

Liste des abréviations

ADME/Tox: Absorption-Distribution-Metabolism-Excretion/Toxicity
ADN: Acide DésoxyriboNucléique
AMBER: Assisted Model Building with Energy Refinement
ARN: Acide RiboNucléique
ATP: Adenosine Triphosphate
Bp: paire de base
BRCA: breast cancer associated
BRCT: BRCA1 C-terminal
CDB: Cassure double-brin
Cer-XLF: Cernunnos-XRCC4-Like Factor
COSY: Correlation Spectroscopy
CPU: Central Processing Unit
CSB: Cassure simple-brin
CSM: Chemical Shift Mapping
C-ter: C-terminus
D2O: Deuterium Oxide
DBD: DNA Binding Domain
DC: Dichroisme Circulaire
DDR: DNA Damage Response
DMSO: DiMethyl SulfOxide
DNA-PK: DNA-dependent protein kinase
dNTP: désoxyriboNucléotide Tri-Phosphate
DO_{600nm}: Densité optique à 600nm
DSF: Differential Scanning Fluorimetry
DTT: DiThioThreitol
EDTA: Ethylene Diamine Tetra acetic Acid
GAFF: General Amber Force Field
HR: Homologous Recombination
HTS: High-Throughput Screening
IC₅₀: half maximal inhibitory concentration
IPTG: IsoPropyl-β-ThioGalactopyranoside
ITC: Isothermal Titration Calorimetry
Kd: dissociation constant
kDa: kilodalton
KO: knock-out
LB: Luria Bertani
Lig4: DNA Ligase IV
MD: Molecular Dynamics
NDSB-195: Non Detergent SulfoBetain-195
NHEJ: Non-Homologous End Joining

Ni-NTA: Nickel-NitriloTriacetic Acid

NOE: Nuclear Overhauser Effect

NOESY: Nuclear Overhauser Effect SpectroscopY

PAGE: PolyAcrylamide Gel Electrophoresis

PCR: Polymerase Chain Reaction

PDB: Protein Data Bank

PMSF: PhenylMethylSulfonyl Fluoride

RCS: Relaxed Complex Scheme

RI : Radiation Ionisante

RMN: Resonance Magnetique Nucléaire

RMSD: Root Mean Square Deviation

SAR: Structure-Activity Relationship

SD: Standard Deviation

SFI: Solubility Forecast Index

STD: Saturation Transfer Difference

TBE: Tris Borate EDTA

TOCSY: Total Correlation SpectroscopY

TrNOESY: Transferred Nuclear Overhauser Effect SpectroscopY

UV: UltraViolet

XIR: XRCC4 Interacting Region

XRCC4: X-ray Repair Cross Complementing group 4

Table des matières

Résumé

Liste des abréviations

Table des matières

Partie I : Chapitre introductif

I – 1 : Les cassures double-brin de l'ADN et leur réparation par NHEJ

I – 2 : Inhibition des voies de réparation des cassures double brin de l'ADN pour une thérapie combinée anti-cancéreuse

I – 3 : Inhibition cellulaire des protéines du complexe XXL de ligature et radiosensibilisation des cellules cancéreuses – Preuves de concept

I – 4 : Les interactions Protéine-Protéine : un défi thérapeutique

I – 5 : Criblage de molécules inhibitrices, dirigées contre les complexes XRCC4/Cer-XLF et XRCC4/C-ter Ligase 4 – Analyse du système d'étude, étude de la faisabilité du projet et de la démarche expérimentale

Partie II : Résultats et discussion

II – 1 : Production des protéines du complexe de ligature et tests de fonctionnalité

II – 1.1 : Expression en système bactérien et purification des protéines XRCC4, Cer-XLF et C-terminal de la Ligase 4 du complexe XXL de réparation de l'ADN

II – 1.2 : Analyse de la thermostabilité et des interactions des protéines purifiées par DSF (Fluorimétrie différentielle à balayage) et gel natif

II – 1.3 : Analyse de la fonctionnalité du complexe XRCC4/Cer-XLF par des tests de stimulation de la ligase T4

II – 2 : Inhibition de l’interaction XRCC4/Cer-XLF

II – 2.1 : Simulation de dynamique moléculaire sur le complexe XRCC4/Cer-XLF

II – 2.2 : Recherche de sites « druggables » sur les protéines XRCC4 et Cer-XLF

II – 2.3 : Echantillonnage conformationnel pour le criblage virtuel –
Simulations de dynamique moléculaire sur la tête XRCC4 et analyse des centroïdes de chaque cluster

II – 2.4 : Préparation et filtrage d’une base de données virtuelle de molécules

II – 2.5 : Docking sur les têtes XRCC4 et Cer-XLF – Tri et analyse des hits potentiels

II – 2.6 : Analyse par dynamique moléculaire de la stabilité et de l’énergie de liaison des complexes XRCC4 hit et Cer-XLF hit sélectionnés –
Modèles d’interaction

II – 2.7 : Criblage haut-débit : principe, mise en œuvre et résultats obtenus

II – 2.8 : Analyse par DLS (Diffusion dynamique de la lumière) et RMN ^1H de la solubilité et intégrité des molécules retenues

II – 2.9 : Validation biophysique des interactions protéine-ligand par DSF et STD (Différence de Transfert de saturation)-NMR

II – 2.10 : Approche peptidique : caractérisation et analyse d’interaction par RMN des peptides XRCC4 et Cer-XLF

II – 2.11 : Validation des inhibiteurs potentiels par des tests de stimulation de la ligase T4

II – 2.12 : Validation des inhibiteurs potentiels sur des essais NHEJ avec extraits cellulaires

II – 2.13 : Clonage et production de la partie N-terminale de XRCC4 (1-112 ; 1-119)

II – 2.14 : Cristallisation de la protéine XRCC4 – Trempage de cristaux

II – 2.15 : Expériences de cocristallisation

II – 2.16 : Conclusions et perspectives – Vers la structure d'un complexe XRCC4/inhibiteur

II – 3 : Inhibition de l'interaction XRCC4/C-terminal de la Ligase 4

II – 3.1 : Simulations de dynamique moléculaire sur le complexe XRCC4/C-ter Ligase 4

II – 3.2 : Dynamique moléculaire et échantillonnage conformationnel de la structure clamp en solution

II – 3.3 : Préparation et filtrage d'une base de données virtuelle de molécules

II – 3.4 : Docking sur les différentes structures du clamp – Tri et analyse des hits potentiels

II – 3.5 : Analyse par dynamique moléculaire de la stabilité et de l'énergie de liaison des complexes C-ter Ligase 4 hit sélectionnés – Nouveaux classements et modèles d'interaction

II – 3.6 : Analyse par DLS et RMN ^1H de la solubilité et intégrité des molécules virtuelles retenues

II – 3.7 : Tests d'interaction C-ter ligase 4-ligand par DSF et STD-NMR

II – 3.8 : Test de l'effet préventif des molécules sur l'association XRCC4/C-ter Ligase 4

II – 3.9 : Modèle d'interaction entre le C-ter ligase 4 et l'inhibiteur #3101

II – 3.10 : Conclusions et perspectives – Vers la structure d'un complexe C-ter Ligase4/inhibiteur

Conclusion générale et Perspectives

Article « Structure-Based Virtual Ligand Screening on the XRCC4/DNA Ligase IV Interface »

Matériels et Méthodes

Minirevue « Looking for human DNA ligases inhibitors for combination anticancer therapy »

Bibliographie

Rapport-Gratuit.com

Partie I :

Chapitre introductif

I – 1 : Les cassures double-brin de l'ADN et leur réparation par NHEJ

Origine des cassures double-brin de l'ADN et réponse cellulaire associée

La molécule d'ADN (Acide DésoxyriboNucléique) porteuse de notre code génétique, est le biopolymère central de la plupart des organismes vivants. L'ADN est copié et distribué de manière fidèle entre les différentes cellules en division mais il est constamment soumis à des agressions endo ou exogènes, pouvant mener à des lésions voire à de graves pathologies si ces dernières ne sont pas correctement réparées. Notre organisme est exposé chaque jour en moyenne à près d'une dizaine de milliers de ces lésions d'ADN par cellule (Ciccia and Elledge, 2010; Lindahl et al., 1993). Elles peuvent être des modifications de sucres ou de bases, des pontages covalents entre les deux chaînes de l'ADN ou intra-chaîne, des pontages entre l'ADN et des protéines ou bien des cassures simple-brin (CSB) ou double-brin (CDB).

Les origines des CDB sont très diverses. Elles se forment lorsque deux CSB sont à proximité l'une de l'autre et sur des brins d'ADN opposés. Les origines intracellulaires peuvent d'être d'ordre physiologique et donc programmées (lors de la recombinaison V(D)J, de la commutation isotypique ou même de la méiose) ou d'ordre spontané et non programmé (au cours par exemple des accidents de réPLICATION ou par l'action d'espèces réactives de l'oxygène (ROS) produits lors de la respiration cellulaire oxidative). Les origines exogènes sont due essentiellement aux rayonnements ionisants (rayons X ou gamma) et à l'effet d'agents chimiques (comme les molécules radiomimétiques ou les inhibiteurs d'ADN topoisomérase). Cet effet exogène et l'extrême toxicité des CDB est d'ailleurs volontairement exploité lors des stratégies de radio- et chimiothérapies anti-cancéreuses (Connell et al., 2004; Jeggo and Lavin, 2009). De plus, plusieurs techniques de correction du génome (nucléases à doigt de zinc, TALENs ou CRISPR-Cas) induisent des intermédiaires de type CDB qui doivent être réparés par la cellule (Cox et al., 2015).

Bien que ces CDB ne sont pas aussi fréquentes que les autres types de dommages à l'ADN rencontrés (environ 10 CDB/cellule/jour), elles restent les lésions les plus cytotoxiques et difficiles à réparer. Une seule CDB peut s'avérer létale pour la cellule (Blocher and Pohlit, 1982). La réparation est donc indispensable à la viabilité cellulaire et assure la transmission du patrimoine génétique. En effet, les CDB non

ou mal réparées peuvent mener à la mort cellulaire, au blocage de la croissance cellulaire par sénescence, ou bien au maintien de la croissance cellulaire mais accompagnée d'une instabilité génomique (aberrations chromosomiques, délétions, translocations) pouvant conduire à la cancérisation par activation de proto-oncogènes et/ou inactivation de gènes suppresseurs de tumeurs (Durante et al., 2013).

Suite à l'induction d'une CDB (comme pour d'autres types de dommages à l'ADN), une réponse cellulaire immédiate se met en place, suivie d'une réponse plus tardive. Cette réponse est appelée la Réponse aux Dommages à l'ADN (DDR). Une véritable cascade de transduction de signaux (déTECTEURS, médiateurs, effecteurs) permet la détection de ces cassures, la signalisation de leur présence et leur réparation. Cette cascade mène également à une activation de points de contrôle spécifiques permettant ainsi l'arrêt de la machinerie transcriptionnelle et de division cellulaire pour une réparation des dommages (Ciccia and Elledge, 2010; Jackson and Bartek, 2009; Shiloh and Ziv, 2013). Ce programme est d'ailleurs commun à tous les types de dommages (Figure 1).

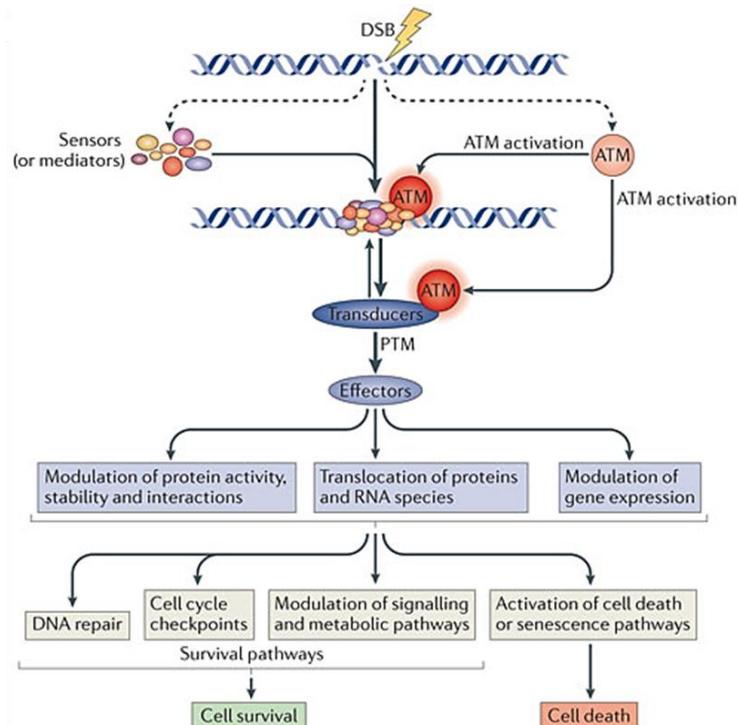


Figure 1 : Schéma de la réponse aux dommages à l'ADN, suite à l'induction d'une cassure double-brin (Shiloh et al, 2013).

Si ces dommages sont trop nombreux, une mort cellulaire programmée par apoptose ou autophagie se met alors en place (Okada and Mak, 2004; Surova and Zhivotovsky, 2013). Les cellules dont la DDR est défectueuse ont généralement une sensibilité moindre aux agents de dommage à l'ADN pouvant alors mener au cancer *via* la réparation cellulaire, malgré une réparation fautive de ces CDB. En effet, l'observation d'un défaut de DDR dans plusieurs cellules cancéreuses souligne la connexion entre défaut de réparation des CDB et carcinogénèse (Bartkova et al., 2005; Gorgoulis et al., 2005). Certaines formes de cancer présentent des mutations ou des altérations dans l'expression de gènes codant pour des protéines impliquées dans le métabolisme de l'ADN (Connell et al., 2006). De plus, dans certains cancers, certaines protéines de réparation peuvent être surexprimées ou bien complètement absentes, modulant ainsi la réponse aux agents de dommage à l'ADN (Rassool and Tomkinson, 2010; Tobin et al., 2012) (Bouwman et al., 2012).

Dans les cellules eucaryotes, deux voies majeures de réparation des CDB peuvent se mettre en place : la recombinaison homologue (RH) et la jonction d'extrémités non-homologue (NHEJ) (Goodarzi and Jeggo, 2013; Thompson, 2012). Le mécanisme de NHEJ est un système de réparation conservé au cours de l'évolution des bactéries à l'homme et qui permet une religation directe des deux segments d'ADN cassé (Lieber, 2010; Wang and Lees-Miller, 2013; Waters et al., 2014). Lorsque les extrémités sont cohésives ou franches, les cassures sont religaturées de façon fidèle, avec une perte minimale d'information (Betermier et al., 2014; Guirouilh-Barbat et al., 2004). Elle est la voie majeure de réparation des CDB (réparation d'environ 70% des CDB ; (Guirouilh-Barbat et al., 2004). Elle intervient lors de toutes les étapes du cycle cellulaire car elle ne nécessite aucune homologie de séquence, à l'inverse de la RH qui se déroule seulement durant les phases S et G2 du cycle cellulaire lorsqu'une matrice homologue (*via* la chromatide sœur) est disponible. Elle intervient dans la réparation des CDB pathologiques mais aussi dans la réparation de certaines CDB programmées physiologiques (recombinaison V(D)J et commutation isotypique). Ainsi, les patients dont le mécanisme NHEJ est défectueux sont en général fortement immunodéficients par défaut de développement des lymphocytes du fait de la non-résolution des CDB programmées, et sensibles aux

rayonnements ionisants par défaut de réparation des CDB pathologiques (RS-SCID) (Woodbine et al., 2014).

Le mécanisme de la C-NHEJ

Nous ne développerons pas ici le mécanisme de réparation des CDB par la NHEJ alternative (Alt-NHEJ), qui est une voie de ligature de secours (Deriano and Roth, 2013; Frit et al., 2014). Nous centrerons notre étude sur les aspects fonctionnels et structuraux des protéines intervenant dans le mécanisme de la NHEJ classique (C-NHEJ). Ce mécanisme se découpe en quatre étapes distinctes et séquentielles :

- La reconnaissance de la cassure ainsi que l'assemblage et la stabilisation du complexe NHEJ au site de cassure,
- Le rapprochement et la stabilisation des deux extrémités de la cassure,
- La maturation des extrémités (élimination des extrémités non ligaturables ou dommages),
- La ligation finale des deux extrémités par le complexe XXL de ligature.

Le modèle séquentiel de ce mécanisme dans la réparation des CDB est présenté dans la figure ci-dessous (Figure 2). Cependant, ce modèle est remis en question par des données plus récentes qui favorisent plutôt un recrutement concomitant de toutes les protéines de C-NHEJ amorcé par la liaison de Ku à l'ADN et conduisant à la formation d'un macro-complexe protéique (Cottarel et al., 2013; Yano et al., 2009).

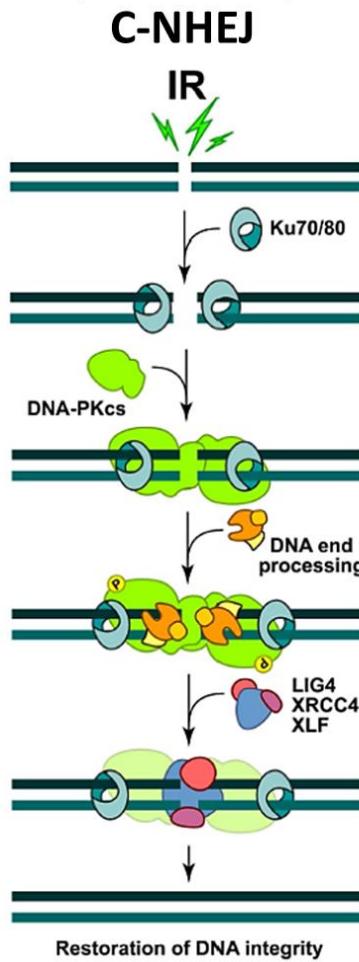


Figure 2 : Schéma de la voie NHEJ classique de réparation des cassures double-brin de l'ADN (Mladenov et al, 2013).

Après induction de CDB (par exemple après rayonnements ionisants), le site est rapidement reconnu par l'hétérodimère Ku70/Ku80. Cet hétérodimère (complexe Ku) permet le recrutement et l'activation de la sous-unité catalytique de la DNA-PK (protéine kinase dépendante de l'ADN). Cette enzyme, couplée à Ku et à l'ADN devient alors pleinement fonctionnelle et va avoir plusieurs rôles (Jette and Lees-Miller, 2015). Elle possède une activité sérine/thréonine kinase indispensable au reste de la réaction (Mahaney et al., 2009). Elle stabilise et protège les deux extrémités d'ADN (Meek et al., 2008). Ce rapprochement des deux extrémités est appelé synapse (DeFazio et al., 2002). La DNA-PK sert de plateforme à l'amarrage de certaines enzymes, dont les enzymes de maturation des extrémités de l'ADN. En effet, les extrémités induites par des agents génotoxiques sont souvent modifiées (site abasique, 3'-phospho-glycolate, complexe ADN-protéine...) et par conséquent,

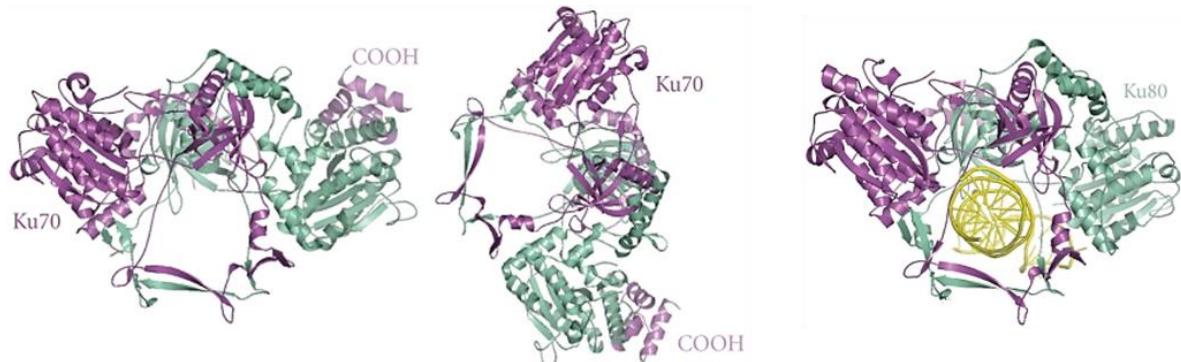
elles nécessitent d'être maturées pour devenir ligaturables (Andres et al., 2015). Selon la nature de la modification aux extrémités de la cassure à réparer, peuvent ainsi intervenir la polynucléotide kinase/phosphatase (PNKP), l'exo/endonuclease Artemis ou trois des ADN polymérasées de la famille X (pol lambda, pol mu et terminal deoxyribonucleotidyltransferase, TdT) (Waters et al., 2014). Ces dernières combinent la brèche d'ADN à la jonction de la cassure selon un gradient décroissant de dépendance vis-à-vis d'une matrice d'ADN, pol lambda étant la plus dépendante et TdT la moins dépendante (Waters et al., 2014). Cette capacité de polymériser des nucléotides sans matrice de la TdT, polymérase spécifiquement exprimée dans les lymphocytes, permet d'augmenter la variation des jonctions dans les régions V(D)J codant pour la partie variable des protéines de reconnaissance des antigènes (Rooney et al., 2004).

Enfin, la DNA-PK est nécessaire au chargement du complexe tripartite de ligature (complexe XXL) formé par les protéines XRCC4 (X-Ray sensitivity Cross Complementing group 4), Cernunnos-XLF (XRCC4-Like Factor) et l'ADN Ligase IV. Le complexe de ligature va assurer la ligature finale des deux extrémités d'ADN. L'accès des enzymes de maturation ainsi que du complexe final de ligature aux extrémités d'ADN est permis grâce à un changement conformationnel de la DNA-PKcs, lui-même induit par un mécanisme d'auto-phosphorylation. Une fois la ligature terminée, la DNA-PKcs ainsi que le complexe de ligature se dissocient du site de la CDB. A ces protéines, il faut ajouter PAXX récemment découverte et qui stabilisera l'interaction de Ku avec l'ADN (Ochi et al., 2015; Xing et al., 2015).

La figure ci-dessous (Figure 2bis) présente les structures tridimensionnelles (modèles ou expérimentales) connues des protéines critiques impliquées dans le mécanisme de C-NHEJ (Protéines du complexe DNA-PK, de maturation et de ligature de l'ADN) (Delarue et al., 2002; Garces et al., 2011; Garcia-Diaz et al., 2007; Moon et al., 2014; Ochi et al., 2015; Ochi et al., 2010; Ochi et al., 2014).

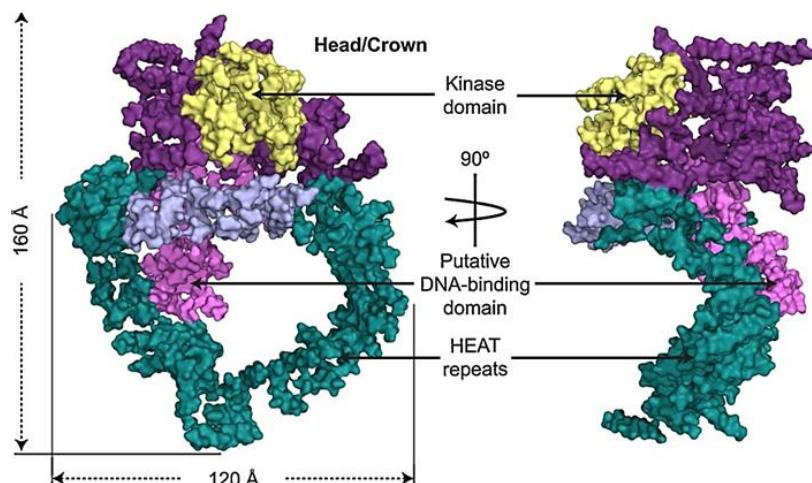
Le complexe DNA-PK

KU



D'après Ochi et al., J Nucleic Acids 2010

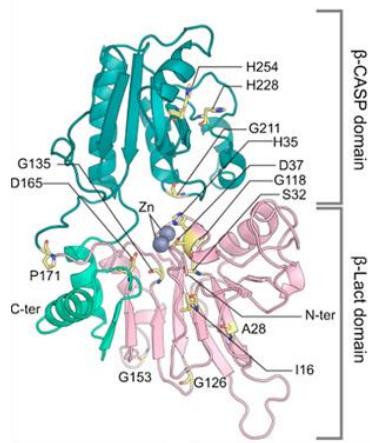
DNA-PKcs



D'après Ochi et al., DNA Repair 2014

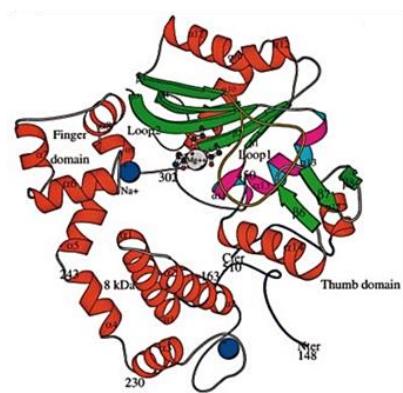
Les protéines de maturation

Artemis (Modèle)



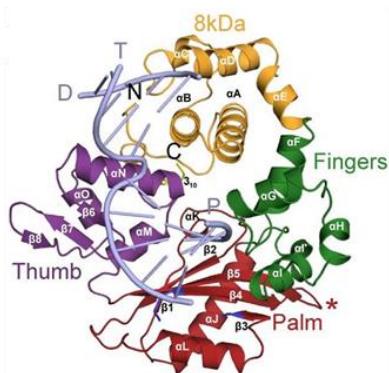
D'après Ochi et al., DNA Repair 2014

TdT



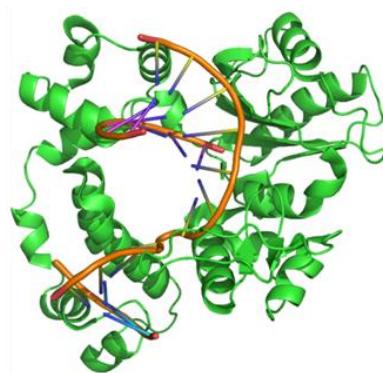
D'après Delarue et al., EMBO J 2002

ADN pol μ



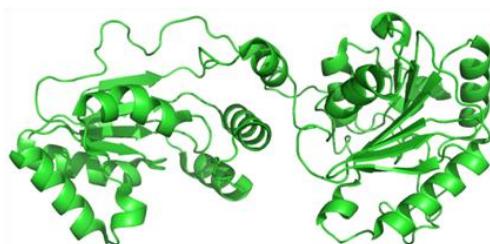
D'après Moon et al., Nat Struct Mol Biol 2014

ADN pol λ



Garcia-Diaz et al., DNA Repair 2007

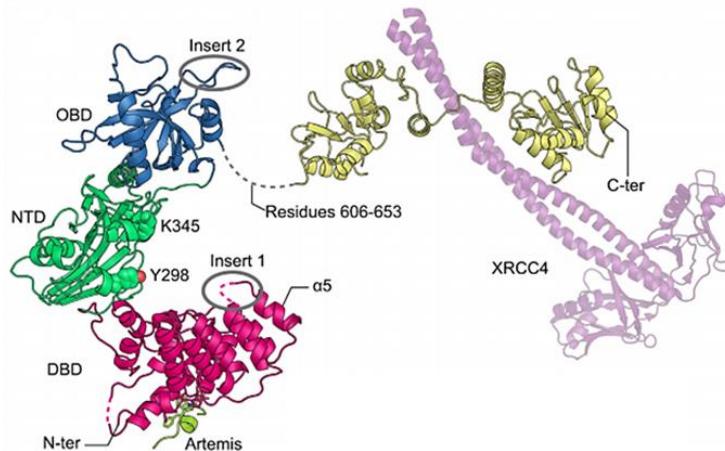
PNKP



Garces et al., Mol Cell 2011

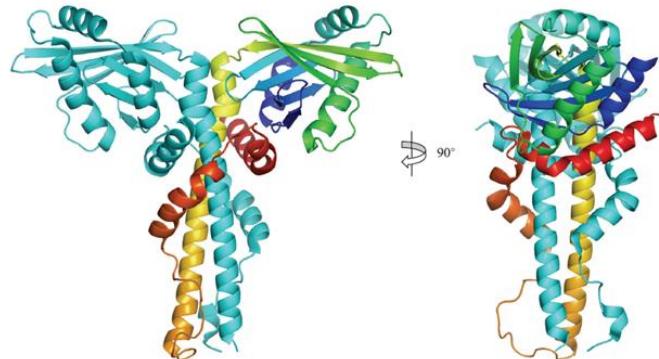
Les protéines de ligature

XRCC4/Ligase4 (Modèle)



D'après Ochi et al., DNA Repair 2014

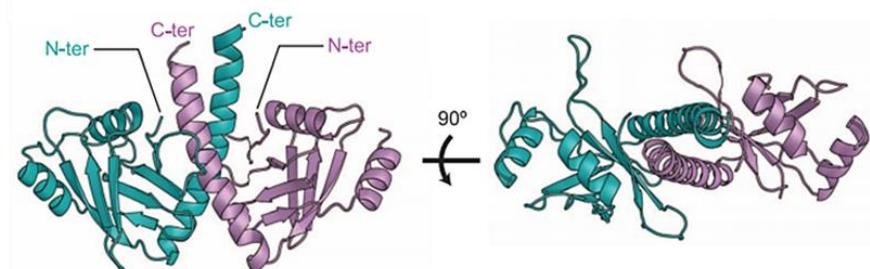
Cer-XLF



D'après Ochi et al., J Nucleic Acids 2010

Autres acteurs

PAXX



D'après Ochi et al., Science 2015

Figure 2bis : structures tridimensionnelles (modèles ou expérimentales) connues des protéines critiques impliquées dans le mécanisme de C-NHEJ.

Le complexe NHEJ de ligature (complexe XXL)

Le complexe XXL (XRCC4/Cer-XLF/ADN Ligase IV) est un des deux complexes majeurs (avec la DNA-PK) impliqué dans la réparation des CDB de l'ADN par la voie NHEJ classique. Il assure l'étape finale de ligature.

La protéine XRCC4 :

Cette protéine de 336 acides aminés (38kDa, migre à 56kDa sur un gel SDS-PAGE) est une protéine nucléaire essentielle à la NHEJ. En effet, un défaut d'expression de cette protéine dans les cellules conduit à une hypersensibilité de ces dernières aux rayons ionisants ainsi qu'à un défaut de recombinaison V(D)J (Lee et al., 2003). Dans le système murin, une déficience en protéine XRCC4 cause une létalité embryonnaire (Gao et al., 1998).

Cette protéine ne possède pas d'activité catalytique propre, mais elle sert de plateforme pour l'amarrage d'autres enzymes de maturation et de religature au cours du mécanisme NHEJ comme la PNK (Koch et al., 2004), APLF (Macrae et al., 2008), Cer-XLF (Ahnesorg et al., 2006) ou la ligase 4 (Grawunder et al., 1997). XRCC4 possède un domaine C-terminal non structuré (contenant une séquence de localisation nucléaire), une tige hélicoïdale (« Stalk » aa 135-233) et une tête globulaire en N-terminal (« Head » aa 1-115) (Figure 3).

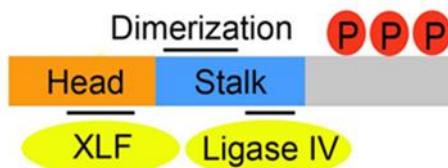


Figure 3 : Schéma des principales caractéristiques structurales de XRCC4. Les domaines « Head » (aa 1-115) et « Stalk » (aa 135-233) sont basés sur la structure cristallographique. La région (aa 119-155) est la région de dimérisation. L'ADN Ligase IV interagit au niveau de la région 173-195. Les régions de phosphorylation sont la thréonine 233, la sérine 260 et la sérine 318. (D'après (Mahaney et al., 2009)).

Elle forme un homodimère via des interactions très stables et hydrophobes entre les deux tiges hélicoïdales, formant ainsi une structure en coiled-coil (Junop et al., 2000) (Figure 3bis). *In vitro*, il a été constaté que deux homodimères étaient également capables de former un homotétramère avec un équilibre dépendant de la concentration en sels (Figure 3bis) (Modesti et al., 2003). Ce complexe tétramérique

est mutuellement exclusif avec celui formé entre un homodimère et l'ADN Ligase IV car leur site d'interaction se chevauchent (Modesti et al., 2003). L'interaction entre les deux homodimères est médiée par des contacts tige-tige tel que cela a été observé par cristallographie (Junop et al., 2000).

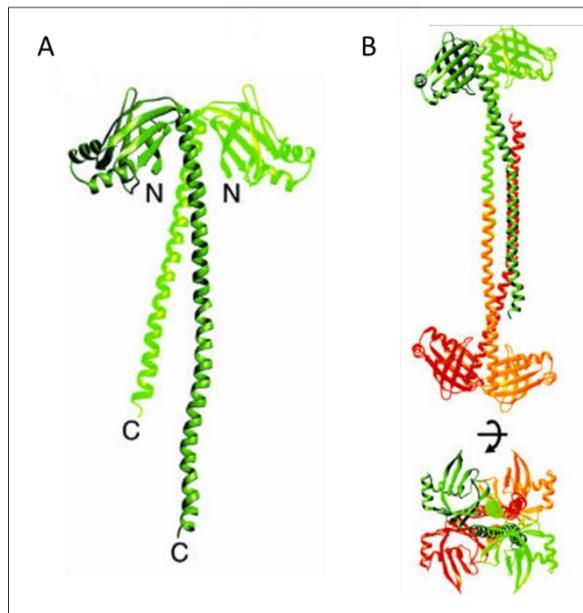


Figure 3bis : Structure cristallographique et oligomérisation de la protéine XRCC4 (Junop et al, 2000). (A) Structure cristallographique de la forme XRCC4 homodimérique (PDB code : 1FU1). (B) vues orthogonales de l'organisation en tétramère de la protéine XRCC4.

