostatique de surface sur les deux faces du PCNA. Les résidus chargés positivement sont représentés en bleu et les résidus chargés négativement sont représentés en rouge.

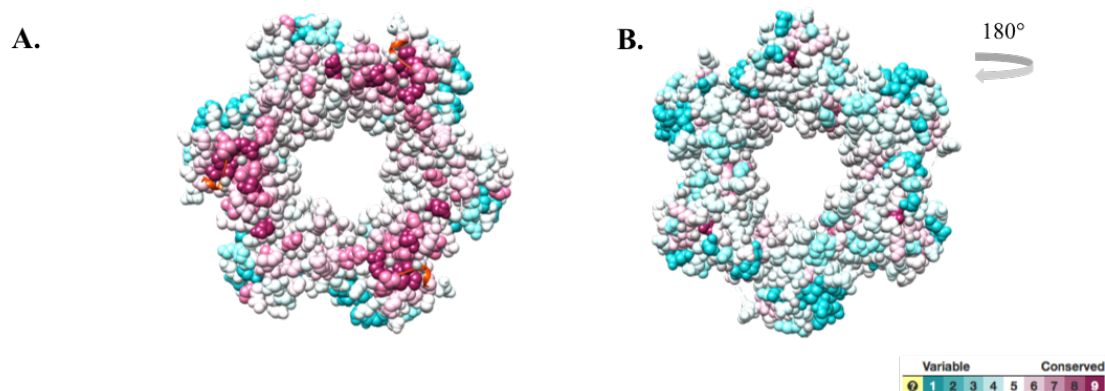


Figure 20. Répartition des régions conservées à la surface du PCNA. Les résidus conservés sont représentés en mauve et les résidus non conservés sont représentés en bleu. (A) Superposition de la structure du PCNA de *Pyrococcus furiosus* lié à un peptide de 11 résidus contenant le motif PIP de la grande sous-unité du facteur de réplication C (RFC) (PDB ID : 1ISQ) (**69**) sur la structure du PCNA de *Pyrococcus abyssi*. Le motif PIP représenté en orange se fixe sur la face très conservée. (B) Face peu conservée du PCNA.

III. Cartographie de l'interaction entre le PCNA et la PolD de *Pyrococcus abyssi*

Des tests d'interaction ont été réalisés par interférométrie à bio-couches sur un Octet RED384 (Pall ForteBio). Cette technique permet d'étudier les interactions entre le PCNA et différentes constructions de la PolD en faisant appel à l'analyse des motifs d'interférence de la lumière blanche réfléchi depuis la surface d'un biocapteur. Ce dernier se compose d'une pointe avec une couche biocompatible et une couche de référence interne. Les molécules de PCNA sont immobilisées sur les pointes, et la variation du nombre de molécules des différentes constructions de la PolD liées au PCNA entraîne un décalage du motif d'interférence, qui est mesuré en temps réel (**Figure 14 A et B** p. 46).

1. Production des différentes constructions utilisées

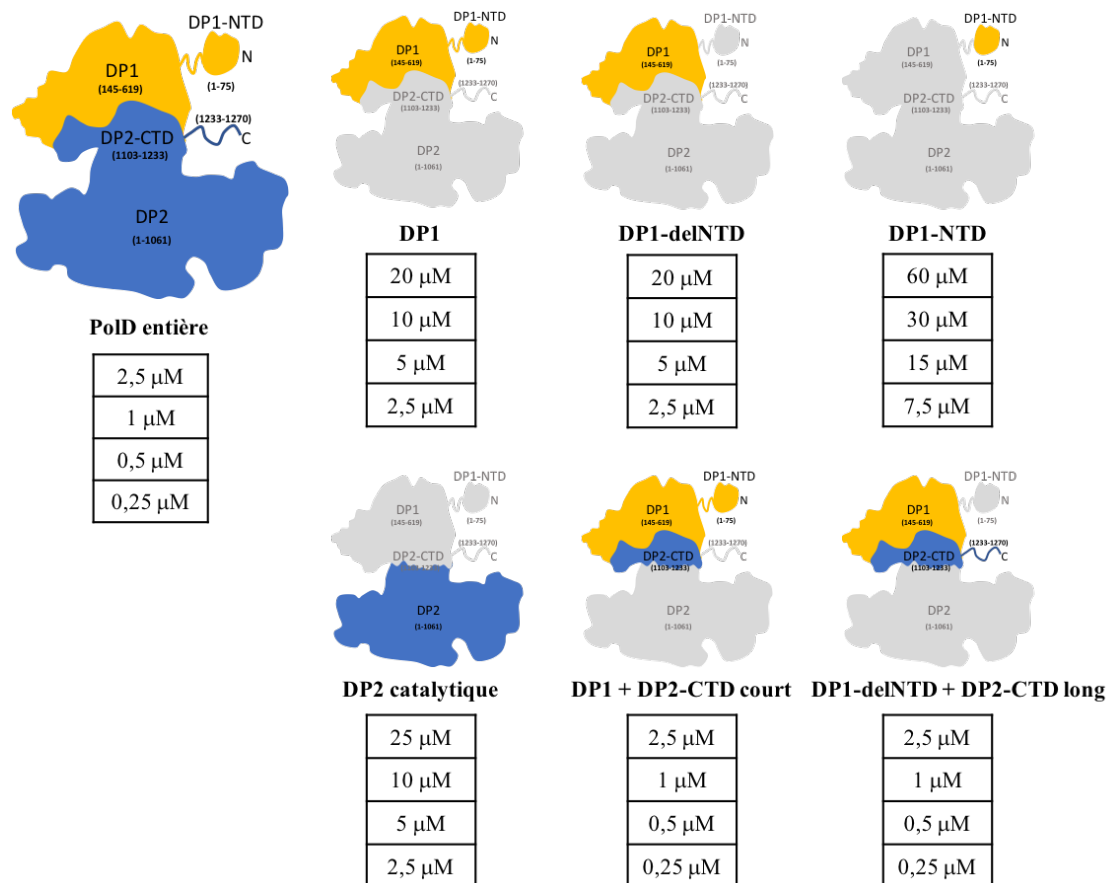


Figure 21. Représentation schématique des différentes constructions de la PolD utilisées lors des tests d'interaction et concentrations choisies pour chacune d'elles. Les domaines non compris dans les constructions sont grisés.

Les constructions testées sont celles de la PolD entière de *P. abyssi*, du domaine DP1 entier (DP1 : résidus 1 à 619), du domaine DP1 tronqué de sa région N-terminale (DP1-

delNTD : résidus 145 à 619), du domaine N-terminal de DP1 seul (DP1-NTD : résidus 1 à 75), du domaine catalytique de DP2 (DP2 catalytique : résidus 1 à 1061), du domaine DP1 entier complexé au domaine C-terminal de DP2 délété des quarante derniers résidus (DP1+DP2-CTDcourt : résidus 1103 à 1233) et de DP1 sans son domaine N-terminal complexé au domaine C-terminal de DP2 entier (DP1-delNTD+DP2-CTDlong : résidus 1103 à 1270). Les caractéristiques (structure et concentration utilisées) pour chacune des constructions sont indiquées dans la **Figure 21**.

Les trois constructions de DP1 avaient été exprimées et purifiées avant mon arrivée au laboratoire. En revanche, j'ai purifié et exprimé le reste des constructions selon un protocole similaire à celui du PCNA. Une étape de clivage de l'étiquette poly-histidines par la protéase TEV a cependant été ajoutée après la seconde étape de chromatographie sur colonne héparine.

2. Tests d'interaction par interférométrie à bio-couches

Mise au point des conditions de l'expérience

La première partie de ces travaux a consisté à optimiser les conditions de l'expérience dans le but de réduire au maximum les interactions non spécifiques. Pour cela, plusieurs étapes de mise au point ont été nécessaires :

- **Choix du type de pointes**

Nous avons testé deux pointes permettant de retenir des partenaires avec une étiquette poly-histidines : les pointes NTA - Ni-NTA et les pointes HIS1K - Anti-Penta - HIS. Le chargement du PCNA étant nettement meilleur sur les pointes NTA - Ni-NTA, ce sont celles-ci que nous avons utilisées pour la suite des expériences.

- **Détermination de la concentration en PCNA à utiliser**

L'objectif est de trouver la concentration minimum permettant le compromis entre une occupation la plus grande possible et une interaction forte et stable du PCNA. Nous avons testé une gamme de quatre concentrations, de 2400 nM à 89 nM. La fixation du PCNA aux pointes était suffisante en se plaçant à 120 nM, nous avons donc conservé cette valeur pour la suite des expérimentations.

- **Choix du tampon**

Nous avons choisi de mener les tests dans le tampon d'élution de gel filtration (200 mM NaCl ; 20 mM HEPES pH 8). Nous avons systématiquement ajouté de la BSA à une

concentration de 0,2 mg/mL dans ce tampon afin de limiter la fixation non-spécifique du PCNA sur les pointes NTA - Ni-NTA.

