

Etude structurale et fonctionnelle des complexes multi-protéiques de la voie de réparation NHEJ chez l'homme

Jérémy Amram

▶ **To cite this version:**

Jérémy Amram. Etude structurale et fonctionnelle des complexes multi-protéiques de la voie de réparation NHEJ chez l'homme. Biochimie [q-bio.BM]. Université Paris Sud - Paris XI, 2015. Français. ⟨NNT : 2015PA114822⟩. ⟨tel-01552301v2⟩

HAL Id: tel-01552301

https://theses.hal.science/tel-01552301v2

Submitted on 4 Jul 2017

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.



HAL Authorization

UNIVERSITÉ PARIS-SUD 11

FACULTÉ DE PHARMACIE DE CHÂTENAY-MALABRY

ECOLE DOCTORALE :

INNOVATION THÉRAPEUTIQUE : DU FONDAMENTAL A L'APPLIQUÉ

PÔLE : INGENIERIE DES PROTEINES ET CIBLES THERAPEUTIQUES

ANNÉE 2014 - 2015 SÉRIE DOCTORAT N°

THÈSE

Présentée

À L'UNITÉ DE FORMATION ET DE RECHERCHE
FACULTE DE PHARMACIE DE CHATENAY-MALABRY
UNIVERSITÉ PARIS-SUD 11

pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ PARIS-SUD 11

par

Jeremy AMRAM

Titre de la thèse :

ETUDE STRUCTURALE ET FONCTIONNELLE DES COMPLEXES MULTI-PROTEIQUES DE LA VOIE DE REPARATION NHEJ CHEZ L'HOMME

Soutenue le 2 Juillet 2015

JURY :	M. Philippe Minard	Président
	M. Marc Delarue	Rapporteur
	M. Bertrand Castaing	Rapporteur
	M. Sophie Quevillon-Cheruel	Examinatrice
	M. Xavier Veauté	Examinateur
	M. Jean-Baptiste Charbonnier	Directeur de thèse

REMERCIEMENTS

Cette thèse s'est déroulée au Laboratoire de Biologie Structurale et de Radiobiologie au sein du SB²SM\iBiTec-S et j'aimerais remercier la directrice du laboratoire Sophie Zinn-Justin de m'avoir accueilli au sein du laboratoire. Elle a été financée par une bourse ministérielle. La prolongation de thèse a été financée par la Fondation pour la Recherche Médicale.

Ma thèse a été dirigée et encadrée par le Dr Jean-Baptiste Charbonnier, qui a porté ce projet avant moi et le poursuivra après moi et je tiens à le remercier pour ses conseils précieux et son grand soutien lors de ces trois années et demie. Je suis extrêmement reconnaissant pour les efforts qu'il a entrepris pour valoriser mon travail et pour la motivation et la détermination qu'il a sollicitée pour mener à bien nos projets.

J'aimerais remercier le Professeur Philippe Minard d'avoir accepté de présider le Jury de ma soutenance ainsi que le Dr Marc Delarue et le Dr Bertrand Castaing qui ont accepté d'évaluer mon travail et je les remercie tout particulièrement pour le temps qu'ils ont passé à la lecture de mon manuscrit. Enfin je souhaiterais remercier le Dr Sophie Quevillon-cheruel et le Dr Xavier Veaute pour avoir accepté d'évaluer mon travail en temps qu'examinateurs.

De nombreuses personnes ont contribué à ce projet et je souhaiterais remercier le Dr Pascal Drevet, qui m'a accompagné lors de nombreuses étapes cruciales. Ses conseils et son expérience m'ont permis d'éviter de commettre de nombreuses erreurs. Le Dr Virginie Ropars qui m'a permis de commencer à travailler sur ce projet avec des bases solides et en me transmettant le plus d'informations possibles pour mener le mien à bien. Enfin, j'aimerais témoigner ma reconnaissance auprès de toutes les autres personnes qui ont collaboré avec moi lors de cette thèse, Laurie Malherbe, Cyril Bigot, Audrey Comte, et Clément Némoz.

J'aimerais aussi remercier les équipes d'Eric LeCam à l'IGR et de Jean-Pierre de Villartay à l'institut Imagine, pour les nombreuses interactions que nous avons pu avoir lors de nos collaborations.

INTRODUCTION

1. Formation accidentelle de cassures double-brin 8
1.1 Dommages induits par les rayonnements ionisants et les espèces réactives de l'oxygène 9
1.2 Dommages induits par les DNA glycosylases de la voie BER 9
1.3 Pontages covalents de type topoisomérases-ADN 10
1.4 Dommages aux extrémités créés par les ligases 12
1.5 Intermédiaires dérivés de l'introduction de ribonucléotides dans la structure de l'ADN 13
2. Enzymes de réparation des extrémités de CDBs en vue de leur ligature 14
2.1 Retrait d'un 5'-adenylate par l'Aprataxin (APTX) 14
2.2 Polynucléotide Kinase Phosphatase (PNK) 16
2.3 Tyrosyl DNA Phosphodiesterase 1 (TDP1) 16
2.4 Tyrosyl DNA Phosphodiesterase 2 (TDP2) 17
2.5 DNA Polymerase β et activité 5'dRP-lyase et AP-lyase 17
2.6 Ku70/80, activité 5'dRP-lyase et AP-lyase 18
3. Cassures programmées de l'ADN 18
3.1 Cassures programmées dans les lymphocytes : V(D)J et CSR 18
3.2 Cassures programmées lors de la recombinaison méiotique 19
4. Voies de réparation des cassures double-brin 22
4.1 Voie NHEJ classique 22
4.2 Voie de réparation par Recombinaison Homologue (HR) 31
4.3 Voie NHEJ alternative (a-NHEJ) 34
5 Caractéristiques des protéines de la voie NHEJ 35
5.1 Hétérodimère Ku70/Ku80 35
5.2 DNA-PKcs 38
5.3 Artemis 41
5.4 La Ligase4 43
5.5 XRCC4 et Cernunnos/XLF 45
5.6 ADN Polymérases X 52
6. Maladies associées à des mutations entrainant un dysfonctionnement de la voie NHEJ 54
6.1 Patients présentant des mutations sur la Ligase 4 54
6.2 Patients présentant des mutations sur Artemis 55
6.3 Patients présentant des mutations sur Cernunnos/XLF 56
6.4 Patients présentant des mutations sur DNA-PKcs 56
7. Etat de l'art de l'expression en cellules d'insecte 57
7.1 Système MultiBac 58
7.2 Structure cristalline de la tête du Médiateur 60
8 Objectifs de la thèse 63

MATERIEL ET METHODES 66

1. Bioinformatique 66
2. Expression en cellules d'insecte des protéines de la voie NHEJ 68
2.1 Principes généraux 68
2.2 Clonages 71
2.3 Entretien des cellules d'insecte 74
2.4 Préparation des Virus 75
2.6 Cultures des cellules d'insecte 78
3. Expression dans *E. coli* de XRCC4, Cernunnos et du domaine CTD de la Ligase4 79
4. Purifications et co-purifications des protéines exprimées 80
4.1 Purifications de XRCC4/Ligase4 80
4.2 Purifications de Ku70-Ku80 81

4.3 Purification de Cernunnos....82
4.4 Purifications de Artemis....83
4.5 Co-purifications....84
4.5 Purification de XRCC4 et de Cernunnos entiers produits en *E.coli*....84
4.6 Purification du complexe Ligase4(CTD)-XRCC4(1-203) et de Cernunnos(1-224) produits en *E.coli*..86
5. Analyses biochimiques....87
5.1 Chromatographies d'exclusions de taille....87
5.2 Caractérisation des modifications post-traductionnelles....87
5.3 Mesure sur gels SDS-PAGE de la stoechiométrie des protéines au sein d'un complexe....88
6. Caractérisations Biophysiques....88
6.1 Caractérisation des interactions par Calorimétrie....88
7. Etudes structurales....91
7.1 Cristallisation des protéines seules et en complexes....91
7.2 Analyse des complexes par SAXS....94
7.3 Etudes de microscopie électronique....94

RESULTATS....97
1. Expression des protéines et complexes du NHEJ en cellules d'insecte....97
1.1. Stratégie de clonage....97
1.2 Prise en main de la technique....98
2. Purification des protéines et complexes de la voie NHEJ....100
2.1 Purifications de Ligase4-XRCC4....101
2.2 Purifications de Ku70-Ku80....103
2.3 Purification de Cernunnos....106
2.4 Purifications d'Artemis....108
2.5 Co-purifications....110
2.6 Les améliorations apportées au laboratoire au système de production en MultiBac....111
2.7 Purifications de XRCC4 et Cernunnos entiers en *E.coli*....114
3. Caractérisations biochimiques....115
3.1. Chromatographies d'exclusion de taille analytique....115
3.2 Analyses sur l'état de phosphorylation....120
3.3 Analyses sur gel SDS-PAGE de la stoechiométrie des complexes....122
4. Mesures d'interactions protéine-protéine par calorimètrie....124
4.1 Mesures de l'interaction entre XRCC4 et Cernunnos entiers....124
4.2 Mesures de l'interaction entre un complexe Ligase4(CTD)-XRCC4 et Cernunnos....124
5. Essais de cristallisation....126
6. Etude par diffusion des rayons X aux petits angles (SAXS)....128
7. Etudes de microscopie électronique....130
7.1 Filaments formés par les protéines XRCC4 et Cernunnos entières produites dans E. coli....130
7.2 Analyses des filaments formés par XRCC4 entier produit dans E. coli....131
7.3 Analyses des filaments formés par XRCC4 et Cernunnos entiers produits dans *E. coli* en présence d'ADN....131
7.4 Analyses des interactions Ku70/Ku80-ADN....132

CONCLUSION ET PERSPECTIVES....134
1. Mise en place de la plateforme de production en cellules d'insecte....135
2. Production en cellules d'insecte des protéines de la voie NHEJ....136
3. Le cas DNA-PKcs....137
4. Etude du réseau d'interactions médiées par Ku70/Ku80....138
5. Dynamique des filaments XRCC4-Cernunnos....139

ANNEXES.....140
Annexe 1 : Protocole de Transfection de cellules d'insecte.....140
Annexe 2 : Réalisation d'un stock de V1 142
Annexe 3 : Table de McGrady 143
Annexe 4 : Purification XRCC4-10His/Ligase4 143
Annexe 5 : Purification Ku70/10His-Ku80 de cultures en cellules d'insecte.....145
Annexe 6 : Purification Cernunnos sans étiquette.....148
Annexe 7 : Purification Artemis 10His.....139

BIBLIOGRAPHIE.....151

FIGURES PARTIE I - INTRODUTION

Figure 1 : Variété des structures chimiques trouvées aux extrémités de cassures simple-brin (CSB) ou double-brin (CDB) de l'ADN.....11
Figure 2 : Mécanismes enzymatiques de préparation des extrémités des CDBs en vue de leur ligature. 12
Figure 3 : Mécanismes enzymatiques directs de préparation des extrémités de CDBs en vue de leur ligature. 15
Figure 4 : Cassures double-brin programmées lors de la recombinaison méiotique 21
Figure 5 : Flexibilité de la voie NHEJ.....24
Figure 6 : Un des schéma de la voie de réparation NHEJ chez l'homme 26
Figure 7 : La recombinaison V(D)J utilise la voie NHEJ.....30
Figure 8 : Voie de Réparation par Recombinaison Homologue.....33
Figure 9 : Hétérodimère Ku70/Ku80 36
Figure 10 : DNA-PKcs.....40
Figure 11 : Structure d'Artemis.....42
Figure 12 : Structure de la Ligase4.....44
Figure 13 : XRCC4 et Cernunnos.....46
Figures 14 : XRCC4 et Cernunnos forment des filaments dans les structures cristallines 49
Figures 15 : Les filaments XRCC4-Cernunnos sont observés in vitro par microscopie électronique à transmission (TEM) et in cellulo par microscopie à super-résolution. 51
Figure 16 : Les Polymérases X 53
Figure 17 Localisation des mutations identifiées chez des patients sur les protéines du NHEJ 55
Figure 18 : L'approche MultiBac a permis de produire différents complexes multi-protéiques pour des études structurales 59
Figure 19 : Les protéines XRCC4 et Cernunnos entières interagissent directement et forment des filaments. 64

FIGURES PARTIE II (MATERIEL ET METHODES, RESULTATS, CONCLUSION)

Figure 20 : Représentation des différentes informations de la littérature sur les domaines et les sites d'intéractions des protéines humaines du NHEJ.....67
Figure 21 : Création de vecteurs de transfert multigènes. 69
Figure 22 : Représentation schématique d'un vecteur donneur et d'un vecteur accepteur.....70
Figure 23 : Cartes des différents plasmides utilisés dans cette étude.....72
Figure 24 : Schéma du protocole de production des vecteurs d'expression en système Multibac.....76
Figure 25 Mesure par Calorimétrie. 89
Figure 26 : Cristallisation des proteines 92
Figure 27 : Analyses par microscopie électronique.....95
Figure 28 : Stratégie de clonage des protéines du NHEJ.....97
Figure 29 : Tests de production sur la Plateforme MultiBac à l'EMBL-Grenoble du complexe Ligase4-XRCC4 et Cernunnos en cellules Sf21. 100

Figure 30 Purification du complexe Ligase4-XRCC4:.....102
Figure 31 : Purification de Ku70/Ku80 105
Figure 32 Purification Cernunnos.....107
Figure 33a Purification d'Artemis.....109
Figure 33b Co-purifications de Ligase4-XRCC4 et Cernunnos et de Ligase4-XRCC4 et Ku70/Ku80 111
Table 1: Rendement de productions des protéines du NHEJ exprimées en cellules Sf1 ou Hi5.....112
Figure 34: Croissance cellulaire et suivi de production de YFP par FACS.....113
Table 2: Titrations de stock de virus V1 par la méthode de titration sur plaque.....114
Figure 35 Purification de XRCC4 et Cernunnos entier.....115
Figure 36 : Chromatographie d'exclusion de taille des protéines Cernunnos, XRCC4 seules et en complexe. 117
Figure 37 : Chromatographie d'exclusion de taille des complexes Ligase4/XRCC4, Ku70/Ku80 seuls et en complexe 119
Figure 38 : Chromatographie d'exclusion de taille du complexe Ligase4/XRCC4 avant et après un cycle (congélation/décongélation) avec 10% v/v glycerol 120
Figure 39 : Analyses de l'état de phosphorylation des protéines produites en cellules d'insectes.....122
Figure 40 : Analyse d'un échantillon de Ligase4/XRCC4 à partir de gels SDS-PAGE contenant un dérivé halogéné-alcane qui permet de titrer les tryptophanes. 123
Table 3 : Intensité relative des bandes de Ligase4, XRCC4 phosphorylé et non phosphorylé à l'aide de la technologie TGX staining de Biorad.....123
Figure 41 Analyse par calorimétrie de l'interaction entre les protéines XRCC4 et Cernunnos 124
Figure 42 : Analyses par calorimétrie des interactions entre Cernunnos et XRCC4(1-203) ou le complexe XRCC4(1-203)-Ligase(CTD).....126
Table 3 : Cribles de cristallisation réalisés sur les protéines du NHEJ produits dans E. coli ou en cellules d'insectes.....127
Figure 43 : Cristaux et pistes de cristallisation obtenus avec les protéines Ligase4, XRCC4, Cernunnos, Ku70 et Ku80.....127
Figure 44 : Etude du complexe Ku70/Ku80-Ligase4/$XRCC4_2$/$Cernunnos_2$ par SAXS.....129
Figure 45 : Images de microscopie électronique à transmission (TEM) avec les protéines XRCC4 et Cernunnos entières 130
Figure 46 : Images de microscopie électronique à transmission (TEM) et reconstructions.....132
Figure 47 Analyses des interactions Ku70/Ku80 – ADN par gels retard et TEM. 133
Figure 48 : Réseau d'interaction autour du complexe Ku70/Ku80 138

INTRODUCTION

L'intégrité de l'information génétique, c'est-à-dire l'ensemble des informations portées et transmises à la descendance est essentielle au bon fonctionnement et au bon développement de la cellule. Cette information génétique est portée par l'ADN. Il est indispensable que cet ADN soit dupliqué et transmis dans son intégralité, aussi intact que possible. Si des dommages à l'ADN apparaissent et sont maintenus au sein de la cellule, des mutations ou altérations chromosomiques sont fixées et transmises aux cellules filles. Ces modifications peuvent être silencieuses, sans phénotype particulier, mais elles peuvent aussi entrainer la cellule vers la mort cellulaire ou aboutir à des phénomènes tumoraux et à l'apparition de cancers.

Au cours de l'évolution, les cellules de tous les organismes vivants ont acquis des processus cellulaires qui assurent la surveillance de l'intégrité de la structure de l'ADN *(Hoeijmakers, 2001)*. Ces processus reconnaissent les dommages subis par l'ADN et prennent en charge la réparation de ces dommages. Les causes pouvant induire des dommages à l'ADN sont nombreuses, et peuvent être le fruit d'agents chimiques, physiques et environnementaux. Ces dommages peuvent être également le résultat de processus cellulaires liés à la vie même de la cellule (métabolisme, trafic moléculaire, division). Les mécanismes de réplication, de recombinaison ou d'une manière plus générale, l'ensemble des mécanismes qui font intervenir l'ADN en tant que substrat, sont notamment des processus qui, s'ils sont interrompus par accident, génèrent des altérations de l'ADN.

Dans cette introduction, nous nous intéresserons principalement aux stress susceptibles de créer des cassures double-brin (CDBs) ainsi qu'à la structure chimique des extrémités générées selon les différents dommages. Nous présenterons ensuite les mécanismes de réparation directe des dommages aux extrémités et des voies de réparations permettant la re-ligature des cassures double-brin avec un accent particulier sur la voie NHEJ. Nous finirons cette introduction par une présentation des méthodes d'expression des complexes multiprotéiques en cellules d'insecte dont la méthode MultiBac que j'ai utilisé pendant ma thèse.

1. Formation accidentelle de cassures double-brin

Des espèces réactives de l'oxygène (ROS en anglais pour reactive oxygen species) sont générées au cours du métabolisme cellulaire et durant les phénomènes d'inflammation. Les ROS représentent une attaque permanente sur le matériel génétique et ils peuvent produire des cassures de l'ADN. Il a été estimé qu'entre 10 000 et 100 000 modifications sont apportées chaque jour par les ROS au sein d'une cellule de mammifère *(Lindahl, 1972) (Roberts, 2006)*. De leur côté, les radiations ionisantes (ou IR pour ionizing radiation) génèrent des cassures simple-brin (CSBs) et des CDBs qui présentent des extrémités « abimées » suite à l'attaque de radicaux libres sur les désoxyriboses de l'ADN. On parle d'extrémités « abimées » dès lors que les cassures présentent des extrémités portant une fonction différente d'un hydroxyl en 3' (3'-OH) ou d'un groupement phosphate en 5' (5'-P) qui sont les extrémités utilisables par les ligases pour re-sceller le squelette interrompu du brin d'ADN. Toute autre fonction en 3' et 5' des extrémités de cassures nécessite un nettoyage préalable par des enzymes.

Il faut noter à cette étape que les cassures simple-brin sont des sources significatives de cassures double-brin pour au moins trois raisons majeures. Premièrement, deux cassures simple-brin générées sur deux brins opposés de l'ADN, mais proches l'une de l'autre, peuvent entrainer un dommage de type cassure double-brin. Deuxièmement, une cassure simple brin peut se trouver en face d'un brin qui va subir une oxydation sur une base et/ou une intervention d'une enzyme de réparation qui va générer une seconde cassure simple-brin ramenant au premier cas. Enfin, troisièmement, la non-réparation d'une cassure simple-brin peut lors d'une étape de réplication entrainer un arrêt de la fourche de réplication et l'apparition d'une cassure double-brin. A partir de l'analyse de cellules primaires humaines ou de souris en division et du dénombrement de cassures chromosomiques en métaphase, il a été proposé une fréquence d'environ 10 cassures double-brin par jour par cellules en absence de stress génotoxique externe. Dans le cadre de l'étude des cassures double-brin, il est important de comprendre les stress à l'origine des cassures, la nature des extrémités générées et leur mode de réparation.

1.1 Dommages induits par les rayonnements ionisants et les espèces réactives de l'oxygène

Au cours du processus normal de respiration oxydative, les mitochondries transforment environ 0,1%-1% de l'oxygène en superoxide ($O_2^{\cdot}$- ou O_2-). Les superoxides dismutases présentes dans les mitochondries (enzyme SOD1) ou dans le cytosol (enzyme SOD2) transforment le superoxyde, O_2-, en radicaux libres hydroxyles susceptibles de générer des cassures simple-brin sur l'ADN. Environ 10^{22} radicaux libres ou espèces réactives de l'oxygène sont créés chaque heure dans le corps humain, ce qui représente 10^9 ROS par cellule par heure *(Lieber, 2010)*. Les rayonnements ionisants, rayons gamma et rayons X représentent une autre source de CDBs. Au niveau de la mer, environ 300 millions de particules de rayonnements ionisants traversent chaque heure chaque personne. Ces particules créent des radicaux libres sur leur chemin, notamment lors de leur rencontre avec des molécules d'eau. On considère que ces particules génèrent une CDB toutes les 25 CSBs qu'elles génèrent suite à la proximité entre deux CSBs.

Parmi les modifications chimiques les plus observés aux extrémités 3' des cassures générées par les ROS et les IR, on trouve les deux modifications suivantes : un 3'-phosphate (3'-P) ou un 3'-phosphoglycolate (3'-PG) (**Figure 1,** bas, 1er et 2ème structures) *(Andres, 2015)*. Aux extrémités 5', le produit stable le plus abondant rencontré suite à l'action des ROS et des IR est le groupe 5'-phosphate (5'-P). On trouve également des groupements 5'-aldehyde et 5'-hydroxyl (5'-OH) (**Figure 1***,* haut, 2ème et 5eme structures).

1.2 Dommages induits par les DNA glycosylases de la voie BER

La voie de réparation par excision de bases (BER pour Bases Excision Repair) est également une source importante de cassures simple-brin. La prise en charge des dommages oxydatifs de l'ADN par les glycosylases mono- et bi-fonctionnelles aboutit à des extrémités variées d'ADN au cours de la réparation des dommages *(Boiteux, 2013)*. Les glycosylases monofonctionnelles créent des sites abasiques (AP pour abasic ou apurinic sites) par hydrolyse de la liaison N-glycosidique. Ces sites AP sont également le produit de dépurinations spontanées générées en grand nombre dans les cellules. Ces sites abasiques subissent une incision en 5' du site AP par l'AP endonuclease 1 (APE1) qui aboutit à une

cassure simple-brin avec une extrémité 5'-deoxyribose-phosphate (5'-dRP) et 3'-OH (**Figure 1**, haut, 3ème structure). Les glycosylases bi-fonctionnelles créent un site abasique puis réalisent une réaction de β,δ-élimination ou β-élimination qui génère une cassure simple-brin avec un 3'-phosphate ou avec un groupement aldéhyde 3'-phospho-α, β insaturé (**Figure 1**, haut, 4ème structure) *(Andres, 2015)*.

1.3 Pontages covalents de type topoisomérases-ADN

Les processus cellulaires comme la réplication, la transcription ou la division cellulaire nécessitent des mécanismes spécifiques pour relâcher les contraintes topologiques de l'ADN et pour résoudre l'enchevêtrement de l'ADN au sein de la chromatine ou suite au passage des ADN ou ARN polymérases. Cette fonction est assurée par les DNA topoisomérases qui génèrent de façon transitoire des CSBs (topoisomérases de type 1) ou des CDBs (topoisomérases de type 2). Ces enzymes créent des liaisons covalentes enzyme-phosphotyrosyl-ADN avec les extrémités 5' ou 3' de l'ADN. Une interruption ou un dysfonctionnement de ces enzymes peut aboutir à la formation d'adduits protéine-ADN (**Figure 1**, haut, 6ème structure et bas 5ème structure).

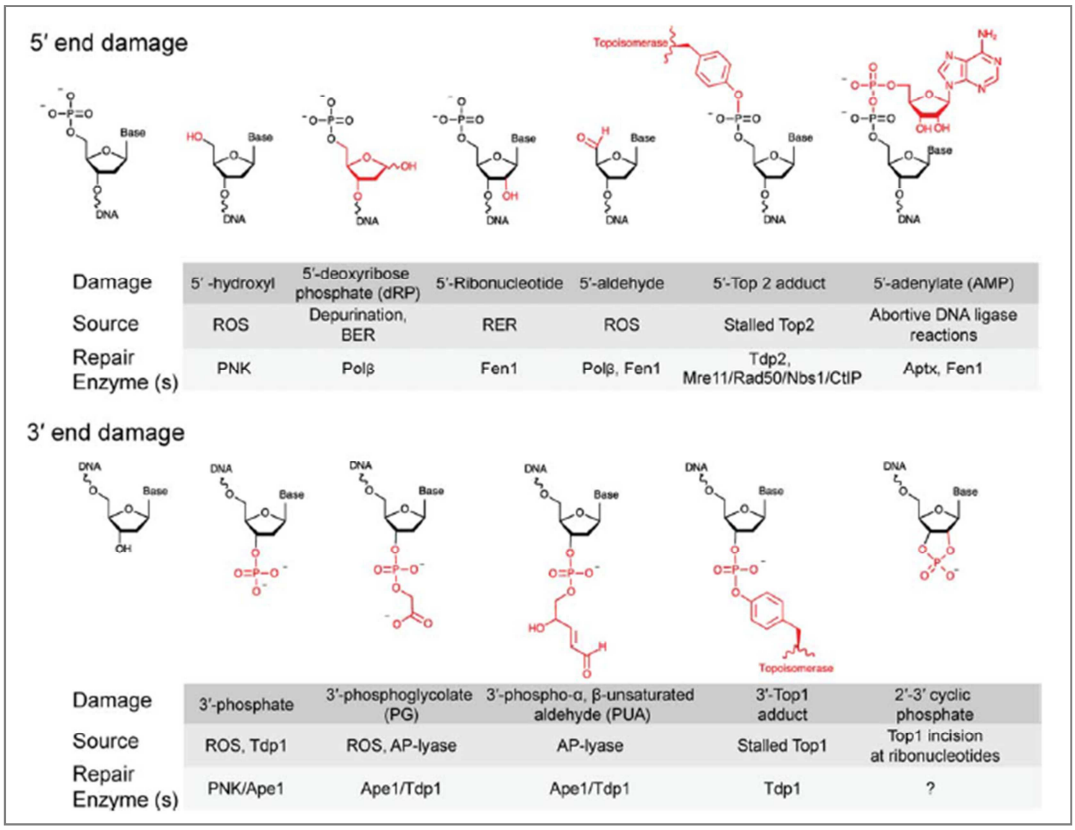
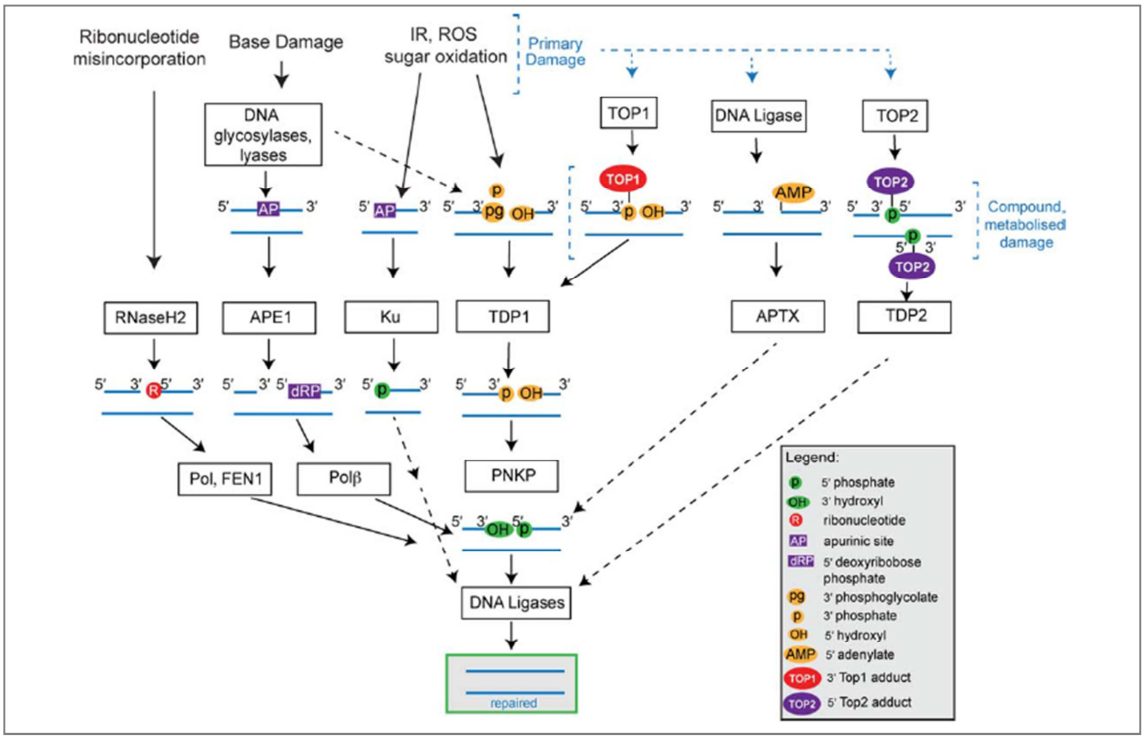


Figure 1 : Variété des structures chimiques trouvées aux extrémités de cassures simple-brin (CSB) ou double-brin (CDB) de l'ADN. *Les cassures générées par les ROS, les IR et d'autres stress portent à leurs extrémités des modifications chimiques (en rouge) soit à l'extrémité 5' (panneau du haut) soit à l'extrémité 3' (panneau du bas). Ces modifications empêchent un déroulement correct de l'étape de ligature. Les sources des dommages à l'origine de chaque modification chimique sont indiquées ainsi que les enzymes susceptibles de réparer les extrémités en vue de leur préparation pour l'étape de ligature (d'après (Andres, 2015)).*

Les topoisomérases de type 1 (TOP1) relâchent les contraintes topologiques sur l'ADN à l'aide d'une tyrosine du site catalytique qui introduit un CSB en créant une liaison covalente avec le 3'-phosphate de l'incision générée. Cette liaison permet de maintenir l'ADN fixé à l'enzyme pendant la rotation de l'ADN et le relâchement des contraintes topologiques. A l'issue de cette étape, TOP1 relie les deux brins d'ADN incisés. Cette réaction est cependant sensible à différents paramètres comme la présence d'un dommage à proximité du site incisé (mésappariement, 8oxoG, site AP ou autre incision) qui peuvent gêner l'étape de re-ligature. De même certaines drogues utilisées en chimiothérapie, comme la camptothecine, peuvent empoisonner la réaction des topoisomérases de type 1 et aboutir à des complexes irréversibles TOP1-ADN. Ces adduits protéines-ADN bloquent la transcription et entrainent un effondrement des fourches de réplication.



<u>Figure 2 : Mécanismes enzymatiques de préparation des extrémités des CDBs en vue de leur ligature</u>. *Schéma reprenant les principales imperfections que l'on peut trouver aux extrémités d'une CDB et indiquant les enzymes qui peuvent prendre en charge ces imperfections de façon à rétablir des extrémités 3'-OH et 5'-P susceptibles d'être utilisées par la ligase pour la dernière étape de réparation. Certains dommages situés à proximité des CDBs et certaines drogues peuvent perturber la fonction des topoisomérases ou des ligases (flèches et crochets en pointillés bleu clair) (d'après (Andres, 2015))*

Les topoisomérases de type 2 (TOP2) relâchent les contraintes topologiques sur l'ADN en créant une cassure double-brin avec un décalage de 4 bases entre les incisions faites sur les 2 brins. Cette CDB permet de passer un duplex d'ADN à travers la « porte » créée puis l'enzyme relie les extrémités clivées. Deux molécules de TOP2 forment chacune une réaction de trans-estérification avec une tyrosine du site actif générant ainsi deux adduits TOP2-ADN de manière similaire aux TOP1 mais avec une cassure double-brin. De la même manière les TOP2 sont sensibles à l'environnement des sites incisés et à des drogues comme l'étoposide, utilisé également comme traitement anti-cancéreux.

<u>1.4 Dommages aux extrémités créés par les ligases</u>

Les mécanismes des 3R, réplication, recombinaison et réparation, présentent une étape finale de ligature au cours de laquelle les extrémités générées dans ces processus sont reliées. Les ligases eucaryotes (Ligase 1, 3 et 4) suivent un mécanisme en trois étapes : (i)

une lysine du site actif est adénylée, (ii) le groupement AMP lié à la lysine est transféré sur le 5'-P de l'ADN substrat, et (iii) la liaison phosphodiester est reformée par attaque nucléophile du 3'-OH sur le 5'-adenylate *(Pascal, 2004, Tomkinson, 2013)*. Une interruption de ce processus catalytique peut, de manière similaire à ce qui a été décrit pour les topoisomérases, aboutir à la présence d'adduits aux extrémités d'ADN. Dans le cas des ligases, il s'agit d'adduits 5'-adenylate (**Figure 1**, haut, 7ème structure).

Différents environnements peuvent donner une ligature interrompue avec un adduit 5'-adenylate relâché par la ligase. Il s'agit notamment des cas où la ligature se déroule à proximité d'un dommage de type 3'-AP (site abasique), d'un 8oxo-dG ou d'un groupement 5'-dRP. Les ligases considérées généralement comme l'ultime étape de la réparation peuvent donc elles-mêmes créer des intermédiaires toxiques d'ADN adénylés.

1.5 Intermédiaires dérivés de l'introduction de ribonucléotides dans la structure de l'ADN

Les travaux du laboratoire de T Kunkel ont montré ces dernières années l'importance de l'introduction de ribonucléotides à la place de désoxyribonucléotides par les polymérases réplicatives Pol ε et Pol δ *(Nick McElhinny, 2010)*. Il a été estimé que 10 000 rNMPs sont incorporés par les ADN polymérases réplicatives à chaque cycle cellulaire chez *S. cerevisiae* et plus de 1 000 000 de rNMPs seraient incorporés à chaque division cellulaire chez la souris. Ces chiffres sont à comparer par exemple aux 10 000 sites AP générés dans une cellule de mammifère par jour. A cela s'ajoute la synthèse par les primases d'environ 5% du brin tardif (lagging strand en anglais) à l'aide de rNTPs.

L'incorporation de ces rNMPs dans l'ADN est associée à des risques majeurs d'instabilité génétique. Premièrement, les ribonucléotides possèdent sur le ribose un groupement hydroxyle en position 2' (2'-OH) qui rend la liaison sucre-phosphate plus sensible à l'hydrolyse dans des conditions physiologiques. Cette hydrolyse crée une cassure simple brin avec un 5'-OH et un groupement phosphate 2'-3'-cylique (**Figure 1**, haut, 2ème structure, bas 6ème structure). Deuxièmement, le retrait des ribonucléotides de l'ADN est pris en charge par une voie appelée RER (pour Ribonucleotide Excision Repair) *(Tumbale, 2014)*. La voie RER est initiée par la RNAse H2, une enzyme hétérotrimerique qui clive du coté 5' du

ribonucléotide. Un intermédiaire de cette voie est donc une CSB avec un 5'-P ribonucléotide et un 3'-OH. La présence en grande quantité de ces ribonucléotides et des intermédiaires de la voie RER semble jouer un rôle majeur dans l'instabilité génétique. A titre d'exemple, il a été observé que l'action de TOP1 en position 3' d'un ribonucléotide peut générer un groupement bloquant de type phosphate 2'-3' cyclique comme suite à l'hydrolyse via le 2'-OH du ribose décrite au-dessus.

2. Enzymes de réparation des extrémités de CDBs en vue de leur ligature

Les nombreux types de dommages rencontrés aux extrémités des CDBs, décrits ci-dessus, sont ôtés par de nombreuses enzymes afin de préparer ces extrémités pour la ligase à savoir sous leur forme 3'-hydroxyl (3'-OH) et 5'-Phosphate (5'-P). Ces enzymes reconnaissent directement et réparent directement ces dommages ou sont recrutées au cours de la succession des étapes d'une voie de réparation des cassures double-brin (§ 3.).

2.1 Retrait d'un 5'-adenylate par l'Aprataxin (APTX)

L'Aprataxin est une enzyme capable d'ôter un 5'-adenylate sur un intermédiaire de ligature non processé pour générer un 5'-Phosphate (**Figure 2, 3**). APTX a été identifié à partir de patients présentant des lésions cérébrales de type ataxie (pathologie neuromusculaire avec manque de coordination fine des mouvements). La maladie liée au défaut d'APTX est appelée ataxia-oculomotor apraxia type 1 ou AOA1. Cette maladie est une maladie neurodégénérative autosomale récessive proche d'une autre maladie très étudiée, l'ataxie télangiectasie (A-T) liée à des mutations dans le gène ATM (Ataxia Telangiectasia Mutated, ATM appartient à la famille des PI-3-kinases et phosphoryle notamment le variant d'histone H2AX en présence de cassures double-brin). Les cellules de ces patients, déficientes pour APTX (cellules A0A1), présentent des niveaux élevés de dommages oxydatifs de l'ADN, une sensibilité accrue à un inhibiteur de la topoisomérase I, la camptothecine (CPT), mais pas de sensibilité accrue aux rayonnements ionisants (IR) suggérant un rôle principal dans la réparation des cassures simple-brin *(Mosesso, 2005)*.

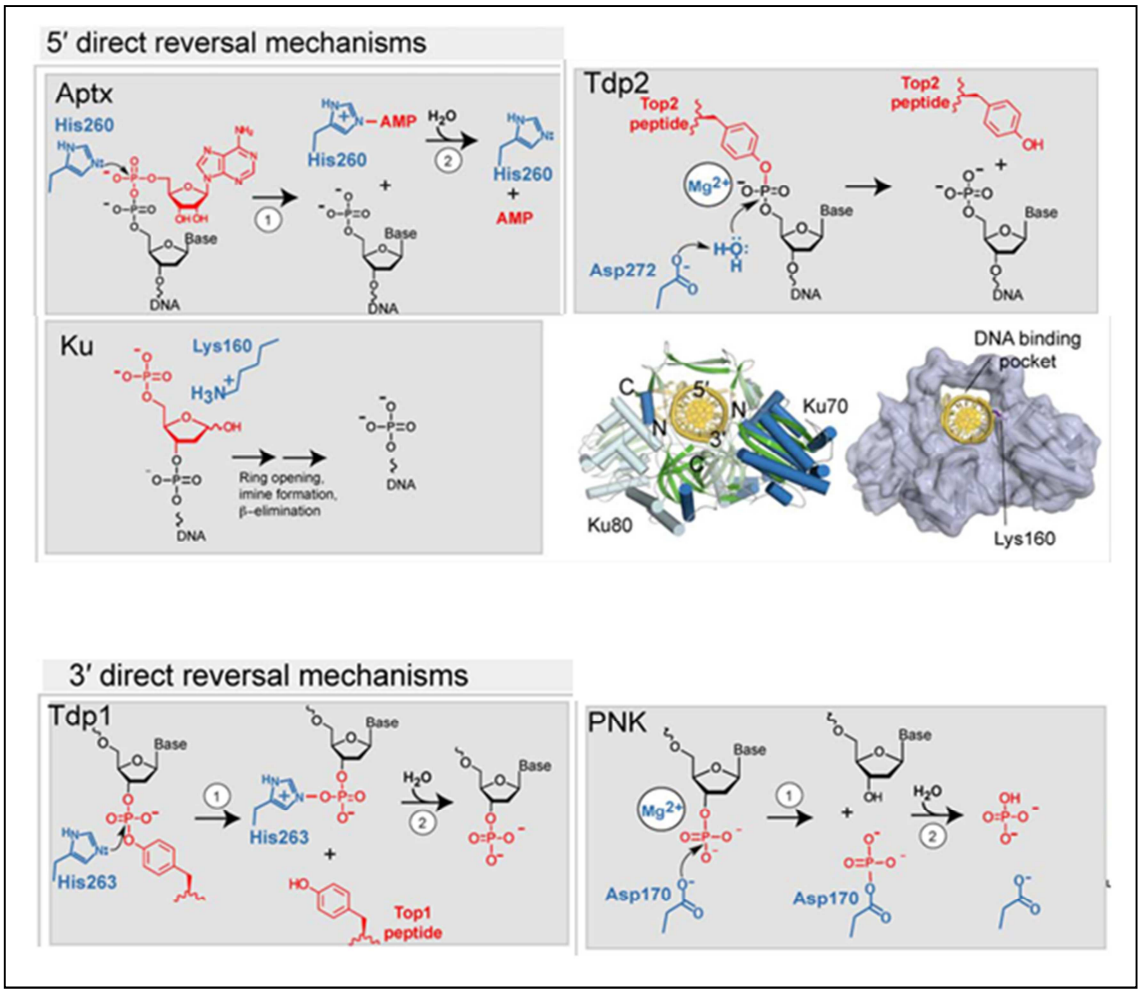


Figure 3 : Mécanismes enzymatiques directs de préparation des extrémités de CDBs en vue de leur ligature. *Schéma de l'action des enzymes nettoyant les extrémités en 5' (haut, APTX, Tdp2, Ku et les extrémités en 3' (bas Tdp1, PNK). Pour la protéine Ku, est présentée la structure cristalline de Ku avec la position de la Lysine 160.*

Les structures cristallines de l'APTX de *S. pombe* et humaine ont été déterminées *(Schellenberg, 2015)*. L'APTX appartient à la famille des hydrolases et transférases de nucléotides appelée HIT pour Histidine Triade de par la présence d'un motif HφHφHφφ (φ étant un acide aminé hydrophobe) qui contient l'histidine nucléophile (H260). L'APTX contient un domaine à doigt de zinc (Zinc Finger ou ZnF) qui est impliqué dans la reconnaissance de l'extrémité 5'-adenylée. APTX contient également un domaine FHA (pour Forkhead domain) qui permet des interactions avec notamment XRCC1, un facteur de coordination du BER et de la réparation des cassures simple-brin (CSB) et avec XRCC4, le co-facteur de la Ligase4 du NHEJ. Les domaines FHA reconnaissent des motifs phosphorylés sur leurs protéines partenaires. APTX interagit plus fortement avec XRCC1 et XRCC4 lorsque

ceux-ci sont phosphorylés par la kinase CK2 suggérant une régulation de l'interaction par l'état de phosphorylation de XRCC4 et XRCC1.

2.2 Polynucléotide Kinase Phosphatase (PNK)

La PNK possède deux activités. Elle a une activité ADN 5'-kinase qui phosphoryle une extrémité 5'-OH et une activité 3'-phosphatase qui hydrolyse les extrémités 3'-phosphorylées (3'-P) produites par les ROS et les ADN glycosylases (**Figure 2, 3**). Une activité importante de la PNK est l'hydrolyse des 3'-P laissés par la Tyrosyl DNA Phosphodiesterase 1 (TDP1) lors de la réparation des intermédiaires des adduits TOP1-ADN (voir plus bas, § 2.3 TDP1). La structure cristalline de la PNK entière murine a été décrite *(Bernstein, 2005)*. L'enzyme présente une structure modulaire avec un domaine N-terminal FHA de liaison aux phosphoprotéines et un domaine C-terminal couplant les activités phosphatase et kinase. Dans les structures de type phosphatase-ADN, le substrat présente une conformation avec une structure étendue en simple brin, le 3'-P est situé dans une poche enfouie, et un aspartate catalytique (Asp170 dans mPNK) activerait une attaque nucléophile par un cation Mg^{2+}. Dans les structure de type kinase-ADN, l'extrémité 5' occuperait le site kinase sous la forme phosphorylée suite à une réaction ATP-dépendante *(Garces, 2011)*.

2.3 Tyrosyl DNA Phosphodiesterase 1 (TDP1)

Une interruption de la réaction catalysée par TOP1 aboutit à la création d'une cassure simple-brin avec un adduit TOP1-ADN avec l'enzyme TOP1 liée de façon covalente en 3' de l'extrémité d'ADN. Cet intermédiaire déclenche une dégradation de TOP1 par la voie ubiquitine-protéasome, l'adduit TOP1-ADN devenant un adduit peptide(TOP1)-ADN avec un lien phospho-tyrosine (**Figure 1**, bas, 5^{e} structure). L'enzyme TDP1 catalyse l'hydrolyse directe de la liaison phospho-tyrosine. La structure 3D d'un complexe ternaire TDP1 humaine, ADN-peptide, vanadate a été décrite *(Davies, 2003)*. La TDP1 contient deux domaines en tandem de structure appartenant à la super famille des phospholipases D (PLD) avec chaque domaine qui reconnait une face de la liaison phosphodiester entre le peptide et le phosphate de l'ADN. Comme les autres PLD, le site actif est constitué par un motif His-Lys-Asp. L'hydrolyse de l'adduit phospho-tyrosine est réalisé en deux temps, (i) attaque

nucléophile par une histidine de la liaison phosphodiester, puis (ii) l'intermédiaire phospho-histidyl est lui-même hydrolysé (**Figure 2, 3**). La TDP1 génère alors une extrémité 3'-P qui doit être reconnu et modifié par la PNK pour donner un 3'-OH. Un dysfonctionnement de la TDP1 chez l'homme peut induire des maladies neurodégénératives de type ataxie cérébelleuse. Une mutation d'une histidine du site actif en arginine a ainsi été observée chez un patient et associée à l'apparition d'intermédiaires phospho-histidyles stabilisés qui transforment le dommage adduit ADN-TOP1 toxique en adduit ADN-TDP1 toxique.

2.4 Tyrosyl DNA Phosphodiesterase 2 (TDP2)

De façon similaire au couple TOP1/TDP1, un adduit TOP2-ADN formé par interruption du cycle catalytique de TOP2 est pris en charge par le protéasome pour dégrader TOP2 puis par l'hydrolase TDP2 pour résoudre l'adduit phospho-tyrosine ADN-peptide dérivé de TOP2. Dans le cas de TOP2, l'adduit phospho-tyrosine se situe sur l'extrémité 5′ de l'ADN. TDP2 génère un produit avec une extrémité 5'-P qui peut être directement utilisé par la Ligase4 du NHEJ (**Figure 2,3**). La structure cristalline d'un complexe TDP2-ADN a permis de mettre en évidence un repliement de type metallo-enzyme de la super famille EEP (Endonucléase/Exonucléase/Phosphatase) *(Schellenberg, 2012)*. La structure permet de proposer un mécanisme avec un aspartate conservé du site actif bien positionné pour jouer le rôle de base générale et activer une molécule d'eau pour une attaque nucléophile sur la liaison phosphodiester et un cation Mg2+ jouerait alors un rôle de stabilisation de l'état de transition de la réaction d'hydrolyse. Ce mécanisme est proche de celui décrit pour l'endonucléase du BER, APE1. Aucune maladie n'est associée à ce jour à un défaut de TDP2. En revanche, les cellules dérivées des modèles animaux dont le gène TDP2 est inactivé présentent une sensibilité accrue à l'étoposide (inhibiteur de TOP2) et un nombre important de foyers γH2AX (produit de la signalisation des CDBs).

2.5 DNA Polymerase β et activité 5'dRP-lyase et AP-lyase

L'incision des nombreux sites AP formés spontanément dans la cellule ou générés lors des premières étapes du BER créent des ADN incisés (nick) présentant des groupements 5'-dRP. Les cellules de mammifères présentent une activité 5'-dRP-lyase portée

principalement par la polymerase β (POLβ). POLβ est la plus petite polymérase multifonctionnelle de mammifère (39kDa) avec une activité polymérase (portée par le domaine C-ter de 31kDa avec 3 sous-domaines classiques : paume, doigt et pouce) et 5'-dRP-lyase (portée par le domaine N-ter 8kDa). Dans la voie BER courte (short-patch BER), la POLβ ôte le groupement dRP laissant une brèche (gap) qui est comblée par l'activité polymérase de POLβ, la Ligase3 refermant les deux extrémités 3'-OH/5'-P. Les structures 3D de complexes entre la POLβ et un ADN contenant une incision ou une brèche suggèrent un mécanisme sans cation métallique avec attaque nucléophile d'une lysine sur l'atome C1' du sucre et la formation d'une base de Schiff, de façon similaire au mécanisme décrit pour les glycosylases bi-fonctionnelles.

2.6 Ku70/80, activité 5'dRP-lyase et AP-lyase

Du fait du nombre élevé de site AP créés dans une cellule, on peut trouver à proximité d'une cassure double-brin certains de ces sites. Des études récentes suggèrent que l'hétérodimère Ku70/Ku80 possèderait une activité dRP-lyase et AP-lyase qui permettrait d'ôter des sites abasiques situés à un ou deux nucléotides d'une extrémité 5'-sortante (5'-overhang) *(Roberts, 2010) (Strande, 2012)*. Cette activité serait restreinte aux sites AP susceptibles de gêner l'étape de ligature du NHEJ.

L'hétérodimère Ku70/Ku80 est décrit plus loin en tant que facteur central de la voie NHEJ. Brièvement, Ku est une des premières protéines présentes aux extrémités d'une CDB de par sa grande abondance dans la cellule et sa forte affinité pour les extrémités d'ADN. De par sa structure en anneau, Ku entre sur l'ADN par l'extrémité et peut glisser vers l'intérieur. Deux lysines situées dans le domaine N-terminal α/β de Ku70, les lysines Lys31 et Lys160 ont été proposées comme impliquées dans l'activité lyase de Ku (**Figure 2,3**). On peut cependant noter que le double mutant K31A/K160A ne s'accompagne que d'une diminution d'un facteur 4 de l'activité lyase de Ku.

3. Cassures programmées de l'ADN

3.1 Cassures programmées dans les lymphocytes : V(D)J et CSR

Le système immunitaire présente un processus de réarrangements de gènes, la recombinaison V(D)J qui est essentiel pour la génération d'un large répertoire de récepteurs des cellules T et B. Ce processus est initié par les nucléases lymphoïdes spécifiques RAG1/RAG2 qui introduisent des cassures double-brin aux bornes des gènes V, D et J. Ces gènes sont alors réarrangés au cours de la recombinaison V(D)J. La ligature des gènes V, D et J utilise la voie de recombinaison NHEJ. Le mécanisme moléculaire de ce processus est décrit plus loin au **paragraphe 4.1.4** comme un exemple de mécanisme de réparation par la voie NHEJ.

Les immunoglobulines sont également l'objet de réarrangements au niveau des chaines lourdes de ces récepteurs pour modifier leur fonction. Ce processus est appelé changement de classe isotypique (CSR pour Class switch recombination) et il met en jeu la désamination de cytidine par l'enzyme AID pour l'incision de l'ADN sur les sites uraciles formés suite à cette action. Lorsque deux incisions sont réalisées à proximité, le processus CSR aboutit à des CDBs qui sont pris en charge par la voie NHEJ. Le mécanisme moléculaire de ce processus est également décrit ci-dessous comme un exemple de réparation NHEJ (**paragraphe 4.1.5).**

3.2 Cassures programmées lors de la recombinaison méiotique

Durant la méiose, la formation de cassures double-brin programmées permet un échange de matériel génétique entre les chromosomes homologues. Ces échanges augmentent la diversité des génomes des organismes qui présentent une reproduction sexuée. Ces échanges sont essentiels pour une ségrégation correcte des chromosomes lors de la première division méiotique *(Subramanian, 2014)*.

La méiose est un processus de ségrégation des chromosomes où une cellule parentale diploïde donne naissance à des gamètes haploïdes. Cette réduction de ploïdie est essentielle pour la gamétogenèse chez tous les organismes qui présentent une reproduction sexuée. Elle est accomplie par un cycle de réplication de l'ADN suivi de deux évènements de ségrégation des chromosomes qui séparent non seulement les chromatides sœurs mais également les chromosomes homologues. La séparation des chromosomes homologues se déroule durant la première division méiotique (méiose I), suivi d'une séparation des

chromatides sœurs durant la méiose II. Le mécanisme de ségrégation des chromosomes nécessite que les paires de chromosomes devant être séparées, soient dans un premier temps connectées entre elles pour assurer leur orientation adéquate le long du fuseau méiotique. Les connections entre chromosomes homologues sont créées suite à la formation d'enjambements par recombinaison (crossover recombination en anglais) (**Figure *4***).

La formation de crossover (CO) est initiée par la formation de nombreuses cassures double-brin de l'ADN programmées. Ces CDBs sont réalisées par l'enzyme SPO11, de la famille des Topoisomérase II *(Baudat, 2013)*. Ces cassures sont générées alors que les chromosomes sont organisés en une structure appelée « bouquet » avec les télomères ancrés dans la membrane nucléaire. Deux molécules de SPO11 effectuent un lien covalent lors de l'introduction de deux cassures simple brin en vis-à-vis. Ces intermédiaires ADN-SPO11 sont résolues par le complexe MRN qui libère des complexes covalents SPO11-oligonucléotides avec des ADN de 12-36 nucléotides liés covalemment à SPO11. L'exonucléase 5'-3' EXO1 ainsi que le complexe MRN procèdent à une résection des extrémités et bloquent la prise en charge de ces cassures par la voie NHEJ. La formation des CO passe par un chargement des protéines d'échange de brins RAD51 et DMC1 puis la formation de jonctions de Holiday qui peuvent être résolues par les nucléases parmi lesquelles on trouve MLH1-MLH3 ou MUS81 *(Zakharyevich, 2012)*. Les intermédiaires formés à cette étape peuvent également être dissous et donner un échange sans enjambement (non crossover) avec une conversion de gène cad insertion d'une courte séquence du chromosome homologue (quelques paires de bases à 300bp).

Deux méthodes ont permis de donner une estimation du nombre de CDBs générées en méiose. Il s'agit du séquençage des oligonucléotides fixés de manière covalente aux molécules SPO11 et du dénombrement de foyers de RAD51/DMC1. Par cette dernière technique, environ 200-400 foyers de RAD51/DMC1 sont identifiés par cellule humaine ou par cellule de souris dans les premières étapes de la méiose I. Ces cassures sont finement régulées de façon à générer au moins un crossover par paire de chromosome homologue. Le nombre de CO est estimé à environ 1/10ème du nombre de CDBs générées par SPO11.

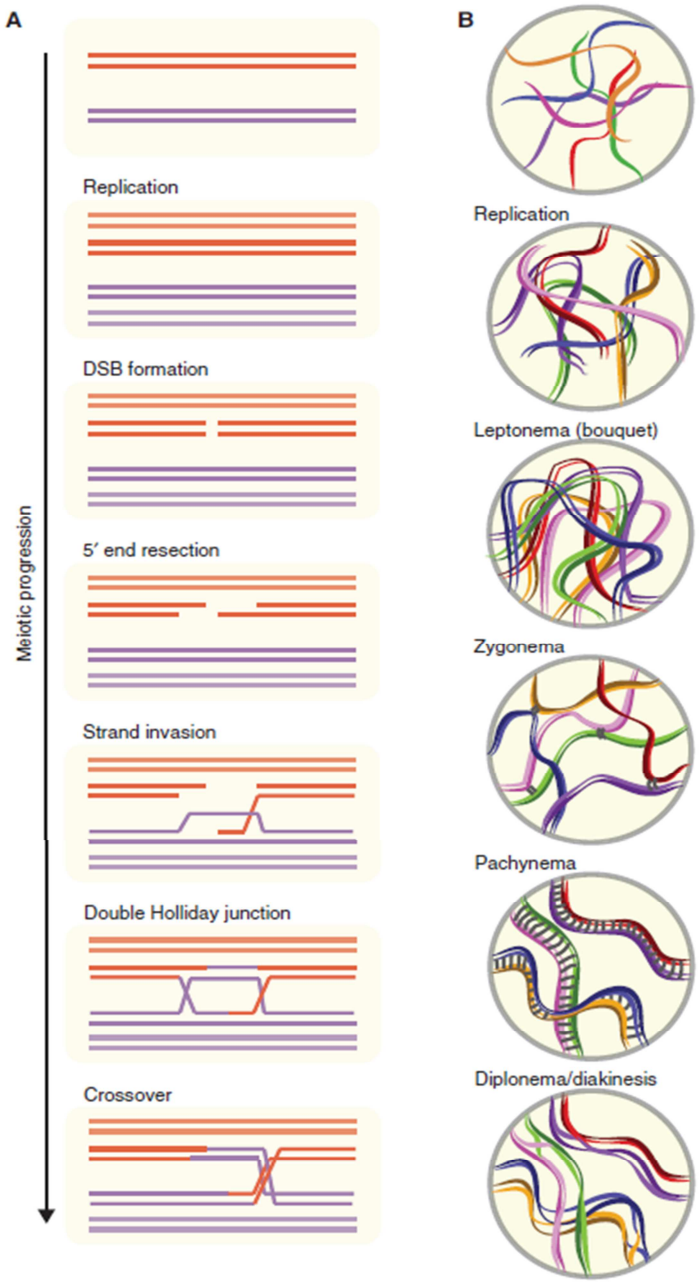


Figure 4 : Cassures double-brin programmées lors de la recombinaison méiotique. *La figure représente(A) les événements de réparation des CDBs en méiose ainsi que (B) les événements d'organisation des chromosomes qui sont concomitants. Les chromosomes homologues sont répliqués pendant la phase S pré-méiotique. A l'étape leptonème, les CDBs sont introduites alors que les télomères des chromosomes viennent d'être ancrés dans la membrane nucléaire formant une structure appelée bouquet. A l'étape du zygonème, des premières structures d'invasion (lignes grises) se forment entre les paires de chromosomes homologues. A l'étape du pachynème, l'ensemble des échanges de brins est réalisé, les chromosomes homologues sont fortement appariés et la recombinaison est à l'étape de la formation des jonctions de Holiday. A l'étape diplonème, les complexes méiotiques se désassemblent et les échanges de type CO sont formés. (D'après (Subramanian, 2014))*

4. Voies de réparation des cassures double-brin

Dans les cellules eucaryotes, deux voies de réparation principales prennent en charge la réparation des cassures double-brin de l'ADN. Il s'agit de la voie de réparation non-homologue par jonction des extrémités (NHEJ pour Non-Homologous End Joining, appelée aussi c-NHEJ pour voie NHEJ classique) et de la voie par recombinaison homologue (HR pour Homologous Recombination) *(Bassing, 2004)*. Ces deux voies agissent de manière compétitive et complémentaire. Compétitive car elles sont initiées par une étape de reconnaissance qui met en jeu chez l'homme les facteurs Ku70/Ku80 pour la voie NHEJ et le complexe MRN (MRE11/RAD50/NBS1) pour la voie HR. Ces complexes entrent en compétition pour orienter la réparation vers la voie NHEJ ou vers la voie HR. Ces deux voies sont également complémentaires pour deux raisons. Premièrement, elles interviennent à des étapes différentes du cycle cellulaire. La voie HR joue un rôle important dans les cellules en phase S/G2 lorsqu'une chromatide sœur est disponible pour servir de modèle. La voie NHEJ est active tout au long du cycle avec un rôle majeur au cours de la phase G1. Elles sont également complémentaires car chacune peut servir en partie de voie de substitution à l'autre en cas d'un dysfonctionnement de l'autre voie. Il est à noter que l'importance de ces deux voies varie selon les organismes : la voie NHEJ prédomine chez les mammifères alors que la voie HR a un rôle très important dans les levures.

Une troisième voie a été mise en évidence plus récemment et est appelée a-NHEJ (pour alternative NHEJ) La voie a-NHEJ comprend elle-même une sous-voie MMEJ (pour Microhomology Mediated End Joining) qui s'appuie sur des recherches d'appariements sur des petites séquences homologues. La voie a-NHEJ n'est pas une voie spécifique à la réparation de cassures double-brin particulières ou à des phases du cycle cellulaire. Elle se définit plutôt comme un ''bricolage'' effectué en cas de déficience de la voie NHEJ classique ou HR pour résoudre les CDBs à partir des enzymes d'autres voies de réparation comme le complexe XRCC1/Ligase3 de la voie BER *(Frit, 2014)*.

4.1 Voie NHEJ classique

Une des particularités de la voie NHEJ est de prendre en charge un large ensemble de dommages regroupés derrière le terme unique de « cassures double-brin » **(Figure 5)** *(Lieber,*

2010). Premièrement comme décrit précédemment les extrémités des cassures peuvent être très hétérogènes d'un point de vue chimique (3'-P, 3'-Phosphoglycolate, adduit ADN-peptide TOP1 etc …) (voir **paragraphe 1.**). La voie NHEJ intervient sur ces cassures aussi bien après une intervention des enzymes de préparation des extrémités (voir **paragraphe 2.**) ou avant leur intervention. Deuxièmement, la voie NHEJ prend en charge les CDBs qui présentent des extrémités de structure différente : cassures double-brin franches (blunt), 3'- ou 5'-sortantes (3' ou 5'-overhang), des extrémités cohésives ou non-cohésives. Troisièmement, ces extrémités sortantes peuvent elles-mêmes être de longueurs plus ou moins grandes. Les travaux des dernières années sur la voie NHEJ ont mis en évidence une grande flexibilité au sein de cette voie de réparation qui répond à la grande variété de cassures double-brin prises en charge par cette voie.