L'affinité entre l'homodimère et la ligase IV est bien supérieure ($K_d \approx 10^{-10}$) en comparaison avec celle mesurée entre deux homodimères ($K_d \approx 10^{-6}$) (Modesti et al., 2003). Ainsi, *in vivo*, aucune preuve n'a été à ce jour apportée ni de l'existence ni d'une possible fonction cellulaire d'un tel arrangement en tétramère. XRCC4 contrôle aussi l'import nucléaire de la ligase IV et sa distribution sub-nucléaire (Berg et al., 2011).

L'interface d'interaction entre XRCC4 et la ligase IV s'étend sur près d'environ 2900 \AA^2 . Cette stabilisation est essentielle car cette protéine est dégradée en l'absence de XRCC4 (Bryans et al., 1999). XRCC4 contrôle aussi l'import nucléaire de la ligase IV (Berg et al., 2011). Cette interaction entre les deux partenaires est également extrêmement stable. En effet, elle résiste à des concentrations salines au-delà de 2M et résiste à 7M urée (Robins and Lindahl, 1996). Elle résiste même à des solvants

organiques (McFadden et al., 2014). Le seul moment où ces deux protéines se retrouvent dissociées est durant la mitose (Przewloka et al., 2003). Pendant cette période, XRCC4 est exclue des chromosomes mitotiques pendant que l'ADN Ligase IV est toujours en association avec ces derniers. Une analyse structurale détaillée de l'interaction entre les deux partenaires sera donnée dans la partie suivante : « la protéine ADN Ligase IV ».

Finalement, XRCC4 subit des modifications post-traductionnelles nécessaires à sa régulation. Ces modifications sont localisées au niveau de sa partie C-terminale. Elle est un bon substrat de la protéine DNA-PKcs, qui la phosphoryle au niveau des régions S260 et S318, régulant ainsi sa liaison à l'ADN (Leber et al., 1998). Cette interaction avec l'ADN ne peut se faire que pour de larges fragments (>100pb) (Modesti et al., 1999). Une sumoylation en K210 par la protéine SUMO-1 est également nécessaire au passage de XRCC4 dans le noyau (Yurchenko et al., 2006). Finalement, XRCC4 est phosphorylée par la protéine CK2 permettant son interaction avec APLF (Macrae et al., 2008) et la PNK (Koch et al., 2004).

La protéine ADN Ligase IV :

L'ADN Ligase IV (Lig4) humaine catalyse l'activité centrale de religature de la NHEJ classique. Elle joue donc un rôle clé au sein de cette voie métabolique (Grawunder et al., 1998). Elle fait partie du complexe XXL de ligature qui module son activité ainsi que sa stabilité (Riballo et al., 2009) (Lu et al., 2007). Cette ligase est intrinsèquement instable et très flexible, la rendant alors difficilement étudiable *in vitro* lorsqu'elle se retrouve isolée de son partenaire XRCC4. La Lig4 est l'une des trois ADN ligases humaines dépendantes de l'ATP (Adénosine TriPhosphate) (LigI, LigIII et Lig4) et est présente chez tous les eucaryotes (Tomkinson et al., 2006). Elle appartient, tout comme les LigI et III à la superfamille des nucleotidyltransferases et catalyse une réaction de transfert de nucléotide en trois étapes (Shuman and Lima, 2004):

- La formation d'un intermédiaire adénylé AMP (Adénosine MonoPhosphate) - enzyme (Lys273) covalent à partir d'une molécule d'ATP et libérant une molécule pyrophosphate,
- Le transfert de cet AMP sur une extrémité 5'-P d'ADN (site de la cassure),

- Attaque de ce 5'-phosphate par l'extrémité 3'-OH du brin adjacent, formant une liaison covalente et libérant de l'AMP.

Ainsi, une pré-adenylation initiale de la Lig4 est indispensable à son activité de ligature. Contrairement aux autres ADN ligases, la Lig4 est pré-adénylée lors de sa purification (Tomkinson et al., 2006). Lors de la libération finale d'AMP, cette réadénylation est à nouveau nécessaire. La protéine Cernunnos-XLF (Cer-XLF – dont nous parlerons dans le dernier paragraphe de cette partie) est connue pour stimuler cette réadénylation (Riballo et al., 2009).

La Lig4 possède donc un domaine catalytique. Ce domaine est conservé chez les trois ADN ligases et il comporte un domaine d'adénylation (AdD, aa 248-454) ou nucleotidyltransférase et un domaine de liaison aux oligonucléotides (Oligo-Binding Domain, OBD, aa 455-596) ou OB-fold domaine. La Lig4 possède également un domaine N-terminal de liaison à l'ADN (DBD, aa 1-241), qui lui est homologue à l'ADN LigI.

En plus d'un domaine catalytique et d'un domaine de liaison à l'ADN, la Lig4 possède un extra-domaine en C-terminal qui lui est propre et donc non présent chez les autres ligases humaines (Tomkinson et al., 2006). Cet extra-domaine (aa 654-911) est un tandem BRCA1 (Breast Cancer Associated) C-terminal (BRCT), composé de deux domaines BRCT1 et BRCT2 (reliés par un linker organisé en motif hélice-boucle-hélice : HLH), responsable de l'interaction avec la protéine XRCC4 (Figure 4) (Marchetti et al., 2006). Ces domaines BRCT ont typiquement quatre brins beta parallèles entourés de trois hélices alpha (Zhang et al., 1998) et sont communs aux protéines de régulation du cycle cellulaire en réponse aux dommages à l'ADN (Bork et al., 1997). Il a été montré que le domaine BRCT1 interagissait avec le complexe hétérodimérique Ku70/80 (Costantini et al., 2007), et possiblement avec Cer-XLF (Wu et al., 2009). La Lig4 interagit aussi avec Artemis et une structure cristalline de cette interaction a été publiée (Figure 5) (Ochi et al., 2013).

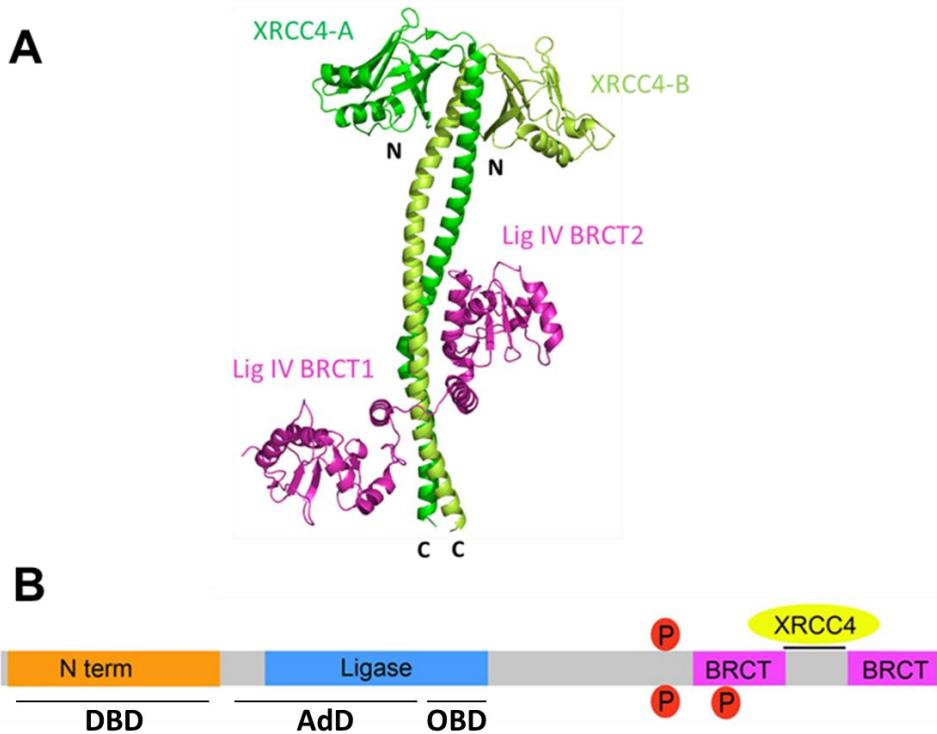


Figure 4 : (A) Structure cristallographique de la protéine XRCC4 en complexe avec le domaine C-terminal (aa 654-911) de la Ligase IV (Wu *et al.*, 2009). **(B)** schéma des principales caractéristiques structurales de la Ligase IV. *In vitro*, la Ligase IV est phosphorylée au niveau de la thréonine 650, de la sérine 668 et de la sérine 672 (D'après (Mahaney *et al.*, 2009)).

Finalement, la Lig4 s'organise en quatre domaines distincts : le DBD, le domaine catalytique, le tandem BRCT en C-terminal et un linker prédit comme majoritairement désordonné, reliant ce tandem BRCT au domaine catalytique (Figure 5). Ce linker semble cependant être important pour l'activité catalytique de la Lig4 (Grawunder *et al.*, 1998) et possède un site de phosphorylation par la DNA-PKcs en T650 pour la stabiliser (Wang *et al.*, 2004a).

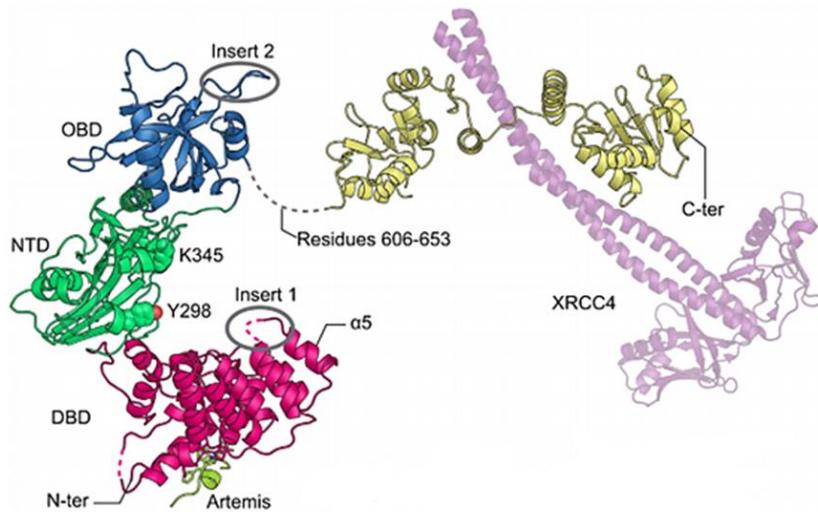


Figure 5 : Structure cristallographique du domaine catalytique de la ligase IV en complexe avec un peptide Artemis (PDB code : 3W1B) et relié par un linker (aa 606-653) au domaine tandem BRCT de la ligase IV en complexe avec XRCC4 (PDB code : 3I16) (Ochi et al, 2014). DBD = Domaine de liaison à l'ADN ; NTD = Domaine d'adénylation ; OBD = Domaine OB-fold.

Le tandem BRCT de la Lig4 interagit avec la structure hélicoïdale de XRCC4 via une interaction par le domaine BRCT2 (hélice $\alpha 1'$) ainsi que le linker hélice-boucle-hélice (HLH) qui vient littéralement « clamer » cette tige autour du coiled-coil (Wu et al., 2009) (Figure 5bis). D'ailleurs, nous désignerons ce domaine HLH comme le domaine « clamp » pour la suite dans cette thèse.

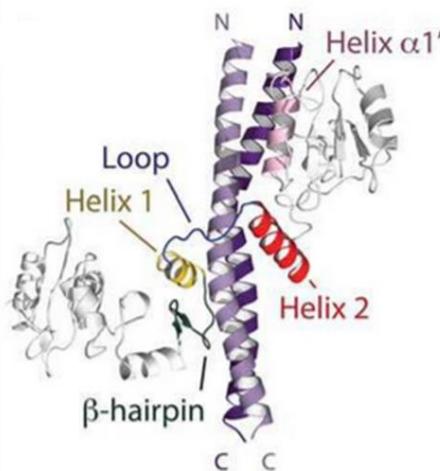


Figure 5bis : Domaine tandem BRCT C-terminal de la ligase IV lié à la protéine XRCC4. Représentation des 4 domaines de la ligase impliqués dans l'interaction : β -hairpin, hélice1, loop, hélice2, hélice $\alpha 1'$ (McFadden et al, 2014).

L'interaction avec XRCC4 est principalement due à ce clamp (Grawunder et al., 1998) (McFadden et al., 2014). Une région minimale d'interaction à XRCC4 (XIR : XRCC4 Interacting Region) avait été dans un premier temps caractérisée par cristallographie en complexe avec la protéine XRCC4 (Sibanda et al., 2001). Cette région minimale est une partie de la structure en hélice-boucle-hélice (contenant également une partie du β -hairpin, entre le clamp et le domaine BRCT1) mais des mutants de délétions ainsi qu'une structure du domaine C-terminal complet a permis par la suite d'impliquer également le domaine BRCT2 dans une interaction hydrophobe stable avec XRCC4 (Wu et al., 2009). Il est à noter que la liaison du domaine XIR ou du C-terminal complet de la Lig4 sur XRCC4 induit une courbure au sein du domaine coiled-coil de XRCC4 et cette courbure tend à être dans des directions opposées en fonction de la liaison avec le XIR ou le tandem BRCT (Wu et al., 2009) (Figure 6). Ainsi, la courbure générée par le XIR pourrait être un état intermédiaire dans l'interaction XRCC4/Lig4. Si tel est le cas, cette dynamique conformationnelle pourrait avoir un rôle significatif *in vivo*.

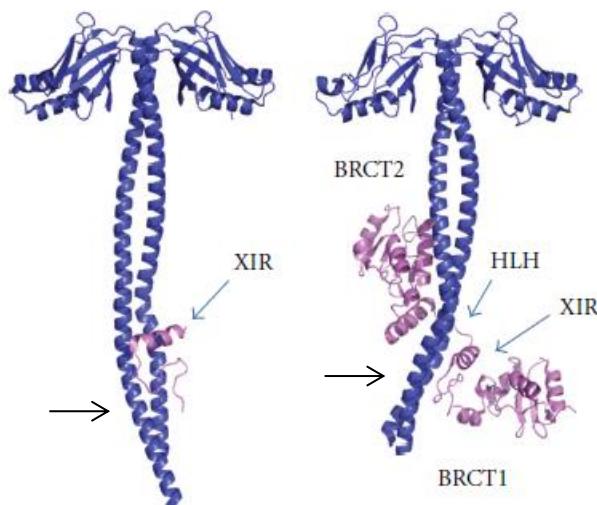


Figure 6 : Structures cristallographiques des complexes XRCC4/Domaine C-terminal de la ligase IV (PDB code : 3II6) et XRCC4/XIR (XRCC4 Interacting Region) (PDB code : 1IK9) (Ochi et al, 2010). La flèche '→' indique la zone de courbure du domaine coiled-coil.

Plus récemment, une étude biophysique menée par McFadden en 2014 a permis de montrer qu'un déplacement compétitif du XIR (aa 759-803) sur XRCC4 n'est significatif qu'avec l'hélice 2 du clamp ou des fragments contenant cette hélice 2 (McFadden et al., 2014).

La protéine Cernunnos-XLF (Cer-XLF) :

Chez des patients (2BN) atteints de microcéphalie, de radiosensibilité et d'une immunodéficience combinée sévère (RS-SCID), il a été observé une complémentation fonctionnelle avec un ADNc humain transduit, et codant pour une protéine que les auteurs ont appelé « Cernunnos » (Buck et al., 2006). En parallèle, une autre équipe a recherché des partenaires protéiques de la protéine XRCC4. Grâce à la technique du double-hybride, ces auteurs ont déterminé une interaction spécifique avec une protéine qu'ils ont baptisé XLF (XRCC4-Like Factor) car cette dernière a été prédite comme possédant les mêmes caractéristiques structurales que XRCC4 (Ahnesorg et al., 2006). Il s'est avéré par la suite que XLF et Cernunnos étaient les mêmes protéines et une complémentation fonctionnelle par XLF de cellules provenant de patients RS-SCID (mutées pour le gène XLF) permet de restaurer la fonction NHEJ. De plus, XLF interagit *in vitro* et *in vivo* avec le complexe XRCC4/Ligase IV.

Cer-XLF est une protéine de 299 acides aminés (33kDa) qui est très bien conservée au cours de l'évolution à travers divers types d'eucaryotes tels que les vertébrés, les insectes et même les fungi filamentueux (Hentges et al., 2006). Elle est localisée au sein du noyau des cellules (Ahnesorg et al., 2006).

La structure cristallographique de cette protéine (tronquée en C-terminal) a pu être déterminée par deux équipes indépendantes (Andres et al., 2007) (Li et al., 2008). Cette structure ainsi que les caractéristiques principales de Cer-XLF sont présentés en figure ci-dessous (Figure 7).

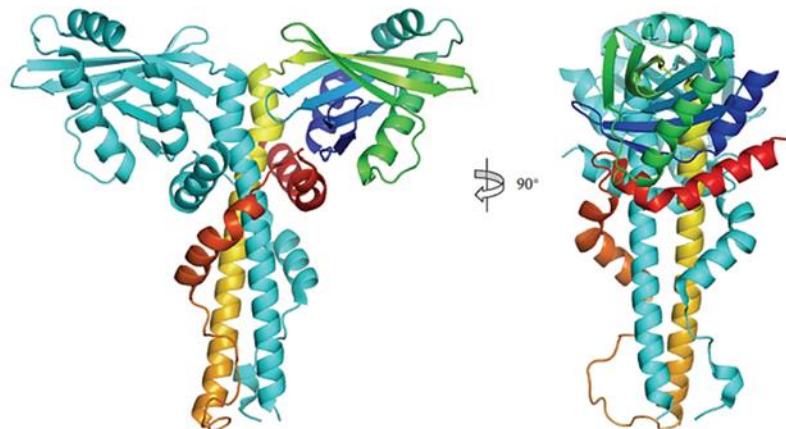
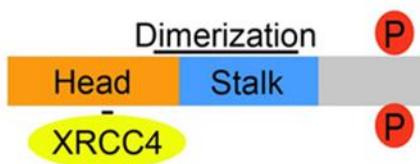
A**B**

Figure 7 : (A) Structure cristallographique de la protéine Cer-XLF (PDB code : 2QM4) en vue de face et de profil (adapté de Li et al, 2008) (B) schéma des principales caractéristiques structurales de Cer-XLF. Les domaines « Head » (aa 1-135) et « Stalk » (aa 135-233) sont basés sur la structure cristallographique. La région d'interaction avec XRCC4 implique notamment le résidu critique Leucine 115. *In vivo*, Cer-XLF est phosphorylée au niveau des résidus sérine 245 et sérine 251 (D'après (Mahaney et al., 2009)).

Ses caractéristiques structurales font de cette protéine un homologue structural de la protéine XRCC4. C'est un homodimère possédant une tête globulaire en N-terminal (« Head » aa 1-135), une queue hélicoïdale (« Stalk » aa 135-233) ainsi qu'une région C-terminale non structurée (aa 245-299) et donc non présente dans la structure cristallographique (environ 70 résidus). Malgré une architecture similaire, une différence majeure par rapport à XRCC4 toutefois, est le repliement de cette queue sur la protéine elle-même, venant ainsi se positionner à la base de la tête globulaire du monomère adjacent. Ce repliement préviendrait peut-être, l'interaction du domaine C-terminal de la Lig 4 avec Cer-XLF. Ce domaine C-terminal porte, tout comme XRCC4, un domaine NLS de localisation nucléaire. Il porte également deux sites de phosphorylation (S245 et S251) par la DNA-PKcs et ATM (Yu et al., 2008) mais dont la substitution par un acide-aminé non modifiable n'influe pas sur l'activité NHEJ *in vivo*. Bien qu'une organisation structurale de Cer-XLF bien plus complexe et

dépendante de la concentration saline ait pu être observée lors d'expériences de gel filtration (Andres et al., 2007), la forme homodimérique est la seule unité fonctionnelle stable.

Les fonctions et mécanismes de Cer-XLF ne sont pas encore clairs. Bien qu'elle ne présente aucune activité enzymatique connue, cette protéine stabilise les complexes XRCC4/Lig4 aux sites de CDB de l'ADN et stimule la ligature d'extrémités non compatibles (Gu et al., 2007) (Tsai et al., 2007). Elle permet également de ré-adényler la Lig4 (Ahnesorg et al., 2006) (Riballo et al., 2009) et est essentielle au processus de remplissage de brèches (« gap-filling ») catalysé par les ADN polymérasées pol μ et pol λ (Akopiants et al., 2009).

Cer-XLF n'est pas nécessaire à l'interaction entre la Lig4 et XRCC4 mais la Lig4, elle, stabilise l'interaction de Cer-XLF avec XRCC4 (Wu et al., 2009). Elle est également capable (via son extrémité C-terminale) de lier de large fragments d'ADN (>80pb) (Lu et al., 2007). L'interaction entre Cer-XLF et XRCC4 est sensible à la concentration saline et n'est pas dépendante de la liaison à l'ADN (Riballo et al., 2009). La première caractérisation du mode d'interaction entre ces deux protéines par des études double-hybride a permis de mettre en évidence une interaction possible médiée par les têtes globulaires de ces dernières (aa 1-119) (Deshpande and Wilson, 2007). De plus, des mutations ponctuelles sur l'une ou l'autre des protéines ont pu confirmer cette hypothèse et permettre de dresser un modèle d'interaction où le sommet de la tête globulaire de Cer-XLF viendrait contacter la partie inférieure de XRCC4 (Andres et al., 2007). Des études de pull-down ont montré que des billes liées à la partie C-terminale de Cer-XLF étaient toujours capable de lier le complexe XRCC4/Lig4, éliminant ainsi l'importance de ce domaine C-terminal dans l'interaction (Lu et al., 2007).

Ainsi, de par la structure homodimérique de XRCC4 et Cer-XLF ainsi que leur interaction médiée par leur tête globulaire, il a été proposé un modèle selon lequel ces protéines pourraient former de larges assemblages en polymères (Malivert et al., 2010). Ces « filaments » de protéines ont par la suite été confirmés expérimentalement à partir de protéines purifiées et via notamment des méthodes

combinées de cristallographie et microscopie électronique à transmission (Hammel et al., 2011) (Ropars et al., 2011) (Figure 8).

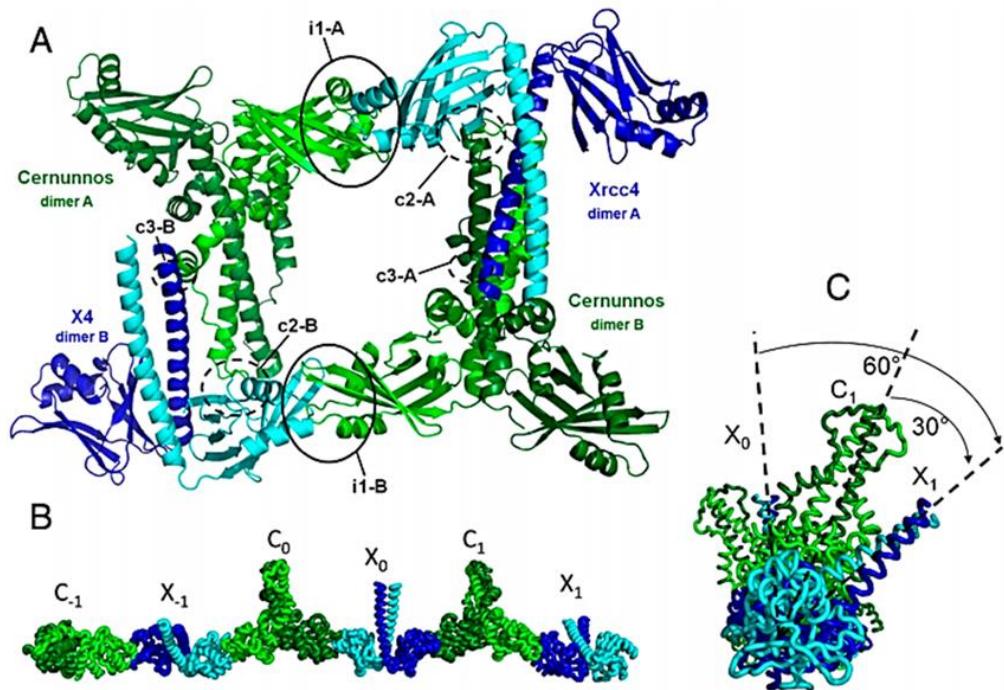


Figure 8 : Structure cristallographique du complexe XRCC4/Cer-XLF et de l'organisation en filament (PDB code : 3Q4F) (Ropars et al, 2011).

Ces structures permettent d'entrevoir l'assemblage complexe entre XRCC4 et Cer-XLF, mais également de cerner les résidus critiques d'interaction entre les deux partenaires. Ces résidus critiques avaient déjà été déterminés dans une étude précédente de mutagénèse (Andres et al., 2007). Plus récemment, une liste précise des résidus clé d'interaction entre les deux partenaires a pu être dressée et corrélée aux études structurales via des méthodes combinées de mutagénèse dirigée et calorimétrie à titration isotherme (ITC) (Ropars et al., 2011). Dans les conditions expérimentales utilisées, les résidus hydrophobes M59, M61, L101, V104 et F106 de XRCC4 sont en interaction avec les résidus hydrophobes L65, A67, F72, L112, L115, P116 et Y118 de Cer-XLF. La figure ci-dessous (Figure 9) met en évidence qu'une mutation ponctuelle sur les résidus E55, D58, M61 ou F106 de XRCC4 ou bien R64 ou L115 de Cer-XLF est suffisante pour perdre complètement l'interaction entre les deux partenaires.

Cernunnos *	Xrcc4 *	Kd (μ M)
WT ¹⁻²²⁴	WT ¹⁻²⁰³	3.9 \pm 0.2
WT ¹⁻²²⁴	WT ¹⁻¹⁵⁷	4.1 \pm 0.2
WT ¹⁻²²⁴	E55R ¹⁻¹⁵⁷	no interaction
WT ¹⁻²²⁴	D58R ¹⁻¹⁵⁷	no interaction
WT ¹⁻²²⁴	M61R ¹⁻¹⁵⁷	no interaction
WT ¹⁻²²⁴	E62R ¹⁻¹⁵⁷	3.8 \pm 0.2
WT ¹⁻²²⁴	K65E ¹⁻¹⁵⁷	>40
WT ¹⁻²²⁴	E69R ¹⁻¹⁵⁷	5.0 \pm 0.2
WT ¹⁻²²⁴	F106E ¹⁻¹⁵⁷	no interaction
WT ¹⁻²²⁴	F106A ¹⁻¹⁵⁷	>40
WT ¹⁻²²⁴	F106Y ¹⁻¹⁵⁷	>40
F111K ¹⁻²²⁴	WT ¹⁻¹⁵⁷	6.9 \pm 0.2
R64E ¹⁻²²⁴	WT ¹⁻²⁰³	no interaction
L65D ¹⁻²²⁴	WT ¹⁻²⁰³	>80
L115D ¹⁻²²⁴	WT ¹⁻²⁰³	no interaction

Figure 9 : Analyse de l'interaction XRCC4/Cer-XLF par calorimétrie à titration isotherme (ITC) et après mutagénèse dirigée (Ropars et al., 2011). Les mutations ponctuelles encadrées en rouge correspondent à des mutations induisant la perte totale d'interaction entre les deux partenaires.

Cette information est primordiale dans une stratégie de design rationnel de molécules inhibitrices de cette interaction car elle souligne d'une part, la faible affinité entre ces deux protéines et d'autre part, les résidus « hot-spot » qu'il serait intéressant par conséquent de cibler dans une telle approche.

Finalement, le rôle physiologique de ces filaments n'est pas encore bien connu à ce jour. Il est cependant admis qu'ils pourraient faciliter l'alignement des extrémités de la cassure au cours de la réparation. Bien que les deux protéines soient capable de lier l'ADN au niveau de zones bien caractérisées, seule la zone d'interaction de Cer-XLF (aa 224-299) à l'ADN semble être absolument nécessaire à l'association du complexe avec l'ADN (Andres et al., 2012). Cette zone d'interaction étant localisée en C-terminal de Cer-XLF, cela suggèrerait que l'ADN est « enveloppé » par le complexe XRCC4/Cer-XLF. Récemment, il a été montré que l'interaction XRCC4/Cer-XLF est plus ou moins requise selon les modèles cellulaires pour la réparation par NHEJ mais qu'elle n'est pas nécessaire à la stimulation de la Lig4 (Roy et al., 2015). Une étude utilisant la microscopie de fluorescence en super résolution a identifié dans les cellules humaines des structures filamenteuses du complexe XXL (Reid et al., 2015). Enfin, la stabilité de ces filaments pourrait être régulée par la phosphorylation des extrémités C-terminales de XRCC4 et Cer-XLF par la protéine DNA-PK (Roy et al., 2012).

I – 2 : Inhibition des voies de réparation des cassures double brin de l'ADN pour une thérapie combinée anti-cancéreuse

Les agents chimiothérapeutiques

La toxicité des CDB de l'ADN est exploitée en thérapie anti-cancéreuse (Connell et al., 2004; Jeggo and Lavin, 2009). Afin de lutter contre le développement et la progression tumorale, la radiothérapie, seule ou en combinaison avec des traitements chimiothérapeutiques, va être utilisée. Ces agents sont en général des radiations ionisantes (RI) (rayons X ou γ), ou des molécules dites radiomimétiques (car elles miment l'effet des rayons) comme par exemple la Calichéamicine, la Néocarzinostatine ou la Bléomycine. Le but de ces radiomimétiques est de pouvoir générer une proportion plus grande de CDB, de manière aléatoire dans le génome, par rapport à l'effet seul des RI. Ces agents induisent donc directement des CDB, ou bien des CSB qui seront ensuite converties en CDB par la réPLICATION. Ce mécanisme réPLICatif est par ailleurs « suractif » dans les tumeurs, où les cellules sont dans un état hyper-prolifératif. Ainsi, les cellules cancéreuses sont plus sensibles face à ces agents. Cependant, les cellules des tissus sains en prolifération (follicule pileux, muqueuse intestinale, lignées hématopoïétiques) sont aussi particulièrement sensibles à ces agents ce qui explique certains de leurs effets secondaires (alopecie, troubles digestifs, aplasie médullaire).

Les cassures générées par des RI sont directes (Siddiqi and Bothe, 1987) ou bien en lieu par l'intermédiaire d'espèces réactives de l'oxygène (ROS), produites par interaction de ces rayons avec les molécules d'eau. (Van Der Schans, 1978). Dans chaque cas, des CSB proches l'une de l'autre mènent à une CDB. Les rayons restent l'un des traitements les plus efficaces sur les tumeurs solides (Delaney et al., 2005). Les cassures générées par des molécules radiomimétiques proviennent d'une attaque radicalaire, par déshydrogénération du désoxyribose sur les deux brins d'ADN. Le site de lésion ainsi créé conduit à une perte de base puis un clivage de la chaîne d'ADN (Chen et al., 2005) (Povirk, 1996).

Les problèmes majeurs rencontrés lors de l'utilisation de ces génotoxiques en thérapie anti-cancéreuse sont le ciblage spécifique de la tumeur (sans affecter non spécifiquement les tissus sains environnants), l'hétérogénéité interindividuelle au

niveau de la sensibilité de chaque patient aux différents agents clastogènes (doses parfois trop fortes ou trop faibles) et pouvant conduire au décès du patient ou parfois même à la réactivation du cancer (Barnett et al., 2009; Guo et al., 2015) et enfin, la sensibilité de chaque tumeur à ces traitements, qui doivent être par conséquent adaptés en fonction des mutations présentes au sein des cellules concernées par exemple (Dietlein et al., 2014; Pearl et al., 2015).

L'avantage d'une utilisation combinée de molécules chimiothérapeutiques en plus des rayons, est d'augmenter la sensibilité de la tumeur face à ces derniers (effet synergique), faisant ainsi face à des mécanismes de radio-résistances acquises ou intrinsèques, et tout en limitant les doses de rayons afin de réduire la cytotoxicité au niveau des cellules saines (Begg et al., 2011). Notamment, l'utilisation d'agents radiosensibilisants permet d'augmenter grandement l'efficacité des radiothérapies et agents génotoxiques anti-cancéreux.

Les résistances développées face aux agents génotoxiques apparaissent à travers divers mécanismes. L'un de ces mécanismes est la sur-expression de certaines protéines de réparation de l'ADN (Begg et al., 2011). Des revues spécialisées décrivent les avantages d'une inhibition des voies de réparation de l'ADN dans le but d'augmenter significativement l'efficacité des drogues antinéoplasiques dans les cellules cancéreuses (Begg et al., 2011; Dietlein et al., 2014; Pearl et al., 2015; Srinivasan and Gold, 2012). Cette stratégie de compromettre certaines protéines de réparation de l'ADN (mais également des protéines impliquées dans la régulation du cycle cellulaire) a été validée pour plusieurs cibles protéiques par modulation de leur niveau d'expression (Begg et al., 2011; Dietlein et al., 2014; Jekimovs et al., 2014; Pearl et al., 2015; Srinivasan and Gold, 2012).