- **Choix d'un contrôle négatif**

Nous avons également introduit un témoin négatif avec une étiquette poly-histidines qui se lie aux pointes mais ne présente aucune interaction spécifique avec la PolD. Après avoir testé plusieurs constructions disponibles dans le laboratoire, nous avons décidé d'utiliser le domaine N-terminal soluble d'une protéine membranaire étudiée au laboratoire, provenant d'une espèce bactérienne qui ne possède pas de PolD qui ne présente à priori aucune possibilité d'interaction avec la PolD. Après vérification, aucune interaction avec la PolD n'est observée ce qui nous a confortés dans le choix de ce contrôle négatif (**Figure 22**).

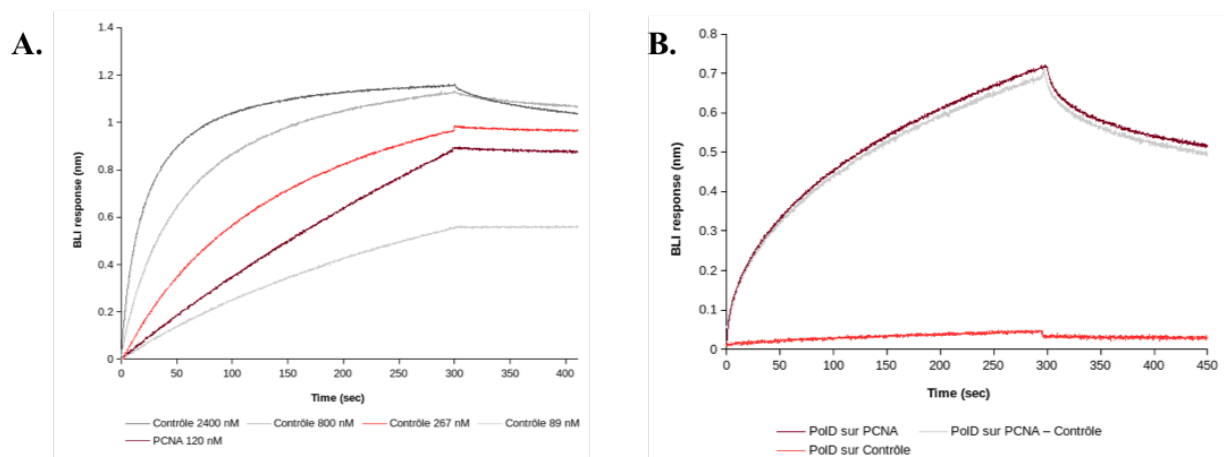


Figure 22. Choix de la protéine contrôle à utiliser. L'objectif est de pouvoir mesurer la réponse non-spécifique (fixation de la PolD aux pointes libres n'ayant pas chargé le PCNA) afin de pouvoir la soustraire à la réponse spécifique. (A) La protéine Contrôle 267 nM (courbe orange) se charge aussi bien sur les pointes que le PCNA (courbe rouge foncé). (B) La PolD ne se fixe que très peu sur le Contrôle (courbe rouge clair) mais se fixe très bien au PCNA (courbe rouge foncé). La réponse non-spécifique peut être soustraite à la réponse spécifique (courbe grise).

- **Régénération des pointes**

Finalement, nous avons observé que les résultats étaient plus reproductibles si les pointes étaient régénérées après chaque mesure de fixation.

Ces mises au point nous ont permis de réaliser la suite des tests d'interaction dans des conditions optimales et de tenir compte de la fixation non-spécifique de la PolD aux pointes libres. Ainsi, nous avons pu soustraire la réponse spécifique à la réponse non spécifique afin de pouvoir observer au mieux les interactions entre le PCNA immobilisée sur les pointes et les différentes constructions de la PolD.

Étude comparative des interactions entre le PCNA et les différentes versions tronquées de la PolD

Le protocole opératoire est identique pour chacune des constructions testées et comprend 5 étapes principales : chargement du PCNA ou de la protéine contrôle sur les pointes, établissement de la ligne de base, association de chaque construction sur le PCNA ou la protéine contrôle, dissociation et enfin régénération des biocapteurs. Les courbes d'association et de dissociation obtenues peuvent alors être retraitées afin d'obtenir des valeurs approximatives des constantes d'association (k_{on}), de dissociation (k_{off}) et de la constante de dissociation à l'équilibre (K_d).

Les valeurs de K_d mesurées ont d'abord été calculées sans faire d'hypothèse sur le type de mécanisme d'interaction impliqué, en se basant uniquement sur la réponse à l'équilibre (**Figure 23**). Toutes les constructions testées, à l'exception du domaine N-terminal de DP1 (DP1-NTD), interagissent avec le PCNA avec des K_d mesurés entre 300 nM et 11,4 μ M, selon la construction.

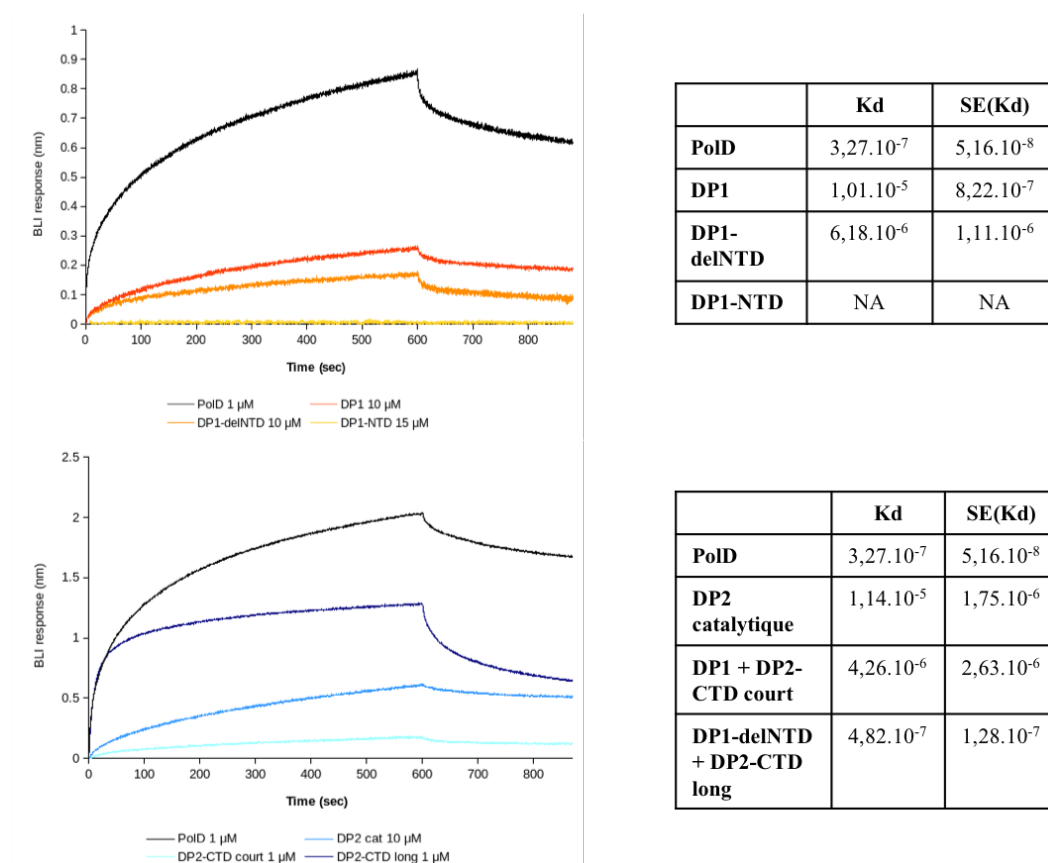


Figure 23. Résultats des tests d'interaction entre les différentes constructions de la PolD et le PCNA interférométrie en bio-couches. L'interaction de chaque construction de la PolD sur le PCNA immobilisé produit des courbes d'association et de dissociation à partir desquelles peuvent être calculées des valeurs de K_d associées à leur marge d'erreur.

L'interaction la plus affine est mesurée pour la PolD entière, avec un K_d de 300 nM. Un K_d de 480 nM est observé pour la construction DP1-delNTD+DP2-CTDlong. Ce K_d est proche de celui mesuré pour la PolD entière suggérant que cette construction couvre la majorité des régions les plus déterminantes pour l'interaction avec le PCNA. Le K_d mesuré pour la construction DP1+DP2-CTDcourt est en revanche de 4,26 μ M, démontrant que la délétion des 37 derniers résidus du domaine C-terminal de DP2 affecte d'un facteur 10 l'affinité d'interaction avec le PCNA. Cette observation est en accord avec des études antérieures menées sur la PolD de *P. furiosus* qui ont démontré l'existence d'un motif PIP localisé à l'extrémité C-terminale de la sous-unité DP2 (65).