Selon la nature des extrémités des cassures double-brin à réparer, la voie NHEJ utilise une ou plusieurs fois des nucléases et des kinases pour préparer les extrémités endommagées d'ADN, des polymérases pour re-synthétiser éventuellement des nucléotides manquants et une ligase pour relier les extrémités. Dans les premières étapes de la voie, les deux extrémités de la CDB sont reconnues par l'hétérodimère Ku70/Ku80 (appelé aussi Ku), avec un ou deux hétérodimère sur chaque extrémité *(Britton, 2013)*. Ce complexe est présent en grand nombre dans la cellule. Avec environ 400 000 molécules par cellule, c'est la molécule de la voie NHEJ la plus abondante et une des protéines les plus abondantes de la cellule. L'hétérodimère Ku possède une forte affinité pour des duplexes d'ADN (Kd ~ 10^{-9}M). Cette reconnaissance se fait de façon séquence indépendante et est peu sensible à la structure de l'extrémité (franche ou bout sortant ou structure en épingle à cheveux) *(Arosio, 2002, Blier, 1993)*. Le complexe Ku-ADN joue alors un rôle central de plateforme de recrutement pour les nucléases, les kinases, les polymérases ou la ligase de la voie NHEJ *(Grundy, 2014)*. De ce point de vue, le rôle de Ku rappelle celui de la protéine PCNA qui joue le rôle de plateforme de recrutement au niveau de la fourche de réplication. PCNA est un homotrimère qui forme un anneau sur l'ADN. L'anneau de PCNA n'est pas pré-formé comme Ku mais il est chargé et formé sur l'ADN par la protéine RFC. Comme Ku, PCNA joue le rôle de plateforme de recrutement pour un nombre important de nucléases, polymérases et ligases *(Moldovan, 2007)*.

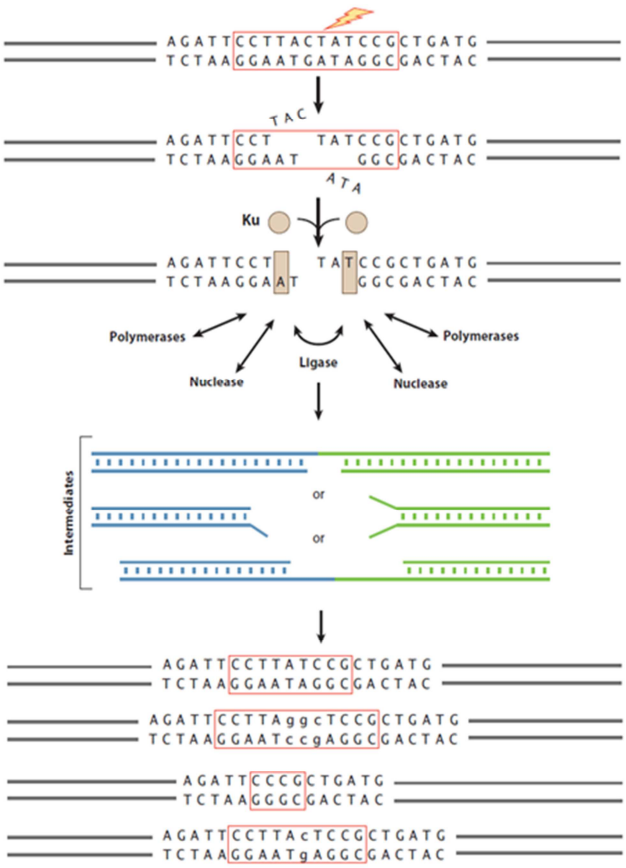


Figure 5 : Flexibilité de la voie NHEJ. *Suite à différents stress génotoxiques endogènes ou exogènes (éclair rouge) des CDBs sont générées sur l'ADN. Ces CDBs peuvent présenter des extrémités endommagées et non-directement utilisables par la ligase du NHEJ, la Ligase4. Un (ou deux) hétérodimères Ku se fixeraient à chaque extrémité de la CDB. Ku recruterait alors dans un ordre qui peut varier selon la nature de la CDB soit une nucléase, soit une polymérase soit une ligase. Des études ont proposé que lors de la fixation à l'ADN, Ku subirait certains changements de conformation pour expliquer l'interaction de Ku avec certains partenaires uniquement en présence d'ADN (ce changement de conformation est représenté sur le schéma par le passage d'une sphère grise en un rectangle gris). La flexibilité de la voie NHEJ est essentielle pour permettre la réparation d'un ensemble très varié de structures des extrémités des CDBs. La figure montre pour un exemple de CDB donné, une grande variété d'intermédiaires possibles de réparation (panneau du milieu). La variété des intermédiaires d'ADN possibles se reflètent dans la diversité des séquences observées au niveau de la jonction suite à la ligature (panneau du bas). (D'après (Lieber, 2010))*

Selon les premiers modèles proposés historiquement, la voie NHEJ se déroulait selon un schéma linéaire en trois étapes principales : (i) reconnaissance de la cassure par Ku et DNA-PKcs, (ii) recrutement d'enzymes de préparation pour la ligature, des nucléases et des kinases et des polymérases pour préparer les extrémités pour la ligature (end processing) et (iii) recrutement toujours par Ku-(DNA-PKcs) du complexe de ligature. Les travaux de ces dernières années ont montré que la voie NHEJ pouvait fonctionner selon des processus beaucoup plus variés que ce strict modèle linéaire. Premièrement, chaque extrémité Ku-ADN

peut recruter dans un ordre variable les nucléases, les polymérases ou la ligase de la voie NHEJ. Deuxièmement, le recrutement à une extrémité d'une de ces enzymes peut être répété une ou plusieurs fois.

Nous décrivons dans les paragraphes suivants quelques exemples d'enchainements d'étapes proposés pour la voie NHEJ. Les scénarios décrits ci-dessous correspondent à des types de CDBs différents. Dans les paragraphes qui suivent, nous n'entrerons pas dans le détail des domaines des protéines et des sites d'interactions entre protéines. Ces informations seront présentées dans les paragraphes du **chapitre 5.** « Caractéristiques des protéines de la voie NHEJ».

4.1.1 Cassures double-brin nécessitant une étape d'excision/incision des extrémités

Comme évoqué dans le paragraphe précédent, la réparation par la voie NHEJ de CDBs débute par la fixation de l'hétérodimère Ku70/Ku80 aux extrémités des CDBs (**Figure 6).** Ce complexe présente une structure en anneau préformé avec deux sous-unités de 70kDa et 86kDa *(Walker, 2001)*. Cet anneau s'enfile par les extrémités de la cassure. Des analyses de microscopies à force atomique *(DeFazio, 2002)* ou à transmission par coloration négative (S Baconnais, en préparation) montrent que Ku70/Ku80 peut occuper une ou deux extrémités d'un ADN et peut également glisser à l'intérieur de l'ADN. Différents travaux ont montré que Ku possède une forte affinité pour l'ADN double-brin (Kd de 0,5nM en présence de 220mM NaCl mesuré par anisotropie de fluorescence) *(Arosio, 2002)*.

Le complexe Ku-ADN est capable de recruter directement la serine/threonine protéine kinase, DNA-PKcs (pour DNA dependent Protein Kinase catalytic subunit). Cette kinase est la plus grande kinase humaine avec 4168 acides aminés (aa) et 469-kDa. Bien que moins abondante que l'hétérodimère Ku, c'est une protéine également très abondante. Lors de son interaction avec Ku, DNA-PKcs se positionnerait au niveau de l'extrémité de la cassure double-brin *(DeFazio, 2002)*. Le complexe ternaire Ku-(DNA-PKcs)-ADN constitue un ensemble appelé holoenzyme DNA-PK. Suite à la formation de cet holoenzyme, la kinase DNA-PKcs est activée et elle phosphoryle plusieurs protéines dont elle-même, Ku70/Ku80, la Ligase4, XRCC4 et Cernunnos. L'holoenzyme DNA-PK joue un rôle majeur dans les étapes du NHEJ qui suivent la reconnaissance de la cassure. Premièrement, il est impliqué dans le maintien des

deux extrémités de la CDB à proximité en vue de leur ligature et pour prévenir une ligature incorrecte avec une autre CDB *(DeFazio, 2002)* (**Figure 6**). Cette étape de maintien des extrémités est appelée formation de la synapse (du grec *syn* ensemble et *haptein* saisir).

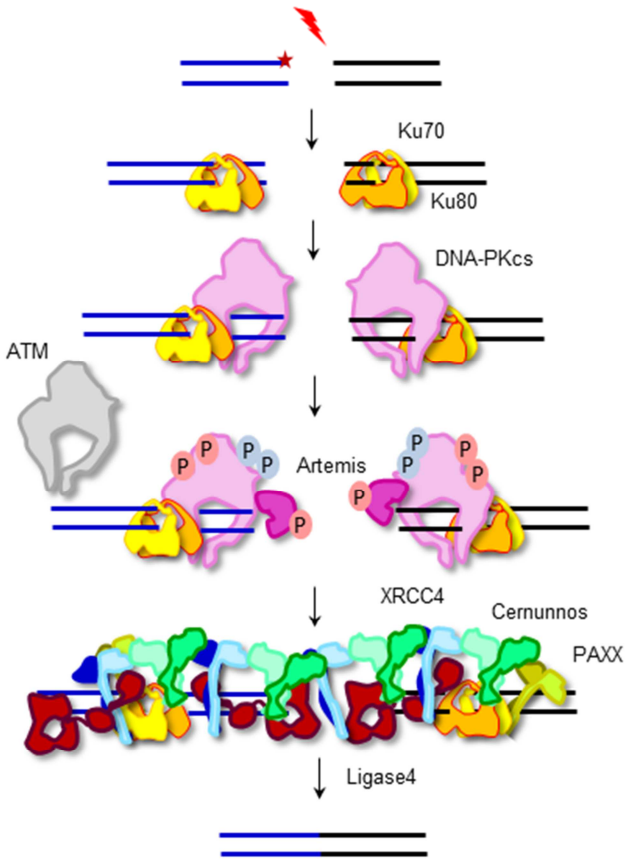


Figure 6 : Un des schéma de la voie de réparation NHEJ chez l'homme*. L'hétérodimère Ku70/Ku80 reconnait avec une forte affinité les deux extrémités d'une CDB et recrute la kinase DNA-PKcs. Le complexe Ku-PKcs-ADN appelé holoenzyme DNA-PK. Il effectue 4 actions : (i) il maintient les extrémités de la CDB à proximité (formation de la synapse), (ii) il active PKcs qui s'autophosphoryle (P en rose), (iii) il recrute Artemis et la phosphoryle et (iv) il recrute le complexe de ligation Ligase4-XRCC4-Cernunnos. Il a été proposé que le complexe XRCC4-Cernunnos forme des filaments pouvant contribuer au maintien des extrémités des CDBs. Récemment, il a été décrit qu'ATM jouerait un rôle central dans l'étape de maturation des extrémités de la CDB tandis que les phosphorylations médiées par PKcs auraient un rôle prépondérant pour le recrutement du complexe de ligation près de la CDB. Un facteur récemment identifié, PAXX, homologue de XRCC4 et Cernunnos, serait impliqué également dans l'étape de ligature suite à son recrutement par Ku.*

Deuxièmement, DNA-PK favorise le recrutement aux extrémités de nucléases comme l'endonucléase/exonucléase Artemis, de kinases et de phosphatases pour préparer les extrémités en vue de la ligature *(Goodarzi, 2006)*. Troisièmement, DNA-PK recrute le complexe de ligature Ligase4-XRCC4 *(Calsou, 2003)* afin de réaliser l'étape finale de ligature. Le complexe Ligase4-XRCC4 est constitué d'un monomère de Ligase4 et d'un dimère de XRCC4, aussi

appelé $LX4_2$ *(Wu, 2009)*. Le complexe $LX4_2$ effectue l'étape finale de ligature au niveau de la synapse à l'aide d'un troisième facteur de ligation, la protéine Cernunnos/XLF *(Ahnesorg, 2006, Buck, 2006)*. La structure cristallographique du complexe XRCC4-Cernunnos a été déterminée par quatre laboratoires indépendamment dont notre équipe *(Andres, 2012, Hammel, 2011, Ropars, 2011, Wu, 2011)*. Ces structures complétées par des analyses SAXS *(Hammel, 2010)* et de microscopies électronique *(Ropars, 2011)* ont révélé une organisation en filaments du complexe XRCC4-Cernunnos. Il a été proposé que ce(s) filament(s) pourrai(en)t contribuer à la formation de la synapse lors de l'étape de ligation **(Figure 6)**. Des analyses *in vitro* sont venues documenter l'activité potentielle de pontage des filaments XRCC4-Cernunnos *(Roy, 2012)*. Plus récemment, une première visualisation *in cellulo* de ces filaments a été reportée par microscopie super-résolution *(Reid, 2015)*.

Des travaux récents ont mis en évidence un rôle d'un nouveau facteur PAXX (Paralog of XRCC4 and XLF) *(Ochi, 2015, Xing, 2015)*. Ce nouveau facteur interagit directement avec Ku. Il a été proposé que PAXX joue un rôle dans l'étape de ligature. Les cellules PAXX-/- sont très peu radiosensibles comparées aux cellules Ligase4-/- ou XRCC4-/- suggérant un rôle secondaire de PAXX dans la voie NHEJ ou un rôle dans une autre voie de réparation *(Ochi, 2015)*. Une seconde étude récente a mis en évidence les rôles respectifs d'ATM et de PKcs dans les étapes de préparation des extrémités et de ligation. En utilisant un modèle de souris présentant un mutant de PKcs sans activité kinase (PKcs kinase dead ou KD ; PKcs avec la mutation D3922A), les auteurs ont montré un rôle de la protéine ATM et de son activité de phosphorylation de PKcs dans la régulation des étapes de maturation (processing) *(Jiang, 2015)* et que le mutant PKcs(KD) effectue correctement les étapes de maturation médiées par Artemis suggérant que cette étape est indépendante de l'activité kinase de PKcs. En revanche, l'absence d'activité kinase sur PKcs semble bloquer le recrutement du complexe de ligation et suggère une régulation de ce recrutement par l'auto-phosphorylation de PKcs **(Figure 6)**.

Dans une dernière étape, le complexe Ku doit être ôté de l'ADN. Il est admis que l'hétérodimère Ku n'existe que sous sa forme complexé en anneau. Sa présence vers l'intérieur de l'ADN au moment de la réparation a longtemps intrigué la communauté du NHEJ quand à la façon d'ôter le ou les molécules de Ku « piégées » suite à la ligature. Des

travaux récents ont montré que suite à la réparation de la CDB, Ku est ubiquitiné par un processus dépendant de RNF8 et que cette réaction permettrait de dégrader les molécules de Ku « piégées » *(Feng, 2012)*. Le schéma présenté de la voie NHEJ ci-dessus, correspond à un modèle possible d'enchaînement des étapes de reconnaissance, préparation et ligature. Nous présentons ci-dessous trois variantes autour de ce modèle qui correspondent à trois types de CDBs différentes : (1) Cassure double-brin prête pour la ligature par la Ligase4, (2) CDBs générées lors d'un réarrangement de gènes V(D)J, (3) CDBs générées lors du processus de changement de classe isotypique. Ces exemples illustrent la grande flexibilité de la voie NHEJ.

4.1.2 Cassure double-brin prête pour la ligature par la Ligase4

Des CDBs préparées par les enzymes décrites au **paragraphe 2** peuvent présenter des extrémités 3'-OH/5'-P compatibles avec une ligature directe par la ligase du NHEJ, la Ligase4 (**Figure 2**). Dans ce contexte, il a été proposé que le complexe Ku-ADN recrute directement la Ligase4 pour réaliser la ligature. Le recrutement du complexe de ligation est rendu possible par deux interactions décrites entre Ku et le complexe ternaire Ligase4-$XRCC4_2$-$Cernunnos_2$. La première interaction est une interaction directe entre le complexe Ku-ADN et la Ligase4, dépendante de la présence d'ADN *(Costantini, 2007).* La seconde interaction est une interaction directe entre Ku et Cernunnos indépendamment de la présence d'ADN *(Grundy, 2013)*.

4.1.3 CDB programmée (I) : réarrangement de gènes V(D)J

Afin de combattre un ensemble large d'agents infectieux possibles, le système immunitaire des vertébrés génère un ensemble très varié d'immunoglobulines et de récepteurs des cellules T. Chez de nombreuses espèces, cette diversité est mise en place par le processus de recombinaison V(D)J. La recombinaison V(D)J est un des deux processus physiologiques de cassures double-brin programmées observés dans des cellules somatiques. En recombinant de nombreuses séquences codant, la V(D)J est capable d'assembler des millions de gènes d'immunoglobulines et de récepteurs des cellules T. La V(D)J a lieu dans les cellules B et T précoces du système immunitaire *(Schatz, 2011)*.

La V(D)J est initiée par un complexe formé par les recombinases RAG1 et RAG2 **(Figure 7).** Ces enzymes sont exprimées seulement dans les cellules B et T précoces. Le complexe RAG1/RAG2 reconnait spécifiquement les séquences signal de recombinaison RSS (pour Recombination Signal Sequences) et introduit une cassure double-brin spécifiquement aux bords de ces séquences RSS. Les séquences RSS sont des séquences consensus de 7 et 9 nucléotides séparées par des séquences non conservées de 12 ou 23 nucléotides. Une réaction de recombinaison V(D)J nécessite strictement un site RSS de chaque type, un site 12-RSS et un site 23-RSS (condition appelée règle 12/23) *(Schatz, 2011)*.

L'étape d'incision débute par la reconnaissance d'un premier RSS (12-RSS ou 23-RSS) suivi de la reconnaissance du second. Les sites RSS sont distribués sur le génome pour permettre de recombiner soit des gènes V et J ou des gènes V, D et J. Les protéines RAG1/RAG2 sont assistées par les protéines de la famille HMGB1 (High Mobility Group Box) qui stimulent l'activité de reconnaissance et de clivage du complexe RAG. Les protéines HMGB1 possèdent deux domaines de 80 acides aminés qui sont capables de tordre la structure d'ADN et de se fixer à ces structures tordues.

Le complexe RAG1/RAG2 forme un complexe appelé complexe synaptique par rapprochement d'une séquence 12-RSS et 23-RSS. Il réalise une première incision (nick) au bord d'une séquence RSS puis utilise l'extrémité de l'incision créée pour faire une attaque nucléophile sur le brin d'ADN complémentaire. Cette réaction aboutit à la formation de deux structures en épingle à cheveux aux extrémités des deux séquences codantes. Les deux autres extrémités des CDBs générées correspondent aux extrémités des régions non codantes également appelées régions signal et sont constituées d'extrémités franches 5'-phosphorylées. A l'issue de cette étape, le complexe RAG1/RAG2 maintient les extrémités des CDBs aux bords des régions codantes ou non-codantes à proximité à travers des complexes appelés complexes synaptiques.

La voie NHEJ intervient à partir de cette étape pour reconnaitre, préparer et relier les cassures double-brin générées. Le complexe PKcs-Artemis est recruté sur les extrémités des séquences codant et l'activité endonucléase structure-spécifique d'Artemis permet l'ouverture de la structure en épingle à cheveux *(Schatz, 2011)*. L'incision réalisée par Artemis n'est pas effectuée à l'apex de la structure en épingle à cheveux et aboutit à une CDB avec

une extrémité 3’-sortante de 4 nucléotides qui nécessite l’ajout de quelques nucléotides par des polymérases *(Lieber, 2010)*.

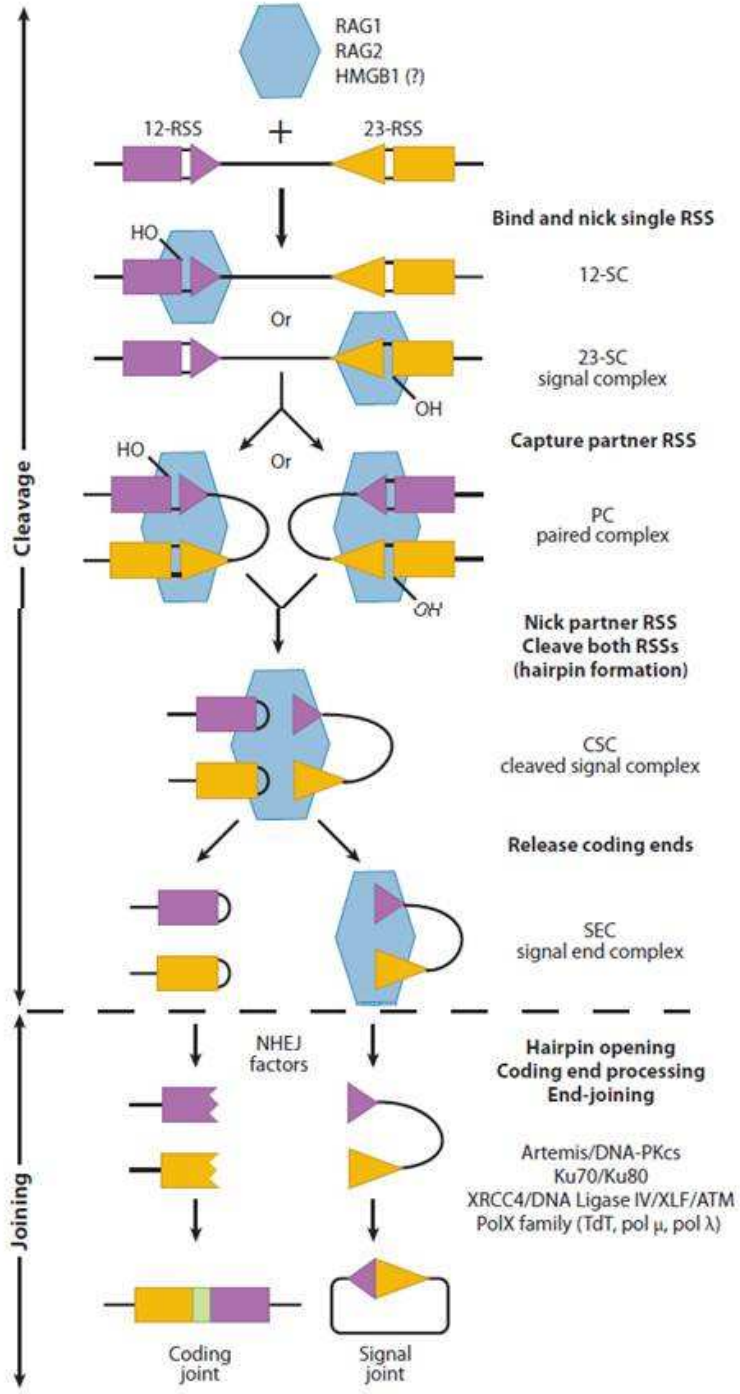


Figure 7 : La recombinaison V(D)J utilise la voie NHEJ*. Dans les premières étapes, le complexe RAG1/RAG2 reconnait successivement une première séquence RSS (12- ou 23-RSS) suivie de la seconde RSS (resp 23- ou 12-RSS). Les recombinases effectuent alors une incision et réalisent la formation d’une structure d’ADN simple brin en épingle à cheveux. Cette étape aboutit à la formation d’une séquence codant en épingle à cheveux et d’une séquence signal à bouts francs. Ces extrémités sont maintenues dans un complexe de clivage transitoire (CSC pour Clivage Sequence Complex). Les séquences codant se dissocient des RAG et les séquences signal restent fixées (SEC pour Signal End*

Complex). Les séquences codant et signal sont réparées de manière asynchrone par la voie NHEJ pour donner une jonction signale réparée de façon fidèle et une jonction codante qui est généralement réparée de manière infidèle (d'après (Schatz, 2011)).

La ligature des extrémités codant et signal met alors en jeu les facteurs classiques de la voie NHEJ : Ku, Ligase4, XRCC4 et Cernunnos. Des polymérases de la famille PolX (TdT, polμ et polλ) insèrent au niveau de la séquence codante des nucléotides de façon indépendante d'une séquence d'ADN matrice *(Lieber, 2010)*. Cette activité des PolX contribue à l'introduction d'une diversité supplémentaire au niveau des jonctions codantes V-J et J-D. La jonction des séquences signal est réalisée de façon précise.

4.1.4 CDB programmée (II) : changement de classe isotypique

Le changement de classe isotypique (CSR pour Class Switch Recombination) est un processus observé dans les cellules B après la fin du processus de recombinaison V(D)J. La CSR représente la seconde forme de CDB programmée physiologique dans des cellules somatiques. Chez les mammifères, la CSR assure le changement de gènes sur la chaine lourde des immunoglobulines. Ce processus permet le changement de gènes Igμ pour les IgM vers un gène Igγ, Igα ou Igε pour former les IgG, IgA ou IgE. La CSR débute avec la conversion par la cytidine déaminase (AID pour Activation-Induced cytidine Deaminase) de cytidine (C) en uracile (U). Suite à cette conversion, la glycosylase UNG convertit ces uraciles en sites abasiques qui sont eux-mêmes convertis en sites d'incision par l'endonucléase APE1. La participation d'autres enzymes comme l'exonucléase 5'-3' EXO1, contribue à convertir ces incisions formées sur les brins supérieurs et inférieurs en cassure double-brin. La voie NHEJ est la voie majoritairement impliquée dans la réparation des CDBs formées lors de ce processus. Il a été montré qu'en absence de Ligase4, les Ligases 1 et 3 pouvaient jouer un rôle de voie de secours (backup) *(Lieber, 2010)*.

4.2 Voie de réparation par Recombinaison Homologue (HR)

La recombinaison homologue (HR pour Homologous Recombination) permet la réparation des cassures double-brin de façon fidèle en s'appuyant sur la présence d'une chromatide sœur non-endommagée. Cette voie prend en charge les CDBs créées par des radiations mais également celles issues de liaisons covalentes inter-brins et de fourches de

réplication bloquées *(San Filippo, 2008).* Chez les vertébrés, les mutations des gènes principaux de la voie HR sont fréquemment létales soulignant l'importance de cette voie.

Les étapes principales de la voie HR sont (**Figure 8) :** (i) la résection des cassures par des nucléases pour donner des extrémités 3' sortantes, (ii) la formation d'un filament nucléo-protéique entre une recombinase et un simple brin d'ADN (ssDNA), (iii) l'invasion d'un brin d'ADN contenant une séquence homologue avec formation d'une boucle intermédiaire appelée D-loop, (iv) l'extension par une ADN polymérase du brin 3' envahissant, (iv) la capture de la seconde CDB par fusion avec la D-loop en cours d'extension, (v) la formation de jonctions de Holiday (HJs pour Holidays Junctions) et finalement (vi) la résolution des HJs pour donner des structures de type non-enjambement ou enjambement (non-crossover ou crossover) *(San Filippo, 2008)*.

Les facteurs au cœur des voies HR eucaryotes ont été pour la plupart identifiés chez *S. Cerevisiae* en premier. La résection est une étape majeure dans l'orientation de la réparation des CDBs vers la voie HR plutôt que vers la voie NHEJ. Le recrutement de nucléases comme le complexe MRN (Mre11-Rad50-Nbs1) ou EXO1 aux extrémités des CDBs initient la résection des extrémités et le chargement de la recombinase RAD51 **(Figure 8)**. Les études structurales par cristallographie et microscopie électronique ont mis en évidence la formation par RAD51 de structures oligomériques en hélice sur de l'ADN simple-brin sur des longueurs importantes (plusieurs milliers de bases). Ce filament est appelé filament pré-synaptique. L'ADN simple brin présente dans ces filaments une structure étendue avec 6 molécules de Rad51 par tour d'hélice du filament et un pas de 100Å *(Ogawa, 1993)*. Le filament pré-synaptique capture l'ADN de la chromatide sœur, recherche des séquences homologues puis réalise un échange de brin avec l'ADN double-brin ciblé. La translocase ATP-dépendante RAD54 stimule cette étape d'échange de brins et la formation de la D-loop.

Différents facteurs jouent un rôle dans la formation du nucléo-filament RAD51-ssDNA et dans l'échange de brin. Il s'agit de l'hétérotrimère RPA, qui présente des propriétés de fixation à l'ADN simple brin. RPA est une protéine abondante du noyau qui peut dissoudre les structures secondaires de l'ADN simple brin. A travers cette activité les protéines RPA facilitent la formation du filament pré-synaptique mais elles entrent également en compétition avec RAD51 pour la fixation à l'ADN simple brin (ADNsb). La protéine RAD52 présente une activité de fixation à RAD51, à RPA et à l'ADNsb. La structure cristallographique

de la région N-terminale conservée de RAD52 a révélé une structure oligomérique en anneau avec 11 sous-unités *(Kagawa, 2002)*. La surface extérieure de cet anneau présente une gorge riche en acides aminés basiques et aromatiques. L'étude par mutagenèse dirigée de cette gorge a montré son rôle dans la reconnaissance de l'ADN simple brin. Il a été proposé que RAD52 joue un rôle dans le dépôt de RAD51 sur un filament RPA-ADNsb et donc favorise l'échange de brin **(Figure 8)**.

Enfin, le gène suppresseur de tumeur BRCA2 a été montré comme important pour la formation des foyers de RAD51. Ce gène présente des tailles très variables chez les eucaryotes, de 3418aa chez l'homme à 383aa chez *C. elegans.* BRCA2 humain interagit avec RAD51 à travers 8 motifs BRC de 30aa environ chacun et un motif en position C-terminale de la protéine. La structure du complexe entre un domaine de RAD51 et un motif BRC de BRCA2 *(Pellegrini, 2002)* a montré que ce motif vient interagir sur le site d'oligomérisation de RAD51. Il a été proposé que la protéine BRCA2 fixe une face d'un monomère de RAD51 et permettrait la nucléation du nucléo-filament sur l'autre face de ce monomère jouant un rôle stimulateur.

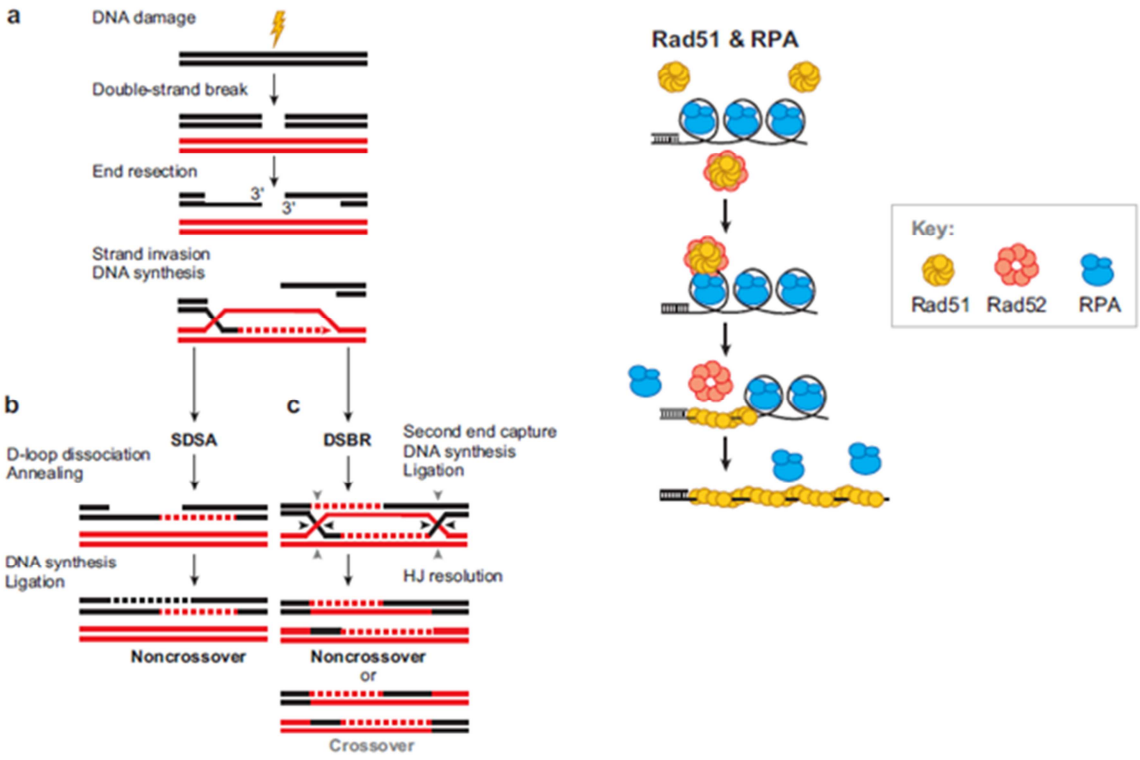


Figure 8 : Voie de Réparation par Recombinaison Homologue. *La voie de réparation par recombinaison homologue.* ***(Droite)*** *La recombinaison homologue regroupe deux voies : une voie par*

*fusion de simple-brin et synthèse d'ADN (SDSA Synthesis-Dependent Strand Annealing) et une voie appelée DSBR pour Double-Strand Break Repair. Les étapes principales décrites dans le texte sont représentées. (**Gauche**) La protéine RAD52 joue un rôle de médiateur des premières étapes d'échange de brin. RAD52 forme un complexe avec RAD51 et le dépose sur un simple brin d'ADN couvert de RPA. Ce dépôt permettrait de former progressivement le filament pré-synaptique et de déplacer les protéines RPA (d'apres (San Filippo, 2008)).*

Dans le cadre de la voie SDSA, la recombinaison s'achève avec la dissolution des intermédiaires et la recapture des ADN simple-brin échangés avec leur ADN complémentaire de départ. Dans le cadre de la voie DSBR, la recombinaison s'achève par dissolution des jonctions de Holiday par des nucléases appelées résolvases et la formation selon le site d'incision à des produits présentant des crossover (échange de matériel génétique) ou des non-crossover.

4.3 Voie NHEJ alternative (a-NHEJ)

L'existence d'une voie alternative à la voie classique NHEJ est venue en premier de travaux utilisant des levures mutées sur les gènes Ku70, Ku80 ou Ligase4 et sur lesquelles était observée une activité résiduelle de jonction des extrémités *(Boulton, 1996)*. L'activité observée montrait des délétions à l'issue de la réparation et semblait dépendre de courtes homologies de séquences aux jonctions. Ces activités résiduelles Ku- et Ligase4-indépendantes furent alors observées dans de nombreux organismes : *A. Thaliana, C. elegans, X. Tropicalis,* chez les rongeurs et chez l'homme *(Wang, 2003)*. La voie a-NHEJ inclut la réparation appelée MMEJ (pour Microhomology-Mediated End Joining) et de façon plus générale les mécanismes de jonction aux extrémités Ku- et Ligase4-indépendants. La voie a-NHEJ est initiée par la poly(ADP-ribose)-polymerase 1 (PARP1). La PARP1 est un senseur des cassures simple- et double-brin. Lors de la fixation sur ces dommages, PARP1 est activée et catalyse la poly(ADP-ribosyl)ation d'elle-même et des protéines présentes sur ces dommages. La PARP1 a été proposée comme la protéine responsable dans les premières étapes de la voie a-NHEJ du maintien des extrémités de la CDB à proximité. Suite à cette étape de reconnaissance, la PARP1 recruterait un certain nombre de facteurs pour la préparation des extrémités comme la PNK, le complexe MRN et des facteurs pour la ligation comme le complexe XRCC1-Ligase3.

5 Caractéristiques des protéines de la voie NHEJ

5.1 Hétérodimère Ku70/Ku80

Ku a été identifié pour la première fois dans les années 1980 comme un auto-antigène ciblé par des auto-anticorps dans le sérum de patients présentant des maladies auto-immunes de la famille des sclérodermies (affections cutanées et vasculaires). Le nom Ku provient des deux premières initiales du premier patient, K.U., dont le sérum permit d'isoler l'autoantigène. Il fut montré dans un premier temps que Ku est une protéine principalement nucléaire et abondante, capable de lier avec une forte affinité les extrémités d'un ADN double brin *(Blier, 1993, Mimori, 1986)*. Le rôle de Ku dans la réparation des CDBs fut mis en évidence par la capacité des gènes de ses deux sous-unités à complémenter des cellules radiosensibles (gène Ku80 initialement appelé XRCC5 pour X-ray Repair Cross-Complementing protein 5 et gène Ku70 initialement appelé XRCC6 pour X-ray Repair Cross-Complementing protein 6).

Les protéines Ku70 et Ku80 contiennent 3 domaines **(Figure 9) :** (i) un domaine N-terminal de type von Willebrand (vWA) constitué d'hélices α et brins β ; (ii) un domaine cœur central qui permet l'hétérodimèrisation ; (iii) un domaine hélical C-terminal. La structure cristallographique de Ku70/Ku80 seul et d'un complexe Ku70/Ku80-ADN a montré que le domaine cœur forme un anneau capable de reconnaître environ 14 paires de bases *(Walker, 2001)*. Cet anneau est formé de structures secondaires entrelacées de Ku70 et Ku80. L'anneau présente une surface d'interaction chargée positivement qui interagit de manière séquence indépendante avec l'ADN. Ku possède une forte affinité pour l'ADN double-brin (Kd de 0,5nM en présence de 220mM NaCl, Kd mesuré par anisotropie de fluorescence) *(Arosio, 2004).* Cette interaction est très sensible à la force ionique, le Kd augmente de 60 fois lorsque la concentration en NaCl passe de 220mM à 300mM (Kd de 27nM en présence de 300mM NaCl). Ku reconnait indifféremment un ADN avec une extrémité franche, 3'-sortante, 5'-sortante ou avec une structure en épingle à cheveux de 8 nucléotides (8 thymidines) *(Arosio, 2002)*. Ku reconnait l'ADN de manière indépendante de la séquence de l'ADN.

La structure cristallographique avec un ADN suggère que Ku70/Ku80 se fixe à une CDB avec une orientation préférentielle. La sous-unité Ku70 serait positionnée vers

l'extrémité de l'ADN et Ku80 vers l'intérieur *(Walker, 2001)*. Il se fixe préférentiellement aux extrémités d'un fragment d'ADN mais il est également capable de glisser vers l'intérieur à des concentrations supérieures *(de Vries, 1989) (Arosio, 2004)*. Toutes les études suggèrent que Ku ne peut se fixer sur l'ADN que part une extrémité et ne se dissocier que par une extrémité ou par dégradation après ubiquitination.

Le domaine N-terminal vWA, aussi appelé α/β, forme un repliement de type repliement de Rossman avec un feuillet β constitué de six brins β. Le domaine vWA est présent dans plusieurs protéines de la matrice extracellulaire. Il est impliqué dans de nombreuses interactions protéine-protéine. Il a été proposé comme jouant un rôle dans l'interaction entre Ku et une sous-unité du complexe shelterin présent au télomère, la protéine TRF2 (Telomere Repeat binding Factor 2) *(Ribes-Zamora, 2007)*. Il a également été proposé que ce domaine participe à l'interaction entre Ku et un facteur de la voie NHEJ, APLF (Aprataxin and PNKP like factor) *(Shirodkar, 2013)*.

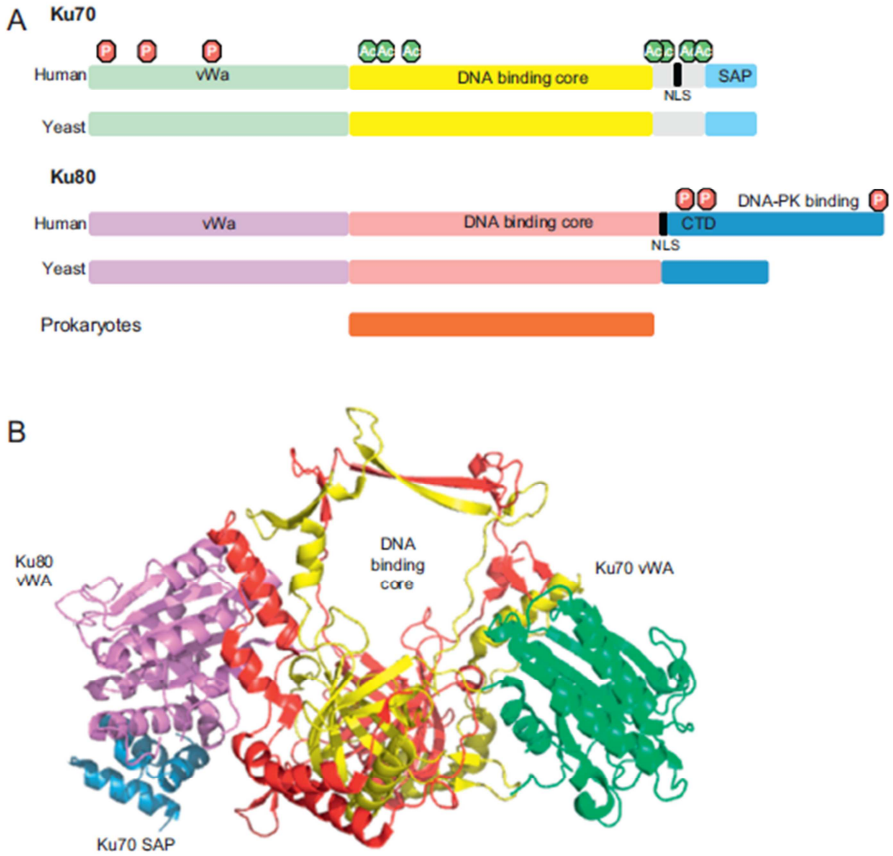


Figure 9 : Hétérodimère Ku70/Ku80 *(A) Représentation schématique des sous-unités Ku70-Ku80. Ces sous-unités contiennent un domaine N-terminal vWA de type α/β, un domaine central de fixation à l'ADN et un domaine C-terminal. Le CTD de Ku80 contient le motif de fixation à DNA-PKcs (12 derniers aa). Le CTD de Ku70 est plus court et contient le domaine SAP. La position des NLS et des sites modifications post-traductionnelles*

(phosphorylation et acetylation) sont indiqués sur les sous-unités humaines. Ku de levure a une structure similaire hormis le domaine CTD de Ku80 (absence de motif de fixation à DNA-PKcs). Les procaryotes possèdent un homodimère Ku dont la séquence est homologue au domaine de fixation à l'ADN. (B) Structure cristalline de Ku (pdb 1JEY) colorié selon les couleurs utilisées en (A). L'hétérodimère forme un panier asymétrique avec un anneau chargé positivement qui reconnait environ 14bp. La région CTD de Ku80 n'est pas présente dans la structure. (d'après (Fell, 2015))

Il a été montré que l'hétérodimère Ku présentait *in vitro* une activité dRP/AP lyase pouvant contribuer à la préparation des extrémités des CDBs pour la ligature *(Strande, 2014)*. Cette activité semble portée principalement par deux lysines du domaine vWA, les lysines 31 et 160. L'activité AP-lyase ôte les sites abasiques (AP) en introduisant une incision en 3' du site AP par un mécanisme faisant intervenir un intermédiaire de type base de Schiff.

Les deux sous-unités de Ku possèdent un domaine C-terminal (CTD) qui est la partie la moins conservée entre les deux séquences. La région CTD de Ku80 a une taille de 15kDa et contient une région hélicale et une région désordonnée. Chez les vertébrés, la région la plus C-terminale de ce CTD et notamment les 12 derniers acides aminés sont impliqués dans le recrutement de DNA-PKcs *(Gell, 1999)*. Des organismes eukaryotes comme *S. cerevisiae* ou *A thaliana,* qui ne possèdent pas d'homologue de DNA-PKcs, présentent une sous-unité Ku80 qui ne possède pas cette région de fixation à DNA-PKcs. La région CTD de Ku70 présente une longue charnière et un domaine structuré de 5kDa en hélice-boucle-hélice appelé SAP domaine (pour SAF-A/B, Acinus et PIAS motifs) qui présente des propriétés de fixation à l'ADN et contribue à améliorer l'affinité de Ku pour l'ADN *(Wang, 1998)*. Cette région C-terminale de Ku70 est le site d'acétylations et de sumoylations qui réguleraient le recrutement de Ku aux CDBs. Enfin, les CTDs de Ku70 et Ku80 possèdent un signal de localisation nucléaire qui régule le transport de Ku au noyau.

L'abondance de Ku et sa forte affinité pour l'ADN permette à ce facteur de s'associer aux CDBs en un laps de temps estimé à 5s *(Mari, 2006)*. La présence de Ku joue un rôle important pour prévenir des processus de résection des extrémités par les enzymes des voies de réparations HR ou a-NHEJ. En effet dans des cellules déficientes en Ku, les extrémités sont soumises à des activités de dégradation *(Mimitou, 2010)*. A l'inverse, Ku présente une faible affinité pour un ADN ayant subi une forte résection et présentant une longue région simple brin. Des études récentes du laboratoire de S Jackson ont permis pour la première fois d'évaluer la quantité d'hétérodimère Ku aux extrémités d'une CDB. L'utilisation conjointe d'un traitement ribonucléase, d'extraction avec un détergent et de

microscopie à super-résolution a permis d'évaluer à deux le nombre de Ku aux bords d'une CDB soit un par extrémité *(Britton, 2013)*. Une fois positionné, Ku semble jouait le rôle de plateforme de recrutement pour l'ensemble des protéines et complexes de la voie NHEJ.

5.2 DNA-PKcs

DNA-PKcs est la plus grande kinase humaine connue avec 4128 acides aminés (469kDa). Elle a une forte affinité pour l'ADN (Kd de 3nM pour l'ADN à bout franc). Cette affinité est encore renforcée en présence de Ku (Kd de 30pM) *(West, 1998)*. Une fois fixé au complexe Ku-ADN, l'activité serine/threonine kinase de DNA-PKcs est activée. Cette activité est dirigée en premier sur la kinase elle-même avec près de 15 sites d'auto-phosphorylation identifiés *(Jette, 2015)*.

DNA-PKcs présente un large domaine N-terminal, constitué des 2800 premiers acides aminés. Ce domaine est formé principalement de motifs HEAT répétés (Huntingtin Elongation factor 3, regulatory subunit A of PP2A, TOR1) et de domaines en hélices additionnels. PKcs possède une région C-terminale avec un domaine FAT (FRAP, ATM, TRRAP), un domaine kinase et un domaine FAT-C. Cette organisation générale est commune aux protéines de la famille PI3K (Phosphatidyl Inositol 3 Kinase) comme ATM et ATR. Les études de cryo-EM ont montré que PKcs a une forme en berceau concave avec une cavité centrale *(Spagnolo, 2006, Williams, 2008)*. L'utilisation d'ADN biotinylé et marqué avec des particules d'or a permis de mettre en évidence la fixation d'ADN dans cette cavité *(Boskovic, 2003)*. Un tour de force remarquable a été en 2010, la publication par le laboratoire de T Blundell de la structure cristalline de la DNA-PKcs à basse résolution (6.6Å, pdb 3kgv) (**Figure 10**) *(Sibanda, 2010)*. Les cristaux ont été obtenus à partir de la DNA-PKcs endogène et de larges cultures de cellules HeLa (100L par purification pour la cristallogenèse). Les formes cristallines étudiées correspondent à des essais de cristallisation de DNA-PKcs en complexe avec des fragments purifiés dans *E. coli* de régions C- terminale de Ku80 (Ku80 ct194 et Ku80 ct140). L'enzyme présente une large structure en berceau ouvert composé des motifs HEAT (**Figure 10**). Le domaine kinase en C-ter est situé au sommet de cette structure. DNA-PKcs mesure 160Å du sommet du domaine kinase à la base de l'anneau. L'anneau mesure environ 120Å de diamètre et est constitué de 66 hélices α qui font le tour de cet anneau. Cet anneau

a une forme de berceau vue de profil et présente une ouverture à proximité du N-terminus de la protéine.

Des analyses par SAXS de la forme phosphorylée ou non-phosphorylée de DNA-PKcs et les enveloppes calculées pour ces deux formes suggèrent que l'autophosphorylation de PKcs s'accompagne d'un changement significatif de conformation et d'un élargissement de l'ouverture située près du N-terminal de l'enzyme *(Hammel, 2010) (Dobbs, 2010)*. Le laboratoire de G Chu de son côté a pu mettre en évidence les capacités de DNA-PKcs a former une synapse entre deux extrémités d'ADN *(DeFazio, 2002)*. A l'aide d'un test astucieux de capture (pull-down) d'un ADN marqué au P32 'donneur' par un ADN biotinylé 'récepteur' en présence ou non de PKcs et/ou de Ku, ce laboratoire a mis en évidence une capacité de PKcs à former un pont (aussi appelé synapse) entre deux ADN. Cette propriété est plus forte chez PKcs que Ku. Ku renforce cette propriété de PKcs à joindre deux ADN si le test est réalisé avec des oligonucléotides de 180bp et n'a pas d'effet si des oligonucléotides de 42bp sont utilisés *(DeFazio, 2002)*. Des études par microscopie électronique ont permis à ces auteurs de confirmer par une seconde technique la propension de DNA-PKcs à former des ponts entre deux ADN avec 2 molécules de PKcs par synapse **(Figure 10)**.

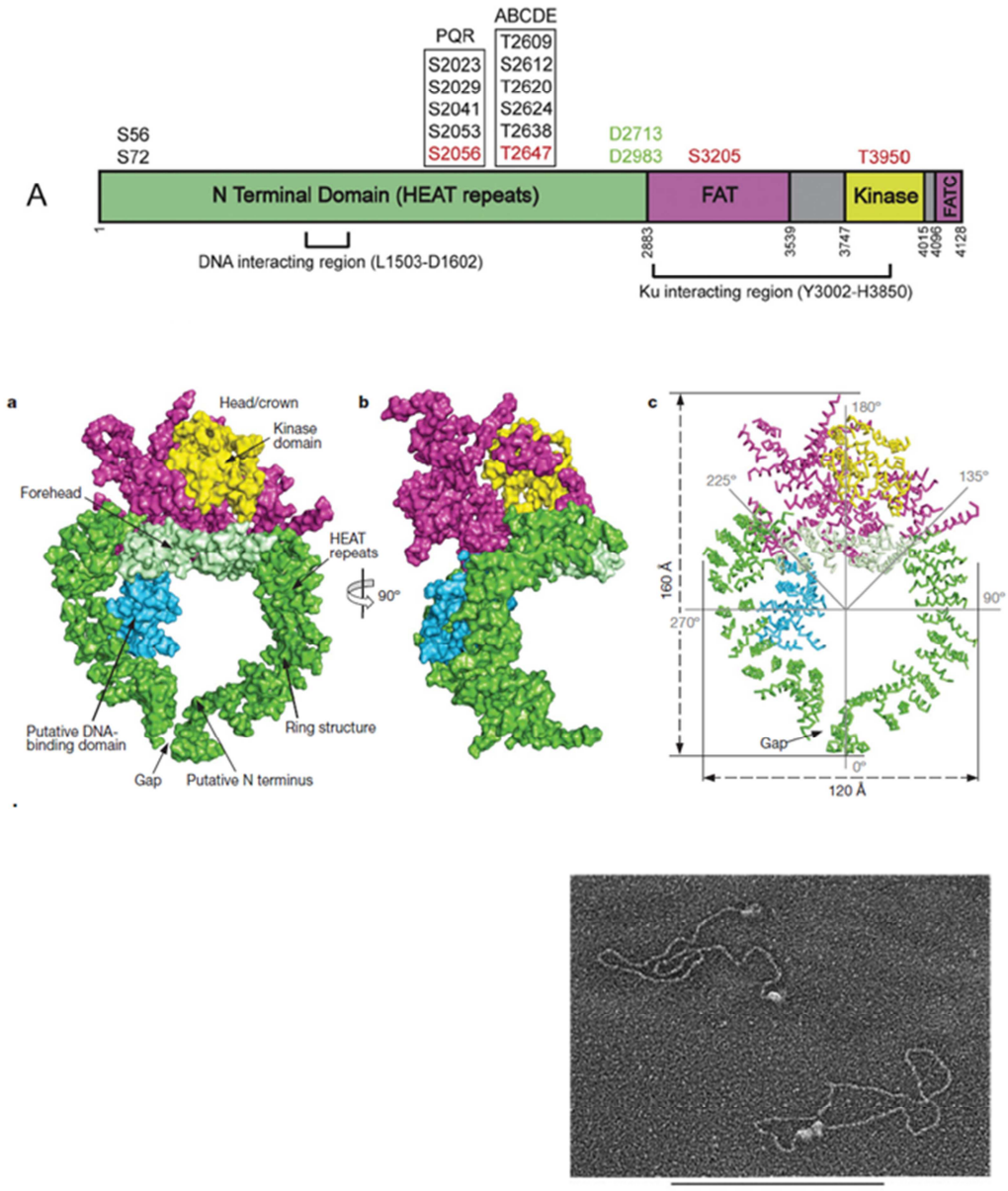


Figure 10 : DNA-PKcs. ***(Haut)*** *Séquence de DNA-PKcs avec les limites des domaines et les principaux sites de phosphorylation observés in vivo (d'après (Jette, 2015)).* ***(Milieu)*** *Structure cristalline de DNA-PKcs (Sibanda, 2010).* ***(Bas)*** *Image de microscopie électronique de pontage des deux extrémités d'un ADN plasmidique linéarisé. Image réalisée en présence du complexe DNA-PKcs/Ku. Les images sont réalisées à partir de 100ng, 2nM de plasmide pBluescript linéarisé incubé avec PKcs (22nM) et Ku (22nM) (Barre= 1kb d'ADN) (d'après (DeFazio, 2002)).*

L'activité kinase de PKcs est essentielle pour la voie NHEJ mais le rôle exact de cette activité n'est pas complètement compris *(Davis, 2014)*. *In vitro,* PKcs est capable de phosphoryler chacun des facteurs de la voie NHEJ et notamment Ku70/Ku80, XRCC4, la Ligase4 et Cernunnos. Cependant aucune de ces phosphorylations n'est indispensable pour la voie NHEJ individuellement. Une étude récente réalisée à partir d'un modèle animal DNA-PKcs(kinase inactive) a permis de découpler le rôle de l'activité kinase de PKcs dans la

préparation des extrémités et dans le recrutement du complexe de ligation *(Jiang, 2015)*. Les auteurs ont montré que les cellules avec une PKcs inactive ont une activité de préparation (end processing) intacte alors que ce mutant bloque le recrutement du complexe de ligation. Cette étude a permis de proposer un modèle selon lequel l'autophosphorylation de PKcs et son activité sur les autres enzymes permettrait de faciliter l'accès du complexe de ligation aux CDBs. Il a également été montré dans cette étude que la phosphorylation de PKcs par ATM serait importante pour l'étape de préparation des extrémités *(Jiang, 2015).*

5.3 Artemis

Artemis est un facteur essentiel de la voie NHEJ et de la recombinaison V(D)J. Artemis est une nucléase structure-spécifique de la superfamille des metallo-β-lactamases dont l'activité est initiée suite à son interaction avec DNA-PKcs *(Kurosawa, 2010)*. Une activité majeure d'Artemis est l'ouverture des structures en épingle à cheveux (hairpin) générées par le complexe RAG1/RAG2 lors de la recombinaison V(D)J **(Figure 7)**. Suite à l'action d'Artemis, la structure en hairpin donne une cassure double brin avec une extrémité 3' sortant de 4 nucléotides. Il a été proposé qu'Artemis possède également une activité 5'-3' exonucléase mais une étude récente suggère que cette activité provient d'exonucléases co-purifiées avec Artemis.

Artemis a été identifié par le laboratoire de JP de Villartay à partir de l'étude de patients présentant un syndrome de déficience sévère du système immunitaire couplé à une radiosensibilité de leurs cellules. Les patients en question étaient des Athabascan (indiens d'Amérique). Suite au clonage du gène déficient chez ces personnes, il a été montré qu'Artemis possède un domaine metallo-β-lactamase (β-lact), un domaine β-CASP (CPSF Artemis SNM1 PSO2) et un domaine C-terminal prédit comme déstructuré *(Callebaut, 2002)*. Les mammifères présentent deux autres nucléases de la famille β-lact/β-CASP, homologues d'Artemis. La première SNM1A interagit avec 53BP1 et serait impliquée dans la réparation des dommages covalents inter-brins (ICL ou inter-strands crosslinks). La seconde, SNM1B aussi appelée Apollo est impliquée dans la protection des télomères et interagit avec TRF2. La troisième, SNM1C, est la protéine Artemis.

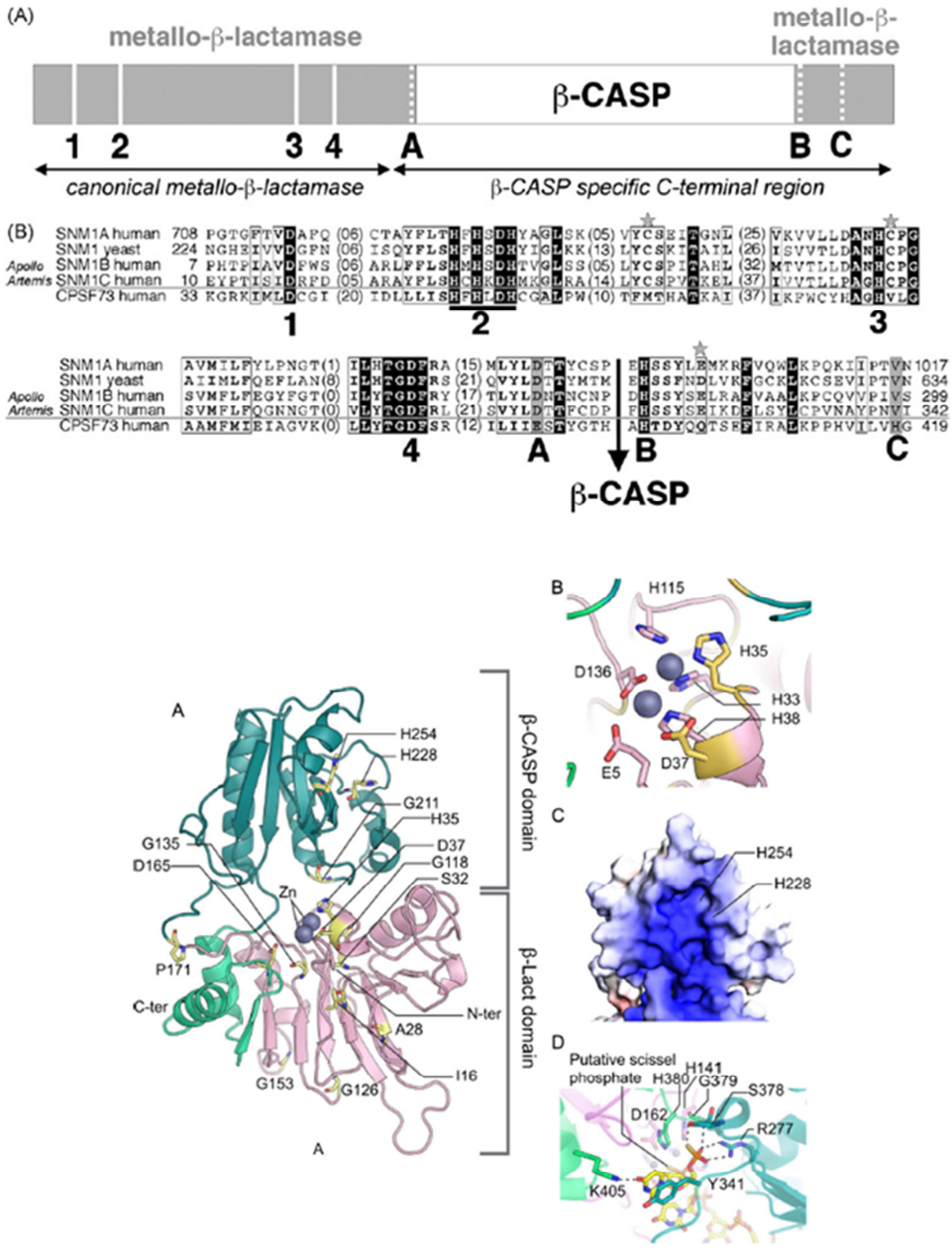


Figure 11 : Structure d'Artemis. *(**Haut**) Séquences des domaines ß-lactamase/ ß-CASP d'Artemis et des protéines SNM1-A et –B. Les motifs conservés 1-4 et A-C sont indiqués (d'après Callebaut (Callebaut, 2002)) (**Bas**) Modèle du domaine ß-lact/ ß -CASP d'Artemis à partir des protéines homologues (bas-B) site actif d'Artemis (bas-C) sillon susceptible de fixer l'ADN (bas-D) vue d'une RNAse proche d'Artemis en complexe avec un ARN (d'après (Ochi, 2014))*

Les structures cristallines de CPSF-73, Apollo et SNM1A ont été déterminées *(Mandel, 2006)(non published, structural genomic consortium)*. Ces protéines présentent entre 26% et 32% d'identité de séquence et permettent de construire des modèles par homologie d'Artemis *(Ochi, 2014)*. Les motifs conservés des protéines Artemis, 1-4 et A-C, sont situés

dans le domaine β-lactamase et semblent correspondre aux positions qui coordonnent les ions métalliques de la nucléase **(Figure 13)**. Ces cations sont des atomes de zinc dans les enzymes cristallisées. Les études biochimiques suggèrent que dans le cas d'Artemis, elle fonctionnerait avec du magnésium et/ou du manganèse. Le domaine β-CASP d'Artemis et des autres protéines SNM1, SNM1-A et-B, présente un sillon chargé positivement qui pourrait être impliqué dans la fixation de l'ADN. La région C-terminale prédite comme déstructurée semble avoir une fonction dans la régulation de l'activité d'Artemis. On y trouve deux motifs, 399-404 et 485-495, qui sont respectivement impliqués dans les interactions d'Artemis avec DNA-PKcs et avec Ligase4.