La NHEJ comme cible dans une approche de chimiothérapie

Une détection et réparation rapide des CDB générés par des rayons ou autres agents, est indispensable à la cellule tumorale afin d'assurer sa survie. La NHEJ est la voie principale de réparation de ce type de dommage et elle intervient dans toutes les phases du cycle cellulaire (Shrivastav et al., 2008). Ce mécanisme NHEJ est donc une cible de choix dans une stratégie de ciblage pour augmenter l'efficacité des

thérapies anticancéreuses. D'ailleurs, il a été prouvé plusieurs fois que le défaut de NHEJ par des mutations du complexe DNA-PK ou XRCC4/Cer-XLF/Lig4 conduisait à une augmentation significative de la sensibilité aux rayons ionisants (Woodbine et al., 2014). Les voies de réparation de secours (NHEJ Backup ou B-NHEJ, non décrite ici) pourraient prendre le relais à l'issue d'une telle inhibition. Cependant, ces mécanismes de secours ne seraient clairement pas suffisants puisque par exemple, les patients atteints du syndrome LIG4 (par défaut de l'ADN ligase IV) sont fortement radiosensibles (Tomkinson et al., 2013). D'ailleurs, le blocage de la C-NHEJ pourrait également prévenir l'accès des CDB aux protéines des autres voies de réparation par l'intermédiaire de la protéine KU liée aux cassures (Cheng et al., 2011). Nous ne re-décrirons pas ici ce mécanisme de C-NHEJ mais nous ferons l'état des lieux concernant l'existence d'inhibiteurs spécifiques de certaines des protéines impliquées dans ce dernier ainsi que les pistes restant à explorer dans cette voie, notamment l'inhibition du complexe XXL de ligature. L'inhibition de protéines impliquées dans la voie de RH ne sera pas abordée au cours de cette étude.

Inhibition de la sous-unité catalytique de la DNA-PK (DNA-PKcs)

La DNA-PKcs est surexprimée dans certaines tumeurs et les lignées cellulaires résistantes aux radiations ionisantes, indiquant ainsi un rôle dans la croissance et la survie tumorale (Ader et al., 2002). Au vu de cela, de son rôle pivot au sein de la NHEJ, et de sa qualité de protéine kinase, la DNA-PKcs semble être une cible d'intérêt dans le développement de radio ou chimio sensibilisants (Hsu et al., 2012). Plusieurs inhibiteurs de la DNA-PKcs ont déjà été identifiés. Les plus efficaces ont pour cible la poche de liaison à l'ATP de son domaine kinase (Allen et al., 2003). Certains de ces inhibiteurs ont atteint l'étape d'évaluation préclinique (Davidson et al., 2013; Dietlein et al., 2014).

Par exemple, la Wortmannin (Figure 10) a été l'un des premiers inhibiteurs découvert pour l'inhibition de cette protéine. Cette molécule possède un IC₅₀ proche de 5nM et a montré une inhibition significative de la DNA-PKcs et donc de la réparation des CDB, conduisant à une sensibilité accrue aux rayons ionisants et à l'étoposide (Collis et al., 2005).

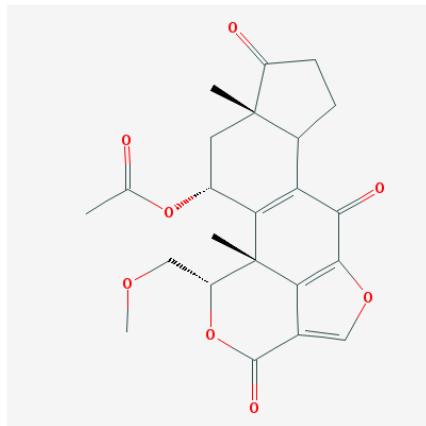


Figure 10 : Structure de la Wortmannin (molécule issue de *penicillium funiculosum*), inhibitrice de la protéine DNA-PK.

Un autre exemple est celui de la molécule NU7026 possédant un IC₅₀ proche de 0,23µM contre la DNA-PKcs (Peddi et al., 2010) (Figure 11). Cet inhibiteur, spécifique et réversible, inhibe la croissance tumorale (arrêt G₂/M) en combinaison avec d'autres agents tels que la daunorubicine, la doxorubicine ou l'étoposide (Willmore et al., 2004).

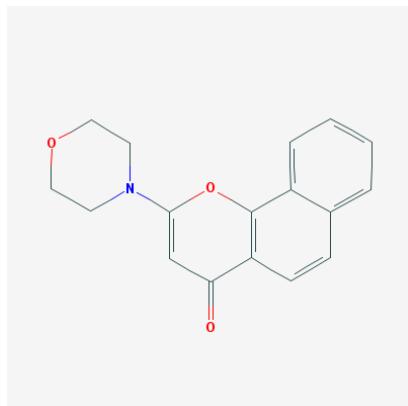


Figure 11 : Structure de la molécule NU7026, inhibitrice de la protéine DNA-PK.

Finalement, deux autres inhibiteurs de la DNA-PK ont également été découvert (SU11752 et OK-1035) et sensibilisent les cellules aux agents de dommage à l'ADN (Kruszewski et al., 1998).

Bien que ces inhibiteurs aient atteint l'étape d'évaluation préclinique, ils se sont avérés être rapidement de mauvais candidats de par des faibles propriétés pharmacocinétiques, un manque de solubilité et de spécificité et parfois même, une

toxicité *in vivo*. Leur instabilité métabolique respective limite la dose pouvant atteindre les cellules cancéreuses (Collis et al., 2005).

Toutefois, l'obtention récente d'une structure de la DNA-PKcs pourrait conduire à un développement futur de nouvelles molécules inhibitrices par une stratégie rationnelle et via une approche computationnelle (Sibanda et al., 2010).

Ciblage de XRCC4

Une seule étude a été conduite jusqu'à présent qui visait à inhiber l'association XRCC4/ADN Ligase IV par une recherche computationnelle de molécules en interaction avec le domaine coiled-coil de XRCC4, au niveau de la région d'interaction de la Ligase 4 (Sun et al., 2011). Trois molécules (Salvianolic acid B, lithospermic acid, 2-O-feruloyl tartaric acid) issues du docking contre la base de données TCM (Traditional Chinese Medicine) ont été caractérisées avec un bon score de docking. Ces molécules ont été modélisées en interaction au niveau des lysines 188 et 187 des chaînes A et B respectives de XRCC4, au niveau de la zone d'interaction du domaine clamp de la Ligase 4. Cependant, aucune information concernant la validation expérimentale (*in vitro* et *in cellulo*) de ces molécules n'a été fournie et l'efficacité de ces composés reste encore à déterminer.

Plus récemment, une étude a permis de démontrer qu'un fragment minimal de la Ligase IV permettait de déplacer de manière compétitive l'interaction XRCC4/XIR (XRCC4 Interacting Region) (McFadden et al., 2014), dont le Kd a pu être estimé autour de $7.8 \pm 4.2\text{nM}$. Cette expérience a été réalisée grâce à la méthode ESI-MS/MS (Spectrométrie de Masse avec Ionisation par ElectroSpray), avec différents fragments issus du domaine β -hairpin-clamp (β -hairpin-hélice1-loop-hélice2) de la Ligase IV. Il a pu ainsi être prouvé que seuls les fragments contenant au moins l'hélice 2 du XIR étaient capables de déplacer ce complexe pré-formé (Figure 5bis).

Cette étude pionnière a démontré que des modulateurs de l'interaction entre XRCC4 et le XIR du domaine C-terminal de la Ligase IV, devraient pouvoir inhiber cette dernière et particulièrement en ciblant spécifiquement cette hélice 2. L'hélice 2 faisant partie du domaine « hélice1-loop-hélice2 » du clamp de la Ligase IV, les régions hélice 1 et loop offrent également des zones de ciblage secondaires. Finalement, une telle étude permet également de mettre en avant le fait qu'un

modulateur peptidique (ou peptido-mimetique) de cette interface serait aussi une méthode de choix dans cette stratégie d'inhibition.

Ciblage de l'hétérodimère Ku70/80

Tout comme pour la DNA-PKcs, le niveau d'expression de Ku70/80 augmente de manière significative dans certains types de tumeurs, ce qui pourrait ainsi indiquer que la survie tumorale dépend également en partie de ce complexe hétérodimérique (Beskow et al., 2009). En effet, il a pu être montré qu'une déplétion de l'une ou l'autre des protéines de ce complexe en utilisant des shRNA, conduisait à une inhibition de croissance de cellules tumorales pancréatiques, ainsi qu'une augmentation de leur sensibilité aux rayons ionisants (Li et al., 2012). Cela suggère, là encore, que ce complexe pourrait être une cible thérapeutique importante en thérapie anti-cancéreuse. A ce jour, aucun inhibiteur n'a pu être identifié (Jekimovs et al., 2014). Cependant, KU possède une fonction essentielle dans les cellules humaines, ce qui pourrait compliquer l'utilisation d'un tel inhibiteur en tant qu'agent thérapeutique (Li et al., 2002).

Inhibition de l'ADN Ligase IV

Il a été montré que des cellules déficientes en ADN Ligases étaient plus sensibles à des agents de dommages à l'ADN (Tomkinson et al., 2006). Une recherche intensive d'inhibiteurs de ligases humaines a pu être permise grâce à l'émergence de trois découvertes indépendantes et primordiales : l'expression et la purification d'ADN Ligases humaines recombinantes, le développement de méthodes de criblage haut-débit d'inhibiteurs, et la structure cristallographique du domaine catalytique de ces ligases, permettant un criblage d'inhibiteurs plus rationnel, basé sur ces structures.

L'ADN Ligase IV est l'unique ligase associée au mécanisme NHEJ de réparation des CDB de l'ADN. Cette ligase n'intervenant dans aucune autre voie métabolique de réparation, son inhibition devient alors extrêmement spécifique de la NHEJ. Des patients atteints d'une déficience en cette Ligase IV (syndrome LIG4) sont pour certains, sévèrement radiosensibles (Ellenberger and Tomkinson, 2008) (Tomkinson et al., 2013). De plus, il a été démontré qu'une expérience de compétition cellulaire par expression ectopiques de mutants de délétion de cette Ligase IV, conduit à une

diminution de la Ligase IV endogène, et une augmentation de la radiosensibilité des cellules (Wu et al., 2009). Tous ces arguments sont en faveur d'une inhibition de la Ligase IV dans une stratégie de radiosensibilisation des cellules tumorales et donc, pour une thérapie anti-cancéreuse combinée.

A ce jour, deux inhibiteurs de cette Ligase IV ont été validés expérimentalement (L189 et SCR7). Ces deux inhibiteurs ont été trouvés de manière rationnelle par une approche de Drug Design assisté par informatique.

L189 : La structure de l'ADN Ligase I (hLig1) en complexe avec un fragment d'ADN au niveau de son domaine de liaison à l'ADN (domaine DBD) a permis d'effectuer un criblage rationnel de petites molécules inhibitrices de cette interaction hLig1/ADN (Chen et al., 2008) (Figure 13).

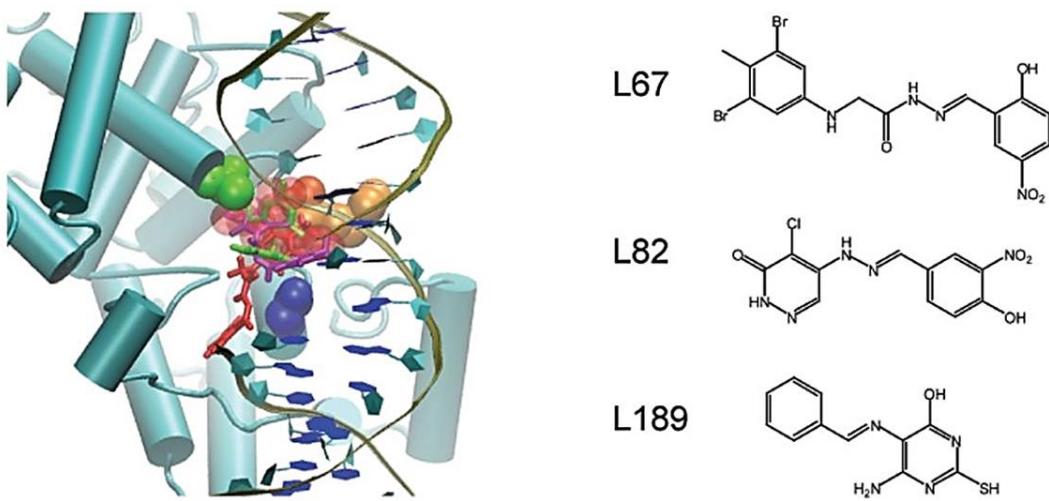


Figure 13 : Criblage rationnel de molécules inhibitrices de l'interaction DBD de hLig1/ADN. A gauche : résidus clés de la poche de liaison à l'ADN (Gly448 (vert) ; Arg451 (orange) ; Ala455 (bleu)) et poses de docking des trois composés caractérisés : L67 ; L82 et L189 (à droite) (Chen et al, 2008).

Parmi les candidats retenus, L189 a pu montrer un effet inhibiteur significatif dans un essai NHEJ avec des extraits cellulaires impliquant l'ADN Ligase IV. L189 inhibe la réaction de ligation au-delà de 90% et augmente la cytotoxicité de certains agents de dommages à l'ADN ainsi que des rayons ionisants dans trois lignées cellulaires différentes mais pas dans des lignées épithéliales normales (Chen et al., 2008).

Cependant, dû à la conservation de ce domaine catalytique à travers les différentes ligases humaines (Figure 14), le composé L189 a démontré également un effet inhibiteur sur les Ligase humaines I et III dans les mécanismes BER et de réPLICATION de l'ADN.

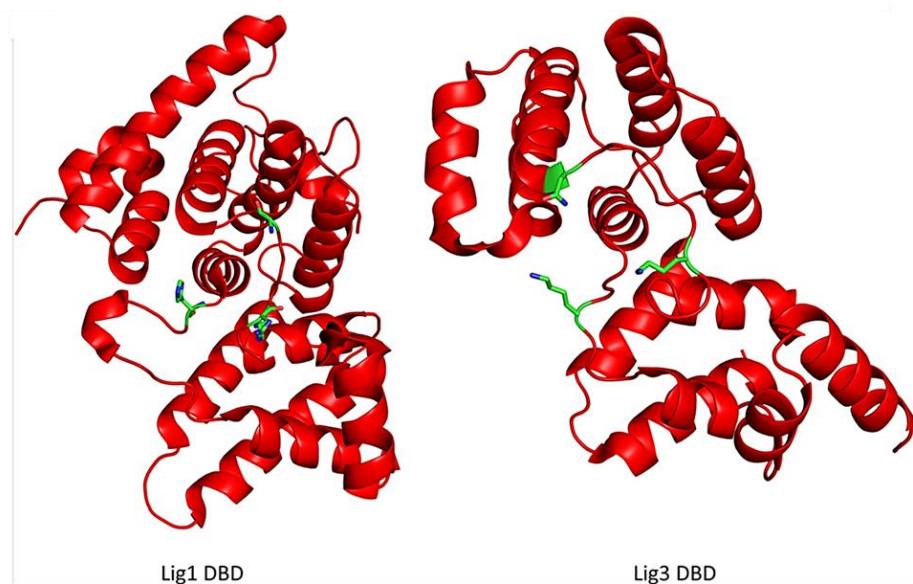


Figure 14 : Résidus impliqués dans la liaison à l'ADN au sein du domaine DBD des ligases humaines 1 et 3 et ciblés dans les stratégies de criblage rationnel d'inhibiteurs de ces interfaces (Tomkinson et al, 2013). Ces structures cristallographiques reflètent la conservation structurale de ces domaines au sein des ligases humaines.

SCR7 : Ce composé est un dérivé de la molécule L189 décrite précédemment. Par une approche de docking, les auteurs ont cherché à mettre au point une molécule spécifique de l'ADN Ligase IV humaine, ciblant le DBD (Srivastava et al., 2012). Cette étude est la première qui établit l'inhibition spécifique d'une ligase humaine par une approche partant du criblage virtuel jusqu'au modèle animal de tumeur (xenogreffe). Basé sur la structure du DBD des autres ligases humaines, un modèle par homologie du domaine DBD de la Ligase IV a été construit (aa 1-241). Ce domaine putatif a été soumis à une étape de criblage virtuel par docking au niveau des résidus critiques d'interaction à l'ADN. Des modifications *in silico* de l'inhibiteur L189 précédemment caractérisé ont permis de mettre au point un nouveau composé

relativement spécifique et énergétiquement favorable, que les auteurs ont baptisé « SCR7 » (Figure 15).

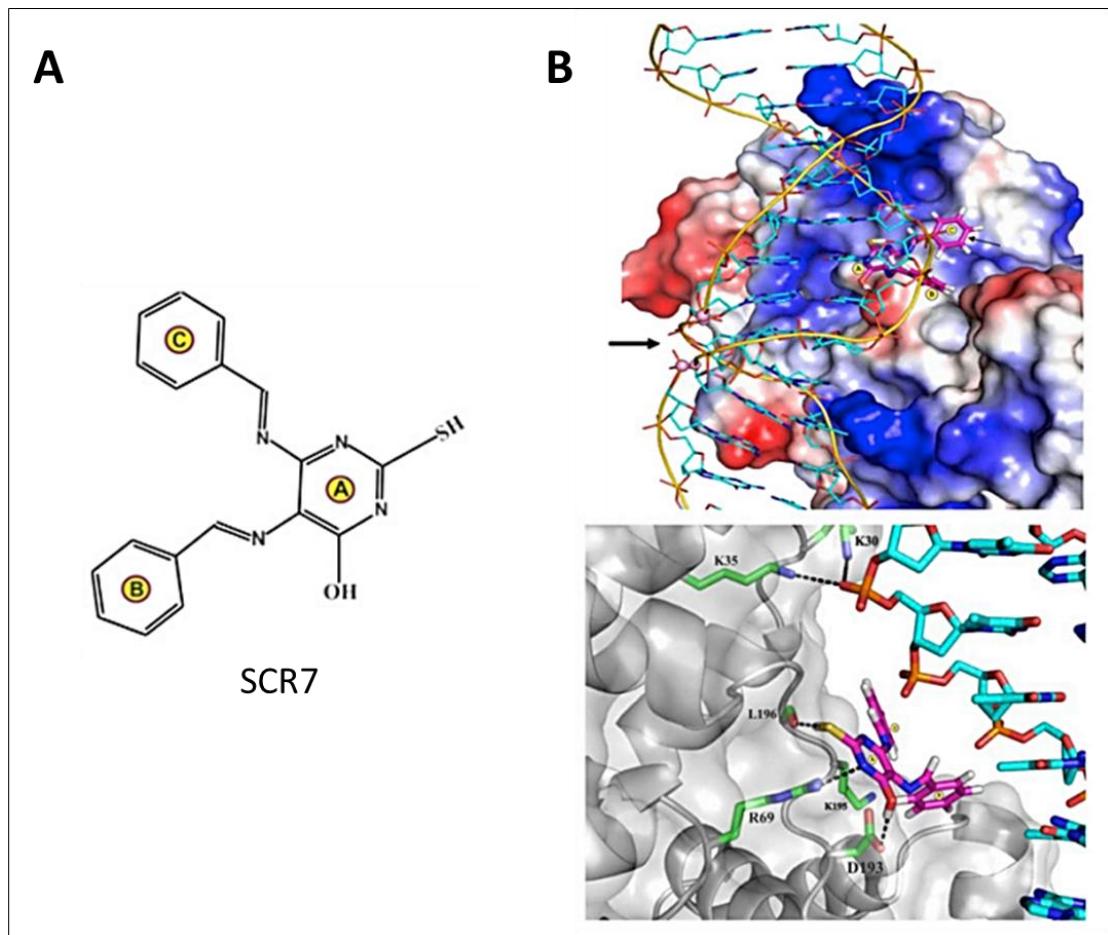


Figure 15 : Schéma et mode de liaison du composé SCR7, inhibiteur de l’interaction DBD de la Ligase IV humaine/ADN (Srivastava et al, 2012). (A) Structure chimique du composé SCR7 [5,6-bis(benzylideneamino)-2-mercaptop-pyrimidin-4-ol]. (B) Docking et analyse du mode de liaison et des interactions théoriques du composé SCR7 au sein du DBD à l’interface DBD/ADN (représentation en surface avec potentiel électrostatique (haut) et en cartoon (bas). Les liaisons hydrogènes impliquées dans l’interaction sont représentées en pointillés noir.

Des essais NHEJ avec des extraits cellulaires ont montré un effet dose-réponse de SCR7 et une activité inhibitrice à partir de 10 μ M. Dans les cellules, ce composé induit à lui seul une augmentation des CDB non réparées suivi d'un mécanisme d'apoptose médié par p53 suite à l'accumulation de ces CDB. De plus, SCR7 a été récemment utilisé pour inhiber la NHEJ dans des stratégies de correction du génome dont l'efficacité dépend de la RH (Maruyama et al., 2015).

Le rôle des ADN Ligases humaines dans les mécanismes de ligation, réPLICATION et voies de réparation de l'ADN les rendent tout à fait intéressantes pour le développement d'inhibiteurs agissant de manière synergique avec d'autres agents génotoxiques. Plusieurs revues récentes ont d'ailleurs permis de mettre en lumière l'utilité de ces ADN Ligases comme cibles thérapeutiques (Tomkinson et al., 2013). A travers les deux exemples précédents, l'inhibition de la Ligase IV humaine a montré un effet bénéfique et pourrait par conséquent ouvrir des pistes quant à la découverte d'autres interactants moléculaires afin de radiosensibiliser les cellules cancéreuses dans une stratégie de thérapie combinée. Cependant, ces études ont été focalisées sur le DBD de la ligase (dont la hLigI), et du fait que le domaine catalytique est conservé à travers les trois ligases humaines aucun de ces inhibiteurs n'est exclusivement spécifique de la ligase IV.

Afin de faire face à ce problème, il serait tout à fait intéressant d'inhiber la Ligase IV indirectement en modulant son interaction avec XRCC4. En effet, rappelons que cette ligase interagit avec XRCC4 via un domaine propre en C-terminal et qui n'est pas présent chez les autres ligases humaines. L'inhibition indirecte de l'activité de ligation via une interférence dans la formation du complexe XRCC4/Cer-XLF préviendrait de manière spécifique le mécanisme NHEJ dans les cellules. Ces interfaces protéine-protéine représentent donc un réel défi (du fait de la difficulté avérée de l'inhibition de telles interfaces) mais également une piste prometteuse dans un but de radiosensibilisation des cellules cancéreuses.

Une mini-revue (en cours de finalisation) a été rédigée au cours de cette thèse et présente les différentes stratégies et résultats obtenus ainsi que les futures perspectives dans la recherche d'inhibiteurs spécifiques de ligases humaine pour une thérapie anticancéreuse combinée. Celle-ci est présentée en p.246 de cette thèse.

Le complexe de ligature comme cible thérapeutique

Il est donc maintenant avéré qu'une inhibition spécifique du mécanisme NHEJ de réparation des CDB de l'ADN en combinaison avec d'autres agents génotoxiques pourrait avoir un bénéfice thérapeutique non négligeable. Si l'on fait le point concernant la recherche d'inhibiteurs pour chacun des acteurs du mécanisme NHEJ,

des failles importantes sont à noter bien que la stratégie fonctionne. Certains inhibiteurs ont un effet inhibiteur spécifique mais ne peuvent être bien exploités en clinique de par une faible pharmacocinétique ou une toxicité apparente (cas des inhibiteurs de DNA-PKcs). De nombreuses études de criblage virtuel ont été conduites dans un but de découverte rationnelle d'interactants moléculaires basée sur la structure tridimensionnelle des protéines ciblées. Dans le cas de XRCC4, des molécules (inhibiteurs potentiels de l'interaction avec l'ADN Ligase IV) ont été proposées mais aucune validation expérimentale n'a été menée à ce jour sur ces dernières. L'ADN Ligase IV humaine est une cible de choix dans ce mécanisme NHEJ car elle est l'unique ligase associée à cette voie et n'intervient dans aucune autre voie métabolique. Des tentatives d'inhibition de cette dernière, toujours via une approche rationnelle et assistée par informatique, a conduit à des composés fonctionnels mais non complètement spécifiques au vu de la conservation du domaine catalytique dans les ligases humaines.

Ainsi, les interfaces protéines-protéines du complexe XXL de ligature des CDB de l'ADN n'ont à ce jour pas encore été exploitées pour la recherche d'interactants capables de radiosensibiliser les cellules tumorales en interférant avec les complexes XRCC4/C-terminal de la ligase 4, et XRCC4/Cer-XLF (Figure 16).

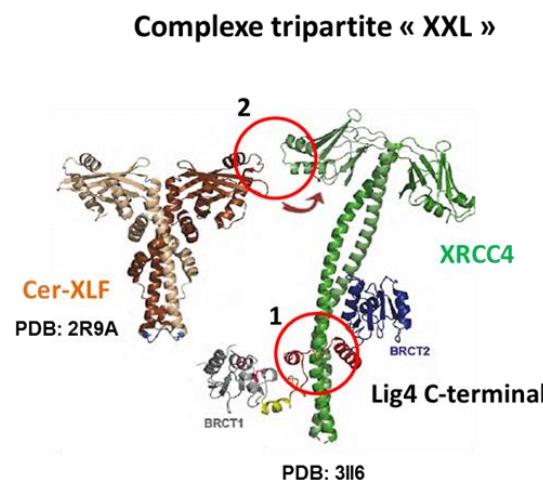


Figure 16 : Représentation des interfaces protéine-protéine du complexe XXL de ligature des CDB de l'ADN. L'interface 1 représente l'interaction de la région clamp de la ligase IV au sein du domaine tandem BRCT avec XRCC4. L'interface 2 représente l'interaction entre les têtes globulaires de XRCC4 et Cer-XLF. Ces deux interfaces ont été spécifiquement ciblées au cours de cette thèse.

S'intéresser à ces interfaces permettrait de s'affranchir du problème de spécificité rencontré en affectant uniquement le mécanisme NHEJ et en inhibant la ligase 4 soit directement via son association avec XRCC4, soit indirectement en interférant avec le complexe XRCC4/Cer-XLF. De plus, d'ores et déjà, nous possédons des preuves de concept sur l'effet thérapeutique de l'inhibition *in cellulo* de ces complexes protéiques. Ces preuves de concept ainsi que le système d'étude seront plus détaillés ultérieurement.

Les avantages d'inhiber l'interface XRCC4/C-terminal de la ligase 4 :

- Cette interface est unique et propre à la ligase 4 humaine et au mécanisme NHEJ. Nous n'attendons donc aucun problème d'effet cellulaire hors-cible comme celui rencontré lors de l'inhibition du domaine catalytique des ligases.
- C'est une interface innovante et encore non explorée dans la stratégie d'inhibition par des petites molécules.
- Une structure de cette interface est présente avec une bonne résolution, permettant la recherche rationnelle d'inhibiteurs potentiels.
- La modulation par compétition avec des peptides de cette ligase a permis de définir la zone prioritaire à cibler dans cette stratégie de criblage rationnel.
- Enfin, la ligase 4 étant instable lorsqu'elle est privée de son partenaire XRCC4, on s'attend à ce que cette interférence conduise à un effet d'autant plus efficace par dégradation consécutive de cette ligase.

Les avantages d'inhiber l'interface XRCC4/Cer-XLF :

- Cette interface est unique et propre au mécanisme NHEJ. Nous n'attendons donc aucun problème d'effet cellulaire hors-cible.
- C'est une interface innovante et encore non explorée dans la stratégie d'inhibition par des petites molécules.
- Une structure de cette interface est présente avec une bonne résolution, permettant la recherche rationnelle d'inhibiteurs potentiels.
- Une faible affinité entre les protéines XRCC4 et Cer-XLF (de l'ordre de 7 μ M) et le fait qu'une seule mutation ponctuelle suffise à compromettre totalement cette interaction est en faveur de l'utilisation d'une petite molécule pour dissocier ce complexe.

- Malgré que ces protéines ne possèdent aucune activité enzymatique connues, un test breveté de stimulation de la ligase du bactériophage T4 via ce complexe (Mauro Modesti, Murray Junop, Sara Andres, 2013 – « Method and kit for dna ligation associating human proteins XRCC4 and XLF with a ligase of microbial origin » - Numéro de brevet : WO 2013057457 A1 (Roy et al., 2015)) permettra de confirmer expérimentalement l'effet inhibiteur de nos molécules sur cet assemblage.

I – 3 : Inhibition cellulaire des protéines du complexe XXL de ligature et radiosensibilisation des cellules cancéreuses – Preuves de concept

Nous avons décrit précédemment les avantages de l'inhibition du complexe tripartite (XXL) de ligature des CDB de l'ADN. Ce complexe de ligature est caractérisé par une première interface protéine-protéine entre XRCC4 et le domaine C-terminal de la ligase 4 (deux en réalité, si l'on compte l'interaction entre le domaine BRCT2 de la ligase et le coiled-coil de XRCC4 en plus du domaine clamp), et une seconde interface protéine-protéine entre XRCC4 et Cer-XLF dont l'interaction s'effectue via leur tête globulaire respective (Figure 16).

Avant d'entamer une recherche d'inhibiteurs spécifiques de ces interfaces il est important d'établir un effet thérapeutique bénéfique lors d'une inhibition de ces dernières directement *in cellulo* (une preuve de concept). Ainsi, en modulant l'interaction entre XRCC4 et le domaine C-terminal de la ligase 4 et entre XRCC4 et Cer-XLF, ou en sous-régulant l'expression de l'une de ces protéines, l'effet thérapeutique majeur recherché va être une augmentation de la radiosensibilité des cellules concernées.

Inhibition *in cellulo* de l'interaction XRCC4/domaine C-terminal de la Ligase 4 :

En 2009, notre équipe a participé à une étude sur la structure cristallographique de ce complexe qui a caractérisé de manière plus précise les interactions critiques au sein de ce complexe (Wu et al., 2009). Nous décrirons plus en détail ce système d'étude dans la prochaine partie. Dans cette étude, elle a analysé l'effet cellulaire de l'expression ectopique de fragments de cette ligase 4. Ces fragments (dont le domaine C-terminal BRCT1-βhairpin-clamp-BRCT2, privé du domaine catalytique et

de liaison à l'ADN), ont montré un effet dominant négatif par compétition avec la ligase 4 endogène et pleine taille, pour la liaison à XRCC4. Cette compétition n'était possible que si ces fragments contenaient au moins le domaine clamp-BRCT2. Ce domaine BRCT2 a donc été démontré comme indispensable à la stabilisation endogène de la ligase 4 sur la protéine XRCC4. D'un point de vu structural, il est nettement visible qu'une partie de ce BRCT2 (hélice alpha1') vient former des contacts avec la partie coiled-coil de XRCC4.

Ce mécanisme intracellulaire de compétition conduit à une dégradation de la ligase 4 endogène (visible en western blot) car cette dernière est instable lorsqu'elle n'est pas en interaction avec son partenaire (Figure 17).

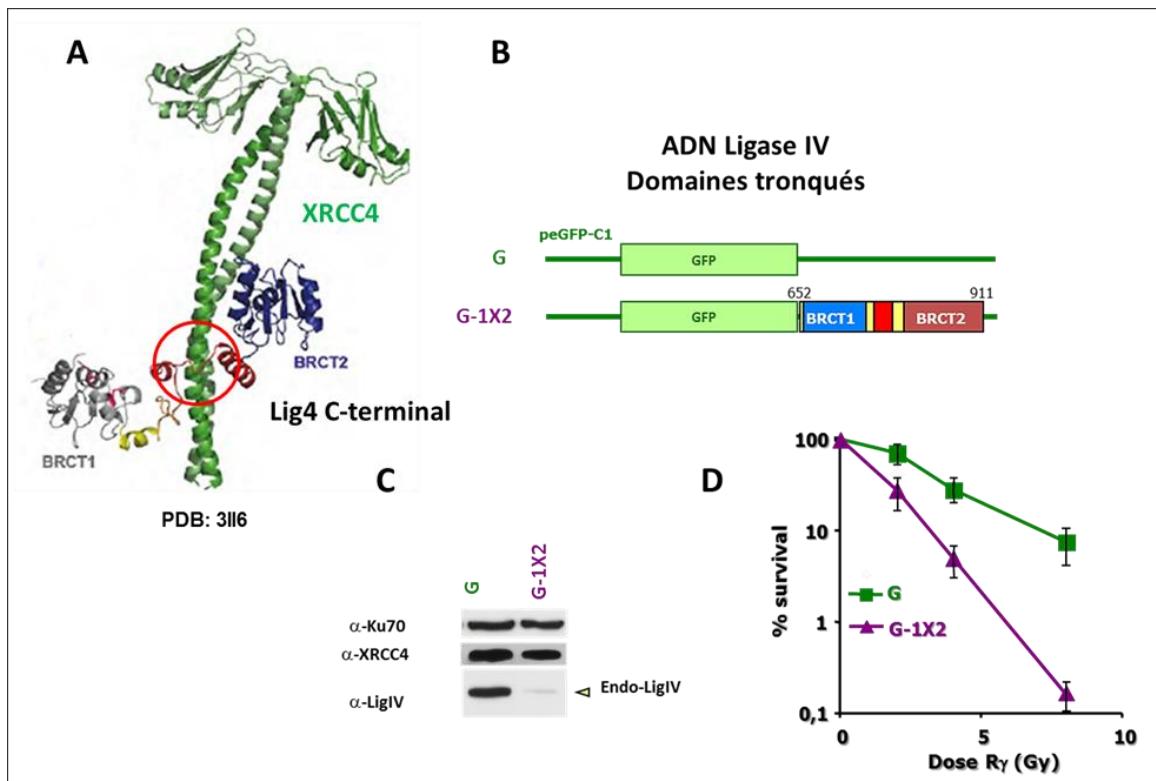


Figure 17 : Déstabilisation cellulaire du complexe endogène XRCC4/Ligase IV par expression ectopique de fragments de la Ligase IV. (A) Structure cristallographique du complexe XRCC4/Domaine C-terminal de la Ligase IV (PDB code : 3II6). (B) Deux exemples de constructions surexprimées dans les cellules. La construction G-1X2 contient le domaine C-terminal tandem BRCT (représenté en (A)), fusionné à la GFP. (C) Expérience de Western blot montrant la dégradation de la protéine Ligase IV endogène suite à l'expression ectopique du fragment G-1X2 (en comparaison avec le fragment contrôle G, correspondant à la GFP seule). (D) Mesure de la survie cellulaire (clones fibroblastiques MRC5 humains) après irradiation γ et en réponse à l'expression ectopique des domaines G et G-1X2.