Nos travaux ont également démontré que les domaines catalytiques des deux sous-unités de la PolD interagissent également avec le PCNA, suggérant qu'il existe de multiples sites d'interaction entre la PolD et le PCNA. Nous avons ainsi mesuré des K_d de 10 μ M, 6,2 μ M, et 11,4 μ M, respectivement pour les constructions DP1, DP1-delNTD, et DP2 catalytique.

L'analyse des courbes de dissociation suggère qu'il existe deux phases de dissociation, une rapide, et une plus lente. Afin de mieux comprendre les participations respectives de ces deux phases, nous avons appliqué un modèle de calcul basé sur l'hypothèse que la dissociation du complexe est biphasique. Ce modèle permet de calculer une valeur de constante de dissociation (k_{off}) pour chacune d'elles et de connaître la proportion de dissociation rapide et lente. Ainsi, chez la PolD entière, on remarque qu'environ 5 fois plus de molécules se dissocient lentement et que la phase de dissociation rapide est par conséquent moins importante. Pour la construction DP1-delNTD+DP2-CTDlong qui présente pourtant un K_d proche de celui de la PolD entière, on observe exactement l'inverse : une phase de dissociation rapide plus importante que la phase de dissociation lente. En effet, nous avons mesuré pour la construction DP1-delNTD+DP2-CTDlong que la dissociation rapide est 15 fois plus importante que la dissociation lente. Ces données nous permettent de faire que l'hypothèse que, si le domaine C-terminal de DP2 joue un rôle primordial dans l'interaction avec le PCNA, les contributions des autres domaines de la PolD sont déterminantes pour que le complexe PolD-PCNA soit stable.

IV. Proposition d'un modèle d'interaction entre le PCNA et la PolD

Récemment, la structure de la PolD entière de *P. abyssi*, en complexe avec l'ADN, a été déterminée au laboratoire par microscopie électronique (15). Dans cette structure, la PolD adopte un mode dit d'élongation, car elle oriente l'ADN dans son site catalytique polymérase. Nous pouvons désormais combiner les données structurales apportées par la structure cristallographique du PCNA et la structure obtenue par cyo-microscopie électronique de la

PolD, avec les données fonctionnelles obtenues lors des tests d'interaction, afin de proposer un modèle d'interaction entre la PolD en complexe avec l'ADN et le PCNA (**Figure 24**).

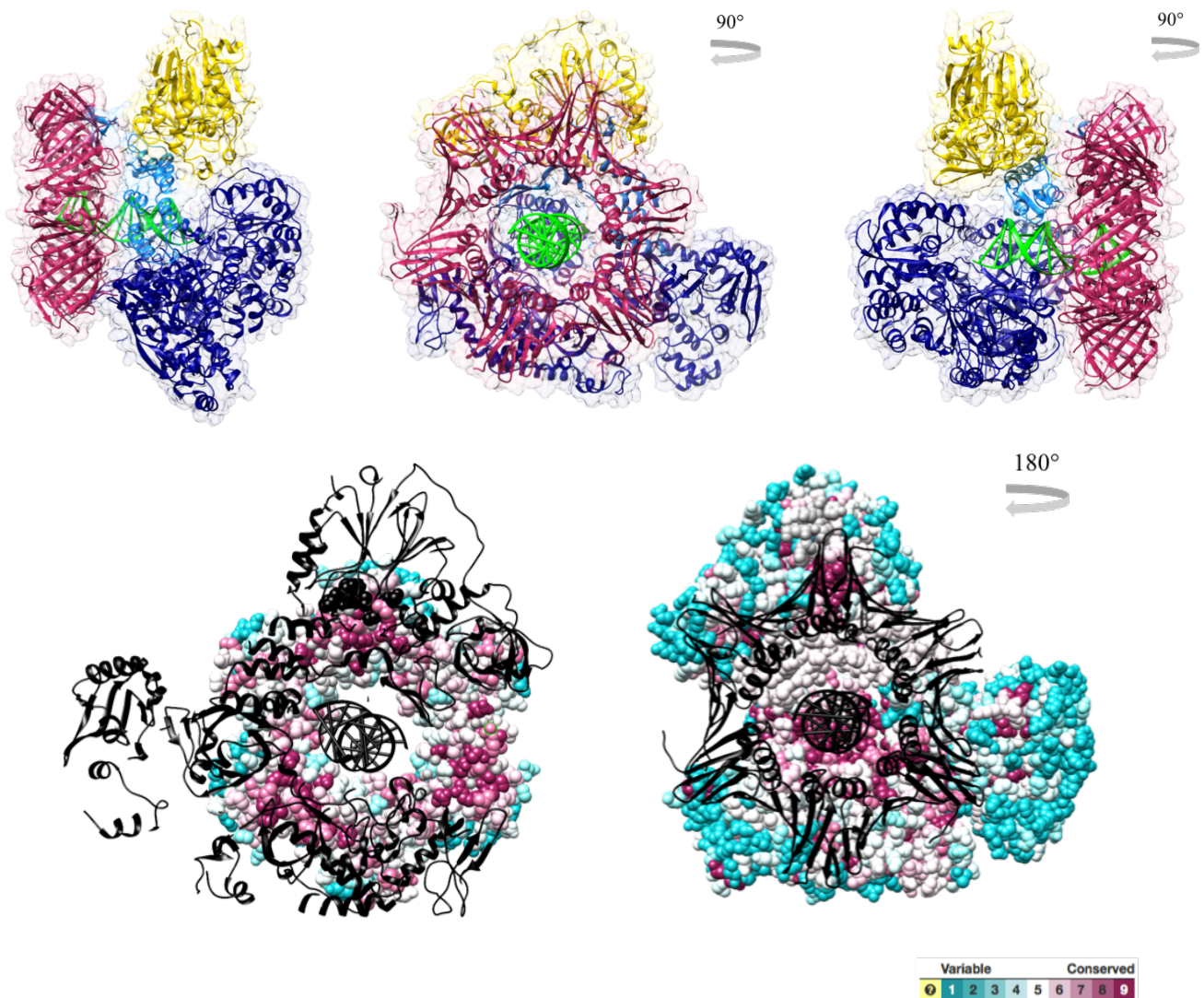


Figure 24. Représentations tridimensionnelles du modèle d'interaction proposé entre la PolD de *Pyrococcus abyssi* en complexe avec l'ADN et le PCNA de *Pyrococcus abyssi*. (A) Vue latérale du complexe. Le PCNA est coloré en rose, l'ADN en vert, la petite sous-unité DP1 de la PolD en jaune, la grande sous-unité DP2 de la PolD en bleu foncé et son domaine C-terminal ainsi que le motif PIP interagissant avec le PCNA en bleu clair. (B) Vue de face du même complexe. (C) Vue latérale du même complexe. (D) Complexe ternaire avec une représentation des régions variables (en bleu) et conservées (en mauve) du PCNA. Une vue partielle de la PolD et l'ADN sont colorés en noir. (E) Complexe ternaire avec une représentation des régions variables (en bleu) et conservées (en mauve) de la PolD. Le PCNA et l'ADN sont colorés en noir.

La structure proposée du complexe ternaire a été modélisée manuellement en se basant sur plusieurs éléments. Tout d'abord, nous connaissons la position de l'ADN par rapport à la PolD et nous savons que le PCNA forme un anneau autour de la double hélice, ce qui limite les

possibilités quant à la position du trimère dans un espace tridimensionnel. La région de fixation du motif PIP sur le PCNA est également connue, elle se situe à l'interface entre les deux sous-domaines de chaque monomère, vraisemblablement au niveau de la boucle connective qui les relie. Nous savons aussi que ce motif est compris dans les derniers résidus du domaine C-terminal de DP2, et nous pouvons donc faire coïncider ces résidus de la PolD avec la zone d'interaction du motif PIP sur le PCNA. Une analyse des régions variables et conservées chez la PolD et le PCNA permet de voir que les régions en contact entre les deux protéines sont conservées.

Dans ce modèle du complexe ternaire PCNA-DNA-PolD, le PCNA se retrouve en contact avec de nombreuses régions qui couvrent non seulement le motif PIP au niveau des derniers résidus C-terminaux de DP2, mais aussi le DP2 catalytique et une partie de DP1. Ces observations sont en accord nos essais d'interaction qui que la stabilité du complexe entre la PolD et le PCNA fait intervenir de multiples régions de la PolD.

La structure de la PolD en microscopie électronique a été déterminée en complexe avec l'ADN dans son mode élongation. Il est probable que la position du PCNA par rapport à la PolD et à l'ADN diffère quand celle-ci opère ses activités de relecture. Ces changements de configuration ont par ailleurs déjà été démontrés par microscopie électronique, avec l'étude du complexe entre une polymérase de *P. furiosus*, l'ADN et le PCNA (70). Le modèle proposé dans cette étude présentait la zone d'interaction avec le motif PIP comme un pivot lors du changement de mode.