5.4 La Ligase4

L'étape de ligature est réalisée par le complexe Ligase4/XRCC4/Cernunnos-XLF. Ces deux derniers facteurs jouent un rôle sur la stabilité et l'activité de la Ligase4 (voir paragraphe suivant). La Ligase4 est une des trois ligases humaines présentes chez tous les eucaryotes *(Ellenberger, 2008)*. La Ligase4 présente en N-terminal un domaine catalytique similaire aux deux autres ligases eucaryotes, la Ligase 1 et la Ligase 3, puis deux domaines BRCT en tandem en position C-terminale, propriété unique de la Ligase4. La région catalytique est constituée d'un domaine de fixation à l'ADN (DBD pour DNA Binding Domain) en position N-ter, d'un domaine d'adenylation ou nucleotide-transférase (NTD pour NucleotidylTransferase Domain) et d'un domaine de repliement OB (OBD pour OB-fold Domain) **(Figure 13)**. La réaction de ligature se déroule en trois étapes (étape 1) l'adénylation (transfert d'un AMP) sur une lysine de la ligase, (étape 2) le transfert de l'AMP sur le 5'-phosphate, (étape 3) la ligature du 5'-phosphate sur le 3'-OH avec relargage de l'AMP. Les études structurales des Ligases 1 et 3 eucaryote et de ligases d'archée, ont montré que les ligases présentent trois conformations différentes, un état libre et ouvert, un état libre et fermé et un état en complexe avec l'ADN.

La Ligase4 est capable d'établir une ligature sur des extrémités non compatibles et à travers des brèches (gap) aux extrémités de l'ADN. Une particularité de la Ligase4 est sa difficulté à être ré-adénylée pour effectuer plusieurs cycles catalytiques, propriété non-observée sur les autres ligases. Les structures cristallines des domaines DBD, AdD et OB-fold

de la Ligase4 déterminées par le laboratoire de T Blundell ont révélé des spécificités au niveau de certaines régions charnières de la Ligase4 par rapports aux Ligases 1 et 3 qui pourraient expliquer leurs différences d'efficacité *(Ochi, 2014)*. De façon intéressante, les Ligases 1 et 3 ne sont pas capables de compenser l'absence de Ligase4 dans des modèles murins déficients en Ligase4.

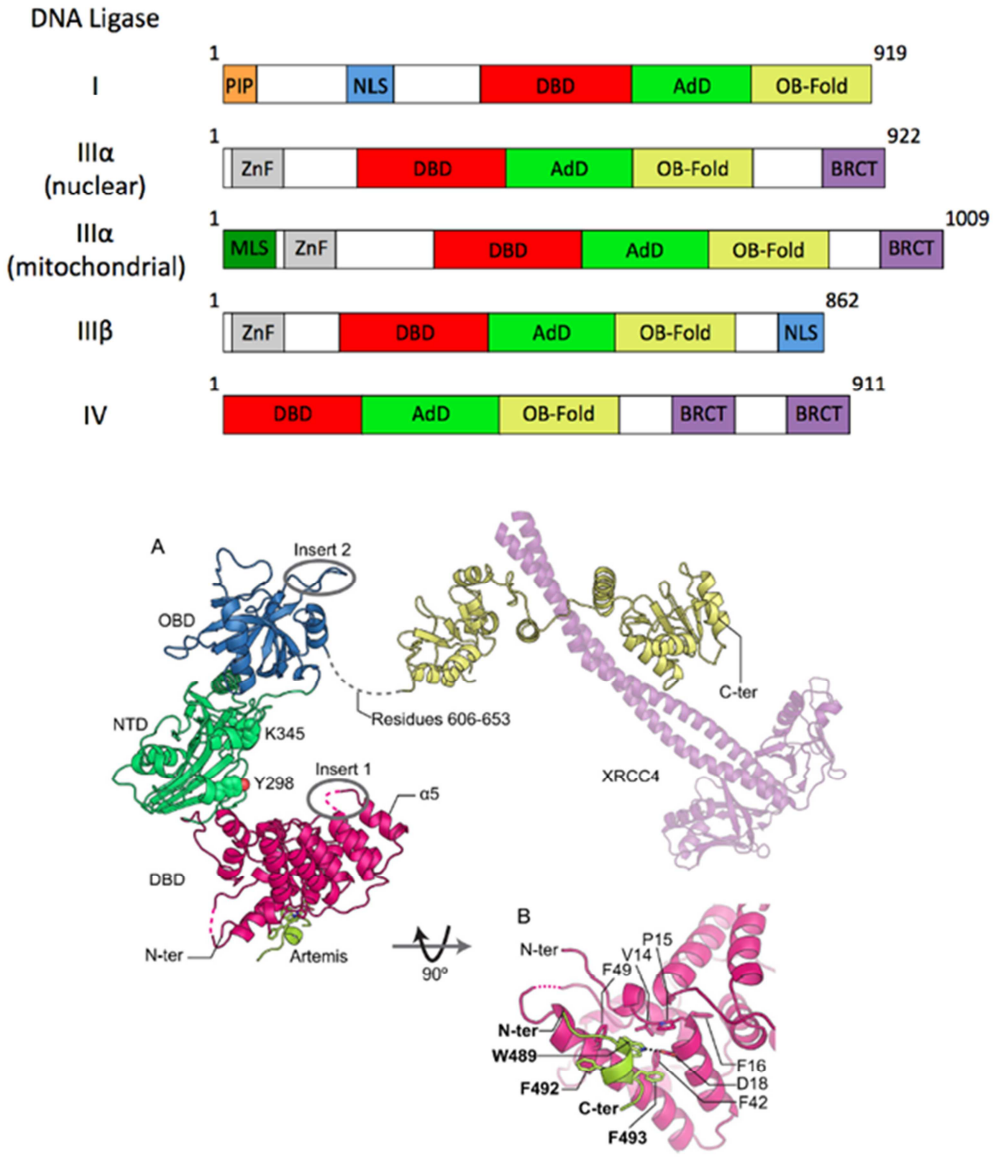


Figure 12 : Structure de la Ligase4. ***(Haut)*** *Séquences des Ligases 1, 3 et 4 eucaryotes. Toutes les ligases présentent un domaine catalytique constitué d'un domaine de fixation à l'ADN, d'un domaine d'adénylation (AdD ou NTD) et d'un domaine de repliement de type OB (OB fold). Le gène de la Ligase3 code pour plusieurs protéines. (d'après (Ellenberger, 2008))* ***(Bas)*** *Modèle de la Ligase4 à partir des structures des différents domaines de Ligase4. La structure du domaine de fixation à l'ADN a été décrite en complexe avec un motif peptidique dérivé de l'endonucléase Artemis (bas B pour détail). La région charnière entre les deux domaines BRCT est au cœur de l'interaction avec la région en faisceau d'hélices torsadées (coiled-coil) de XRCC4 (d'après (Ochi, 2014)).*

Une étude récente a montré que la Ligase4 présente une activité importante pour la voie NHEJ indépendante de son activité catalytique *(Cottarel, 2013)*. La Ligase4 joue un rôle très tôt dans l'autophosphorylation de DNA-PKcs et dans la formation de la synapse lors des premières étapes de la voie NHEJ.

La Ligase4 présente également une interaction avec une affinité modérée (5µM) avec l'endonucléase Artemis (région 485-495). Le rôle de cette interaction dans l'activité de la ligase ou de l'endonucléase n'est pas bien compris mais cette interaction pourrait jouer un rôle dans le recrutement de la ligase par Artemis aux extrémités des CDBs en cours de préparation.

5.5 XRCC4 et Cernunnos/XLF

La protéine XRCC4 forme un complexe très stable avec Ligase4 et contribue à la stabilité et l'activité de la Ligase4 *(Grawunder, 1997) (Critchlow, 1997)*. Les cellules déficientes en XRCC4 ne présentent pas de Ligase4 stable *(Bryans, 1999)*. La protéine XRCC4 est une protéine dimérique, constituée d'une région globulaire N-terminale, d'un long faisceau d'hélices torsadées (coiled-coil) et d'une région C-terminale prédite comme non-structurée. Les études structurales ont mis en évidence une stœchiométrie du complexe Ligase4-XRCC4 de 1:2 avec une interaction entre le coiled-coil d'un dimère de XRCC4 et la région C-terminale de la Ligase4. L'interaction est médiée principalement par la région charnière (linker) entre les deux domaines BRCT de la Ligase4 *(Sibanda, 2001, Wu, 2009)*.

Un second co-facteur du complexe de ligation, la protéine Cernunnos, a été identifié par le laboratoire de JP de Villartay par complémentation de cellules de fibroblastes de patients présentant à la fois une immunodéficience sévère, des cellules radio-sensibles et une microcéphalie *(Buck, 2006)*. La même protéine a été identifiée par un second laboratoire à partir d'un crible de double-hybride grande échelle des partenaires protéiques potentiels de XRCC4 humain. La protéine identifiée a été appelée XLF pour XRCC4-Like Factor *(Ahnesorg, 2006)* correspond au gène de Cernunnos. La structure 3D de Cernunnos a mis en évidence un repliement proche de celui de XRCC4 avec une structure dimérique, un domaine globulaire en N-ter de structure superposable au domaine équivalent de XRCC4, un coiled-coil central semblable à celui de XRCC4 mais plus court. Au niveau du milieu du coiled-coil de XRCC4,

Cernunnos présente 3 hélices qui rebroussent chemin et rejoignent le N-ter de la protéine *(Andres, 2007) (Li, 2008)*.

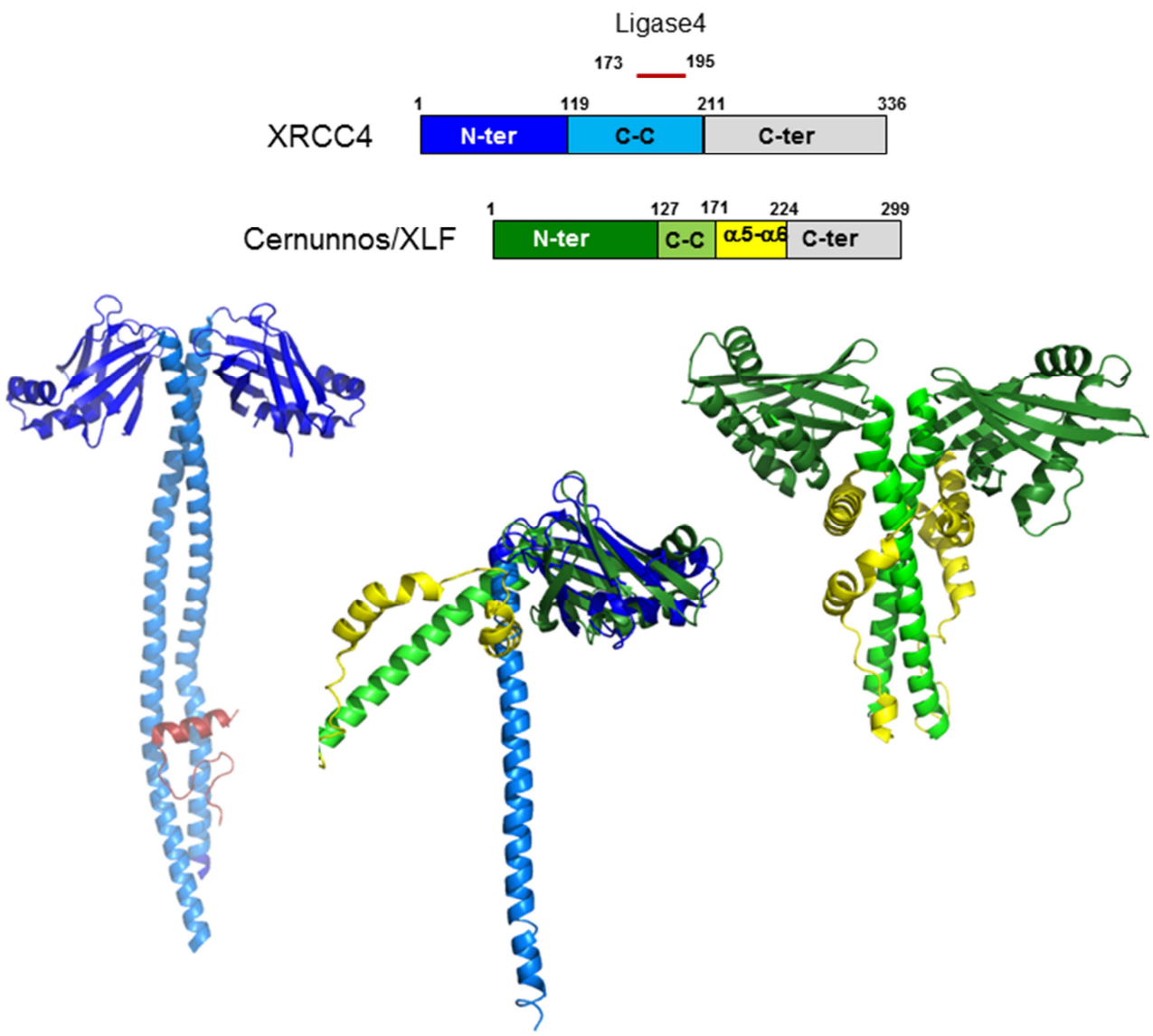


Figure 13 : XRCC4 et Cernunnos. *(**Haut**) Représentation des domaines des protéines XRCC4 et Cernunnos/XLF. (**Bas de gauche à droite**) Structure de XRCC4 avec la région charnière de la Ligase4 (Sibanda, 2001), superposition d'un monomère XRCC4 et Cernunnos par alignement des domaines globulaires N-terminaux, structure de Cernunnos (Andres, 2007) (Li, 2008). Le schéma des séquences montre que le coiled-coil de Cernunnos change de direction avant la région du coiled-coil qui chez XRCC4 sert de site d'interaction avec la Ligase4. La superposition de XRCC4 et Cernunnos montre que les orientations des coiled-coil de XRCC4 et de Cernunnos par rapport au domaine globulaire N-terminale sont différentes (angle plus grand chez Cernunnos). Cette différence est due à l'insertion de l'hélice α6 entre la tête et le coiled-coil chez Cernunnos.*

Différentes études *in vitro* ont montré que Cernunnos stimule l'activité de ligature de la Ligase4 en facilitant l'étape de ré-adénylation selon un mécanisme non encore élucidé

(Riballo, 2009, Tsai, 2007). Dans une étude fonctionnelle sur Cernunnos, le laboratoire de JP de Villartay a construit une vingtaine de mutants de cette protéine. 6 de ces mutants affectent fortement la réparation par la voie NHEJ et le réarrangement de gènes V(D)J sur un ADN extra-chromosomique *(Malivert, 2010)*. Deux autres mutants ont été inclus dans cette étude, les mutants R57G et C123R, qui correspondent aux mutations observées sur deux patients. Ces mutants présentent une activité de recombinaison V(D)J très faible. Notre laboratoire en collaboration avec celui de Necker a caractérisé les 8 mutants de Cernunnos qui présentent une activité V(D)J fortement réduite. Sur ces 8 mutants, 5 dont les 2 de patients présentent une expression très faible de la protéine Cernunnos et une perturbation de la localisation nucléaire de ce facteur. L'altération de la recombinaison V(D)J est liée selon toute vraisemblance à une déstabilisation de la structure de Cernunnos. 3 mutants, R64E, L65D et L115D, localisés dans un même site de la région N-terminale sont exprimés correctement et localisés au noyau.

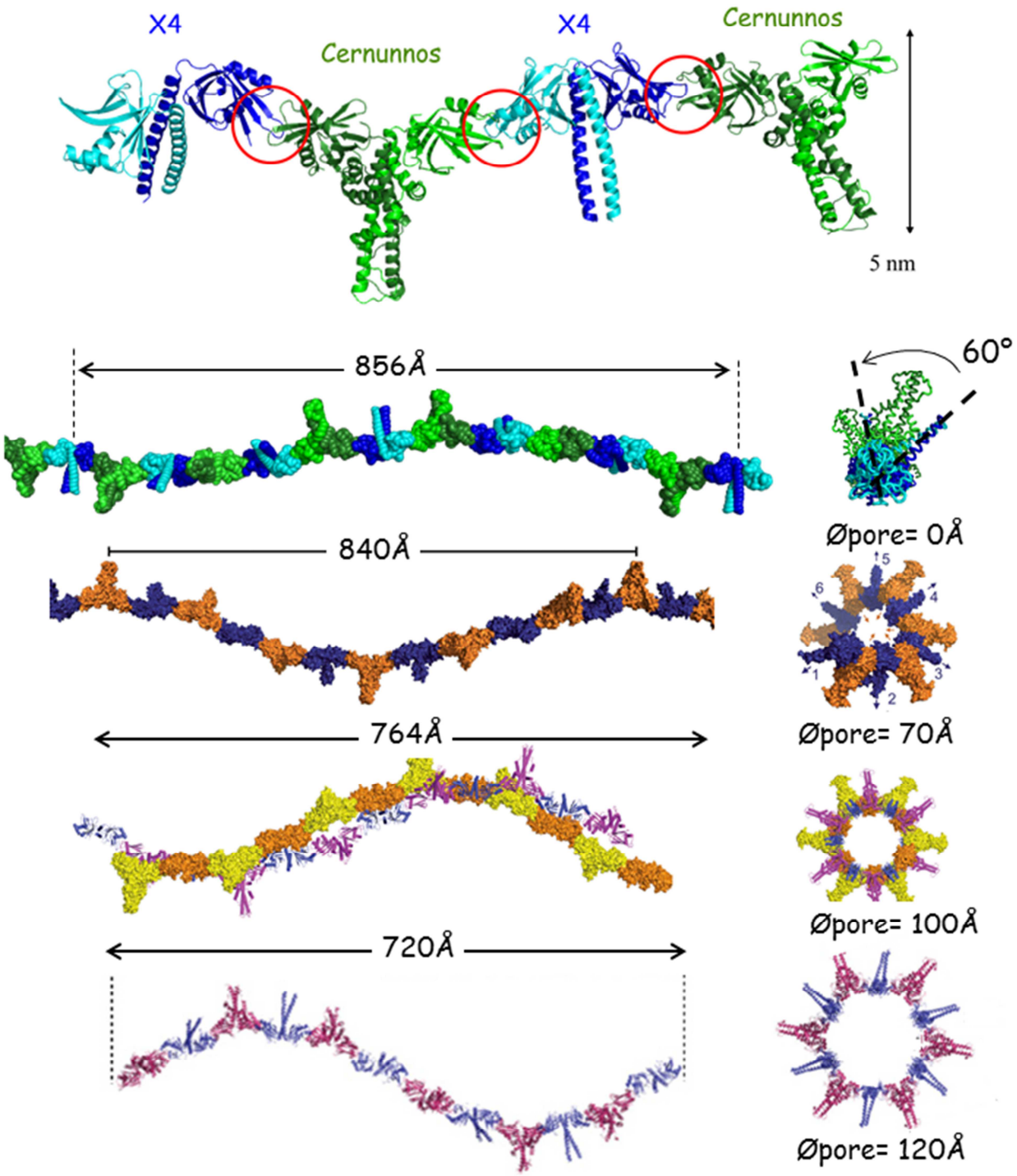
Nous avons pu mettre en évidence au laboratoire que ces mutations altèrent l'interaction entre Cernunnos et XRCC4 par des mesures de microcalorimétrie *(Malivert, 2010)*. Les protéines Cernunnos et XRCC4 délétées de leur région C-terminale, prédite comme désordonnée, ont été sur-exprimées et purifiées dans *E. coli*. Les protéines sauvages ainsi produites présentent une interaction endothermique (conduite par l'entropie) avec une stœchiométrie 1:1, une affinité moyenne avec constante de dissociation Kd de 3.9µM ± 0.2µM. De façon intéressante, les 3 mutants correctement repliés mais inactif dans le test V(D)J ne présentent plus d'interaction avec XRCC4 (mesuré par calorimétrie). Cette étude, nous a permis de caractériser le site d'interaction de Cernunnos avec XRCC4 et de montrer le rôle majeur *in cellulo* de cette interaction dans la recombinaison V(D)J et dans la voie NHEJ *(Malivert, 2010)*. De par la nature dimérique des deux protéines et la position des sites d'interactions caractérisées, nous avons proposé un modèle selon lequel Cernunnos était susceptible d'effectuer des pontages entre des complexes XRCC4-Ligase4 *(Malivert, 2010)*.

V Ropars, post-doctorante au laboratoire, a alors obtenu les premiers cristaux du complexe entre les formes tronqués de XRCC4 et Cernunnos. Les cristaux obtenus présentaient dans leur ensemble un très faible pouvoir de diffraction avec une forte anisotropie (diffraction dans la meilleure direction entre 7Å et 12Å). Afin d'obtenir de déterminer la structure de ce complexe, un ensemble important de cristaux (environ 400)

ont été testés systématiquement et ont permis d'obtenir un seul jeu de données à mieux que 6Å de résolution. Nous avons pu collecter un jeu à 5,5Å de résolution avec des statistiques satisfaisantes. La maille du cristal étudié présente une symétrie $P6_522$ et un paramètre de maille importante (856Å) sur la direction de l'axe 6_5. La qualité du faisceau de la ligne Proxima1 (SOLEIL, Gif) et l'expertise de P Legrand sur cette ligne ont été deux éléments majeurs dans la conduite de ce projet. La structure a été résolue par remplacement moléculaire à partir des structures cristallines de XRCC4 et Cernunnos seules. Les cartes de densité électronique obtenues après un cycle d'affinement en corps rigide (programme Buster) ont permis d'identifier sans ambiguïté les hélices principales de XRCC4 et Cernunnos et notamment les hélices α5 et α6 de Cernunnos qui remontent le long du coiled-coil dans cette protéine. Néanmoins, afin de lever toute ambiguïté sur l'identification de XRCC4 et Cernunnos, la protéine Cernunnos a été marquée avec des seleno-méthionines et des données de diffractions à 6.6Å ont été obtenues. Les cartes à basse résolution confirment la caractérisation des protéines proposées avec l'identification des seleno-methionines de Cernunnos *(Ropars, 2011).*

La structure du complexe obtenu a permis de confirmer un modèle proposé par R Guerois lors de notre précédente étude *(Malivert, 2010)*. Les deux complexes possèdent un RMSD (Root Mean Square Deviation) de 2,1Å sur les acides aminés à l'interface montrant la qualité des procédés de modélisation. La structure montre une surface d'interaction de taille réduite avec une contribution majeure des acides aminés hydrophobes de XRCC4 et Cernunnos. Cette dernière caractéristique est en accord avec les mesures de calorimétrie car les interactions conduites par l'entropie et d'affinité moyenne sont souvent la signature d'une interface médiée par des acides aminés hydrophobes et de faible surface d'interaction.

Une observation importante de cette structure cristalline a été l'arrangement des protéines XRCC4 et Cernunnos en long filament hélicoïdaux avec une symétrie 6_5 qui correspond au paramètre de maille de 856Å. Ce paramètre correspond au pas de l'hélice formée par 6 molécules de complexes XRCC4-Cernnunos.

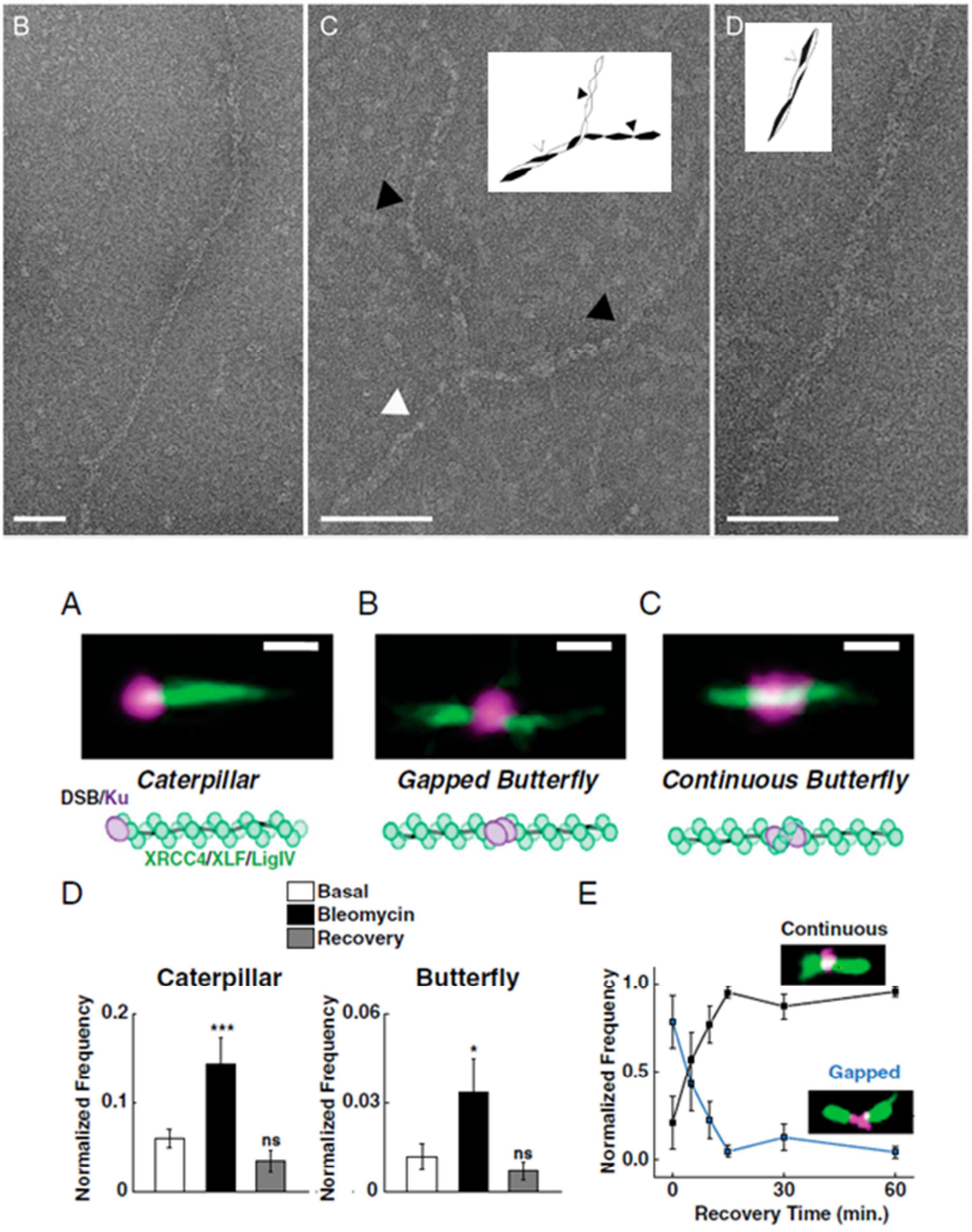


Figures 14 : XRCC4 et Cernunnos forment des filaments dans les structures cristallines*. (Haut) Structure cristalline déterminée au laboratoire (Ropars 2011 PNAS). Les protéines dimériques XRCC4 et Cernunnos interagissent par leur région globulaire N-terminale. La position du site d'interaction et l'orientation des protéines permet un arrangement en filaments des deux facteurs dans le cristal avec un pas de 856Å avec 6 complexes et une symétrie 6_5. (Bas) Comparaison des 4 filaments XRCC4-Cernunnos. Ces filaments présentent des courbures différentes. (Bas, droite) Une vue à 90° de ces filaments, le long de leur axe, montre qu'ils s'enroulent autour d'un tube dont le diamètre varie de 0Å à 120Å. Ces différents diamètres suggèrent que les filaments pourraient s'enrouler autour de structures allant d'un ADN nu à des assemblages plus grands comme des nucléosomes.*

Afin de confirmer cette observation par d'autres méthodes biophysiques, notre laboratoire a collaboré avec le laboratoire d'E Le Cam (IGR, Villejuif) pour analyser le complexe XRCC4-Cernunnos par microscopie électronique à transmission (TEM) par coloration négative. Il a ainsi été mis en évidence la formation de filaments sur les grilles de microscopie électronique. Les structures observées présentent une épaisseur compatible avec environ 2 filaments de XRCC4-Cernunnos (10nm pour une épaisseur de 5nm de chaque filament dans le cristal) **(Figure 15)**. Les images de TEM montrent une certaine variabilité dans la structure globale de ces filaments. Cette observation est en bon accord avec l'ensemble des structures cristallines des complexes XRCC4-Cernunnos reportées par trois autres laboratoires entre 2011 et 2012. Suite à nos observations, le laboratoire de J Tainer *(Hammel, 2011)* puis de T Blundel *(Wu, 2011)* et enfin de M Junop *(Andres, 2012)* ont publié les structures cristallines de XRCC4-Cernunnos avec des fragments plus ou moins tronqués de ces protéines. Les 4 structures sont en accord sur la localisation du site d'interaction entre les têtes globulaires N-ter de XRCC4 et Cernunnos. Les filaments observés dans ces 4 cristaux ont des géométries hélicoïdales différentes. Les différences majeures entre ces 4 structures portent principalement sur la taille du « pore » dessiné par les structures hélicoïdales. Le pore de l'hélice a un diamètre presque nul dans notre structure, de 70Å dans celle de Junop, de 100Å dans celle de Tainer, et de 120Å dans celle de Blundell. Ceci suggère une flexibilité de ces structures en filaments et la possibilité qu'elles entourent éventuellement soit de l'ADN (diamètre 20Å), de l'ADN double-brin avec de la protéine Ku ou un nucléosome (diamètre 110Å).

Afin d'évaluer la pertinence physiologique de ces filaments XRCC4-Cernunnos observés par cristallographie et microscopie électronique, différents laboratoires ont effectué des analyses additionnelles notamment en présence d'ADN. Le laboratoire de K Meek a pu mettre en évidence *in vitro* la capacité du complexe XRCC4-Cernunnos a effectué un pontage entre deux molécules d'ADN de 500bp et 1000bp *(Roy, 2012)*. Les protéines seules n'ont pas cette propriété de pontage et des mutants du site d'interaction soit du côté de XRCC4 ou du côté de Cernunnos n'ont plus des propriétés de pontage (bridging) de l'ADN. Dans une seconde étude, Tsai et al ont décrit l'assemblage coopératif d'une ou de deux

molécules de Ku avec de 1 à 5 molécules de XRCC4-Ligase4 et Cernunnos (mesures par gels retard) *(Tsai, 2013)*.



Figures 15 : Les filaments XRCC4-Cernunnos sont observés in vitro par microscopie électronique à transmission (TEM) et in cellulo par microscopie à super-résolution. *(**Haut**) Différents filaments sont observés par TEM avec les protéines XRCC4(1-157) et Cernunnos (1-224) utilisées en cristallographie. (B) On observe des filaments réguliers avec une épaisseur de 10nm qui correspondraient à deux filaments simples côte à côte. (C, D) On observe des filaments de 10nm qui semblent constitués de deux filaments de 5nm (en noir et blanc dans le schéma) qui peuvent se séparer ou s'enrouler l'un sur l'autre. La barre fait 50nm (d'après (Ropars, 2011))(**Bas**) Analyses des images de microscopie in cellulo de complexes Ku (magenta), XRCC4 ou Ligase4 ou Cernunnos (vert) en présence d'ADN. Les cellules sont fixées avec du paraformaldehyde et les protéines sont révélées par immunofluorescence. Trois évènements principaux sont observés : (i) chenille : Ku est à une extrémité et le filament Ligase4/XRCC4/Cernunnos couvre l'ADN (les auteurs suivent soit la Ligase4, soit XRCC4, soit Cernunnos). (ii) papillon avec une brèche : Ku fait un pontage entre deux ADN couverts du filament*

L4-X4-Cernunnos (iii) papillon continu : Ku est au niveau de la jonction entre les deux extrémités d'ADN mais l'ensemble est recouvert du filament L4-X4-Cernunnos (d'après (Reid, 2015))

Plus récemment, une étude du laboratoire de E Rothenberg a décrit pour la première fois par des approches de molécules uniques la présence *in cellulo* et *in vitro* de filaments de Ligase4-XRCC4-Cernunnos au niveau d'une cassure double-brin *(Reid, 2015)*. Par microscopie fluorescente à super-résolution, ils ont mis en évidence la présence de ces protéines et leur propriété de pontage de CDBs **(Figure 15)**. Par des mesures de FRET, ils ont pu caractériser la dynamique du complexe multiprotéique Ku-Ligase4-XRCC4-Cernunnos. Ces dernières travaux confortent l'existence des filaments XRCC4-Cernunnos *in cellulo* et permettent de proposer un nouveau modèle selon lequel un filament Ligase4-XRCC4-Cernunnos lié à une molécule de Ku viendrait interagir avec un second filament de même nature. Les deux filaments glisseraient dans un premier temps l'un sur l'autre avant de positionner de façon optimale la Ligase4 pour son action de ligature. Ce travail majeur amène à considérer un assemblage plus dynamique de Ku avec l'ADN et avec le complexe de ligation Ligase4-XRCC4-Cernunnos.

5.6 ADN Polymérases X

Les polymérases de mammifères de la famille X sont constituées des ADN polymérases β, λ, mu et TdT (pour Terminal Deoxy-ribonucleotidyl-Transferase). Ces polymérases sont de petite taille. Elles sont impliquées dans des processus de réparation de l'ADN comme la voie BER et la réparation de cassures double-brin *(Moon, 2007)*. La fonction principale de ces enzymes est de combler des brèches (gap) de un ou deux nucléotides au cours de la réparation de l'ADN **(Figure 16)**. Les polymérases λ et mu ont été montrées comme étant impliquées dans la voie NHEJ de façon générale. A l'inverse, la TdT est exprimée spécifiquement dans les cellules du système immunitaire (lymphocyte B et T). La TdT contribue à la génération de la diversité des récepteurs des cellules B et T à travers le processus de recombinaison V(D)J. Au cours de ce processus, la TdT ajoutent de façon aléatoire des nucléotides (processus appelé N-addition) à la fin des segments codants avant qu'ils soient reliés. La TdT a la propriété unique de pouvoir ajouter des nucléotides de manière indépendante d'une matrice (template-independent). L'activité de la TdT est dépendante de Ku et de DNA-PKcs.

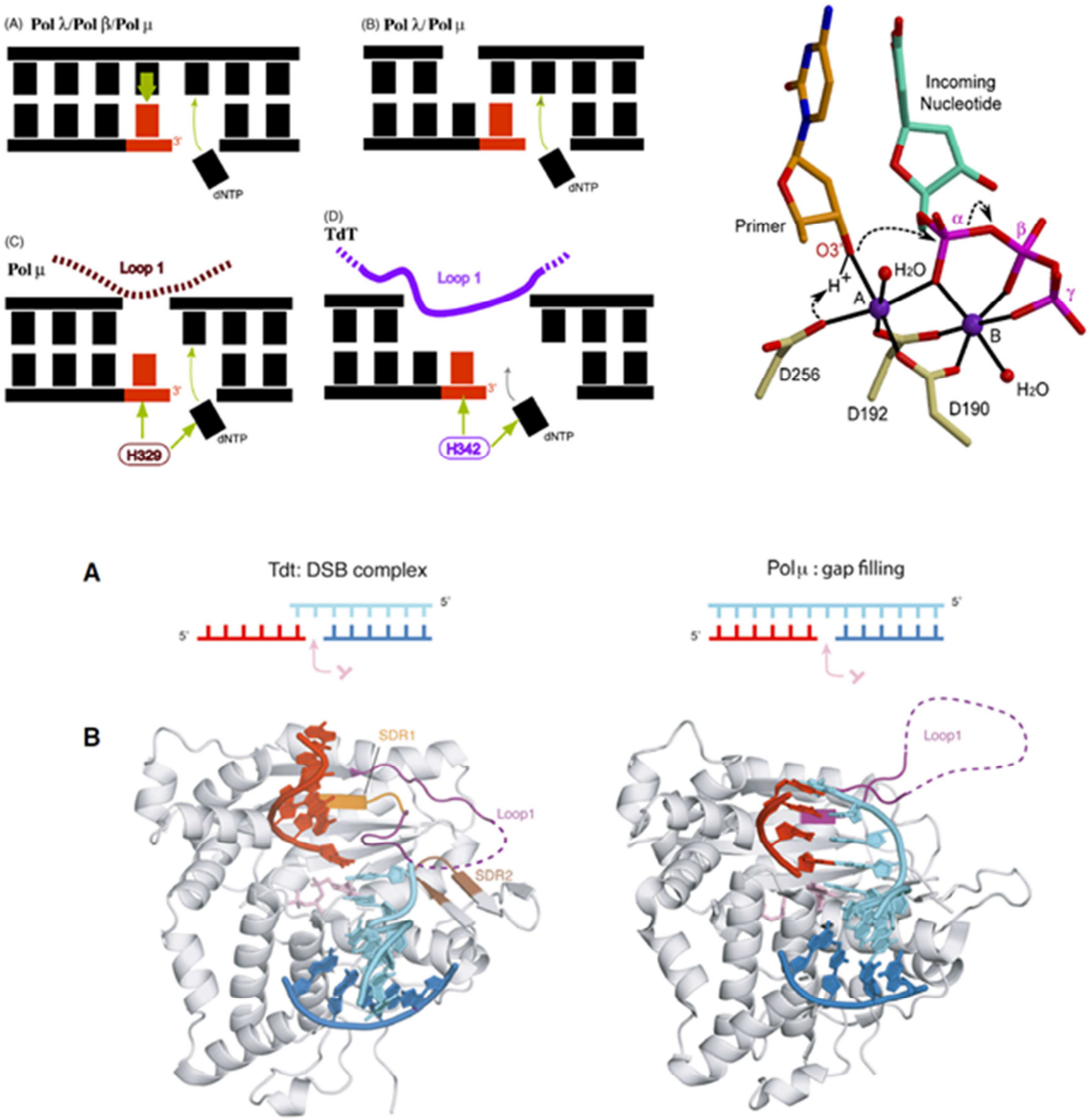


Figure 16 : Les Polymérases X. *(**Haut à gauche**) Représentation schématique des substrats préférentiels des différentes polymérases de la famille X. Le nucléotide terminal de l'amorce est en rouge et le nucléotide entrant en noir (**Haut droite**) Mécanisme catalytique de transfert de nucléotide, le nucléotide entrant est en cyan avec ses phosphates en magenta. Les ions Mg2+ sont en pourpre. (D'après (Moon, 2007)) (**Bas**) Structure de la TdT en complexe avec un ADN de type CDB (gauche) et comparaison avec la structure de la Pol mu dans un complexe de type brèche. L'ADN en aval en duplex (bleu clair et foncé) est hybridé avec l'ADN en aval (rouge) avec une seule paire de base.*

Des structures cristallographiques des 4 types de polymérases X ont été déterminées. La première fut celle de la Pol β qui présente comme les autres polymérases une organisation en trois domaines : paume, pouce et doigts *(Pelletier, 1994)*. Les structures des autres polymérases ont également été résolues et ont mis en évidence un repliement et une organisation des sous-domaines assez similaires (TdT *(Delarue, 2002)*, Pol λ *(Garcia-Diaz, 2005)*,

Pol mu *(Moon, 2007)*). Les polymérases λ, mu et TdT possèdent en N-terminal un domaine BRCT qui permet leur recrutement aux niveaux des extrémités des CDBs par des protéines de la voie NHEJ comme Ku, XRCC4 et la Ligase4 *(Mahajan, 2002)*.

Ces polymérases ont un mécanisme catalytique avec deux ions métalliques *(Moon, 2007)*. Le métal catalytique (site A, **Figure 16**) interagit avec le phosphate-α, trois résidus aspartates conservés et le nucléophile de la réaction catalytique, le 3'-OH du sucre deoxyribose terminal de l'amorce. Le second métal (site B) est coordonné par les atomes d'oxygène des 3 phosphates du nucléotide entrant et par deux aspartates conservés. Récemment, la structure cristallographique de la TdT complexé avec un brin d'ADN amorce, un nucléotide ddCTP entrant, un duplex en aval avec une extrémité 3' a été déterminée *(Gouge, 2015)*. Cette structure montre la capacité intrinsèque de TdT à effectuer un assemblage synaptique entre deux extrémités d'ADN. Cette étude montre également que le modèle présentée pour la TdT dans le cadre de la recombinaison V(D)J pourrait s'étendre à la Pol mu dans le cadre de la voie NHEJ

6. Maladies associées à des mutations entrainant un dysfonctionnement de la voie NHEJ

De par le rôle majeur des protéines de la voie NHEJ dans la réparation des cassures double-brin générées par des rayonnements ionisants et dans la recombinaison V(D)J, les patients déficients pour ces protéines présentent des cellules radio-sensibles et des défauts sévères du système immunitaire (appelés SCID pour Severe Combined ImmunoDeficiencies). A ce jour des mutations ont été identifiés sur principalement 4 protéines de la voie NHEJ : DNA-PKcs, Artemis, Ligase4 et Cernunnos *(Woodbine, 2014)*. Les patients présentant de telles mutations ont également d'autres défauts majeurs comme des microcéphalies et des problèmes de retard de croissance.

6.1 Patients présentant des mutations sur la Ligase 4

Une disruption complète du gène de la Ligase4 s'accompagne d'une létalité embryonnaire au stade E13.5 avec une apoptose très importante au niveau neuronal. En conséquence les mutations observées chez les patients sont hypomorphes (avec une activité

résiduelle de Ligase4) **(Figure 17A)**. Le premier patient identifié avec une mutation Ligase4 développa une leucémie à l'âge de 14ans et succomba à une radiothérapie lors d'un traitement *(O'Driscoll, 2001)*. Ce patient présentait une mutation homozygote R278H entrainant une réduction de 90% de l'activité catalytique de la Ligase4. Cette mutation est localisée dans le site de fixation de l'ATP de la ligase et il a été proposé qu'elle entraine une réduction de la réaction d'adénylation de la ligase. Les autres patients avec un syndrome Ligase4 (OMIM606593) présentent des défauts plus ou moins sévères du système immunitaire et généralement des altérations de type microcéphalie, retard de croissance et asymétrie faciale.

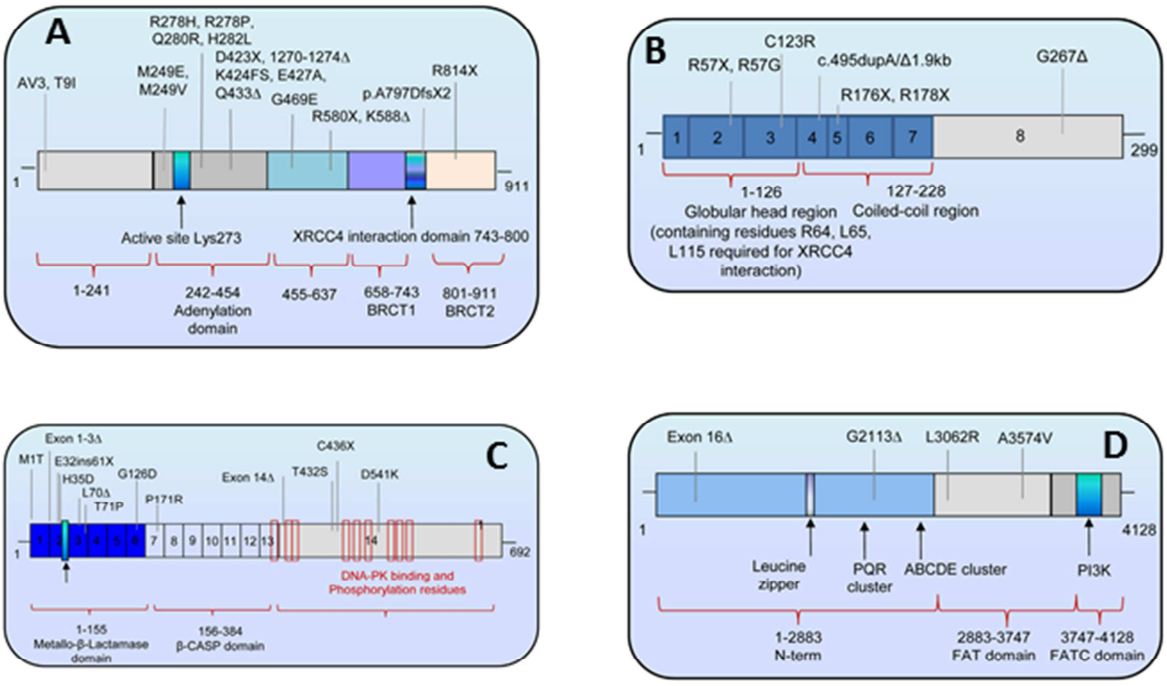


Figure 17 Localisation des mutations identifiées chez des patients sur les protéines du NHEJ *respectivement sur (A) la ligase4, (B) Cernunnos/XLF, (C) Artemis et (D) DNA-PKcs. Les domaines des protéines sont indiqués. Pour Artemis, les sites de phosphorylations par DNA-PKcs sont indiqués en rouge (d'après (Woodbine, 2014)).*

6.2 Patients présentant des mutations sur Artemis

Les mutations sur le gène Artemis (OMIM 602450) sont une des causes les plus fréquentes de syndrome d'immunodéficiences sévères liées à une radiosensibilité (rs-scid pour radiosensitive severe combined immuno-deficiency) **(Figure 17C)**. Artemis a un rôle

essentiel dans l'ouverture des hairpin formées lors de la V(D)J et aucun autre rôle majeur lors du développement. Les souris dont le gène Artemis a été ôté (souris KO pour Artemis) sont viables. Artemis a un rôle différent des autres protéines « cœur » de la voie NHEJ et il a été proposé qu'elle contribue à la réparation d'environ 15% des CDBs induites par des radiations *(Riballo, 2004)*. En conséquence, l'exposition de cellules de patients Artemis-nulles à des radiations X ou γ se traduit par un défaut de réparation de seulement une fraction de CDBs sur une longue durée (21 jours) contrairement aux cellules de patients Ligase4 ou Cernunnos pour lesquels le défaut de réparation concerne l'ensemble des CDBs formées. Un patient a été caractérisé qui présente une mutation sur une histidine du site actif et dont l'activité catalytique est fortement réduite *(de Villartay, 2009)*.

6.3 Patients présentant des mutations sur Cernunnos/XLF

Les mutations sur le gène Cernunnos (OMIM 611291) entrainent comme pour les mutations des patients Ligase4 une immunodéficience, des retards de croissances et une microcéphalie **(Figure 17B)**. Les cellules de ces patients sont également radiosensibles. Le laboratoire de JP de Villartay et P Revy a identifié le gène Cernunnos à partir de cellules de patients comme le gène humain capable de complémenter le phénotype de radiosensibilité de cellules de ces patients *(Buck, 2006)*. Contrairement à ce qui est observé avec XRCC4 et la Ligase4, les souris Cernunnos-/- ne présentent pas de létalité embryonnaire même si une instabilité génétique est observée ainsi qu'un niveau élevé de translocation. L'impact de deux mutations, R57G et C123R, a été étudié *(Malivert, 2010)*. Ces deux mutations entrainent une expression réduite des protéines, une mauvaise localisation dans la cellule et un défaut de NHEJ et V(D)J. Les fibroblastes de patients Cernunnos-/- présentent une radiosensibilité et un niveau de CDBs non réparés (suivi par la présence de foyers γH2AX) similaires à ceux observés chez les patients atteints d'un syndrome Ligase4.

6.4 Patients présentant des mutations sur DNA-PKcs

Le premier patient radiosensible et immunodéficient avec une mutation sur DNA-PKcs a été identifié en 2009 *(van der Burg, 2009)* **(Figure 17D)**. Il présente des caractéristiques proches d'un patient Artemis (immunodéficience sévère sans problème de développement).

De façon intéressante, la mutation observée n'a pas d'impact sur l'activité kinase de PKcs mais sur la fonction d'Artemis. La délétion du gène PKcs dans la souris entraine une immundéficience mais pas d'autres phénotypes majeurs. La délétion de PKcs dans des lignées cellulaires humaines a en revanche un impact très fort sur la capacité de prolifération des cellules. Un écart semblable a été observé entre les cellules humaines et les cellules de rongeurs sur la délétion de Ku qui est létale chez l'homme et non chez les rongeurs.

6.5 Pourquoi une déficience du NHEJ entraine-t-elle des microcéphalies ?

L'immunodéficience observée chez les patients ayant des mutations sur les facteurs de la voie NHEJ est en bon accord avec l'utilisation de cette voie dans le cadre de la recombinaison V(D)J. L'étude approfondie des patients ayant des mutations dans les facteurs de la voie NHEJ a mis en évidence des défauts majeurs d'apoptose au stade embryonnaire. Les cellules souches neuronales embryonnaires sont situées dans les zones ventriculaires et sub-ventriculaires (VZ et SVZ). Les études chez la souris montrent que ces cellules se divisent extrêmement rapidement du stade E8 à E16.5, puis la division cellulaire cesse. Ces cellules sont très sensibles à l'apoptose. On peut déclencher l'apoptose des cellules dans la région VZ-SVZ avec des doses d'irradiations de 0,5Gy alors que l'apoptose n'est pas déclenché dans un cerveau adulte à des doses 10Gy. Les embryons Ligase4-/- meurent entre E13.5 et E16.5 d'une apoptose neuronale massive, qui suggère aux auteurs de ces études que des CDBs se forment probablement spontanément suite à la division cellulaire rapide à cette étape du développement embryonnaire *(Frank, 1998)*.

7. Etat de l'art de l'expression en cellules d'insecte

Un des objectifs majeur de ma thèse a porté sur la reconstitution de complexes multi-protéiques de la voie de réparation NHEJ chez l'homme. Nous avons choisi de nous orienter vers une expression en cellules d'insecte de ces protéines et complexes de la voie NHEJ et d'utiliser la méthode MultiBac *(Bieniossek, 2012)*. Dans ce paragraphe, nous avons souhaité faire un point sur l'apport des méthodes d'expression en cellules d'insecte dans le contexte des études structurales et fonctionnelles.

De nombreux processus cellulaires mettent en jeu des complexes multi-protéiques aussi appelées machineries qui sont constituées de plusieurs sous-unités protéiques et d'autres biomolécules (acides nucléiques, sucres, lipides et petites molécules). A ce jour, nous ne disposons d'informations structurales que pour un petit sous-ensemble de ces machineries. Une des raisons de cette lacune provient de la difficulté à obtenir des échantillons de complexes multi-protéiques homogènes et en quantité suffisante pour des études structurales et fonctionnelles *(Bieniossek, 2012)*. Beaucoup de ces complexes ne peuvent être directement purifiés à partir de leur forme endogène car ils sont présents en trop faible quantité dans la cellule. La production recombinante de ces complexes a pour objectif de résoudre cette limite. Récemment, un système d'expression a été proposé pour co-exprimer dans *E. coli* plusieurs protéines en utilisant un transcrit mRNA polycistronique et deux ou plusieurs plasmides dans la même cellule *(Busso, 2011)*. Cependant de nombreuses protéines ne peuvent être exprimées efficacement dans *E. coli*. Elles peuvent avoir des domaines de taille trop importante, nécessitant des chaperons spécifiques pour se replier correctement ou des modifications (phosphorylation, acétylation) que *E. coli* ne peut apporter. Des systèmes d'expression en cellules eucaryotes, comme le système MultiBac et d'autres, ont pour objectif de débloquer le verrou que représente l'expression de certains de ces complexes multi-protéiques.

7.1 Système MultiBac

Nous décrirons dans Matériel et Méthode ce système. Brièvement, ce système consiste en (i) un ensemble de petits plasmides synthétiques, (ii) un génome de baculovirus dérivé du virus nucléaire polyhedral de *Autographa californica* (AcNPV) qui est utilisé pour infecter des cellules d'insecte de la chenille *Spodoptera frugiperda* et (iii) un ensemble de protocoles détaillant chaque étape depuis l'insertion des gènes dans les plasmides jusqu'à la production des complexes dans les cultures de cellules d'insecte fonctionnelles *(Barford, 2013, Bieniossek, 2012)*. L'insertion successive d'un nombre important de gènes dans un même plasmide est une étape délicate. Le système MultiBac permet de manipuler des plasmides plus petits et de les recombiner progressivement à partir de deux types de vecteurs appelés « accepteurs» et « donneurs» (**Figure 18**). L'insertion des gènes du complexe sur le génome viral est réalisée dans une souche bactérienne spécifique qui contient le génome viral

comme un chromosome artificiel ainsi qu'un plasmide codant pour le complexe enzymatique du transposon Tn7. Ce système enzymatique insère les cassettes d'expression présentes sur le plasmide de fusion « accepteur-donneur» sur le site d'attachement Tn7 introduit par ingénierie sur le génome viral. Le succès de l'étape de transposition s'accompagne de l'interruption d'un gène codant pour LacZα et permet un crible blanc/bleu des colonies.

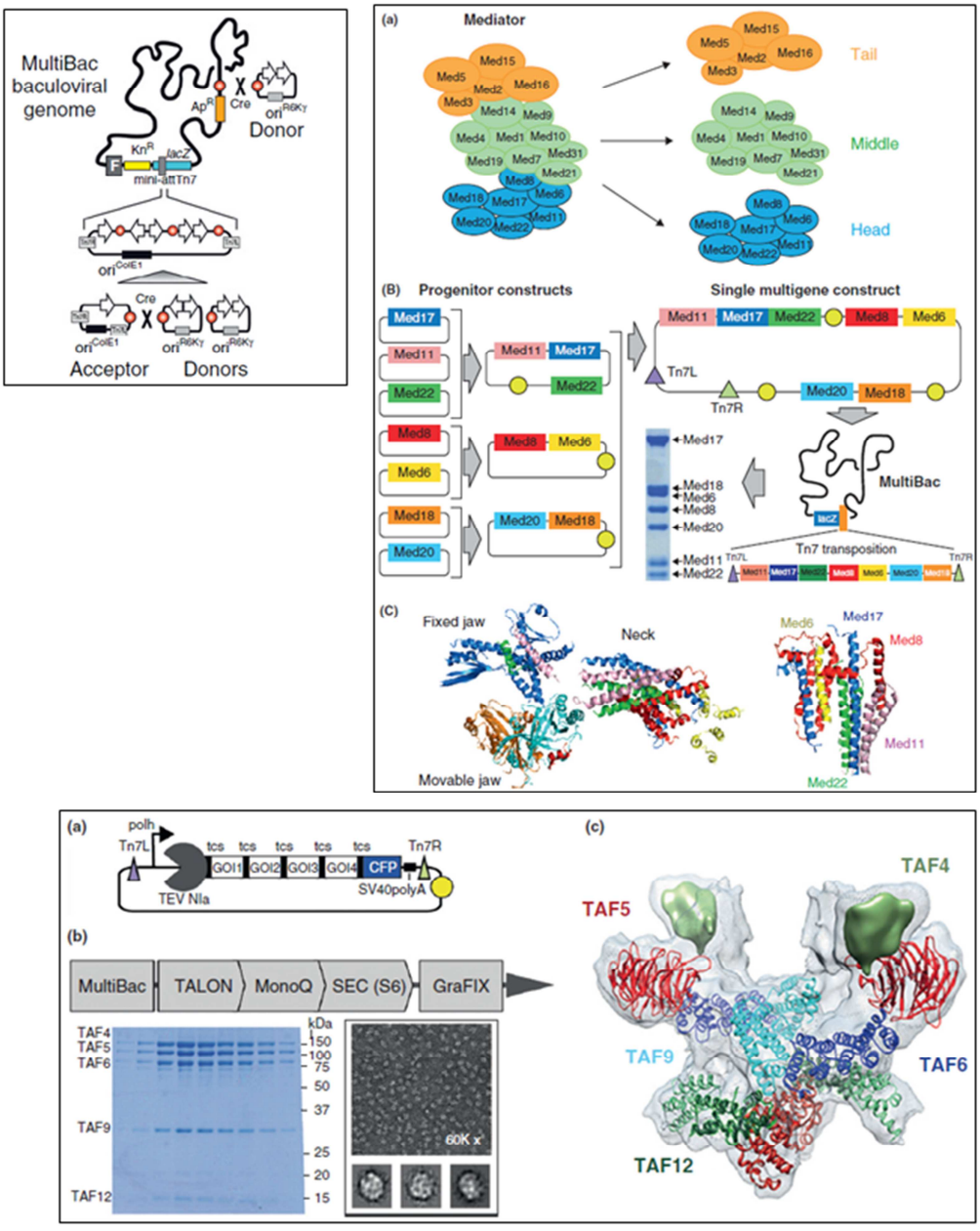


Figure 18 : L'approche MultiBac a permis de produire différents complexes multi-protéiques pour des études structurales. (**Haut Gauche**) *Approche MultiBac développée par I Berger. Les gènes d'intérêt sont insérés dans des vecteurs accepteurs et donneurs. La fusion de ces vecteurs avec une recombinase Cre permet de former un plasmide multi-gènes qui est inséré dans le génome viral à l'aide d'un site de transposition Tn7 pour former un Bacmide. Des gènes hétérologues additionnels*

*peuvent être insérés sur le site LoxP. Une protéine YFP (non représentée) est également présente sur le Bacmide comme rapporteur pour suivre l'expression des protéines d'intérêt au cours de la culture (d'après (Bieniossek, 2012)) (**Gauche**) Complexe du Médiateur de S. cerevisae (a) Organisation en 3 modules du Médiateur (b) les gènes codant pour les 7 sous-unités sont progressivement assemblés sur un plasmide unique, produits en cellules d'insecte et le complexe multiprotéique est purifié (gel SDS). (c) Structure cristallographique de la tête du Médiateur Med17 (bleu), Med11 (pourpre), Med22 (vert), Med6 (jaune), Med8 (rouge), Med18 (cyan), et Med 20 (orange). Le domaine "cou" est fait d'un faisceau d'hélices impliquant 5 sous-unités (d'après (Imasaki, 2011)) (**Bas gauche**) Complexe cœur du facteur général de transcription humain TFIID (a) Stratégie polyprotéine et application à la production du TFIID cœur. Cette stratégie a pour but de produire les différentes sous-unités avec une stœchiométrie correcte. Les gènes d'intérêt (GOI), la protéase TEV, qui cible le site 'tcs' et la protéine fluorescente CFP sont présents sur un seul ORF sous le contrôle du promoteur du baculovirus polh. Expression et purification du TFIID-cœur humain. Les images de microscopie électronique par coloration négative sont utilisés pour optimiser la purification jusqu'à l'obtention d'images homogènes (c) Structure du complexe cœur TFIID à 11.6Å par cryo-microscopie électronique et amarrage des structures cristallographiques connues des sous-unités dans les enveloppes de Cryo-EM (d'après (Bieniossek, 2013))*

Le génome du baculovirus MultiBac a été modifié par ingénierie pour la production en retirant le gène de la protéase virale et les activités apoptotiques pour réduire la protéolyse des gènes hétérologues et retarder la lyse des cellules infectées. Un site d'entrée LoxP a été introduit en plus du site Tn7 pour ajouter des fonctionnalités supplémentaires. Par exemple, un gène codant pour la λ-phosphatase peut-être inséré dans ce site pour dé-phosphoryler les protéines du complexe si nécessaire.

Le génome viral est utilisé pour transfecter des cellules d'insecte et générer des virus appelés V0 qui eux-mêmes sont utilisés pour transfecter des cellules d'insecte pour générer des virus appelés V1 qui serviront de stock pour infecter et cultiver des cellules lors des cultures. La production de protéines est réalisée à partir de cultures de cellules d'insecte en suspension et infectées par du virus V1. La production des protéines dans les cellules d'insecte peut être suivie grâce à la YFP introduit dans le génome viral sous le même promoteur que les protéines d'intérêts. On suit ainsi la mise en route de l'expression directement sur les cellules non lysées. Le système MultiBac a été développé par le laboratoire d'Imre Berger et a été appliqué à de nombreux systèmes à travers la plateforme Instruct de ce laboratoire.

7.2 Structure cristalline de la tête du Médiateur

Le Médiateur est un large complexe multi-protéique qui régule la transcription des gènes par l'ARN polymérase II (Pol II). C'est un assemblage conservé chez les eucaryotes. Il contient 21 sous-unités chez *S. cerevisiae,* qui combinées forme une machinerie de plus de 1MDa. Le Médiateur effectue une interface entre les facteurs d'activations de la transcription qui se lient à l'ADN et la Pol II. Des mutations chez l'homme sur plusieurs sous-unités du Médiateur ont été associées à des désordres neurologiques et des cancers. Les études génétiques et biochimiques ont montré que le Médiateur était formé de 3 modules (Tête, Milieu et Queue), chacune constituée de 7-9 sous-unités (**Figure 18**)

La structure du module Tête du Médiateur a été déterminée en 2011 à 4,4Å de résolution *(Imasaki, 2011)* par le laboratoire d'Y Takagi. Ce module est constitué de 7 sous-unités (Med17, Med6, Med8, Med11, Med22, Med18, Med20) avec un poids moléculaire de 223kDa. Les nombreux essais effectués par co-transfections de cellules d'insecte avec des virus contenant chacun une sous-unité n'ont pas permis d'obtenir d'échantillons satisfaisants. L'étape de production a été débloquée par construction d'un plasmide contenant les 7 sous-unités et insertion de ce plasmide dans le génome viral (**Figure 18**). Une seconde étape importante a été l'ingénierie des sous-unités Med17 et Med18 afin d'ôter les régions flexibles de ces sous-unités pour obtenir des monocristaux bien ordonnés. Les échantillons ont également pu être marqués à la séléno-méthionine. La structure permet d'identifier les zones d'interactions des différentes sous-unités et de rationaliser les nombreuses données génétiques accumulées sur ce Médiateur. La région Cou (Neck) de ce module Tête du Médiateur est un élément central dans la stabilité et l'intégrité de ce module. Il forme un faisceau d'hélices impliquant 5 des 7 sous-unités de la Tête en une seule structure compacte *(Imasaki, 2011)*.