L'expression du fragment G-1X2 augmente de manière significative la radiosensibilité des cellules. D'après (Wu et al., 2009).

De façon consécutive, la diminution de ligase 4 endogène conduit à un défaut de recrutement d'un complexe de ligature pleinement fonctionnel aux sites de CDB. Les extraits cellulaires correspondants ont été soumis à des essais NHEJ *in vitro* dans lesquels ils se sont avérés inefficaces. Finalement, cet effet dominant négatif, conduit à une augmentation de la radiosensibilité des cellules en réponse à des radiations γ (Figure 17). Cette étude démontre clairement la faisabilité d'une stratégie de radiosensibilisation des cellules par interférence avec le complexe XRCC4/domaine C-terminal de la ligase 4 et appuie l'hypothèse selon laquelle l'effet d'une petite molécule inhibitrice de cette interface (par liaison sur l'une ou l'autre des protéines au site d'interaction), pourrait conduire au même effet biologique. Cela suppose également que cette molécule doit être capable de franchir les membranes biologiques.

De la même manière, la sur-expression *in cellulo* des fragments de la protéine XRCC4 via un adénovirus, conduit à une inhibition de l'interaction XRCC4/Ligase 4 endogène et donc une augmentation de la radiosensibilité de cellules tumorales (Jones et al., 2005).

Inhibition *in cellulo* des protéines Cer-XLF et XRCC4:

La protéine Cer-XLF interagit directement avec le complexe XRCC4/Ligase 4 *in vitro* et *in vivo*. De plus, il a été montré qu'un knock-down de Cer-XLF par siRNA dans des cellules de mammifères, conduisait à une augmentation de la radiosensibilité de ces cellules ainsi qu'un défaut de NHEJ. La réintroduction de cette protéine sauvage corrige le phénotype observé (Ahnesorg et al., 2006).

En 2015, une équipe a caractérisé une nouvelle mutation au niveau du domaine N-terminal de XRCC4 (W43R). Cette substitution, qui remplace un résidu hydrophobe par un résidu chargé positivement semble affecter la stabilité intrinsèque de la protéine en perturbant un motif hélice-tour-hélice au sein d'un cœur hydrophobe de la tête globulaire (distorsion). Des études cellulaires et biochimiques montrent que cette déstabilisation, induisant principalement un défaut de solubilité de la protéine,

cause une réduction de l'activité de la ligase 4 et une réparation médiée par la NHEJ compromise après traitement par des radiations ionisantes (Murray et al., 2015). La sous régulation des protéines XRCC4 et Cer-XLF indique clairement un défaut de réparation NHEJ conséquent. Cependant, notre stratégie ne sera pas ici de sous-réguler l'une ou l'autre de ces protéines mais plutôt d'intervenir au niveau de leur interaction propre.

Inhibition *in cellulo* de l'interaction XRCC4/Cer-XLF:

A la suite de CDB, Cer-XLF s'accumule aux sites de cassure via une interaction avec la tête globulaire de XRCC4, déjà impliqué dans le complexe XRCC4/Ligase 4. A la suite de cela, Cer-XLF va stimuler la liaison d'extrémités d'ADN complémentaires ou non-complémentaires via ce complexe.

Cer-XLF stimule la réadénylation de la ligase 4 (Riballo et al., 2009). Des expériences de co-immunoprecipitation avec des fragments de la ligase 4 ont montré que Cer-XLF interagissait avec cette dernière via le domaine BRCT1 (Wu et al., 2009). Il n'est cependant pas très clair à ce jour, si les interactions entre les partenaires du complexe tripartite se mettent en place dans un contexte de formation de filaments entre XRCC4 et Cer-XLF. Un argument en faveur de ce scénario est la mise en évidence de filaments XXL en cellules via de la microscopie de fluorescence super-résolution (Reid et al., 2015). On pourrait cependant supposer que ces interactions permettent la stabilisation ou un changement conformationnel de la ligase IV aidant ainsi à stimuler le mécanisme d'adenylation. Cependant, l'interaction XRCC4/Cer-XLF ne semble pas nécessaire à la stimulation de la ligase IV (Roy et al., 2015).

Une étude de microscopie électronique en 2009 rapporte d'ailleurs la structure basse résolution d'un complexe XRCC4/Ligase 4 et suggère que le large domaine catalytique de la ligase 4 se positionnerait au niveau de l'une des deux têtes globulaires de XRCC4 (Recuero-Checa et al., 2009).

La preuve de concept d'un défaut de réparation NHEJ après dommage à l'ADN par une perte d'interaction entre les protéines XRCC4 et Cer-XLF a été apportée en

2010 via une étude structurale et de mutagénèse dirigée (Malivert et al., 2010). Des mutants de Cer-XLF, localisés au niveau du site d'interaction avec XRCC4 sur la tête globulaire de Cer-XLF (résidus critiques : R64, L65 et L115), n'induisaient pas de changement d'expression et de localisation intracellulaire de la protéine mais un défaut clair de réparation de l'ADN après induction de CDB et une diminution de la survie cellulaire (Figure 18).

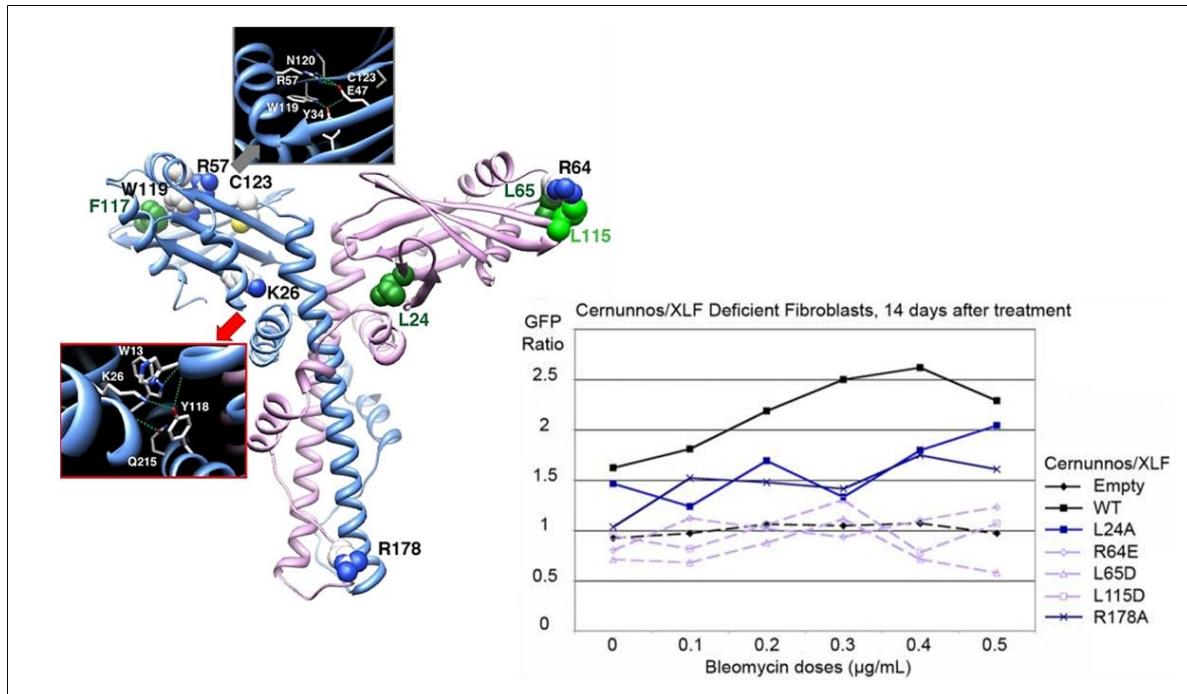


Figure 18 : Structure cristallographique de la protéine Cer-XLF (PDB code : 2R9A) soulignant les résidus critiques d'interaction à XRCC4 et survie cellulaire pour des cellules déficientes en Cer-XLF après traitement à la bléomycine (Malivert et al, 2010). A gauche : structure en cartoon de la protéine Cer-XLF et résidus d'interaction (R64, L65 et L115). A droite : Courbes de survies pour des cellules fibroblastiques déficientes en Cer-XLF et transduites avec un vecteur codant pour la protéine Cer-XLF sauvage ou mutée.

Ainsi, et de la même manière que pour l'interférence entre XRCC4 et le domaine C-terminal de la ligase 4, il serait tout à fait envisageable qu'une petite molécule inhibant l'interaction XRCC4/Cer-XLF provoque le même phénotype. A ce jour, aucun inhibiteur de l'interface XRCC4/Cer-XLF n'est connu. A l'inverse de l'interface précédente, les résidus critiques intervenant dans cette interaction sont connus et ont été caractérisés par mutagénèse dirigée (Ropars et al., 2011). Ces résidus clé sont les résidus à cibler dans une approche de criblage rationnel d'inhibiteurs. Dans le cas d'un criblage virtuel par docking, la recherche aléatoire devra s'effectuer

autour de ces résidus afin de trouver des molécules capables d'interagir avec ces derniers et avec une bonne affinité théorique.

En sachant l'affinité faible entre ces deux protéines, les résidus critiques impliqués dans cette interaction et la preuve de concept d'une radiosensibilisation des cellules par interférence avec ce complexe, la recherche de petites molécules inhibitrices de ce complexe pourrait être (en plus de l'inhibition XRCC4/C-terminal de la ligase 4) une toute nouvelle stratégie dans le fait de compromettre la NHEJ et radiosensibiliser les cellules cancéreuses.

I – 4 : Les interactions Protéine-Protéine : un défi thérapeutique

Les interactions protéine-protéine (PPI) sont les bases moléculaires à l'origine de presque tous les processus physiologiques et sont souvent dérégulées lors de certaines maladies graves. Dans les cancers, les PPI peuvent avoir des rôles régulateurs importants comme dans le cycle de division cellulaire ou la signalisation cellulaire. Ainsi, des changements d'affinité ou de spécificité au sein de ces interactions peuvent conduire à des malfonctions telles qu'une croissance cellulaire non contrôlée (Nero et al., 2014). Une protéine ne fonctionne jamais en tant qu'entité unique mais est au cœur d'un vaste réseau d'interactions avec d'autres composants cellulaires. On estime à environ 650.000 PPI pour environ 100.000 protéines produites au sein de la cellule (l' « interactome » protéine-protéine) (Kelly and Stumpf, 2008). Aujourd'hui, plus de 300.000 paires d'interactions ont été identifiées dans le génome humain (Nero et al., 2014). L' « interactome » humain représente ainsi un vaste champ de cibles thérapeutiques intra et extracellulaires potentielles et donc une opportunité certaine pour des interventions pharmacologiques. Ces PPI sont l'un des enjeux majeurs aujourd'hui en recherche post-génomique pour comprendre, représenter et inhiber ces dernières et restent une branche encore peu exploitée en recherche d'inhibiteurs pharmacologiques.

Caractéristiques structurales et ciblage des PPI :

Pendant longtemps, les PPI ont été considérées comme des cibles non accessibles à des molécules thérapeutiques (non « druggables ») (Thompson et al., 2012) (et tout particulièrement par des petites molécules synthétiques jusqu'alors utilisées pour

d'autres types de cibles thérapeutiques) au vu du fait de la nature de leur interface (large, plate et sans cavité apparente, hautement flexible). Cependant, plusieurs succès récents ont permis aux chercheurs d'être plus confiants quant à la possibilité d'un tel ciblage (Chene, 2006; Fry, 2006; Kuenemann et al., 2015; Villoutreix et al., 2014; Wells and McClendon, 2007). D'ailleurs, on dénombre actuellement plus de 12 petites molécules modulatrices de PPI (oncogéniques) en développement clinique (quelques exemples ci-dessous; Figure 19).

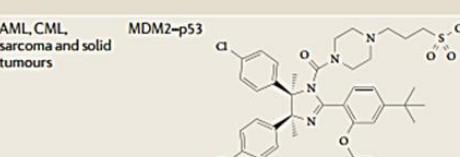
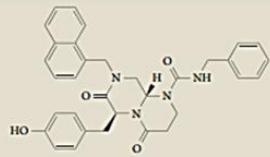
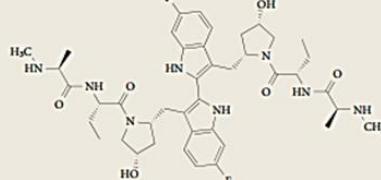
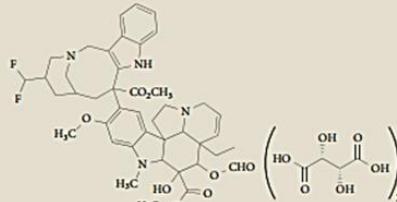
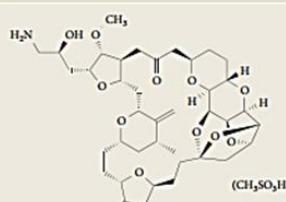
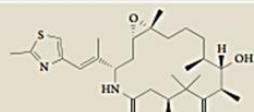
Cancer type	Target PPI	Chemical structure	Molecular mass (cLogP)	Generic name (type of modulator)	Clinical phase
AML, CML, sarcoma and solid tumours	MDM2-p53		728 Da (10.6)	RG7112 or RO504537 (orthosteric inhibitor)	Phase I
AML, CML, pancreas and solid tumours	β -catenin-CREB-binding protein		549 Da (6.1)	PRI-724 or ICG-001 (orthosteric inhibitor)	Phase I/II
AML, lymphoma and solid tumours	IAP family-SMAC		807 Da (2.8)	TL32711 (orthosteric inhibitor)	Phase I/II
Bladder	Tubulin- α -tubulin- β		817 Da (5.4) free base; 1117 Da, tartrate salt	Vinflunine ditartrate or PM391 (allosteric inhibitor)	In the clinic
Breast	Tubulin- α -tubulin- β		730 Da (1.2) free base; 826 Da, mesylate salt	Eribulin mesylate or E7389 (allosteric inhibitor)	In the clinic
Breast	Tubulin- α -tubulin- β		507 Da (3.1)	Ixabepilone (allosteric stabiliser)	In the clinic

Figure 19 : Exemple de modulateurs ortho- ou allostériques pour des interfaces protéine-protéine oncogéniques et en développement clinique. (Nero, Morton et al, 2014).

Ces différents cas de succès ne seront pas détaillés ici mais l'existence de ces modulateurs (naturels ou synthétiques) valide l'approche générale d'une interférence pharmacologique avec ce type de cible. Ces PPI sont très étudiées depuis des années et ont fait l'objet de nombreuses revues (Bahadur and Zacharias, 2008) (Reichmann et al., 2007). Elles forment un réseau très complexe et utilisent différents modes d'interaction en fonction du complexe protéique (complexe permanent ou transitoire, interaction à caractère obligatoire ou non-obligatoire, complexe homo ou heteromultimérique...). Ces interfaces sont beaucoup plus larges (de l'ordre de 1500 à 3000 Å²) en comparaison avec les interfaces d'interaction protéine-ligand (de l'ordre de 300 à 1000 Å²). Des méthodes originales et récentes permettent de caractériser les associations protéine-protéine en définissant de manière relativement précise le site et les résidus impliqués dans cette interaction. Cette caractérisation peut s'avérer essentielle dans le cas d'un désir de vouloir inhiber spécifiquement ladite interaction. C'est le cas par exemple du « footprinting » par marquage au tritium (Mousseau et al., 2010) ou de la méthode dite de « protein painting » (Luchini et al., 2014).

Des caractéristiques bien particulières définissent ces PPI et doivent par conséquent être prises en compte lors d'un criblage rationnel d'inhibiteurs. Ces caractéristiques concernent par exemple la taille, la géométrie, la flexibilité (plasticité), les régions « hot-spot » ainsi que le rôle des molécules d'eau à l'interface des partenaires. La notion d'index de planarité des interfaces (interface plus ou moins plane ou incurvée) a également été étudiée (Jones and Thornton, 1996). En fonction du type de complexe, la nature de l'interface peut différer de manière significative en terme par exemple de ratio résidus hydrophobes et aromatiques/ résidus polaires. En général, ces interfaces sont enrichies en acides-aminés aliphatiques et aromatiques (bien que des charges puissent être présentes) et consistent en un réseau de résidus « hot-spots » non connectés (Chene, 2006; Wells, 1996). Une des caractéristiques observées pour ces interfaces est la présence d'un cœur conservé et enrichi en acides aminés hydrophobes et aromatiques, entouré par un anneau (« rim »), accessible au solvant (Janin et al., 2008).

Les molécules d'eau ont un rôle primordial dans certaines PPI. En effet, l'exclusion des molécules d'eau consécutive à l'interaction entre les partenaires engendre un

effet hydrophobe favorable à cette interaction. De même, certaines de ces molécules d'eau restent présentes à l'interface et stabilisent l'interaction entre les deux protéines via un réseau de liaisons hydrogènes.

Finalement, la notion de flexibilité de ces interfaces est primordiale. Les interfaces protéine-protéine ont effectivement une grande plasticité et une nature adaptative. Des sillons ou cavités jusqu'alors invisibles sur des partenaires isolés (au sein d'une structure expérimentale) (Cozzini et al., 2008) peuvent apparaître au cours d'une simulation de dynamique moléculaire ou en conséquence de la liaison d'une molécule. On parle alors de cavités cryptiques. Ces cavités, proviennent d'un réarrangement de certaines chaînes latérales après, par exemple, la liaison d'un ligand, créant certaines cavités dans le but de mieux accommoder ce dernier (Arkin et al., 2003) (Figure 20).

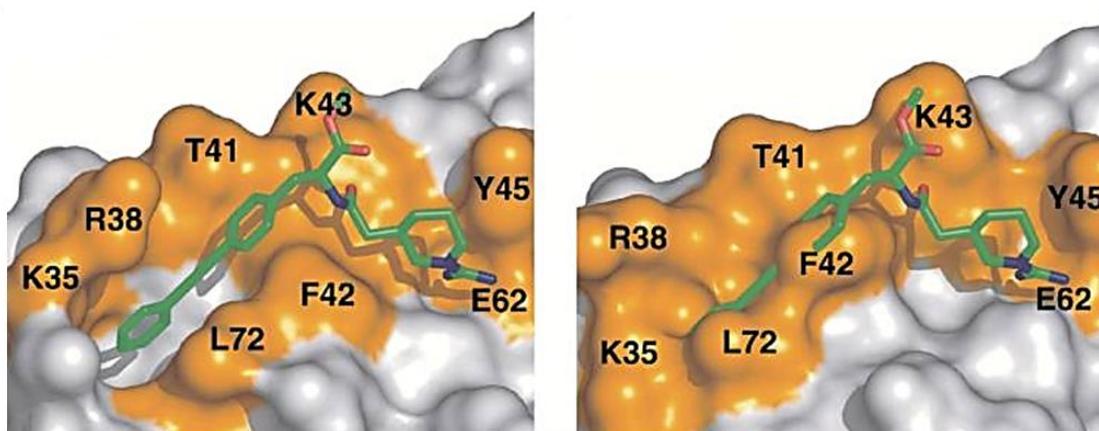


Figure 20 : Représentation de la surface adaptative de la protéine IL-2 suite à la liaison d'un inhibiteur spécifique (Arkin et al, 2003). La dynamique de la surface de liaison est représentée avec les résidus hot-spot colorés en orange. A gauche : détails du mode d'interaction du composé au sein du complexe IL-2/inhibiteur. Les résidus accommodent l'inhibiteur qui vient alors se loger parfaitement au sein du site. A droite : superposition du composé précédent sur la structure de la protéine IL-2 apo. Le site de liaison présente une topologie de surface tout à fait différente en comparaison avec la forme liée soulignant la plasticité du site de liaison.

Des petites molécules peuvent donc sélectionner une conformation protéique de basse énergie (ce que l'on appelle la « sélection de conformation »), et induire un changement conformationnel après liaison à sa cible (qui est donc un mécanisme d'adaptation de la protéine au ligand) (Wells and McClendon, 2007). Certaines études visent à prédire l'apparition de ces cavités par des calculs de simulation de

dynamique moléculaire. Ces prédictions peuvent être une piste dans la tentative de démontrer la « druggabilité » d'une cible protéique potentielle (Sugaya, 2009), en particulier si ces cavités sont au niveau ou à proximité de résidus hot-spot.

En conclusion, nous pouvons tout à fait admettre que les PPI sont, certes, un réel défi mais restent pour certaines des cibles accessibles à des petites molécules conduisant à un effet thérapeutique. Ces molécules peuvent être stabilisatrices ou déstabilisatrices des interfaces protéine-protéine (Thiel et al., 2012). La recherche d'inhibiteurs pharmacologiques spécifiques de certaines de ces interfaces a pu conduire à des molécules aujourd'hui en développement clinique.

Au-delà de ces récents succès, plusieurs indications nous permettent d'affirmer que ces interfaces sont effectivement « druggables ». Par exemple, l'analyse comparative des PPI et des interactions protéine/peptide montre des comportements thermodynamiques très similaires malgré la différence de taille des interfaces (Gonzalez-Ruiz and Gohlke, 2006). Les interfaces ne présentant aucun sillon ou cavité distincte peuvent être soumises à une analyse computationnelle et laisser entrevoir au cours de la simulation, des cavités cryptiques et transitoires et pouvant avoir une validité biologique. Il a été montré que ce type de cavité pouvait d'ailleurs avoir une taille parfois similaire à une cavité créée par un ligand spécifique (Eyrisch and Helms, 2009). Ces cavités cryptiques peuvent alors être ciblées lors d'une campagne de criblage par docking et afin de trouver des composés potentiellement de bonne affinité. Au niveau des régions « hot-spot », il a été noté que les résidus concernés à l'interface protéine/ligand corrélaient souvent avec ceux impliqués dans les PPI (les mêmes acides aminés peuvent être impliqués dans les deux types d'interaction) (Jubb et al., 2012). Ces régions « hot-spot » sont donc une cible thérapeutique avec une grande propension pour la liaison de petites molécules. Par ailleurs, l'aire de la région contenant les résidus « hot-spot » est en moyenne $\approx 600\text{\AA}^2$, compatible avec la taille d'une entité de bas poids moléculaire ; des études de mutagenèse dirigée (« alanine-scanning ») ont montré que l'on pouvait complètement déstabiliser une PPI en mutant seulement quelques résidus « hot-spot » impliqués dans cette dernière (Clackson and Wells, 1995) voire un seul

(Ropars et al., 2011). La stabilité de ces PPI n'est donc pas dépendante de la taille de l'interface et le ciblage par une petite molécule est tout à fait envisageable.

Quelles chimiothèques pour les PPI ?

Un modulateur d'une PPI est défini comme un composé capable de directement entrer en compétition avec l'un des deux partenaires protéiques au sein d'une interface discontinue (Wells and McClendon, 2007). La vaste majorité des modulateurs actuellement connus sont des biomolécules telles que des anticorps ou des peptides, ou bien des petites molécules naturelles ou de synthèse. Les antibiotiques sont hautement spécifiques et stables dans la circulation sanguine mais leurs problèmes majeurs sont une faible biodisponibilité, pas de perméabilité cellulaire (ciblage uniquement de structures extracellulaires accessibles) et un coût très élevé de production. Les peptides sont peu stables, ont également une faible biodisponibilité et peuvent induire des réactions immunitaires (comme les petites molécules). Cependant, ces derniers restent un outil de choix en biologie et particulièrement dans une approche de ciblage thérapeutique car ils apportent, en amont, une preuve de concept quant au ciblage de certaines cibles avant une campagne de criblage coûteuse (Neduva and Russell, 2006). Des modifications chimiques de ces peptides permettent de stabiliser la structure secondaire de ces derniers et prévenir par conséquent leur dégradation. Des études dans cette voie ont par exemple permis le développement de peptide « agrafé » (les « stapled peptides ») (Walensky et al., 2004). Finalement, l'utilisation de composés de bas poids moléculaire (petites molécules naturelle ou synthétique), reste une bonne option dans le ciblage thérapeutique des PPI. Ces petites molécules peuvent alors agir via différents mécanismes : orthostérique (si la liaison du composé se situe à l'interface) ou allostérique (liaison sur un site éloigné de l'interface protéine-protéine mais induisant un changement conformationnel affectant la liaison du partenaire protéique au niveau de cette interface). Nous ne considérerons dans cette thèse que les molécules agissant via un mécanisme orthostérique.

Ces molécules de bas poids moléculaire, doivent être des molécules dites « Drug-Like », c'est-à-dire posséder des propriétés ADME/Tox (Absorption, Distribution, Métabolisme, Excrétion / Toxicité) satisfaisantes. Le premier problème rencontré est

que toutes les molécules virtuellement disponibles au sein des bases de données ne rentrent pas dans ces critères et un filtrage minutieux devra alors être considéré afin de pouvoir proposer des candidats en adéquation avec un programme de développement clinique. Cependant, les molécules inhibitrices ne rentrant pas dans ces critères ADME/Tox pourrait s'avérer tout de même être un bon outil biologique afin de sonder et comprendre certains mécanismes moléculaires.

Un autre problème, qui lui est spécifique du type d'interface ciblé (en l'occurrence des PPI), est que les collections de composés actuellement disponibles ont été depuis des années enrichies par des molécules visant à inhiber des sites catalytiques plutôt que des PPI, les rendant ainsi pour la plupart non adaptées à ces PPI. En effet, les enzymes et récepteurs de la surface cellulaire représentent plus de 80% des cibles aujourd'hui ciblées par des petites molécules approuvées par la FDA (Hopkins and Groom, 2002). Typiquement, une drogue administrée par voie orale, possède un poids moléculaire < 500Da. Cette taille est idéale pour le ciblage d'une petite cavité au sein d'un site catalytique. Mais les interfaces protéine-protéine sont beaucoup plus larges et n'ont pas évoluées dans le but d'accompagner de telles molécules. Jusqu'à récemment, aucune stratégie n'avait vraiment été développée pour la création de bases de données dédiées aux PPI à la différence des bases de données dédiées au ciblage des enzymes, RCPGs (Récepteurs Couplés aux Protéines G), canaux ioniques et autres cibles (Sperandio et al., 2010). Des lois empiriques ont été créées afin de définir les propriétés physico-chimiques nécessaires pour une bonne biodisponibilité des molécules ciblant ce type de récepteur. Mais l'espace chimique représentatif de ces molécules ne couvre certainement pas celui nécessaire au ciblage des PPI.

Malgré cela, il est à noter que certaines interfaces protéine-protéine peuvent être adressées avec ce type de composé filtré par exemple via des critères en accord avec les lois de Lipinski (« la règle des 5 » (Lipinski et al., 2001)), ce qui rend tout de même possible, dans certains cas, l'utilisation de ces bases de données (Pagliaro et al., 2004).

Du fait du faible taux de hit rencontré lors des campagnes de criblage contre ces PPI, des auteurs ont proposé qu'il été nécessaire de centraliser les collections de composés spécifiques des PPI ainsi que leur analogues structuraux (Sperandio et al., 2010). Certaines études visaient à définir des descripteurs moléculaires permettant de discriminer un modulateur de PPI d'un autre type de modulateur pour enrichir des bases de données (Neugebauer et al., 2007). Aujourd'hui, ce type de banque centralisant des inhibiteurs d'interfaces protéine-protéine existe. C'est le cas par exemple de la base de données « 2P2I », une base de donnée structurale annotée manuellement et qui centralise plus de 250 modulateurs orthostériques, correspondant à 17 familles de cibles biologiques et 29 complexes protéine-protéine (version 2015) (Morelli et al., 2011) (Sperandio et al., 2015). Une analyse statistique de ces inhibiteurs de PPI a permis à ses auteurs de définir ce qu'ils ont appelés la « règle des 4 » (MW>400Da ; ALogP>4 ; #Rings>4 et #HBA>4). Cette règle pourrait définir le profil générique d'un composé inhibiteur d'une PPI et pouvant ainsi entrer dans une phase d'optimisation chimique. Il est à noter que ces paramètres contredisent le dogme traditionnel de la chimie médicinale qui est cette « règle des 5 ». Un autre exemple de base de données est appelée « TIMBAL » (Higueruelo et al., 2013), qui regroupe 7000 inhibiteurs de 50 PPI connues. D'autres études centrées sur ce type d'inhibiteur ont montré que ces molécules étaient en général plus hydrophobes, aromatiques et larges en comparaison avec d'autres drogues connues et utilisées (Sperandio et al., 2010). De plus, certaines drogues naturelles (paclitaxel, rapamycine...) ou synthétiques (ABT 263) ne s'accordent pas avec les règles traditionnelles d'ADME/Tox mais restent cependant oralement biodisponibles. Un critère, qui n'est pas toujours rencontré, mais présent pour beaucoup de modulateurs d'interfaces protéine-protéine est la notion de tridimensionnalité de la molécule. Cette caractéristique est assez fréquente au niveau des inhibiteurs avérés et validés (cas par exemple d'un inhibiteur de la protéine MDM2 (Vassilev, 2004)). Une molécule plane et hydrophobe peut cependant atteindre une bonne affinité si l'on en croit des études de docking réalisées sur le site de liaison de la protéine MDM2 (Vassilev, 2004). La « globularité » de ces molécules peut avoir deux origines :

- Le composé peut avoir une partie 'œur', d'où projettent différents 'bras' dans des plans multiples (exemple du complexe Nutlin-3-MDM2) (Figure 21) ;

- Ou bien posséder un scaffold non nécessairement tridimensionnel mais dont le caractère tridimensionnel apparaît via un stacking intramoléculaire au sein de la molécule (exemple de l'inhibiteur de TNF- α) (Figure 21).

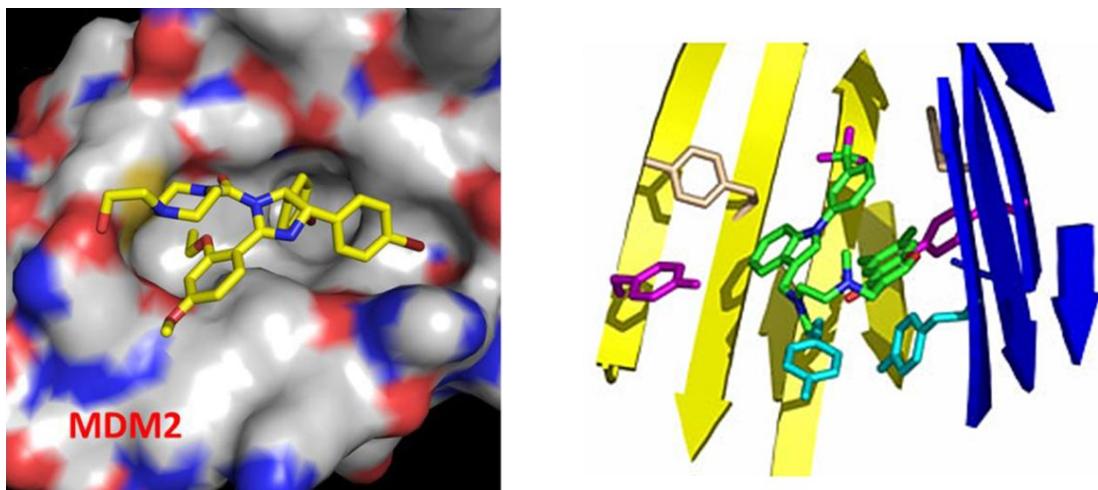


Figure 21 : Illustration des composés inhibiteurs des protéines MDM2 et TNF α soulignant le caractère tridimensionnel de certains modulateurs d'interactions protéine-protéine. A gauche : structure en surface avec potentiel électrostatique de la protéine MDM2 complexée à un inhibiteur (système « cœur » et « bras ») (Vassilev et al, 2004). A droite : structure en cartoon de la protéine TNF α complexée à un inhibiteur (système stacking intramoléculaire) (He et al, 2005).

Un avantage évident de ce comportement serait que cette tridimensionnalité préexisterait en solution afin de minimiser les contacts hydrophobes avec le solvant. Mais également, le composé étant dans une conformation bioactive, il éviterait une certaine pénalité énergétique associée à une perte d'entropie conformationnelle (favorisant ainsi la liaison à la protéine). Un bénéfice additionnel pourrait être une capacité plus grande à pouvoir déplacer des portions flexibles de la protéine, induisant la formation de poches de liaison (exemple des modulateurs de la famille Bcl-2, (Feng et al., 2007)). Des molécules non suffisamment rigides n'auraient peut-être pas cette influence.