DISCUSSION

I. Développement du potentiel biotechnologique de la PoID

1. Le rôle clé des ADN polymérases

Les ADN polymérases sont des acteurs clés du processus de réplication de l'ADN et des outils précieux utilisés dans de nombreuses techniques de biologie moléculaire. Ces nano-machines sont capables de synthétiser l'ADN à partir de nucléotides et ont pu être regroupées en sept familles (A, B, C, D, X, Y et transcriptase inverse) sur la base d'homologies de séquences et de structures. Au-delà de leurs fonctions biologiques essentielles, les ADN polymérases sont des outils précieux nécessaires à de nombreuses applications en biologie moléculaire. Parmi toutes ces applications, la Réaction de Polymérisation en Chaîne (PCR) compte parmi les plus anciennes et les plus connues. Au départ développée pour amplifier un fragment d'ADN de façon spécifique pour des clonages, elle possède désormais de nombreuses applications, notamment en mutagenèse dirigée, en diagnostic génétique, en thérapie génique ou encore en police scientifique. Dans une réaction de PCR, l'ADN est amplifié par des ADN polymérases thermostables. Ces dernières se distinguent par leurs propriétés enzymatiques telles que la spécificité, la fidélité ou encore la résistance à divers contaminants. Le développement de la PCR en médecine clinique et en médecine légale représente un défi supplémentaire, car elle requiert des ADN polymérases capables d'amplifier de l'ADN à partir de tissus biologiques, de sang, et d'autres fluides biologiques.

2. Les ADN polymérases thermostables utilisées en PCR

Pour obtenir les meilleurs résultats possibles, le choix de l'ADN polymérase à utiliser lors d'une réaction de PCR est crucial et dépend de son application (PCR de routine, PCR sur colonies, incorporation de nucléotides modifiés, PCR haute-fidélité...). Aujourd'hui, toutes les ADN polymérases commercialisées en PCR appartiennent invariablement aux familles A (provenant de bactéries thermophiles et hyperthermophiles, comme la Taq-PolA de *Thermus aquaticus*) ou B (provenant d'Archées hyperthermophiles, comme la Pfu-PolB de *Pyrococcus furiosus*). L'ADN polymérase de la famille Y Sso-PolY, provenant de l'Archée hyperthermophile *Sulfolobus solfataricus*, est également commercialisée pour des applications plus spécifiques (amplification d'ADN anciens ou endommagés).

Le développement constant des applications de la PCR entraîne un besoin croissant d'ADN polymérases optimisées pour chacune d'entre elles. La tendance actuelle est donc à la découverte de nouvelles ADN polymérases ou à l'ingénierie d'ADN polymérases déjà connues dans le but d'accroître leurs performances. On peut citer par exemple les polymérases thermo-

activées (hot start polymérase), les polymérase à la processivité augmentées ou résistantes à l'inhibition par l'uracile. De plus, les applications de la PCR en diagnostic clinique requièrent l'utilisation d'ADN polymérase de plus en plus rapides et capables d'amplifier directement des échantillons de tissus, de sang ou de différents fluides corporels. Récemment, une nouvelle famille d'ADN polymérase, la Famille D, appelée PolD, a été découverte chez des Archées thermophiles. Cette PolD est une ADN polymérase impliquée dans la réplication de l'ADN chez les Archées, et possède des propriétés biochimiques intéressantes, exploitables en PCR.

3. Une famille d'ADN polymérase au potentiel biotechnologique inexploité

La PolD est présente chez toutes les Archées sauf celles appartenant au phylum des Crénarchées, qui n'en possèdent pas. Elle est composée d'une grande sous-unité catalytique DP2 et d'une petite sous-unité DP1 qui porte l'activité de relecture exonucléase. Les PolD des Archées appartenant à l'ordre des Thermococcales, stables à de très hautes températures, possèdent des propriétés de processivité, de spécificité et de thermostabilité élevées, ce qui leur confère un potentiel biotechnologique important. Ce potentiel demeure néanmoins inexploité jusqu'à présent en raison du manque d'informations détaillées sur sa structure. Les bases structurales de la liaison à l'ADN, de la reconnaissance et de la sélection des nucléotides sont bien documentés chez les autres ADN polymérase grâce à leurs structures déjà résolues et bien connues. Ces informations structurales ont largement contribué à l'exploitation de leur potentiel biotechnologique, en permettant de concevoir des variants optimisés par ingénierie enzymatique pour diverses applications en biologie moléculaire.

Mon laboratoire d'accueil a déterminé les structures cristallographiques des deux sous-unités isolées de la PolD de *Pyrococcus abyssi* et plus récemment la structure de la PolD entière en complexe avec l'ADN par cryo-microscopie électronique. Ces structures ont montré que la PolD se distingue structurellement des autres ADN polymérase et ont permis de mettre en évidence une parenté évolutive inattendue avec les ARN polymérase à sous-unités multiples responsables de la transcription dans toutes les formes de vie cellulaire. Le cœur catalytique à deux tonneaux β de la PolD, porté par la grande sous-unité DP2, se distingue totalement de ceux des autres ADN polymérase qui appartiennent tous à une des deux superfamilles : Klenow (partagé par toutes les polymérase thermostables commercialisées pour leurs application en PCR) ou Pol β (**Figure 25**).

Par bien des aspects, la PolD est un modèle idéal pour exploiter le potentiel biotechnologique du cœur catalytique atypique à deux tonneaux β . En effet, sa structure est beaucoup plus simple que celle des ARN polymérase auxquelles elle s'apparente, mais qui possèdent de multiples sous-unités. De plus, les activités exonucléase et polymérase de la

PolD sont portées par deux sous-unités distinctes, ce qui permet de travailler sur chacun des deux mécanismes séparément. Enfin, contrairement à la plupart des polymérase au cœur catalytique Klenow, la PolD est capable d'étendre indifféremment des amorces ADN et ARN.

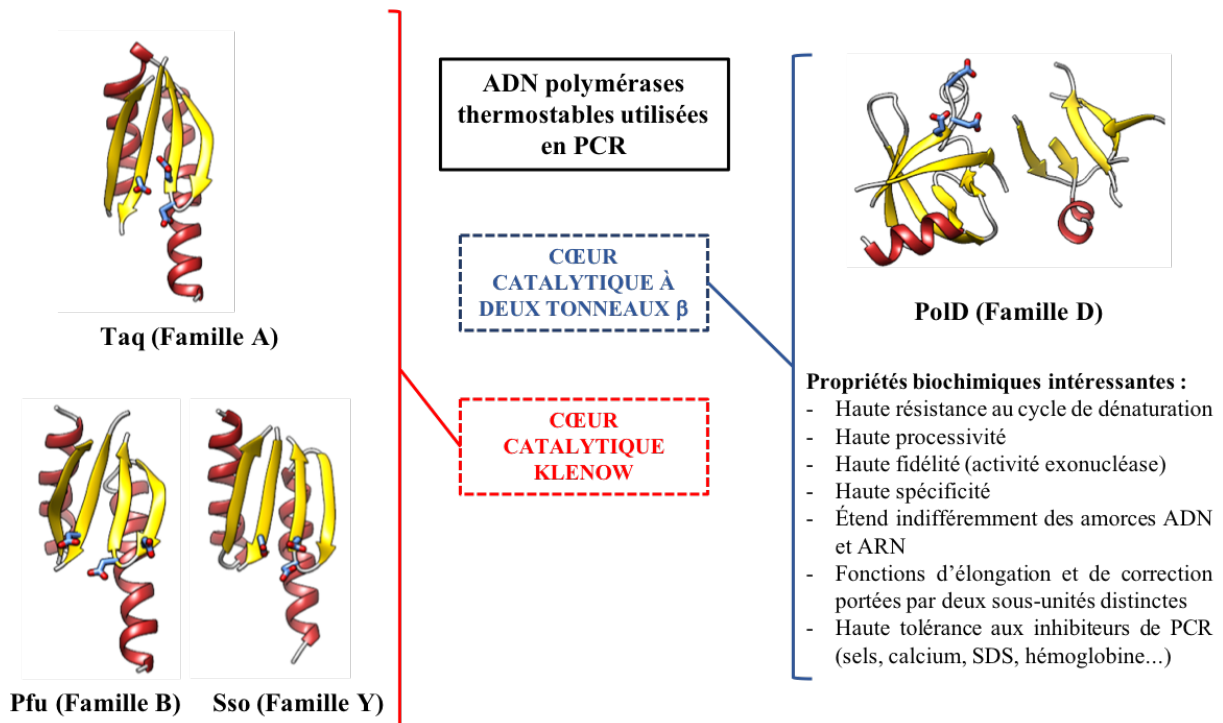


Figure 25. Cœurs catalytiques des ADN polymérase thermostables utilisées en PCR.