8.3 Structure du facteur de transcription TFIID

L'initiation de la transcription par l'ARN polymerase II (Pol II) nécessite un assemblage progressif bien régulé d'un complexe de pré-initiation (PIC) qui est constitué de Pol II et des facteurs génériques de transcription (TFIIA, B, D, E, F, H). Le complexe TFIID joue un rôle central dans l'assemblage de cette machine moléculaire que constitue le PIC. Chez l'homme, le TFIID comprend 14 sous-unités protéiques : la protéine TBP (Tata-box binding protein) et 13 facteurs associés à TBP (TAFs) qui ont des tailles allant de 250 kDa à moins de 20kDa. Les

TAFs et TBP s'assemblent dans un complexe multi-protéique avec environ 20 sous-unités et un poids total de ~1.6MDa et plusieurs copies de certaines des TAFs. Des structures globales du TFIID humain ou de levure ont été proposées à partir d'images de microscopie électronique et ont mis en évidence une structure tri-lobée. Cependant l'hétérogénéité des échantillons obtenus à partir des protéines endogènes et les faibles quantités de matériel ont limité les analyses structurales (résolution de ~30-40Å pour le TFIID humain)

La structure d'un sous-complexe multi-protéique du TFIID, appelé TFIID-cœur, a été étudiée récemment par une approche combinant cryo-EM (résolution de 11.6Å), données de cristallographie et modélisation par homologie *(Bieniossek, 2013)*. La structure TFIID-cœur a été identifiée pour la première fois dans le noyau de la drosophile et elle contient 2 copies des sous-unités TAF4, 5, 6, 9 et 12. Initialement, ce complexe cœur TFIID a été produit avec une stratégie similaire à celle adoptée pour la Tête du Médiateur avec un baculovirus contenant les cassettes pour les différentes sous-unités. Il s'est alors avéré que le niveau d'expression des sous-unités était très variable, réduisant le niveau de production final au niveau de l'élément le moins bien exprimé. Ce verrou a été levé sur ce projet en adoptant une stratégie différente de type poly-protéine suivi d'un clivage par une protéase (**Figure 18**). Certains virus comme les Coronavirus expriment de long ORFs (cadre ouvert de lecture) qui sont traduits en poly-protéines. Ces poly-protéines sont transformées en protéines individuelles par des protéases très spécifiques qui coupent entre les protéines individuelles. Cette protéase est typiquement partie intégrante de la poly-protéine et se libère toute seule. Cette stratégie a été implémentée sur la méthode MultiBac. Un long ORF composé des gènes TAFs a été construit et le gène codant pour la protéase Nia du virus Tobacco etch (TEV) a été ajouté en 5' ainsi que le site de clivage par la TEV entre les gènes TAFs. Afin de suivre la production de la poly-protéine durant l'expression un gène codant pour une protéine fluorescente (CFP) a été ajouté en 3'. Cette approche a permis d'obtenir des échantillons en quantité suffisante et avec une bonne stœchiométrie et homogénéité des sous-unités TAFs. La structure par cryo-EM révèle une organisation avec une symétrie 2 de l'assemblage **(Figure 18).** Les protrusions observées dans les enveloppes de microscopie sont en bon accord avec les structures cristallines des protéines individuelles connues. Ces données ont permis pour la première fois de proposer un modèle du TFIID-cœur *(Bieniossek, 2013)*. Deux

sous-unités additionnelles, TAF8 et TAF10 ont pu être ensuite ajoutées et analysées par cryo-EM.

8 Objectifs de la thèse

A mon arrivée au laboratoire en Master2, mon projet a porté sur la caractérisation du complexe XRCC4-Cernunnos formé avec les protéines entières pour compléter les informations structurales et fonctionnelles obtenues sur leur forme tronquée. Ce travail avait pour objectif de montrer que les formes physiologiques de ces protéines étaient capables d'interagir directement et de former également des filaments. J'ai exprimé dans *E. coli* et purifié les formes entières de XRCC4 et Cernunnos ($X4^{FL}$ et $Cernu^{FL}$ pour FL pour Full Length) (**Figure 19**). J'ai mis en évidence par chromatographie d'exclusion de taille l'interaction directe en ces deux protéines.

Après avoir essayé sans succès de cristalliser le complexe $X4^{FL}$-$Cernu^{FL}$, ce complexe a été analysé par microscopie électronique par transmission. J'ai participé à ces analyses réalisées par S Baconnais à l'Institut Gustave Roussy, Villejuif. Nous avons pu mettre en évidence la formation de filaments torsadés du complexe $XRCC4^{FL}$-$Cernu^{FL}$ avec une épaisseur supérieure (20-25nm) à celle observée avec les formes tronquées des protéines (**Figure 19**) *(Ropars, 2011)*. $XRCC4^{FL}$ seul et $Cernunnos^{FL}$ seuls forment également des filaments mais de structures très différentes. Ces travaux ont permis de répondre à une question des examinateurs du manuscrit, soumis par notre laboratoire à PNAS sur la description de la structure cristallographique et les analyses par TEM du complexe XRCC4-Cernunnos. Lors de la révision à PNAS, les examinateurs ont en effet demandé de préciser si les formes entières de XRCC4 et Cernunnos formaient des filaments pour améliorer la relevance physiologique de nos observations (J Amram dans *(Ropars, 2011)*)

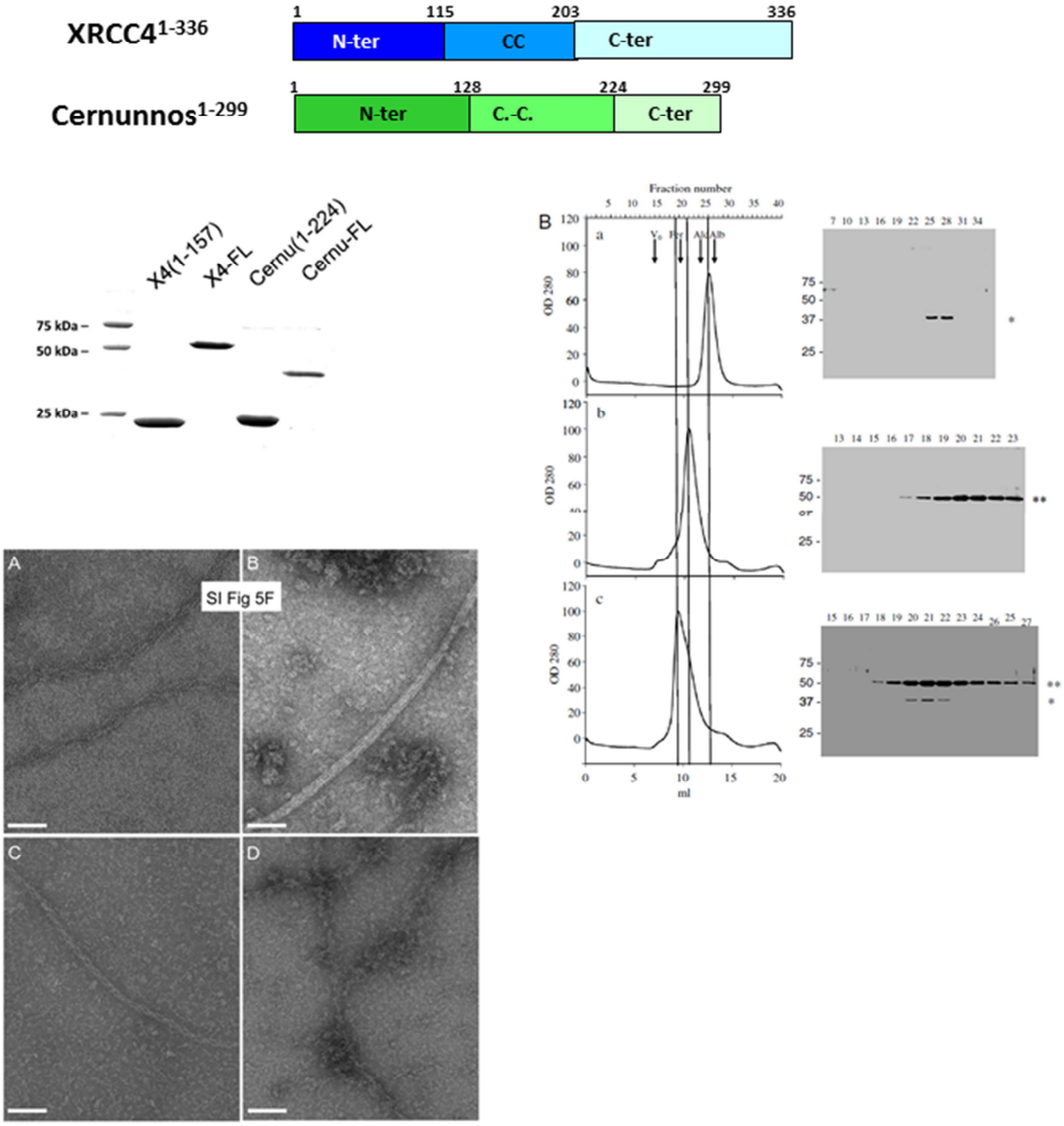


Figure 19 : Les protéines XRCC4 et Cernunnos entières interagissent directement et forment des filaments. *(**Haut**) Les protéines XRCC4 et Cernunnos entières avec leur région prédite comme non-structurée de respectivement 135 et 75 acides aminés sont exprimées dans E. coli et purifiées à homogénéité. (**Droite**) Caractérisation de l'interaction XRCC4-Cernunnos par chromatographie d'exclusion de taille (Superdex 200 10/300GL Ge-Healthcare, vol total 24ml). Chromatogrammes et gels SDS correspondant à (haut) Cernunnos 6 nmoles, (milieu) XRCC4 15 nmoles (bas) Cernunnos-XRCC4 avec un ratio 1 :5 (3nmoles : 15 nmoles). Calibration : Vo bleu dextran (>2Mda); Ferritin (440 kD); Aldolase (158 kD); Albumine (67 kD). (**Bas gauche**) Images de microscopie électronique à transmission (TEM) avec les protéines XRCC4 et Cernunnos entières (a) filament long typique de XRCC4 entier seul sous forme de tresses relâchées (b) Cernunnos entier ne fait pas de filament ou de très rares réguliers (épais 20-25nm). (c et d) filaments observés avec XRCC4 et Cernunnos entiers. Ces filaments co-existent avec des agrégats de un, deux ou plusieurs protéines (voir dans d). Les filaments XRCC4-Cernunnos sont présents en grand nombre et sous forme de tresses très serrées, différentes de celles observées avec XRCC4 seul (d'après (Ropars, 2011)). La barre fait 100nm.*

Dans la continuité de ce travail, l'objectif de ma thèse était de reconstituer la voie NHEJ humaine et les complexes multi-protéiques principaux de cette voie en utilisant l'expression en cellules d'insecte et d'initier les études structurales et fonctionnelles de ces complexes. Au cours de ce travail, j'ai acquis dans le laboratoire d'I Berger à l'EMBL Grenoble une expertise en expression des protéines recombinantes en cellules d'insecte. J'ai pu implémenter cette technique au laboratoire avec P Drevet et nous avons alors progressivement crée les plasmides, bacmides et virus afin de produire principalement les 5 protéines et complexes de la voie NHEJ suivants : (1) l'hétérodimère Ku70/Ku80, (2) la nucléase Artemis, (3) et (4) les cofacteurs de Ligase4, XRCC4 et Cernunnos, ainsi que (5) le complexe Ligase4-XRCC4. Ces protéines ont été produites en cellules d'insecte avec des volumes allant de 400mL à 3,2L. J'ai mis en place des protocoles robustes de purification de ces 5 protéines et complexes afin d'obtenir des échantillons à l'échelle de plusieurs milligrammes et de pureté compatible avec des études structurales ou de mesures biophysiques d'interaction des protéines.

Le manuscrit de thèse comprend après cette introduction une première partie Matériel et Méthodes où je décris les protocoles utilisés au cours de ma thèse ainsi que les informations principales sur les protéines et complexes étudiés. La partie suivante décrit les résultats obtenus au cours de ma thèse en ce qui concerne la production des 5 protéines et complexes de la voie NHEJ étudiés et la caractérisation de ces facteurs du NHEJ principalement par chromatographie d'exclusion de taille, diffusion de lumière, microscopie électronique et les essais de cristallisation entrepris. Dans une dernière partie, discussion et perspectives, les principaux résultats de cette thèse sont rappelés et mis en perspectives des travaux qui ont suivi au laboratoire et des études à venir.

MATERIEL ET METHODES

1. Bioinformatique

Les protéines de la voie NHEJ étudiées dans cette thèse ont fait l'objet de d'études structurales et biochimiques qui ont permis de définir les domaines principaux de ces protéines (**Figure 20**). Nous avons choisi au laboratoire de travailler dans un premier temps sur les formes entières de ces protéines pour éventuellement construire dans un second temps des formes plus courtes, notamment s'il s'avérait utile pour la cristallogenèse d'ôter des régions peu structurées. Les caractéristiques biochimiques des protéines ont été calculées à l'aide du programme ProtParam (http://web.expasy.org/protparam/).

Ku70 est une protéine de 609 acides aminés d'un poids moléculaire de 69,8 kDa possédant un pI (point isoélectrique) de 6,23. Ku80 est une protéine de 732 acides aminés d'un poids moléculaire de 85,7 kDa et possédant un pI théorique de 5,55. La structure du complexe Ku70/Ku80 en complexe avec de l'ADN a été résolue par cristallographie pour les 609 résidus de Ku70 et pour les 565 premiers résidus de Ku80 (PDB : 1JEY) *(Walker, 2001)*. Ku70 est composé d'un domaine vWA en N-ter, d'un domaine de dimérisation avec Ku80 et d'un domaine SAP en C-ter. Ku80 est composé d'un domaine vWA en N-ter, d'un domaine de dimérisation avec Ku70, puis d'un domaine C-terminal prédit comme étant intrinsèquement désordonné.

XRCC4 est une protéine de 336 acides aminés d'un poids moléculaire de 38,3 kDa et possédant un pI de 4,91. La structure des 203 premiers acides aminés de sa séquence a été résolue en dimère par cristallographie (PDB : 1FU1) *(Junop, 2000)*. Les 133 résidus C-terminaux sont prédits comme étant désordonnés, et leur structure n'est pas connue. La structure de XRCC4 a aussi été résolue en complexe avec un peptide de la région charnière entre les 2 domaines BRCT de la Ligase4 (PDB : 1IK9) *(Sibanda, 2001)* ou avec les deux domaines BRCT (3II6) *(Wu, 2009)*. La structure de XRCC4 en complexe avec Cernunnos a également été résolue notamment au laboratoire (PDB : 3Q4F, 3RWR, 3SR2, 3W03) *(Ropars, 2011) (Hammel, 2011) (Andres, 2012) (Wu, 2011)*.

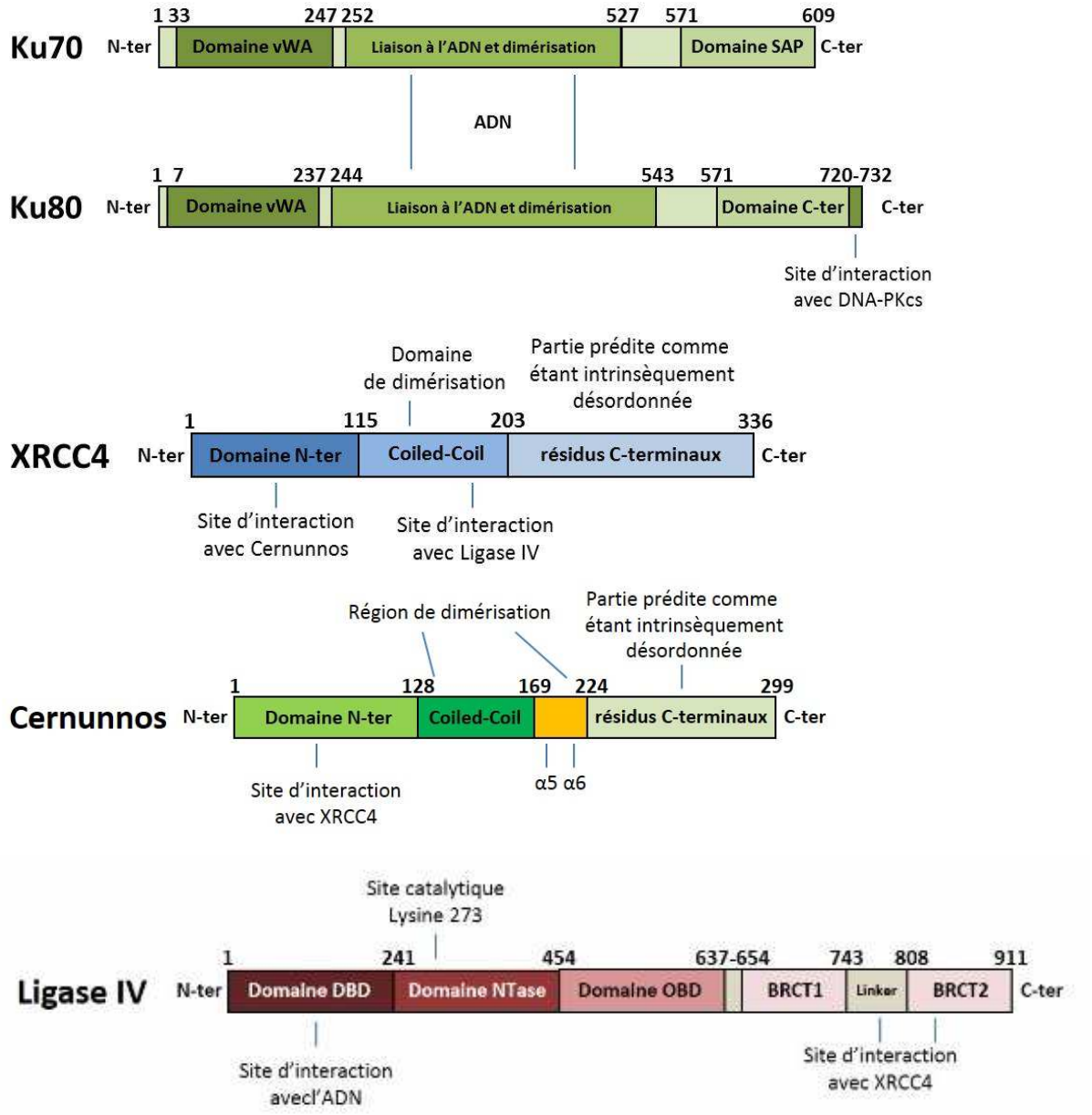


Figure 20 : Représentation des différentes informations de la littérature sur les domaines et les sites d'intéractions des protéines humaines du NHEJ.

Cernunnos-XLF est une protéine de 299 acides aminés, d'un poids moléculaire de 33,3 kDa et d'un pI théorique de 5,65. La structure de ses 224 premiers résidus a été résolue par cristallographie (2R9A) *(Andres, 2007)* et (2QM4) *(Li, 2008)*. La structure de ses 75 derniers acides aminés est prédite comme étant désordonnée et n'a pas été résolue. La structure de Cernunnos a aussi été résolue en complexe avec XRCC4.

La ligase4 est une protéine de 911 acides aminés, d'un poids de 103 kDa et d'un pI théorique de 8,17. Elle est composée d'un domaine DBD pour DNA Binding Domain, d'un

domaine d'adénylation ou NTase pour nucleotide-transférase, d'un domaine OBD pour OB-fold domaine, puis d'une région charnière flexible qui relie ces trois domaines à deux domaines BRCT. La structure cristallographique des trois premiers domaines a été résolue par cristallographie (PDB : 3W1G, 3VNN) *(Ochi, 2012)*. La structure des deux domaines BRCT a été résolue en complexe avec XRCC4 ainsi que celle entre le domaine catalytique de la Ligase4 et un peptide d'Artemis ((aa 485 à 495) (3W1G, 3W5O) *(Ochi, 2013)*.

2. Expression en cellules d'insecte des protéines de la voie NHEJ

2.1 Principes généraux

Comme détaillé précédemment, l'objectif central de ma thèse a porté sur la reconstitution des protéines centrales de la voie NHEJ humaine. La stratégie choisie a été de purifier ou co-purifier en cellules d'insecte ces protéines dont la taille varie de 299 à 4128 acides aminés. Dans un premier temps, nous avons choisi de produire l'ensemble des protéines sous leur forme entière pour pouvoir mener des études fonctionnelles (mesures d'interactions protéine-protéine, tests catalytique, analyses par microscopie électronique à transmission) avec des formes physiologiquement pertinentes. Vers la fin de ma thèse, des formes additionnelles ont été construites au laboratoire afin d'ôter des régions prédites comme désordonnées afin de faciliter l'étape de cristallisation.

Pour mener à bien ce projet, nous avons décidé de collaborer avec la plateforme MultiBac d'Imre Berger (EMBL, Grenoble) et d'utiliser sa méthode d'expression MultiBac de protéines en cellules d'insecte infectées par un Baculovirus. Cette méthode a été développée par l'équipe d'Imre Berger à l'EMBL afin de produire des complexe multi-protéiques et des assemblages macromoléculaires de grande taille *(Fitzgerald, 2006) (Bieniossek, 2013)*. Ces vecteurs permettent le clonage de séquences d'intérêt pour les sur-exprimer en cellules d'insecte (sf9, sf21, Hi5) sous le contrôle de promoteurs de baculovirus et ils permettent une combinatoire des gènes de façon à exprimer plusieurs protéines simultanément (**Figure 21**).

Les vecteurs Multibac possèdent deux sites de clonages multiples, MCS1 et MCS2, qui permettent l'insertion de séquences codantes sous la dépendance de deux promoteurs viraux différents : le promoteur de la protéine P10 et celui de la polyhédrine. De plus, Les

vecteurs proposés par cette méthode sont divisés en deux catégories, vecteurs accepteurs (pFL et pKL) et vecteurs donneurs (pUCDM et pSPL). Les différents vecteurs Multibac portent tous un premier gène de résistance à un antibiotique, la gentamycine. Ils possèdent également un second gène de résistance à un second antibiotique, différent pour chaque vecteur, à savoir ampiciline, kanamycine, chloramphenicol et spectinomycine pour pFL, pKL, pUCDM et pSPL respectivement.

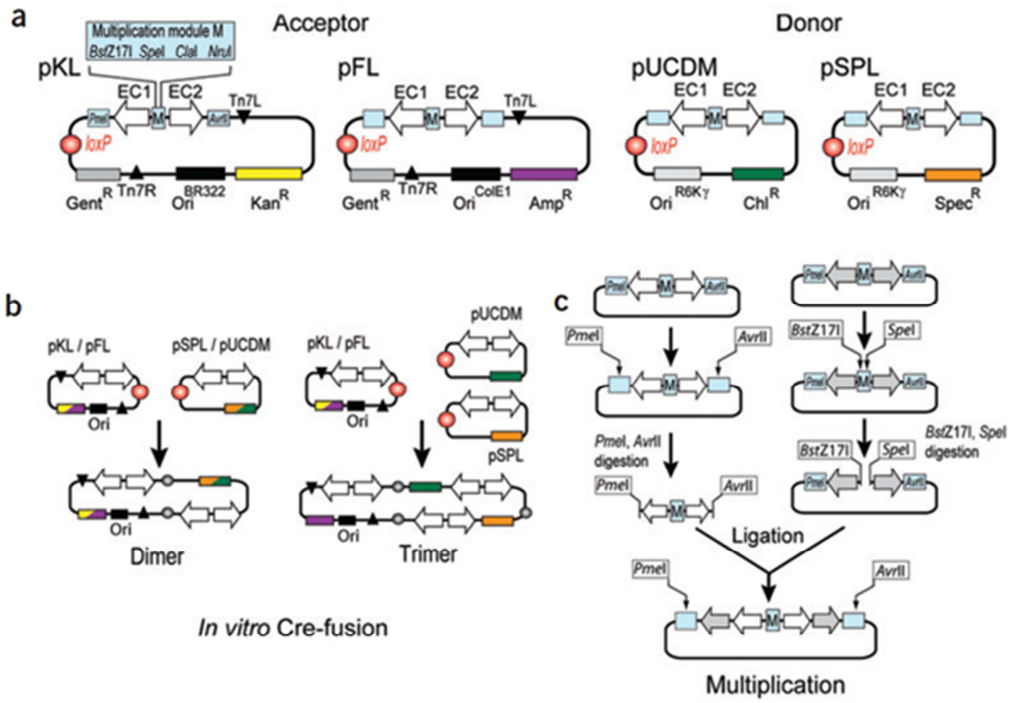


Figure 21 : Création de vecteurs de transfert multigènes.
(a) Carte des vecteurs (les flêches blanches représentent les cassettes d'expression. (b) Assemblage des vecteurs de transfert multigene par fusion in vitro de plasmides de type accepteur (pFL, pKL) et de plasmides de type donneur (pUCDM, pSPL) à l'aide de l'enzyme Cre recombinase. Les plasmides accepteurs et donneurs contiennent un site loxP (cercle rouge). Des plasmides dérivés des vecteurs accepteurs peuvent être fusionnés avec un ou deux plasmides donneurs. Les cassettes d'expression entre deux séquences de transposition Tn7 (triangles noirs) sont intégrées dans le génome baculovirus de type MultiBac. (c) Assemblage multigenes en utilisant le module de multiplication M présents sur tous les vecteurs de transfert. Les deux méthodes d'assemblages peuvent être utilisées conjointement. (D'après (Fitzgerald, 2006)*)*

Chaque vecteur possède une cassette de multiplication (Multiplication Module M) située entre les deux sites de multiclonage MCS1 et MCS2 et contenant deux sites de restriction Bstz 171 et SpeI (**Figure 21 et 22**). L'ouverture de ces deux sites génère des extrémités compatibles avec les sites PmeI et AvrII situés au-delà des sites MCS1 et MCS2. Ainsi, un fragment généré par une coupure PmeI/AvrII peut être inséré dans un vecteur

ouvert par Bstz 171/SpeI. Un nouveau module MM est ainsi inséré avec le fragment ce qui permet l'itération du processus.

Les vecteurs utilisés portent tous une séquence LoxP permettant leur recombinaison modulaire les uns avec les autres en utilisant la recombinaison CreLox. Cette possibilité permet de générer un plasmide contenant l'ensemble des gènes portés par les deux plasmides parentaux afin de produire des complexes multi-protéiques. La sélection des complexes se fait en utilisant un mélange adéquat d'antibiotiques par référence aux résistances portées par les plasmides parentaux.

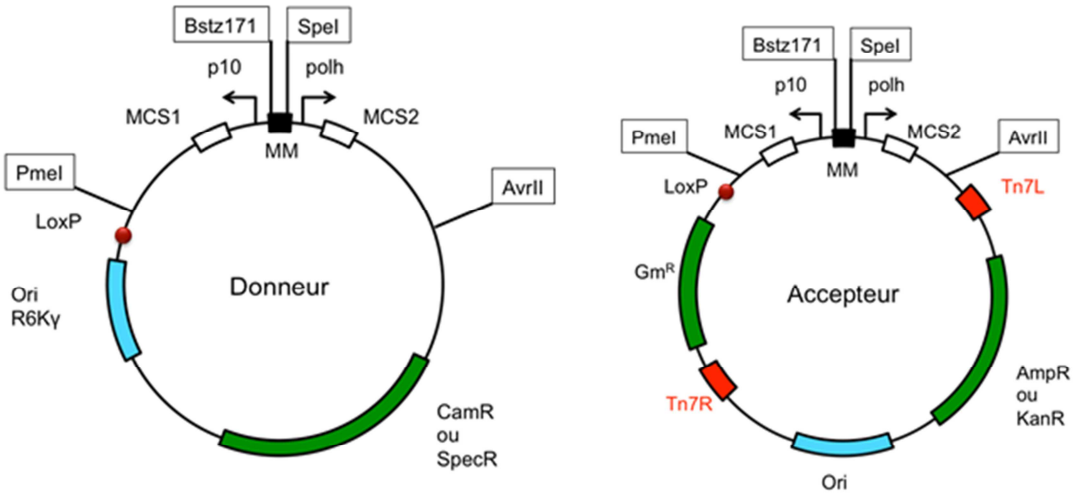


Figure 22 : Représentation schématique d'un vecteur donneur et d'un vecteur accepteur.
Les sites MCS1 et MCS2 contiennent un ensemble de sites uniques permettant le clonage sous la dépendance des promoteurs p10 et polh. Les vecteurs existent en version résistant à l'ampiciline (pFL), à la kanamycine (pKL), au chlormaphenicol (PUCDM) ou à la spectinomycine (PSPL).

Seuls les vecteurs accepteurs sont porteurs de sites Tn7L et Tn7R qui permettent la recombinaison de la portion du vecteur contenant les séquences codantes pour les protéines clonées dans le bacmide présent dans la souche bactérienne EMBacY. Un bacmide est un fragment d'ADN hybride contenant à la fois le génome du baculovirus et des séquences propres aux bactéries (origine de réplication, résistances à des antibiotiques, etc..) afin de permettre sa multiplication au sein de celles-ci. Le bacmide présent dans les cellules EMBacY a été mis au point par l'équipe d'Imre Berger et code pour la protéine fluorescente YFP sous contrôle d'un promoteur viral. L'expression constitutive de la YFP est utilisée comme rapporteur de l'activité de synthèse des protéines à partir des promoteurs viraux. On estime que l'expression des protéines d'intérêt suit une progression proche de

celle de la YFP avec classiquement une expression croissante entre les jours J+3 et J+4 après le jour d'arrêt de prolifération appelé DPA pour Day after Proliferation Arrest. Cette information est utilisée empiriquement pour déterminer le meilleur moment pour collecter les cellules d'insecte.

Une fois les vecteurs de clonage recombinés au sein du bacmide, ce dernier est purifié, et utilisé pour transfecter des cellules d'insecte qui produisent la première génération de virus (V0). Les virus V0 sont alors utilisés pour infecter de nouvelles cellules d'insecte afin d'amplifier le virus et augmenter son titre. On obtient ainsi une seconde génération de virus (V1). C'est cette dernière population de virus qui est stocké plusieurs mois à 4°C et qui est utilisée pour infecter de larges volumes de cultures cellulaires en production.

2.2 Clonages

2.2.1 Clonage des séquences de XRCC4 et de la Ligase4 dans pKL.

La séquence codante pour la protéine humaine entière Ligase4 (aa 1-911) (gène fourni par JP de Villartay, Institut Imagine, Paris) a été clonée dans le vecteur pKL dans le site MCS1 en utilisant les sites de restrictions BamHI et XbaI. Après transformation de cellules d'*E. coli* XL1Blue par choc thermique, puis sélection des colonies obtenues sur LB supplémenté de kanamycine et de gentamycine, le succès du clonage a été vérifié par PCR sur colonies. Le vecteur pKL-LigIV a ensuite été amplifié et purifié par miniprep pour insérer la séquence codante de la protéine humaine entière XRCC4 (X-ray repair cross-complementing 4) (aa 1-336) sur le site MCS2 en utilisant les sites de restrictions XhoI et SphI à partir de l'amplification d'un gène présent sur un plasmide fourni par l'équipe de JP de Villartay. La séquence codante pour une étiquette 10 histidines a été ajoutée en C-ter de XRCC4, sans site de clivage entre l'étiquette et la protéine. Aucune étiquette n'a été ajoutée à la séquence de la Ligase4. Après transformation par choc thermique et vérification de la présence des séquences codantes pour la Ligase4 et pour XRCC4 dans les clones par PCR sur colonies, le vecteur pKL-XRCC4-LigIV est amplifié et purifié par miniprep **(Figure 23)**.

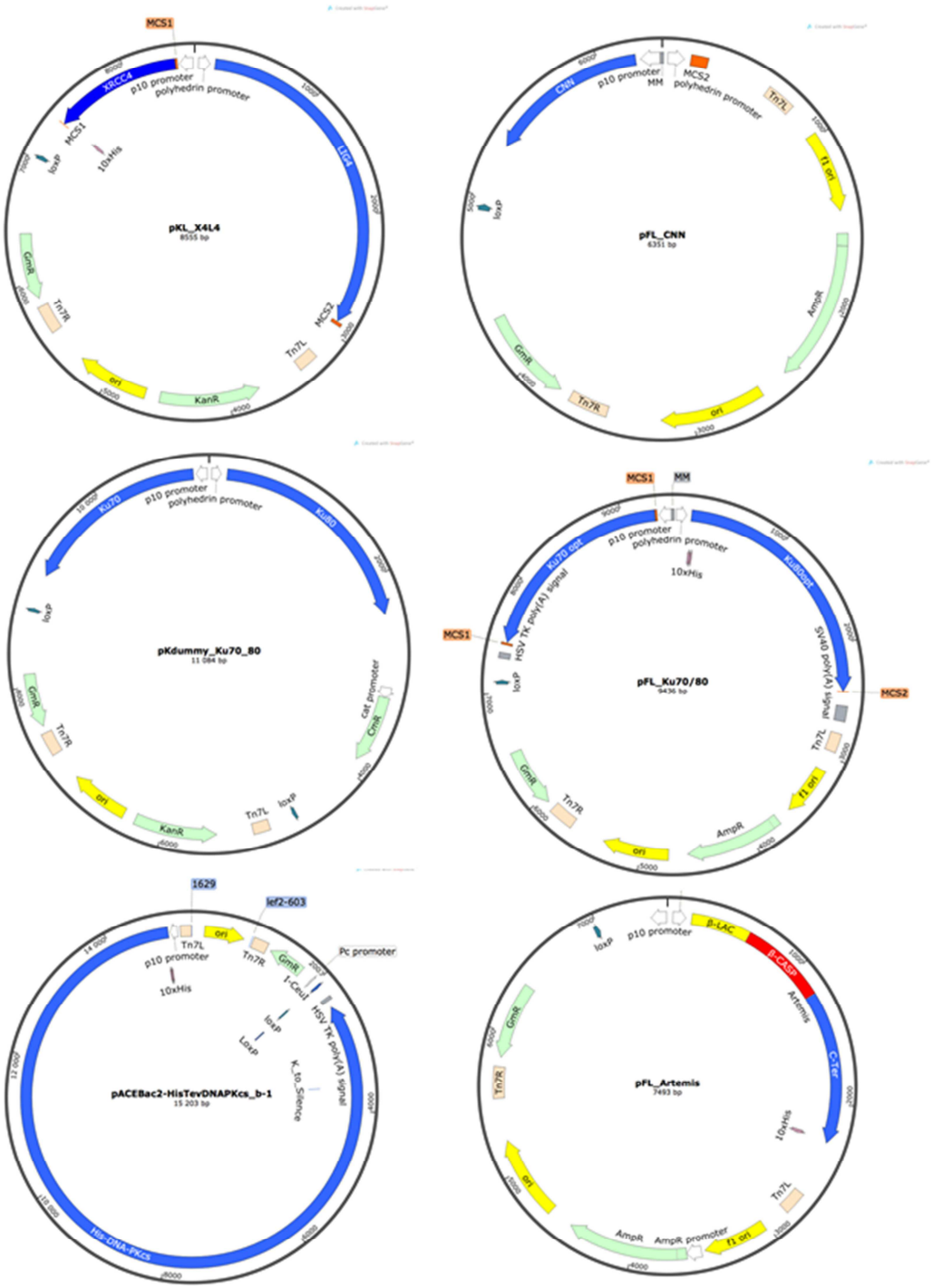


Figure 23 : Cartes des différents plasmides utilisés dans cette étude.

Les séquences codant pour les protéines sont en bleu (sauf pour Artemis où les 3 domaines sont coloriés). Les résistances sont représentées en vert. Le plasmide construit pour DNA-PKcs est présenté mais n'a pas été utilisé dans la thèse. Il est utilisé actuellement au laboratoire pour essayer de produire la kinase.

2.2.2 Clonage des séquences de Cernunnos dans pFL

La séquence codant pour la protéine humaine entière (aa 1 à 299) a été amplifiée par PCR à partir d'un plasmide donné par JP de Villartay et clonée dans le vecteur pFL en utilisant les sites XbaI et XhoI **(Figure 23)**.

2.2.3 Clonage des séquences de Ku70/80 avec ou sans étiquette dans pUCDM et pFL.

Les séquences sauvages de Ku70 et Ku80 humain sont amplifiées à partir de plasmides fournis par l'équipe de Patrick Calsou (IPBS, Toulouse) puis clonées dans le vecteur pUCDM en utilisant respectivement les enzymes XmaI/NheI et BssHI/SalI. Le vecteur pUCKu70/80 obtenu est séquencé pour vérifier l'intégrité des séquences. Le vecteur pUCDM étant un vecteur donneur, les séquences de Ku70 et Ku80 pourront être aisément transférées dans des vecteurs accepteurs comportant déjà d'autres protéines du NHEJ par la suite. En revanche, la production du complexe Ku70/80 ne peut être faite directement à partir de ce vecteur, c'est pourquoi nous avons réalisé une fusion de plasmide grâce la recombinase CreLox, en utilisant les vecteur pUCKu70/80 et pKDummy. Le vecteur pKDummy est une version simplifiée du vecteur pKL ne comportant par de sites de clonage mais un site loxP et les sites de recombinaison Tn7L et Tn7R. Le vecteur obtenu a été nommé pKdum_Ku70/80 **(Figure 23)**.

Une autre version d'un vecteur codant pour le complexe Ku70/80 est obtenue à partir de gènes synthétiques. Ces gènes, commandés chez Genscript.Cette compagnie a développé un algorithme, OptimumGene qui permet d'optimiser une séquence nucléotidique sur différents paramètres en vue d'améliorer l'expression dans un hôte donné. Les paramètres optimisés à travers des mutations silencieuses dans le gène sont l'adaptation aux codons les plus fréquents dans l'hôte utilisé et surtout la réduction des structures secondaires susceptibles d'être formées par les mRNA (http://www.genscript.com/codon_opt.html). Dans cette version, le gène codant pour Ku80 est précédé d'une séquence de 10 Histidines suivie d'un site de coupure par la protéase TEV, que nous produisons au laboratoire. Les séquences de Ku70 et Ku80 ont été clonées dans le vecteur pFL par l'utilisation des enzymes XhoI/NheI et BamHI/HindIII respectivement conduisant à l'obtention du vecteur pFL Ku70_80_opt **(Figure 23)**.

2.2.4 Clonage de la séquence d'Artemis dans pFL.

Un gène synthétique, codant pour la protéine Artemis, a été commandé chez Genscript, les données de la littérature suggérant une faible expression d'Artemis en cellules d'insecte. Le gène synthétique est construit par Genscript à partir de la séquence nucléotidique d'Artemis, déterminée par JP de Villartay lors de la découverte de la protéine *(Moshous, 2001)*. Comme pour Ku70 et Ku80, cette séquence est optimisée en vue de la production en cellules d'insecte. Une séquence additionnelle de 10 Histidine est ajoutée à l'extrémité C-terminale de la protéine, sans site de coupure entre l'étiquette et la protéine. Artemis a été clonée dans le vecteur pFL grâce à l'utilisation des enzymes BamHI et XbaI.

Deux versions plus courtes de la protéine ont été obtenues par insertion d'une cassette comportant une séquence codant pour 10His suivi d'un codon stop après les positions 370 et 500. Dans chaque cas, une mutation silencieuse a été faite afin de révéler un site de coupure unique dans la région où la cassette devait être introduite. Les versions 1-370 et 1-500 d'Artemis correspondent à deux versions de la protéine contenant les domaines β-lact/βCasp mais des régions C-terminales de tailles réduites, cette région ayant été prédite comme étant peu structurée *(Callebaut, 2002)*.

2.2.5 Clonage des séquences de DNA-PKcs dans pACBAC2.

Un gène synthétique codant pour la séquence de DNA-PKcs a été commandé chez Genscript par notre collaborateur I Berger dans le cadre d'un projet ANR. Les données de la littérature suggèrent comme pour Artemis des niveaux d'expression très faible de DNA-PKcs par voie recombinante. Le gène synthétique est construit par Genscript à partir de la séquence nucléotidique de DNA-PKcs, clonée pour la première fois par le laboratoire de S Jackson *(Hartley, 1995)*. Le gène synthétique est introduit dans le vecteur pACBAC2. Une séquence aditionnelle de 10His suivie d'un site de coupure par la protéase TEV a été ajoutée en position N-terminale (**Figure 23**).

2.3 Entretien des cellules d'insecte

Les cellules SF21 utilisées au laboratoire (don du laboratoire d'I Berger ou SF21 Gibco en SF-900 II) et High five (don du laboratoire de M Mirande, I2BC, Gif) sont entretenues en

milieu SF-900 II, en maintenant en permanence un stock de cellules en petit volume (généralement 50mL) dans une gamme de concentrations où elles se portent bien ou « Healthy Range », à savoir entre $0,5x10^6$ et $2x10^6$ cellules par mL. Ce maintien se fait par dilution dans du milieu neuf sans sérum. Toutes les manipulations sont réalisées sans antibiotique, sous une hotte à flux laminaire.

2.4 Préparation des Virus

Le vecteur donneur contenant les séquences des protéines à produire est utilisé pour transformer des cellules EMBacY par électroporation. Ces bactéries transformées sont étalées sur LB supplémenté de X-gal 40µg/mL, d'IPTG 0,1mM, de kanamycine 30µg/ml, de tetracycline 20µg/ml, de gentamycine 20µg/ml. L'insertion du vecteur dans le bacmide se fait par recombinaison *in vivo* via les sites Tn7 et grâce à la présence de T7 recombinase codée par un plasmide présent dans les cellules. Les sites Tn7 portés par le bacmide se trouvent à l'intérieur de la séquence codant pour le fragment lacZ (**Figure 24**). Lorsque les séquences de protéines recombinantes sont insérées dans le bacmide, le fragment lacZ est interrompu et ne peut plus complémenter la ß-galactosidase déficiente produite par la bactérie. La bactérie ne peut plus métaboliser le X-gal et libérer le produit de couleur bleue. Les colonies blanches, ayant inséré le plasmide, sont isolées et striées sur LB supplémenté de la même façon.

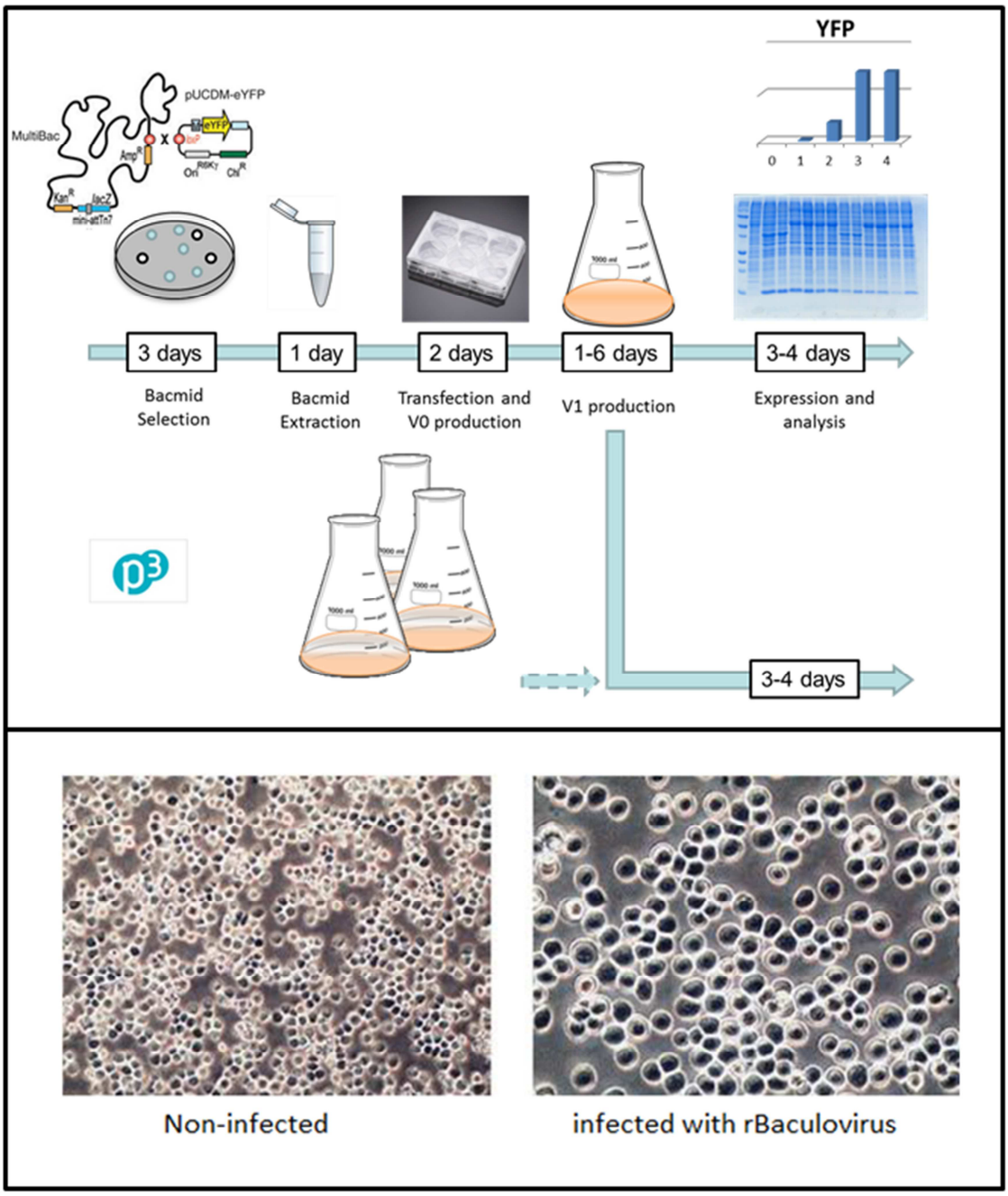


Figure 24 : Schéma du protocole de production des vecteurs d'expression en système Multibac.
(Haut) Description des étapes principales de la production en cellules d'insecte. (Bas) Aspect morphologique de cellules infectées et non-infectées par un baculovirus (issus du Web)

Ces clones sont mis en culture (2mL) en LB supplémenté en antibiotiques (Kan 30 µg/ml, Tet 20 µg/ml, Gent 20 µg/ml) une nuit à 37°C afin d'obtenir une préparation de bacmide (cf **Annexe 1**). Ce bacmide est être utilisé pour transfecter $1x10^6$ cellules Sf21 dans du milieu de culture SF900 II, disposées dans une boîte de culture 6 puits. Il est primordial d'effectuer la purification du bacmide juste avant de réaliser la transfection. Ce paramètre est déterminant dans le succès de la transfection. Les cellules ne sont pas adhérentes, et

doivent être déposées la veille ou quelques heures avant la transfection de façon à ce qu'elles se déposent au fond des boîtes 6 puits. Un taux de confluence compris entre 70% et 90% de confluence est recommandé.

Les boîtes sont inspectées chaque jour pour vérifier l'absence de contamination (**Figure 24**). Au bout de 3 jours, le surnageant de culture des puits manifestant des signes d'infection est collecté et conservé à 4°C à l'abri de la lumière, il s'agit du stock V0 de virus. 3mL de milieu de culture neuf sont ajoutés afin de permettre aux cellules de proliférer 3 jours supplémentaires. Les cellules sont ensuite collectées en raclant le fond des plaques 6 puits.

Les cellules d'insecte récoltées lors cette première étape servent à réaliser un premier test d'expression lors de la production d'une nouvelle protéine ou d'un nouveau complexe. On peut ainsi vérifier à l'aide d'un western blot la présence de la protéine d'intérêt. La concentration finale des cellules est calculée en comptant les cellules par microscopie optique à l'aide d'une cellule de Malassez ou dans un compteur de cellule BioRad TC10. Environ $1x10^6$ cellules sont centrifugées à 3000G, à 4°C pendant 5 minutes, le surnageant est retiré et les cellules sont re-suspendues dans 1mL PBS 1X. Les cellules sont lysées par sonication. 10µL du lysat sont prélevés et conservés à 4°C et constituent un échantillon de la fraction totale des protéines. Le reste du lysat est centrifugé à 20000G à 4°C pendant 15 minutes. 10µL de surnageant sont prélevés et conservés à 4°C, il s'agit de la fraction soluble des protéines. Les fractions totales et solubles prélevées sont déposées sur gel SDS-PAGE. Le reste de la fraction soluble du lysat est analysée par fluorimétrie pour détecter la présence ou non de YFP en utilisant comme longueur d'onde d'excitation 488nm et 520nm comme longueur d'onde d'émission.

Le volume de surnageant V0 étant relativement faible, il est nécessaire de l'amplifier une fois, en infectant un volume plus important de culture cellulaire, pour collecter la seconde génération de virus, appelée V1. La totalité du V0 (3 ml) est utilisé pour infecter 25 ml de cellules à $0,5x10^6$ cellules/ml. Les cellules sont comptées toutes les 24h et le volume est ajusté afin de conserver une concentration de $1x10^6$ cellules/ml tant que les cellules continuent de doubler. Lorsque les cellules croissent sans doubler, on a atteint le jour appelé DPA (Day after Proliferation Arrest) et le prélèvement du surnageant de culture se fait le

lendemain (dpa+24) par centrifugation et constitue le stock V1. Cette génération de virus servira pour toutes les infections et productions suivantes.

2.5. Titration des virus

La titration des particules virales est effectuée par la méthode de dilution limite en plaque 96 puits. Une culture de Sf21 à $2*10^6$ cellules/ml en phase exponentielle est diluée à $0,5*10^6$ cellules/ml puis distribuée à raison de 100µL par puits. Après 1h d'incubation à 27°C le surnageant est éliminé et remplacé par 100µL de milieu frais, 100µL de solution virales est alors ajoutée. Une dilution en cascade de 2 en 2 est alors réalisée à partir du puits initial. 3 puits par dilution sont ainsi effectués. On laisse incuber pendant 6 jours à 27°C sans agitation. La lecture se fait au microscope à fluorescence. Chaque puits est noté positivement ou négativement en fonction de la présence ou non de cellules produisant de la YFP. On groupe alors les scores par 3 dilutions successives (ex: 321 signifie 3 tubes (+) dans la première dilution, 2 dans la suivante et 1 dans la troisième). On prend en compte la suite de dilutions donnant le score le plus élevé mais inférieur à 330. On consulte alors la table de Mc Grady (**Annexe 3**) qui donne le nombre le plus probable de particules dans la première dilution en fonction du score.

2.6 Cultures des cellules d'insecte

Une culture de cellule Sf21 ou Hi5 est amplifiée afin d'obtenir de 1 à 8 fioles de 400ml à $0,5x10^6$ cellules /ml. Chaque fiole est infectée par la dilution de V1 calculée précédemment et la culture se poursuit jusqu'à dpa + 3 jours ou 4 jours. La culture est suivie en croissance et en production de YFP comme décrit dans le paragraphe précédent. Le signal d'arrêt de la culture est l'entrée en phase plateau de la production de la YFP. Les cellules sont alors récupérées par centrifugation à basse vitesse (800rpm, 10min), resuspendues dans du PBS et réparties en tube Facon à raison de 200ml d'équivalent de culture initiale et stockées à -80°C.

3. Expression dans *E. coli* de XRCC4, Cernunnos et du domaine CTD de la Ligase4

3.1 Expression de XRCC4 et Cernunnos entiers

Les protéines XRCC4 et Cernunnos entières ont été surexprimées et purifiées en premier à partir de cultures d'*E. coli*. Nous utilisons un vecteur pETM-13Lic avec une étiquette 6His en Cter. Les plasmides sont introduits dans des cellules chimio-compétentes, Rosetta(Novagen), préparées au laboratoire à partir de cellules commerciales. Les plasmides sont introduits par choc thermique (45s à 42°C) après une incubation d'1h dans la glace avec 1µL de plasmide à 100ng/µL. Après transformation, les cellules sont agitées à 37°C dans 250µL de SOC pendant 1h. Les cellules transformées sont étalées sur boite LB/agar contenant 50µg/mL de chloramphénicol (Cam) et 30µg/mL de kanamycine (Kan). Des cellules provenant d'une colonie issue de l'un des étalements sont ensuite inoculées dans une préculture de 50mL de LB + Cam 50µg/mL + Kan 30 µg/mL ou Amp 100µg/mL et mises à incuber à 37°C avec agitation. La protéine Cernunnos entière est clonée dans un autre vecteur pETM13-Lic et une préculture est préparée de la même façon.

Les cultures de ces deux protéines sont réalisées par ensemencement de 8mL de la préculture dans 0,8L de (LB + Cam 50µg/mL + Kan 30µg/mL) dans des fioles de type Fernbach de 2L, placées à 37°C sous agitation. Lorsque les cultures atteignent une DO^{600nm} de 0,6, les cultures sont placées à 20°C sous agitation. Les cultures sont induites à une DO^{600nm} comprise entre 0,8 et 1 avec 0,5mM d'IPTG et sont laissées à 20°C sous agitation pendant une nuit. Les cultures sont ensuite centrifugées à 3000g pendant 45 minutes à 4°C.

3.2 Expression de XRCC4(1-203), Cernunnos (1-224) et le domaine CTD de la Ligase4 (654-911)

Nous avons souhaité évaluer l'influence que pouvait avoir la fixation du domaine C-ter de la Ligase4 sur XRCC4, sur l'interaction entre XRCC4 et Cernunnos. XRCC4 interagit en deux sites éloignés avec ces deux protéines. Il interagit par son coiled-coil avec la region 654-911 de la Ligase4 et par sa tête globulaire en N-ter avec Cernunnos. Pour évaluer l'influence de la fixation de la Ligase4 sur l'interaction XRCC4-Cernunnos, nous avons exprimé et purifié le complexe Ligase4(CTD)-XRCC4(1-203) et Cernunnos(1-224). Les conditions d'expression

de Cernunnos et XRCC4 sont semblables à celles décrites dans *(Ropars, 2011)*. Le plasmide contenant L4(CTD) nous a été fourni par M Modesti (Instabilité du Génome et Cancérogenèse, Marseille). Les cultures de Ligase4(CTD) sont réalisées de la même façon.

4. Purifications et co-purifications des protéines exprimées

Les purifications décrites dans les paragraphes suivants correspondent principalement à des purifications réalisées à partir de culots de culture en cellules d'insecte. Seules certaines purifications de XRCC4 et Cernunnos entières ont été réalisées à partir de culots d'*E. coli*.

4.1 Purifications de XRCC4/Ligase4

Un culot de 200mL de culture de cellules d'insecte est repris dans 25mL de tampon de Lyse-XL4 (25mM Tris pH8, 150mM KCl, 150mM NaCl, 10mM 2-Me-EtOH, 5% Glycérol, 50mM Imidazole) (voir **Annexe 4**). Après ajout d'inhibiteurs de protéases (Pepstatine (1µg/ml) / Chymostatine (1µg/ml) / Leupeptine (0,5µg/ml) / AEBSF ou PMSF (1mM) / Benzamidine (34µg/ml) final), 10µL sont prélevés pour le premier gel SDS-PAGE (échantillon extrait total noté ET). Après re-suspension à la Propipette ou au Vortex, les cellules sont lysées par sonication (3x1 minute à 60% d'amplitude, (1s on/1s off)) en veillant à maintenir le lysat entre 0°C et 4°C, en plongeant le tube récepteur (de préférence métallique) dans un bain de glace / carboglace / eau / éthanol autour du récipient contenant le lysat. 0,02U/µl de Benzonase Merck + 1mM de $MgCl^2$ final sont ajoutés.

Le lysat est clarifié par centrifugation, 50000g à 4°C pendant 30 minutes. Le surnageant est conservé et 100µL sont conservés pour analyse sur gel SDS-PAGE (échantillon noté S). Le culot est resuspendu dans 20 mL solution 2% de SDS, 100µl de culot resuspendu sont conservés pour analyse sur SDS-PAGE (échantillon noté C). Le surnageant est ensuite filtré à 0.45µm. Après équilibration de 8 mL de résine Ni-NTA (Qiagen) avec du tampon TpLyse-XL4, le surnageant est mis à incuber 2h à 4°C dans une colonne avec cette résine. Après incubation, la fraction non-retenue (FT ou Flow Through) est collectée, et la colonne est lavée en utilisant 20 volumes de colonne de TpLavage-XL4 (Tampon de Lyse avec 850mM NaCl et 10% glycerol) (**Annexe 4**). Le lavage est collecté et analysé sur gel SDS-PAGE.

L'élution des protéines est réalisée avec 2 à 3 volumes de colonne de Tampon d'Elution NiNTA-XL4 (Tampon de Lyse avec 300mM imidazole et 10% glycerol) par fraction de 1mL. Les fractions sont analysées sur gel SDS-PAGE. Après analyse sur le gel de tous les échantillons collectés au cours des différentes étapes de purifications, les fractions d'intérêt sont dosées par Bradford, analysées par spectrométrie UV et regroupées en fonction du gel SDS-PAGE.

Les fractions d'intérêt sont dialysées contre du tampon TpQA-XL4 (25mM Tris pH8, 100mM KCl, 10mM 2-Me-EtOH), 3 fois 1h ou 1 fois 2h puis 1 fois sur la nuit à 4°C avec agitation. 50µl de dialysat sont mis de côté pour doser la concentration protéique par Bradford et mesurer le rendement de l'étape de dialyse. Une colonne Ressource Q 6mL (GE-Healthcare) est équilibrée avec du TpQA-XL4 et les fractions dialysées sont chargées. La fraction non-retenue (FT) est collectée et analysée par gel SDS-PAGE (échantillon noté FT-Q). Les protéines sont éluées en utilisant un gradient de tampon TpQB-XL4 (TpQA-XL4 avec 850mM NaCl en plus) de 0 à 40% en 120 minutes à un débit de 2mL par minute en collectant des fractions de 2mL. Le chromatogramme présente deux pics sur la courbe de suivi de l'UV à 280nm, un premier pic correspond au complexe Ligase4-XRCC4 et un second pic au dimère XRCC4 seul. Les deux pics sont analysés en déposant des échantillons des fractions correspondantes sur gel SDS-PAGE.

Les fractions d'intérêts sont rassemblées et dialysées dans un tampon adapté à la manipulation ultérieure envisagée avec cette production. Le dialysat est ensuite concentré sur système d'ultrafiltration de seuil de coupure 10kDa et de volume maximum de 4 ou de 15 mL. Les protéines sont utilisées immédiatement en cristallogenèse ou conservées à 4°C. Les échantillons purifiés qui ne sont pas destinés à une utilisation immédiate sont congelés après ajout de 10% de Glycérol (final) dans de l'azote liquide par aliquotes de 0.5mg.

4.2 Purifications de Ku70-Ku80

Les essais de purifications sur le complexe sans étiquette de Ku70/Ku80 n'ont pas donné de résultats concluant hormi leur de la co-purification avec le complexe Ligase4/XRCC4 (voir Résultats § 2.5.2 plus loin)

Ku80 présente une étiquette 10His suivie d'un site de coupure à la TEV (Tobacco Etch Virus) en position N-terminale. Ku70 est sans étiquette. Un culot de 200mL de culture est

repris dans 25mL de tampon de TpLyseNiNTA-Ku (20mM Tris pH8, 75mM KCl, 425mM NaCl, 10 mM β-Mercapto, 40mM Imidazole) (**Annexe 5**). La lyse ainsi que la première étape de purification, une chromatographie d'affinité par passage sur résine Ni-NTA, sont réalisées dans les mêmes conditions que pour le complexe Ligase4/XRCC4 à quelques exceptions. Le tampon de Lyse TpLyseNiNTA-Ku et le tampon d'élution TpElutionNiNTA-Ku contiennent tous les deux 425mM de NaCl et 75mM de KCl, afin de dissocier Ku de l'ADN libre, qui autrement contamine les protéines purifiées.

Les fractions d'intérêt sont dialysées contre du tampon TpQA-Ku (25mM Tris pH8, 100mM KCl, 10 mM β-Mercapto) à 4°C avec agitation. 50µl de dialysat sont mis de cotés pour doser la concentration protéique par Bradford et mesurer le rendement de l'étape de dialyse. Une colonne Ressource Q 6mL (GE-Healthcare) est équilibrée avec du TpQA-KuTag avant d'y faire passer le dialysat. Les protéines sont éluées en utilisant un gradient de tampon TpQB-Ku (25mM Tris pH8, 150mM KCl, 850mM NaCl, 10 mM β-Mercapto) de 0 à 60% en 120 minutes à un débit de 2mL par minute en collectant des fractions de 2mL. Le chromatogramme présente un pic principal sur la courbe d'UV à 280nm. Les fractions sont analysées sur gel SDS-PAGE. Toutes les analyses sur SDS-PAGE sont réalisées en doublant le rapport Laemmli sur échantillon protéique habituel, et en déposant deux fois moins d'échantillon. En effet, Ku80 a tendance à présenter une migration diffuse sur gel SDS-PAGE dans les conditions classiques de dénaturation. Les échantillons sont dialysés, concentrés et conservés de façon semblable au complexe Ligase4/XRCC4 pour la cristallogenèse et les autres utilisations.

4.3 Purification de Cernunnos

Un culot de 200mL de cellules infectées par les virus V1 de Cernunnos est resuspendu dans 35mL de Tp-Lyse-Cer (Tris pH=8, 20mM, NaCl 350mM, KCl 150mM, ßMercapto 10mM Glycerol 10%) (**Annexe 6**). La lyse est effectuée dans les mêmes conditions que pour la purification de L4-X4. Après clarification par centrifugation et filtration, le lysat est chargé à 2mL par minute sur une colonne HiTrap Q (GE Healthcare) équilibrée dans du tampon TpQA-Cer (Tris pH 8, 20mM, NaCl 50mM, KCl 50mM, ßMercapt 10mM). La fraction non-retenue est collectée pour analyse sur gel SDS-PAGE. La colonne est lavée avec du tampon TpQA jusqu'à

retour à la ligne de base en UV sur le chromatogramme. Les protéines sont éluées en faisant passer un gradient de TpQB-Cer (Tris pH 8, 20mM, NaCl 850mM, KCl 150mM, ßMercapt 10mM) sur la colonne, 150mL à 2mL/min, de 0 à 40% de TpQB. Des fractions de 1mL sont collectées. Cernunnos sort sur le chromatogramme dans le premier pic (UV 280nm) à ~30% de TpQB. Les fractions contenant Cernunnos sont rassemblées et dialysées contre du TpQA-Cer.