Un logiciel en libre accès, appelé « PPI-HitProfiler » permet de filtrer une chimiothèque à partir des descripteurs définis précédemment afin de ne sélectionner que des molécules ayant un fort potentiel pour inhiber une interaction protéine-protéine (Sperandio et al., 2015). Ce programme a été validé sur des données

expérimentales et est implémenté au sein de l'outil « FAF-drugs3 » permettant l'application de filtres selon notamment des critères d'ADME-Tox.

Au-delà du problème de devoir s'adapter à des nouveaux chemotypes pour l'inhibition des PPI, nous nous confrontons également à un défi plus technique : celui de devoir trouver des techniques haut-débit adaptées au criblage d'inhibiteurs de ces interfaces. La mise en place de tests habituels (FRET, BRET, SPR, RMN...) est parfois difficile et mal adaptée à cette notion de haut-débit (Congreve et al., 2005) (Colas, 2008).

Ainsi il apparaît évident qu'une recherche de modulateurs de PPI basée sur la structure, via un criblage *in silico* avec différentes stratégies chemo-informatiques, peut s'avérer être une bonne option.

Le criblage de modulateurs de PPI par des approches computationnelles :

1. Approches basées sur les pharmacophores

Un pharmacophore est défini par l'IUPAC (Wermuth et al, 1998) comme « un ensemble de caractéristiques stériques et électroniques nécessaire à une interaction supramoléculaire optimale avec une cible biologique spécifique et dans le but de moduler sa réponse biologique ». Bien que les pharmacophores soient d'ordinaire utilisés pour aligner et comparer des ligands ciblant la même protéine (Leach et al., 2010), ce concept peut être transféré aux PPI pour lesquelles l'un des partenaires sera considéré comme le récepteur et le second sera le ligand. Le pharmacophore résultant peut alors être utilisé afin d'identifier des touches (« hits ») remplissant ces critères pharmacophoriques à partir d'une chimiothèque virtuelle. Plusieurs programmes comme par exemple « Pocket Query » (Koes and Camacho, 2012), sont utilisés dans ce but. En général, le pharmacophore sera déterminé après identification des résidus « hot-spots » au sein de l'interface protéine-protéine. Si plusieurs structures sont disponibles (cristallographiques et/ou issues de la RMN), un modèle consensus sera déterminé après alignement de chaque pharmacophore correspondant. Par exemple, des auteurs ont utilisés cette technique afin d'identifier des inhibiteurs de l'interaction p53-MDM2 (Xue et al., 2013).

2. Approches basées sur le docking

Le docking est l'approche la plus intuitive afin de prédire la pose d'un ligand au sein d'une structure protéique. Cependant, cette technique est confrontée à quelques faiblesses telles que l'incapacité de prédire avec une relative exactitude, l'énergie libre de liaison du composé mais également l'incapacité de classer de manière exacte les différents « hits » par ordre d'affinités décroissantes (Damm-Ganamet et al., 2013). De plus, les programmes de docking ont, encore une fois, été particulièrement adaptés à la prédiction de pose d'une molécule au sein de cavités bien définies telles que dans les RCPGs ou les enzymes.

Pour ce criblage virtuel, une base de données virtuelle de parfois plus d'un million de molécules (souvent commercialement disponibles) sont « amarrées » au sein d'un site de liaison putatif au niveau de l'interface protéine-protéine considérée et un score leur est affecté permettant de retrouver les molécules se liant avec la meilleure énergie théorique (définie par une fonction score qui dépend du programme de docking utilisé). Les « hits » possédant les meilleurs scores sont alors testés expérimentalement. Récemment, des outils sont à disposition des utilisateurs désirant mener des expériences de criblage virtuel dans des projets de chimie-biologie et de découverte de médicaments. Ces outils, tel que le serveur web « MTiOpenScreen » (Labbe et al., 2015), permettent un accès facilité au docking de ligand en proposant également des banques de données de molécules (dont une base de données enrichie en molécules adaptées pour l'inhibition des interactions protéine-protéine). Plusieurs études menées sur des PPI pour l'identification de composés inhibiteurs et de bas poids moléculaire par docking ont été achevées avec succès: c'est le cas par exemple du facteur TNF- α (He et al., 2005), du domaine SH2 de la protéine p561ck (Huang et al., 2004) ou bien encore de l'interface oncogénique XIAP-caspase 9 (Nikolovska-Coleska et al., 2004).

Il est à noter que pour la recherche et la validation *in silico* d'interactants moléculaires, les programmes de docking sont très souvent complémentés par des simulations de dynamique moléculaire. Ces simulations sont effectuées sur chaque complexe protéine/ligand et apportent une vision plus dynamique de ces interactions, au sein d'un environnement solvaté. En particulier, le mouvement des chaînes latérales de la protéine ainsi que l'adaptation de cette dernière à la liaison du ligand

sont observés. L'échantillonnage d'une trajectoire stable pour un complexe protéine/ligand permet alors de prédire de manière plus exacte que pour le docking, une énergie théorique de liaison (Stjernschantz and Oostenbrink, 2010).

De même, dû aux faiblesses de la fonction de score du programme de docking et l'incapacité de pouvoir échantillonner toutes les conformations du ligand et du récepteur au cours de cette étape, un taux faible de « hits » est en général observé ainsi qu'un grand nombre de faux positifs et faux négatifs. Dans le but d'augmenter ce taux de « hits », une méthode appelée RCS (« Relaxed Complex Scheme ») a été proposée (Lin et al., 2003). Cette méthode utilise le docking sur un ensemble de conformations du récepteur, extraites d'une simulation de dynamique moléculaire. Par cet échantillonnage de conformations du récepteur, nous prenons en compte une flexibilité totale du récepteur au cours du docking et cela permet, en théorie, d'augmenter nos chances de trouver une molécule touche tout en tenant compte de phénomène de sélection de conformation par le ligand. Cette méthode a été utilisée avec succès dans plusieurs cas (Durrant and McCammon, 2010) (Schames et al., 2004) (Barakat et al., 2012). Finalement, il peut être très utile de considérer une étape préliminaire de filtrage d'une base de données par docking afin d'enrichir par exemple une librairie de molécules physiquement testées lors d'un test HTS par exemple.

Un filtrage en amont peut également s'avérer utile afin de constituer une base de données de molécules prédites comme « Drug-Like » et respectant ainsi un certain nombre de règles indispensables à l'optimisation et l'élaboration d'un futur candidat médicament. Un serveur web cité précédemment et appelé « FAF-Drugs3 » (Lagorce et al., 2015), permet par exemple d'effectuer ce type de calcul des propriétés physicochimiques et d'aider le biologiste dans la constitution et le filtrage d'une base de données enrichie en composés respectant les règles d'ADMET, en supprimant les molécules dites « PAINS » et les groupes fonctionnels non désirés, en proposant des indices de solubilité etc...

L'identification de protéines « druggables » par des approches computationnelles est parfois indispensable afin d'assurer le bon développement d'un inhibiteur de PPI. Ces interfaces étant en général relativement planes sur les protéines isolées, les

sites de liaison putatifs sont souvent exposés après changement conformationnel associé à la formation des multimères (Archakov et al., 2003). L'implémentation des calculs de docking et de dynamique moléculaire (afin de définir des zones, cavités ou sillons, potentiellement « druggables » ou bien des cavités cryptiques) couplée à des études de mutagénèse dirigée (pour définir les résidus hot-spot, impliqués dans l'interaction protéine-protéine), permet de pouvoir cibler virtuellement des inhibiteurs potentiels de PPI au sein de sites de liaison putatifs.

I – 5 : Criblage de molécules inhibitrices, dirigées contre les complexes XRCC4/Cer-XLF et XRCC4/C-ter Ligase 4 – Analyse du système d'étude, étude de la faisabilité du projet et de la démarche expérimentale

Les objectifs de cette thèse sont multiples. Il s'agit de pouvoir démontrer, d'une part, la « druggabilité » des protéines impliquées dans le complexe XXL de ligature des CDB (et par conséquent définir quelles seraient les interfaces à cibler, et quels seraient les sites de liaison putatifs) et d'autre part, de proposer une stratégie originale basée sur un criblage rationnel de molécules inhibitrices de ces interfaces protéines-protéines respectives. L'objectif à court terme serait d'isoler à l'issue de cette étude, des composés au moins d'affinité modérée (des premiers « hits ») et possédant une activité inhibitrice *in vitro*, afin d'apporter une base dans un programme ultérieur d'optimisation par chimie médicinale (stratégie « Hit-to-Lead ») guidée par la structure. Les objectifs à plus long terme seraient d'inhiber spécifiquement et au niveau cellulaire, le complexe tripartite de ligature des CDB de l'ADN avec ces molécules dérivées. Ces dernières auraient une plus forte affinité et spécificité et provoqueraient une radiosensibilisation cellulaire à l'instar des phénotypes observés lors de l'inhibition *in cellulo* précédemment décrite de l'interaction entre ces deux protéines.

Le système d'étude :

Les deux interfaces XRCC4/domaine C-terminal de la ligase 4 et XRCC4/Cer-XLF représentent un réel défi et s'inscrivent dans le désir actuel et grandissant de vouloir inhiber des interaction protéine-protéine. La preuve de concept sur la faisabilité

d'inhiber *in cellulo* ces deux interfaces et l'effet thérapeutique qui en découle (à savoir un effet cellulaire radiosensibilisant) a été démontré précédemment. L'enjeu de cette thèse est maintenant de pouvoir mettre au point des inhibiteurs de bas poids moléculaires, spécifiques de l'une de ces interfaces et ce par une approche combinée de criblage virtuel, biophysique et biochimique.

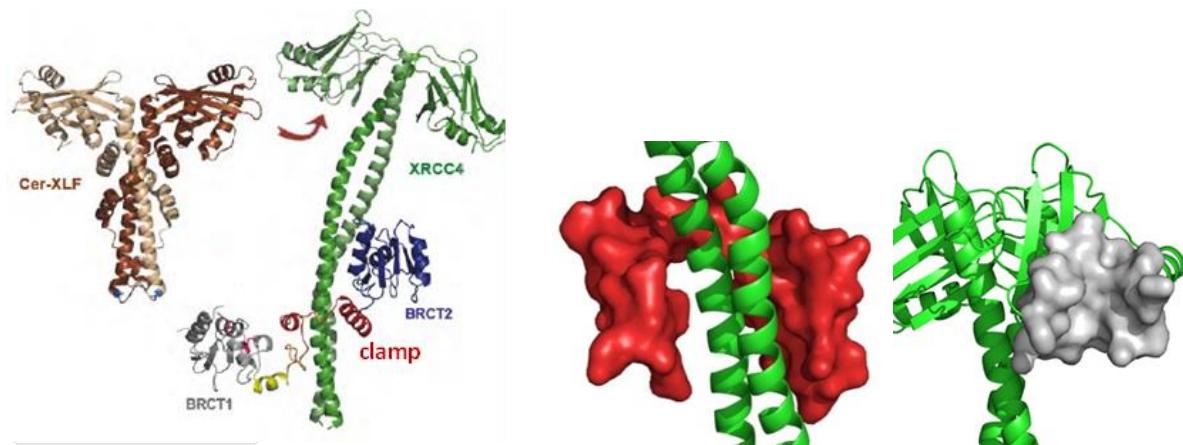


Figure 22 : Complexe XXL de réparation des CDB de l'ADN et illustration des interfaces protéine-protéine ciblées. A gauche : Structures cristallographiques des protéines Cer-XLF (Ref. PDB : 3W03) et XRCC4 en interaction avec le domaine C-terminal de la ligase 4 (BRCT1-clamp-BRCT2) (PDB code: 3I16). XRCC4 et Cer-XLF présentent une structuration très similaire et interagissent via leur tête globulaire. A droite : Représentation en surface des zones ciblées lors de mon criblage virtuel. En rouge, le domaine clamp de la ligase 4 ; En gris, l'interface d'interaction avec la protéine Cer-XLF sur la tête XRCC4.

En observant ce complexe tripartite de ligature, plusieurs interfaces peuvent être considérées pour la recherche d'interactants moléculaires. L'interaction entre les protéines XRCC4 et Cer-XLF s'effectue via leur tête globulaire respective mais seule la protéine XRCC4 présente au niveau de cette interface un sillon potentiellement bien adapté à la liaison d'une molécule et à proximité immédiate des résidus impliqués dans l'interaction avec Cer-XLF (Figure 22). L'interaction entre les protéines XRCC4 et domaine C-terminal de la ligase 4 fait apparaître deux interfaces potentiellement accessibles à de petites molécules. Ces deux interfaces sont un contact entre une hélice alpha du domaine BRCT2 et le domaine coiled-coil de XRCC4 et un domaine clamp (hélice-boucle-hélice) autour de ce domaine coiled-coil (Figure 22). Les interfaces coiled-coil de XRCC4 ainsi que l'hélice alpha du domaine BRCT2 sont planes et il paraît difficile d'envisager un criblage virtuel sur ces dernières. La surface du domaine clamp est également plane mais on pourrait

envisager que la structure en hélice-boucle-hélice accommode un ligand en son sein, à l'instar d'une cavité, avec ce dernier se stabilisant par contact à la fois des deux hélices et de la boucle. Par rapport à l'interface précédente, nous ne disposons cependant pas de la structure isolée de ce domaine C-terminal de la ligase 4. Par la suite, il faudra alors envisager des simulations sur ce domaine, séparé de XRCC4 afin d'en extraire des structures du clamp plus adaptées à l'étape de docking.

L'affinité extrêmement forte entre les deux partenaires XRCC4 et ligase 4 nous permet d'anticiper une stratégie où l'on cherchera plutôt à prévenir la formation de cette PPI en pré-formant un complexe ligase/inhibiteur stable, plutôt que de la déplacer. Néanmoins, l'affinité entre XRCC4 et Cer-XLF étant plus faible (de l'ordre de 7 μ M – estimée par des techniques d'ITC et biacore), ce complexe transitoire pourrait quant à lui être déplacé par un composé de bonne affinité se liant sur la tête globulaire de XRCC4. Un autre argument est qu'une mutation ponctuelle (au niveau de certains résidus critiques) suffit à elle seule pour dissocier complètement ce complexe. Ainsi, il est tout à fait envisageable de reproduire cet effet inhibiteur par liaison à l'interface et à proximité de ces régions hot-spot.

Etude de la faisabilité du projet et de la démarche expérimentale envisagée :

Approche computationnelle ou criblage cellulaire ? Les deux approches ici citées ont chacune leurs forces et leurs faiblesses et permettraient potentiellement de trouver des premiers inhibiteurs de bas poids moléculaire. L'avantage d'un criblage cellulaire est que l'on observe directement l'effet recherché (la radiosensibilisation des cellules par exemple). Ainsi, on s'affranchit du problème de la perméabilité membranaire et toxicité cellulaire tout en répondant à la question : le composé est-il actif *in cellulo* ? D'ailleurs, des études récentes ont permis d'identifier des nouveaux agents radiosensibilisants par recherche d'inhibiteurs de réparation des CDB par un criblage cellulaire (Goglia et al., 2015; Shahar et al., 2014). Cependant, cette approche nécessite la mise au point du test cellulaire lui-même qui peut s'avérer parfois difficile. De plus, le phénotype recherché permet de valider un ou des composés mais ne permet pas d'identifier avec certitude la cible de ces derniers. Cette identification est pourtant primordiale si l'on veut prouver la spécificité d'une molécule et améliorer son affinité pour la cible par chimie médicinale et par des approches

rationnelles. Des méthodes lourdes de click pull-down peuvent être toutefois employées pour répondre à la question : quelle est la cible ? (Larrieu et al., 2014).

Dans notre cas, nous disposons des structures cristallographiques de chacune des cibles. Ces structures ont une résolution suffisamment bonne pour envisager un criblage virtuel. N'ayant aucun inhibiteur déjà connu pour ces interfaces, une approche basée sur le ligand n'est pas possible (de même que des expériences de compétition). Une approche basée sur la structure est alors la meilleure option pour démarrer. La démarche expérimentale pour les deux interfaces ciblées sera la suivante :

- Analyse des structures disponibles et validations des régions potentiellement « druggables » pour orienter notre campagne de criblage virtuel (hot-spot connus ? présence d'une cavité ou d'un sillon à proximité des résidus hot-spot ?),
- Eventuellement une étape de « blind-docking » avec des fragments permettra de valider ces sites potentiellement druggables en fonction de la localisation, du nombre, du mode de liaison et du recouvrement des différents hits (Cette technique peut également être effectuée expérimentalement par trempage des cristaux de protéine avec des cocktails de fragments solubles),
- Echantillonnage conformationnel des protéines et préparation de bases de données virtuelles exhaustives et consciencieusement filtrées de molécules,
- Docking et validation *in silico* des molécules touchées par une analyse combinée visuelle et de dynamique moléculaire des complexes sélectionnés (contacts persistants, médiés par des molécules d'eau, effets d'ajustements induits, énergies théoriques de liaison, stabilité de la trajectoire), et éventuellement de la superposition au pharmacophore protéique,
- Les candidats virtuellement sélectionnés seront commandés (ou synthétisés), et intégrés à des tests de solubilité en tampon aqueux afin de s'affranchir d'éventuels faux-positifs causés par des phénomènes d'agrégation,
- L'interaction à la cible ainsi que la spécificité d'interaction seront validés par des méthodes biophysiques relativement sensibles et orthogonales (DSF + RMN),

- L'effet inhibiteur des composés spécifiques sera ensuite évalué dans des tests biochimiques adaptés. Les protéines que nous possédons n'ont pas d'activité catalytique propre. Des tests indirects et d'interactions seront alors envisagés afin de déterminer l'effet inhibiteur des molécules touchées. Également, un test HTS sera mis au point (ce test de fluorescence, basé sur du FRET, permettra de suivre l'interaction entre les protéines XRCC4 et Cer-XLF et sera détaillé plus loin) et permettra d'intégrer ces molécules à la banque physiquement ciblée,
- Les ou les inhibiteurs validés entreront dans un programme de détermination structurale des complexes afin d'optimiser chimiquement et de manière rationnelle, l'affinité et la spécificité de ces derniers pour la cible. Connaissant les conditions cristallographiques pour l'obtention de la structure de ces protéines, ces conditions seront ainsi exploitées dans ce but.

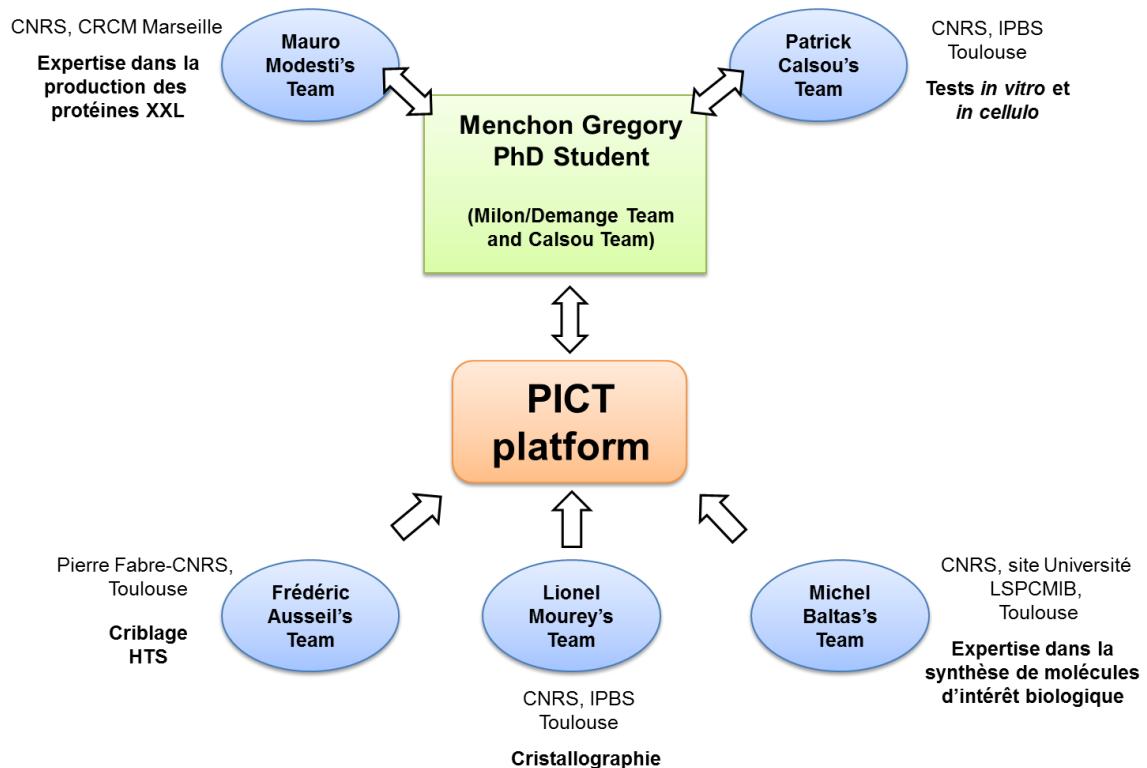
Les limitations majeures de ce projet pourraient être : les limites de concentration et la stabilité de nos protéines isolées, l'absence d'inhibiteurs de forte affinité connus pour ces cibles rendant ainsi les expériences de compétition impossible (pour la validation indirecte de la localisation de la molécule notamment), l'absence d'activité enzymatique et les surfaces planes caractéristiques des PPI (en particulier pour l'interface XRCC4/domaine C-terminal de la ligase 4).

Finalement, j'ai dressé ci-dessous un tableau comparatif des trois approches (HTS, criblage virtuel et approche par fragments) envisageables avec leurs avantages et inconvénients respectifs (cette liste de critères est non exhaustive).

	HTS	Criblage Virtuel	Fragments
Avantages	<ul style="list-style-type: none"> - Haut-débit, ultra haut-débit; - Information directe sur l'inhibition des PPI - Essai direct possible en cellules 	<ul style="list-style-type: none"> - Possibilité de filtrer des molécules plus larges, 3D et aromatiques (PPI) - Par rapport aux fragments, plus de molécules d'affinité μM pour les tests directs d'activité - Pré-criblage permet d'enrichir une librairie HTS - Evaluation de la druggabilité d'une cible (plus rapide que la cristallographie et le trempage de fragments) - Haut-débit et filtrage possible en amont 	<ul style="list-style-type: none"> - Détermination rapide et expérimentale de la druggabilité d'une cible protéique - Plus grand échantillonnage de l'espace chimique - Plus haut taux de hits avec une bonne valeur de LE - Détermination possible de nouveaux sites de liaison potentiels - Optimisation chimique tout en contrôlant les propriétés physico-chimique de la molécule finale - Possibilité de faire du linking, merging, growing
Inconvénients	<ul style="list-style-type: none"> - Fort taux de faux-positifs - Développement d'un test d'activité - Coût de production et développement 	<ul style="list-style-type: none"> - LE faible - Fort taux de faux positifs et négatifs - Un échantillonnage plus faible de l'espace chimique (par rapport aux fragments) 	<ul style="list-style-type: none"> - Affinités faibles, utilisation de fortes concentrations dans les essais biochimiques - Problèmes de solubilité - Besoin de méthodes de détection très sensibles - Besoin de connaissance en chimie médicinale - Règle « des 3 » peu adaptée aux PPI

Le consortium autour du projet « XXL-screen » :

Nous disposons au sein de l'IPBS mais également via plusieurs collaborations, de l'expertise (plateforme de criblage PICT), du matériel et de toutes les compétences requises pour mener à bien ce projet de thèse. Ci-dessous, un schéma représentatif de ce consortium.



Partie II :

Résultats et discussion

II – 1 : Production des protéines du complexe de ligature et tests de fonctionnalité

II – 1.1 : Expression en système bactérien et purification des protéines XRCC4, Cer-XLF et C-terminal de la Ligase 4 du complexe XXL de réparation de l'ADN

Dans cette partie, les principes de production et de purification de ces protéines sont détaillés. Cependant, tous les détails techniques ainsi que la constitution de chacun des tampons et les détails des constructions utilisées sont explicités en partie Matériels et méthodes. Chacune des protéines possède des spécifications de production et purification qui lui sont propres. Cela conditionne notamment sa stabilité et donc sa viabilité pour nos tests ultérieurs d'interaction et d'activité *in vitro*. La production des protéines XRCC4, Cer-XLF et C-terminal de la ligase 4 s'effectue en système hétérologue bactérien après transformation par un plasmide porteur de la séquence génétique correspondante. Chacune des protéines possède une étiquette clivable de 6 histidines, nécessaire à l'une des étapes de purification chromatographique ainsi qu'un gène de résistance permettant la sélection des clones correctement transformés. Les protéines pures seront par la suite intégrées à nos tests de validation biochimiques et biophysiques des inhibiteurs potentiels.

Un stock glycérol de bactéries transformées et utilisé pour l'ensemencement d'une pré-culture en milieu riche LB. La culture bactérienne s'effectue jusqu'à une phase exponentielle de croissance, phase à laquelle l'induction de la protéine sera déclenchée par ajout d'IPTG (dont la concentration a été au préalable optimisée, tout comme la température à laquelle s'effectue la sur-expression). L'IPTG active le promoteur de la T7 polymérase par déplacement du répresseur *lac*. L'induction à basse température (notamment pour le domaine C-terminal de la ligase 4 et le complexe XRCC4/C-terminal ligase 4) favorise l'expression douce des protéines, évitant ainsi la formation des corps d'inclusion. Durant l'induction de l'expression des protéines, le taux de croissance bactérienne est ralenti. En fin d'induction, la DO₆₀₀ est généralement proche de 2. Un échantillon bactérien est prélevé et comparé à un même échantillon non induit afin d'évaluer sur gel, l'efficacité de la sur-expression ainsi que l'état de solubilité de la protéine avant l'étape de purification.

L'étape de purification démarre par la lyse bactérienne dans un tampon de lyse contenant notamment du lysozyme, qui désorganise la paroi bactérienne, un détergent non-ionique pour perméabiliser la membrane et de la benzonase afin de réduire la viscosité du milieu par dégradation des acides nucléiques. Des cycles de sonication alternés par des périodes de refroidissement sur glace permettent de finir la lyse des bactéries en évitant la dénaturation de la protéine par l'échauffement. Le tampon de lyse contient également des inhibiteurs de protéases afin de protéger de toute dégradation possible notre protéine. La fraction soluble de la protéine est séparée par centrifugation à haute vitesse et le surnageant obtenu est encore filtré pour obtenir un lysat clair, prêt pour les étapes de purification chromatographiques.

Parmi les étapes de purification utilisées, la première étape sera une purification par chromatographie d'affinité au nickel (de type HisTrap HP) pour les trois protéines. A un pH>6.0, l'étiquette histidine (chargée) permet à la protéine de se fixer sur la résine conditionnée de la colonne. Après des étapes de lavage de la résine afin d'éliminer tout contaminants, la protéine est élueée via un gradient d'imidazole. L'imidazole entre en effet en compétition avec l'étiquette histidine pour l'affinité aux ions nickel et va permettre par conséquent de décrocher la protéine qui s'en retrouvera pratiquement pure. Les fractions collectées et correspondant au pic d'élution (pic d'absorption à 280nm) sont ensuite dosées séparément et déposées sur un gel d'électrophorèse (gel de polyacrylamide SDS-PAGE à 10%) afin d'évaluer la présence, la pureté et l'efficacité de la chromatographie d'affinité. A ce stade, une seule étape de purification par cette méthode n'est évidemment pas suffisante. En effet, des contaminants protéiques subsistent en dépit d'un enrichissement certain en notre protéine d'intérêt. Cela s'explique notamment par une affinité moindre d'autres protéines du lysat bactérien pour les ions nickel. Une seconde étape de purification (par gel filtration ou une colonne échangeuse d'ion) sera alors nécessaire afin d'obtenir une protéine complètement pure.

1. Purification du domaine C-terminal de la ligase 4

Ci-dessous (Figures 23 et 24) sont présentées les gels SDS-PAGE effectués suite aux étapes successives de purification de ce domaine par chromatographie d'affinité

au nickel (HisTrap HP) et par chromatographie échangeuse d'anion (HiTrap Q HP). Les fractions 1 à 5 présentées en Figure 23 ont été rassemblées et chargées sur la seconde colonne conditionnée.

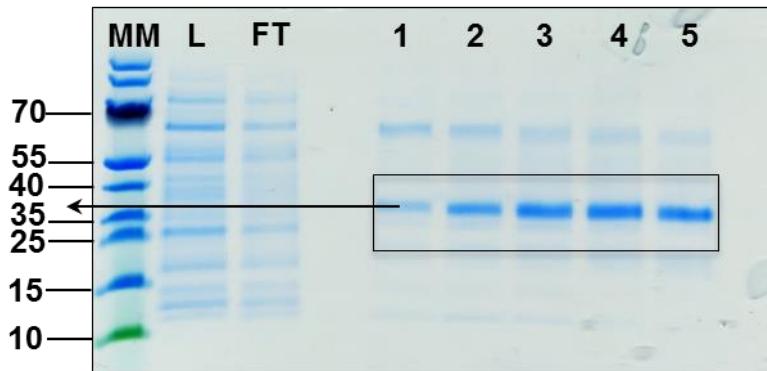


Figure 23: Gel SDS-PAGE 10% - Chromatographie d'affinité au nickel pour la purification du C-terminal de la ligase 4 étiqueté histidine en N-terminal.

La piste annotée « MM » correspond aux marqueurs de poids moléculaire. La piste « L » est le lysat bactérien avant purification. Le « FT » représente les protéines qui ne se sont pas accrochées à la résine lors du chargement de la colonne. Les pistes 1 à 5 sont les fractions éluées lors du gradient d'imidazole. La bande majoritaire est le domaine C-terminal de la ligase 4 qui a une taille d'environ 29kDa et qui migre autour de 35kDa. Coloration à l'« Instant Blue ».

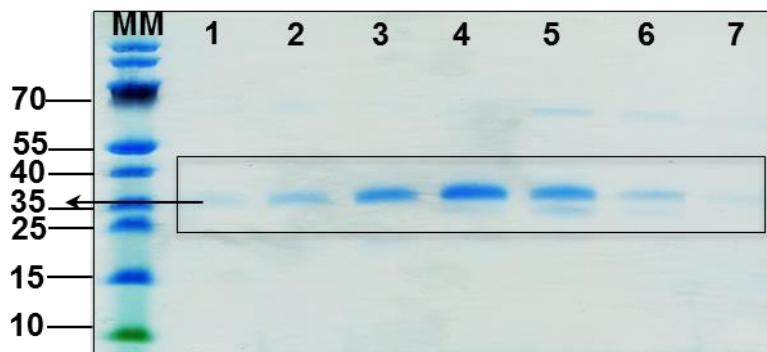


Figure 24: Gel SDS-PAGE 10% - Chromatographie d'échange d'ions pour la purification finale du C-terminal de la ligase 4.

La piste annotée « MM » correspond aux marqueurs de poids moléculaire. Les pistes 1 à 7 sont les fractions éluées lors du gradient salin. La bande majoritaire est le domaine C-terminal de la ligase 4 pur. Coloration à l'« Instant Blue ».

Une fois purifiée, la protéine est concentrée (habituellement autour de 10 μ M), aliquotée, plongée dans l'azote liquide et stockée à -80°C (en présence de glycérol à

10%) en attendant les tests fonctionnels. Le rendement final est environ de 1-2mg de protéine recombinante par litre de culture.

2. Purification de la protéine Cer-XLF

Ci-dessous (Figures 25 et 26) sont présentées les gels SDS-PAGE effectués suite aux étapes successives de purification de cette protéine par chromatographie d'affinité au nickel (HisTrap HP) et par chromatographie d'exclusion. Les fractions 1 à 3 présentées en Figure 25 ont été rassemblées et chargées sur la seconde colonne conditionnée.

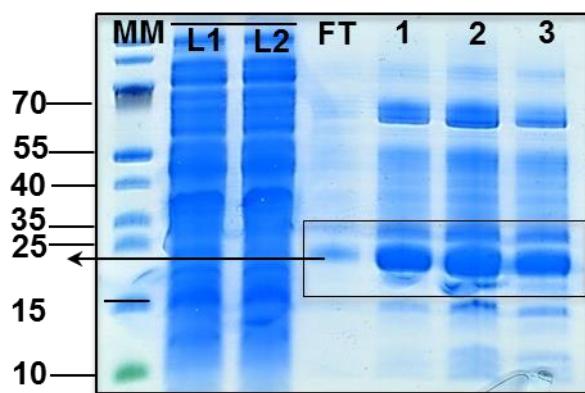


Figure 25: Gel SDS-PAGE 10% - Chromatographie d'affinité au nickel pour la purification de la protéine Cer-XLF étiquetée histidine en C-terminal.

La piste annotée « MM » correspond aux marqueurs de poids moléculaire. Les pistes « L1 » et « L2 » sont respectivement les lysats bactériens des deux cultures de 2L avant purification. Le « FT » représente les protéines qui ne se sont pas accrochées à la résine lors du chargement de la colonne. Les pistes 1 à 3 sont les fractions éluées lors du gradient d'imidazole. La bande majoritaire est la protéine Cer-XLF qui a une taille d'environ 25kDa et qui migre autour de 25kDa. Coloration à l' « Instant Blue ».