4. Intérêts du couplage de la PolD avec le PCNA

La PolD est capable de contourner certains obstacles rencontrés par d'autres polymérase couramment utilisées en PCR, comme par exemple dans l'amplification d'échantillons riches en calcium. Cependant, des études antérieures ont montré que le rendement des PCR était drastiquement réduit lorsque la concentration en ADN était inférieure à 0,1 ng, ou quand la taille du fragment à amplifier excédait les 3500 paires de bases. Même si c'est aussi le cas pour d'autres polymérase thermostables déjà commercialisées, ces caractéristiques pourraient être limitantes dans certains domaines d'application de la PCR, notamment en police scientifique où l'ADN peut être présent en très petites quantités.

L'association de la PolD avec le PCNA, un de ses partenaires clés au sein de la fourche de réplication, pourrait permettre de surmonter ces limitations en l'aidant à synthétiser de plus longs fragments d'ADN et en augmentant sa sensibilité à l'ADN. Le PCNA orchestre la coopération entre les différents partenaires de la fourche de réplication et des études antérieures ont déjà montré qu'il stimulait considérablement l'activité de la PolD in vitro. A ce jour, il

n'existe aucune structure du complexe à haute résolution entre une polymérase répliquative et son PCNA. La connaissance des bases moléculaires de l'interaction entre la PolD et le PCNA, largement documentée par la structure du complexe qu'elles forment, est indispensable à la compréhension des différents mécanismes mis en jeu lors de la réplication, mais aussi à l'exploitation du potentiel biotechnologique de cette polymérase atypique.

En effet, les données apportées par la structure du complexe entre la PolD et le PCNA permettront de disposer des bases structurales nécessaires à la mise au point d'un couplage optimal entre ces deux protéines par des approches d'ingénierie moléculaire, afin d'augmenter la processivité et la sensibilité à l'ADN de la PolD en conditions de PCR. La structure du complexe PolD-ADN-PCNA dans ses modes élongation (polymérase) et édition (exonucléase) par cryo-microscopie électronique est en cours de détermination au laboratoire. Ces données vont bientôt permettre non seulement de caractériser les bases moléculaires du couplage entre ces deux protéines dans le détail mais aussi les changements de conformation entre les deux modes. Des images à 4,2 Å de résolution du complexe en mode élongation sont déjà disponibles et présentées dans la **Figure 26**.

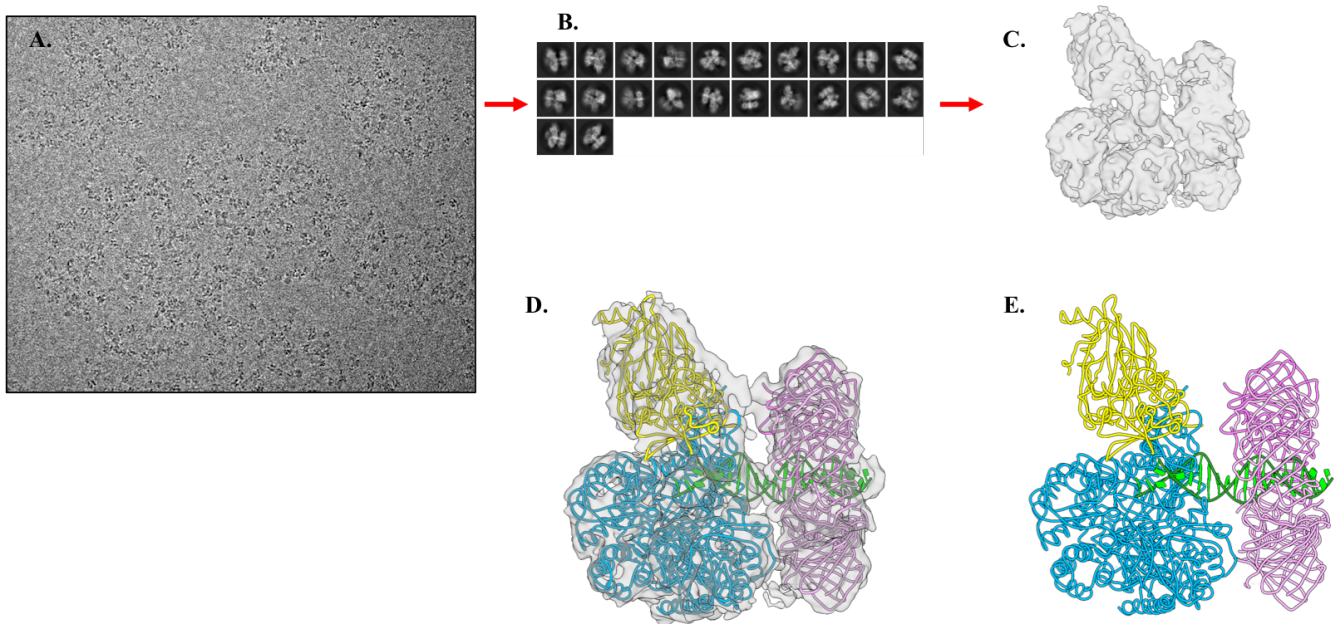


Figure 26. Structure du complexe PolD-ADN-PCNA de *Pyrococcus abyssi* en mode élongation en cryo-microscopie électronique (en cours de résolution au laboratoire). (A) Image du complexe sur une grille de cryo-microscopie électronique. (B) Classes 2D sélectionnées qui représentent le complexe. (C) Carte de densité calculée du complexe. (D) Structure en cours de résolution dans la densité. La petite sous-unité DP1 de la PolD est colorée en jaune, la grande sous-unité DP2 de la PolD en bleu, l'ADN en vert et le PCNA en rose. (E) Structure du complexe en cours d'affinement.

5. Le PCNA, véritable maestro de la fourche de réplication

Le PCNA, souvent considéré comme le chef d'orchestre de la fourche de réplication, interagit avec un grand nombre de protéines impliquées dans de multiples mécanismes cellulaires vitaux, comme cela a déjà été montré chez les Eucaryotes : les polymérases, les protéines en charge de la réparation de l'ADN, les régulateurs du cycle cellulaire, les facteurs d'assemblage et de remodelage de la chromatine, les protéines de cohésion des chromatides sœurs, les facteurs de transcription et de traduction, des enzymes du métabolisme et même des protéines de la membrane et du cytosquelette (71). Lors de la réplication, la progression de la fourche de réplication peut être bloquée suite à des lésions sur l'ADN entraînées par des dommages endogènes ou l'action d'agents extérieurs (comme des anti-cancéreux). La réponse cellulaire aboutit le plus souvent à un changement de la fonction du PCNA, déclenché soit par des modifications post-traductionnelles spécifiques, soit par des changements d'interaction avec ses différents partenaires. Au cœur de ces mécanismes, le PCNA tient un rôle central dans la détermination du sort de la fourche de réplication et, par la même occasion, de la progression d'une tumeur. L'étude facilitée de ces mécanismes chez les protéines très stables d'Archées hyperthermophiles, comme notre modèle *Pyrococcus abyssi*, permettrait de mieux comprendre leur fonctionnement chez les Eucaryotes, et notamment chez l'être humain.

II. Perspectives

Dans la continuité du projet présenté, plusieurs travaux sont actuellement en cours ou envisagés au laboratoire. L'étude du réplisome de *Pyrococcus abyssi* s'étend désormais à d'autres acteurs clés de la fourche de réplication : d'un côté la primase, qui joue un rôle central dans le complexe d'initiation de la réplication, et d'un autre côté le complexe formé par l'hélicase MCM et les protéines GINS et Cdc45 (GAN chez les Archées), éléments indispensables du complexe d'élongation.

1. Sur le complexe d'initiation

La primase est un des acteurs clés du réplisome : en effet, la PolD est incapable d'initier la synthèse d'ADN *de novo* et a besoin d'une amorce ARN synthétisée par la primase. Chez les Archées, elle est composée de deux sous-unités : une petite sous-unité PriS, portant les résidus catalytiques, et une sous-unité plus large appelée PriL, régulant son activité. Le dimère PriS/PriL est conservé chez toutes les Archées et partage une homologie avec la primase eucaryote, également appelée PriS/PriL. À noter que dans le groupe des Sulfolobales (des

Crenarchaeota), le dimère PriS/PriL est associé à une troisième sous-unité régulatrice, PriX, et dans le groupe des Nanoarchaeales (des DPANN), PriS et PriL sont fusionnées (72).