La seconde étape de purification est à nouveau un passage sur une colonne échangeuse d'anion de type Resource Q 6 mL. Le dialysat est chargé sur la colonne équilibrée en TpQA-Cer à une vitesse de 2mL/min. La colonne est lavée avec 150 mL de tampon TpQA jusqu'à retour à la ligne de base en UV sur le chromatogramme. L'élution de Cernunnos est réalisée en passant un gradient de TpQB-Cer sur la colonne, de 0 à 40 % en 2 heures à 2mL/min. Le gradient est arrêté lorsqu'une première hausse du spectre UV est observée. On laisse l'UV redescendre à la ligne de base avant de recommencer le gradient. Les fractions éluées sont ensuite dialysées contre un tampon de conservation, identique à celui utilisé pour conserver L4-X4 puis congelées dans l'azote liquide pour une utilisation ultérieure.

4.4 Purifications de Artemis

Le protocole pour la purification d'Artemis est très proche de celui décrit pour le complexe Ligase4-XRCC4 avec quelques adaptations (**Annexe 7**). On effectue la lyse d'un culot de 200ml de cultures de cellules d'insecte dans un tampon contenant 40mM imidazole, 75mM KCl et 425mM NaCl. Le lysat est chargé sur une résine Ni-NTA et la résine est lavée avec 1M NaCl. L'élution est réalisée avec 2-3 volumes de colonne et un tampon contenant 300mM imidazole. Les fractions sont regroupées en fonctions du gel SDS-PAGE et dialysées contre un tampon contenant 100mM KCl. Artemis est éluée avec un gradient de 0 à 40% d'un tampon contenant 150mM KCl, 850mM NaCl sur 20 volumes de colonnes. Les fractions d'intérêts sont regroupées et dialysées dans un tampon adapté à la manipulation ultérieure.

4.5 Co-purifications

4.5.1 Co-purification de Ligase4/XRCC4 et Cernunnos

XRCC4 et la Ligase4 sont co-exprimés avec XRCC4 qui possède une étiquette 10His en position N-terminale. Cernunnos n'a pas d'étiquette. Un culot de 50 mL de cellules d'insecte exprimant la Ligase4 et XRCC4 est mélangé avant l'étape de lyse avec un culot de 50 mL de cellules d'insecte exprimant Cernunnos. Le tampon utilisé pour resuspendre les culots cellulaires est le même que celui utilisé dans la purification de L4-X4. La première étape de purification, une colonne d'affinité par passage sur une résine NiNTA est effectuée de la même façon que pour la purification de L4-X4. Les étapes de lyse et de lavage sont cependant réalisées 100mM KCl et 50 mM NaCl. La seconde étape de purification est une colonne échangeuse d'anions (Ressource Q), réalisée dans les mêmes conditions que pour la purification de L4-X4. Nous observons deux pics sur le chromatogramme. Un premier pic qui est un complexe de L4-X4 et Cernunnos, et un second qui ne contient que du complexe L4-X4. Les fractions d'élutions correspondantes sont séparées et rassemblées, et les protéines purifiées sont congelées dans les mêmes conditions que pour L4-X4.

4.5.2 Co-purification de Ligase4/XRCC4 et Ku70/Ku80

XRCC4 et la Ligase4 sont co-exprimés avec XRCC4 qui possède une étiquette 10His en position N-terminale. Cernunnos n'a pas d'étiquette. Ku70/Ku80 est exprimé sans étiquette (premier vecteur du complexe). Un culot de 50 mL de cellules d'insecte exprimant la Ligase4 et XRCC4 est mélangé avant l'étape de lyse avec un culot de 50 mL de cellules d'insecte exprimant Ku70/Ku80 sans étiquette. La purification est effectué selon le même protocole que pour le complexe L4-X4 avec une première étape sur colonne de. Les étapes de lyse et de lavage sont cependant réalisées 75mM KCl et 425mM NaCl pour séparer l'ADN de Ku.

4.5 Purification de XRCC4 et de Cernunnos entiers produits en *E.coli*

Les protéines XRCC4 et Cernunnos entières présentent une étiquette 6His en Cter. Les deux protéines sont purifiées de manière semblable. Un culot correspondant à 800ml de culture dans *E. coli* de XRCC4 ou 3,2L de Cernunnos est re-suspendu au 1/20e dans le

tampon de lyse TpLyse-Coli (20mM Tris-HCl pH8, 10% glycerol, 10mM 2-Me-EtOH, 100mM KCl, 1mM PMSF, 34µg/mL Benzamidine, 5mM EDTA). Les culots sont congelés en l'état dans l'azote liquide et conservés à -80°C. Le même protocole a été utilisé pour purifier les deux protéines. On ajoute 1 mM de PMSF et du lysozyme à une concentration finale de 2mg/mL aux culots resuspendus qui sont mis à incuber dans la glace pendant 45 minutes. La lyse est réalisée par sonication avec une amplitude 60% pendant 1min avec des cycles de 1sec on/1sec off et une température maximale de l'échantillon de 15°C. Cette opération est répétée 3 fois. Après la lyse, 1mM de PMSF sont à nouveau ajoutés ainsi que 125 unités de Benzonase par litre de culture lysé et 10mM de $MgCl_2$. Enfin, le lysat est centrifugé 45 minutes à 50000g à 4°C.

Le surnageant de lyse cellulaire de XRCC4 ou Cernunnos, filtré à 0,45µm, est chargé sur une colonne échangeuse d'anion (HiTrap Q HP, 5 mL, GE-Healthcare). La colonne est équilibrée dans un tampon QA (tampon QA : 20mM Tris-HCl pH8, 100mM KCl, 2mM EDTA, 10% Glycérol, 10mM 2-Me-EtOH, 34µg/mL Benzamidine). Jusqu'à 50mL de surnageant sont chargés à un débit de 3mL/min. Un gradient est appliqué de 0% à 50% de tampon QB (tampon QB = tampon A avec 1M NaCl) sur 40 volumes de colonne (soit 200mL). Des fractions de 3mL sont collectées. Les fractions retenues lors de l'échangeuse d'anions sont déposées sur gel SDS-PAGE. Le gel est analysé par Western-Blot avec des anticorps anti-6His. Le transfert du gel SDS-PAGE sur membrane de nitrocellulose (PROTRAN BA83, Whatman) est effectué pendant 1h15 à 90mA (générateur Electrophoresis Power Supply 601). La membrane de nitrocellulose est saturée pendant 1h à température ambiante avec une solution contenant de la BSA (solution de saturation : PBS 1X, 0,1% Tween, 2% BSA). La membrane est ensuite incubée dans une solution d'anticorps anti-HisTag (His-probe H-3 HRP, IgG1 monoclonal de souris, Santa Cruz Biotechnology) pendant 1h à température ambiante (solution d'anticorps : PBS 1X, 0,1% Tween, 5ng d'anticorps). La membrane est lavée 3 fois pendant 5 mn avec un tampon PBS-T (tampon PBS-T : PBS 1X, 0,1% Tween) puis 1mL de chaque réactif d'un kit ECL (Western Bloting Detection Reagents, GE-Healthcare) sont ajoutés. La membrane est alors révélée contre un film photographique (Amersham Hyperfilm MP, GE-Healthcare). Cette étape permet l'identification des fractions éluées d'échangeuses d'anions qui contiennent des protéines avec une étiquette 6His.

Les fractions issues d'échangeuse d'anions qui contiennent les protéines d'intérêt d'après le Western Blot sont dialysées dans des boudins de dialyse Spectra/Por (seuil de coupure 6-8,000 Da, contenance 1,7mL/cm). Ces fractions sont dialysées contre 50 fois le volume du dialysat dans le tampon NiA (tampon NiA : 20mM Tris-HCl pH8, 500mM KCl, 10% Glycérol, 10mM 2-Me-EtOH, 34µg/mL Benzamidine, 20mM Imidazole). La seconde étape de purification est un passage sur colonne de Nickel Ni-NTA. Les protéines dialysées contre le tampon NiA sont passées sur 5mL de résine Ni-NTA équilibrée avec 5 volumes de colonne de tampon NiA. Un lavage est effectué avec 10 volumes de colonnes (soit 50mL) de tampon Ni-40 (tampon NiA avec 40mM Imidazole). Les protéines fixées sont éluées avec 3 volumes de colonne de tampon Ni-300 (tampon NiA avec 300mM Imidazole) et collectées par fractions de 1mL.

4.6 Purification du complexe Ligase4(CTD)-XRCC4(1-203) et de Cernunnos(1-224) produits en *E.coli*

L'expression et la purification de L4(CTD 654-911) seule dans *E. coli* ne nous a pas permis d'obtenir une quantité suffisante de L4(CTD) pure. Nous avons donc choisi de co-purifier L4(CTD) en présence de XRCC4(1-203). On peut noter que la construction utilisée de XRCC4 est plus longue de 46 acides aminés par rapport à la forme tronquée XRCC4(1-157) utilisée précedemment *(Ropars, 2011)*. La forme XRCC4(1-203) contient le site d'interaction avec L4(CTD). Les deux protéines L4(CTD) et XRCC4(1-203) sont exprimées séparément et les cellules correspondant à 1 litre de culture de chaque protéine sont mélangées avant la lyse cellulaire. Les cellules de XRCC4(1-203) et L4(CTD) sont lysées ensemble et centrifugées. La première étape de purification est une étape de fixation par l'étiquette 6His de chacune des deux protéines, XRCC4(1-203) et L4(CTD), sur une colonne de nickel. Après une étape de lavage avec 0,5M KCl et 40 mM imidazole, les protéines sont éluées avec 300mM d'imidazole. La seconde étape de la purification est un passage sur une colonne échangeuse d'anions (ResourceQ, 1mL, GE-Healthcare). Le complexe L4(CTD)-XRCC4(1-203) est élué entre 200 et 250mM NaCl. Ce protocole permet d'obtenir 5 mg de complexe L4(CTD)-XRCC4(1-203) par litre de culture. La protéine Cernunnos(1-224) est purifiée selon le protocole décrit dans *(Ropars, 2011)* avec une étape de fixation sur colonne de Nickel et une étape de passage sur colonne échangeuse d'anions. Le protocole de purification de Cernunnos(1-224) permet d'obtenir 10 mg de protéine par litre de culture.

5. Analyses biochimiques

5.1 Chromatographies d'exclusions de taille

Les échantillons purifiés sont analysés par chromatographie d'exclusion de taille en utilisant des colonnes Superdex 10/300 GL et Superose 6 10/300 GL (GE Healthcare). Les colonnes sont calibrées avant utilisation en utilisant des protéines commerciales : Ovalbumine 44kDa, Conalbumine 75kDa, Aldolase 158kDa, Ferritine 440kDa, Thyroglobuline 669kDa et Blue Dextran >2M Da. Après équilibration des colonnes, les échantillons sont chargés sur les colonnes à 0,5mL/min. Les protéines sont collectées par fractions de 0,5mL et anaysées par gel SDS-PAGE.

5.2 Caractérisation des modifications post-traductionnelles

5.2.1 Coloration au Pro-Q Diamond

La présence de modifications post-traductionnelles dans nos échantillons de protéines purifiées, et plus particulièrement de phosphorylations, est analysée en utilisant le kit Pro-Q Diamond Phosphoprotein Gel Stain (Life Technologies). Cette méthode permet de colorer en 4-5h spécifiquement les protéines phosphorylées dans un gel SDS-PAGE et de détecter 1-16ng de protéine phosphorylée par bande selon l'état de phosphorylation. Le colorant utilisé a des maxima d'émission et d'excitation à respectivement 555nm et 580nm. Nous suivons au laboratoire le signal à l'aide d'un Typhoon (GE-Healthcare) qui est un système de visualisation flexible avec plusieurs lasers pour la quantification et l'imagerie des protéines, acides nucléique et sondes de fluorescence.

5.2.2 Déphosphorylation à l'aide de phosphatases

Nous avons utilisé au laboratoire jusqu'à maintenant deux phosphatases pour déphosphoryler nos protéines : la Lambda Protein Phosphatase et la Phosphatase Alcaline de veau. La Lambda Protein phosphatase ou Lambda PP (New England Biolabs 400kU/ml) est une protéine phosphatase à Mn^{2+} avec une activité dirigée contre les serines, thréonines et tyrosines phosphorylées. C'est le gène ORF221 de 221 acides aminés du bactériophage lambda. Une unité est définie comme la quantité d'enzyme nécessaire pour hydrolyser 1nmol de p-Nitrophenyl Phosphate en 1 minute à 30°C dans un volume de 50µl et dans le

tampon (50 mM HEPES, 1mM $MnCl_2$, 10 mM NaCl, 2 mM DTT, 0.01% Brij 35, pH 7.5 @ 25°C). L'enzyme est inhibée avec 50mM EDTA ou en chauffant 60min à 65°C.

La Phosphatase Alcaline utilisée (Ozyme, TAK2250A, 10x, 1000U/1ml) est l'enzyme originaire de l'intestin de veau et produite en levure. C'est une glycoprotéine dimèrique formée de 2 chaines de 50kDa. Elle hydrolyse presque tous les mono-esters de phosphate mais n'agit pas sur les liaisons di- ou tri-phosphate. Une unité est définie comme la quantité nécessaire pour produire 1 µmol de p-nitrophenol par minute à 37°C et pH 9.8, avec du p-nitrophenyl-phosphate comme substrat. Le tampon de réaction 10x est constitué de 500mM HCl pH9, 10mM $MgCl_2$.

5.3 Mesure sur gels SDS-PAGE de la stoechiométrie des protéines au sein d'un complexe

Pour évaluer la stoechiométrie de nos complexes protéiques, nous utilisons au laboratoire les gels Mini-PROTEAN® TGX Stain-Free™ et le système de lecture Gel Doc EZ de Biorad. Les gels TGX contiennent un composé organique de type halogénoalcane ou dérivé halogéné d'alacane. Un de ces composés, breveté par BioRad (WO 2013163451 A1) réagit avec les tryptophanes et modifie leur propriété de fluorescence. L'utilisation des gels contenant du trihalo permet de quantifier la quantité de protéines présente dans les différentes bandes du gel SDS-PAGE à condition qu'elles contiennent des Tryptophanes. Le signal émis par les bandes est proportionnel au nombre de tryptophane de chaque protéine et à la stoechiométrie des sous-unités présentes dans le complexe déposé sur gel. Nous avons utilisé dans un premier temps, cette technique sur le complexe Ligase4-XRCC4 purifié, avant et après congélation. La Ligase4 possède 10 Trp et XRCC4, 4 Trp et la stoechiométrie attendue est de 1 Ligase 4 pour 2 XRCC4.

6. Caractérisations Biophysiques

6.1 Caractérisation des interactions par Calorimétrie

Les mesures de calorimétrie sont très sensibles aux variations de tampons entre les solutions des deux partenaires sur lesquels on mesure une interaction. Les solutions

protéiques sont dialysées 3 fois pendant 2h et 100ml du dernier bain de dialyse est conservé pour les lavages et dilutions éventuelles. Les deux protéines sont concentrées. Lors des premiers essais de concentrations, les rendements de concentration sont déterminés afin d'évaluer entre les deux partenaires lequel est le plus soluble. On analyse également sur colonne d'exclusion de taille l'apparition d'agrégats ou d'oligomères lors de la concentration. Selon ces résultats, la protéine, appelée « ligand » qui se concentre le plus facilement sera mise dans la seringue à une concentration d'environ 100 fois le Kd attendu et un volume de 500µl est préparé pour une courbe de titration. La seconde protéine, appelée « récepteur », est préparée à une concentration de 10 fois le Kd attendu et un volume de 2,2ml pour aller dans la cellule du calorimètre. Dans le cas, de l'interaction XRCC4-Cernunnos, XRCC4 est mis dans la cellule à une concentration de 10-20µM et Cernunnos dans la seringue à une concentration de 100-200µM.

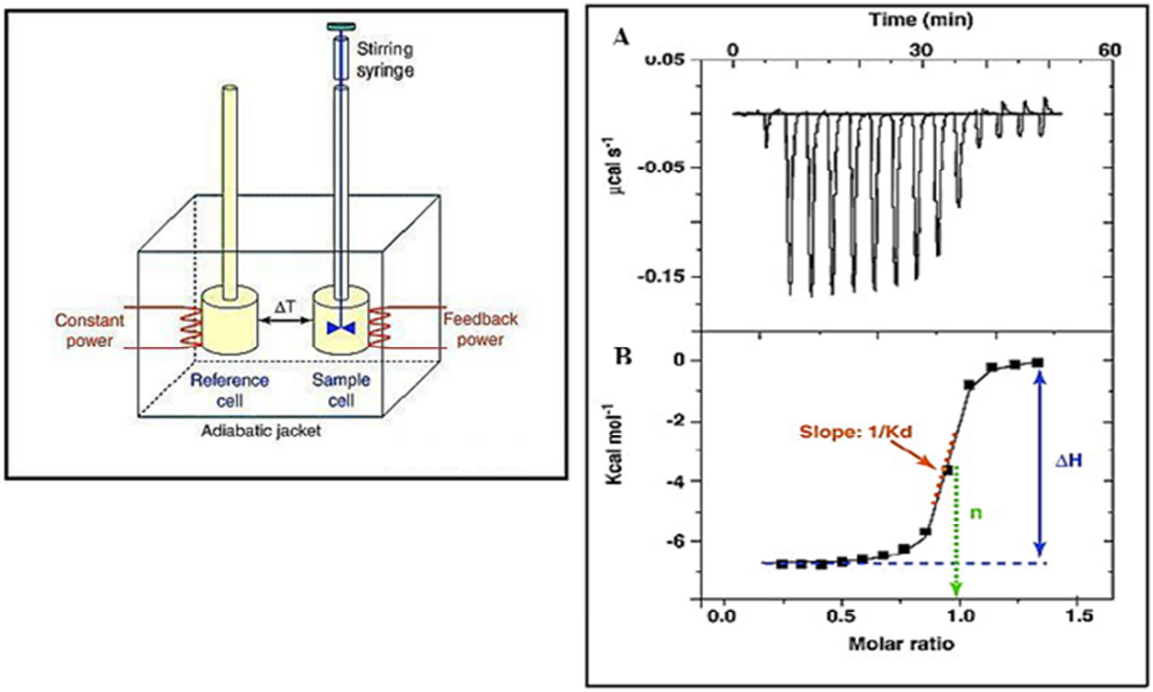


Figure 25 Mesure par Calorimétrie.

*(**Gauche**) Le calorimètre est constitué d'une enceinte adiabatique, d'une cellule de référence, d'une cellule de mesure, et d'une seringue munie d'un système d'agitation. Les deux cellules sont maintenues à température identique. Tout dégagement ou absorption de chaleur suite à une interaction entre un récepteur et un ligand se traduit par une réduction ou une augmentation de la puissance de chauffage appliquée à la cellule de mesure. (**Droite**) Le thermogramme (haut) présente sous forme de pics les microcalories dégagées ou absorbées suite à l'injection de 5-10µl de ligand dans les 1,8ml de récepteur présent dans la cellule. L'intégration de ces pics constitue l'isotherme de titration, dont le plateau inférieur correspond à l'enthalpie de la réaction en kcal.mol-1, la pente de la courbe donne la constante thermodynamique d'association, Ka en M-1 et la valeur du ratio molaire à mi-pente (nmole ligand/nmole récepteur) donne la stoechiométrie de la réaction (d'après G Craescu, Institut Curie, Orsay)*

Les deux solutions de protéines sont dégazées sous vide pendant 10min. Le calorimètre utilisé est un système VP-ITC (MicroCal, Malvern). Il mesure le dégagement ou l'absorption de chaleur (réaction exothermique ou endothermique, respectivement) pendant l'ajout progressif de ligand au récepteur. Le récepteur est placé dans la cellule de mesure (de volume ~1.8ml) sous agitation continue et le ligand y est ajouté à intervalle régulier à l'aide d'une seringue de 300µL par une série d'injections de 10 ou 12µl espacées d'un temps suffisant (240s). A chaque injection, l'interaction entre le récepteur et le ligand (si elle a lieu) s'accompagne d'un dégagement ou d'une absorption de chaleur qui entraîne une variation de la température de la cellule de travail. Afin de ré-équilibrer la température entre les deux cellules, l'appareil applique une puissance électrique W (en µcal.s-1) qui est mesurée par le microcalorimètre (**Figure 25**).

Au début de la titration, le récepteur est essentiellement libre et les molécules de ligands injectées se fixent sur le récepteur. On observe des pics d'échange de chaleur de taille similaire. Progressivement, les injections de ligand saturent le récepteur et, dans le cas d'une réaction exothermique (endothermique), des quantités positives (négatives) de chaleur sont dégagées indiquant la formation du complexe protéine-protéine. En fin de réaction, tous les sites de liaison du récepteur sont occupés et chaque nouvel ajout de ligand a comme seule conséquence sa dilution dans la cellule de travail.

L'énergie dégagée par l'interaction est enregistrée en fonction du temps puis intégrée par le logiciel Origin 7.0 (MicroCal). L'isotherme de titration qui s'ajuste sur la courbe expérimentale permet de déterminer la variation d'enthalpie de la liaison, $\Delta H°$ (kcal/mol), la constante d'association, Ka (M-1) ainsi que la stœchiométrie d'interaction (n) (**Figure 25**). Les autres paramètres thermodynamiques, à savoir les variations d'entropie ($\Delta S°$) et d'enthalpie libre ($\Delta G°$) ainsi que la constante de dissociation (Kd) peuvent être déduits des relations: $\Delta G° = \Delta H° - T\Delta S° = -RT\ln(Ka) = RT \ln(Kd)$ avec T (K), la température absolue et R, la constante des gaz parfaits, 8.314 $J.mol^{-1}.K^{-1}$. L'énergie libre $\Delta G°$ est donc la résultante de deux contributions énergétiques *(Holdgate, 2001)* (i) une contribution enthalpique reflétant principalement des liaisons hydrogène et des interactions de Van der Waals; (ii) Une contribution entropique reflétant principalement les interactions hydrophobes, des changements conformationnels dus à l'interaction ainsi que le relargage de molécules d'eau ou d'ions du fait de l'interaction.

Les principaux avantages de la méthode de microcalorimétrie isotherme de titration sont l'efficacité et la capacité à fournir tous les paramètres thermodynamiques d'une interaction. L'inconvénient majeur de cette technique est qu'elle nécessite une forte quantité de protéine et ligand, allant jusqu'à plusieurs milligrammes pour une seule courbe de titration pour de faible affinité (autour de 10µM).

7. Etudes structurales

7.1 Cristallisation des protéines seules et en complexes

La cristallisation d'une macromolécule est le passage de cette dernière d'un état liquide vers un état solide organisé grâce à des interactions non-covalentes. Ce passage d'un état vers un autre est un phénomène complexe. La cristallisation d'une protéine, d'un complexe protéine/protéine ou d'un complexe nucléo-protéique est encore aujourd'hui une étape largement empirique et souvent limitante dans la détermination de structure des protéines par cristallographie. La solubilité d'une protéine est influencée par de nombreux facteurs parmi lesquels le pH du tampon dans lequel elle se trouve, sa concentration, la température, la présence de solvant organique, de sels, de polymères organiques. La mise au point de nouvelles conditions de cristallisation et l'amélioration de celles-ci est donc un mélange d'essais-erreurs et de contrôle rationnel de la croissance du cristal.

Malgré le caractère empirique de la cristallisation, un certain nombre d'observations ont pu être dégagées suite à la cristallisation des quelques 97 084 structures cristallines présentes dans la PDB (base RCSB, Mai 2015) : (i) Les chances de succès d'un nouveau projet augmentent avec la qualité de la purification (échantillon pur à 95% ou plus), l'homogénéité (absence d'agrégat et de fragments de dégradation) et la stabilité dans le temps des macromolécules (même si certains cristaux apparaissent en une journée à 4°C). Les projets de génomique structurale de par le grand nombre de cibles étudiés et les protocoles utilisés ont confirmé cette première tendance (**Figure 26**). (ii) L'utilisation des conditions de cristallisation ayant eu le plus de succès sur un grand nombre de protéines ont des chances importantes de permettre la cristallisation d'une nouvelle protéine. En 1991, Jancarik et Kim ont ainsi proposé un crible reprenant les conditions de cristallisation retrouvées les plus fréquemment dans les structures cristallines de l'époque *(Jancarik, 1991)*. Plusieurs années

après, le consortium de génomique structural JCSG à Toronto a montré que sur 755 protéines étudiées par eux, les seules 24 premières conditions du crible Jancarik permettent de cristalliser 318 protéines soit 42%. Des cribles additionnels ont été proposés pour certaines applications spécifiques : complexes nucléo-protéiques, protéines membranaire, etc…

A

Protein property	Threshold value	Proteins	Crystals	%
Purity	>95%	229	134	59
	<95%	27	10	37
Monodispersity	>85%	94	58	62
	<85%	24	4	17
	Not determined	138	82	59
Folding state	<60% random coil	18	13	72
	>60% random coil	4	2	50
	Not determined	234	129	55
Folded	Yes	1	1	100
	Not determined	255	143	56
Identity	Determined	102	65	64
	Not determined	154	79	51
Total		256	144	56

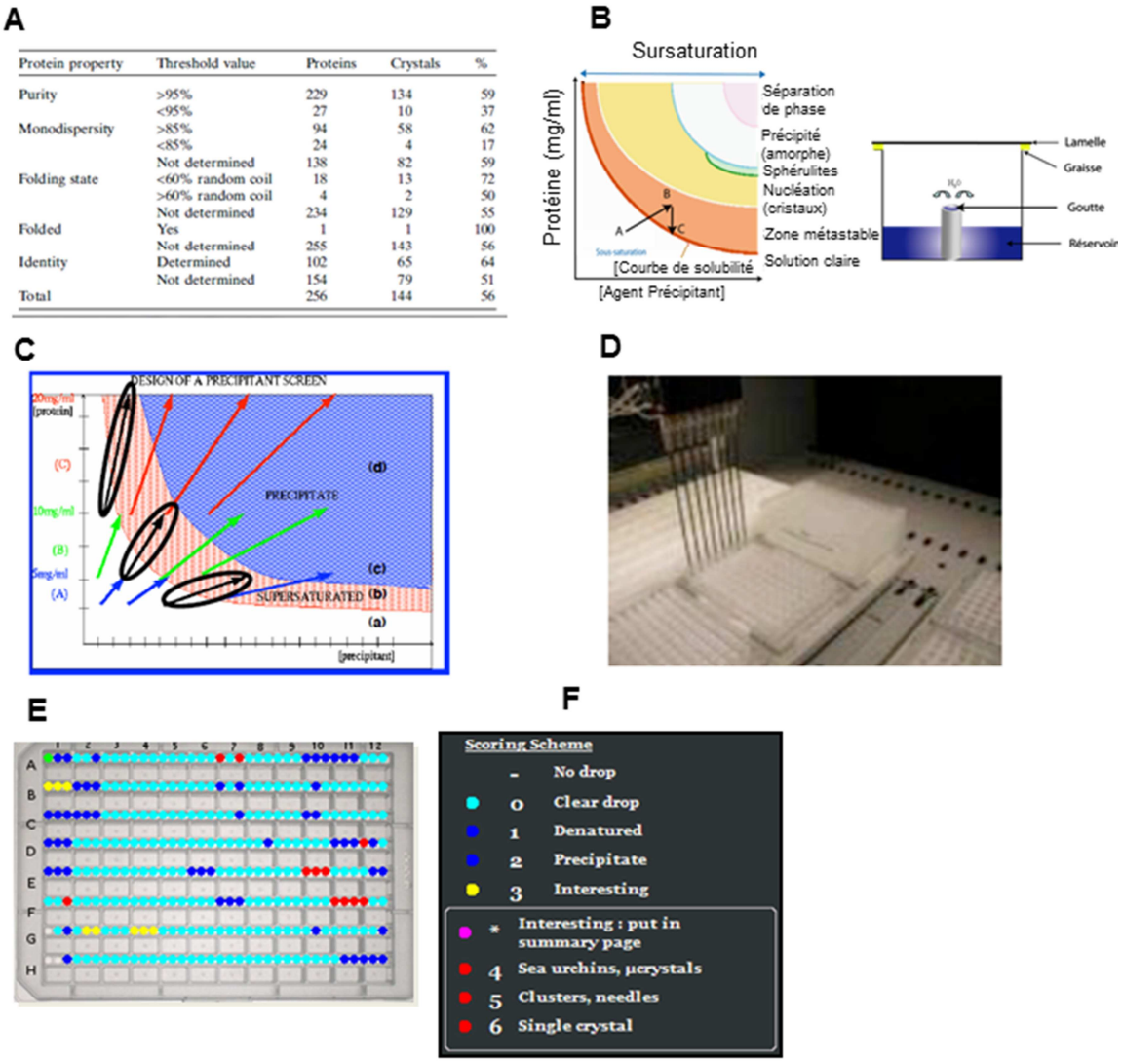


Figure 26 : Cristallisation des proteines

(A) Influence de différents paramètres sur la cristallisation d'une protéine (B) Diagramme de solubilité et méthode de diffusion de vapeur en goutte assise. (C) Crible matriciel. 3 concentrations en protéine sont testées avec 4 concentrations en agent précipitant. L'objectif est d'avoir un certain nombre d'essais dont le point final d'équilibre est dans la zone de nucléation ou sur-saturation (flèches noires)(D) Robot de cristallisation (E) Visualisation des scores d'une boite de 3 fois 96 essais avec le logiciel Crims (HTX, EMBL, Grenoble) (F) Score des gouttes de cristallisation.

Les protéines et complexes purifiées au cours de cette thèse ont été soumis à des essais de cristallisation à la fois au laboratoire et sur la plateforme de cristallisation haut-débit HTX (J Marquèz, EMBL, Grenoble) avec laquelle nous collaborons depuis la structure cristalline du complexe XRCC4(1-157)-Cernunnos(1-203) *(Ropars, 2011)* et celle du complexe du Mlh1-Pms1 de la voie de réparation MMR *(Gueneau, 2013)*. Les échantillons sont concentrés entre 5mg/ml et 20mg/ml selon leur solubilité. Pour les cribles sur HTX, les échantillons sont envoyés dans la glace en colis express et sont reçus à Grenoble sous 24h dans la glace. Nous effectuons classiquement 6 cribles de 96 conditions par la méthode de diffusion de vapeur. La goutte est réalisée en mélangeant 0.1µl de la solution de protéine et 0.1µl de la solution du réservoir. La goutte est donc deux fois plus diluée que le réservoir en agent de cristallisation. Comme le tout est fermé hermétiquement par un film, les composés volatiles de la goutte (essentiellement l'eau) vont s'évaporer jusqu'à équilibration de la pression de vapeur saturante. Au cours de cette évaporation, la concentration de la protéine et de l'agent précipitant vont progressivement augmenter et dans le cas favorable, des cristaux vont se former et grossir jusqu'à ce que la concentration en protéine atteigne la courbe de solubilité.

Nous réalisons en standard les cribles suivants : (1) Classics Suite (d'après Jancarik et Kim),(2) JCSG, (3) PACT (combinaison de PEG/ion et études d'effets des anions et cations à des pH donnés), (4) PEG-I (PEG-ions, PEGS + sels peu concentrés, 4 pH différents), (5) Salt Grid (Cribles en sels dont Ammonium Sulfate, Sodium Malonate, Sodium Formate) (6) Wizard I et II. Les gouttes de cristallisation sont réalisées à température ambiante avec en standard 3 concentrations en protéine par puits (par exemple 20mg/ml, 10mg/ml et 5mg/ml). Les 6 boites sont ensuite stockées dans un robot de visualisation Formulatrix à 4°C. Des prises d'images sont effectuées par le robot le lendemain (J+1) puis aux jours J+2, J+4, J+8, J+15, J+30 et J+60. Les images sont visualisées et annotées au laboratoire par internet à l'aide du logiciel CRIMS développé à Grenoble. Les gouttes sont annotées selon 4 codes couleurs et des notes de 1 à 8. Le logiciel permet de visualiser la progression des gouttes sur les différents jours.

Au laboratoire, nous effectuons également des gouttes par diffusion de vapeur. Les gouttes sont réalisées en goutte assise avec des volumes de 1,5µl de protéine et 1,5µl de réservoir à 20°C ou à 4°C. Nous réalisons des cribles complémentaires de ceux faits sur la plateforme HTX. Nous utilisons notamment un crible développé par un collègue du CEA de

Saclay, E. Stura *(Stura, 1994)*. Ce crible initialement conçu pour des fragments d'anticorps, permet de tester 12 conditions de cristallisation classiques avec pour chacune 4 concentrations en agents précipitant. Ce crible appelé matriciel permet d'identifier la zone située entre la région soluble et précipité de la protéine, zone où les chances d'apparitions de cristaux sont les plus fortes (**Figure 26C**). Ce crible a été décrit dans un projet de génomique structurale réalisé à l'AFMB (Marseille) sur une centaine de protéine d'*E. coli* comme étant le plus efficace parmi une série de cribles *(Sulzenbacher, 2002)*. Le laboratoire a également eu de nombreux succès avec ce crible matriciel *(Charbonnier, 2007, Gueneau, 2013)*.

7.2 Analyse des complexes par SAXS

Les protéines purifiées Ku70/80, Ligase4/XRCC4$_2$ et Cernunnos$_2$ sont mélangées de manière équimolaire en tampon S(20 mM Tris pH7.5, 75 mM KCl, 75 mM NaCl, 0.1 mM EDTA, 2 mM DTT, 20% glycérol) puis concentrées à l'aide d'un système d'ultrafiltration Amicon© 30K (Millipore) jusqu'à ce que la concentration du complexe soit de 5mg/mL. Le rendement de l'étape de concentration (masse de protéine avant concentration/ masse de protéine après concentration) est mesuré afin d'évaluer la quantité de protéine perdue sur la membrane d'ultrafiltration ou précipitée lors de cette étape. L'acquisition des données se fait sur la ligne de lumière SWING du synchrotron SOLEIL après un passage par une colonne de chromatographie d'exclusion de taille haute performance (colonne SEC3, Agilent Technologies™) préalablement équilibrée en tampon S, branchée sur un appareil HPLC Agilent propre à la ligne. Les expériences de filtration sur gel sont réalisées à l'avance au laboratoire sur la même colonne afin de mettre au point les conditions optimum lors de la session sur SWING. Le détecteur utilisé est une caméra PCCD170170 (AVIEX™) placée à 5m de l'échantillon.

7.3 Etudes de microscopie électronique

Au cours de la thèse, j'ai pu passer quelques jours au sein du laboratoire de E Le Cam (IGR, Villejuif) afin de me familiariser avec la préparation de grilles en coloration négative et

postive à l'acquisition de micrographies. J'ai notamment pu réaliser avec S Baconnais les analyses sur les protéines XRCC4 et Cernunnos entières seules ou en complexe.

Pour la préparation des grilles, nous utilisons des grilles de cuivre de 3mm de diamètre possédant un maillage de 300 de réseau (mesh) (84 microns entre deux nœuds). Ces grilles sont disposées face polie contre un film de collodion 2% polymérisé à la surface d'un bain d'eau. (Cette étape permet de créer un support fin et d'éviter que le film de carbone dont il est question par la suite ne se dépose que sur le maillage de la grille). Les grilles sont séchées sur du papier filtre sans cendres puis sont placées à 6 cm sous une tresse de carbone (BAL-TEC) qui est connectée aux électrodes d'un générateur (MED010 Balzers Union). Un vide de $3x10^{-4}$ millibars est appliqué et la tresse de carbone est dégazée tout en protégeant les grilles afin d'éliminer toutes les impuretés de la tresse. Le vide se dégrade et il faut attendre un vide de $5x10^{-5}$ pour retirer la protection des grilles et appliquer un courant jusqu'à incandescence de la tresse de carbone. Ce courant est maintenu jusqu'à ce qu'un film de carbone recouvre les grilles (~10nm). Enfin, les grilles seront effluvées dans la demi-heure avant le dépôt d'échantillon par ionisation de l'air (2-5mA) sous vide ($7x10^{-1}$ millibars) pendant 40 à 60 secondes.

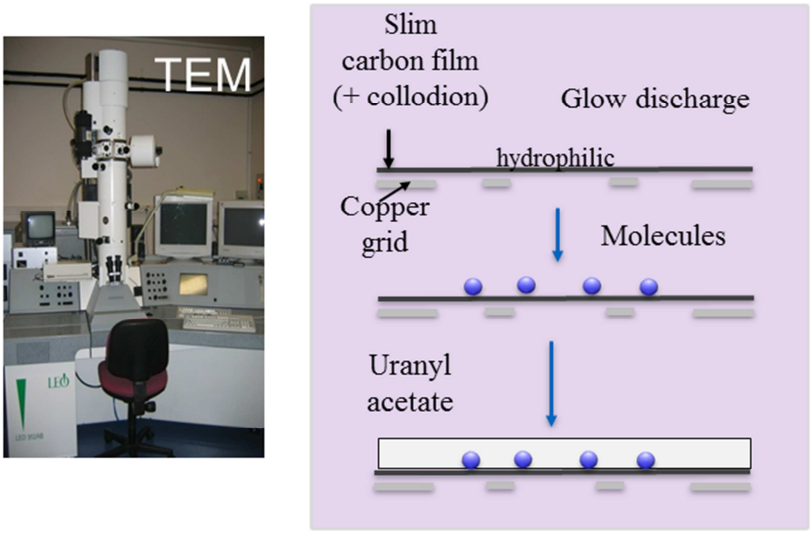


Figure 27 : Analyses par microscopie électronique
(Gauche) Microscope électronique (Zeiss 912AB) utilisé à l'IGR (Villejuif) (Droite) Préparation d'une grille de microscopie électronique.

L'étape suivante consiste à déposer 5µL d'échantillon à une concentration de 1,5µM pendant 30 secondes sur une grille préparée selon le protocole décrit ci-dessus. Les échantillons utilisés dans notre cas sont soit XRCC4 entier, soit Cernunnos entier soit le

mélange XRCC4-Cernunnos 1:1, à une concentration de 15µM. La grille, où les protéines se sont absorbées, est alors inclinée à 45° et une goutte d'acétate d'uranyle 2% masse/volume est utilisée pour laver la grille. Après 30 secondes, les grilles sont séchées de sorte que l'acétate d'uranyl enrobe les protéines dans un fin film de contrastant. Elles et sont prêtes à être analysées par microscopie électronique à transmission.

Les grilles sont insérées dans la colonne du microscope maintenue sous vide (**Figure 27**). Le faisceau est alimenté par un filament LaB_6 (mini-cristal d'hexaborure de lanthane). Les premiers réglages de l'appareil consistent à centrer le faisceau, à régler l'astigmatisme jusqu'à l'obtention d'une projection ronde, à contrôler la présence d'anneaux de diffractions provoqués par la présence du film de carbone, et à ne sélectionner que le pic sans perte (i.e les électrons qui ont conservé leur énérgie initiale après transmission au travers de l'échantillon). Les micrographies de microscopie électronique à transmission présentées ont été réalisées à partir d'une caméra CCD ProScan 1024 HSC après un réglage de la focalisation permettant une acquisition équilibrée entre résolution et contraste. Les clichés ont été pris avec des grandissements compris entre x10 000 et x63 000.

RESULTATS

1. Expression des protéines et complexes du NHEJ en cellules d'insecte

1.1. Stratégie de clonage

L'objectif général du laboratoire est de reconstituer progressivement le ou les complexes participant au processus NHEJ. Pour cela, nous avons décidé d'utiliser la technique MultiBac mis au point par Imre BERGER (EMBL, Grenoble). Un des avantages de ce système est de permettre l'addition itérative de gènes dans un même vecteur. Pour tirer profit de cet avantage, nous avons défini une stratégie de clonage des différents partenaires du NHEJ telle qu'illustrée dans la **Figure 28**. L'idée était de cloner dans un même vecteur les complexes correspondant aux trois étapes du NHEJ : le complexe Ku70/80 pour la reconnaissance des extrémités des cassures double-brin, le complexe DNA-PKcs/Artemis, seul ou avec Ku70/Ku80, pour la préparation des extrémités et le complexe Lig4/XRCC4/Cernunnos pour la ligature des extrémités. L'objectif ultime étant de préparer un vecteur avec les 7 sous-unités étudiées.

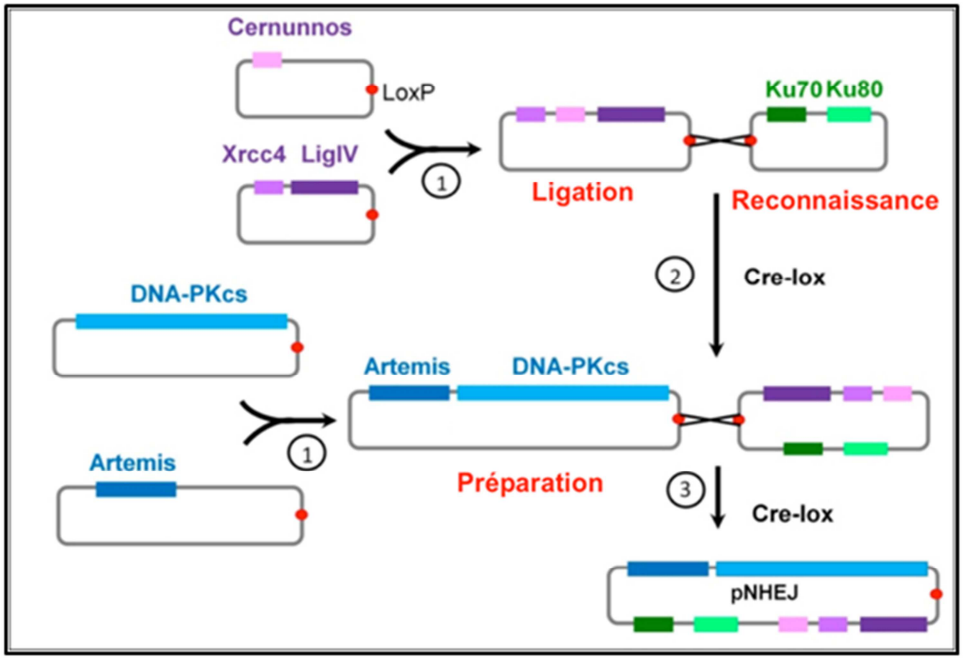


Figure 28 : Stratégie de clonage des protéines du NHEJ.
Les complexes protéiques correspondant à chacune des étapes de la voie NHEJ cœur (reconnaissance, préparation des extrémités, ligature) sont clonés ensemble sur un même vecteur. Les vecteurs contenant des complexes protéiques peuvent eux-mêmes être combinés en utilisant les sites Cre-lox.

Dans un premier temps, le complexe Ligase4/XRCC4 a été cloné dans le vecteur accepteur pKL en ajoutant une séquence de 10 Histidines (10His) en C-ter de XRCC4, cela afin de permettre la synthèse du complexe seul et de préparer le vecteur à une addition ultérieure. Le vecteur pKL_XL4 a ainsi été obtenu. Parallèlement, Cernunnos a été cloné dans un vecteur accepteur différent (pFL) mais sans étiquette, afin de permettre la production de la protéine seule et de prévoir la production/purification du complexe Lig4-X4-Cernunnos en utilisant l'étiquette 10His de XRCC4. Ce second vecteur a été appelé pFL_CNN.

Dans le même état d'esprit, la séquence sauvage de l'hétéroduplexe Ku70/80 a été cloné dans le vecteur donneur PUCDM sans étiquette dans le but de réaliser ultérieurement une fusion avec le vecteur accepteur Lig4-X4-Cernunnos à l'aide de la CreLox. Toutefois, pour permettre la production de Ku70/80 seul, une fusion CreLox a été réalisée avec le vecteur pK-dummy donnant un vecteur appelé pKL_Ku. Par la suite il est apparu que cette construction ne donnait pas des résultats satisfaisants et c'est un vecteur pFL contenant une séquence synthétique optimisée de Ku70/80 (Genscript) avec une étiquette 10His en position Nter de Ku80 qui a été utilisé en production. Enfin les gènes des protéines Artemis et DNA-PKcs ont été obtenus par voie synthétique et cloné dans des vecteurs accepteurs pFL dans un premier temps afin de tester leur capacité de production. Le gène DNA-PKcs est un gène particulier à manipuler car il fait 12 384 paires de bases.

1.2 Prise en main de la technique

Suite à la mise en place de la collaboration entre notre équipe et celle d'Imre Berger à l'EMBL, j'ai suivi une formation pour la culture et l'expression de protéines recombinantes en cellules d'insecte sur le site de l'EMBL de Grenoble. Muni des plasmides pKL_XL4 et pFL_CNN, j'ai pu me familiariser avec P Drevet à l'ensemble des étapes de la production en cellules d'insecte. Ainsi après transformation dans les cellules EMBacY, Nous avons obtenu environ 50% de clones blancs mettant en évidence l'insertion de nos plasmides dans le bacmide présent dans la bactérie (**Figure 24**). Nous avons préparé du bacmide à partir de plusieurs clones différents puis nous avons réalisé une transfection dans les cellules Sf21. Nous avons ainsi obtenu une première génération de baculovirus appelé CNN pour l'expression de Cernunnos et X4L4 pour la production du complexe Ligase4-XRCC4. Nous

avons ensuite amplifié cette première génération de virus pour obtenir une deuxième génération, V1, qui nous a servi de stock viral afin de production.

Nous avons alors procédé à la réalisation de productions d'essais en 50ml en utilisant les 2 baculovirus CNN et les 4 baculovirus X4L4 comme source d'infection. Nous avons suivi la production de YFP par mesure de la fluorescence présente dans les extraits de cellules. Sur la **Figure 29A**, on peut voir un graphique typique de l'évolution de la fluorescence dans les extraits cellulaire. La fluorescence apparaît 24h après l'arrêt de croissance des cellules (Day after Proliferation Arrest = DPA) et continue d'augmenter pendant 3 jours puis marque un plateau. Les 6 cultures ont montré des profils similaires. Lorsque la production de YFP atteint une phase plateau les cellules sont récoltées par centrifugation pour analyse.

Les cellules ont été soumises à un traitement aux ultrasons puis nous avons séparé la fraction soluble par centrifugation. Nous avons ensuite réalisé un gel SDS-PAGE à partir des extraits cellulaires afin de vérifier la production de mes protéines. La **Figure 29B** indique une très bonne production de chacune des protéines et sous forme essentiellement soluble. J'ai pu confirmer la présence de protéine XRCC4 par immunorévélation grâce à la séquence 10His placée en C-ter de XRCC4 (**Figure 29C**). On remarque que la protéine XRCC4 migre vers 50KDa alors que son poids moléculaire théorique est de 39,6KDa. Cette aberration de migration a déjà été constatée dans la littérature *(Recuero-Checa, 2009)* et serait imputable à la forme allongée de la protéine et la présence d'une grande région prédite comme non-structurée (175 acides aminés).

Nous avons ensuite procédé à l'EMBL à une purification à partir d'un culot de cellules ayant produit X4L4 à l'aide d'une colonne NiNTA sépharose. La **Figure 29D** montre le résultat de cette première purification. On peut observer sur le gel, la bande correspondant à la protéine XRCC4 et une bande dont le poids moléculaire correspond à la Ligase 4. XRCC4 étant la seule protéine à posséder une séquence poly histidine c'est bien le complexe formé que nous avons purifié de cette façon.

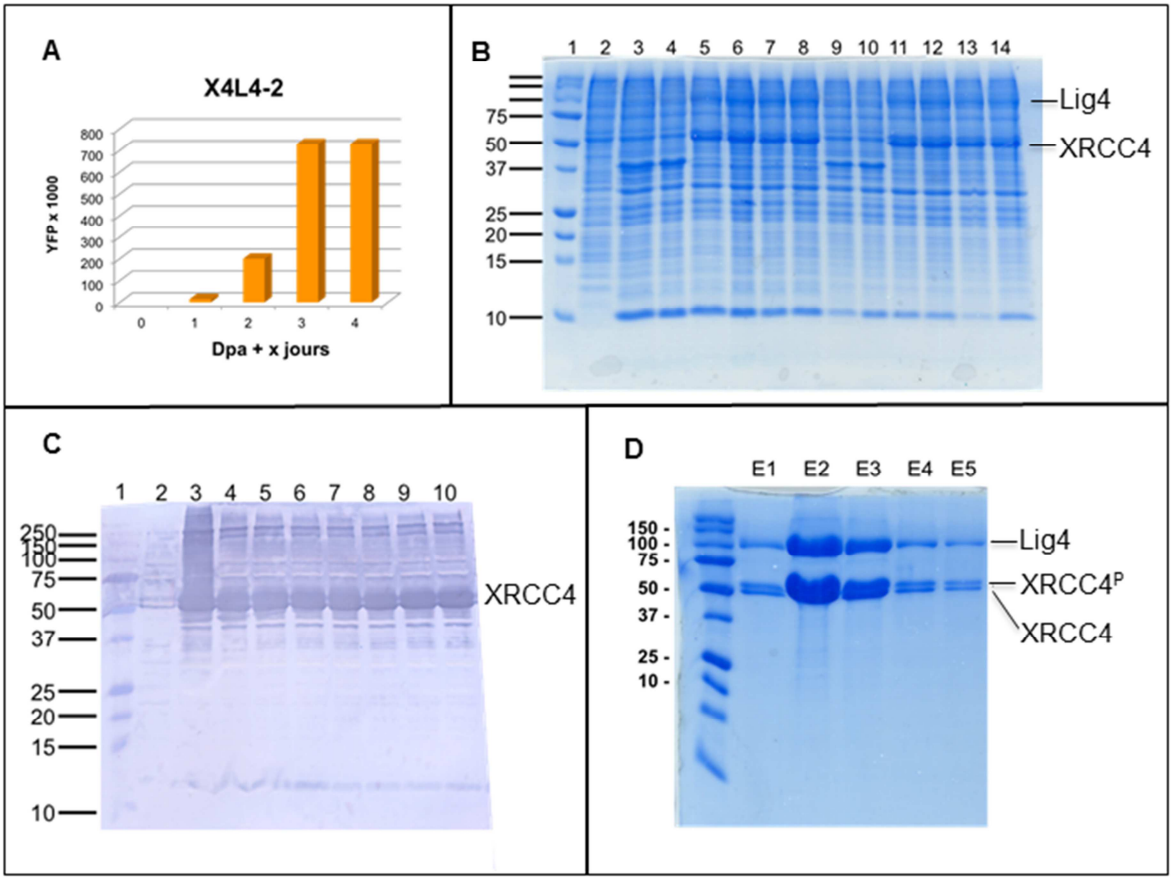


Figure 29 : Tests de production sur la Plateforme MultiBac à l'EMBL-Grenoble du complexe Ligase4-XRCC4 et Cernunnos en cellules Sf21.

(A) Profil de production de YFP par des cellules infectées. Mesure de la fluorescence présente dans 1x10^6 cellules après l'arrêt de croissance (dpa) (B) Gel SDS-PAGE des extraits totaux de cellules (piste 2 à 8) et des fractions solubles (9 à 14). Cellules non infectées : piste 2. Cernunnos clone 1 et 2 : piste 3, 4, 9 et 10. Lig4X4 clone 1, 2, 3 et 4 : pistes 5 à 8 et 11 à 14 (C) Immunorévélation à l'aide d'anticorps anti-polyHistidine sur des extraits totaux (pistes 3 à 6) et des extraits solubles (pistes 7 à 10) des clones Lig4X4 1 à 4. (D) Purification du complexe Lig4/XRCC4 sur colonne NI-NTA sépharose.

2. Purification des protéines et complexes de la voie NHEJ

L'expertise acquise à Grenoble nous a permis d'implémenter la technique au CEA de Saclay et de l'appliquer à la production de plusieurs protéines de la voie NHEJ que ce soit

sous forme de protéine isolée, comme dans le cas de Cernunnos et Artemis ou sous forme de protéines co-exprimées comme pour Ligase4-XRCC4 ou pour l'hétérodimère Ku70/80. Les résultats présentés correspondent à un exemple représentatif de purification réalisé au laboratoire. Les exemples décrits correspondent à des protocoles optimisés après plusieurs essais de purification.

2.1 Purifications de Ligase4-XRCC4

Les culots de Ligase4-XRCC4 (L4X4) sont purifiés au départ par culot de 200ml puis lorsque le protocole est bien établi, nous purifions le complexe en prenant des culots de 800ml de cultures de cellules d'insecte. Le choix du volume de cultures purifiées est également fait en fonction du niveau final de fluorescence (YFP) mesuré en fin de culture et de la densité de cellules atteintes.

Après lyse par sonication et centrifugation, le surnageant est mis en contact avec de la résine Ni-NTA, 2h en vrac avec agitation douce en chambre froide. Nous avons également testé la purification par gravité en chambre froide. Ces deux procédés se sont avérés efficaces et plus intéressants pour ce complexe qu'un passage sur une colonne pré-coulée de Nickel (type Hi-Trap 5ml, GE-Healthcare). Le gel SDS-PAGE correspondant à cette étape montre que le dernier lavage ne contient plus de protéines et que l'élution se fait sur environ 45ml pour 8ml de résines (**Figure 30A**). Sur les fractions moins chargées, on note l'existence de deux bandes à la hauteur attendue pour la Ligase4 dont une bande supérieure plus intense. On observe également deux bandes d'intensité équivalente vers 50kDa. Les analyses par Western Blot indiquent qu'il s'agit de XRCC4 (39kDa) dans les deux cas.

Il y a également une bande vers 75kDa entre les bandes attribuées à la Ligase4 et à XRCC4. L'ensemble des fractions est regroupée en un lot unique et dialysé contre un tampon (25mM Tris pH8, 100mM KCl, 10mM 2-Me-EtOH). La mesure du spectre UV de cet échantillon, réalisée au Nanodrop, indique une concentration en protéine de 0,75mg/ml pour un volume de 47ml soit 33mg de protéine. Le surnageant des 800ml de cultures est passé sur deux résines en parallèle. Avec la seconde résine, sur laquelle est passé un plus petit volume de surnageant, nous obtenons un gel similaire et 5,2mg de protéine. Nous avons donc à cette première étape un rendement de 38,2mg de protéine pour 800ml de cellules d'insecte.

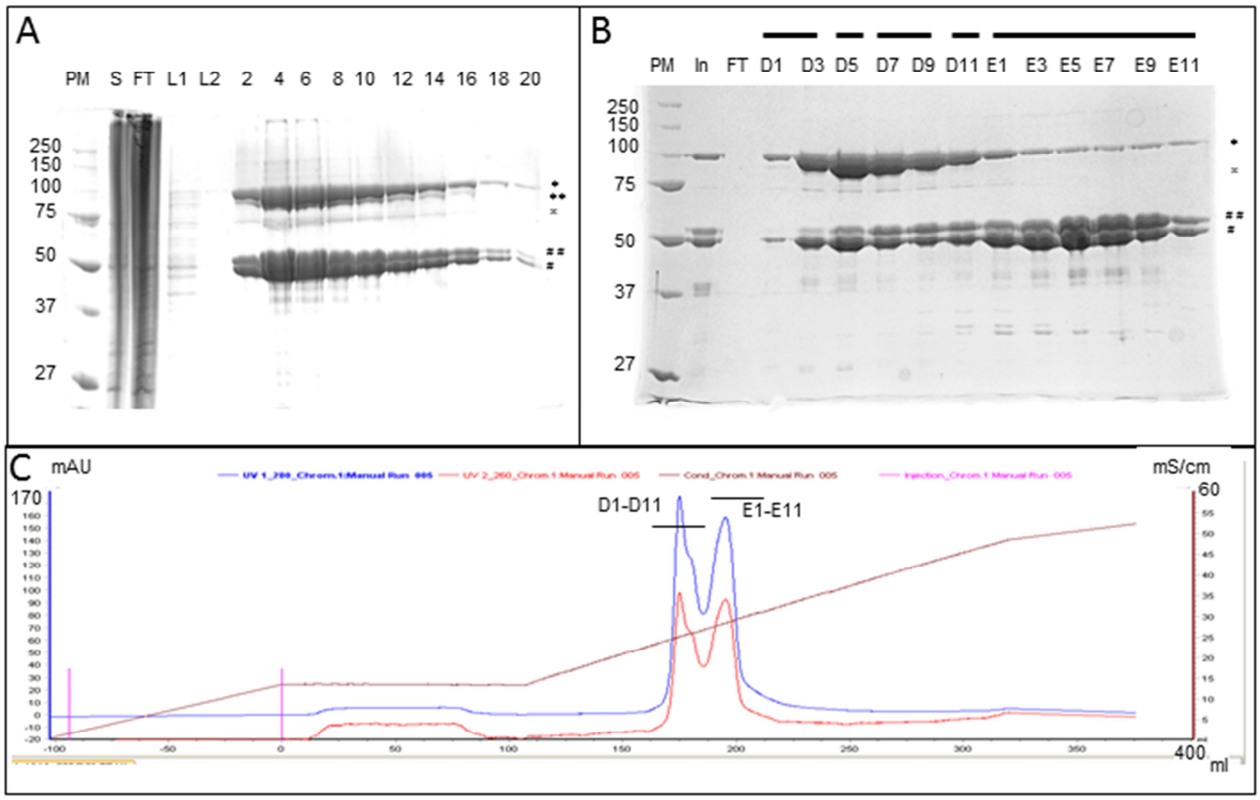


Figure 30 Purification du complexe Ligase4-XRCC4:
(A*) Gel SDS-PAGE 10% de la 1er étape de purification du complexe Ligase4-XRCC4 sur colonne de Nickel. Légende : (* et **) deux bandes attribuées à la Ligase4 ; (# et ##) deux bandes attribués à XRCC4 ; ° bande entre XRCC4 et la Ligase4. (B) Chromatogramme et (C) Gel SDS Page 10% de l'étape de polissage du complexe Lig4-XRCC4 sur Resource Q 6ml.*

L'échantillon de L4X4 dialysé est passé sur une ResourceQ pour fixer le complexe en utilisant les propriétés acides de XRCC4 (pI 5.3). Le chromatogramme et le gel SDS-PAGE correspondant sont présentés **Figures 30B et 30C**. Nous pouvons remarquer que sur ce gel, nous n'avons plus qu'une bande pour la Ligase4 sur la fraction chargée (In). La fraction non-retenue présentre une absorbance faible et supérieure à 260nm sur le chromatogramme. L'échantillon est élué sur deux pics principaux. Un premier pic avec une conductimétrie de ~25mS/cm (soit ~300mM sels d'après la conductimétrie de 100% de tampon QB). Ce premier pic correspond au complexe Ligase4-XRCC4. Un second pic avec une conductimétrie légèrement supérieure, de ~28mS/cm. Ce second pic correspond à la protéine XRCC4. XRCC4 est toujours présent sous la forme de deux bandes. Le complexe L4X4 n'est pas dissocié par le gradient de sel. Ceci est en bon accord avec l'interaction forte décrite entre XRCC4 et la Ligase4 *(Grawunder, 1997)*. Les fractions sont regroupées en fonction du gel (voir les traits au-dessus des numéros de fractions) (**Figure 30B**). Le spectre UV des 5 lots ainsi constitués est mesuré sur Nanodrop et ces lots sont analysés sur gel SDS-PAGE. Les lots D5-D6 et D7-D9

présentent une stœchiométrie en Ligase4 et XRCC4 proche de 1:2. Nous mesurons 7,9mg de protéine sur ces deux lots ce qui correspond à notre rendement final de concentration (~10mg/L cellules d'insecte). On obtient également 12,9mg de XRCC4 légèrement contaminé par de la Ligase4.

Le protocole actuel peut probablement être encore amélioré. (i) Les cellules d'insecte sont beaucoup plus fragiles que les bactéries comme *E. coli*. Notre méthode de sonication est probablement un peu forte. Une optimisation consisterait à diminuer le temps et la puissance de sonication. Il serait intéressant de voir si cela réduit la fraction de protéines dans la fraction insoluble après centrifugation du lysat. Nous avons observé lors de certaines purifications qu'une fraction significative de L4X4 se retrouve dans la fraction insoluble après centrifugation du lysat (données non montrées). (ii) Une seconde étape de polissage pourrait être introduit. Plusieurs possibilités s'offrent à nous. Une étape de chromatographie par exclusion de taille ou une chromatographie échangeuse de cations pour tirer avantage de la nature relativement basique de la Ligase4 (pI 8.2).