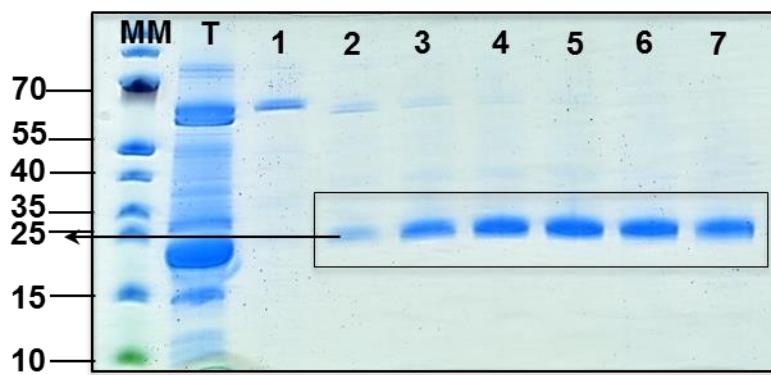


Figure 26: Gel SDS-PAGE 10% - Chromatographie d'exclusion pour la purification finale de la protéine Cer-XLF.

La piste annotée « MM » correspond aux marqueurs de poids moléculaire. La piste « T » est un aliquot du pool total injecté dans la colonne. Les pistes 2 à 7 sont les fractions correspondant au pic d'élution de Cer-XLF pur. Coloration à l'« Instant Blue ».

Une fois purifiée, la protéine est concentrée (habituellement entre 100 et 200 μ M de monomères), aliquotée, plongée dans l'azote liquide et stockée à -80°C (en présence de glycérol à 10%) en attendant les tests fonctionnels. Le rendement final est environ de 2-3mg de protéine recombinante par litre de culture.

3. Purification de la protéine XRCC4

La production et la purification de cette protéine (Figure 27) ont été effectué par le collaborateur Mauro Modesti (CRCM – CNRS, Marseille). Du fait de la similarité entre les protéines XRCC4 et Cer-XLF, les protocoles de production et de purification sont identiques.

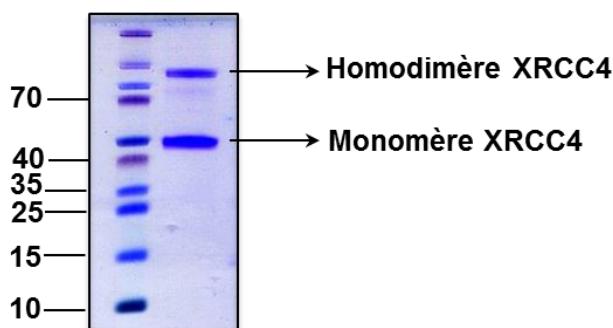


Figure 27: Gel SDS-PAGE 10% - Protéine XRCC4 humaine purifiée. Coloration au bleu de Coomassie.

4. Purification du complexe XRCC4/C-terminal de la ligase 4

Ci-dessous (Figures 28 et 29) sont présentées les gels SDS-PAGE effectués suite aux étapes successives de purification de ce complexe par chromatographie d'affinité au nickel (HisTrap HP) et par chromatographie échangeuse d'anion (HiTrap Q HP). Les fractions 3 à 8 présentées en Figure 28 ont été rassemblées et chargées sur la seconde colonne conditionnée.

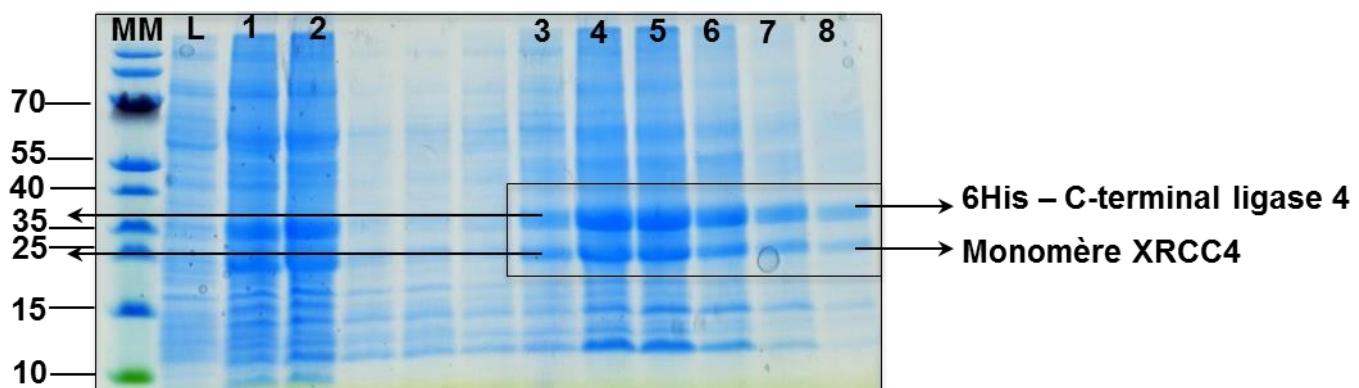


Figure 28: Gel SDS-PAGE 10% - Chromatographie d'affinité au nickel pour la purification du complexe XRCC4/C-terminal de la ligase 4.

La piste annotée « MM » correspond aux marqueurs de poids moléculaire. La piste « L » est le lysat bactérien avant purification. Les pistes 1 à 8 sont les fractions éluées lors du gradient d'imidazole. Les pistes 1 à 2 et 3 à 8 correspondent à deux pics distincts d'absorption à 280nm sur le chromatogramme. Les deux bandes majoritaires sont les protéines XRCC4 et le C-terminal, qui migrent aux tailles attendues. Coloration à l'« Instant Blue ».

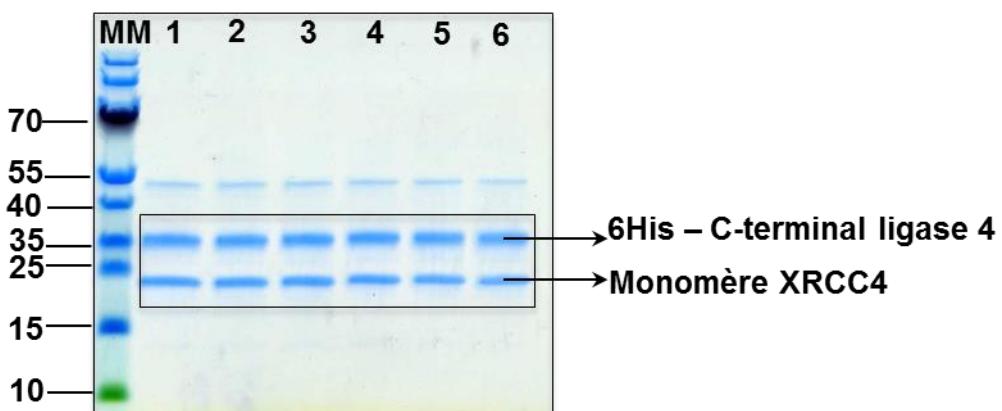


Figure 29: Gel SDS-PAGE 10% - Chromatographie d'échange d'ions pour la purification finale du complexe XRCC4/C-terminal de la ligase 4.

La piste annotée « MM » correspond aux marqueurs de poids moléculaire. Les pistes 1 à 6 sont les fractions éluées lors du gradient salin. Les bandes majoritaires sont les protéines XRCC4 et le domaine C-terminal de la ligase 4 pures. La bande entre 40 et 55kDa correspond à un dimère de protéine XRCC4 ou du C-ter causé par un pont cystine car elle n'est plus présente après un traitement supplémentaire au DTT. Coloration à l' « Instant Blue ».

La bande observée autour de 55kDa est très certainement un dimère de XRCC4 induit par un pont cystine car un traitement par 100mM DTT et analyse ultérieure sur gel SDS-PAGE montre une disparition de cette dernière. Le complexe ainsi purifié est par conséquent très pur.

Une fois purifié, le complexe est immédiatement dialysé et concentré à hauteur de 6 à 8mg/ml dans un tampon de cristallisation (voir partie Matériels et Méthodes).

II – 1.2 : Analyse de la thermostabilité et des interactions des protéines purifiées par DSF et gel natif

1. Profil DSF des protéines XRCC4/Cer-XLF et C-terminal de la ligase 4.

La fluorescence à balayage différentiel (DSF) va nous permettre de valider le repliement, évaluer la thermostabilité ainsi que mesurer potentiellement la fonctionnalité de nos protéines purifiées du complexe de ligature (voir partie Matériels et Méthodes). Les protéines purifiées ci-dessus sont testées à une concentration de 5 μ M, dans un tampon PBS 1X à pH7.4 et dans un volume final de 20 μ L. La molécule fluorescente Sypro orange a été utilisée à une concentration finale de 10X. Ces valeurs ont été optimisées afin d'obtenir un bon signal de fluorescence. Ci-dessous est représenté le thermogramme pour les trois constructions : Le C-terminal de la ligase 4 (654-911) étiqueté histidine, XRCC4 (1-336) étiqueté histidine et Cer-XLF (1-224) étiqueté histidine (figure 30). Le fait d'enlever l'étiquette de 6 histidines pour les trois protéines n'affecte pas le profil DSF observé.

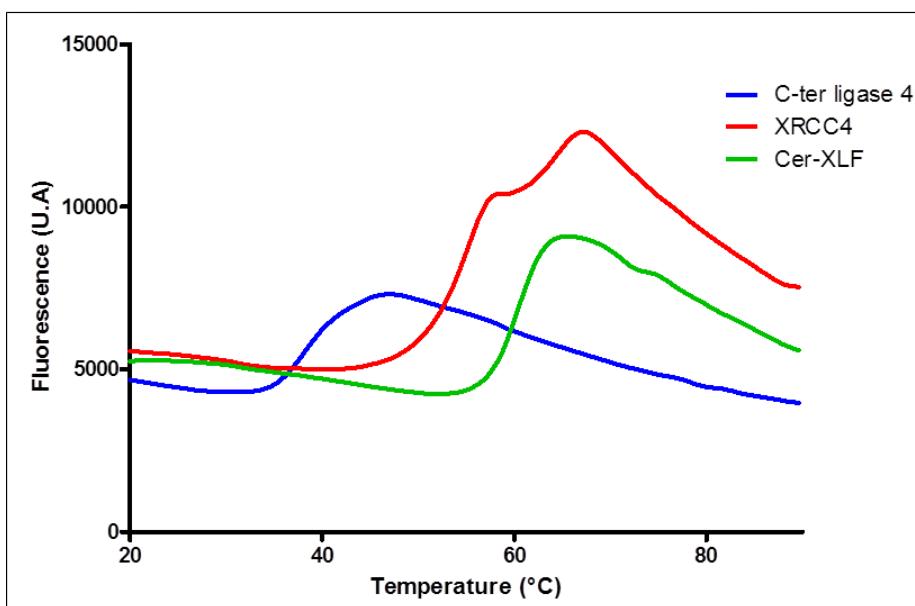


Figure 30: Profil DSF des protéines XRCC4, Cer-XLF et du domaine C-terminal de la ligase 4 à 5µM et dans un tampon PBS 1X pH 7.4

Ce profil DSF nous permet de tirer plusieurs informations concernant nos protéines recombinantes. A savoir, les courbes des trois protéines ont un profil sigmoïdal similaire et un point de fusion (Tm) caractéristique témoignant de la présence d'un repliement initial (même si cela ne nous donne aucune indication concernant un bon ou mauvais repliement). Le Tm est la température à partir de laquelle 50% de notre protéine se retrouve dénaturée.

La structuration de XRCC4, protéine homodimérique, est très similaire à Cer-XLF ; Cependant, par rapport à cette dernière le profil DSF de XRCC4 est biphasique. Une des hypothèses serait que le premier palier correspondrait à une séparation des deux monomères et que le palier suivant serait la dénaturation finale des deux monomères séparés. Or, nous savons qu'il existe *in vitro*, un équilibre entre une forme dimérique et tétramérique de XRCC4 avec un Kd autour de 22µM (Modesti et al., 2003). Une autre théorie serait alors qu'on observe un Tm différent pour ces deux formes. Les Tm de XRCC4 et Cer-XLF dimériques sont respectivement de 57 et 60°C. Ceci témoigne de la bonne stabilité de ces protéines dans ces conditions. Dans les mêmes conditions, le domaine C-terminal de la ligase 4 est un peu plus instable. Son Tm est autour de 38°C. Cela confirme la littérature et atteste de

l'instabilité intrinsèque de ce domaine lorsque celui-ci n'est pas lié à son partenaire XRCC4 (Bryans et al., 1999).

Finalement, un test de fonctionnalité par cette technique de DSF a été effectué afin de pouvoir mesurer la formation des complexes et les changements de thermostabilité en conséquence. Ci-dessous, un thermogramme DSF montre la liaison du domaine C-terminal de la ligase 4 à XRCC4 et l'augmentation consécutive de leur thermostabilité respective. Un contrôle de dissociation du complexe par modification covalente des cystéines (par le cacodylate de sodium), restaure les profils des protéines individuelles (Figures 31). Dans ce thermogramme, la dérivée première du signal de fluorescence est représentée cette fois-ci pour plus de clarté en fonction de la température. Les points de fusions respectifs de chaque protéine sont représentés par le sommet de ces pics.

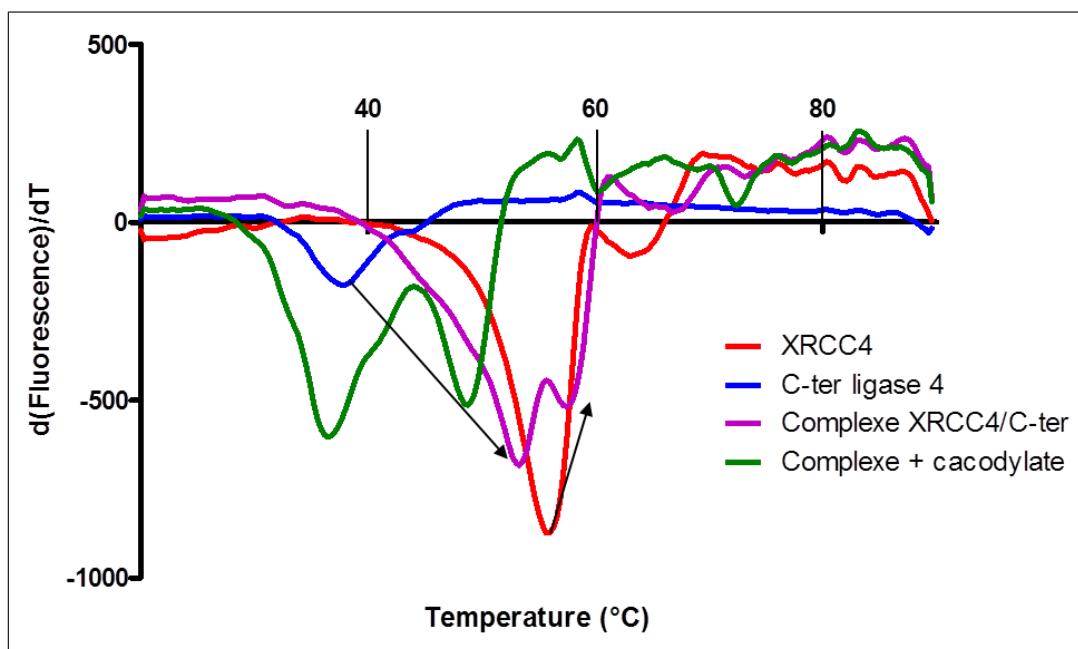


Figure 31: analyse par DSF de la formation du complexe XRCC4/C-terminal de la ligase 4 à 5 μ M et dans un tampon PBS 1X pH 7.4.

Lorsque l'on incube les protéines recombinantes et purifiées XRCC4 et le C-terminal de la ligase 4, on forme le complexe ce qui se traduit par deux éléments :

- Un shift positif des températures de fusion des deux protéines (dont un shift de presque 10°C pour le C-terminal) traduisant ainsi une augmentation de la

stabilité de ces deux protéines après complexation (flèches noires sur la figure 31).

- Le second pic de XRCC4 (pic présumé de la forme tétramérique) diminue suite à la liaison de la ligase qui entre en compétition avec ce site de tétramérisation.

Evidemment, ces théories restent à prouver mais apportent un bon modèle pour tenter d'expliquer les thermogrammes obtenus. Le traitement au cacodylate de sodium couplé au DTT va permettre de modifier de façon covalente les résidus cystéine et de rompre le complexe. Cela se traduit par le thermogramme vert ci-dessus où l'on restaure les pics individuels du domaine C-terminal et de XRCC4. Néanmoins, on note que ce traitement affecte le profil DSF de XRCC4 qui, après traitement, retrouve une température de fusion plus faible. Le thermogramme DSF du complexe XRCC4/Cer-XLF est moins convaincant (Figure 32).

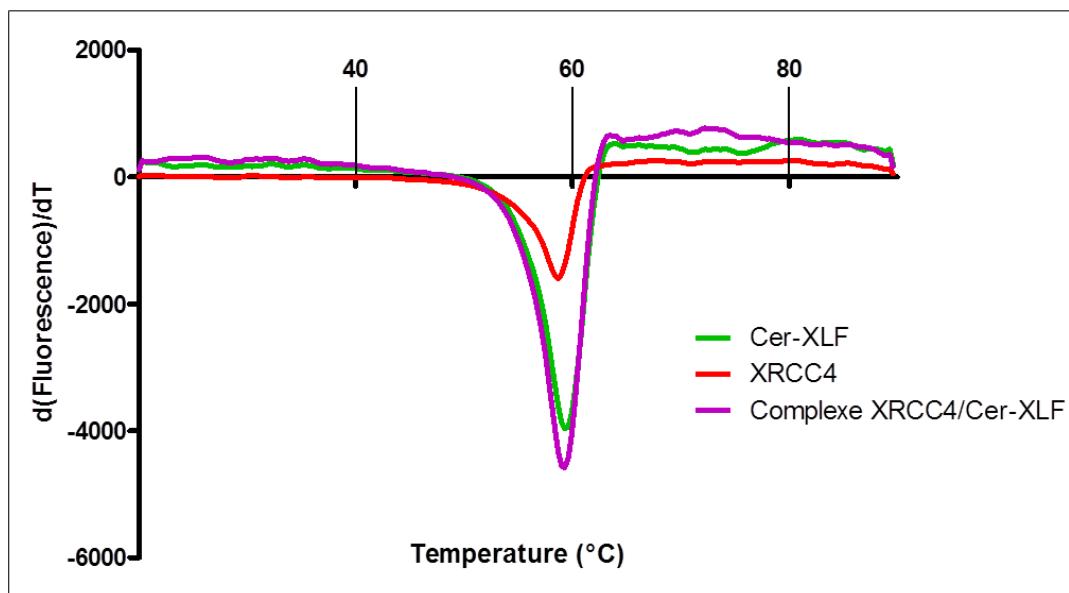


Figure 32: analyse par DSF de la formation du complexe XRCC4/Cer-XLF à 5 μ M et dans un tampon PBS 1X pH 7.4

Dans ce cas précis, aucune augmentation de T_m n'est à noter. Il est vrai que le K_d entre XRCC4 et Cer-XLF est plus élevé de plusieurs ordres de grandeurs que celui entre XRCC4 et le C-terminal de la ligase 4 (qui lui est nanomolaire). Cette plus faible affinité entre les deux partenaires est peut-être la raison pour laquelle nous

n'observons pas de profil caractéristique d'un complexe. Dans ce cas, un autre test d'interaction est nécessaire comme par exemple un gel d'électrophorèse en conditions natives.

2. Analyse des interactions protéine-protéine par électrophorèse en gel natif.

Ci-dessous sont représentés les tests d'interaction en gel natif 6% (gels retards) pour les complexes XRCC4/C-terminal Ligase 4 et XRCC4/Cer-XLF (Figure 33). Les protéines sont évaluées pour leur capacité de liaison dans des quantités stœchiométriques. Les conditions expérimentales de préparation des gels, de l'interaction et d'électrophorèse sont détaillées en partie Matériels et Méthodes. Pour ces tests, les protéines sont incubées dans leur tampon de conditionnement respectif.

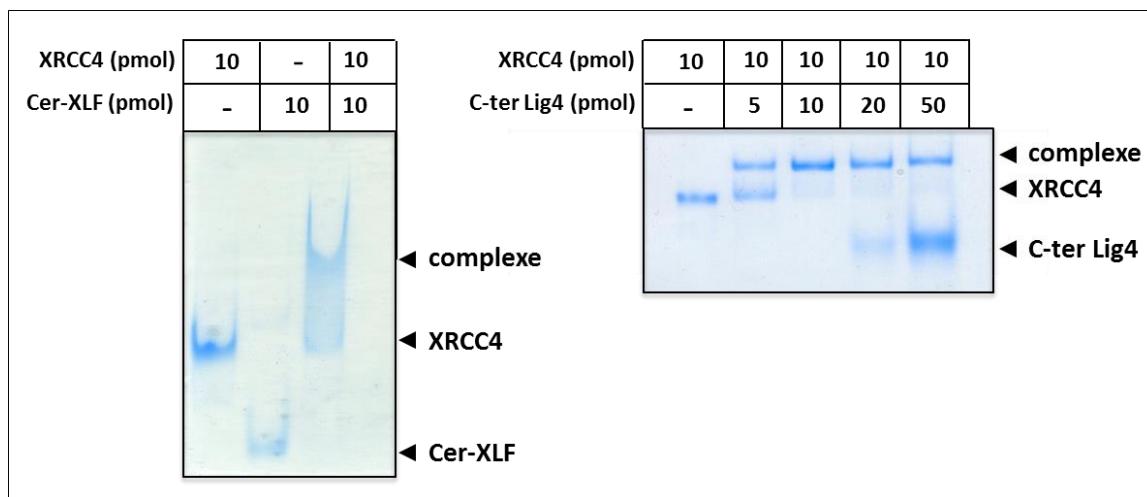


Figure 33: Gels natifs 6% - Analyse des interactions en conditions natives entre les protéines XRCC4/Cer-XLF et XRCC4/C-terminal de la ligase 4.

Sur ces gels, les profils de migration des protéines seules et en complexes sont montrés. En quantité équimolaire, on observe un retard de migration pour chacune d'entre elles témoignant de la formation des complexes. Coloration à l'« Instant Blue ».

Sur ces gels retard, nous prouvons définitivement la fonctionnalité de nos protéines recombinantes par un simple test d'interaction. Dans les deux cas, l'incubation à température ambiante d'une quantité équimolaire de chaque partenaire permet de former un complexe total. Le « smear » observable dans le cas du complexe XRCC4/Cer-XLF est peut-être dû à l'extension importante et déstructurée du C-terminal pour XRCC4. Une manière de le prouver serait d'effectuer ce même test en incubant Cer-XLF avec la construction XRCC4 plus courte (1-203). Ce test

d'interaction en gel natif est l'unique test direct de fonctionnalité de ces protéines du fait de l'absence d'une activité catalytique. Néanmoins, il peut être intégré à des tests de validation biochimique d'inhibiteur potentiels d'interaction protéine-protéine, axé sur ces interfaces. En effet, dans l'optique d'une prévention de la formation d'un de ces complexes, la pré-incubation d'une molécule inhibitrice avec l'une des protéines ciblées empêcherait la liaison du second partenaire protéique si celle-ci se liait à l'interface ou si elle modifiait de façon allostérique cette zone d'interaction.

II – 1.3 : Analyse de la fonctionnalité du complexe XRCC4/Cer-XLF par des tests de stimulation de la ligase T4

Les complexes protéiques XRCC4/Cer-XLF et XRCC4/C-ter ligase 4 utilisés n'ont pas d'activité catalytique propre. Le premier possède une fonction de « scaffold » protéique dans le mécanisme NHEJ de réparation des cassures double brin de l'ADN (permettant notamment l'amarrage d'autres facteurs de réparation) et Le domaine C-terminal de la ligase ne contient pas son domaine catalytique. Le test d'interaction précédent en conditions natives permet d'apporter un premier élément de confirmation quant à l'activité putative d'un inhibiteur issu d'un criblage sur la prévention de l'association entre les deux protéines.

Dans le cas du complexe XRCC4/Cer-XLF, un test indirect d'activité catalytique a pu cependant être développé (Roy et al., 2015). Ce test, dont le principe est schématisé sur la figure ci-dessous (Figure 34), ne mesure pas une activité directe de ce complexe mais son effet stimulateur sur la ligase du bactériophage T4. En effet, il a été observé qu'une incubation en présence du complexe XRCC4/Cer-XLF stimulait fortement l'activité de ligation intermoléculaire de la ligase T4 sur des plasmides d'ADN linéarisés. Ainsi, on observe que la stimulation induite de la ligase T4 favorise fortement la formation de « concatémères » d'ADN (multimères), des larges fragments d'ADN linéaires re-ligués.

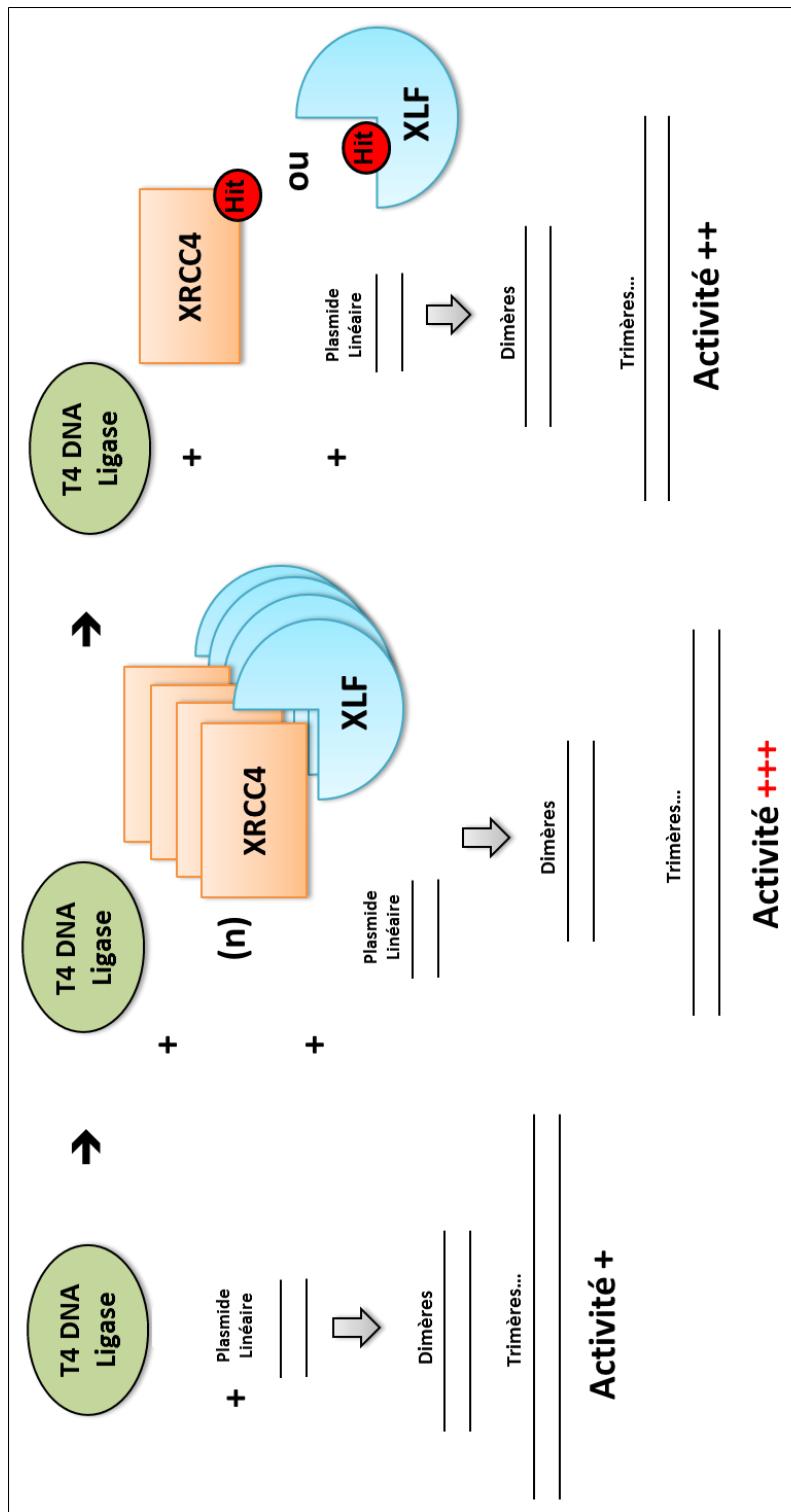


Figure 34: Schéma du principe de la stimulation de la ligase T4 par le complexe XRCC4/Cer-XLF.

L'activité basale de la ligase T4 favorise la formation d'espèces dimériques et trimériques d'ADN. Une incubation du complexe XRCC4/Cer-XLF stimule fortement cette activité de liaison intermoléculaire et davantage d'espèces multimériques sont présentes. Cette stimulation est fortement diminuée par un inhibiteur de l'interaction entre XRCC4 et Cer-XLF.

A l'issu de ce test d'activité, on mesure donc sur un gel d'agarose, un pourcentage de ligation intermoléculaire par rapport à une quantité totale d'ADN initiale. La stimulation de la ligase T4 par le complexe favorise fortement la formation de produits de ligation à haut poids moléculaires. La figure ci-dessous (Figure 35) présente l'effet observé lors d'une stimulation de la ligase T4 et dans des conditions optimisées (plus de détails dans partie Matériels et Méthodes).

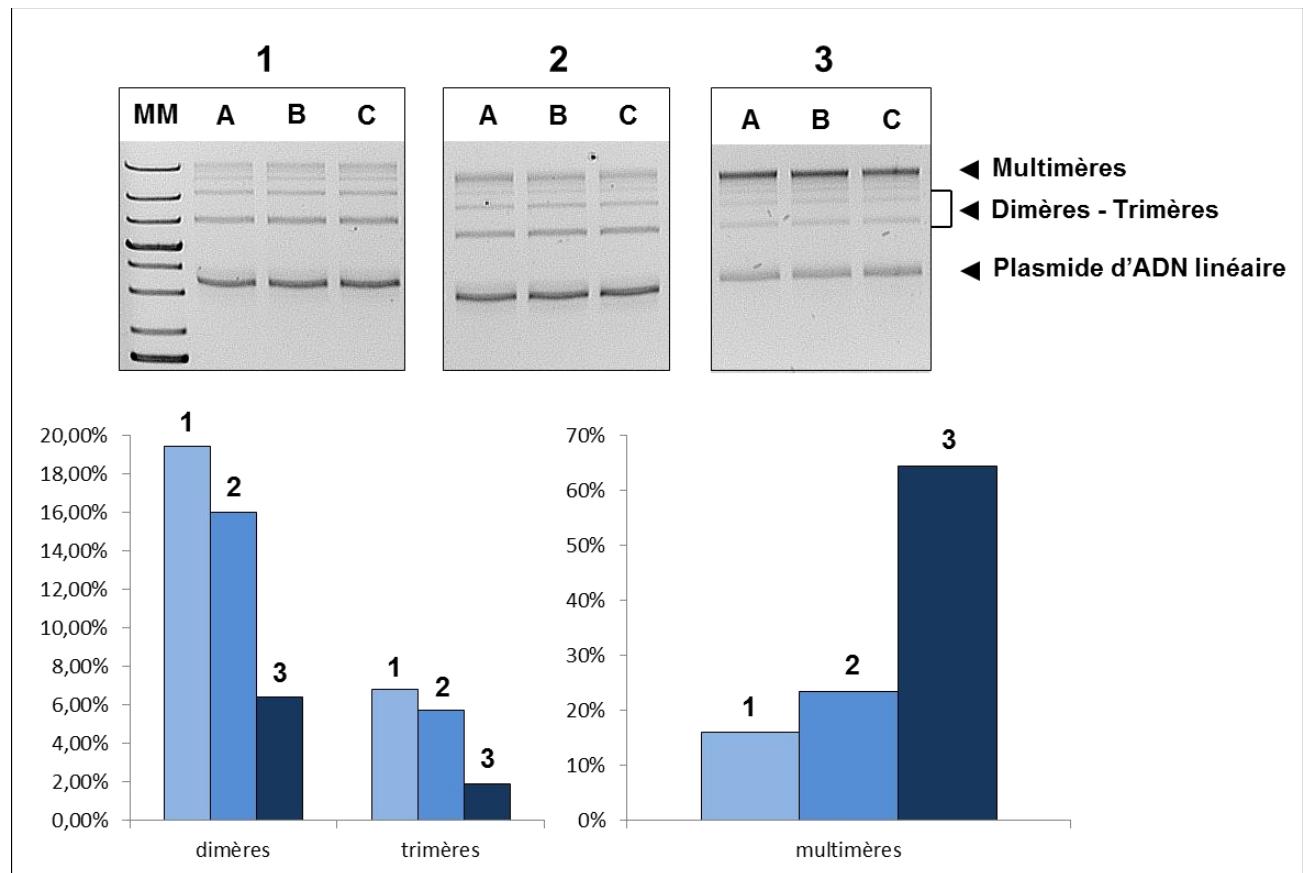


Figure 35: Optimisation des conditions expérimentales pour la stimulation de la ligase T4 par le complexe XRCC4/Cer-XLF.