Peu avant mon arrivée dans l'équipe, la structure de PriS en complexe avec PriL déléguée de son domaine C-terminal a été déterminée par cristallographie aux rayons X à 1,6 Å de résolution au laboratoire (Figure 27). Ce domaine C-terminal contient un assemblage organisé d'atomes de fer et de soufre, appelé cluster Fer-Soufre, sensible à l'oxydation. Cela rend donc la purification de la primase entière très difficile en présence d'oxygène, ce cluster ayant tendance à s'oxyder et à déstabiliser le domaine C-terminal de PriL qui le contient. J'ai cependant purifié la primase entière en conditions aérobie et obtenu des cristaux (Figure 28 A). Si le centre Fer-Soufre est intact, ces travaux permettront de décrire pour la première fois la structure d'une primase hétérodimérique d'Euryarchées (Figure 28 B).

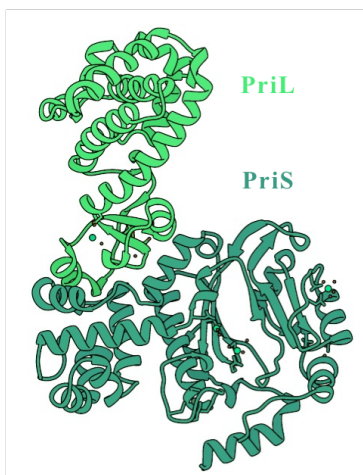
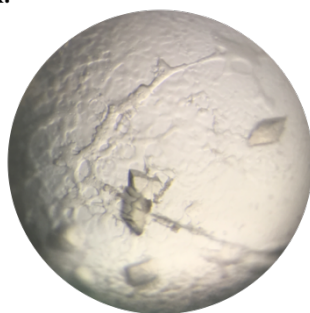


Figure 27. Structure de PriS en complexe avec PriL sans son domaine C-terminal obtenue par cristallographie aux rayons X à 1,6 Å de résolution. La grande sous-unité PriL sans son domaine C-terminal est colorée en vert clair tandis que la petite sous-unité PriS est colorée en vert foncé.

A.



B.

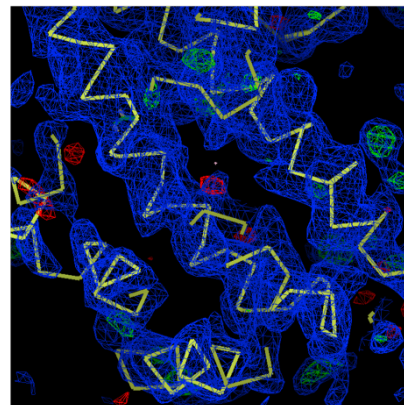


Figure 28. Résolution de structure de la primase entière de *Pyrococcus abyssi* en cours. (A) Cristaux obtenus au bout de 4 jours à la plateforme de cristallographie de l'Institut Pasteur observés au microscope. (B) Densité électronique (en bleu) obtenue après diffraction des cristaux au Synchrotron SOLEIL et traitement des jeux de données obtenus. Des hélices α peuvent déjà être observées mais des cycles d'affinement et de reconstruction seront nécessaires pour déterminer la présence ou non du cluster Fer-Soufre.

De nouveaux résultats obtenus par *bio-layer interferometry* au laboratoire suggèrent que l'interaction entre la PolD et la primase serait dépendante de la présence d'un motif PIP. La question se pose alors de savoir comment se déroule la coopération entre le PCNA et la primase au sein de la fourche de réplication.

2. Sur le complexe d'élongation

Les homologies entre les protéines d'Archées et d'Eucaryotes se retrouvent également au niveau du complexe hélicase CMG. Pour rappel, l'hélicase MCM requiert le recrutement des protéines Cdc45 (*Cell Division Control protein 45*) et GINS (*Go-Ichi-Nii-San*) pour former un complexe hélicase actif, appelé CMG (Cdc45-MCM-GINS). Ce complexe CMG est capable d'ouvrir la double hélice d'ADN et de conduire la progression de la fourche de réplication le long du génome. Chez les Archées, MCM est un homo-hexamère, tandis qu'il s'agit d'un hétéro-hexamère chez les Eucaryotes. Les deux structures demeurent cependant très proches, à tel point que la structure cristallographique du MCM d'Archées a servi de modèle pour l'étude du MCM dans le réplisome de levures. De la même manière, les structures du complexe GINS sont très proches chez les Archées et les Eucaryotes. Chez les Archées, GINS est associé à une exonucléase appelée GAN (GINS-Associated Nuclease) qui pourrait constituer un intermédiaire entre la protéine bactérienne RecJ et la protéine eucaryote Cdc45. Elle semble avoir un rôle, avec GINS, dans l'activation du complexe MCM, et participer aux mécanismes de réparation de l'ADN (73).

Un des objectifs futurs du laboratoire serait d'étudier la structure du complexe entre l'hélicase, GAN, GINS, la PolD, le PCNA et l'ADN afin de comprendre de quelle façon tous ces acteurs de la réplication coopèrent pour mener à bien ce mécanisme complexe et néanmoins indispensable à la vie. Ces travaux permettraient d'aller encore plus loin dans la compréhension de la réplication chez les Archées et de l'histoire évolutive des trois domaines du vivant tout en développant un modèle d'étude simplifié extrapolable aux mécanismes eucaryotes.

CONCLUSION

L'évolution rapide, au cours des dernières décennies, des différentes approches de biologie structurale permet aujourd'hui d'espérer faire le lien entre les événements ayant lieu aux différentes échelles du vivant. Cette démarche implique de déterminer la structure des biomolécules seules et en complexe avec leurs partenaires, d'étudier la dynamique et l'énergétique de leurs interactions et de leurs réactions ainsi que leur organisation spatio-temporelle dans la cellule.

Au cours de mes travaux, j'ai utilisé cette approche intégrative pour étudier la PolD et le PCNA de *Pyrococcus abyssi* sous plusieurs angles. Tout d'abord en travaillant avec le PCNA seul, en l'exprimant et en le purifiant, avec pour objectif d'obtenir des cristaux pour la détermination de sa structure par cristallographie aux rayons X. Peu après, j'ai exprimé et purifié différentes constructions de la PolD, dont la structure entière a récemment été résolue par microscopie électronique dans le laboratoire. Disposant de ces données structurales, j'ai ensuite étudié les deux protéines dans le cadre de leur complexation afin d'établir une cartographie de leurs interactions et d'être en mesure de proposer un modèle du complexe ternaire qu'elles forment avec l'ADN. Cette étude ouvre la voie à plusieurs perspectives.

Dans un premier temps, la structure du complexe ternaire PolD-ADN-PCNA déterminée en cryo-microscopie électronique permettra d'aller encore plus loin dans la compréhension de leurs mécanismes d'interaction. Les données structurales viendront compléter les données d'interaction obtenues par interférométrie à bio-couches pour cibler les régions impliquées. Par la suite, l'étude d'autres acteurs clés de la fourche de réplication comme GINS et la primase, permettrait d'avoir une vue d'ensemble du réplisome. Comprendre la façon dont ces mécanismes complexes sont orchestrés chez les Archées paverait la voie à une meilleure compréhension des mécanismes eucaryotes dont ils sont proches. De plus, l'étude du réplisome des Archées hyperthermophiles présente un grand intérêt sur le plan biotechnologique. Les propriétés des enzymes thermostables peuvent en effet être utilisées pour améliorer les performances de réplication de l'ADN in vitro dans les techniques de PCR.

BIBLIOGRAPHIE

1. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A*. 1990 Jun;87(12):4576–9.
2. Olsen GJ, Woese CR, Overbeek R. The winds of (evolutionary) change: breathing new life into microbiology. *J Bacteriol*. 1994 Jan;176(1):1–6.
3. Crick FH. On protein synthesis. *Symp Soc Exp Biol*. 1958;12:138–63.
4. Crick F. Central dogma of molecular biology. *Nature*. 1970 Aug 8;227(5258):561–3.
5. Delarue M, Poch O, Tordo N, Moras D, Argos P. An attempt to unify the structure of polymerases. *Protein Eng*. 1990 May;3(6):461–7.
6. Ito J, Braithwaite DK. Yeast mitochondrial DNA polymerase is related to the family A DNA polymerases. *Nucleic Acids Res*. 1990;18(22):6716–6716.
7. Burgers PM, Koonin EV, Bruford E, Blanco L, Burtis KC, Christman MF, et al. Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J Biol Chem*. 2001 Nov 23;276(47):43487–90.
8. Grohmann D, Werner F. Recent advances in the understanding of archaeal transcription. *Curr Opin Microbiol*. 2011 Jun;14(3):328–34.
9. Petrov AS, Gulen B, Norris AM, Kovacs NA, Bernier CR, Lanier KA, et al. History of the ribosome and the origin of translation. *Proc Natl Acad Sci*. 2015 Dec 15;112(50):15396–401.
10. Filée J, Forterre P, Sen-Lin T, Laurent J. Evolution of DNA Polymerase Families: Evidences for Multiple Gene Exchange Between Cellular and Viral Proteins. *J Mol Evol*. 2002 Jun;54(6):763–73.
11. Imamura M, Uemori T, Kato I, Ishino Y. A non-alpha-like DNA polymerase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biol Pharm Bull*. 1995 Dec;18(12):1647–52.
12. Uemori T, Sato Y, Kato I, Doi H, Ishino Y. A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. *Genes Cells Devoted Mol Cell Mech*. 1997 Aug;2(8):499–512.
13. Sauguet L, Raia P, Henneke G, Delarue M. Shared active site architecture between archaeal PolD and multi-subunit RNA polymerases revealed by X-ray crystallography. *Nat Commun*. 2016;7:12227.
14. Raia P, Carroni M, Henry E, Pehau-Arnaudet G, Brûlé S, Béguin P, et al. Structure of the DP1-DP2 PolD complex bound with DNA and its implications for the evolutionary history of DNA and RNA polymerases. *PLoS Biol*. 2019 Jan;17(1):e3000122.
15. Raia P, Delarue M, Sauguet L. An updated structural classification of replicative DNA polymerases. *Biochem Soc Trans*. 2019 28;47(1):239–49.

16. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*. 1977 Nov;74(11):5088–90.
17. Magrum LJ, Luehrsen KR, Woese CR. Are extreme halophiles actually “bacteria”? *J Mol Evol*. 1978 May 12;11(1):1–8.
18. Hafenbradl D, Keller M, Dirmeier R, Rachel R, Rossnagel P, Burggraf S, et al. *Ferroglobus placidus* gen. nov., sp. nov., A novel hyperthermophilic archaeum that oxidizes Fe²⁺ at neutral pH under anoxic conditions. *Arch Microbiol*. 1996 Nov;166(5):308–14.
19. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. *Nat Microbiol*. 2016 Apr 11;1:16048.
20. Castelle CJ, Banfield JF. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. *Cell*. 2018 08;172(6):1181–97.
21. DeLong EF, Pace NR. Environmental diversity of bacteria and archaea. *Syst Biol*. 2001 Aug;50(4):470–8.
22. Fiala G, Stetter KO. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C. *Arch Microbiol*. 1986 Jun;145(1):56–61.
23. Erauso G, Reysenbach A-L, Godfroy A, Meunier J-R, Crump B, Partensky F, et al. *Pyrococcus abyssi* sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch Microbiol* [Internet]. 1993 Nov [cited 2019 May 22];160(5). Available from: <http://link.springer.com/10.1007/BF00252219>
24. Blöchl E, Rachel R, Burggraf S, Hafenbradl D, Jannasch HW, Stetter KO. *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extrem Life Extreme Cond*. 1997 Feb;1(1):14–21.
25. López-García P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*. 2001 Feb;409(6820):603–7.
26. Ciaramella M, Napoli A, Rossi M. Another extreme genome: how to live at pH 0. *Trends Microbiol*. 2005 Feb;13(2):49–51.
27. DeLong EF. Everything in moderation: archaea as “non-extremophiles.” *Curr Opin Genet Dev*. 1998 Dec;8(6):649–54.
28. Eckburg PB, Lepp PW, Relman DA. Archaea and their potential role in human disease. *Infect Immun*. 2003 Feb;71(2):591–6.
29. Chaudhury P, Quax TEF, Albers S-V. Versatile cell surface structures of archaea: Cell surface structures of archaea. *Mol Microbiol*. 2018 Feb;107(3):298–311.
30. Koonin EV, Wolf YI. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res*. 2008 Oct 23;36(21):6688–719.

31. Peeters E, Driessen RPC, Werner F, Dame RT. The interplay between nucleoid organization and transcription in archaeal genomes. *Nat Rev Microbiol*. 2015 Jun;13(6):333–41.
32. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*. 2015 May;521(7551):173–9.
33. Barry ER, Bell SD. DNA Replication in the Archaea. *Microbiol Mol Biol Rev*. 2006 Dec 1;70(4):876–87.
34. Kelman LM, Kelman Z. Archaeal DNA Replication. *Annu Rev Genet*. 2014 Nov 23;48(1):71–97.
35. Ausiannikava D, Allers T. Diversity of DNA Replication in the Archaea. *Genes*. 2017 Jan 31;8(2):56.
36. Nagata M, Ishino S, Yamagami T, Ogino H, Simons J-R, Kanai T, et al. The Cdc45/RecJ-like protein forms a complex with GINS and MCM, and is important for DNA replication in *Thermococcus kodakarensis*. *Nucleic Acids Res*. 2017 Oct 13;45(18):10693–705.
37. Bell SD. Initiation of DNA Replication in the Archaea. In: Masai H, Foiani M, editors. *DNA Replication* [Internet]. Singapore: Springer Singapore; 2017 [cited 2019 May 22]. p. 99–115. Available from: http://link.springer.com/10.1007/978-981-10-6955-0_5
38. Cann IKO, Ishino S, Yuasa M, Daiyasu H, Toh H, Ishino Y. Biochemical Analysis of Replication Factor C from the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *J Bacteriol*. 2001 Apr 15;183(8):2614–23.
39. Wu S, Beard WA, Pedersen LG, Wilson SH. Structural Comparison of DNA Polymerase Architecture Suggests a Nucleotide Gateway to the Polymerase Active Site. *Chem Rev*. 2014 Mar 12;114(5):2759–74.
40. Ollis DL, Brick P, Hamlin R, Xuong NG, Steitz TA. Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature*. 1985 Mar 28;313(6005):762–6.
41. Bailey S, Wing RA, Steitz TA. The Structure of *T. aquaticus* DNA Polymerase III Is Distinct from Eukaryotic Replicative DNA Polymerases. *Cell*. 2006 Sep;126(5):893–904.
42. Gueguen Y, Rolland J, Lecompte O, Azam P, Le Romancer G, Flament D, et al. Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*: DNA polymerases from *Pyrococcus abyssi*. *Eur J Biochem*. 2001 Nov 15;268(22):5961–9.
43. Shen Y, Musti K, Hiramoto M, Kikuchi H, Kawarabayashi Y, Matsui I. Invariant Asp-1122 and Asp-1124 Are Essential Residues for Polymerization Catalysis of Family D DNA Polymerase from *Pyrococcus horikoshii*. *J Biol Chem*. 2001 Jul 20;276(29):27376–83.
44. Shen Y. A 21-amino acid peptide from the cysteine cluster II of the family D DNA polymerase from *Pyrococcus horikoshii* stimulates its nuclease activity which is Mre11-

- like and prefers manganese ion as the cofactor. *Nucleic Acids Res.* 2004 Jan 2;32(1):158–68.
45. Jokela M. Characterization of the 3' exonuclease subunit DP1 of *Methanococcus jannaschii* replicative DNA polymerase D. *Nucleic Acids Res.* 2004 Apr 28;32(8):2430–40.
 46. Matsui I, Urushibata Y, Shen Y, Matsui E, Yokoyama H. Novel structure of an N-terminal domain that is crucial for the dimeric assembly and DNA-binding of an archaeal DNA polymerase D large subunit from *Pyrococcus horikoshii*. *FEBS Lett.* 2011 Feb 4;585(3):452–8.
 47. Yamasaki K, Urushibata Y, Yamasaki T, Arisaka F, Matsui I. Solution structure of the N-terminal domain of the archaeal D-family DNA polymerase small subunit reveals evolutionary relationship to eukaryotic B-family polymerases. *FEBS Lett.* 2010 Aug 4;584(15):3370–5.
 48. Yao NY, O'Donnell ME. Evolution of replication machines. *Crit Rev Biochem Mol Biol.* 2016 May 3;51(3):135–49.
 49. Aravind L, Koonin EV. Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res.* 1998 Aug 15;26(16):3746–52.
 50. Tahirov TH, Makarova KS, Rogozin IB, Pavlov YI, Koonin EV. Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol ϵ and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol Direct.* 2009;4(1):11.
 51. Makarova KS, Krupovic M, Koonin EV. Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Front Microbiol* [Internet]. 2014 Jul 21 [cited 2019 May 22];5. Available from: <http://journal.frontiersin.org/article/10.3389/fmicb.2014.00354/abstract>
 52. Ishino S, Ishino Y. DNA polymerases as useful reagents for biotechnology – the history of developmental research in the field. *Front Microbiol* [Internet]. 2014 Aug 29 [cited 2019 May 22];5. Available from: <http://journal.frontiersin.org/article/10.3389/fmicb.2014.00465/abstract>
 53. Terpe K. Overview of thermostable DNA polymerases for classical PCR applications: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol.* 2013 Dec;97(24):10243–54.
 54. Killelea T, Ralec C, Boss A, Henneke G. PCR performance of a thermostable heterodimeric archaeal DNA polymerase. *Front Microbiol* [Internet]. 2014 May 7 [cited 2019 May 22];5. Available from: <http://journal.frontiersin.org/article/10.3389/fmicb.2014.00195/abstract>
 55. Zhuang Z, Ai Y. Processivity factor of DNA polymerase and its expanding role in normal and translesion DNA synthesis. *Biochim Biophys Acta BBA - Proteins Proteomics.* 2010 May;1804(5):1081–93.
 56. Krishna TS, Kong XP, Gary S, Burgers PM, Kuriyan J. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell.* 1994 Dec 30;79(7):1233–43.