2.2 Purifications de Ku70-Ku80

Nous avons travaillé au laboratoire sur deux constructions du complexe Ku70/Ku80. Dans l'objectif de purifier le complexe Ku70/Ku80/Ligase4/XRCC4, la première stratégie choisie a été d'utiliser deux vecteurs, un vecteur Ligase4/XRCC4 avec une étiquette 10His en C-ter de XRCC4 et aucune sur la Ligase4 (vecteur pKL-X4L4) et un second vecteur Ku70/Ku80 (vecteur pKdummy-Ku70/80) sans étiquette sur aucune des deux sous unités. L'idée était de co-purifier ou de co-exprimer les 4 protéines en attrapant par XRCC4 le complexe (L4-$X4_2$-Ku70-Ku80). Les premiers essais de purification de Ku70/Ku80 seul ont été réalisés avec ce vecteur. Malgré différentes production en cultures d'insectes et différentes stratégies de purification inspirées notamment de la littérature *(DeFazio, 2002)*, nous n'avons pas réussi à purifier Ku70/Ku80 de façon satisfaisante. Nous avons cependant mis en évidence la production du complexe par immuno-révélation du contenu cellulaire à l'aide d'anticorps spécifiques anti-Ku70 et anti-Ku80 fournis par P. Calsou. Nous avons cependant constaté un assez faible rendement de production mais surtout une mauvaise stœchiométrie dans la formation du complexe dans la mesure où le monomère Ku70 semblait plus abondant que Ku80. Il est aussi envisageable que ce problème de stoechiométrie apparent sur SDS-Page

soit en fait imputable à la migration abbérante de Ku80 sur SDS-Page dans des conditions de dénaturation classiques comme cité plus haut.

Nous avons alors choisi de construire un second vecteur nous permettant de produire le complexe Ku70/Ku80 en grande quantité et de le purifier indépendamment du complexe Lig4-XRCC4. Nous avons fait produire par Genscript des gènes synthétiques de Ku70 et Ku80. Ces gènes synthétiques ont alors été insérés au laboratoire dans un vecteur baptisé pFL_Ku70/80opt avec une étiquette 10His-TEV en N-ter de Ku80 et aucune étiquette sur Ku70. Les culots de Ku70/Ku80 sont purifiés par culot de 400ml. Le protocole de purification suivi pour Ku70/Ku80 est proche de celui suivi pour XRCC4/Ligase4. Il se déroule en deux étapes une étape de purification sur colonne de Nickel où le complexe est retenu par Ku80 suivie d'une étape de polissage sur échangeuse d'anions (Resource Q). Les différences principales sont la présence de 500mM de sels dans le tampon de lyse (75mM KCl, 425mM NaCl). Ku présente une forte affinité pour l'ADN y compris des petits fragments de 20-40bp. Lors des premières purifications, nous avons pu observer une faible rétention du complexe sur la colonne de Nickel et la présence simultanée d'une forte absorbance à 260nm de nos échantillons. L'ajout de sels, 1M dans les premiers protocoles puis 500mM dans les protocoles finaux, permet de recouvrer une bonne fixation sur la colonne de Nickel, probablement par dissociation de l'ADN fixé en grande quantité.

Nous avons dû également légèrement adapter le protocole de gels SDS-PAGE pour le complexe Ku70/Ku80. Lors des premières purifications, seule la bande de Ku70 était résolue sur le gel avec une trainée présente au-dessus. En dénaturant plus longtemps, l'échantillon chargé et en déposant moins de matériel, la bande de Ku80 est à nouveau visible (**Figure 31A**). Un western blot réalisé à cette étape indique que la bande supérieure correspond à la protéine Ku80 en accord avec la hauteur de migration observée (anticorps anti-His couplé à la peroxydase, Sigma Aldrich). Nous observons à cette étape un contaminant qui migre vers 50kDa. Les fractions 2 à 25 sont regroupées, dialysées contre un tampon 25mM Tris pH8, 100mM KCl, 10 mM β-mercaptoethanol et un spectre UV de ce lot est réalisé sur un Nanodrop. Ce lot présente une contribution importante à 260nm du fait de la présence de fragments d'ADN complexés à Ku70/Ku80. Nous évaluons la quantité de matériel par test de Bradford et obtenons une concentration de 0,1mg/ml pour un volume de 45ml, soit 4,5mg pour 400ml de culture de cellules d'insecte à cette première étape.

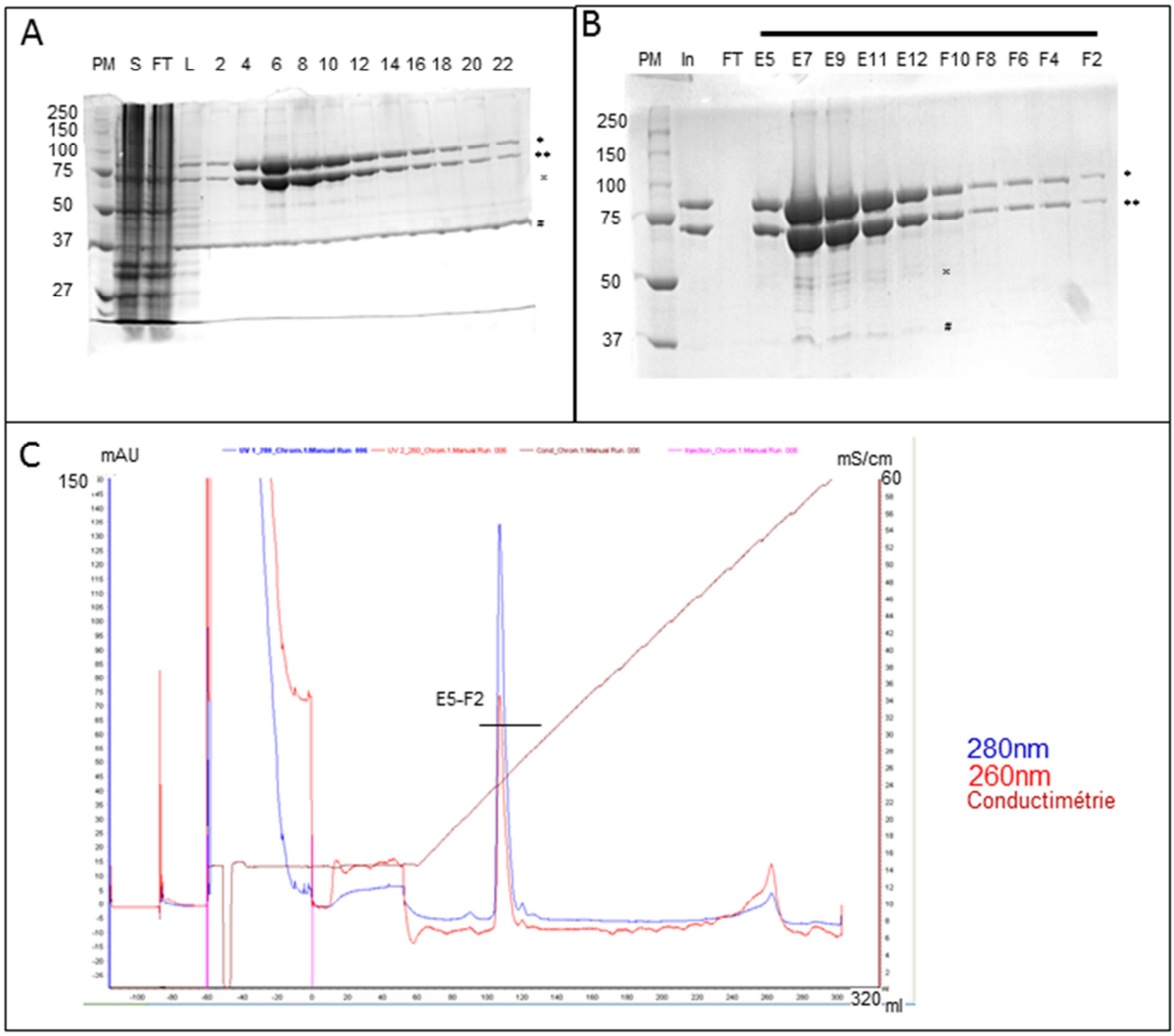


Figure 31 : Purification de Ku70/Ku80

(A) Gel SDS-PAGE 10% de la 1er étape de purification du complexe Ku70/Ku80 sur colonne de Nickel. Légende : (*) Ku80 et (**) Ku70; (x) contaminant ; (#) front de migration du bleu ; (C) Chromatogramme et (B) Gel SDS Page 10% de l'étape de polissage sur Resource Q 6ml. (x) et (#) deux contaminants ; sur le chromatogramme, l'injection de l'échantillon Ku70/Ku80 démarre à la seconde marque rose.

L'échantillon de Ku70/Ku80 purifié sur colonne de Nickel et dialysé, est injecté sur une colonne Resource Q à pH8 afin de fixer le complexe Ku70/Ku80 dont les sous-unités ont un pI légèrement acide, de respectivement 5,8 et 6,2. Nous observons une légère absorbance dans la fraction non-retenue et un pic principal dans le gradient à une conductimétrie de ~24mS/cm (soit ~280mM en sels, mélange KCl-NaCl). Le gel SDS-PAGE correspondant à cette étape indique que le complexe Ku70/Ku80 est élué autour de ce pic **(Figure 31B et 31C)**. Les deux sous-unités semblent présentes avec la bonne stœchiométrie 1:1. On peut observer que les fractions correspondant au sommet du pic, E7-E9, sont très

concentrées. Ceci montre que la chromatographie d'échange d'anions couplée à une dialyse peut être utilisée efficacement comme étape de concentration de Ku70/Ku80 avant la cristallogenèse. Sur les fractions les plus concentrées, on observe deux bandes faibles autour de 50kDa. Une analyse par spectrométrie de masse devrait permettre par la suite au laboratoire d'évaluer s'il s'agit de contaminants ou de bandes de dégradation du complexe Ku70/Ku80. Dans ce dernier cas, il sera utile de prévoir une étape de polissage additionnelle pour ôter ces fragments de dégradation qui peuvent gêner la cristallisation en bloquant la croissance cristalline. Un pic est observé en fin de gradient avec une absorbance plus forte à 260nm qu'à 280nm, qui correspond vraisemblablement à de l'ADN co-purifié lors de la première étape sur colonne de Nickel. En effet, les spectres UV réalisés sur les fractions de Ku70/Ku80 après la ResourceQ ne montre plus d'absorbance à 260nm contrairement à la fraction chargée sur ResourceQ. Les fractions E4 à F2 sont regroupées. Le spectre UV de ce lot indique une concentration de 0,34mg/ml pour un volume de 17ml soit 5,8mg de protéine pour 400ml de culture de cellules d'insecte. Le rendement de purification à l'issue de la ResourceQ est donc d'environ 14,5mg/L de cultures de cellules d'insecte.

2.3 Purification de Cernunnos

La protéine Cernunnos est exprimée en cellules d'insecte sf21 sans étiquette, en vue d'une co-purification avec le complexe XRCC4-Cernunnos. Nous utilisons 4 culots de 50ml de cultures d'insectes sf21 avec un nombre total de cellules de $532x10^6$. Ces cellules sont reprises au $1/8^{ème}$ dans le tampon de lyse et soniquées. La première étape de purification est une étape de chromatographie sur échangeuse d'anions de type Hi-TrapQ (Ge-Healthcare volume 5ml) pour profiter des propriétés acides de Cernunnos (pI de 5,6). La protéine est éluée dans un gradient de 0% à 50% d'un tampon 20mM Tris pH8 contenant 150mM KCl/850mM NaCl sur 10 volumes de colonne (50ml). Les fractions sont analysées sur gel SDS-PAGE et regroupées en fonction de la présence de Cernunnos (PM~ 37kDa) (**Figure 32A**). On peut remarquer que sur la fraction de surnageant de lysat déposée, Cernunnos est bien sur-exprimée. L'éluat principal présente un contaminant vers 30kDa, qui peut correspondre à un contaminant ou à une forme dégradée de Cernunnos, probablement dans sa région C-terminale prédite comme non-structurée. Le lot de Cernunnos obtenu est dialysé contre un tampon contenant les sels suivants 50mM NaCl/50mM KCl. Cet échantillon est chargé sur

une colonne ResourceQ 6ml et un gradient de 0% à 50% sur 10 volumes de colonnes (60ml) du tampon 20mM Tris pH8, 150mM KCl/850mM NaCl. On obtient un pic principal qui est élué assez rapidement avec une conductimétrie de 14 mS/cm. Ce pic principal est entouré de deux pics plus petits. Le gel montre que cette technique permet d'ôter un contaminant vers 43kDa. En revanche, elle ne permet généralement pas d'obtenir un échantillon de pureté satisfaisante et l'échantillon présente souvent de nombreux contaminants dont une protéine vers 30kDa. Nous obtenons en fin de purification 5,2mg de Cernunnos à 1,04mg/ml avec un spectre UV présentant un ratio 260/280 de 0,63 (soit un rendement de 26mg de Cernunnos par litre de culture de cellules d'insecte).

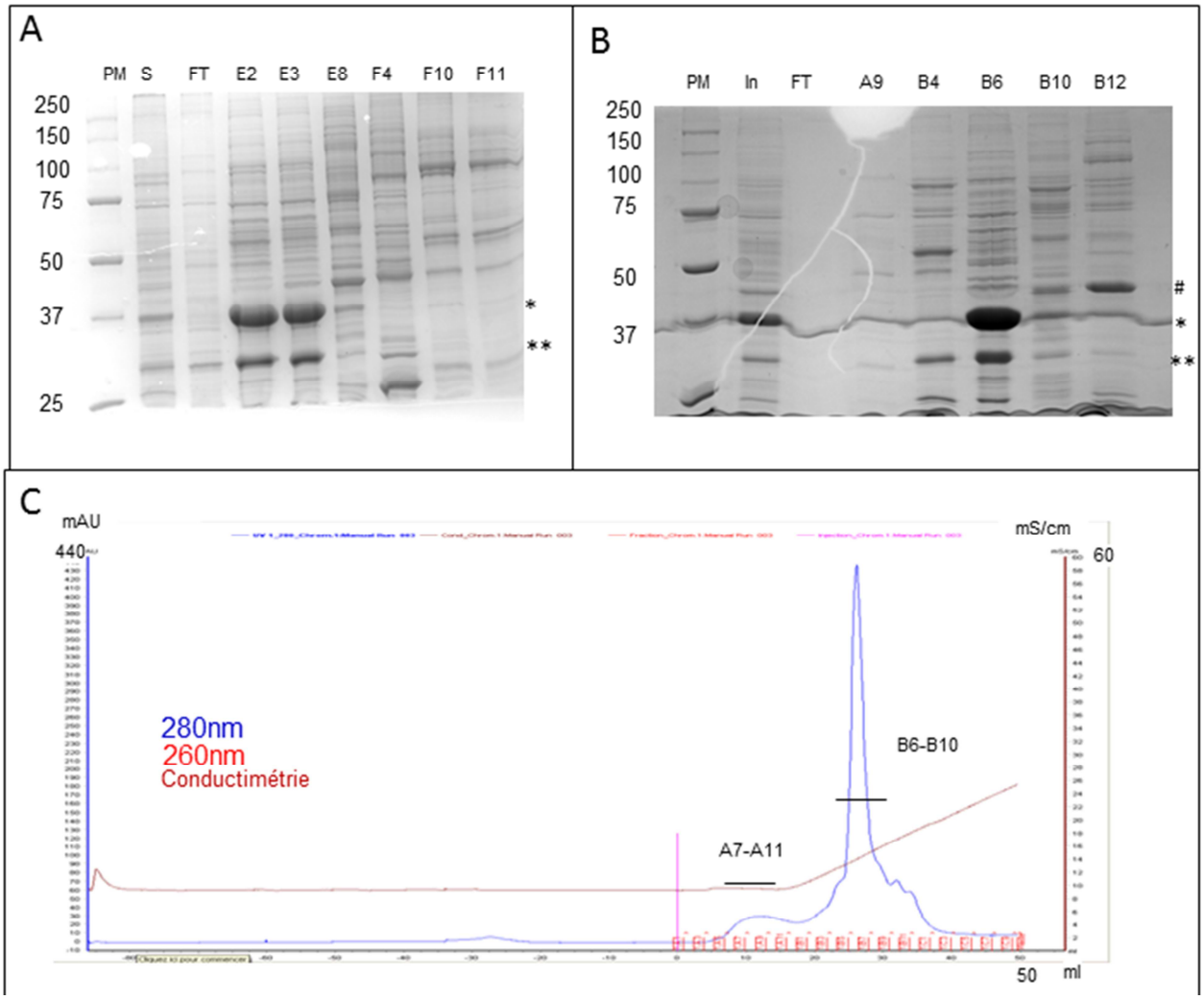


Figure 32 Purification Cernunnos.

(A*) Gel SDS-PAGE 10% de la 1er étape de purification de Cernunnos sur colonne de Nickel. Légende : (*) Cernunnos ; (**) contaminant vers ~30Kda; (C) Chromatogramme et (B) Gel SDS Page 10% de l'étape de purification sur Resource Q 6ml ; (#) contaminant vers ~43Kda*

Cette purification a été optimisée par la suite au laboratoire car les deux étapes utilisées utilisent le principe d'échange d'anions et donc ne correspondent pas réellement à deux étapes de purification différentes. La seconde étape sur échangeuse d'anions sur ResourceQ a été remplacée par une étape d'échangeuse de cations à pH8 sur ResourceS de 0 à 40% du même tampon (**Annexe 6**).

2.4 Purifications d'Artemis

La protéine Cernunnos est exprimée en cellules d'insecte sf21 avec une étiquette 10His en Cter. Nous utilisons 4 culots de 50ml. Ces cellules sont reprises au 1/8ème dans le tampon de lyse et soniquées. La première étape de purification est une chromatographie d'affinité sur colonne de Nickel. Le gel SDS-PAGE (**Figure 33a**) indique qu'une protéine vers 100kDa est bien sur-exprimée. Un western blot réalisé avec des anticorps dirigé contre l'étiquette 10His nous montre que cette protéine est bien Artemis. Les fractions éluées en dernier sont très enrichies en Artemis. Les fractions sont regroupées en fonction du gel SDS-PAGE puis dialysées contre un tampon contenant 100mM KCl et chargées sur une colonne échangeuses d'anions Resource Q à pH8 (Artemis a un pI de 6). On obtient un pic principal lors du gradient en sels avec une conductimétrie de 14 mS/cm (donnée non montrée). Le gel SDS-PAGE montre que la protéine ainsi purifiée présente un degré de pureté satisfaisant. Nous obtenons à l'issu de la première étape, un lot de 3,2mg d'Artemis pour 200 ml de cultures (soir un rendement à cette étape d'environ 16mg d'Artemis par litre de culture). Une fraction seulement a été passée lors de cette purification sur ResourceQ ne permettant pas de définir à ce stade un rendement global de cette purification. L'essai réalisé sur une fraction suggère cependant que le rendement de l'étape de Resource Q est bon et que cette étape pourra être appliquée à l'avenir à l'ensemble du lot issu de colonne de Nickel. Les premières tentatives de concentration d'Artemis ont révélé une solubilité faible de l'enzyme. Les travaux à venir au laboratoire viseront à améliorer cette solubilité en ajoutant des protéines partenaires d'Artemis ou de l'ADN.

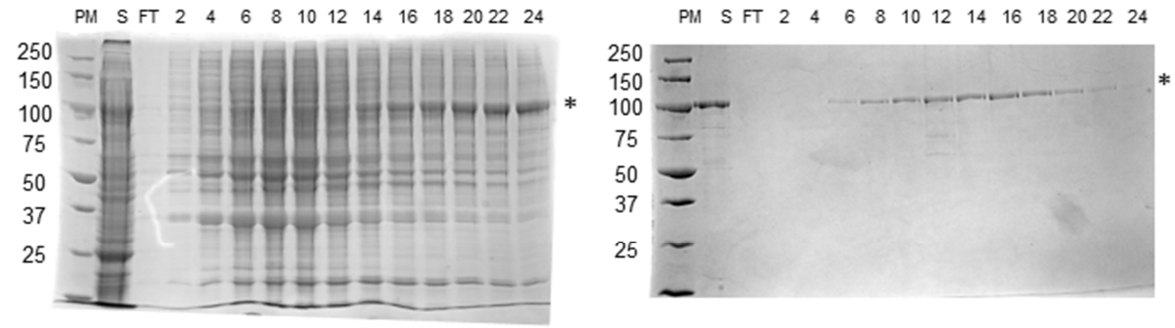


Figure 33a Purification d'Artemis.
*(**Gauche**) Gel SDS-PAGE de l'étape de chromatographie par affinité sur colonne de Nickel, * protéine Artemis. Une dizaine de fractions de 25 à 34, présentes sur un second gel montre des fractions très enrichies en Artemis, comme la fraction 24 (données non montrées) (**Droite**) Gel SDS-PAGE de l'étape d'échangeuse d'anions Resource Q*

Les travaux décrits dans la littérature ces dernières années sur Artemis et les essais de productions réalisés précédemment au laboratoire ont souligné les difficultés liées à la production de mg ou même de centaines de µg d'Artemis purifiée. La protéine a été purifiée à l'échelle de quelques µg pour des essais enzymatiques en exprimant la protéine en cellules 293T *(Ma, 2002)*. Artemis n'a pas été cristallisée (hormi un peptide de 11 acides aminés, région 485-495, en complexe avec le domaine NTase de la Ligase4) *(Ochi, 2013)*. On ne trouve également aucune étude biochimique ou biophysique nécessitant quelques mg de l'enzyme.

Des études de production ont été commandées par JP de Villartay auprès de Protein Expert (devenue PX'Therapeutics (Grenoble), http://www.px-therapeutics.com). Cette compagnie effectue entre autre, la production de protéines recombinantes à demande dans *E. coli* ou en cellules d'insectes. Le rapport établi par cette compagnie indique que les essais de production dans *E.coli* avec ou sans chaperons co-exprimés ne permettent pas d'obtenir de façon satisfaisante de la protéine purifiée. Les essais de production par Protein Expert d'Artemis exprimée en corps d'inclusion et renaturée sont également négatifs. Enfin, les essais par cette compagnie de production en cellules d'insectes permettent de purifier avec un degré de pureté très moyen environ 100µg d'Artemis.

Des essais ont également été réalisés au CEA de Saclay en collaboration avec le laboratoire de M Gondry en exprimant Artemis dans *E. coli* en fusion avec 7 partenaires de fusion différents dont la GST, MBP, Trx et NusA. Les résultats les plus encourageants sont obtenus avec la MBP qui permet d'exprimer et de purifier quelques mg de MBP-Artemis. Cette fusion s'avère insoluble et l'étiquette MBP ne peut être retirée. Dans ce contexte, les

résultats obtenus sur Artemis en cellules d'insecte au laboratoire sont extrêmement encourageants et seront poursuivis.

2.5 Co-purifications

2.5.1 Co-purifications de Ligase4-XRCC4 et Cernunnos

En parallèle, des purifications du complexe Ligase-XRCC4 et de Cernunnos séparément, nous avons réalisé des co-purifications en partant de 2 culots de 50ml de cultures de cellules d'insecte infectées par du virus de Ligase4-XRCC4 (1 culot) ou par du virus Cernunnos (1 culot). Les culots sont re-suspendus séparément puis lysés ensemble. La première étape est une colonne de Nickel pour fixer le complexe Ligase4-XRCC4 par l'étiquette 10His en Cter de XRCC4. La résine est lavée avec le tampon LA (100mM KCl, 50mM NaCl, 50mM imidazole). Le gel SDS-PAGE montre que dans la fraction du surnageant, les 3 protéines sont fortement sur-exprimées (**Figure 33b Haut)**. Lors de la chromatographie sur la colonne de Nickel, Cernunnos est retenu par le complexe Ligase4-XRCC4. Une certaine fraction de Cernunnos est décrochée dans les lavages à 100mM KCl/50mM NaCl. Cette fraction peut correspondre à un excès de Cernunnos mais également à l'affinité moyenne mesurée entre Cernunnos et XRCC4 (Kd=4.1µM pour les formes tronquées XRCC4(1-157)-Cernunnos(1-224)) *(Ropars, 2011)*. A cette étape, nous obtenons 1,9mg du complexe pour 100ml de culture de cellules d'insectes.

2.5.2 Co-purification de Ligase4-XRCC4 et Ku70/Ku80

En parallèle, des purifications du complexe Ligase-XRCC4 et Ku70/Ku80 séparément, nous avons réalisé des co-purifications en partant de 2 culots de 50ml de cultures de cellules d'insecte infectées par du virus de Ligase4-XRCC4 (1 culot) ou par du virus Ku70/Ku80 sans étiquette (1 culot). Le protocole suivi est semblable à celui décrit ci-dessus pour la co-purification de L4-X4 et Cernunnos. Les culots sont re-suspendus séparément puis lysés ensemble. La première étape est une colonne de Nickel pour fixer le complexe Ligase4-XRCC4 par l'étiquette 10His en Cter de XRCC4. Le gel SDS-PAGE de cette étape montre que Ku70/Ku80 est retenu par le complexe Ligase4-XRCC4 (Figure 33b Bas). Lors du lavage, une certaine fraction de Ku70/Ku80 est décrochée. Cette fraction peut correspondre à un excès de Ku70/Ku80. On peut noter que la co-purification nous permet de purifier du complexe

Ku70/Ku80 dans la fraction non-retenue de cette étape alors que les essais de purification de Ku70/Ku80 sans étiquette à partir de culots de Ku70/Ku80 ne nous ont pas permis de purifier le complexe. A cette étape, nous obtenons 1,9mg du complexe L4-X4-Ku70-Ku80 pour 100ml de culots.

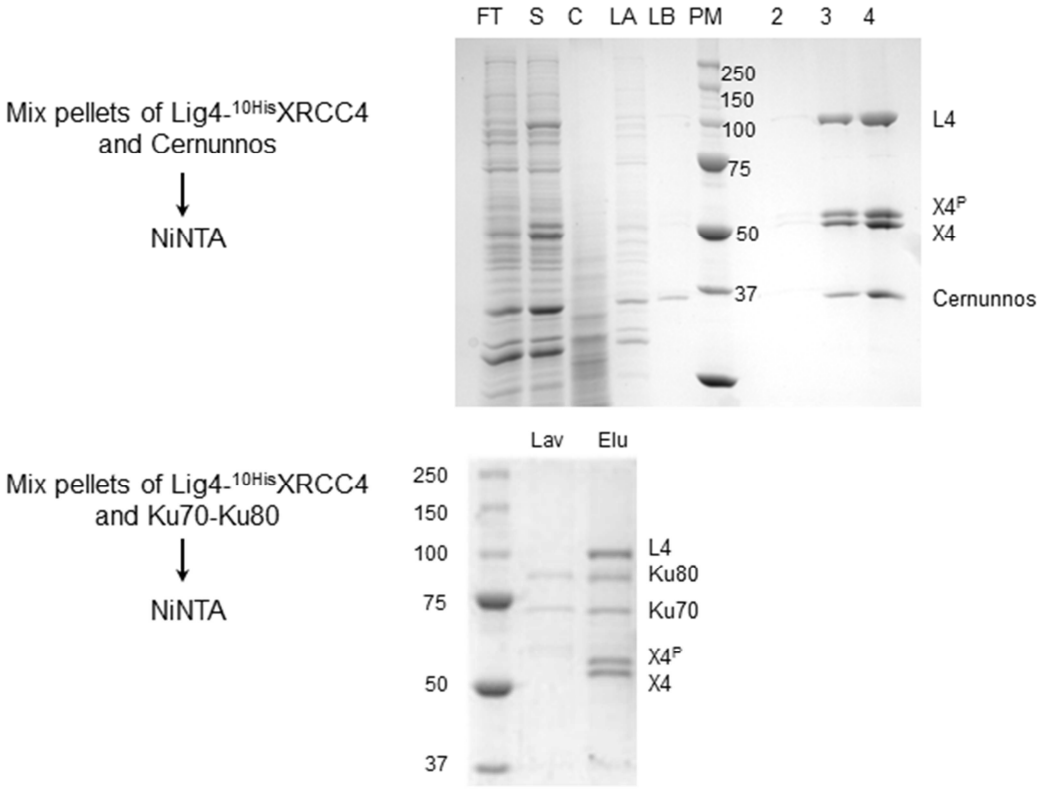


Figure 33b Co-purifications de Ligase4-XRCC4 et Cernunnos et de Ligase4-XRCC4 et Ku70/Ku80
(Haut) Co-purification de Ligase4-XRCC4(10his) et Cernunnos à partir d'un culot de 50ml de cultures d'insectes chacun. Protocole (gauche) et gel SDS-PAGE (droite). (Bas) Co-purification de Ligase4-XRCC4(10his) et Ku70/Ku80 à partir d'un culot de 50ml de culture de cellules d'insectes pour chacun. Protocole (gauche) et gel SDS-PAGE (droite).

2.6 Les améliorations apportées au laboratoire au système de production en MultiBac

La souche de cellules Sf21, dérivée de la souche Sf9 *(Vaughn, 1977)*, est l'une des plus utilisée pour la production en cellules d'insecte. Elle est facile à cultiver même en l'absence de sérum et montre un temps de doublement d'environ 24h. Nous avons toutefois voulu tester la souche High five car elle est connue pour les bons rendements obtenus notamment pour des protéines difficiles. Les résultats obtenus avec les complexes Ligase4-XRCC4 et Ku70/80 confirment un intérêt de la souche High five en termes de quantité de protéines exprimées sur certains de nos projets (**Tableau 1**). Toutefois il convient de signaler que ces

résultats sont issus d'expériences unitaires et qu'une analyse similaire sur 2-3 cultures et purification sera nécessaire pour statuer sur le gain réel apporté par la souche High five. En effet, bien qu'elle se cultive aussi sans sérum et que son temps de doublement soit proche de 24h, la souche High five est dans nos mains plus délicate à manipuler que la souche Sf21. En particulier, elle présente une forte tendance à former des agrégats. Pour l'instant, nous avons choisi au laboratoire d'utiliser principalement la souche sf21. Nous gardons la souche Hi5 pour les protéines peu sur-exprimées comme une alternative possible.

Protéine	Sf21	Hi5
Cernunnos	26mg/L	
Ligase4-XRCC4	10mg/L	60mg/L
Ku70/80 opt	14,5mg/L	31mg/L
Artemis	16mg/L	

Table 1: Rendement de productions des protéines du NHEJ exprimées en cellules Sf1 ou Hi5

La définition d'un rapport optimum (particules virales/cellules) lors de l'infection est un autre aspect important dans la production en cellules d'insecte. Pour définir ce rapport, nous avons d'abord utilisé la technique employée à l'EMBL. Celle-ci consiste à tester plusieurs dilutions de virus sur un ensemble de cultures de cellules établies à $0,5x10^6$ cellules/ml. On évalue alors la croissance à l'aide d'un compteur de cellules et l'expression de la YFP à l'aide d'un fluorimètre après une étape de lyse cellulaire. Nous avons mis à profit la présence d'un FACS au laboratoire pour fusionner ces deux mesures en une seule. Les critères de sélection d'une dilution de virus idéale sont l'obtention d'un DPA 2 jours maximum après l'infection et une synthèse maximale de la YFP en 3 ou 4 jours après le DPA. La satisfaction de ces critères permet d'obtenir une production maximale de protéines. Les **Figures 34A et 34B** montrent les courbes typiques obtenues pour une expérience. Comme attendu, une trop forte dilution de virus permet la croissance mais n'entraine qu'une faible production de la YFP. A l'inverse, la présence d'une trop forte charge virale dans le milieu entraine un brusque arrêt de croissance mais une production de la YFP qui n'est pas optimale. Dans le cas représenté, La valeur idéale se situe entre ces deux extrêmes, en l'occurrence $5x10^{-3}$ % (en volume de culture de virus/volume de culture de cellules d'insecte).

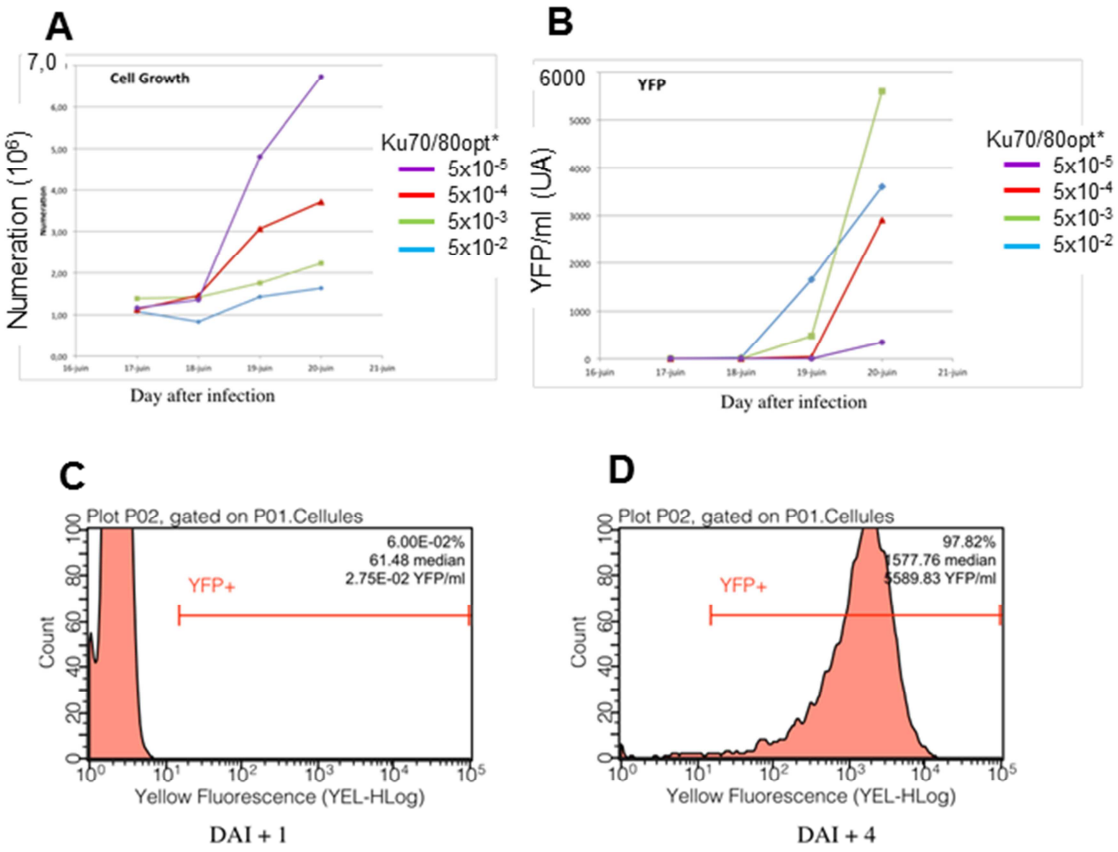


Figure 34: Croissance cellulaire et suivi de production de YFP par FACS
(A)Courbe de croissance de cellules Sf21 infectées par des dilutions croissantes de virus (% en volume virus sur volume de culture) (B) Evolution de la production de YFP dans ces mêmes cellules (C) Comptage de cellules en fonction de leur contenu en YFP 1 jour après l'arrêt de croissance. En encart le pourcentage de population considéré comme YFP positif. (D) Même représentation qu'en(C) mais 4 jours après l'arrêt de croissance.

L'utilisation du FACS nous a également permis de mesurer le pourcentage de cellules infectées et le taux de YFP présent dans chacune (Figures 34C et 34D). On remarque que 4 jours après l'infection, la quasi-totalité des cellules sont infectées et produisent de la YFP. Cette mesure du taux d'infection en temps réel a constitué un outil de surveillance efficace du bon déroulement d'une production.

Nous avons également tiré profit de la présence de la YFP sur les bacmides pour mettre au point une mesure de titration de nos solutions stock de virus par dilution limite. En effet, nous disposons au laboratoire de microscope à fluorescence avec objectif inversé qui permet la lecture de plaque 96 puits. Ainsi il est facile de détecter les cellules infectées par leur fluorescence et donc d'apprécier ainsi la présence de particules virales dans le puits. Cette méthode s'est avérée très efficace et facile à mettre en œuvre car plusieurs solutions virales peuvent être testées en parallèle, une numération ne demandant qu'une plaque de

96 puits au lieu d'un ensemble important d'Erlen. Les numérations effectuées montrent qu'une préparation fraiche de V1 se situe aux environs de $1x10^7$ pfu/ml (**Tableau 2**). Nous avons pu également constater une diminution importante du titre des solutions conservées à 4°C au cours du temps de l'ordre d'un log par trimestre.

Protéine	Date Titration	pfu/mL
Artémis	27/04/15	$1,28 \times 10^6$
Cernunos	17/04/15	$1,54 \times 10^7$
DNAPKcs	17/04/15	$9,22 \times 10^6$
Ku 70/80 Opt	19/11/14	$2,56 \times 10^7$
	03/04/15	$1,90 \times 10^6$
X4L4	03/04/15	$3,60 \times 10^4$

Table 2: Titrations de stock de virus V1 par la méthode de titration sur plaque

2.7 Purifications de XRCC4 et Cernunnos entiers en *E.coli*

Les protéines entières ont été clonées dans un vecteur d'expression pETM13Lic avec une étiquette 6His en position C-terminale. L'étiquette a été positionnée en C-terminale car des essais antérieurs au laboratoire sur une construction GST-Cernunnos ont mis en évidence une protéolyse de la protéine au niveau de sa région C-terminale lors de l'expression dans *E. coli*.

Les purifications sont réalisées à partir de 0,8L de culture de XRCC4 et de 3,2L de culture de Cernunnos. Après l'étape de lyse cellulaire en présence d'inhibiteurs de protéases (EDTA et PMSF), les surnageants sont passés dans un premier temps sur une colonne échangeuse d'anions (colonne Hi-Trap Q HP de 5ml, GE-Healthcare). Les gels SDS-PAGE et les Western blot montrent que XRCC4 et Cernunnos sont éluées entre 200 et 250 mM NaCl en tampon Tris-HCl pH8 (**Figure 35 Haut**). Les fractions sont dialysées pour ôter l'EDTA, puis fixées sur résine Ni-NTA. La résine est lavée avec 0,5M de NaCl puis avec 50mM imidazole afin d'éliminer les contaminants. Les protéines XRCC4 et Cernunnos sont éluées avec 300mM d'imidazole. Ce protocole permet d'obtenir 20 mg de XRCC4 et 1 mg de Cernunnos par litre de culture d'*E. coli* avec un degré de pureté satisfaisant.

Différentes analyses sont effectuées afin de s'assurer de l'intégrité des protéines, de leur pureté, et de leur repliement correct. Un Western Blot est réalisé en fin de purification

pour vérifier la présence de l'étiquette 6His et donc de la présence des derniers acides aminés des protéines. Le gel des protéines purifiées indique la présence d'un faible contaminant vers 70 kDa. L'analyse par digestion et spectrométrie de masse des bandes correspondantes permet d'identifier ce contaminant comme étant une protéine d'*E. coli*, la protéine bifonctionelle de résistance à la Polymixine, ArnA.

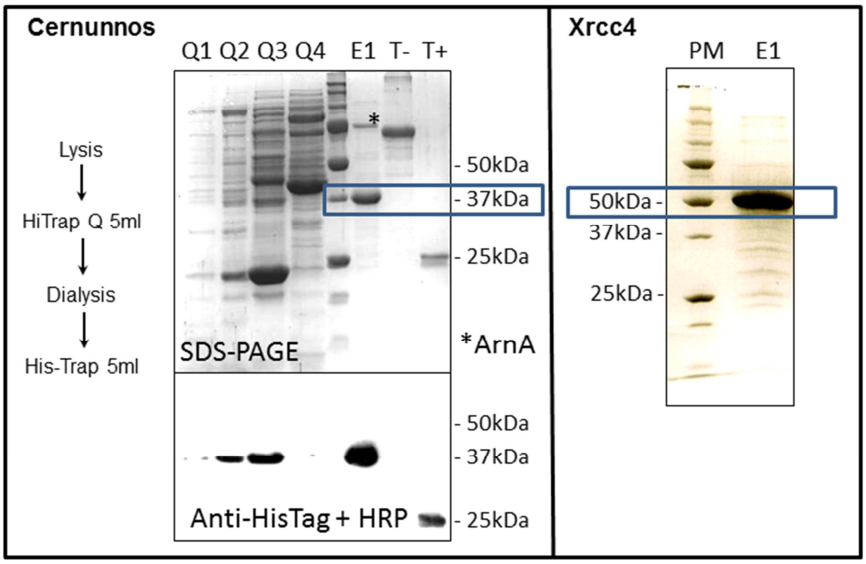


Figure 35 Purification de XRCC4 et Cernunnos entier
(***Haut***) *Gel SDS-PAGE de la purification de Cernunnos entier et XRCC4 entier sur HiTrapQ (1er étape) (**gauche**) puis sur HisTrap (2nd étape) (**droite et gauche**). Cernunnos est élué dans la fraction Q3 du gradient de sels sur la HiTrapQ. T- et T+ sont des protéines témoins sans ou avec une étiquette His6 pour le Western Blot. Le faible contaminant observé avec les protéines Cernunnos correspond à la protéine bifonctionelle de résistance à la Polymixine, ArnA d'E. coli d'après une analyse par spectro de masse.*

3. Caractérisations biochimiques

3.1. Chromatographies d'exclusion de taille analytique

Des analyses par chromatographie d'exclusion de taille (ou SEC pour size exclusion chromatography) analytique ont été effectuées (i) pour évaluer la présence de protéines agrégées dans nos échantillons, (ii) pour évaluer le poids moléculaire apparent, en amont d'études par SEC-MALS (Size Exclusion Chromatography coupled with Multiple Angle Light Scattering) ou UCA (Ultra-Centrifugation Analysis), (iii) pour déterminer des conditions optimales de congélation, et (iv) enfin pour détecter des interactions protéine-protéine fortes. Nous avons utilisé principalement deux colonnes : une Superose6 10/300 (24ml) et une Superdex200 10/300 GL (24ml) (GE-Healthcare)

3.1.1 Analyse par SEC des protéines XRCC4 et Cernunnos entières

Les protéines XRCC4 et Cernunnos entières produites dans *E. coli* sont analysées par chromatographie d'exclusion de taille sur une colonne Superdex 200 (HR 10/300 GL, volume total 24ml, GE-Healthcare). 6 nmoles de Cernunnos sont injectées (**Figure 36 Haut**). On observe un seul pic sur le chromatogramme avec un volume d'élution (Ve) de 12,6ml. On peut noter l'absence de protéine aggrégée, détectable au niveau du volume mort de la colonne (V0) de 7,5ml. Le volume d'élution de Cernunnos se situe entre celui observé sur la même colonne pour l'Albumine (Ve de 13,5 ml pour un Poids Moléculaire (PM) de 67 kDa) et celui observé pour l'Aldolase (Ve de 12,2 ml pour un PM 158 kDa). Ce volume d'élution est compatible avec un dimère de Cernunnos (74 kDa) avec une forme allongée due à la présence d'une région C-terminale prédite comme non-structurée de 75aa ou éventuellement avec un tétramère.

15 nmoles de XRCC4 sont injectées sur la même colonne. Elles sortent sous la forme d'un pic principal avec un Ve de 10,5 ml et deux légers épaulements avec des Ve de 7,5 et 14,5ml. L'épaulement avec un Ve de 7,5ml suggère la présence d'une fraction de la protéine XRCC4 sous la forme d'agrégats. Le volume d'élution observé pour le pic principal de XRCC4 se situe entre celui de l'Aldolase (Ve de 12,2ml pour un PM 158 kDa) et celui de la Ferritine (Ve de 10,0ml pour un PM de 440 kDa). Ce résultat reflete soit la présence d'une forme oligomérique du dimère de XRCC4 de taille plus importante (tétramères, hexamères, ...) soit la présence de la région C-terminale de 133aa prédite comme non-structurée. Des expériences de SEC-MALS ont été réalisées à Gif (I2BC/B3S, A Labarde, réfractomètre tREX Wyatt, colonne HPLC KW-804). Les mesures réalisées montrent la présence d'un pic principal avec un indice de polydispersité Mw/Mn de 1,012 et une masse de 76 kDa, en bon accord avec la masse attendue du dimère de XRCC4 de 78 kDa.

Nous avons alors analysé la formation du complexe XRCC4-Cernunnos par SEC. La Figure 36 Bas montre le chromatogramme obtenu en injectant 15 nmoles de XRCC4 et 3 nmoles de Cernunnos (ratio 5:1) sur la colonne Superdex 200 utilisée précédemment pour analyser les protéines seules. On observe un pic principal, un peu élargi, avec un maximum autour d'un volume d'élution de 9,5ml, entre le volume d'élution de la Ferritine et le volume mort. Ce pic présente deux légers épaulements avec des Ve de 7,5 et 14,5ml. L'analyse sur gel SDS-PAGE des fractions d'élution montre que la protéine Cernunnos migre dans les

fractions 19 à 21 qui correspondent à des volumes d'élution entre 9,5 et 11 ml. On ne trouve donc plus de protéine Cernunnos au volume d'élution observé pour la protéine seule (Figure 36 Haut). Cette expérience montre donc une interaction directe entre les protéines purifiées XRCC4 et Cernunnos entières. Nous avons par la suite confirmé cette interaction par calorimétrie.

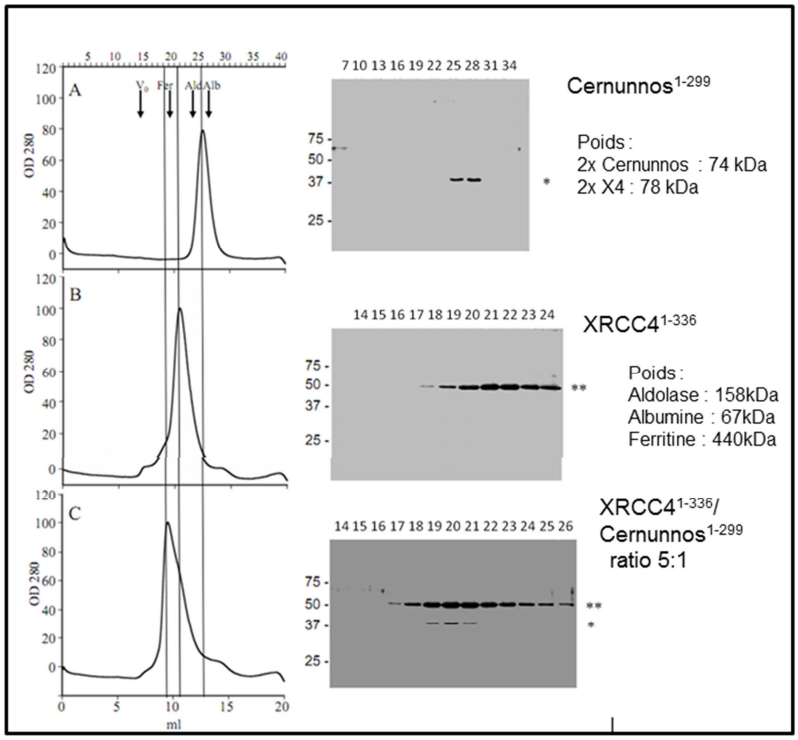


Figure 36 : Chromatographie d'exclusion de taille des protéines Cernunnos, XRCC4 seules et en complexe.

Les protéines sont injectées sur une colonne Superdex 200-24ml et les fractions éluées sont analysées sur gel SDS-PAGE. La colonne est calibrée avec du Bleu Dextran (V0, poids moléculaire >2000kDa), de la Ferritine (Fer, 440kDa), de l'Aldolase (Ald, 158kDa), de l'Albumine (Alb, 67kDa) dont les volumes d'élution sont reportés au-dessus du chromatogramme. Les barres verticales sont des repères des pics principaux obtenus après injection de Cernunnos (Haut), XRCC4 (Milieu) et du mélange Cernunnos-XRCC4 (Bas). Les gels SDS-PAGE correspondants aux deux chromatogrammes sont reportés à leur droite (=Cernunnos, **=XRCC4).*

3.1.2 Analyse par SEC des complexes Ligase4/XRCC4 et Ku70/Ku80 seuls et en complexe

Nous avons analysé selon un protocole similaire à celui décrit pour XRCC4 et Cernunnos entiers, le profil des complexes Ligase4/XRCC4 et Ku70/Ku80 seuls ou en complexe sur une colonne Superdex200 10/300 (24ml). Le tampon utilisé dans ces expériences est un tampon (20mM Tris HCl pH8, 300mM NaCl). Le chromatogramme du complexe L4X4, obtenu en injectant 0,96nmole de complexe, montre un pic principal à un

volume de 10,37ml (volume d'élution proche de celui de la Ferritine, Ve de 10,43ml pour 440KDa) (**Figure 37A**). Le complexe Ligase4/XRCC4$_2$ ayant un poids moléculaire de 183kDa, ceci suggère une forme oligomérique ou une forme allongée du complexe. Le chromatogramme du complexe Ku70/Ku80, obtenu en injectant 0,96nmole de complexe, montre un pic principal à un volume d'élution de 11,61ml (volume d'élution situé entre celui de la Ferritine et celui de l'Aldolase, Ve 12,35ml pour 158kDA) (**Figure 37B**). Ku70/Ku80 ayant un poids moléculaire de 155kDa, ceci suggère une forme monomérique avec un retard généré par des régions dépliées comme le Cter de Ku80 ou une forme dimérique. Des analyses par SEC-MALS et UCA seront réalisées au laboratoire pour préciser ce point.

Nous avons alors analysé l'existence ou non d'une interaction directe entre les complexes Ligase4/XRCC4 et Ku70/Ku80. La **Figure 37C** montre le chromatogramme obtenu en injectant 0,96 nmole de Ligase4/XRCC4 et 0,96 nmole de Ku70/Ku80 (ratio 1:1). On observe deux pics avec des volumes d'élutions, Ve 10,44ml et 11,58ml, proches de ceux observés pour les complexes seuls. Nous n'observons pas de pic supplémentaire de volume d'élution plus petit et le gel SDS PAGE des fractions correspondantes ne montre pas de déplacement de fractions de Ligase4/XRCC4 et de Ku70/Ku80 vers des volumes d'élutions plus petits (**Figure 37E)**. La même expérience est réalisée avec 0,96nmole de Ligase4/XRCC4 et 4,8nmoles de Ku70/Ku80 (ratio 1:5) (**Figure 37D)**. Nous observons également deux pics avec des volumes d'élutions, Ve 10,47ml et 11,76ml, proches de ceux observés pour les complexes seuls. De la même façon, le gel SDS PAGE ne montre pas de déplacement de fractions de Ligase4/XRCC4 et Ku70/Ku80 vers des volumes d'élutions plus petits. Cette expérience suggère l'absence d'interaction directe entre les deux complexes. Des expériences réalisées ultérieurement au laboratoire par microcalorimétrie ont confirmé l'absence d'interaction directe entre Ligase4/XRCC4 et Ku70/Ku80. Les données de la littérature notamment par co-immuno-précipitation indiquent que l'interaction entre Ligase4/XRCC4 et Ku70/Ku80 est en partie dépendante de la présence d'ADN *(Costantini, 2007)*. Les expériences de SEC et de calorimétrie seront réalisées au laboratoire en présence d'un complexe Ku70/Ku80-ADN pour évaluer l'interaction en présence d'ADN.

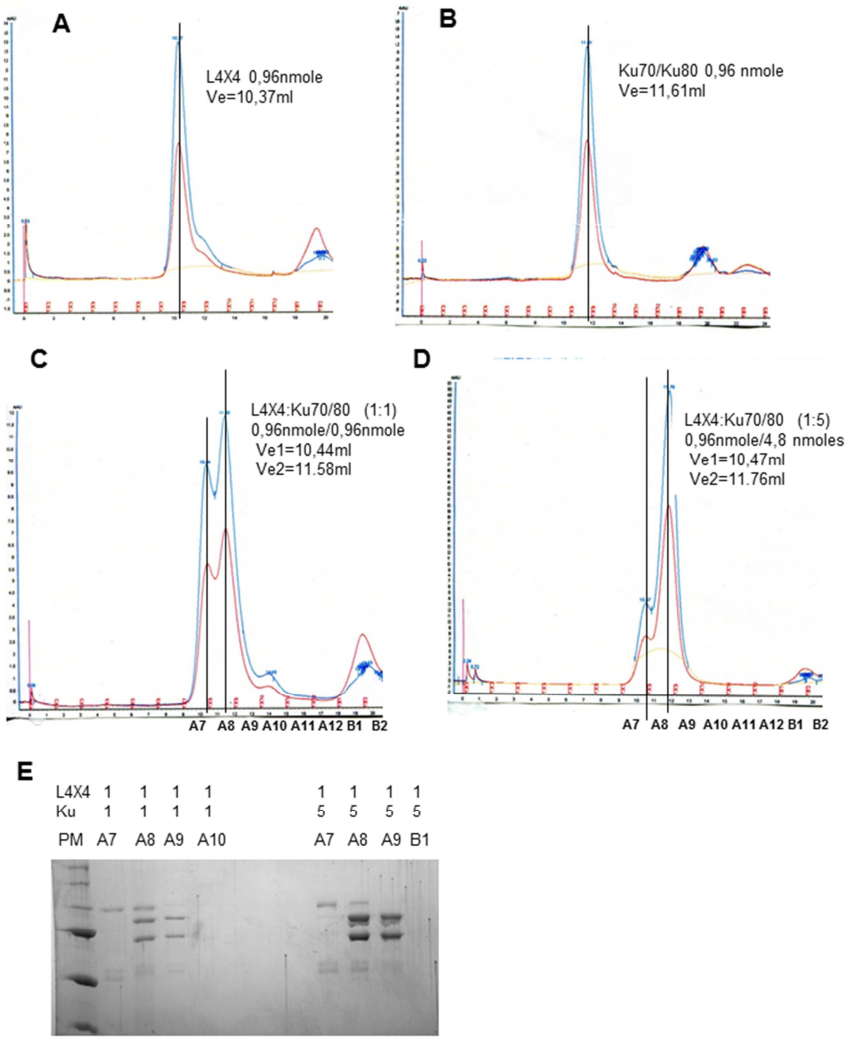


Figure 37 : Chromatographie d'exclusion de taille des complexes Ligase4/XRCC4, Ku70/Ku80 seuls et en complexe

(A*) et (B) analyses par SEC des complexes Ligase4/XRCC4 et Ku70/Ku80. (C) et (D) Analyses par SEC d'un mélange entre les complexes Ligase4/XRCC4 et Ku70/Ku80 dans des ratio 1:1 (C) et 1:5 (D). Gel SDS-PAGE des fractions issues des SEC (C) et (D). Le ratio des deux complexes utilisé lors de l'injection est indiqué au-dessus du gel.*

3.1.3 Analyse par SEC pour évaluer la qualité de la congélation

Les différentes protéines et complexes purifiés à partir de cultures en cellules d'insecte ou dans *E. coli* sont analysés de manière systématique au laboratoire par chromatographie d'exclusion de taille analytique. Les protéines fraichement purifiées ne présentent pas ou très peu de fractions agrégées. Nous utilisons également cette méthode pour évaluer le pourcentage de glycerol qui doit être ajouté afin de congeler les protéines sans affecter leur repliement. Nous analysons le chromatogramme sur SEC analytique avant et après un cycle de congélation/décongélation sans glycerol ou avec des pourcentages

croissants. La condition choisie est le pourcentage de glycerol le plus faible qui ne change pas le chromatogramme avant ou après congélation. La **Figure 38** montre un exemple de résultats obtenus avec le complexe Ligase4/XRCC4. Le complexe est passé ici sur une Superose6 10/300 (24ml). Les échantillons de Ligase4/XRCC4 conservés à 4°C sont injectés dans la semaine sur la Superose6. Un second échantillon est congelé dans l'azote liquide avec 10% glycerol, décongélé et injecté sur la Superose6. On peut voir sur les deux chromatogrammes que les deux échantillons présentent le même profil avec un pic principal avec un volume d'élution Ve de 13,1ml ± 0,1ml. La congélation avec 10% glycerol ne semble pas affecter l'intégrité du complexe.

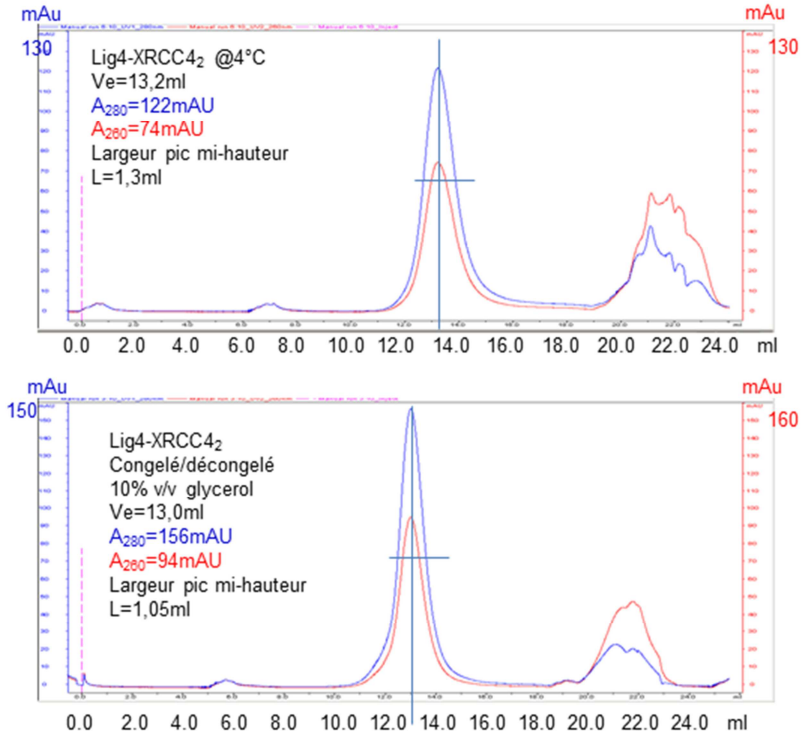


Figure 38 : Chromatographie d'exclusion de taille du complexe Ligase4/XRCC4 avant et après un cycle (congélation/décongélation) avec 10% v/v glycerol
Les deux profils sont similaires. Les pics à 21ml correspondent à de l'ADN ou à des petites molécules.

3.2 Analyses sur l'état de phosphorylation

Les protéines produites en cellules d'insectes sont susceptibles d'être phosphorylées par des kinases d'insectes sur leurs sites physiologiques de phosphorylation et/ou sur des sites consensus de kinases. Les purifications de Ligase4/XRCC4 ont permis de mettre évidence la présence de deux bandes pour la protéine XRCC4. Nous avons réalisé des essais de déphosphorylation à l'aide de deux phosphatases, la protéine lambda phosphatase

(spécifique des phosphorylations sur les acides aminés Ser, Thr et Tyr) et la phosphatase alcaline d'intestin de veau (non spécifique). Les **Figures 39A et B** montrent les essais avec la lambda phosphatase sur le complexe L4/X4 et la **Figure 39C**, les essais avec la phosphatase alcaline sur L4/X4, Ku70/Ku80 et Cernunnos. Les deux phosphatases sont capables de déphosphoryler la bande supérieure de XRCC4. On n'observe alors plus qu'une seule bande, la bande inférieure. Les essais à l'aide de la phosphatase alcaline sur Ku70/Ku80 et Cernunnos n'ont aucun effect sur la migration des bandes de ces protéines. Le second essai réalisé est un essai de coloration des protéines phosphorylées à l'aide du kit pro-Q Diamond. Cet essai réalisé sur le complexe L4/X4 révèle une coloration principalement à la hauteur de XRCC4 (**Figure 39B**).

XRCC4 étant un dimère, les deux bandes observées sur les gels SDS-PAGE peuvent correspondre à un mélange de protéines XRCC4 non-phosphorylées, complétement phosphorylées, ou phosphorylées sur une des deux chaines. Nous pouvons noter que sur les gels SDS-PAGE du complexe L4X4 purifié, les deux bandes de XRCC4 semblent d'égale intensité après coloration au bleu de Coomassie. Une hypothèse, non encore confirmée, est que XRCC4 est systématiquement phosphorylé sur une des deux chaines, ce qui expliquerait cette égale intensité des deux bandes. L'hypothèse est que la kinase responsable, dans les cellules d'insecte, de la phosphorylation de XRCC4 sur une première chaine serait gênée pour phosphoryler la seconde chaine de XRCC4. Cette gêne viendrait du fait de la présence de la Ligase4 ou d'un effet de répulsion de charge ou de changement de conformation induit par la première phosphorylation.

Afin d'identifier le ou les positions phosphorylées sur XRCC4 lors de la production en cellules d'insecte, nous avons découpé les bandes supérieures et inférieures attribuées à XRCC4 et nous les avons envoyées à la plateforme de spectrométrie de masse du SIMOPRO (R Thai, iBiTec-S, Saclay). Les protéines ont été éluées des bandes et soumises à une digestion trypsique puis passées sur un Maldi TOF-TOF. Cette analyse met en évidence la disparition du signal attribué aux peptides contenant la Ser320 ou les Ser327/Ser328. L'absence de ces peptides, que l'on identifie par spectrométrie de masse dans l'échantillon issu de la bande inférieure, pourrait correspondre à une phosphorylation sur 2 ou 3 de ces serines. Ces positions sont en bon accord avec les données de la littérature *(Lee, 2004)*. Il s'agit de sites proposés comme phosphorylés par K-J Lee à partir d'analyses par spectrométrie de masse.