L'activité basale de la ligase T4 favorise la formation de dimères, trimères et multimères d'ADN. A, B et C représentent respectivement l'ajout de 0.5, 0.75 et 1 μ L de ligase T4 NEB (d'une dilution 1/10 soit 40U). 1, 2 et 3 représentent respectivement l'ajout de 0, 0.5 et 1 μ M de complexe XRCC4/Cer-XLF. Une incubation du complexe XRCC4/Cer-XLF stimule fortement cette activité de liaison intermoléculaire et davantage d'espèces multimériques sont présentes. Une augmentation de la ligase T4 ne modifie pas l'activité mais une quantité de 1 μ M de complexe protéique est optimale pour l'effet de stimulation.

Ce mécanisme de stimulation est encore mal connu. Il est supposé que des filaments en formation de XRCC4/Cer-XLF sur l'ADN, faciliteraient l'alignement et donc la religation de ce dernier. Ainsi, un inhibiteur de l'interaction XRCC4/Cer-XLF

empêcherait la formation du complexe et donc le mécanisme de stimulation. Un retour à l'état basal de l'activité de la ligase T4 avec une perte très significative d'espèce multimériques d'ADN serait une preuve indirecte de l'effet inhibiteur de nos molécules. Ce test sera donc utilisé ultérieurement dans ce but.

II – 2 : Inhibition de l'interaction XRCC4/Cer-XLF

Résumé :

Par une approche rationnelle basée sur l'utilisation combinée d'un criblage virtuel (docking et simulations de dynamique moléculaire) et de techniques biophysiques et biochimiques, nous avons caractérisé à l'issu de cette étude le premier inhibiteur de l'interaction entre XRCC4 et Cer-XLF. L'affinité modérée entre les deux partenaires ($K_d \approx 7\mu M$) et la connaissance des résidus hot-spot intervenant dans l'interaction laissent entrevoir la possibilité d'une inhibition de cette interaction protéine-protéine par une petite molécule. L'inhibiteur caractérisé interagit de manière spécifique avec la protéine XRCC4 (avec un $K_d \approx 200\mu M$, déterminé par des expériences de STD-NMR), et inhibe la formation du complexe XRCC4/Cer-XLF dans des expériences *in vitro* de stimulation de la ligase T4 (test breveté par le collaborateur Mauro Modesti : Numéro de brevet : WO 2013057457 A1 (Roy et al., 2015)). Au sein d'un environnement plus complexe tel qu'un extrait cellulaire, nous avons également montré que cet inhibiteur inhibe de manière spécifique et dose-dépendante le mécanisme NHEJ de réparation de l'ADN, avec une activité sub-millimolaire. Notre étude a donc permis d'apporter un premier chemotype efficace et spécifique et donc une base solide pour l'optimisation rationnelle par chimie médicinale d'un inhibiteur de plus haute affinité. Dans cette optique d'optimisation, des expériences de cristallographie ont été initiées et ont permis l'obtention de cristaux de la protéine XRCC4, ouvrant ainsi la possibilité d'effectuer des trempages de cristaux avec notre inhibiteur.

II – 2.1 : Simulation de dynamique moléculaire sur le complexe XRCC4/Cer-XLF

Des travaux récents (cristallisation, microscopie électronique) sur la structure du complexe XRCC4/Cer-XLF nous permettent d'entrevoir l'organisation spatiale ainsi que les résidus impliqués dans cette interaction protéine-protéine (Hammel et al., 2011). Les protéines XRCC4 et Cer-XLF sont de nature homodimérique et interagissent via leur domaine globulaire en partie N-terminale. Ainsi, il a été proposé que ce complexe s'organise en structures filamenteuses avec une disposition en alternance entre XRCC4 et Cer-XLF. D'autres études (Ropars et al., 2011) viennent confirmer cette organisation spatiale. Notamment, la structure cristalline de ce filament a été obtenue à une résolution de 5.5 Å et montre une inclinaison de 30°C entre les tiges en hélice des deux protéines (figure 36).

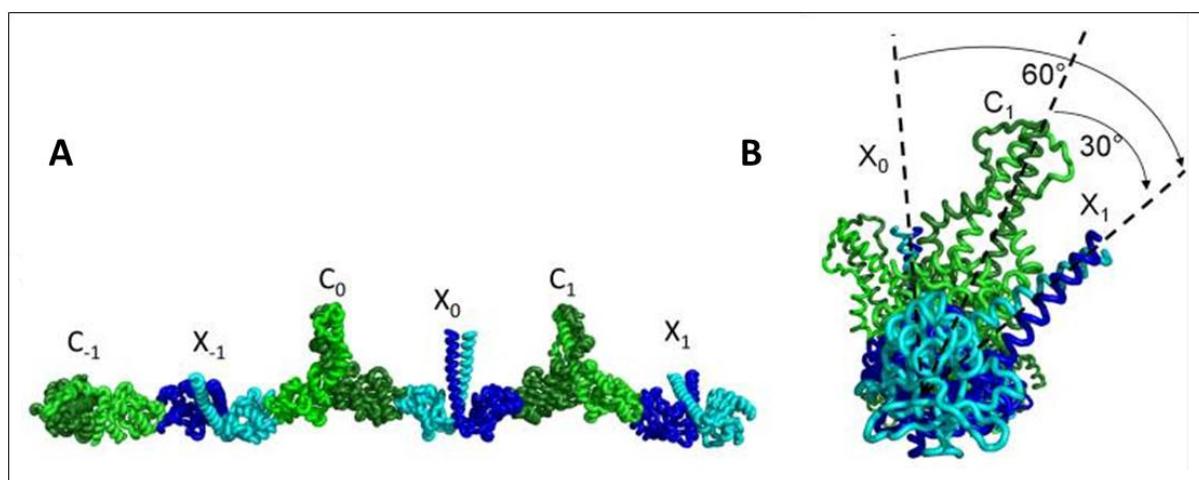


Figure 36: Structure tridimensionnelle cristalline à 5.5 Å, de l'organisation en filaments de XRCC4/Cer-XLF (d'après Ropars et al., 2011).

(A) Un filament XRCC4/Cer-XLF constitué par trois homodimères de XRCC4 (en bleu : X-1, X0 et X1) et trois homodimères de Cer-XLF (en vert : C-1, C0 et C1). Chaque homodimère interagit via son domaine N-terminal. (B) La structure en A est agrandie et représentée avec une rotation de 90° le long de l'axe vertical. On observe une inclinaison d'un angle de 60° entre chaque XRCC4 ou Cer-XLF consécutif et un angle de 30° entre les tiges en hélice de deux homodimères adjacents.

Le rôle de cette organisation filamenteuse n'est pas encore clair aujourd'hui mais elle pourrait faciliter l'alignement et la réparation des deux brins de l'ADN endommagé. Grâce à cet apport structural ainsi qu'à des études de mutagénèse dirigée et de calorimétrie isotherme de titration (ITC) sur ces différents mutants, nous savons que

l'affinité (K_d) entre les deux partenaires est proche de $7\mu M$ et que les résidus critiques dans l'interaction appartiennent à un patch majoritairement hydrophobe. Ce patch est formé par les résidus E55, D58, M59, M61, L101, V104 et F106 de XRCC4 et R64, L65, A67, F72, L112, L115, P116 et Y118 de Cer-XLF (Figure 37). Au sein de ce complexe transitoire, une mutation ponctuelle sur un seul des résidus E55, D58, M61 ou F106 de XRCC4 et R64 ou L115 de Cer-XLF est suffisante pour perdre complètement cette interaction *in vitro* (Ropars et al., 2011). Un autre résidu de XRCC4 (K99) a été caractérisé dont la mutation ponctuelle impacte négativement sur l'interaction XRCC4/Cer-XLF (Roy et al., 2012).

Ainsi, en sachant l'affinité « modérée » entre ces deux protéines et l'effet d'une seule mutation sur la rupture du complexe, une approche d'inhibition rationnelle par une petite molécule organique nous était tout à fait envisageable. Une de ces structures du complexe issue de la PDB (3SR2) a été soumise à une simulation de dynamique moléculaire (DM) de dix nanosecondes. Le principe et l'utilité d'introduire de la DM au sein d'un criblage virtuel a été discuté en introduction. A l'issue de cette dynamique, une analyse des contacts persistants (hydrophobes, ioniques, ponts hydrogènes...) entre les deux protéines avec un seuil de 90% sur toute la trajectoire confirme les expériences de mutagénèse et souligne en particulier les résidus D58 et F106 sur XRCC4 et R64 sur Cer-XLF comme étant des résidus très persistants au cours de la trajectoire et donc critiques dans cette interaction. Ces derniers sont donc ce que l'on pourrait nommer des résidus « hot-spot », concentrant, avec d'autres résidus, la majeure partie de l'énergie de liaison entre les deux protéines. (Figure 37).

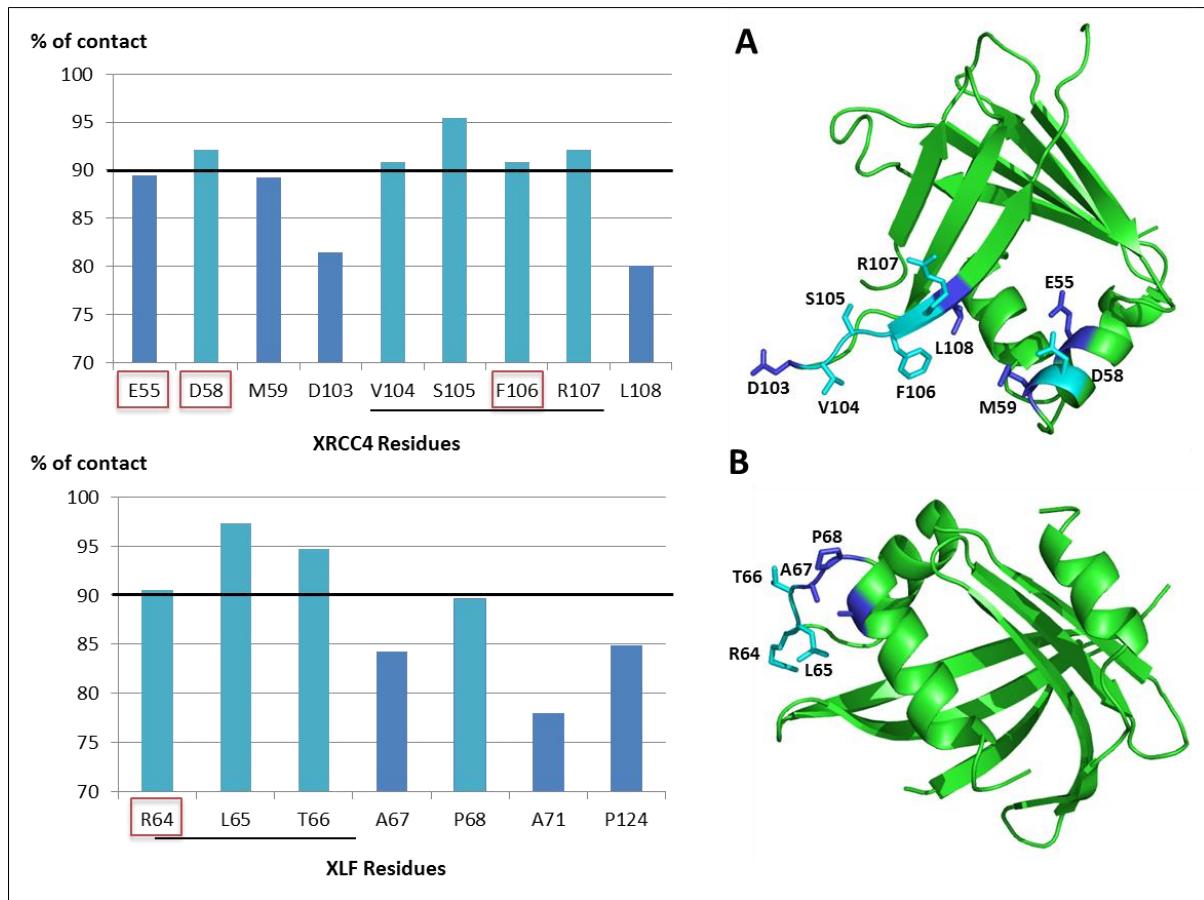


Figure 37: Analyse des contacts persistants entre les protéines XRCC4 et Cer-XLF au cours d'une trajectoire de DM de dix nanosecondes.

(A) Partie N-terminale de XRCC4. Les résidus très persistants (qui atteignent un seuil de 90%), sont représentés en cyan sur l'histogramme et la structure. (B) Partie N-terminale de Cer-XLF. Les résidus très persistants (qui atteignent un seuil de 90%), sont représentés en cyan sur l'histogramme et la structure. Les résidus encadrés en rouges correspondent aux résidus dont la mutation ponctuelle expérimentale prévient complètement l'interaction entre XRCC4 et Cer-XLF.

Dans l'optique d'un criblage virtuel basé sur la structure, je me suis donc donné comme objectif de trouver des molécules capable d'entrer en interaction avec, au moins, l'un de ces résidus au sein de l'interface d'interaction entre les deux protéines. A l'instar d'une mutation, le recrutement de l'un de ces résidus par une molécule organique et l'encombrement stérique pourrait inhiber la formation du complexe. Après analyse de ces contacts persistants, on note également un patch de trois ou quatre résidus successifs qui se détache en particulier pour chacune des protéines. Pour XRCC4, le résidu F106 fait partie d'un tétrapeptide V104-S-F-R107 dont l'interaction est très persistante (souligné en noir sur la figure 37). Pour Cer-XLF, il en est de même avec le tripeptide R64-L-T66. Ces clusters étant formés de

résidus successifs, j'ai par la suite envisagé de faire synthétiser ces derniers afin d'évaluer leur capacité à interagir avec les protéines XRCC4 et Cer-XLF et d'explorer une approche complémentaire d'inhibition potentielle mais basée sur des peptides. Cette partie sera développée un peu plus loin dans ce manuscrit.

II – 2.2 : Recherche de sites « druggables » sur les protéines XRCC4 et Cer-XLF

Comme évoqué dans le chapitre introductif, une surface protéique « druggable » doit posséder certaines caractéristiques (interface flexible, résidus hydrophobes, taille de la surface accessible au solvant...). Lors d'un criblage rationnel basé sur la structure, une étape importante est la sélection d'une ou des structures protéiques représentatives sur lesquelles sera lancée l'étape de docking. Il est donc important de prendre en considération toutes les structures disponibles issues de la base de données PDB (provenant par exemple d'un cristal ou d'une structure par RMN), d'une simulation de dynamiques moléculaires ou même d'une modélisation par homologie. Ces différentes structures obtenues expérimentalement ou *in silico* peuvent, pour une même protéine, avoir des différences mineures de conformation mais pouvant influencer fortement le résultat d'un criblage virtuel. L'utilité de ces variations structurales est de multiplier nos chances d'obtenir des molécules touches potentielles en explorant la variabilité conformationnelle de notre récepteur.

Les structures protéiques utilisées dans mon étude ont toutes une origine cristalline et sont présentes sous forme apo- ou complexée à leur partenaire XRCC4 ou Cer-XLF. J'ai tout d'abord effectué une analyse des différentes structures disponibles dans la PDB. Puis, j'ai comparé les formes apo- ou complexée de mes protéines afin d'en tirer des conclusions quant aux modifications structurales induites et à la flexibilité de mon interface. Le but étant de définir des sites potentiellement « druggables » sur lesquels j'ai effectué mes expériences de docking. Les structures étudiées sont représentées dans le tableau ci-dessous :

Code PDB	XRCC4	XLF	Complexe	Résolution	Date
1FU1	*			2,7Å	2000
3II6	*			2,4Å	2009
1IK9	*			2,3Å	2001
2QM4		*		2,3Å	2008
2R9A		*		2,5Å	2007
3WO3			*	8,5Å	2011
3RWR			*	3,9Å	2012
3SR2			*	3,9Å	2011
3Q4F			*	5,5Å	2011

Figure 38: Tableau représentant les différentes structures PDB étudiées de XRCC4 et Cer-XLF.

La date, la résolution du cristal et le code d'accession PDB pour chaque structure est représenté. Une étoile indique si la structure correspond à une protéine XRCC4 ou Cer-XLF seule ou bien en complexe XRCC4/Cer-XLF.

Pour mon étude de criblage virtuel par docking, je n'ai considéré que les têtes globulaires (partie N-terminale) de XRCC4 et de Cer-XLF, qui sont les zones d'interaction entre les deux partenaires. J'ai par conséquent recherché des cavités ou sillons apparents autour des résidus critiques dans lesquels des petites molécules pourraient se loger (sites druggables). Mon choix s'est donc orienté vers la recherche d'inhibiteurs orthostériques potentiels, sur ces deux protéines, et capables de rompre la liaison entre les deux têtes globulaires de XRCC4 et XLF et donc, la formation des filaments.

La superposition des structures 1FU1, 3II6 et 1IK9 de XRCC4 montre que des changements structuraux majeurs sont surtout localisés au niveau de la partie C-terminale avec les tiges en hélices (RMSD 3II6/1FU1 = 2.682 ; RMSD 1IK9/1FU1 = 1.795). Ces courbures sont surtout dues à la liaison de la protéine ligase IV (« XRCC4 Interacting Region » (XIR) pour la structure 1IK9 et C-terminal complet de la ligase IV pour 3II6). En N-terminal, là où sera centré le criblage virtuel, les structures se superposent mieux et avec des différences mineures (Région 1-226 : RMSD 3II6/1FU1 = 0.744 ; RMSD 1IK9/1FU1 = 0.641). La superposition des deux structures apo de Cer-XLF (2R9A et 2QM4) donne un RMSD de 1.067. En N-terminal, là où sera centré le criblage virtuel, les structures se superposent également avec des différences mineures (Région 1-227 : RMSD 2R9A/2QM4 = 0.619) (Figure 39).

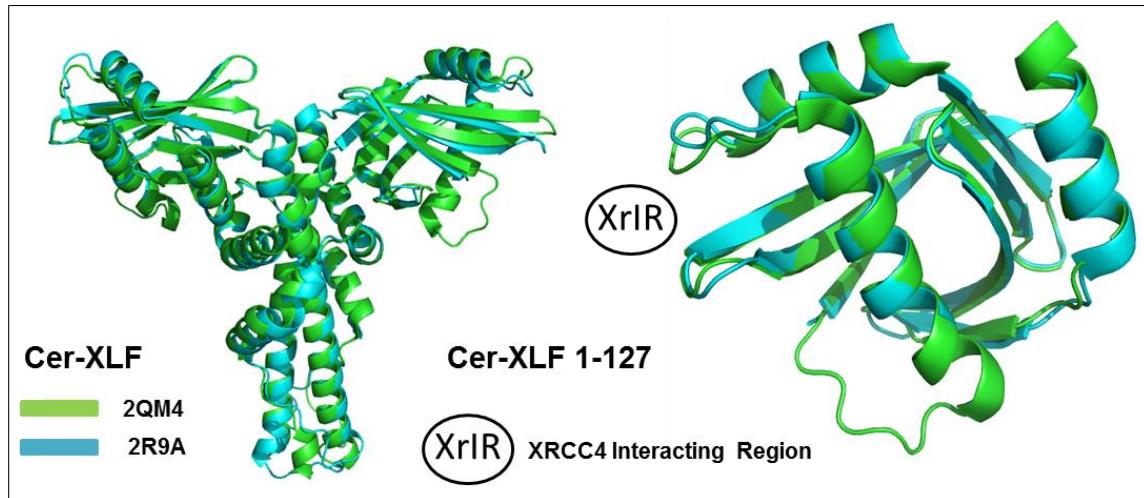
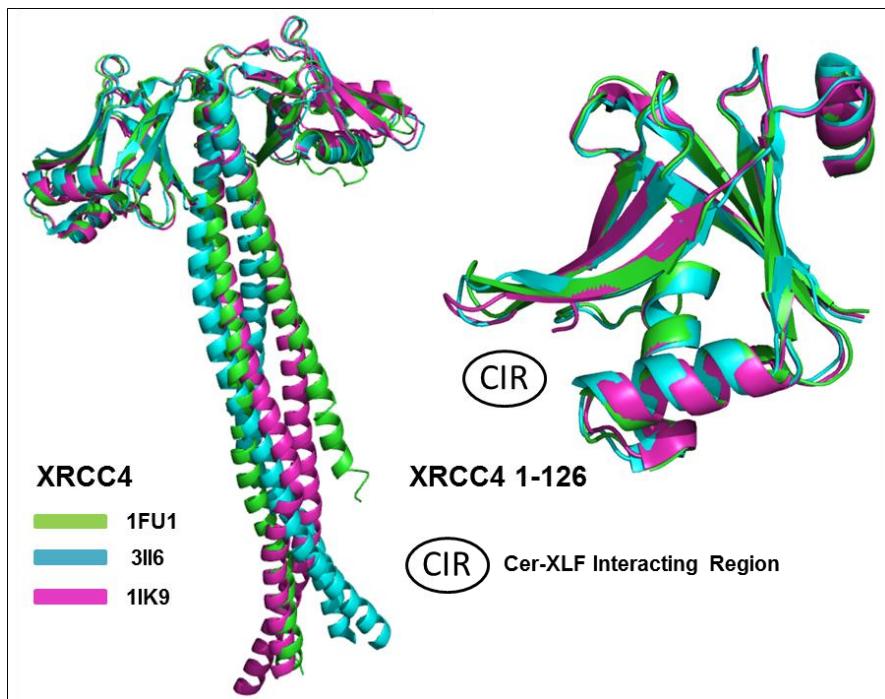


Figure 39: Alignement structural sur PyMOL des protéines XRCC4 et Cer-XLF.

La superposition des structures 1FU1 et 3SR2 de XRCC4 sur PyMOL montre que la forme complexée de XRCC4 à Cer-XLF ne subit pas de changements majeurs de conformation au niveau de son squelette (région 1-126 : RMSD = 0.804). Cependant, la liaison de Cer-XLF introduit de la flexibilité particulièrement au niveau de la zone d'interaction (tête globulaire, région L101 à F106). Dans le cas de Cer-XLF, il n'y a pas non plus de changements majeurs de conformation du squelette (région 1-127 : RMSD = 1.050). Cependant, une flexibilité dans la zone d'interaction de XRCC4 permet une « ouverture » facilitant la liaison à cette dernière (région E111 à L115).

Ce changement conformationnel majeur permet en effet à Cer-XLF d'amarrer sa leucine 115 dans une poche hydrophobe de la tête globulaire de XRCC4 formée par les résidus M59, M61 et F106 (Figure 40).

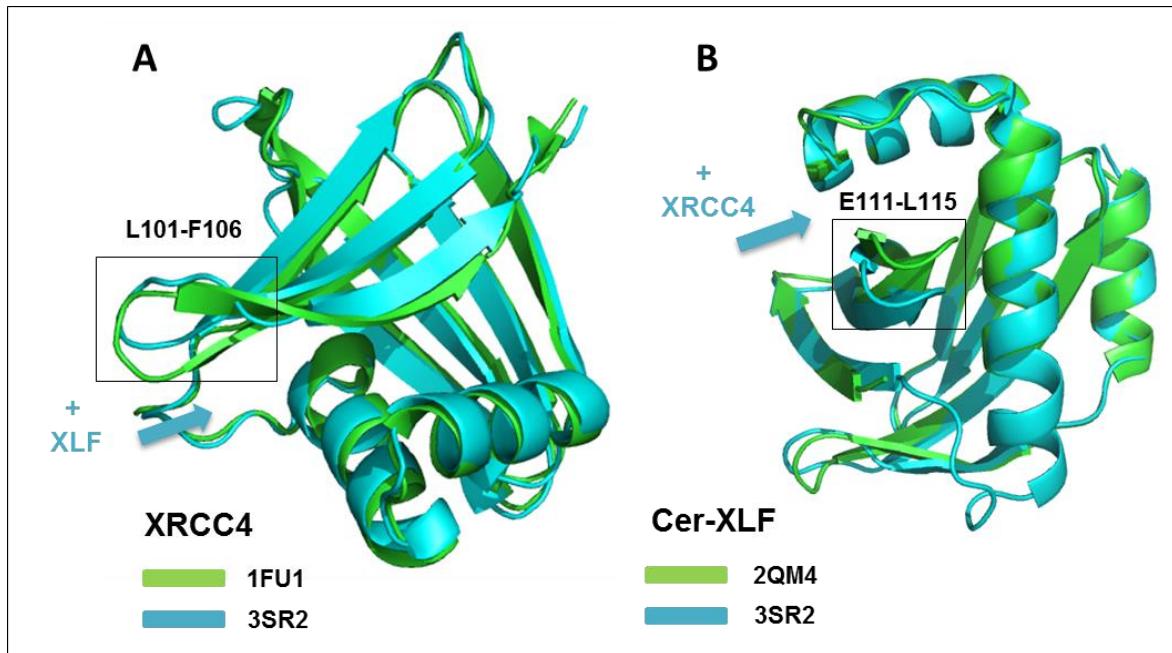


Figure 40: Alignement structural sur PyMOL des protéines XRCC4 et Cer-XLF sous forme seule ou en complexe.

(A) Partie N-terminale de XRCC4. L'encadré noir représente une zone de changement conformationnel consécutif à la liaison de Cer-XLF. (B) Partie N-terminale de Cer-XLF. L'encadré noir représente une zone de changement conformationnel consécutif à la liaison de XRCC4.

La zone d'interaction de Cer-XLF sur XRCC4 est un large sillon majoritairement hydrophobe exposé au solvant (site 1), sans cavité profonde et dont le ciblage pourrait s'avérer difficile mais intéressant (Figure 41). On remarque en plus de cette surface d'interaction une cavité plus profonde et également hydrophobe en amont (site 2). Cette cavité est non loin des résidus critiques et peut servir de point d'ancrage à des molécules venant s'insérer et former des liaisons avec ces derniers en aval. Le site 1 offre la possibilité de cribler des molécules plus grandes. Ces composés peuvent être des molécules organiques ou bien des peptides linéaires ou cycliques. Concernant le site 2, la cavité est présente dans toutes les structures cristallines disponibles de XRCC4 (1FU1 ; 3II6 ; 1IK9).

J'ai donc choisi de cibler les sites 1 et 2 lors de mon criblage virtuel dirigé contre XRCC4.

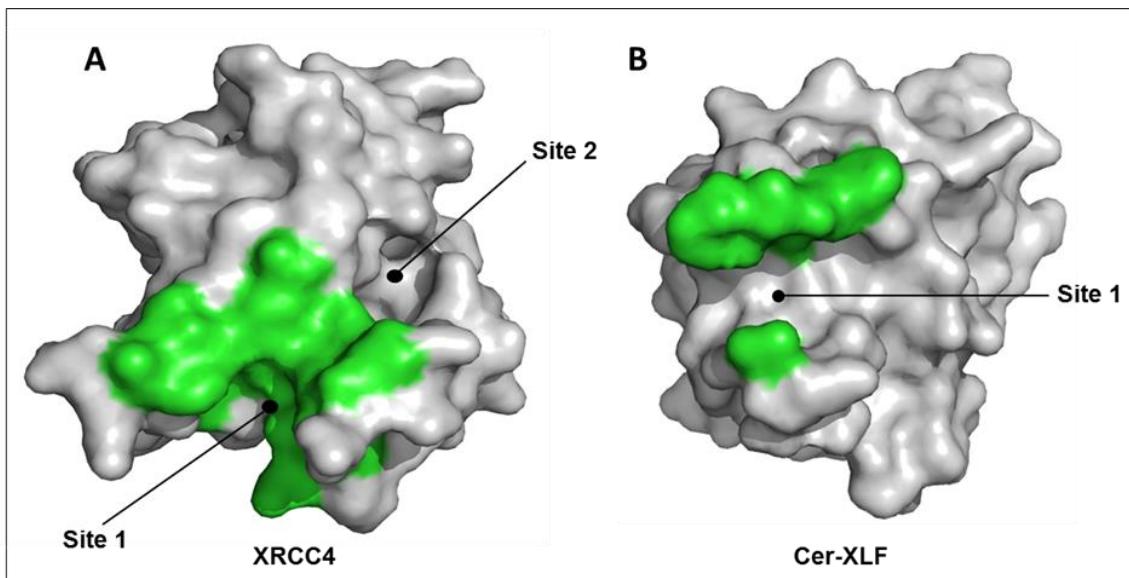


Figure 41: Représentation en surface des domaines N-terminaux globulaires de XRCC4 et Cer-XLF – Localisation des sites druggables.

(A) Partie N-terminale de XRCC4. Les sites potentiellement druggables 1 et 2 sont distribués autour des résidus critiques d'interaction à Cer-XLF (en vert). (B) Partie N-terminale de Cer-XLF. Le site potentiellement druggable 1 se localise autour des résidus critiques d'interaction à XRCC4 (en vert).

Du côté de Cer-XLF, la structure cristalline de la protéine seule non complexée à XRCC4 (2QM4) ne présente aucune cavité de taille intéressante. Cependant, nous venons d'observer que son interface est également dynamique et qu'elle adopte une conformation plus « ouverte » en présence de XRCC4 lorsqu'elle est co-cristallisée avec cette dernière (3SR2). La capacité de Cer-XLF à adopter cette conformation particulière m'a conduit à me demander si une même dynamique pouvait se présenter pour accommoder un ligand de bonne affinité. J'ai donc choisi de cibler cette structure ouverte (site 1) lors de mon criblage virtuel dirigé contre Cer-XLF. Sur la figure ci-dessous, les sites considérés pour notre docking sont précisés sur les têtes de XRCC4 et Cer-XLF et les résidus critiques précédemment cités sont représentés en vert autour de ces sites.

Des sites potentiellement « druggables » sur chacune des protéines ont donc été définis grâce à la disponibilité des structures cristallographiques. Il fallait à présent choisir des structures représentatives pour notre docking et déterminer une méthode de criblage virtuel adaptée en tenant compte du paramètre de flexibilité. Nous partons du principe que ces structures représentatives sont évidemment des

structures parmi beaucoup d'autres mais permettent une étape de pré-criblage par docking pour la sélection de molécules touchées intéressantes (étape d'enrichissement). Nous introduisons de la flexibilité au sein des complexes protéine-molécule ainsi trouvés lors d'une étape ultérieure de DM. Par ailleurs, la flexibilité des chaînes latérales de certains des résidus critiques sera prise en compte pendant le docking et les ligands seront eux, complètement flexibles. Afin de sélectionner des structures représentatives, une simulation de DM de dix nanosecondes a d'abord été lancée sur la tête globulaire de XRCC4 (résidus 1-126) à partir de la structure 1FU1. Cette DM nous apporte deux éléments importants :

- L'obtention de familles énergétiquement « relaxée » de cette tête globulaire que l'on utilisera ensuite comme structures représentatives,
- Déterminer si le site 2 sur XRCC4 constitue une cavité persistante au cours du temps ou bien si c'est une conséquence des conditions de cristallisation.

II – 2.3 : Echantillonnage conformationnel pour le criblage virtuel – Simulations de dynamiques moléculaires sur la tête XRCC4 et analyse des centroides de chaque cluster

Durant cette trajectoire, la cavité (site 2) se referme et se stabilise dans une conformation moins accessible pour une petite molécule par l'intermédiaire d'une liaison hydrogène entre la chaîne latérale du glutamate 55 et le squelette de la serine 110 (figure 42). L'analyse des centroides (de chaque cluster de conformations) au cours de cette trajectoire ne me permet d'obtenir qu'une population de conformères largement majoritaire (centroïde 1 : 6143 structures). Une forme relaxée représentative de ce centroïde majoritaire a donc été extraite. L'analyse comparative du plot de Ramachandran entre cette forme tronquée relaxée et la forme cristalline par le programme PROCHECK (Morris et al., 1992) montre que cette forme relaxée est très stable au cours de la trajectoire et possède 86,7% de résidus dans des régions favorables en fin de DM contre 76,1% avant la DM. Cette forme relaxée de XRCC4 présente des variations au niveau du site 2 à cibler mais en dépit d'une cavité moins accessible par rapport à la structure cristalline de départ, elle représente une seconde structure possible pour le criblage virtuel. Ces deux

structures de XRCC4 (cristalline 1FU1 et relaxée par DM) ont donc été utilisées dans ce but :

- Structure relaxée par DM : recherche de molécules ciblant le site 1.
- Structure cristalline 1FU1 : recherche de molécules ciblant le site 2.

Concernant Cer-XLF, la structure « ouverte » 3SR2 issue du complexe avec XRCC4 sera utilisée pour cibler le site 1.

Trois docking semi-flexibles seront donc lancés sur trois structures représentatives et dans le but de trouver des molécules inhibitrices de l'interaction XRCC4/Cer-XLF.