57. Williams GJ, Johnson K, Rudolf J, McMahon SA, Carter L, Oke M, et al. Structure of the heterotrimeric PCNA from *Sulfolobus solfataricus*. Acta Crystallograph Sect F Struct Biol Cryst Commun. 2006 Oct 1;62(10):944–8.
58. Matsumiya S, Ishino Y, Morikawa K. Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from *Pyrococcus furiosus*. Protein Sci Publ Protein Soc. 2001 Jan;10(1):17–23.
59. O'Donnell M, Kuriyan J, Kong XP, Stukenberg PT, Onrust R. The sliding clamp of DNA polymerase III holoenzyme encircles DNA. Mol Biol Cell. 1992 Sep;3(9):953–7.
60. Patoli AA, Winter JA, Bunting KA. The UmuC subunit of the *E. coli* DNA polymerase V shows a unique interaction with the β -clamp processivity factor. BMC Struct Biol. 2013;13(1):12.
61. Boehm EM, Washington MT. R.I.P. to the PIP: PCNA-binding motif no longer considered specific: PIP motifs and other related sequences are not distinct entities and can bind multiple proteins involved in genome maintenance. BioEssays. 2016 Nov;38(11):1117–22.
62. Henneke G, Flament D, Hübscher U, Querellou J, Raffin J-P. The Hyperthermophilic Euryarchaeota *Pyrococcus abyssi* Likely Requires the Two DNA Polymerases D and B for DNA Replication. J Mol Biol. 2005 Jul;350(1):53–64.
63. Castrec B, Rouillon C, Henneke G, Flament D, Querellou J, Raffin J-P. 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. 2009 Nov;394(2):209–18.
64. Cann IK, Ishino S, Hayashi I, Komori K, Toh H, Morikawa K, et al. Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. J Bacteriol. 1999 Nov;181(21):6591–9.
65. Tori K, Kimizu M, Ishino S, Ishino Y. DNA Polymerases BI and D from the Hyperthermophilic Archaeon *Pyrococcus furiosus* Both Bind to Proliferating Cell Nuclear Antigen with Their C-Terminal PIP-Box Motifs. J Bacteriol. 2007 Aug 1;189(15):5652–7.
66. Tang X-F, Shen Y, Matsui E, Matsui I. Domain Topology of the DNA Polymerase D Complex from a Hyperthermophilic Archaeon *Pyrococcus horikoshii* †. Biochemistry. 2004 Sep;43(37):11818–27.
67. Fernandez-Leiro R, Conrad J, Scheres SH, Lamers MH. cryo-EM structures of the *E. coli* replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and τ . eLife [Internet]. 2015 Oct 24 [cited 2019 May 22];4. Available from: <https://elifesciences.org/articles/11134>
68. Mayanagi K, Kiyonari S, Nishida H, Saito M, Kohda D, Ishino Y, et al. Architecture of the DNA polymerase B-proliferating cell nuclear antigen (PCNA)-DNA ternary complex. Proc Natl Acad Sci. 2011 Feb 1;108(5):1845–9.
69. Matsumiya S, Ishino S, Ishino Y, Morikawa K. *Pyrococcus furiosus* PCNA complexed with RFCL PIP-box peptide. (:unav) [Internet]. 2002 Oct 23 [cited 2019 May 22]; Available from: <ftp://ftp.wwpdb.org/pub/pdb/data/structures/divided/pdb/is/pdb1isq.ent.gz>

70. Nishida H, Mayanagi K, Kiyonari S, Sato Y, Oyama T, Ishino Y, et al. Structural determinant for switching between the polymerase and exonuclease modes in the PCNA-replicative DNA polymerase complex. *Proc Natl Acad Sci U S A*. 2009 Dec 8;106(49):20693–8.
71. Stoimenov I, Helleday T. PCNA on the crossroad of cancer. *Biochem Soc Trans*. 2009 Jun;37(Pt 3):605–13.
72. Cooper CDO. Archaeal DNA polymerases: new frontiers in DNA replication and repair. *Emerg Top Life Sci*. 2018 Dec 14;2(4):503–16.
73. Oyama T, Ishino S, Shirai T, Yamagami T, Nagata M, Ogino H, et al. Atomic structure of an archaeal GAN suggests its dual roles as an exonuclease in DNA repair and a CMG component in DNA replication. *Nucleic Acids Res*. 2016 Nov 2;44(19):9505–17.

RÉSUMÉ

Dans toutes les formes de la vie, les ADN polymérase jouent un rôle central dans la réplication du génome, sa maintenance et sa réparation. Ce sont également des outils précieux nécessaires à de nombreuses applications en biologie moléculaire. Toutes les ADN polymérase ont été classées en sept familles distinctes en se basant sur leur homologie de séquence : PolA, PolB, PolC, PolD, PolX, PolY et les reverse transcriptases. L'ADN polymérase de la famille D, appelée PolD et dont la structure a récemment été résolue, demeure la moins bien caractérisée. Elle a été découverte chez des Archées thermophiles, et présente des propriétés biochimiques intéressantes, exploitables en PCR (réaction de polymérisation en chaîne). C'est une protéine hétérodimérique composée d'une sous-unité DP1 responsable de l'activité de relecture exonucléase, et d'une sous-unité DP2 responsable de l'activité de polymérisation. Les structures de DP1 et DP2 révèlent que la PolD est une ADN polymérase atypique. En effet, la structure des sites actifs polymérase et exonucléase diffère de celle des autres ADN polymérase caractérisées à ce jour. Le PCNA (*Proliferating Cell Nuclear Antigen*) est un acteur clé de la fourche de réplication qui améliore considérablement la processivité de la PolD à la fois *in vivo* et *in vitro*. Au cours de mes travaux, j'ai pu étudier la PolD et le PCNA de l'Archée *Pyrococcus abyssi* sous plusieurs angles. Dans un premier temps, j'ai travaillé sur le PCNA seul pour déterminer sa structure par cristallographie. Peu après, j'ai exprimé et purifié différentes constructions de la PolD afin d'étudier les deux protéines dans le cadre de leur complexation en utilisant la méthode d'interférométrie à bio-couches. Le croisement des données obtenues m'a permis d'établir une cartographie de leurs interactions et de proposer un modèle du complexe ternaire qu'elles forment avec l'ADN.

[POLD, AN ATYPICAL ARCHAEAL DNA POLYMERASE AND A PROMISING TOOL FOR BIOTECHNOLOGIES]

ABSTRACT

In all forms of life, DNA polymerases play a central role in genome replication, maintenance and repair. They are also valuable tools needed for many applications in molecular biology. All DNA polymerases were classified into seven distinct families based on their sequence homology: PolA, PolB, PolC, PolD, PolX, PolY and reverse transcriptases. The DNA polymerase family D, called PolD and whose structure has recently been resolved, remains the least well characterized. It has been discovered in thermophilic Archaea, and has interesting biochemical properties, exploitable by PCR (*Polymerase Chain Reaction*). It is an heterodimeric protein composed of a DP1 subunit responsible for the exonuclease activity, and a DP2 subunit responsible for the polymerization activity. The structures of DP1 and DP2 reveal that PolD is an atypical DNA polymerase. Indeed, the structure of the polymerase and exonuclease active sites differs from that of the other DNA polymerases characterized to date. PCNA (*Proliferating Cell Nuclear Antigen*) is a key player in the replication fork that dramatically improves the processivity of PolD both *in vivo* and *in vitro*. In this work, I was able to study the PolD and the PCNA of the Archaea *Pyrococcus abyssi* from several angles. At first, I worked on the PCNA alone to determine its structure by crystallography. Shortly after, I expressed and purified different PolD constructs in order to study the two proteins as part of their complexation using the bio-layer interferometry method. Crossing the data obtained allowed me to map their interactions and to propose a model of the ternary complex they form with DNA.

Mots clés : PolD, ADN polymérase, PCNA, réplication, réplisome, cristallographie aux rayons X, biologie structurale

Discipline : Pharmacie

Intitulé et adresse du laboratoire : Unité Dynamique Structurale des Macromolécules, UMR 3528 (CNRS)
Institut Pasteur, 25 rue du Docteur Roux,
75015 Paris