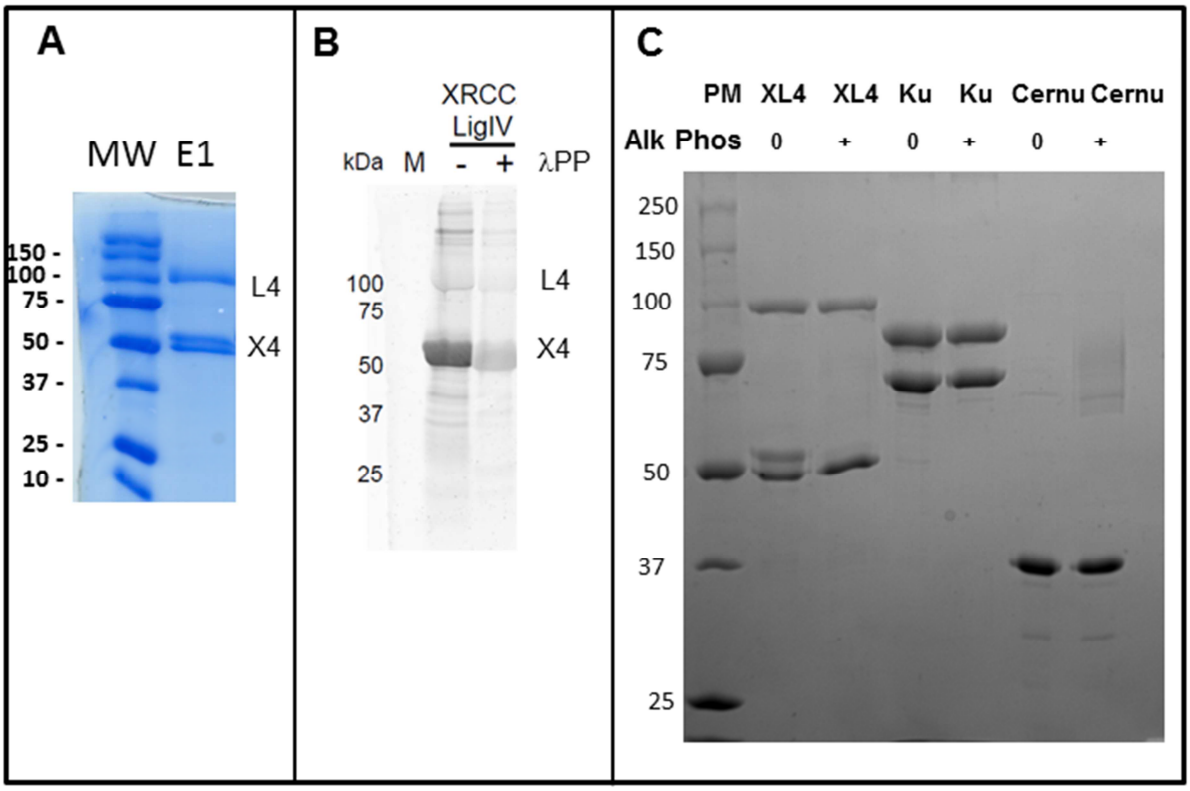


Figure 39 : Analyses de l'état de phosphorylation des protéines produites en cellules d'insectes.
(A) Gel SDS-PAGE des échantillons produits à l'EMBL Grenoble. On distingue les deux bandes à la hauteur de XRCC4. (B) Un test Pro-Q Diamond est réalisé sur ces fractions éluées de colonne de Nickel. Le test met en évidence la présence d'acides aminés phosphorylés au niveau de XRCC4. L'intensité de la bande diminue suite à un traitement à la λ Protein Phosphatase (λ PP). (C) Traitement à la phosphatase alcaline d'intestin de veau des complexes Ligase4/XRCC4, Ku70/Ku80 et de la protéine Cernunnos.

3.3 Analyses sur gel SDS-PAGE de la stoechiométrie des complexes

Nous avons analysé le complexe Ligase4/XRCC4 à l'aide de gels Mini-PROTEAN® TGX Stain-Free™ et du système de lecture Gel Doc EZ de Biorad. Nous avons utilisé les gels, appelés ''any Kd'', qui offrent une gamme de résolution de poids moléculaires adaptée à notre étude. Les tryptophanes de la Ligase4 (au nombre de 10) et ceux de XRCC4 (au nombre de 3 par chaine soit 6 en tout par dimère) réagissent dans le gel avec un dérivé halogéné-alcane qui change la fluorescence du Tryptophane. La **Figure 40** montre le résultat obtenu pour un des échantillons de complexe Ligase4/XRCC4, avant et après un cycle de congélation décongélation. On peut noter que les deux profils sont similaires avec ou sans congélation. La congélation ne fait pas précipiter une sous-unité en particulier.

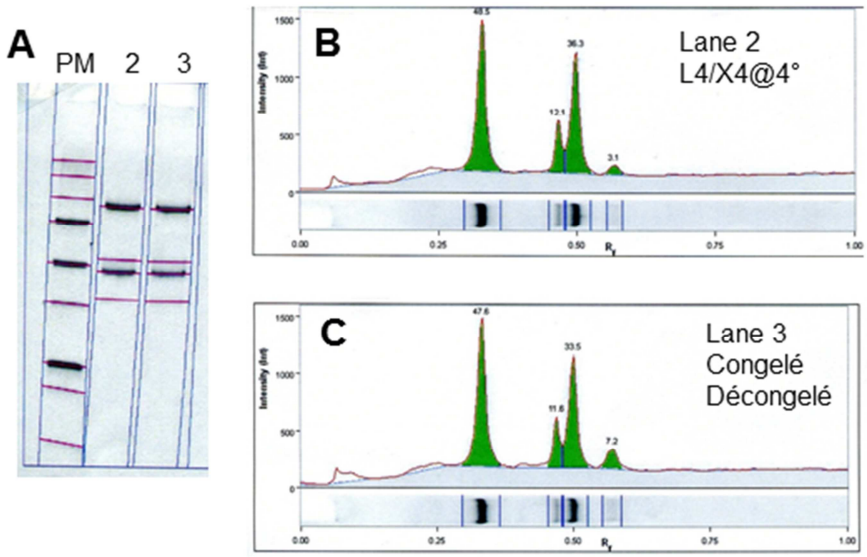


Figure 40 : Analyse d'un échantillon de Ligase4/XRCC4 à partir de gels SDS-PAGE contenant un dérivé halogéné-alcane qui permet de titrer les tryptophanes.

(A) Gel SDS-PAGE de type Any Kd, TGX Staining avec les PM, un échantillon de Ligase4/XRCC4 conservé à 4°C, le même échantillon ayant subi un cycle de congélation/décongélation.

Sur cet échantillon, on peut voir que la bande de XRCC4 phosphorylé représente en intensité un tiers environ de celle de XRCC4 phosphorylé (**Table 3**). Enfin, si on considère le signal de fluorescence de la Ligase4 et son nombre de tryptophane, on peut estimer une stoechiométrie des bandes phosphorylée et non phosphoylée de XRCC4. On obtient 0,83 sous-unité XRCC4 phosphorylée et 2,5 sous-unités non-phosphorylées de XRCC4 par sous-unité de Ligase4 (**Table 3**).

Band	Relative Front	Volume Intensity (10^3)	Protein	Trp Number	Stoichio
1	0,329	1 211	L4	10	1
2	0,467	301	X4(P)	3	0,83
3	0,499	906	X4	3	2,50
4	0,568	78	nd	nd	nd

Table 3 : Intensité relative des bandes de Ligase4, XRCC4 phosphorylé et non phosphorylé à l'aide de la technologie TGX staining de Biorad

Ces résultats sont les premiers obtenus au laboratoire avec cette technologie. Ils doivent être reproduits pour le complexe Ligase4/XRCC4 et d'autres complexes de la voie NHEJ. Cette approche est une méthode rapide pour analyser des résultats de purifications de complexes à plusieurs sous-unités et pour évaluer l'homogénéité de ces complexes.

4. Mesures d'interactions protéine-protéine par calorimètrie

4.1 Mesures de l'interaction entre XRCC4 et Cernunnos entiers

Des mesures d'interactions par calorimétrie ont été présentées dans le manuscrit sur la structure cristalline et par TEM du filament XRCC4-Cernunnos *(Ropars, 2011)*. Ces mesures avaient pour objectif de documenter la nature de l'interaction entre les formes tronquées des protéines XRCC4(1-157) et Cernunnos(1-224) utilisées dans cette étude. Dans le cadre de ma thèse, nous avons évalué la force de l'interaction entre les protéines entières produites dans *E. coli* en utilisant en premier la chromatographie d'exclusion de taille (§ 5.1.1). Nous avons également quantifié cette interaction par calorimétrie à l'aide d'un VP-ITC en titrant un échantillon de Cernunnos entier dans la cellule par un échantillon de XRCC4 dans la seringue. La **Figure 41** montre le thermogramme obtenu avec les protéines entières. Nous observons une interaction endothermique comme avec les formes tronquées de XRCC4 et Cernunnos. La constante de dissociation, Kd, estimée à partir de cette expérience est de 3,9µM, une valeur très proche de celle mesurée avec XRCC4(1-157) et Cernunnos(1-224) qui était de 4,2µM. La présence des régions désordonnées en C-ter de XRCC4 (133aa) et de Cernunnos(75aa) n'a pas d'influence sur l'interaction entre les deux protéines.

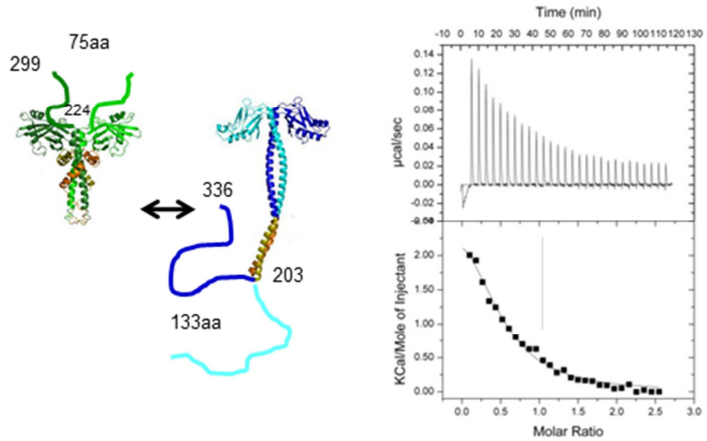


Figure 41 Analyse par calorimétrie de l'interaction entre les protéines XRCC4 et Cernunnos
(Gauche) Structure des dimères de XRCC4 et Cernunnos avec une représentation schématique des régions prédites comme désordonnées et situées en C-ter de ces protéines. Il s'agit de 133aa pour XRCC4 et 54aa pour Cernunnos (Droite) Thermogramme et isotherme de titration obtenus avec les protéines XRCC4 et Cernunnos entiers.

4.2 Mesures de l'interaction entre un complexe Ligase4(CTD)-XRCC4 et Cernunnos

Afin d'évaluer l'influence de l'interaction entre le domaine C-ter de la Ligase4, et XRCC4 sur la force de l'interaction entre XRCC4 et Cernunnos, nous avons purifié un

complexe entre XRCC4(1-203) et le CTD de la Ligase4(654-911) et la protéine Cernunnos (1-224). L'analyse du gel SDS-PAGE (**Figure 42C**) montre le degré de pureté du complexe L4(CTD)-XRCC4(1-203). Ce gel montre l'excès de XRCC4 en bon accord avec la stœchiométrie L4(CTD)-X4 de 1:2 attendue d'après la structure cristallographique du complexe L4(CTD)-X4(1-203)$_2$ *(Wu, 2009)*.

Nous avons choisi de garder le même schéma expérimental pour mesurer l'interaction entre le complexe L4(CTD)-XRCC4(1-203) et Cernunnos(1-224) que celui utilisé auparavant au laboratoire pour mesurer les interactions entre XRCC4(1-203) et Cernunnos(1-224) ou entre XRCC4(1-157) et Cernunnos(1-224) à savoir : (i) les protéines sont dialysées contre le même tampon I (tampon I : 10mM Na2HPO4, 150mM NaCl, 10mM 2-Me-EtOH) (ii) la protéine Cernunnos(1-224), concentrée à 78µM, est injectée par une série de 30 injections de 10µl dans la cellule du calorimètre contenant 1,4ml de XRCC4(1-157) ou XRCC4(1-203) concentré à 20µM. Les données obtenues dans ces expériences montrent que l'interaction entre XRCC4(1-157) et Cernunnos(1-224) est conduite par l'entropie avec une entropie favorable (-TΔS= -8.2 kcal/mol) et une enthalpie faible et défavorable (ΔH= +1,3 kcal/mol). Cette interaction présente une constante de dissociation, Kd, de 4.6µM et une stœchiométrie de 1:1. Les résultats obtenus entre XRCC4(1-203) et Cernunnos(1-224) sont proches de ceux obtenus avec XRCC4(1-157) avec un Kd de 4.2 µM et une réaction conduite par l'entropie.

Le complexe L4(CTD)-XRCC4(1-203) et la protéine Cernunnos(1-224) sont dialysés contre le même tampon I. La protéine Cernunnos(1-224), concentrée à 157µM, est injectée par une série de 30 injections de 10µl dans la cellule du calorimètre contenant 1.4ml de Ligase4(CTD)-XRCC4(1-203) concentré à 16,5µM (**Figure 42.E**). La courbe d'isotherme de titration obtenue est proche de celle observée lors de la mesure réalisée dans les mêmes conditions de concentration avec XRCC4(1-203) et Cernunnos(1-224) *(Malivert, 2010)*. L'interaction est également conduite par l'entropie (-TΔS= -8.4kcal/mol) avec une enthalpie faible (ΔH= +1,3kcal/mol). Cette interaction présente un Kd de 2,7µM et une stœchiométrie entre L4(CTD)-XRCC4(1-157)$_2$ et Cernunnos(1-224)$_2$ de 1:1. L'amélioration de l'affinité d'un facteur 1,7 en présence du domaine de la Ligase4 suggère une légère influence et donc une possible interaction directe entre Cernunnos et le L4(CTD). Cette donnée sera confirmée par

d'autres mesures et par des expériences similaires avec les protéines entières produites en cellules d'insecte.

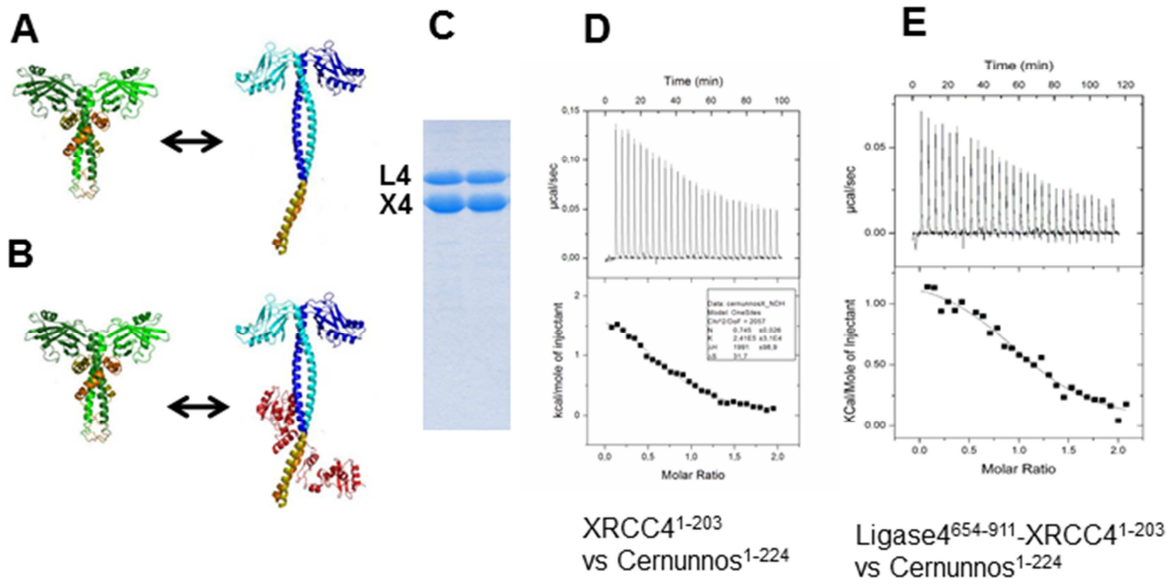


Figure 42 : Analyses par calorimétrie des interactions entre Cernunnos et XRCC4(1-203) ou le complexe XRCC4(1-203)-Ligase(CTD)

(A*) Structure des dimères de XRCC4(1-203) et Cernunnos(1-224). (B) Même vue avec le complexe X4(1-203)-Ligase4(654-911) au lieu de X4(1-203). Cette construction de la Ligase4 contient 2 domaines BRCT en tandem. (C) Gel SDS Page de deux fractions de XRCC4(1-203)-Ligase4(654-911) purifiés. Thermogrammes et isothermes de titration de l'interaction entre (D) XRCC4(1-203) et Cernunnos et (E) entre X4(1-203)-Ligase4(CTD) et Cernunnos(1-224).*

5. Essais de cristallisation

Les protéines purifiées à homogénéité et concentrées à au moins 5mg/ml ont été envoyées sur la plateforme HTX à Grenoble afin de réaliser en moyenne 6 cribles de 96 conditions à 1, 2 ou 3 concentrations. Les images prises par le robot de visualisation ont été observées grâce au logiciel CRIMS via internet. Les gouttes sont annotées à différents moments de J+1 à J+60 selon un code couleur et des numérotations bien définies. La **Table 3** résume les résultats obtenus sur 7 projets étudiés au cours de ma thèse et deux projets réalisés par la suite par C Némoz. Les essais réalisés à ce jour n'ont pas permis d'obtenir de cristaux présentant un pouvoir de diffraction utilisable. Un complexe a donné beaucoup de pistes de cristallisation, le complexe Lig4-XRCC4-Cernunnos (**Table 3**). Un second complexe, Ligase4(654-911)-XRCC4(1-203)-Cernunnos(1-224) a donné des monocristaux qui n'ont pas donné de tâches de diffraction suite aux premiers essais de congélation (**Figure 43E**)

Proteins	Drops (Temp)	Hits (Diffracting)
X4(1-157)-Cer(1-224)	2304 (4℃)	10 (1 at 5.5Å)
L4(654-911)-X4(1-203)-Cer(1-224)	1152 (4℃)	5
L4(FL)-X4(FL)	1152 (4℃)	0
L4(FL)-X4(FL)-Cer(FL)	1152 (4℃)	18
L4(FL)-X4(FL)-Ku70(FL)-Ku80(FL)	576 (4℃)	2
L4(FL)-X4(FL)-Cer(FL)-Ku70(FL)-Ku80(FL)	576 (4℃)	0
Ku70(FL)-Ku80(FL)-ADN(TB)	1152 (4℃)	2
Ku70(FL)-Ku80(FL)-Peptide *	1152 (4℃)	3
X4(FL)-Cer(FL) proteolysis *	96 (20℃)	1

Table 3 : Cribles de cristallisation réalisés sur les protéines du NHEJ produits dans E. coli ou en cellules d'insectes.

Les échantillons envoyés sur la plateforme de cristallisation HTX de l'EMBL sont indiqués avec le nombre de conditions testées et le nombre de pistes observées et testées au laboratoire. Les deux derniers cribles (*) ont été réalisés par C Nemoz.

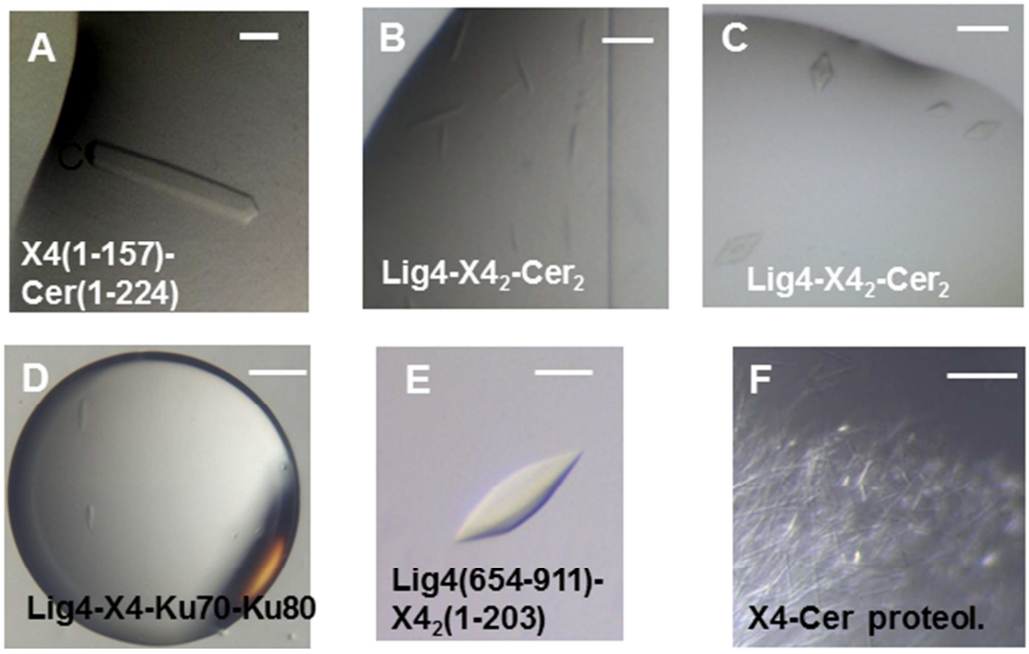


Figure 43 : Cristaux et pistes de cristallisation obtenus avec les protéines Ligase4, XRCC4, Cernunnos, Ku70 et Ku80.

(A*) Forme cristalline du complexe X4(1-157)-Cernunnos(1-224) décrite précedemment* (Ropars, 2011) *La barre mesure 100µm dans toutes les images sauf en D où elle mesure 300µm (B) et (C) deux pistes de cristallisation du complexe Lig4-X4-Cer (D) Piste de cristallisation obtenue avec le mélange Lig4-X4-Ku70-Ku80. (E) Cristaux obtenus avec les protéines utilisées pour la calorimétrie (F) Petites aiguilles obtenues par C Nemoz avec les formes entières de X4 et Cernunnos produits en cellules d'insecte avec une protéolyse in situ.*

Le premier bilan de ces essais de cristallisation est décevant mais il n'est pas complétement surprenant. La stratégie choisie au laboratoire de travailler sur les formes entières des protéines a été pertinente pour initier des études fonctionnelles, des mesures d'interaction ou des analyses par microscopie électronique. En revanche, la présence des régions désordonnée de Ku70, Ku80, XRCC4 et Cernunnos, en Nter ou Cter de ces protéines selon le cas, a probablement gêné l'obtention de formes cristallines bien ordonnées. Dans la suite de ce projet, des constructions nouvelles des protéines du NHEJ seront utilisées pour purifier des protéines ne présentant plus ou peu de régions désordonnées. Il s'agit en premier de l'hétérodimère Ku70/Ku80(1-551) où le domaine Cter de la sous-unité Ku80 est ôté. Cette construction correspond à celle cristallisée avec de l'ADN en 2001 *(Walker, 2001)*. D'autres constructions seront utilisées dont des versions tronquées de Ku70, XRCC4 et Cernunnos ainsi que deux constructions correspondant aux domaines β-lact/ β-CASP d'Artemis, Artemis(1-500) et Artemis(1-370). Une autre direction important qui sera suivie à l'avenir au laboratoire consiste à utiliser des oligonucléotides qui permettent de stabiliser des complexes comprenant à la fois l'hétérodimère Ku70/Ku80 et le complexe Ligase4/XRCC4-Cernunnos.

6. Etude par diffusion des rayons X aux petits angles (SAXS)

Nous avons réalisé des premières études de diffusion aux petits angles qui seront poursuivies au laboratoire. L'objectif était d'effectuer une première analyse sur un mélange stoechimétrique de Ku70/Ku80, Ligase4/$XRCC4_2$ et $Cernunnos_2$. Le mélange formé est analysé sur la ligne SWING (SOLEIL) après passage du complexe sur chromatographie d'exclusion de taille (colonne SEC3). Cette étape permet de séparer les différentes espèces en solution juste avant le passage devant le faisceau de rayons X et ainsi ne mesurer, si possible, la diffusion que d'une unique espèce (échantillon monodisperse). Le chromatogramme d'exclusion de taille du passage sur SEC3 du complexe Ku70/80-XRCC4/Lig4-Cernunnos indique la présence d'agrégats qui sont élués à un volume de 10.96 mL (**Figure 44**). Notre complexe est élué principalement à un volume de 13.27 mL avec un léger épaulement à un volume d'élution plus faible. Des analyses sur SEC3 sont réalisées en amont des expériences sur SWING au laboratoire en utilisant la même colonne SEC3 et un système de purification de type Ethane LC (GE-Healthcare).

La courbe de diffusion obtenue pour le pic principal du passage sur SEC3 est présentée sur le **Figure 44**. Les données sont coupées à q=0.14Å^{-1} car au-delà de cette valeur, la soustraction du tampon pose problème. Cette courbe permet de calculer la courbe P(r) (**Figure 44**), grâce au software DAMMIF (D. Franke, EMBL, Hamburg) et ainsi d'avoir accès à la distribution des distances au sein du complexe diffusant. On voit que le complexe présent dans ce pic principal a une distance moyenne de 50nm et une distance maximum de 175-200nm. De nouvelles expériences sont prévues à la fois au laboratoire pour la mise au point des passages sur SEC3 et sur la ligne SWING pour passer certains sous-complexes de tailles plus petites comme Ku70/Ku80-ADN et Ligase4/XRCC4. Le SAXS pourra être utilisé au laboratoire en complément d'études par cristallographie ou par microscopie electronique de certains complexes multi-protéiques. Le SAXS pourra par exemple permettre de rapidement comparer les enveloppes de complexes en présence d'ADN de structures différentes.

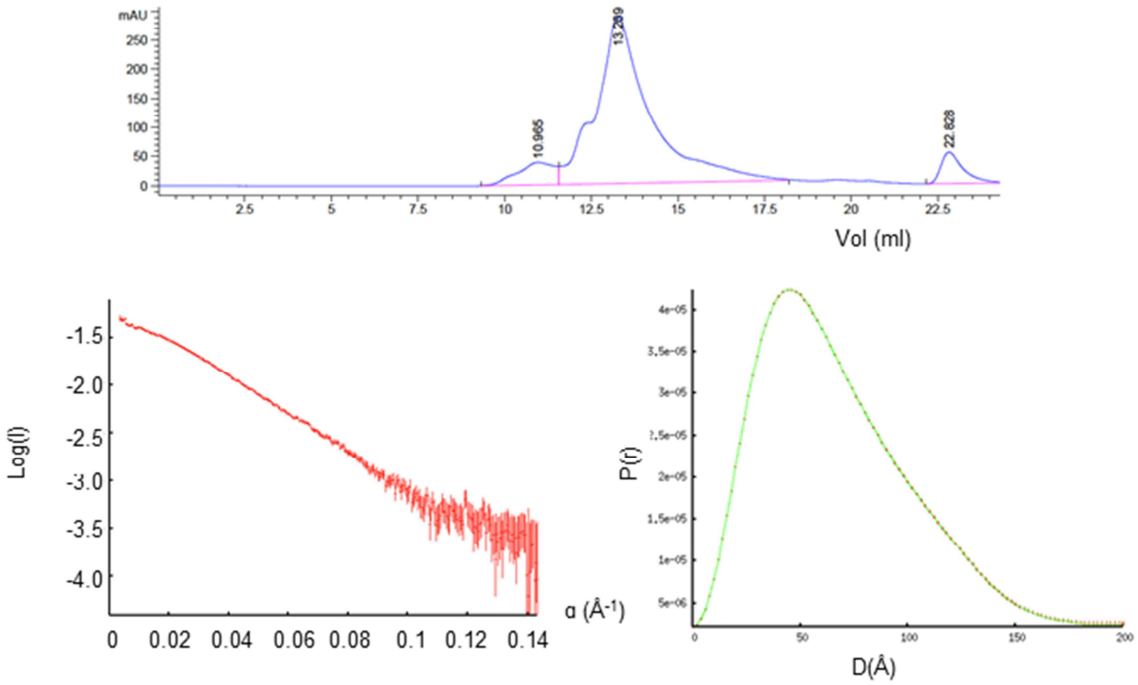


Figure 44 : Etude du complexe Ku70/Ku80-Ligase4/XRCC4$_2$/Cernunnos$_2$ par SAXS.
(Haut) L'étude de diffusion aux petits angles se fait sur la ligne SWING (SOLEIL) après passage du mélange Ku70/80-XRCC4-Lig4-Cernunnos sur une colonne d'exclusion de taille (colonne SEC3). (Bas, gauche) Courbe de diffusion des rayons X du complexe protéique Ku70/80-XRCC4-Lig4-Cernunnos. (Bas, droite) Courbe P(r)= f (D) calculée à partir des données de diffusion.

7. Etudes de microscopie électronique

7.1 Filaments formés par les protéines XRCC4 et Cernunnos entières produites dans E. coli

L'étude structurale des protéines tronquées XRCC4(1-157) et Cernunnos(1-224) a montré que le complexe formé par ces deux protéines adopte dans la structure cristalline un arrangement sous forme de filaments *(Ropars, 2011)*. Cet arrangement a été confirmé par une analyse en microscopie électronique à transmission par coloration négative *(Ropars, 2011)*. A l'issue de la purification des protéines XRCC4 et Cernunnos entières et de la mise en évidence d'une interaction directe entre les deux protéines, nous avons voulu regarder si ces protéines étaient capables ou non de former des filaments et si ces filaments étaient spécifiques du complexe comme cela avait été observé pour les protéines tronquées.

Différentes grilles ont été réalisées à partir de solutions contenant soit Cernunnos entier, soit XRCC4 entier soit le mélange XRCC4-Cernunnos. Les expériences sont réalisées sur des grilles de carbone. On dépose 5µl des protéines préparées à 15µM sur ces grilles puis une goutte d'acétate d'uranyle qui vient enrober les protéines. Les grilles sont analysées sur un cryomicroscope LEO912AB 120kV avec des grandissements compris entre x20 000 et x63 000.

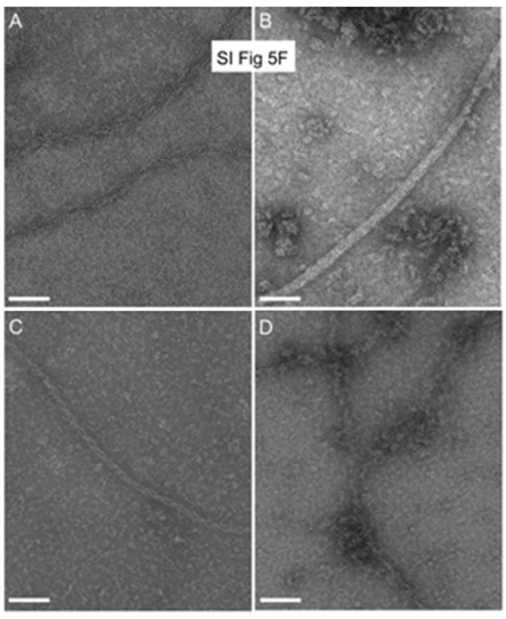


Figure 45 : Images de microscopie électronique à transmission (TEM) avec les protéines XRCC4 et Cernunnos entières

(A*) filaments longs typiques de XRCC4 entier seul sous forme de tresses relâchées (B) Cernunnos entier ne fait pas de filament ou de très rares réguliers (épais 20-25nm). (C et D) filaments observés avec XRCC4 et Cernunnos entiers. Les filaments XRCC4-Cernunnos sont présents en grand nombre et sous forme de tresses très serrées, différentes de celles observées avec XRCC4 seul. La barre fait 100nm*

L'analyse des grilles de Cernunnos entier montre comme pour la forme tronquée essentiellement l'absence de filaments et la présence sur certaines grilles d'aggrégats de protéines. On trouve dans de très rares cas avec Cernunnos des filaments de 40nm de largeur très réguliers en forme d'arc de cercle **(Figure 45B)**. L'analyse des grilles de XRCC4

montre un certain nombre de filaments entrelacés **(Figure 45A)**. La forme tronquée XRCC4(1-157) avait déjà mis en évidence une propension significative de XRCC4(1-157) à former des oligomères sous forme de bâtonnets d'une centaine de nm de long. L'ajout de la région 158-336 de XRCC4 semble avoir accentué la propension de la protéine à faire des filaments. Enfin, le mélange XRCC4-Cernunnos présente des filaments de structures différentes sous la forme de tresses serrées et régulières (**Figure 45C)**. Les filaments observés avec le complexe XRCC4-Cernunnos présentent une structure différente de ceux observés avec le complexe XRCC4(1-157)-Cernunnos(1-224). Les résultats de microscopie électronique suggèrent la présence de filaments spécifiques du complexe XRCC4-Cernunnos.

7.2 Analyses des filaments formés par XRCC4 entier produit dans E. coli

Nous avons dans un second temps étudié avec l'équipe d'E Le Cam (IGR, Villejuif) les filaments réguliers de XRCC4 entier pour mettre au point des méthodes de cryoEM devant servir par la suite aux autres complexes de la voie NHEJ. Cette étude a donné lieu à une collaboration avec le Prof Edward Egelman (University of Virginia). Nous avons ainsi produit régulièrement au laboratoire de la protéine XRCC4 entière à partir de cultures d'*E. coli* et j'ai pu suivre à l'IGR l'acquisition de certaines images comme celle sur la **Figure 46A**. A partir de ces images, E Egelman a sélectionné des régions régulières de ces filaments sur lesquelles il a pu appliquer des algorithmes développés dans son laboratoire pour la reconstrution de filaments hélicoidaux de protéines comme ceux de RecA ou d'actine *(Egelman, 2015)*. Les **Figures 46B et 46C** montre le traitement à l'aide de transformées de Fourier de l'image en **Figure 46A**, l'extraction des informations de symétrie et la reconstruction d'un premier modèle du filament XRCC4. Ce modèle suggère que les filaments de XRCC4 consisteraient en l'enroulement de deux filaments l'un sur l'autre.

7.3 Analyses des filaments formés par XRCC4 et Cernunnos entiers produits dans *E. coli* en présence d'ADN

Afin de mieux comprendre le rôle des filaments XRCC4-Cernunnos, des analyses par microscopie électronique à transmission (TEM) par coloration négative ont été réalisés par S. Baconnais à l'IGR. L'expérience a été réalisée en présence de 3nM du complexe XRCC4-Cernunnos. Dans ces conditions, les filaments XRCC4-Cernunnos ne sont pas observés.

L'ajout de fragments d'ADN dans ces conditions stimule la formation de filaments XRCC4-Cernunnos (**Figure 46D**). On peut noter sur ces images que des fragments d'ADN (structures fines) partent des filaments en différents endroits suggérant une interaction par une extrémité de l'ADN. Ces images sont les premières images par microscopie électronique de l'interaction du filament XRCC4-Cernunnos avec l'ADN.

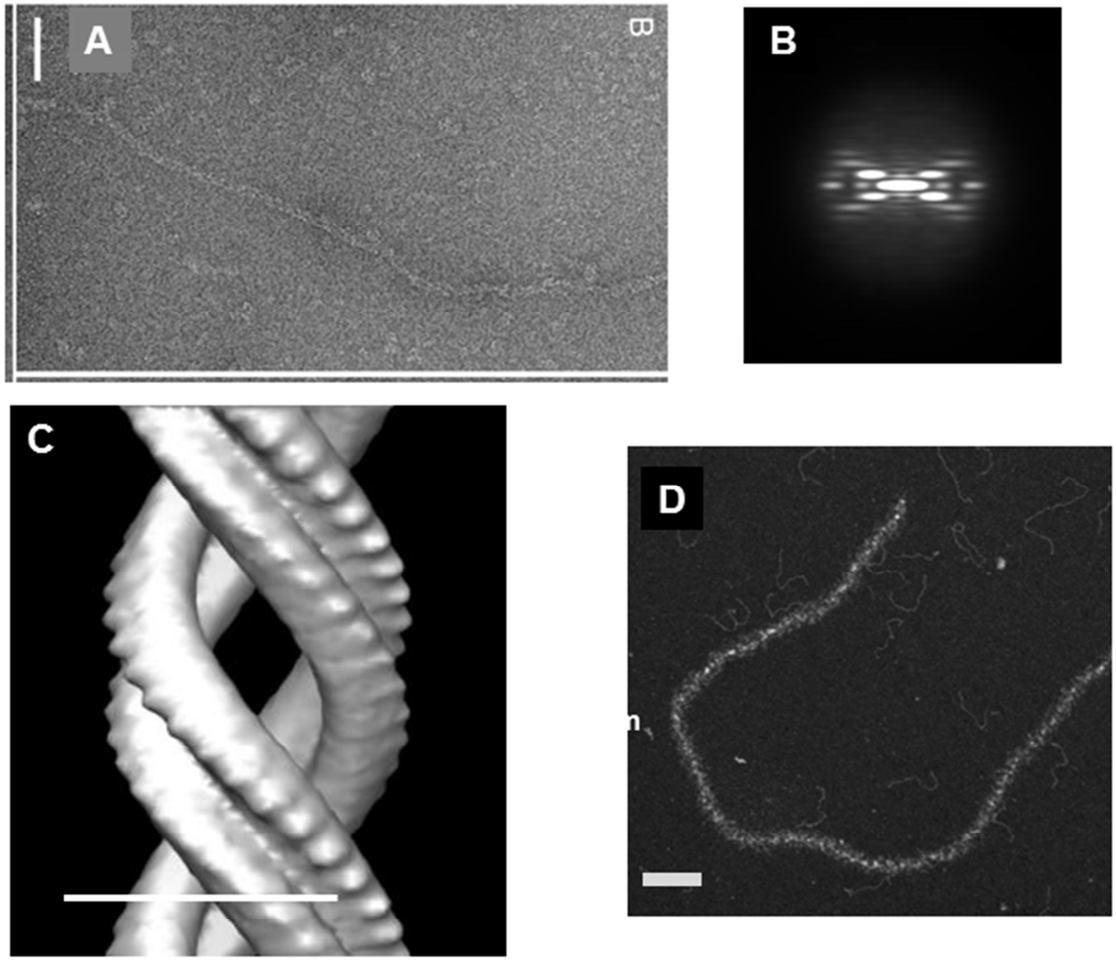


Figure 46 : Images de microscopie électronique à transmission (TEM) et reconstructions
(A*) Filament formé par la protéine XRCC4 entière. (B) et (C) Traitement par transformée de Fourier de l'image 46A et première reconstruction des filaments à l'aide d'algorithmes développés par le Prof E Egelman (Univ of Virginia) La barre fait 20nm. Le pas du filament fait 27nm. (D) Filaments formés par XRCC4-Cernunnos observés à une faible concentration des deux protéines (3nM). La formation de ces filaments est stimulée par la présence d'ADN. On observe plusieurs filaments entrelacés de XRCC4-Cernunnos (structure principale) et des fragments d'ADN seuls ou attachés aux filaments (structures fines) La barre fait 200nm et les fragments d'ADN sont des 600paires de bases*

7.4 Analyses des interactions Ku70/Ku80-ADN

Le complexe Ku70/Ku80 purifié à partir de cellules d'insecte a été utilisé par l'équipe d'E Le Cam. L'objectif était de vérifier l'intégrité du complexe et notamment ses propriétés

de fixation de l'ADN par gels retard et par TEM. La **Figure 47** montre la fixation de plusieurs complexes Ku70/Ku80 sur un même fragment d'ADN de 400 paires de base. On peut observer des complexes réguliers Ku-ADN avec de 1 à 5 molécules de Ku70/Ku80. Les premiers complexes sont observés en gels retard dès une concentration de Ku70/Ku80 de 62nM. Ces mêmes complexes sont analysés sur des grilles de microscopie électronique en reprenant deux des rapports molaires utilisés dans le gel retard, à savoir les rapports molaires Ku/ADN de 8 et de 40.

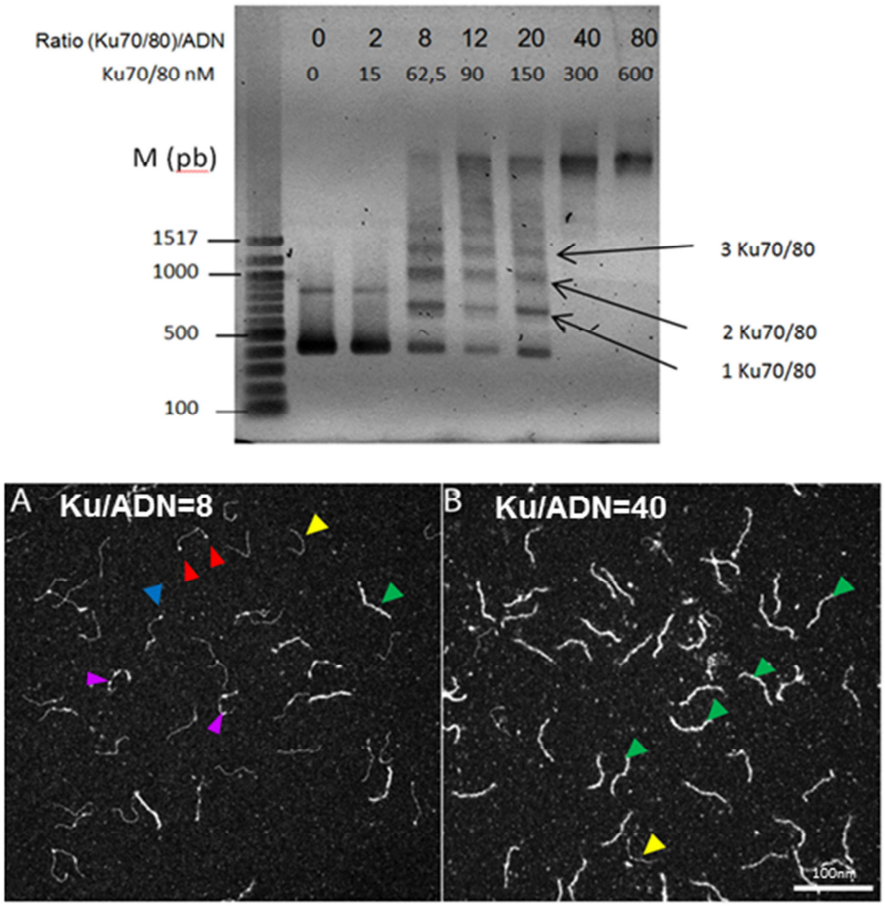


Figure 47 Analyses des interactions Ku70/Ku80 – ADN par gels retard et TEM.
(A*) Un gel retard effectué entre Ku70/Ku80 et un ADN de 400 paires de bases met en évidence des complexes discrets entre une molécule d'ADN et des quantités croissante de Ku qui se forment dès 60nM en Ku. (B) Analyses par TEM des complexes Ku-ADN en se plaçant à deux rapports molaires utilisés dans le gel retard, des rapports Ku/ADN de 8 ou de 40. Les flèches de couleur indiquent les différents complexes ADN-Ku observés et décrits dans le texte.*

Les images obtenues montrent que dans le cas d'un rapport Ku/ADN de 8, on observe clairement différents évenements : (i) une molécule de Ku70/Ku80 fixée à une extrémité (flèche bleue), (ii) une molécule de Ku70/Ku80 fixée à chaque extrémité (flèche rouge), (iii) un ADN recouvert de Ku (flèche verte), (iv) des ADN libres (flèches jaunes) et (v) des ADN qui

semblent avoir des molécules de Ku en dehors des extrémités (flèches magenta). Ces premières images sont très intéressantes car elles ouvrent la possibilité d'étudier l'influence des protéines purifiées au laboratoire, Ligase4, XRCC4, Cernunnos et Artemis, sur le positionnement de l'hétérodimère Ku70/Ku80 sur différents structures d'ADN.

CONCLUSION ET PERSPECTIVES

Les cassures double-brin de l'ADN sont parmi les lésions les plus toxiques de l'ADN. Chez les eucaryotes supérieurs, les CDBs sont réparées principalement par la voie NHEJ (pour Non-Homologous End Joining). La voie NHEJ est également impliquée dans les processus de réarrangement des gènes des immunoglobulines et des récepteurs des cellules T, appelé recombinaison V(D)J. La voie NHEJ présente un premier cœur de sept protéines: l'hétérodimère de reconnaissance des extrémités Ku70/Ku80, la kinase DNA-PKcs impliquée notamment dans la jonction des extrémités et le recrutement du complexe de ligature Ligase4/Xrcc4/Cernunnos, la nucléase Artémis qui clive entre autres des structures en épingle à cheveux générés lors de la recombinaison V(D)J et le complexe de ligature Ligase4/Xrcc4/Cernunnos qui relie les extrémités. A ces protéines cœur de la voie NHEJ s'ajoutent des kinases (PNK, ATM), des phosphatases (PNK, CK2) pour la préparation des extrémités et la régulation de l'activité des facteurs NHEJ. Il faut également inclure des ADN polymérases X (pol λ, pol mu et TdT) qui viennent ajouter des nucléotides dans des brèches ou sur les extrémités et une hélicase associée à la voie NHEJ (WRN).

Les études structurales et fonctionnelles de chacune des protéines pris individuellement ont permis de dresser une première vue générale des propriétés des acteurs de la voie NHEJ. Exceptées les protéines Artemis et WRN, nous disposons des structures cristallines individuelles des différents facteurs cités ci-dessus. Les structures de quelques complexes sont également connus (XRCC4/Cernunnos, Ligase4(CTD)/XRCC4, Ku70/Ku80, Ku70/Ku80/ADN). Il nous manque cependant des structures cristallines d'un

certain nombre de complexes de taille supérieure important pour comprendre des mécanismes moléculaires comme par exemple le recrutement par Ku70/Ku80 du complexe de ligature Ligase4/XCRCC4/Cernunnos ou le recrutement d'Artemis par DNA-PKcs. Ces études cristallographiques sont indissociables d'analyses biophysiques complémentaires comme des mesures d'interaction, de SAXS et de microscopie électronique et de validations par nos collaborateurs biologistes des modèles proposés.

1. Mise en place de la plateforme de production en cellules d'insecte

Afin de mener ces études structurales, nous avons au cours de cette thèse mis en place des protocoles robustes de production en cellules d'insectes des protéines et complexes suivant : Ligase4/XRCC4, Cernunnos, Ku70/Ku80 et Artemis. Nous avons bénéficié pour cela de l'expertise unique du laboratoire d'I Berger et de l'efficacité de son système de production MultiBac. Nous avons bénéficié du programme européen PCube qui a financé notre séjour de 15jours avec P Drevet à l'EMBL Grenoble pour nous former au système MultiBac, à la préparation, au maniement des virus et à la manipulation des cultures en cellules d'insecte. En quinze jours, nous avons pu, grâce à une organisation parfaite proposée par le laboratoire d'I Berger, voir l'ensemble de ces points et nous avons pu à notre retour implémenter cette expertise au CEA de Saclay. Nous avons reproduit à Saclay les résultats obtenus à Grenoble dans un délai très rapide. Nous avons alors progressivement mis en place des protocoles de production en cellules d'insecte sur les protéines de la voie NHEJ

Ce travail a abouti à la création d'un plateau technique de production en cellules d'insectes au CEA de Saclay piloté par notre laboratoire. Ce plateau technique animé par P Drevet a été mis en place au cours des années 2012-2013 sur le projet NHEJ. Nous avons transformé ce plateau technique en plateforme FRISBI au cours de l'année 2014 et avons pris en charge 6 projets externes dès 2014. J'ai moi-même participé à deux de ces projets externes : un premier sur l'acetyl-transférase Nat10 (D Larrieu, ICSN, Gif et S Jackson, University of Cambridge) et un second sur la protéine de l'enveloppe nucléaire l'Emerine (S Zinn-Justin, I2BC, Saclay). Dans les deux cas, j'ai participé à la création des bacmides, à la préparation des virus V0 et V1 et aux premières productions en cellules d'insectes. Les productions effectuées actuellement au laboratoire sont classiquement de 3,2L de culture

par semaine. L'objectif pour le laboratoire est de maintenir en parallèle de ses propres sujets, une activité d'environ 6 projets externes par an prenant en charge la préparation des bacmides, des virus et la production en cellules d'insecte.

2. Production en cellules d'insecte des protéines de la voie NHEJ

Une retombée majeure de ce travail a été la mise en place de protocoles d'expression en cellules d'insecte et de purification à homogénéité de 6 des 7 protéines cœur de la NHEJ humaine avec des rendements élevés (entre 10mg et 60mg par litre de cellules d'insecte). Ce résultat a occupé une partie importante de la durée de la thèse car il a fallu optimiser différents paramètres notamment lors de la culture des cellules et lors de la purification des complexes. Ceci a limité l'étendu des travaux réalisés dans cette thèse sur la partie caractérisation biochimiques et biophysiques des protéines et complexes du NHEJ. Néanmoins, le bilan du travail de production et de purification est extrêmement positif car il définit des bases biochimiques solides pour les études structurales à venir sur la voie NHEJ au laboratoire. Nous avons été notamment très agréablement surpris des rendements obtenus en cellules d'insectes sur des protéines comme la protéine Artemis, réfractaire jusqu'ici aux essais de production. Nous avons également été intéressés par les rendements obtenus de façon générale sur les différentes protéines comparés à ceux classiquement obtenus dans *E. coli* (entre 10mg et 60mg par litre de culture en cellules d'insecte, **Table 1, Résulat §2.6**). Nous avons pu comparer pour une protéine, Cernunnos entier, les rendements obtenus dans *E. coli* (1mg par L de culture) et en Sf21 (26mg par L). Différents développements sont prévus dans la suite de ce travail. Le premier consiste à préparer des plasmides contenant le complexe Ku70/Ku80/Ligase4/XRCC4. Pour réaliser ce complexe, l'étiquette 10His a été déplacée de XRCC4 vers la Ligase4 et les séquences sans étiquette de Ku70/Ku80 sont utilisées. Ce vecteur en cours de construction permettra de purifier en une seule production le complexe Ku70/Ku80/Ligase4/XRCC4 en fixant la Ligase4 sur la colonne de Nickel. Le second développement consiste à mettre au point des protocoles de marquage avec de la sélénométhionine sur un premier complexe, par exemple le complexe Ligase4/XRCC4, afin de pouvoir le cas échéant marquer une protéine présente dans un cristal.

3. Le cas DNA-PKcs

Pendant la durée de ma thèse, des travaux ont été effectués sur la kinase DNA-PKcs. Cette protéine est un cas très particulier avec ses 4128 acides aminés. Elle a été purifiée jusqu'à maintenant principalement à partir de larges volumes de cultures de cellules HeLa. Le laboratoire de T Blundell utilise en effet 100 litres de cultures de cellules HeLa par préparation destinée à la cristallogenèse (communication personnelle) *(Sibanda, 2010)*). Un premier essai de production de DNA-PKcs a été réalisé par C Bigot et C Nemoz lors d'un stage de quinze jours à l'EMBL et un second essai a également été réalisé au laboratoire par C Bigot pour la production et C Tellier-Lebegue pour la purification. L'essai réalisé à Grenoble semble encourageant avec la présence d'une bande de haut poids moléculaire sur les extraits totaux et sur des extraits nucléaires (données non montrées). Les premiers Western blot n'ont pas permis de confirmer la présence de DNA-PKcs mais nous ne pouvons exclure des problèmes techniques lors de la réalisation du Western blot avec une protéine de cette taille ou sa dégradation au cours du processus de production. L'essai réalisé au laboratoire a été fait en mélangeant deux culots de cellules, l'un provenant d'une culture de production de Ku70/Ku80 et l'autre de DNA-PKcs. L'idée est de stabiliser DNA-PKcs avec un de ses partenaires. Les gels SDS-PAGE des extraits faits au CEA de Saclay et les premières étapes de purification sur colonne de Nickel, montrent la présence de Ku70/Ku80 mais pas de bandes visibles autour de 470kDa. Des essais d'activités kinase sur une cible classique de DNA-PKcs, le variant d'histone H2AX, ont été réalisés par C Mann (I2BC, Saclay) avec les fractions faites à Grenoble et avec celles faites à Saclay. Ces tests n'ont pas permis de mettre en évidence une activité kinase. Ces premiers résultats bien que négatifs nous ont permis d'identifier des problèmes techniques inhérents à la taille de DNA-PKcs lors de la réalisation des gels SDS-PAGE, des Western blot et des tests enzymatiques. Dans ce contexte, le laboratoire utilisera des cellules HeLa et commandera de l'enzyme DNA-PKcs pour faire des controles positifs dans les expériences réalisées. Le projet DNA-PKcs est un projet délicat mais l'enjeu en cas de succès dans la production de cette protéine est majeur. Nous utiliserons également, si nécessaire, des vecteurs MultiBac récemment créés par I Berger. Ces vecteurs possèdent 3 MBP consécutives en Nter suivi d'un site TEV et ont donné des résultats intéressants dans le laboratoire d'I Berger sur d'autres enzymes de taille importante de la famille PI3K.

4. Etude du réseau d'interactions médiées par Ku70/Ku80

Les travaux de ces dernières années ont clairement montré que l'hétérodimère Ku70/Ku80 joue au sein de la voie NHEJ un role qui dépasse celui de la reconnaissance des extrémités des cassures double-brin de l'ADN. Ku70/Ku80 est au cœur d'un réseau d'interactions qui impliquent des protéines de la voie NHEJ, d'autres voies de réparation, de la maintenance des télomères et de la chromatine, et d'autres protéines impliquées notamment dans l'apoptose ou la transcription (**Figure 48**). La voie NHEJ est connue pour avoir une grande flexibilité qui lui permet de prendre en charge des cassures double-brin de l'ADN de structures très différentes. Cette flexibilité pourrait provenir de l'hétérodimère Ku70/Ku80 qui recruterait alternativement des kinases, phosphatases, nucléases, polymérases et ligases au niveau de l'extrémité d'une cassure en fonction de sa nature *(Lieber, 2008)*. Ku70/Ku80 apparait ainsi comme un chef d'orchestre de la voie NHEJ. Un objectif majeur du laboratoire dans les années à venir sera de documenter par des structures cristallines la nature des interactions entre Ku70/Ku80 et ses partenaires. Il s'agit notamment de comprendre la compétitivité qui existe entre certains partenaires, de localiser les sites d'interactions par rapport aux extrémités d'ADN. Le laboratoire pourra s'appuyer pour cela sur les protocoles robustes de purification de Ku70/Ku80 établis pendant cette thèse et sur les premiers travaux biochimiques et biophysiques réalisés.

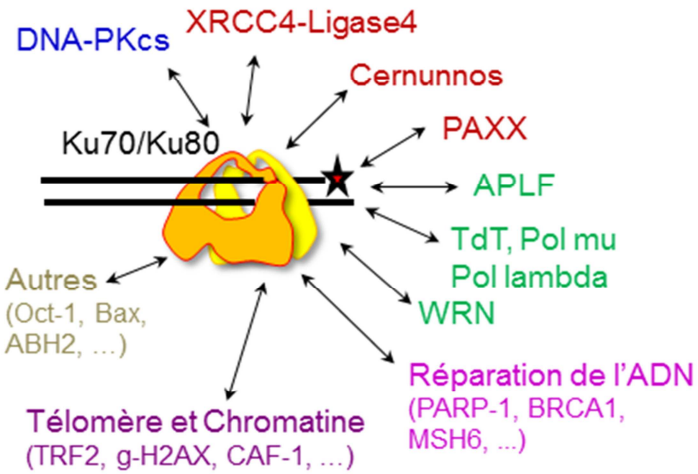


Figure 48 : Réseau d'interaction autour du complexe Ku70/Ku80
Réseau d'interactions décrit autour de l'hétérodimère Ku70/Ku80 (d'après (Grundy, 2014)*). Certaines de ces interactions ont été montrées comme directes (Cernunnos, PAXX, APLF, WRN par exemple), d'autres sont dépendantes ou stimulées par la présence d'ADN sur Ku (DNA-PKcs, XRCC4-Ligase4)*

5. Dynamique des filaments XRCC4-Cernunnos

Les travaux du laboratoire en 2011 ainsi que ceux de trois autres laboratoires ont stimulé la communauté scientifique travaillant sur la voie NHEJ en proposant l'existence de structures d'ordre supérieur du complexe formée par les protéines XRCC4 et Cernunnos. Ces résultats *in cristallo* ont été étayés progressivement par d'autres résultats *in vitro.* Les travaux du laboratoire de J Tainer ont ainsi montré des courbes de SAXS du complexe suggérant l'existence d'oligomères de taille supérieure à l'hétéro-tétramère $XRCC4_2$-$Cernunnos_2$ et la présence d'héxamères ou d'octamères *(Hammel, 2010)*. L'équipe d'E Le Cam, en collaboration avec notre laboratoire, a montré l'existence de filaments formés par le complexe XRCC4-Cernunnos par TEM avec les protéines tronquées de leur région C-ter ou avec les protéines entières *(Ropars, 2011)*. Plus récemment, un travail en microscopie optique à super-résolution a été présenté par Reid et al *(Reid, 2015)*. Ce travail majeur montre l'existence *in cellulo* d'un filament constitué de XRCC4, de Cernunnos mais aussi de la Ligase4. Ce filament recouvre l'ADN et interagit avec un seul hétérodimère de Ku70/Ku80, situé près de l'extrémité de l'ADN. De façon intéressante, des mesures par FRET en molécule unique (smFRET) suggèrent dans ce même travail que deux filaments $(Ligase4\text{-}XRCC4\text{-}Cernunnos)_n$-Ku70/Ku80 seraient capables de glisser l'un à côté de l'autre. Ce travail donne pour la première fois une existence physiologique aux filaments proposés à partir de structures cristallines. De façon intéressante, il pose de nouvelles questions sur l'insertion de la Ligase4 dans les filaments XRCC4-Cernunnos. Il suggère la présence d'interactions labiles entre Ku70/Ku80, positionné en tête de ces filaments et les bords d'un second filament constitué du complexe Ligase4/XRCC4/Cernunnos. Dans ce contexte, un des objectifs du laboratoire est de cristalliser un complexe décrivant les interactions entre Ku70/Ku80 et Ligase4/XRCC4/Cernunnos et de caractériser par des approches biophysiques complémentaires les interactions suggérées par ce dernier modèle.

ANNEXES

Annexe 1 : Protocole de Transfection de cellules d'insecte

1- **Jour 1:**
   a. Transformation de bactéries DH10EMBacY électrocompétentes par électroporation. 3 µl de plasmide (1-50 ng/µl) dans 100 µl de cellules. **Attention : le plasmide doit être dans une solution exempte de sel)**
   b. Reprendre dans 900 µl de SOB et incuber O/N (au moins 4h) à 37°C sous agitation.

2- **Jour 2**
   a. Etaler les bactéries sur milieu LB agar Kan/Tet/Gen/IPTG/XGal
      i. Composition:
         1. Kan 30 µg/ml
         2. Tet 20 µg/ml
         3. Genta 20 µg/ml
         4. IPTG 0,1 mM
         5. XGal 40 µg/ml
      ii. Etaler plusieurs dilutions de bactéries (1/1; 1/10; 1/100; 1/1000). 150 µl par boites
      iii. Incuber o/N 37°C

3- **Jour 3**
   a. Les colonies blanches ont intégré le plasmide dans le bacmide. On étale de nouveaux plusieurs colonies blanches sur le milieu LB Agar K/T/G/I/X, ainsi qu'une colonie bleue comme témoin négatif
   b. O/N 37°C

4- **Jour4**
   a. Ensemencer 2 tubes de 2ml de LB Kan/Tet/Gen avec 2 colonies différentes.
   b. O/N 37°C. **Attention : ne pas dépasser une nuit d'incubation**.

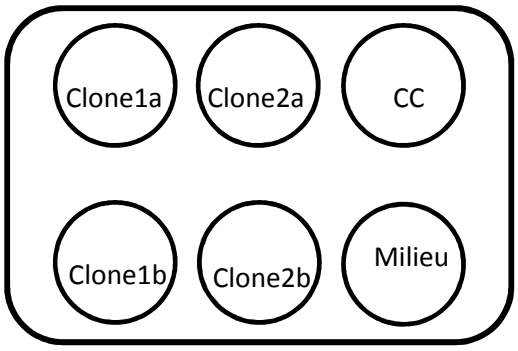
5- **Jour 5**
   1. Centrifuger 5min 13 000rpm
   2. Lyse alcaline:
      i. Reprendre les culots dans 300 µl de solution 1 (Tris/glucose)
      ii. Ajouter 300 µl de solution 2 (SDS). **Agitation TRES douce.**
      iii. Ajouter 300 µl de solution 3 (acétate). **Agitation TRES douce.**
   3. Centrifuger 10 min 13 000 rpm.
   4. Garder le surnageant
   5. Centrifuger 3 min 13 000 rpm.
   6. Garder le surnageant
   7. Ajouter 700 µl d'Isopropanol
   8. Centrifuger 10 min 13 000rpm
   9. Eliminer **complètement** le surnageant (le culot est très peu visible).
   10. Ajouter **avec précaution** 200 µl d'éthanol 70%.

11. Centrifuger 5 min 13 000rpm
12. Eliminer le surnageant
13. Ajouter 50 µl d'éthanol 70%

    A partir de ce moment on manipule sous le PSM.

    Bien décontaminer l'extérieur des tubes avant de les introduire sous le PSM

14. Eliminer l'éthanol
15. Laisser sécher les culots (bacmide) tube ouvert pendant 10 à 15min
16. Pendant ce temps préparer les cellules:
    i. Répartir 2 ml de milieu par puits
    ii. Compter les cellules en culture. Diluer à 0,5-1*10^6 cell/ml
    iii. Répartir 1 ml de dilution par puits (sauf 1 = Témoin Milieu) en dispersant de façon la plus homogène possible. **Faire très attention aux contaminations à ce stade, ne pas passer les mains au-dessus de la plaque.**
    iv. Schémas de la plaque:



    v. Laisser 15 min à RT
17. Ajouter 20 µl d'eau stérile aux culots de bacmides secs. Tapoter doucement les tubes sur la paillasse pour remettre en suspension.
18. Ajouter 200 µl de milieu dans chaque tube
19. Pour 2 bacmides préparer un tube Eppendorf avec
    200 µl de milieu + 20 µl de X-TremeGENE (ROCHE) = Transfectant
20. Distribuer 100 µl de transfectant dans chaque tube de bacmide remis en suspension dans 200 µl de milieu
21. Distribuer 150 µl du mélange précédent dans les puits appropriés, goutte à goutte, en dispersant de façon homogène. **Attention aux contaminations**.
22. Incuber les plaques à 27°C

6- Jours suivants (environ 60h)

a. Prélever les surnageants des 6 puits d'essais **= V_0**. Conserver à 4°C à l'abri de la lumière
b. Ajouter 3 ml de milieu dans les puits.
c. Incuber 3 jours et faire un test d'expression YFP (Fluo) et protéines clonées (SDS-PAGE)

Annexe 2 : Réalisation d'un stock de V1

1 Infecter une flasque de 25ml à $0,5*10^6$ cellules/ml avec 3ml de V0

2 Compter toutes les 24h.
   a. Si les cellules ont doublé (≥ $1*10^6$ cellules/ml) ⇒ Diluer la culture par addition de milieu afin de revenir à $1*10^6$ cellules/ml.
   b. Si les cellules n'ont pas doublé (< $1*10^6$ cellules/ml) ⇒ On est en **dpa** (Day after Proliferation Arrest). Continuer au point 3

3 Faire un prélèvement pour mesure de fluorescence:
   a. Centrifuger 2min 12 000rpm
   b. Eliminer le surnageant
   c. Reprendre le culot dans 500µl de PBS
   d. Casser les cellules aux ultra-sons (30sec)
   e. Centrifuger 2 min 12 000rpm
   f. Mesurer la fluo dans le surnageant

4 Au bout de 12h de culture répéter le point 3 (dpa+12)

5 Au bout de 24h de culture (dpa+24)
   a. Faire un prélèvement pour mesure de fluorescence (pt 3)
   b. Centrifuger la culture 3 min 800rpm
   c. **Prélever le surnageant = V1**. Conserver à 4°C à l'abri de la lumière
   d. Reprendre le culot de cellules avec précaution dans 25 ml de milieu
   e. Compléter au volume initial avec du milieu neuf

6 Continuer la culture et les mesures de fluorescence jusqu'à obtenir un plateau. Vérifier éventuellement la présence de la protéine recombinante dans le culot cellulaire par SDS-PAGE.

Annexe 3 : Table de McGrady

3 tubes par dilution					
Nombre caractéristique	Nombre de cellules	Nombre caractéristique	Nombre de cellules	Nombre caractéristique	Nombre de cellules
0	0.0	201	1.4	302	6.5
1	0.3	202	2.0	310	4.5
10	0.3	210	1.5	311	7.5
11	0.6	211	2.0	312	11.5
20	0.6	212	3.0	313	16.0
100	0.4	220	2.0	320	9.5
101	0.7	221	3.0	321	15.0
102	1.1	222	3.5	322	20.0
110	0.7	223	4.0	323	30.0
111	1.1	230	3.0	330	25.0
120	1.1	231	3.5	331	45.0
121	1.5	232	4.0	332	110.0
130	1.6	300	2.5	333	140.0
200	0.9	301	4.0		

Annexe 4 : Purification XRCC4-10His/Ligase4

Matériel : Protocole pour un culot de 200ml de cultures cellules Sf1 (ou Hi5) infectées par des Baculovirus V1 pKL-X4(10His-Cter)-L4

Resource Q 6ml (Ge-Healthcare); Résine Ni-NTA 8ml en vrac Qiagen

			PM (Da)	Eps (M-1.Cm-1)	pI	Abs (1mg/ml)
XRCC4	fl, Cter 10His	XRCC4fl10H	39658	25690	5.33	0.648
Lig4	fl	Lig4	103971	102440	8.17	0.985

Protéines

XRCC4 : (étiquette en gras) >XRCC4fl_10His

MERKISRIHLVSEPSITHFLQVSWEKTLESGFVITLTDGHSAWTGTVSESEISQEADDMAMEKGKYVGELRKALLSGAGPADVYTFNFSKESCYF
FFEKNLKDVSFRLGSFNLEKVENPAEVIRELICYCLDTIAENQAKNEHLQKENERLLRDWNDVQGRFEKCVSAKEALETDLYKRFILVLNEKKTKIR
SLHNKLLNAAQEREKDIKQEGETAICSEMTADRDPVYDESTDEESENQTDLSGLASAAVSKDDSIISSLDVTDIAPSRKRRQRMQRNLGTEPKM
APQENQLQEKENSRPDSSLPETSKKEHISAENMSLETLRNSSPEDLFDEI**HHHHHHHHHH****

Number of amino acids: 346; **Molecular weight:** 39658.1 Da; **Theoretical pI:** 5.33

Lig4 (sans étiquette) >Lig4-pKL

MAASQTSQTVASHVPFADLCSTLERIQKSKGRAEKIRHFREFLDSWRKFHDALHKNHKDVTDSFYPAMRLILPQLERERMAYGIKETMLAKLYI
ELLNLPRDGKDALKLLNYRTPTGTHGDAGDFAMIAYFVLKPRCLQKGSLTIQQVNDLLDSIASNNSAKRKDLIKKSLLQLITQSSALEQKWLIRMII
KDLKLGVSQQTIFSVFHNDAAELHNVTTDLEKVCRQLHDPSVGLSDISITLFSAFKPMLAAIADIEHIEKDMKHQSFYIETKLDGERMQMHKDG
DVYKYFSRNGYNYTDQFGASPTEGSLTPFIHNAFKADIQICILDGEMMAYNPNTQTFMQKGTKFDIKRMVEDSDLQTCYCVFDVLMVNNKKL
GHETLRKRYEILSSIFTPIPGRIEIVQKTQAHTKNEVIDALNEAIDKREEGIMVKQPLSIYKPDKRGEGWLKIKPEYVSGLMDELDILIVGGYWGKG
SRGGMMSHFLCAVAEKPPPGEKPSVFHTLSRVGSGCTMKELYDLGLKLAKYWKPFHRKAPPSSILCGTEKPEVYIEPCNSVIVQIKAAEIVPSDM
YKTGCTLRFPRIEKIRDDKEWHECMTLDDLEQLRGKASGKLASKHLYIGGDDEPQEKKRKAAPKMKKVIGIIEHLKAPNLTNVNKISNIFEDVEFC
VMSGTDSQPKPDLENRIAEFGGYIVQNPGPDTYCVIAGSENIRVKNIILSNKHDVVKPAWLLECFKTKSFVPWQPRFMIHMCPSTKEHFAREYD
CYGDSYFIDTDLNQLKEVFSGIKNSNEQTPEEMASLIADLEYRYSWDCSPLSMFRRHTVYLDSYAVINDLSTKNEGTRLAIKALELRFHGAKVVSC
LAEGVSHVIIGEDHSRVADFKAFRRTFKRKFKILKESWVTDSIDKCELQEENQYLI*

Number of amino acids: 911; **Molecular weight:** 103970.7Da; **Theoretical pI:** 8.17

Tampons

Tampons Lyse NiNTA-XL4 :
25mM Tris pH8
150mM KCl
150mM NaCl
10mM 2-Me-EtOH
5% Glycérol
50mM Imidazole

Tampon Elution NiNTA-XL4 :
25mM Tris pH8
150mM KCl
150mM NaCl
10mM 2-Me-EtOH
10% Glycérol
300mM Imidazole

Tampon QB-XL4 :
25mM Tris pH8
150mM KCl
850mM NaCl
10mM 2-Me-EtOH

Tampons Lavage NiNTA-XL4 :
25mM Tris pH8
150mM KCl
850mM NaCl
10mM 2-Me-EtOH
10% Glycérol
50mM Imidazole

Tampon QA-XL4 :
25mM Tris pH8
100mM KCl
10mM 2-Me-EtOH

Préparation :

Préparer la résine Ni-NTA

Préparer les différents tampons (ajouter le 2-Me-EtOh extemporanément, filtrer à 0,2µ, vérifier les stocks de Benzonase, d'inhibiteurs de protéases,

Faire un cycle à blanc sur la Ressource Q 6mL : 5 CV de tampon QA-XL4, puis un gradient 0-100% de tampon QB-XL4 sur 5 CV puis 5 CV de Tampon QB-XL4, et enfin 5 CV de tampon QA-XL4

Jour1

❑Reprise de 1 à 4 culots de 200mL de culture au 1/8^{e} dans le tampon de Lyse NiNTA-XL4, ajout d'inhibiteurs de protéases (Pepstatine (1µg/ml) / Chymostatine (1µg/ml) / Leupeptine (0,5µg/ml) / AEBSF ou PMSF (1mM) / Benzamidine (34µg/ml) final). Garder 10µL pour le premier gel SDS-PAGE à réaliser après l'étape de NiNTA (échantillon extrait total ET).

❑Après resuspension à la Propipette ou au Vortex, lyse par sonication (3x1 minute à 60% d'amplitude 1 sec On / 1 sec Off) en veillant à maintenir le lysat entre 0 et 4°C

❑Ajout de Benzonase (0,02U/µl de Benzonase Merck + 1mM de $MgCl^2$ final).

❑Clarification par centrifugation, 50 000g à 4°C pendant 30 minutes (20 000rpm sur JA25.50) ; garder le surnageant ; mettre 100µL de côté pour le 1er gel (échantillon surnageant S); resuspendre le culot dans 20 mL solution 2% de SDS et garder 100µl pour déposer sur gel (échantillon culot C);. Filtrer le surnageant sur 0.45µ

❑Equilibration 8 mL de résine QIAGEN NiNTA neuve ou recyclée avec du tampon de Lyse NiNTA-XL4, puis incubation du surnageant de lysat 2h à 4°C, en batch avec agitation douce sur une roue. On peut aussi passer doucement le surnageant passer par gravité.

❑Récupération de la Fraction non-retenue (FT), laver avec 20 CV de Tampon de Lyse NiNTA-XL4, élution avec 2 à 3 volumes de colonne de Tampon d'Elution NiNTA-XL4 par fraction de 1mL ; garder le FT et les lavages jusqu'à la fin des premiers gels (L1, L2, …).

❑SDS-PAGE + Dosage Bradford : Déposer sur gel dans l'ordre : Poids Moléculaire (MW), 1µg de XRCC4-L4 et de XRCC4 purifié provenant d'une purification précédente, ET (extraits totaux), C (culot après lyse et centrifugation), S (surnageant de centrifugation), FT, Lavage, Fractions d'élutions.

❑Dialyser les fractions d'intérêt contre du tampon QA-XL4 3 fois 1h ou 1 fois 2h puis 1 fois sur la nuit à 4°C avec agitation. Mettre de côté 50µl de dialysat pour faire un Bradford, mesurer le rendement de l'étape de dialyse (rendement= masse de protéine après dialyse / masse de protéines avant dialyse).

Jour2

❑Equilibration d'une Ressource Q 6 mL (1mL à tester) avec du tampon QA-XL4 jusqu'à obtenir une ligne de base en absorbance stable. Si possible, activer la détection des UV à 280 et 260nm.

❑Passage du dialysat sur la Ressource Q 0,5mL.min^{-1} avec du tampon QA-XL4, collecte du FT, jusqu'à retour à la ligne de base de l'absorbance.

❑Elution avec un gradient de 0 à 50% tampon QB-XL4 en 120 minutes à 2ml.min^{-1}, collecte de fractions de 1.5mL. Deux pics sont attendus, un premier pic de mélange XRCC4-LigaseIV, puis un second pic de XRCC4.

❑SDS-PAGE : Faire un gel en déposant dans l'ordre, protéines dialysées et chargée sur la Ressource Q, FT, les fractions correspondant aux deux pics

❑Selon le résultat de l'analyse par SDS-PAGE, regrouper les fractions en séparant bien les deux pics ; Faire un spectre sur le nanodrop des fractions regroupées. Selon la quantité obtenue et le gel, selon la concentration des différents lots, envisager une concentration par ultrafiltration sur amicon ultra 10 kDa de 4mL ou de 15mL selon le volume. La concentration idéale pour des stocks intermédiaires est de 0,5mg/mL dans un tampon adapté aux utilisations ultérieures

❑Congeler par aliquotes de 500µL avec 10% de glycérol en notant s'il s'agit d'XRCC4-Ligase IV ou du second pic, XRCC4. Congeler rapidement. Référencer le tube à l'aide d'un numéro de lot unique dont on fera mention dans le cahier de laboratoire.

❑Evaluer la qualité de la congélation en comparant un passage sur filtration sur gel avant et après congélation.

Commentaires:

Dosage de bradford en direct : Lors de l'étape de NiNTA, il est possible d'ajouter 150µL de Bradford 1X dans une plaque 96 puits et d'y déposer 5 à 10µL des échantillons issus de l'étape de lavage ou de l'étape d'élution au fur et à mesure de la purification. Cette manipulation est particulièrement utile pour savoir quand le lavage est terminé (le tampon de lavage colore autant le bradford en sortie qu'en entrée de colonne) et pour savoir quand arrêter de faire des fractions d'élution ainsi que pour identifier les fractions les plus concentrées.

Gradient de la ResourceQ: Les différentes purifications de ce complexe ont permis de constater que plus le gradient est réalisé lentement, plus il est aisé de séparer les deux espèces contenues dans l'échantillon, XRCC4-LigaseIV, et XRCC4 seul, qui s'accroche mieux sur la colonne car il présente un pI plus acide.

Annexe 5 : Purification Ku70/10His-Ku80 de cultures en cellules d'insecte

Matériel : ❑Protocole pour un culot de 200ml de cultures cellules Sf1 ou Hi5 infectées par des Baculovirus V1 pUCDM-Ku70/(10His-Nter)-Ku80 synthétique (Genescript)

❑Résine Ni-NTA en vrac Qiagen ; ❑Colonne Resource Q (Ge-Healthcare)

Prot	forme	Nom		MW	E280	E0,1%	PI	complexe
Ku80	fl, 10 His TEV Nter	Ku80_opt		84928	42860	0.505	5.81	Ku70/Ku80
Ku70	Fl	Ku70_opt		69843	42290	0.606	6.23	Ku70/Ku80

Protéines

❑**Ku80 :** (10his en Nter, site TEV souligné, coupure entre Q et G)

MHHHHHHHHHHE<u>NLYFQG</u>VRSGNKAAVVLCMDVGFTMSNSIPGIESPFEQAKKVITMFVQRQVFAENKDEIALVLFGTDGTDNPLSGGDQ
YQNITVHRHLMLPDFDLLEDIESKIQPGSQQADFLDALIVSMDVIQHETIGKKFEKRHIEIFTDLSSRFSKSQLDIIIHSLKKCDISLQFFLPFSLGKED
GSGDRGDGPFRLGGHGPSFPLKGITEQQKEGLEIVKMVMISLEGEDGLDEIYSFSESLRKLCVFKKIERHSIHWPCRLTIGSNLSIRIAAYKSILQER
VKKTWTVVDAKTLKKEDIQKETVYCLNDDDETEVLKEDIIQGFRYGSDIVPFSKVDEEQMKYKSEGKCFSVLGFCKSSQVQRRFFMGNQVLKVF
AARDDEAAAVALSSLIHALDDLDMVAIVRYAYDKRANPQVGVAFPHIKHNYECLVYVQLPFMEDLRQYMFSSLKNSKKYAPTEAQLNAVDALI
DSMSLAKKDEKTDTLEDLFPTTKIPNPRFQRLFQCLLHRALHPREPLPPIQQHIWNMLNPPAEVTTKSQIPLSKIKTLFPLIEAKKKDQVTAQEIFQ
DNHEDGPTAKKLKTEQGGAHFSVSSLAEGSVTSVGSVNPAENFRVLVKQKKASFEEASNQLINHIEQFLDTNETPYFMKSIDCIRAFREEAIKFSE
EQRFNNFLKALQEKVEIKQLNHFWEIVVQDGITLITKEEASGSSVTAEEAKKFLAPKDKPSGDTAAVFEEGGDVDDLLDMI*

Number of amino acids: 749; **Molecular weight**: 84927.8Da; **Theoretical** pI: 5.81

❑**Ku70** (sans étiquette)

MSGWESYYKTEGDEEAEEEQEENLEASGDYKYSGRDSLIFLVDASKAMFESQSEDELTPFDMSIQCIQSVYISKIISSDRDLLAVVFYGTEKDKNS
VNFKNIYVLQELDNPGAKRILELDQFKGQQGQKRFQDMMGHGSDYSLSEVLWVCANLFSDVQFKMSHKRIMLFTNEDNPHGNDSAKASRA
RTKAGDLRDTGIFLDLMHLKKPGGFDISLFYRDIISIAEDEDLRVHFEESSKLEDLLRKVRAKETRKRALSRLKLKLNKDIVISVGIYNLVQKALKPPPI
KLYRETNEPVKTKTRTFNTSTGGLLLPSDTKRSQIYGSRQIILEKEETEELKRFDDPGLMLMGFKPLVLLKKHHYLRPSLFVYPEESLVIGSSTLFSAL
LIKCLEKEVAALCRYTPRRNIPPYFVALVPQEEELDDQKIQVTPPGFQLVFLPFADDKRKMPFTEKIMATPEQVGKMKAIVEKLRFTYRSDSFENP
VLQQHFRNLEALALDLMEPEQAVDLTLPKVEAMNKRLGSLVDEFKELVYPPDYNPEGKVTKRKHDNEGSGSKRPKVEYSEEELKTHISKGTLGK
FTVPMLKEACRAYGLKSGLKKQELLEALTKHFQD*

Number of amino acids: 609; Molecular weight: 69843.0Da; Theoretical pI: 6.23

Tampons

Tampons Lyse NiNTA-Ku :
20mM Tris pH8
75mM KCl
425mM NaCl
10 mM β-Mercaptol
40mM Imidazole

Tampon lavage NiB :
20 mM Tris pH 8.0
150 mM KCl
850 mM NaCl
10 mM β-Mercapto
40 mM Imidazole

Tampon Elution NiNTA-Ku :
20mM Tris pH8
75mM KCl
425mM NaCl
10 mM β-Mercapto
300mM Imidazole

Tampon QA-Ku :
25mM Tris pH8
100mM KCl
10 mM β-Mercapto

Tampon QB-Ku :
25mM Tris pH8
150mM KCl
850mM NaCl
10 mM β-Mercapto

Préparation:

- Préparer la résine Ni- ; Préparer les différents tampons. (voir protocole XRCC4-Ligase4)
- Faire un passage à blanc sur la Ressource Q 6mL (voir protocole XRCC4-Ligase4)

Jour 1

❑Reprise d'un culot de 200mL de culture dans 25mL (au 1/8^{e}) de tampon de Lyse NiNTA-Ku sur la glace, ajout d'inhibiteurs de protéases (Pepstatine (1µg/ml) / Chymostatine (1µg/ml) / Leupeptine (0,5µg/ml) / AEBSF ou PMSF (1mM) / Benzamidine (34µg/ml) final). Garder 10µL pour le premier gel SDS-PAGE à réaliser après l'étape de NiNTA (échantillon extrait total ET).

❑Après resuspension à la Propipette ou au Vortex, lyse par sonication (3x1 minute à 60% d'amplitude 1 sec On / 1 sec Off) en veillant à maintenir le lysat entre 0 et 4°C

❑Ajout de Benzaonase (2µl de Benzonase Merck source) + 1mM de $MgCl_2$ final, Incubation 15 min sur glace. Clarification par centrifugation, 50 000g à 4°C pendant 30 minutes (20 000rpm sur JA25.50) ; garder

le surnageant ; mettre 50µL de côté pour le 1er gel (échantillon surnageant S); resuspendre le culot dans 20 mL solution 2% de SDS et garder 50µl pour déposer sur gel (échantillon culot C). Filtrer le lysat à 0.45µ puis 0.22µ si le lysat est trop dense. (garder 10 µL avant chaque filtration pour vérifier si on perd du matériel)

❑Equilibration de 12 mL de résine QIAGEN NiNTA neuve ou recyclée avec du tampon de Lyse NiNTA-Ku, Charger la colonne avec le lysat clarifié et laisser passer par gravité ; Récupération du Flow Though (FT), laver avec 5 CV de Tampon de Lyse NiNTA-Ku puis laver avec 5 CV de tampon de lavage NiB, ré-équilibrer avec 2 CV de tampon NiA. Elution avec 2 à 3 volumes de colonne de Tampon d'Elution NiNTA-Ku par fraction de 1mL garder le FT et les lavages jusqu'à la fin des premiers gels (L1, L2, …) et jusqu'à ce que la quantité attendue de protéine soit obtenue Sinon repasser le FT sur la colonne ré-équilibrée.

❑SDS-PAGE + Dosage Bradford ; Déposer sur gel dans l'ordre : Poids Moléculaire (MW), 1µg de Ku purifié provenant d'une purification précédente, ET (extraits totaux), C (culot après lyse et centrifugation), S (surnageant de centrifugation), FT, Lavage, Fractions d'élutions

❑Dialyser les fractions d'intérêt contre du tampon QA-Ku 3 fois 2h ou 1 fois 2h puis 1 fois sur la nuit à 4°C avec agitation. Mettre de côté 50µl de dialysat pour faire un Bradford, mesurer le rendement de l'étape de dialyse

Jour 2

❑Equilibration d'une Ressource Q 6 mL avec du tampon QA-Ku jusqu'à obtenir une ligne de base en absorbance stable. Si possible, activer la détection des UV à 280 et 260nm. Auparavant, un run à blanc aura été effectué (voir préparation)

❑Passage du dialysat sur la Ressource Q à 1mL.min^{-1} avec du tampon QA-Ku, collecte du FT, jusqu'à retour à la ligne de base de l'absorbance. ; Elution avec un gradient de 0 à 40% tampon QB-Ku en 20CV, collecte de fractions de 1.5mL.

❑SDS-PAGE : Faire un gel en déposant dans l'ordre, protéines dialysées et chargée sur la Ressource Q, FT, les fractions éluées d'intérêt et fractions avec de la DO à 280 ; Selon le résultat de l'analyse par SDS-PAGE, regrouper les fractions selon leur pureté et leur concentration ; Faire un spectre sur le nanodrop des fractions regroupées. Selon la quantité obtenue et le gel, envisager de garder le FT plus longtemps pour un éventuel nouveau passage sur colonne.

❑Selon la concentration des différents lots, envisager une concentration par ultrafiltration sur amicon ultra 10 kDa (30kDa à tester) de 4mL ou de 15mL selon le volume. La concentration idéale pour des stocks intermédiaires est de 1mg/mL dans un tampon adapté aux utilisations ultérieures

❑Congeler par aliquotes de 500µL. Congeler rapidement. Référencer le tube à l'aide d'un numéro de lot unique dont on fera mention dans le cahier de laboratoire. Evaluer la qualité de la congélation en comparant un passage sur filtration sur gel avant et après congélation.

Commentaires :

Gel SDS-PAGE de Ku70/Ku80 : Les analyses sur SDS-PAGE du complexe Ku70/Ku80 doivent être faites en dénaturant les échantillons plus longtemps que d'ordinaire (15-30minutes) et avec 2 à 3 fois plus de Laemmli que d'habitude et des échantillons par trop chargé en protéine (2-4µg de Ku70/Ku80). Sinon, Ku80 forme une trainée diffuse et on peut avoir l'impression d'avoir perdu cette sous-unité.

Fraction de Ku70/Ku80 présente dans le FT de Ni-NTA : Il a été observé qu'une partie significative de Ku70/Ku80 (environ 1/3 du total) passe dans le FT de Ni-NTA bien que nous soyons en excès de résine. Ce complexe peut être récupéré par un nouveau passage sur la colonne. Il reste à déterminer si cette seconde fraction est bien repliée.

Annexe 6 : Purification Cernunnos sans étiquette

Matériel : -Protocole pour un **culot de 200ml** de cultures cellules Sf1 (ou Hi5) infectées par des Baculovirus V1 pFL_CNN ; Colonne **HiTrapQ 5ml** (GE-Healthcare) ; **Resource S 6ml** (Ge-Healthcare)

		PM (Da)	Epsilon (M-1. Cm-1)	pI	Abs (1mg/ml)
Cernunnos	fl	33337	28335	5,65	0,84

>Cernunnos Full length
MEELEQGLLMQPWAWLQLAENSLLAKVFITKQGYALLVSDLQQVWHEQVDTSVVSQRAKELNKRLTAPPAAFLCHLDNLLRPLLKDAA
HPSEATFSCDCVADALILRVRSELSGLPFYWNFHCMLASPSLVSQHLIRPLMGMSLALQCQVRELATLLHMKDLEIQDYQESGATLIRDR
LKTEPFEENSFLEQFMIEKLPEACSIGDGKPFVMNLQDLYMAVTTQEVQVGQKHQGAGDPHTSNSASLQGIDSQCVNQPEQLVSSAPT
LSAPEKESTGTSGPLQRPQLSKVKRKKPRGLFS*
Number of amino acids: 299; **Molecular weight:** 33 337.2Da; **Theoretical pI:** 5.65

Tampons Lyse-Cer:
Tris pH=8, 20mM
NaCl 350mM /KCl 150mM
ßMercapto 10mM
Glycerol 10%

Tp-SA-Cer:
Tris pH 8, 20mM
NaCl 50mM/ KCl 50mM
ßMercapt10mM

Tp-SB-Cer:
Tris pH= 8 20mM
NaCl 850mM/ KCL 150mM
ßMercapto 10mM

Tampon de conservation-Cer:
Tris pH=8 20mM
NaCl 75mM / KCl 75mM
ßMercapto 10mM
Glycerol 10%

Methode

1- Reprise d'un culot de 200mL de culture dans 25mL de tampon de Lyse, ajout d'inhibiteurs de protéases (Pepstatine (1µg/ml) / Chymostatine (1µg/ml) / Leupeptine (0,5µg/ml) / AEBSF ou PMSF (1mM) / Benzamidine (34µg/ml) final) (en général, dilution des stocks au 1/1000). Garder 10µL pour le premier gel SDS-PAGE (échantillon extrait total ET).

2- Après re-suspension à la Propipette ou au Vortex, lyse par sonication (3x1 minute à 60% d'amplitude 1 sec On / 1 sec Off) en veillant à maintenir le lysat entre 0 et 4°C, ce qui peut être réalisé en plongeant le tube récepteur dans un bain de glace/carboglace/eau/éthanol autour du récipient contenant le lysat. Attention à ce que le lysat ne congèle pas.

3- Ajout de Benzonase (0,02U/µl de Benzonase Merck + 1mM de $MgCl_2$ final).

4- Clarification par centrifugation, 50 000g à 4°C pendant 30 minutes (20 000rpm sur JA25.50) ; garder le surnageant ; mettre 20µL de côté pour le 1er gel (échantillon surnageant S); resuspendre le culot dans 20 mL solution 2% de SDS et garder 100µl pour déposer sur gel (échantillon culot C);.

5- Filtrer le surnageant sur 0.22µ

6- Charger HitrapQ équilibrée en Tampon SA-Cer à 2ml/min avec l'extrait clarifié.

7- Laver en SA-Cer à 2ml/min jusqu'à retour à la ligne de base

8- Elution par gradient : 0-40% de tampon SB-Cer en 25CV à 2ml/min. La protéine se trouve dans le premier amas (sortie au environ de 30%).

9- Dialyser les fractions contre le tampon SA-Cer

10- Charger Resource S 6ml équilibrée en Tampon SA-Cer (20 à 30 ml) à 2ml/min.

11- Elution par gradient 0-40% de Tampon SA-Ceren 25 CV à 2ml/min. La protéine sort dans les premières fractions.

12- Dialyser contre tampon de conservation

13- Aliquoter et congeler en N2 liquide puis conserver à -80°C.

Annexe 7 : Purification Artemis 10His

Matériel : -Protocole pour un **culot de 600ml** de cultures cellules Sf1 infectées par des Baculovirus V1 pFL-Artemis ; Résine Nickel Ni-NTA(GE-Healthcare) ; **Resource Q 6ml** (Ge-Healthcare)

		PM (Da)	Epsilon (M-1. Cm-1)	pI	Abs (1mg/ml)
Artemis	FL, 10His Cter	79 808	74 260	5,96	0,93

>Artemis Full length

MSSFEGQMAEYPTISIDRFDRENLRARAYFLSHCHKDHMKGLRAPTLKRRLECSLKVYLYCSPVTKELLLTSPKYRFWKKRIISIEIETPTQIS
LVDEASGEKEEIVVTLLPAGHCPGSVMFLFQGNNGTVLYTGDFRLAQGEAARMELLHSGGRVKDIQSVYLDTTFCDPRFYQIPSREECLS
GVLELVRSWITRSPYHVVWLNCKAAYGYEYLFTNLSEELGVQVHVNKLDMFRNMPEILHHLTTDRNTQIHACRHPKAEEYFQWSKLPC
GITSRNRIPLHIISIKPSTMWFGERSRKTNVIVRTGESSYRACFSFHSSYSEIKDFLSYLCPVNAYPNVIPVGTTMDKVVEILKPLCRSSQSTEP
KYKPLGKLKRARTVHRDSEEEDDYLFDDPLPIPLRHKVPYPETFHPEVFSMTAVSEKQPEKLRQTPGCCRAECMQSSRFTNFVDCEESNS
ESEEEVGIPASLQGDLGSVLHLQKADGDVPQWEVFFKRNDEITDESLENFPSSTVAGGSQSPKLFSDSDGESTHISSQNSSQSTHITEQGS
QGWDSQSDTVLLSSQERNSGDITSLDKADYRPTIKENIPASLMEQNVICPKDTYSDLKSRDKDVTIVPSTGEPTTLSSETHIPEEKSLLNLST
NADSQSSSDFEVPSTPEAELPKREHLQYLYEKLATGESIAVKKRKCSLLDTHHHHHHHHHH*

Number of amino acids: 702; **Molecular weight:** 79 808Da; **Theoretical pI:** 5,96

Tampons **Tampons Lyse NiNTA-Artemis :**
20mM Tris pH8
75mM KCl
425mM NaCl
10 mM β-Mercaptol
40mM Imidazole

Tampon lavage NiB :
20 mM Tris pH 8.0
150 mM KCl
850 mM NaCl
10 mM β-Mercapto
40 mM Imidazole

Tampon Elution NiNTA-Artemis :
20mM Tris pH8
75mM KCl
425mM NaCl
10 mM β-Mercapto
300mM Imidazole

Tampon QA-Artemis :
25mM Tris pH8
100mM KCl
10 mM β-Mercapto

Tampon QB-Artemis :
25mM Tris pH8
150mM KCl
850mM NaCl
10 mM β-Mercapto

Préparation:

- Préparer la résine Ni- ; Préparer les différents tampons. (voir protocole XRCC4-Ligase4)
- Faire un passage à blanc sur la Ressource Q 6mL (voir protocole XRCC4-Ligase4)

Jour 1

❑Reprise d'un culot de 200mL de culture dans 25mL (au 1/8^{e}) de tampon de Lyse NiNTA-Artemis sur la glace, ajout d'inhibiteurs de protéases (Pepstatine (1µg/ml) / Chymostatine (1µg/ml) / Leupeptine (0,5µg/ml) / AEBSF ou PMSF (1mM) / Benzamidine (34µg/ml) final).

❑Après resuspension à la Propipette ou au Vortex, lyse par sonication (3x1 minute à 60% d'amplitude 1 sec On / 1 sec Off) en veillant à maintenir le lysat entre 0 et 4°C

❑Ajout de Benzaonase (2µl de Benzonase Merck source) + 1mM de $MgCl_2$ final, Incubation 15 min sur glace. Clarification par centrifugation, 50 000g à 4°C pendant 30 minutes (20 000rpm sur JA25.50) ; Filtrer le lysat à 0.45µ puis 0.22 µ si le lysat est trop dense.

❑Equilibration de 12 mL de résine QIAGEN NiNTA neuve ou recyclée avec du tampon de Lyse NiNTA-Artemis, Charger la colonne avec le lysat clarifié et laisser passer par gravité ; Récupération du Flow Though (FT), laver avec 5 CV de Tampon de Lyse NiNTA-Artemis puis laver avec 5 CV de tampon de lavage NiB, ré-équilibrer avec 2 CV de tampon NiA. Elution avec 2 à 3 volumes de colonne de Tampon d'Elution NiNTA-Artemis par fraction de 1mL

❑SDS-PAGE + Dosage Bradford ; Déposer sur gel dans l'ordre : Poids Moléculaire (MW), 1µg de Artemis purifié provenant d'une purification précédente, ET (extraits totaux), C (culot après lyse et centrifugation), S (surnageant de centrifugation), FT, Lavage, Fractions d'élutions

❑Dialyser les fractions d'intérêt contre du tampon QA-Artemis

Jour 2

❑Equilibration d'une Ressource Q 6 mL avec du tampon QA-Artemis jusqu'à obtenir une ligne de base en absorbance stable.

❑Passage du dialysat sur la Ressource Q à 1mL.min^{-1} avec du tampon QA-Artemis, collecte du FT, jusqu'à retour à la ligne de base de l'absorbance. ; Elution avec un gradient de 0 à 40% tampon QB-Artemis en 20CV, collecte de fractions de 1.5mL.

❑SDS-PAGE : Faire un gel en déposant dans l'ordre, protéines dialysées et chargée sur la Ressource Q, FT, les fractions éluées d'intérêt et fractions avec de la DO à 280 ;

❑Selon la concentration des différents lots, envisager une concentration par ultrafiltration sur amicon ultra 10 kDa (30kDa à tester) de 4mL ou de 15mL selon le volume. La concentration idéale pour des stocks intermédiaires est de 1mg/mL dans un tampon adapté aux utilisations ultérieures (typiquement, 20mM Tris HCl pH 7.5, 150mM NaCl, 10% Glycérol (non obligatoire), 1mM DTT ou 5mM 2-Me-EtOH). Pour cela, dialyser les lots avant de les concentrer contre le tampon en question.

BIBLIOGRAPHIE

Ahnesorg P, Smith P, & Jackson SP (2006) XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* 124(2):301-313.

Andres SN, Modesti M, Tsai CJ, Chu G, & Junop MS (2007) Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. *Mol Cell* 28(6):1093-1101.

Andres SN, Vergnes A, Ristic D, Wyman C, Modesti M, & Junop M (2012) A human XRCC4-XLF complex bridges DNA. *Nucleic Acids Res*.

Andres SN, Schellenberg MJ, Wallace BD, Tumbale P, & Williams RS (2015) Recognition and repair of chemically heterogeneous structures at DNA ends. *Environ Mol Mutagen* 56(1):1-21 .

Arosio D, Cui S, Ortega C, Chovanec M, Di Marco S, Baldini G, Falaschi A, & Vindigni A (2002) Studies on the mode of Ku interaction with DNA. *J Biol Chem* 277(12):9741-9748.

Arosio D, Costantini S, Kong Y, & Vindigni A (2004) Fluorescence anisotropy studies on the Ku-DNA interaction: anion and cation effects. *J Biol Chem* 279(41):42826-42835 .

Barford D, Takagi Y, Schultz P, & Berger I (2013) Baculovirus expression: tackling the complexity challenge. *Curr Opin Struct Biol* 23(3):357-364 .

Bassing CH & Alt FW (2004) The cellular response to general and programmed DNA double strand breaks. *DNA Repair (Amst)* 3(8-9):781-796.

Baudat F, Imai Y, & de Massy B (2013) Meiotic recombination in mammals: localization and regulation. *Nat Rev Genet* 14(11):794-806.

Bernstein NK, Williams RS, Rakovszky ML, Cui D, Green R, Karimi-Busheri F, Mani RS, Galicia S, Koch CA, Cass CE, Durocher D, Weinfeld M, & Glover JN (2005) The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Mol Cell* 17(5):657-670 .

Bieniossek C, Imasaki T, Takagi Y, & Berger I (2012) MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem Sci* 37(2):49-57.

Bieniossek C, Papai G, Schaffitzel C, Garzoni F, Chaillet M, Scheer E, Papadopoulos P, Tora L, Schultz P, & Berger I (2013) The architecture of human general transcription factor TFIID core complex. *Nature* 493(7434):699-702.

Blier PR, Griffith AJ, Craft J, & Hardin JA (1993) Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *J Biol Chem* 268(10):7594-7601.

Boiteux S & Jinks-Robertson S (2013) DNA repair mechanisms and the bypass of DNA damage in Saccharomyces cerevisiae. *Genetics* 193(4):1025-1064 .

Boskovic J, Rivera-Calzada A, Maman JD, Chacon P, Willison KR, Pearl LH, & Llorca O (2003) Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. *Embo J* 22(21):5875-5882 .

Boulton SJ & Jackson SP (1996) Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res* 24(23):4639-4648 .

Britton S, Coates J, & Jackson SP (2013) A new method for high-resolution imaging of Ku foci to decipher mechanisms of DNA double-strand break repair. *J Cell Biol* 202(3):579-595 .

Bryans M, Valenzano MC, & Stamato TD (1999) Absence of DNA ligase IV protein in XR-1 cells: evidence for stabilization by XRCC4. *Mutat Res* 433(1):53-58.

Buck D, Malivert L, de Chasseval R, Barraud A, Fondaneche MC, Sanal O, Plebani A, Stephan JL, Hufnagel M, le Deist F, Fischer A, Durandy A, de Villartay JP*, et al.* (2006) Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* 124(2):287-299.

Busso D, Peleg Y, Heidebrecht T, Romier C, Jacobovitch Y, Dantes A, Salim L, Troesch E, Schuetz A, Heinemann U, Folkers GE, Geerlof A, Wilmanns M*, et al.* (2011) Expression of protein complexes

using multiple Escherichia coli protein co-expression systems: a benchmarking study. *J Struct Biol* 175(2):159-170 .

Callebaut I, Moshous D, Mornon JP, & de Villartay JP (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res* 30(16):3592-3601.

Calsou P, Delteil C, Frit P, Drouet J, & Salles B (2003) Coordinated assembly of Ku and p460 subunits of the DNA-dependent protein kinase on DNA ends is necessary for XRCC4-ligase IV recruitment. *J Mol Biol* 326(1):93-103.

Charbonnier JB, Renaud E, Miron S, Le Du MH, Blouquit Y, Duchambon P, Christova P, Shosheva A, Rose T, Angulo JF, & Craescu CT (2007) Structural, thermodynamic, and cellular characterization of human centrin 2 interaction with xeroderma pigmentosum group C protein. *J Mol Biol* 373(4):1032-1046.

Costantini S, Woodbine L, Andreoli L, Jeggo PA, & Vindigni A (2007) Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair (Amst)* 6(6):712-722.

Cottarel J, Frit P, Bombarde O, Salles B, Negrel A, Bernard S, Jeggo PA, Lieber MR, Modesti M, & Calsou P (2013) A noncatalytic function of the ligation complex during nonhomologous end joining. *J Cell Biol* 200(2):173-186.

Critchlow SE, Bowater RP, & Jackson SP (1997) Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr Biol* 7(8):588-598.

Davies DR, Interthal H, Champoux JJ, & Hol WG (2003) Crystal structure of a transition state mimic for Tdp1 assembled from vanadate, DNA, and a topoisomerase I-derived peptide. *Chem Biol* 10(2):139-147 .

Davis AJ, Chen BP, & Chen DJ (2014) DNA-PK: a dynamic enzyme in a versatile DSB repair pathway. *DNA Repair (Amst)* 17:21-29 .

de Villartay JP, Shimazaki N, Charbonnier JB, Fischer A, Mornon JP, Lieber MR, & Callebaut I (2009) A histidine in the beta-CASP domain of Artemis is critical for its full in vitro and in vivo functions. *DNA Repair (Amst)* 8(2):202-208.

de Vries E, van Driel W, Bergsma WG, Arnberg AC, & van der Vliet PC (1989) HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex. *J Mol Biol* 208(1):65-78 .

DeFazio LG, Stansel RM, Griffith JD, & Chu G (2002) Synapsis of DNA ends by DNA-dependent protein kinase. *Embo J* 21(12):3192-3200.

Delarue M, Boule JB, Lescar J, Expert-Bezancon N, Jourdan N, Sukumar N, Rougeon F, & Papanicolaou C (2002) Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. *Embo J* 21(3):427-439 .

Dobbs TA, Tainer JA, & Lees-Miller SP (2010) A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA Repair (Amst)* 9(12):1307-1314 .

Egelman EH (2015) Three-dimensional reconstruction of helical polymers. *Arch Biochem Biophys* .

Ellenberger T & Tomkinson AE (2008) Eukaryotic DNA ligases: structural and functional insights. *Annu Rev Biochem* 77:313-338 .

Fell VL & Schild-Poulter C (2015) The Ku heterodimer: function in DNA repair and beyond. *Mutat Res Rev Mutat Res* 763:15-29 .

Feng L & Chen J (2012) The E3 ligase RNF8 regulates KU80 removal and NHEJ repair. *Nat Struct Mol Biol* 19(2):201-206 .

Fitzgerald DJ, Berger P, Schaffitzel C, Yamada K, Richmond TJ, & Berger I (2006) Protein complex expression by using multigene baculoviral vectors. *Nat Methods* 3(12):1021-1032.

Frank KM, Sekiguchi JM, Seidl KJ, Swat W, Rathbun GA, Cheng HL, Davidson L, Kangaloo L, & Alt FW (1998) Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature* 396(6707):173-177 .

Frit P, Barboule N, Yuan Y, Gomez D, & Calsou P (2014) Alternative end-joining pathway(s): bricolage at DNA breaks. *DNA Repair (Amst)* 17:81-97 .

Garces F, Pearl LH, & Oliver AW (2011) The structural basis for substrate recognition by mammalian polynucleotide kinase 3' phosphatase. *Mol Cell* 44(3):385-396 .

Garcia-Diaz M, Bebenek K, Krahn JM, Kunkel TA, & Pedersen LC (2005) A closed conformation for the Pol lambda catalytic cycle. *Nat Struct Mol Biol* 12(1):97-98 .

Gell D & Jackson SP (1999) Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. *Nucleic Acids Res* 27(17):3494-3502 .

Goodarzi AA, Yu Y, Riballo E, Douglas P, Walker SA, Ye R, Harer C, Marchetti C, Morrice N, Jeggo PA, & Lees-Miller SP (2006) DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *Embo J* 25(16):3880-3889.

Gouge J, Rosario S, Romain F, Poitevin F, Beguin P, & Delarue M (2015) Structural basis for a novel mechanism of DNA bridging and alignment in eukaryotic DSB DNA repair. *Embo J* 34(8):1126-1142 .

Grawunder U, Wilm M, Wu X, Kulesza P, Wilson TE, Mann M, & Lieber MR (1997) Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* 388(6641):492-495.

Grundy GJ, Rulten SL, Zeng Z, Arribas-Bosacoma R, Iles N, Manley K, Oliver A, & Caldecott KW (2013) APLF promotes the assembly and activity of non-homologous end joining protein complexes. *Embo J* 32(1):112-125 .

Grundy GJ, Moulding HA, Caldecott KW, & Rulten SL (2014) One ring to bring them all-The role of Ku in mammalian non-homologous end joining. *DNA Repair (Amst)* .

Gueneau E, Dherin C, Legrand P, Tellier-Lebegue C, Gilquin B, Bonnesoeur P, Londino F, Quemener C, Le Du MH, Marquez JA, Moutiez M, Gondry M, Boiteux S*, et al.* (2013) Structure of the MutLalpha C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. *Nat Struct Mol Biol* 20(4):461-468.

Hammel M, Yu Y, Fang S, Lees-Miller SP, & Tainer JA (2010) XLF Regulates Filament Architecture of the XRCC4.Ligase IV Complex. *Structure* 18(11):1431-1442.

Hammel M, Yu Y, Mahaney BL, Cai B, Ye R, Phipps BM, Rambo RP, Hura GL, Pelikan M, So S, Abolfath RM, Chen DJ, Lees-Miller SP*, et al.* (2010) Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. *J Biol Chem* 285(2):1414-1423 .

Hammel M, Rey M, Yu Y, Mani RS, Classen S, Liu M, Pique ME, Fang S, Mahaney BL, Weinfeld M, Schriemer DC, Lees-Miller SP, & Tainer JA (2011) XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair. *J Biol Chem* 286(37):32638-32650 .

Hartley KO, Gell D, Smith GC, Zhang H, Divecha N, Connelly MA, Admon A, Lees-Miller SP, Anderson CW, & Jackson SP (1995) DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell* 82(5):849-856 .

Hoeijmakers JH (2001) Genome maintenance mechanisms for preventing cancer. *Nature* 411(6835):366-374.

Holdgate GA (2001) Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics. *Biotechniques* 31(1):164-170.

Imasaki T, Calero G, Cai G, Tsai KL, Yamada K, Cardelli F, Erdjument-Bromage H, Tempst P, Berger I, Kornberg GL, Asturias FJ, Kornberg RD, & Takagi Y (2011) Architecture of the Mediator head module. *Nature*.

Jancarik J & Kim S-H (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *Journal of Applied Crystallography* 24(4):409-411.

Jette N & Lees-Miller SP (2015) The DNA-dependent protein kinase: A multifunctional protein kinase with roles in DNA double strand break repair and mitosis. *Prog Biophys Mol Biol* 117(2-3):194-205 .

Jiang W, Crowe JL, Liu X, Nakajima S, Wang Y, Li C, Lee BJ, Dubois RL, Liu C, Yu X, Lan L, & Zha S (2015) Differential Phosphorylation of DNA-PKcs Regulates the Interplay between End-Processing and End-Ligation during Nonhomologous End-Joining. *Mol Cell* 58(1):172-185 .

Junop MS, Modesti M, Guarne A, Ghirlando R, Gellert M, & Yang W (2000) Crystal structure of the Xrcc4 DNA repair protein and implications for end joining. *Embo J* 19(22):5962-5970.

Kagawa W, Kurumizaka H, Ishitani R, Fukai S, Nureki O, Shibata T, & Yokoyama S (2002) Crystal structure of the homologous-pairing domain from the human Rad52 recombinase in the undecameric form. *Mol Cell* 10(2):359-371 .

Kurosawa A & Adachi N (2010) Functions and regulation of Artemis: a goddess in the maintenance of genome integrity. *J Radiat Res* 51(5):503-509 .

Lee KJ, Jovanovic M, Udayakumar D, Bladen CL, & Dynan WS (2004) Identification of DNA-PKcs phosphorylation sites in XRCC4 and effects of mutations at these sites on DNA end joining in a cell-free system. *DNA Repair (Amst)* 3(3):267-276 .

Li Y, Chirgadze DY, Bolanos-Garcia VM, Sibanda BL, Davies OR, Ahnesorg P, Jackson SP, & Blundell TL (2008) Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ. *Embo J* 27(1):290-300 .

Lieber MR (2008) The mechanism of human nonhomologous DNA end joining. *J Biol Chem* 283(1):1-5.

Lieber MR (2010) NHEJ and its backup pathways in chromosomal translocations. *Nat Struct Mol Biol* 17(4):393-395.

Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 79:181-211.

Lindahl T & Nyberg B (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry* 11(19):3610-3618 .

Ma Y, Pannicke U, Schwarz K, & Lieber MR (2002) Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* 108(6):781-794 .

Mahajan KN, Nick McElhinny SA, Mitchell BS, & Ramsden DA (2002) Association of DNA polymerase mu (pol mu) with Ku and ligase IV: role for pol mu in end-joining double-strand break repair. *Mol Cell Biol* 22(14):5194-5202 .

Malivert L, Ropars V, Nunez M, Drevet P, Miron S, Faure G, Guerois R, Mornon JP, Revy P, Charbonnier JB, Callebaut I, & de Villartay JP (2010) Delineation of the Xrcc4-interacting region in the globular head domain of cernunnos/XLF. *J Biol Chem* 285(34):26475-26483.

Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, & Tong L (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 444(7121):953-956.

Mari PO, Florea BI, Persengiev SP, Verkaik NS, Bruggenwirth HT, Modesti M, Giglia-Mari G, Bezstarosti K, Demmers JA, Luider TM, Houtsmuller AB, & van Gent DC (2006) Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc Natl Acad Sci U S A* 103(49):18597-18602 .

Mimitou EP & Symington LS (2010) Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *Embo J* 29(19):3358-3369.

Mimori T & Hardin JA (1986) Mechanism of interaction between Ku protein and DNA. *J Biol Chem* 261(22):10375-10379 .

Moldovan GL, Pfander B, & Jentsch S (2007) PCNA, the maestro of the replication fork. *Cell* 129(4):665-679.

Moon AF, Garcia-Diaz M, Batra VK, Beard WA, Bebenek K, Kunkel TA, Wilson SH, & Pedersen LC (2007) The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair (Amst)* 6(12):1709-1725.

Moon AF, Garcia-Diaz M, Bebenek K, Davis BJ, Zhong X, Ramsden DA, Kunkel TA, & Pedersen LC (2007) Structural insight into the substrate specificity of DNA Polymerase mu. *Nat Struct Mol Biol* 14(1):45-53.

Mosesso P, Piane M, Palitti F, Pepe G, Penna S, & Chessa L (2005) The novel human gene aprataxin is directly involved in DNA single-strand-break repair. *Cell Mol Life Sci* 62(4):485-491 .

Moshous D, Callebaut I, de Chasseval R, Corneo B, Cavazzana-Calvo M, Le Deist F, Tezcan I, Sanal O, Bertrand Y, Philippe N, Fischer A, & de Villartay JP (2001) Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* 105(2):177-186.

Nick McElhinny SA, Kissling GE, & Kunkel TA (2010) Differential correction of lagging-strand replication errors made by DNA polymerases {alpha} and {delta}. *Proc Natl Acad Sci U S A* 107(49):21070-21075.

O'Driscoll M, Cerosaletti KM, Girard PM, Dai Y, Stumm M, Kysela B, Hirsch B, Gennery A, Palmer SE, Seidel J, Gatti RA, Varon R, Oettinger MA*, et al.* (2001) DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. *Mol Cell* 8(6):1175-1185 .

Ochi T, Wu Q, Chirgadze DY, Grossmann JG, Bolanos-Garcia VM, & Blundell TL (2012) Structural insights into the role of domain flexibility in human DNA ligase IV. *Structure* 20(7):1212-1222 .

Ochi T, Gu X, & Blundell TL (2013) Structure of the catalytic region of DNA ligase IV in complex with an Artemis fragment sheds light on double-strand break repair. *Structure* 21(4):672-679 .

Ochi T, Wu Q, & Blundell TL (2014) The spatial organization of non-homologous end joining: from bridging to end joining. *DNA Repair (Amst)* 17:98-109 .

Ochi T, Blackford AN, Coates J, Jhujh S, Mehmood S, Tamura N, Travers J, Wu Q, Draviam VM, Robinson CV, Blundell TL, & Jackson SP (2015) DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science* 347(6218):185-188 .

Ogawa T, Yu X, Shinohara A, & Egelman EH (1993) Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science* 259(5103):1896-1899 .

Pascal JM, O'Brien PJ, Tomkinson AE, & Ellenberger T (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* 432(7016):473-478.

Pellegrini L, Yu DS, Lo T, Anand S, Lee M, Blundell TL, & Venkitaraman AR (2002) Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature* 420(6913):287-293 .

Pelletier H, Sawaya MR, Kumar A, Wilson SH, & Kraut J (1994) Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science* 264(5167):1891-1903 .

Recuero-Checa MA, Dore AS, Arias-Palomo E, Rivera-Calzada A, Scheres SH, Maman JD, Pearl LH, & Llorca O (2009) Electron microscopy of Xrcc4 and the DNA ligase IV-Xrcc4 DNA repair complex. *DNA Repair (Amst)* 8(12):1380-1389.

Reid DA, Keegan S, Leo-Macias A, Watanabe G, Strande NT, Chang HH, Oksuz BA, Fenyo D, Lieber MR, Ramsden DA, & Rothenberg E (2015) Organization and dynamics of the nonhomologous end-joining machinery during DNA double-strand break repair. *Proc Natl Acad Sci U S A* .

Riballo E, Kuhne M, Rief N, Doherty A, Smith GC, Recio MJ, Reis C, Dahm K, Fricke A, Krempler A, Parker AR, Jackson SP, Gennery A*, et al.* (2004) A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to gamma-H2AX foci. *Mol Cell* 16(5):715-724.

Riballo E, Woodbine L, Stiff T, Walker SA, Goodarzi AA, & Jeggo PA (2009) XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucleic Acids Res* 37(2):482-492.

Ribes-Zamora A, Mihalek I, Lichtarge O, & Bertuch AA (2007) Distinct faces of the Ku heterodimer mediate DNA repair and telomeric functions. *Nat Struct Mol Biol* 14(4):301-307.

Roberts KP, Sobrino JA, Payton J, Mason LB, & Turesky RJ (2006) Determination of apurinic/apyrimidinic lesions in DNA with high-performance liquid chromatography and tandem mass spectrometry. *Chem Res Toxicol* 19(2):300-309 .

Roberts SA, Strande N, Burkhalter MD, Strom C, Havener JM, Hasty P, & Ramsden DA (2010) Ku is a 5'-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature* 464(7292):1214-1217 .

Ropars V, Drevet P, Legrand P, Baconnais S, Amram J, Faure G, Marquez JA, Pietrement O, Guerois R, Callebaut I, Le Cam E, Revy P, de Villartay JP*, et al.* (2011) Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining. *Proc Natl Acad Sci U S A* 108(31):12663-12668.

Roy S, Andres SN, Vergnes A, Neal JA, Xu Y, Yu Y, Lees-Miller SP, Junop M, Modesti M, & Meek K (2012) XRCC4's interaction with XLF is required for coding (but not signal) end joining. *Nucleic Acids Res*.

San Filippo J, Sung P, & Klein H (2008) Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* 77:229-257 .

Schatz DG & Swanson PC (2011) V(D)J recombination: mechanisms of initiation. *Annu Rev Genet* 45:167-202.

Schellenberg MJ, Appel CD, Adhikari S, Robertson PD, Ramsden DA, & Williams RS (2012) Mechanism of repair of 5'-topoisomerase II-DNA adducts by mammalian tyrosyl-DNA phosphodiesterase 2. *Nat Struct Mol Biol* 19(12):1363-1371 .

Schellenberg MJ, Tumbale PP, & Williams RS (2015) Molecular underpinnings of Aprataxin RNA/DNA deadenylase function and dysfunction in neurological disease. *Prog Biophys Mol Biol* .

Shirodkar P, Fenton AL, Meng L, & Koch CA (2013) Identification and functional characterization of a Ku-binding motif in aprataxin polynucleotide kinase/phosphatase-like factor (APLF). *J Biol Chem* 288(27):19604-19613 .

Sibanda BL, Critchlow SE, Begun J, Pei XY, Jackson SP, Blundell TL, & Pellegrini L (2001) Crystal structure of an Xrcc4-DNA ligase IV complex. *Nat Struct Biol* 8(12):1015-1019 .

Sibanda BL, Chirgadze DY, & Blundell TL (2010) Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. *Nature* 463(7277):118-121 .

Spagnolo L, Rivera-Calzada A, Pearl LH, & Llorca O (2006) Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Mol Cell* 22(4):511-519.

Strande N, Roberts SA, Oh S, Hendrickson EA, & Ramsden DA (2012) Specificity of the dRP/AP lyase of Ku promotes nonhomologous end joining (NHEJ) fidelity at damaged ends. *J Biol Chem* 287(17):13686-13693 .

Strande NT, Carvajal-Garcia J, Hallett RA, Waters CA, Roberts SA, Strom C, Kuhlman B, & Ramsden DA (2014) Requirements for 5'dRP/AP lyase activity in Ku. *Nucleic Acids Res* 42(17):11136-11143 .

Stura EA, Satterhwait AC, Calvo JC, Kaslow DC, & Wilson IA (1994) Reverse Screening. *Acta Cryst.* D50:448-455.

Subramanian VV & Hochwagen A (2014) The meiotic checkpoint network: step-by-step through meiotic prophase. *Cold Spring Harb Perspect Biol* 6(10):a016675 .

Sulzenbacher G, Gruez A, Roig-Zamboni V, Spinelli S, Valencia C, Pagot F, Vincentelli R, Bignon C, Salomoni A, Grisel S, Maurin D, Huyghe C, Johansson K*, et al.* (2002) A medium-throughput crystallization approach. *Acta Crystallogr D Biol Crystallogr* 58(Pt 12):2109-2115.

Tomkinson AE, Howes TR, & Wiest NE (2013) DNA ligases as therapeutic targets. *Transl Cancer Res* 2(3) .

Tsai CJ, Kim SA, & Chu G (2007) Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc Natl Acad Sci U S A* 104(19):7851-7856.

Tsai CJ & Chu G (2013) Cooperative assembly of a protein-DNA filament for nonhomologous end joining. *J Biol Chem* 288(25):18110-18120 .

Tumbale P, Williams JS, Schellenberg MJ, Kunkel TA, & Williams RS (2014) Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity. *Nature* 506(7486):111-115 .

van der Burg M, Ijspeert H, Verkaik NS, Turul T, Wiegant WW, Morotomi-Yano K, Mari PO, Tezcan I, Chen DJ, Zdzienicka MZ, van Dongen JJ, & van Gent DC (2009) A DNA-PKcs mutation in a radiosensitive T-B- SCID patient inhibits Artemis activation and nonhomologous end-joining. *J Clin Invest* 119(1):91-98 .

Vaughn JL, Goodwin RH, Tompkins GJ, & McCawley P (1977) The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). *In Vitro* 13(4):213-217 .

Walker JR, Corpina RA, & Goldberg J (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412(6847):607-614.

Wang H, Perrault AR, Takeda Y, Qin W, & Iliakis G (2003) Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Res* 31(18):5377-5388 .

Wang J, Dong X, & Reeves WH (1998) A model for Ku heterodimer assembly and interaction with DNA. Implications for the function of Ku antigen. *J Biol Chem* 273(47):31068-31074 .

West RB, Yaneva M, & Lieber MR (1998) Productive and nonproductive complexes of Ku and DNA-dependent protein kinase at DNA termini. *Mol Cell Biol* 18(10):5908-5920 .

Williams DR, Lee KJ, Shi J, Chen DJ, & Stewart PL (2008) Cryo-EM structure of the DNA-dependent protein kinase catalytic subunit at subnanometer resolution reveals alpha helices and insight into DNA binding. *Structure* 16(3):468-477.

Woodbine L, Gennery AR, & Jeggo PA (2014) Reprint of "The clinical impact of deficiency in DNA non-homologous end-joining". *DNA Repair (Amst)* 17:9-20 .

Wu PY, Frit P, Meesala S, Dauvillier S, Modesti M, Andres SN, Huang Y, Sekiguchi J, Calsou P, Salles B, & Junop MS (2009) Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4. *Mol Cell Biol* 29(11):3163-3172.

Wu Q, Ochi T, Matak-Vinkovic D, Robinson CV, Chirgadze DY, & Blundell TL (2011) Non-homologous end-joining partners in a helical dance: structural studies of XLF-XRCC4 interactions. *Biochem Soc Trans* 39(5):1387-1392, suppl 1382 p following 1392 .

Xing M, Yang M, Huo W, Feng F, Wei L, Jiang W, Ning S, Yan Z, Li W, Wang Q, Hou M, Dong C, Guo R*, et al.* (2015) Interactome analysis identifies a new paralogue of XRCC4 in non-homologous end joining DNA repair pathway. *Nat Commun* 6:6233 .

Zakharyevich K, Tang S, Ma Y, & Hunter N (2012) Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase. *Cell* 149(2):334-347.

ETUDE STRUCTURALE ET FONCTIONNELLE DES COMPLEXES MULTI-PROTEIQUES DE LA VOIE DE REPARATION NHEJ CHEZ L'HOMME

Résumé :

La voie de réparation NHEJ (Non-Homologous End-Joining) est une voie majeure de réparation des cassures double-brin chez l'homme. Les protéines de cette voie interagissent et forment des complexes dynamiques dont les mécanismes moléculaires sont encore largement méconnus. Nous avons dans un premier temps mis au point des protocoles de production à l'échelle de plusieurs milligrammes des protéines cœur de la voie NHEJ en cellules d'insecte à l'aide du système MultiBac. Nous avons ainsi purifié les complexes Ku70/Ku80 et Ligase4/XRCC4 et les protéines Cernunnos et Artemisà homogénéité. Des essais de cristallisation, des études par SAXS et des analyses par microscopie électronique ont été réalisés sur différents complexes formés par ces protéines cœur du NHEJ. Nous avons également caractérisé par chromatographie d'exclusion de taille et calorimétrie, les interactions effectuées entre les protéines de la voie NHEJ. L'ensemble de ces travaux a permis d'établir des bases biochimiques solides en vue des études structurales et fonctionnelles de la voie NHEJ chez l'homme.

STRUCTURAL AND FUNCTIONAL STUDIES OF THE MULTI-PROTEIN COMPLEXES OF THE DNA REPAIR NHEJ PATHWAY IN HUMAN

Summary :

The DNA repair pathway NHEJ (Non-Homologous End-Joining) is a major pathway of double-strand breaks repair in human. The proteins involved in this pathway interact and form dynamic complexes with molecular mechanisms largely unknown. Firstly, we established protocols to purify milligrams of the core proteins of the NHEJ pathway by using the MultiBac system for insect cells production. We succeeded in purifying the Ku70/Ku80 and Ligase4/XRCC4 complexes and the Artemis and Cernunnos proteins in milligram amounts and with high degree of purity. Crystallization screenings, SAXS experiments and Transmission Electronic Microscopy experiments have been performed on several complexes formed by these core NHEJ proteins. We also characterized the interactions between these proteins by Size Exclusion Chromatography and Isothermal Calorimetry. This PhD study has opened the route for structural and functional studies of the NHEJ multi-protein complexes thank to robust biochemical protocols.