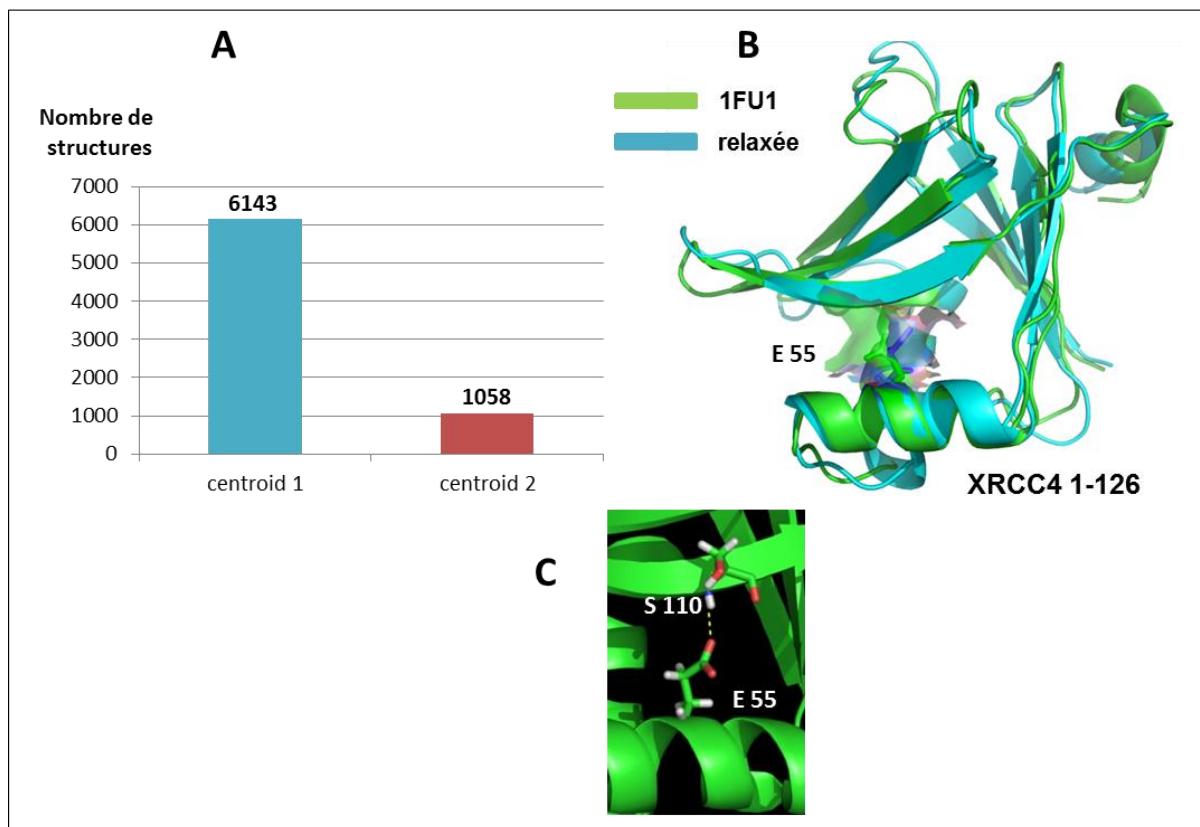


Figure 42: Simulation de DM de dix nanosecondes sur le domaine XRCC4 N-terminal 1-226 – Analyse des centroïdes et de la dynamique du site 2.

(A) Analyse des centroïdes obtenus et leur population de conformères. Le centroïde 1 est largement majoritaire avec 6143 structures. (B) Superposition entre la structure cristalline 1FU1 du domaine XRCC4 1-126 (vert) et la forme relaxée par DM (cyan) représentant la structure représentative du centroïde 1. On observe le mouvement et la stabilisation du glutamate 55, fermant ainsi la cavité du site 2 dans la structure relaxée. (C) stabilisation du glutamate 55 au sein du site 2 par un pont hydrogène avec la serine 110.

II – 2.4 : Préparation et filtrage d'une base de données virtuelle de molécules

Les molécules que nous avons choisies pour notre criblage devaient être accessibles sur internet, annotées, avec une grande diversité chimique et commercialement disponibles afin de pouvoir rapidement tester les molécules touches prioritaires dans nos essais de validation biochimique et biophysique. Les molécules provenant des bases de données : Asinex, Chemblock, Enamine, Chembridge, CNE, Maybridge, NCI, Otava, Prestwick, Sinova, TCM, Zenobia et un sous-jeu de données de ZINC ont été téléchargées dans un format classique .sdf puis traitées grâce à des scripts (non développés) afin :

- De les convertir au format .mol2,
- D'enlever les sels (nettoyage),
- De rajouter les hydrogènes manquants pour un pH donné à 7.2,
- De vérifier l'état d'ionisation,
- De générer une conformation 3D énergétiquement favorable,
- Enfin, de les convertir au format .pdbqt (format nécessaire au lancement du docking sur Autodock Vina).

Au final, pour un temps de calcul raisonnable, un total d'environ deux millions de molécules d'une taille comprise entre 350 et 800 Daltons ont été sélectionnées. Les interfaces protéiques étudiées étant assez larges (notamment les sites 1) et en sachant que le profil physico-chimique caractéristique d'un modulateur d'interaction protéine-protéine inclus également une taille plus importante des molécules par rapport à des modulateurs de cibles conventionnelles (Pagliaro et al., 2004 ; Reynes et al., 2010), j'ai choisi de ne pas me limiter à des molécules d'une taille inférieure ou égale à 500 Daltons, une taille habituellement choisie en accord avec les règles de Lipinski (Lipinski et al., 2001). Ces molécules ont ensuite été filtrées sur un critère de solubilité avec un LogP < 2 à pH 7.2. Seront enfin supprimés, les dupliques, les molécules possédant un groupement toxique (mutagènes), les faux positifs connus (avec des fonctions réactives, des « warheads » ou des motifs d'agrégation). Le logiciel « Screening Assistant 2 » (Guilloux et al., 2012) sera utilisé à cet effet. Un

total de 520.446 molécules a été obtenu pour le docking sur les protéines XRCC4 et Cer-XLF.

II – 2.5 : Docking sur les têtes XRCC4 et Cer-XLF – Tri et analyse des hits potentiels

Le criblage que nous souhaitons effectuer est un criblage virtuel basé sur la structure. Les différentes étapes sont schématisées ci-dessous (Figure 43). Nos expériences de docking sont lancées sur des structures cristallographiques ou issues de simulations de dynamiques moléculaires avec une chimiothèque virtuelle préalablement filtrée (discuté ci-dessus). Ce docking est semi-flexible ce qui signifie que certaines chaînes latérales de résidus clés au niveau du site d'intérêt seront rendus flexibles. A la fin de ce docking, les poses théoriques les plus énergétiquement stables d'après la fonction de score du programme seront extraites. Le centre de gravité de chacune des molécules sera comparé au centre de gravité du site d'intérêt et les ligands les plus proches de ce site seront conservés. Les 200 meilleures molécules d'après leur score seront extraites et seront finalement conservées uniquement les molécules qui forment au moins une liaison hydrogène et un contact hydrophobe avec les résidus clés identifiés précédemment.

L'étape de docking, bien que sujette à de nombreux faux-positifs, est une étape de pré-criblage permettant un enrichissement en molécules touchées potentiellement actives et permettant de réduire l'espace chimique à un nombre de candidats dont les tests biophysiques et biochimiques sont physiquement possibles.

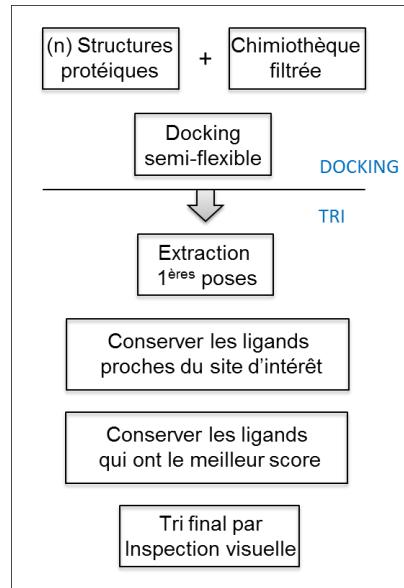


Figure 43: Schéma représentatif des différentes étapes nécessaires au lancement et à l'analyse d'un criblage virtuel basé sur la structure.

1. Docking sur le domaine N-terminal globulaire de XRCC4 (1-126) :

Un alignement de séquences grâce au programme BLAST a été effectué (Figure 44) afin de s'assurer que la protéine XRCC4 que nous possédons physiquement correspond parfaitement à la protéine sur laquelle nous lançons le programme de docking (PDB : 1FU1), et plus particulièrement au niveau de la zone d'intérêt qui concerne la tête globulaire (résidus 1-126) :

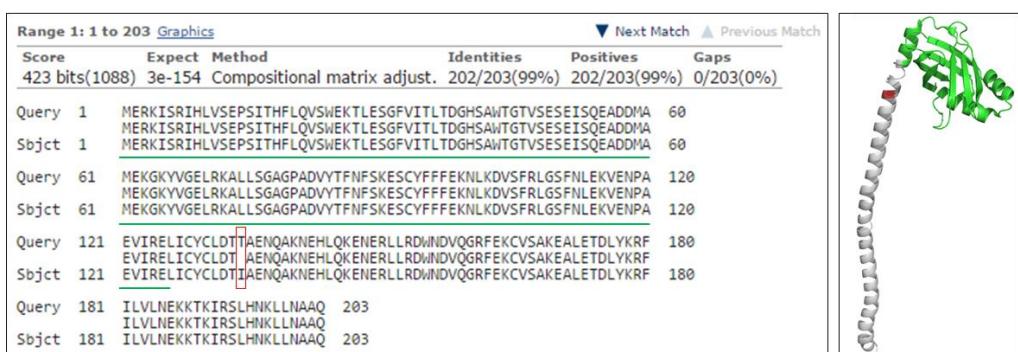


Figure 44: Alignement de séquence par le programme BLAST pour la protéine XRCC4. La séquence protéique (monomère) de XRCC4 utilisée expérimentalement au cours de ma thèse est comparée à la séquence de la protéine cristallisée 1FU1. La séquence 1-126 (en vert) représente la zone considérée pour le criblage. Une mutation ponctuelle (en rouge) ne concerne pas la zone ciblée.

L'alignement de séquences a été effectué sur un seul des monomères de XRCC4. La séquence « Subject » correspond à l'insert XRCC4 contenu dans le plasmide Pet28a que nous possédons (voir partie Matériels et Méthodes). La séquence « Query » est celle de la protéine 1FU1 cristallisée. Mise à part une substitution T134>I (en rouge) située au niveau de la région coiled-coil, la séquence d'intérêt (1-126 représentée en vert) est parfaitement identique.

- Ciblage virtuel du site 1 : (Structure relaxée par DM)

Les détails structuraux de la protéine XRCC4 (1-126) relaxée par dynamique moléculaire sont représentés dans la figure ci-dessous (figure 45). Les molécules candidates issues du docking devaient, dans une conformation énergétiquement stable, contacter un résidu du patch hydrophobe constitué par M59, M61, L101, V104 et F106 et seraient plus particulièrement favorisés les molécules contactant M61 ou F106 (en rouge sur la représentation) qui sont des résidus clés dans l'interaction avec Cer-XLF, définis par les expériences de mutagénèse dirigée. De la même manière, ces molécules favorites devaient en même temps contacter ou être à proximité des résidus clés E55 et D58 (en rouge sur la représentation).

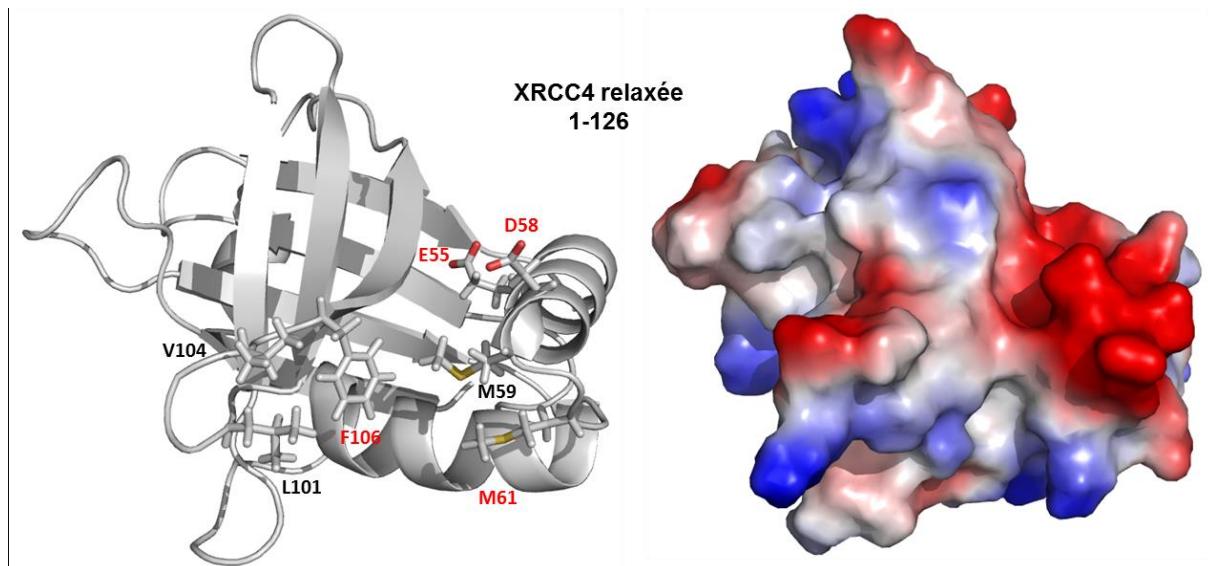


Figure 45: Analyse structurale du domaine N-terminal (1-126) de XRCC4 relaxé par DM – détails du site 1.

La protéine relaxée est représentée en structures secondaires et en surface électrostatique. Les résidus en rouge correspondent aux résidus dont la mutation ponctuelle expérimentale abolie complètement la liaison à Cer-XLF.

Durant l'étape de docking semi-flexible, la libre rotation des chaînes latérales des résidus représentés ci-dessus a été permise. Simuler la flexibilité de la protéine augmente de façon exponentielle le temps de calcul mais bien que chronophage cela permet une meilleure exploration de l'espace conformationnel de cette dernière, la sélection par les molécules d'une conformation particulière ainsi qu'une plus grande liberté dans les poses des molécules. Une conformation pour chacune des molécules (considérée comme la plus stable énergétiquement d'après le programme de docking) a été extraite et toutes les molécules se localisant au-delà de 5 angströms du centre de gravité du site d'intérêt ont été exclues. Sur 6632 ligands, les 200 meilleures en terme de score ont été gardées et l'histogramme de distribution de ces scores montre des énergies prédites allant de -9kcal/mole pour les meilleurs candidats à -7.8kcal/mole (figure 46).

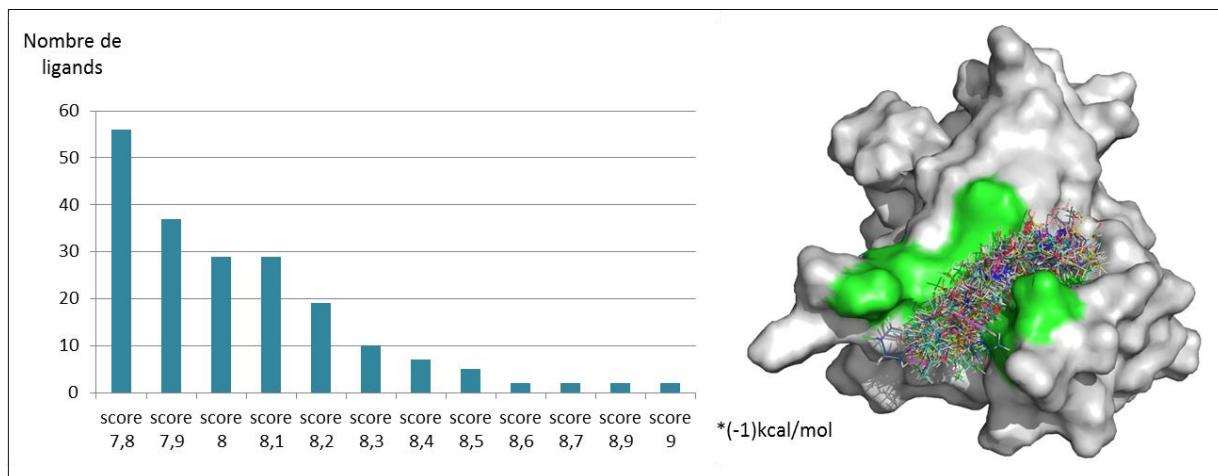


Figure 46: Docking semi-flexible sur le site 1 du domaine XRCC4 (1-126) relaxé par DM – histogramme de distribution des scores de docking et du nombre de ligands correspondant.

La protéine relaxée est représentée en surface et les résidus critiques d'interaction à Cer-XLF sont en vert. Les premières poses de docking pour chacune des molécules retenues sont superposées sur la région d'intérêt.

Les molécules qui forment au moins une liaison hydrogène et un contact hydrophobe avec les résidus clés identifiés précédemment ont été conservées. Après élimination des duplicats, des molécules avec un ou des groupements toxiques et des faux-positifs, nous avons obtenu un total de 82 molécules. Une projection (TPSA v.s MW) de ces 82 candidats, obtenue par le programme Screening Assistant 2 est représentée ci-dessous (figure 47).

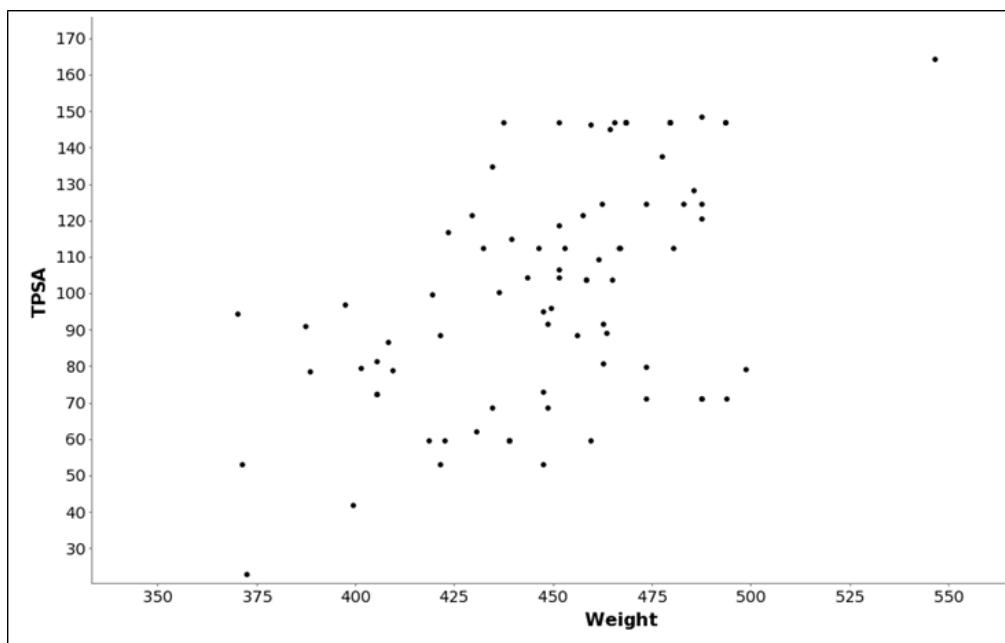


Figure 47: Projection (surface polaire totale v.s poids moléculaire) des 82 candidats retenus à l'issu du docking semi-flexible sur le site 1 du domaine XRCC4 (1-126) relaxé par DM, obtenu par le programme « screening Assistant ».

Cette projection rend compte de la taille et de l'aire totale de la surface polaire de nos molécules. Avec pour quasiment toutes ces molécules, une taille ≤ 500 Da, une TPSA $\leq 140\text{\AA}^2$ et un LogP ≤ 2 , nous réunissons au moins trois des critères de Lipinski (Lipinski et al., 2001) en terme de biodisponibilité. Ces 82 molécules ont été également superposées sur le site d'intérêt (en vert, figure 46) et possèdent pour la plupart un scaffold commun avec un nombre important de contacts (polaires et hydrophobes) au sein du site 1 avec des résidus critiques. Ces scaffolds sont représentés en Figure 48 et apportent un modèle d'interaction prédictive de ces molécules avec XRCC4. De plus, il est possible de rechercher à partir de ces scaffolds, des molécules similaires lors d'une étape ultérieure d'optimisation. Certaines molécules n'ont cependant rien à voir avec ce scaffold mais représentent des candidats potentiels également très intéressants (voir exemple du candidat #378660 en figure 49).

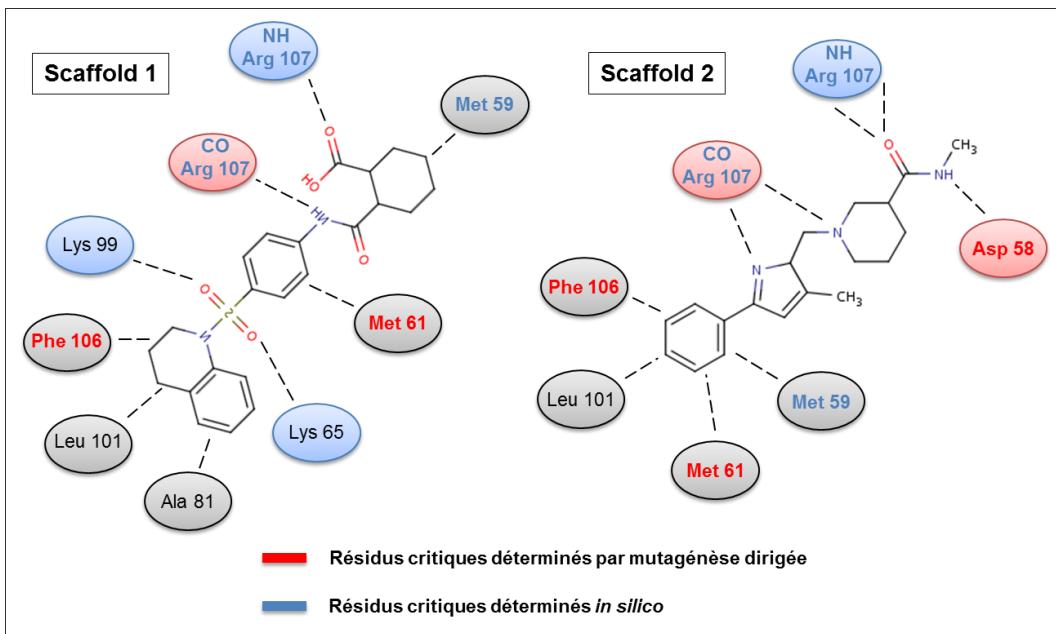


Figure 48: Docking semi-flexible sur le site 1 du domaine XRCC4 (1-126) relaxé par DM – représentation des scaffolds chimiques communs observés pour certains candidats retenus.

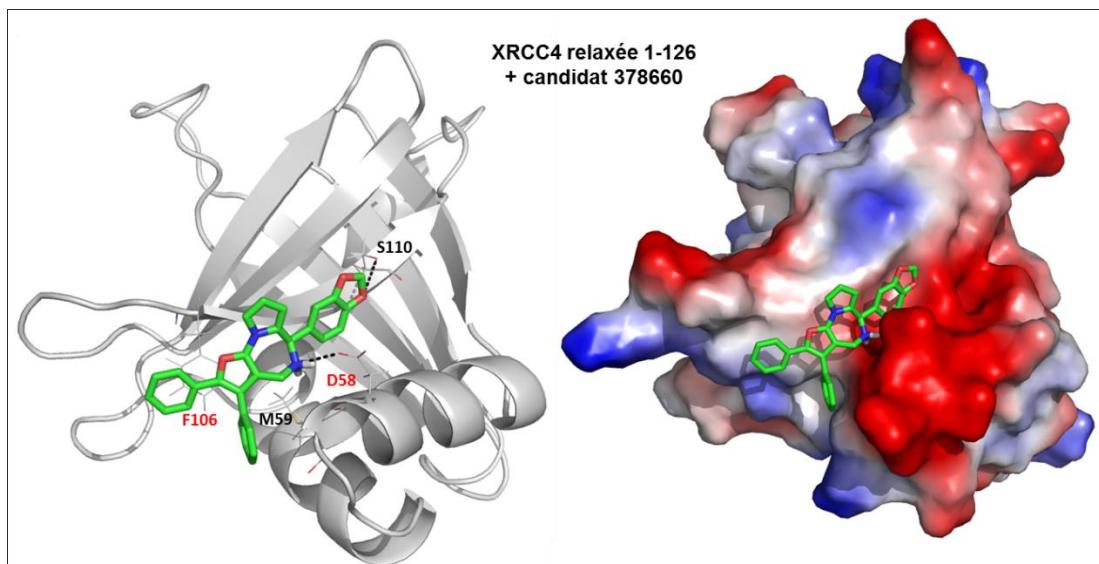


Figure 49: Docking semi-flexible sur le site 1 du domaine XRCC4 (1-126) relaxé par DM – conformation stable et détails des interactions du candidat #378660 avec sa cible.

Parmi les résidus contactés par le candidat, D58 et F106 (en rouge) sont des résidus critiques d'interaction avec Cer-XLF. Les liaisons hydrogènes sont représentées en pointillés noir.

Finalement, cet enrichissement à partir d'une base de données initiale de plus de deux millions de molécules constitue une seconde base beaucoup plus réduite avec

un nombre d'inhibiteurs potentiels dont la validation biologique devient physiquement possible. La prochaine étape consiste à soumettre chacun des 82 complexes protéine-ligand à une simulation de dynamique moléculaire afin d'évaluer leur stabilité et énergie d'interaction.

- Criblage virtuel du site 2 : (Structure cristallographique 1FU1)

Les détails structuraux de la structure 1FU1 de la protéine XRCC4 (1-126) sont représentés dans la figure ci-dessous (figure 50). Les molécules candidates issues du docking devaient, dans une conformation énergétiquement stable, se stabiliser au sein de la poche hydrophobe servant de point d'ancrage constitué par F32, V47, L70 et F111 et contacter en même temps un des résidus critiques E55 ou D58 (en rouge sur la représentation) qui sont des résidus clés dans l'interaction avec Cer-XLF, définis par les expériences de mutagénèse dirigée. Durant l'étape de docking semi-flexible, la libre rotation des chaînes latérales des résidus cités ci-dessus devait permettre à la poche d'avoir un diamètre variable et ainsi accommoder plus facilement des ligands potentiels.

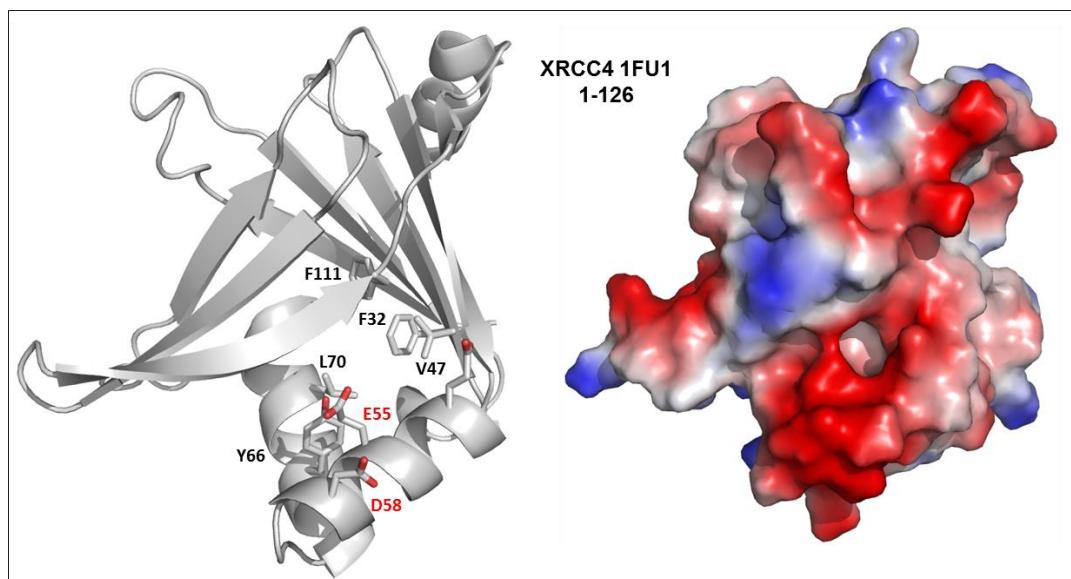


Figure 50: Analyse structurale du domaine N-terminal (1-126) de XRCC4 (structure cristallographique 1FU1) – détails du site 2.

La protéine 1FU1 (1-126) est représentée en structures secondaires et en surface électrostatique. Les résidus en rouge correspondent aux résidus dont la mutation ponctuelle expérimentale abolie complètement la liaison à Cer-XLF.

De façon surprenante, peu de candidats en fin de docking ont pu être retenus au vu des caractéristiques citées ci-dessus. Ceci est surprenant car une poche hydrophobe au sein de la protéine semble beaucoup plus propice à la découverte de molécules touchées qu'un sillon sans cavité apparente.

Un seuil de score choisi de -8.5kcal/mole a permis la sélection de 17 molécules candidates dont 4 ont été finalement retenues après élimination des dupliques, groupements toxiques et molécules dont le nombre de contacts polaires et hydrophobes a été considéré comme insuffisant. Les 4 molécules sélectionnées (appelées #43335, #43741, #60309 et #61284) possèdent toutes la caractéristique commune de s'insérer via un cycle phényle au sein de cette cavité servant de point d'ancrage, et de contacter des résidus hydrophiles plus en aval dont le résidu critique E55 (figure 51). La grande similarité entre ces quatre composés nous permettra par la suite d'entamer une étude de structure-activité.

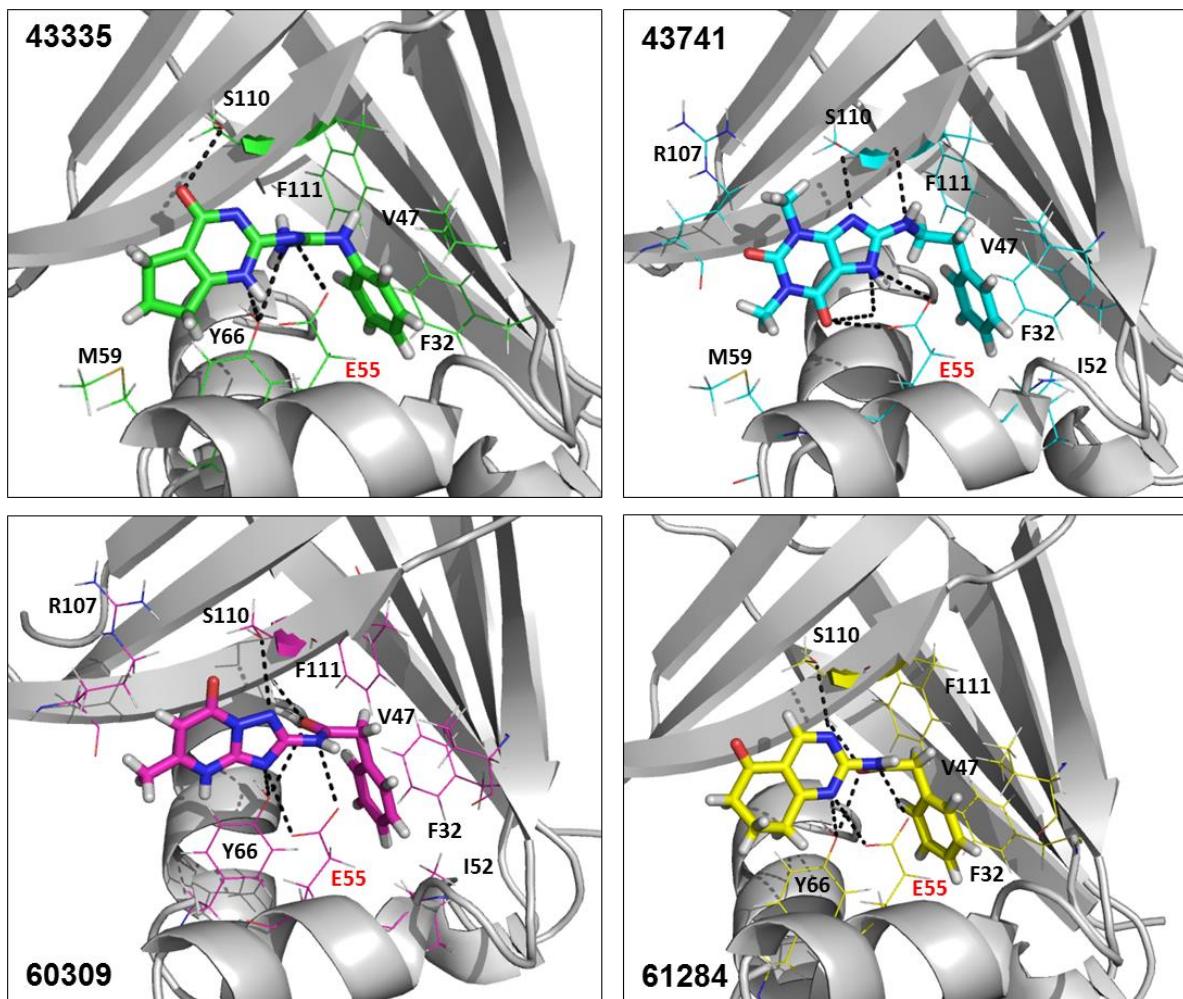


Figure 51: Docking semi-flexible sur le site 2 du domaine N-terminal (1-126) de XRCC4 (structure cristallographique 1FU1) – conformation stable et détails des interactions des 4 candidats retenus avec leur cible.

Parmi les résidus contactés par les candidats, E55 (en rouge) est un résidu critique d'interaction avec Cer-XLF. Les liaisons hydrogènes sont représentés en pointillés noir.

La prochaine étape consiste à soumettre ces 4 complexes protéine-ligand à une simulation de dynamique moléculaire afin d'évaluer leur stabilité et énergie d'interaction.

2. Docking sur le domaine N-terminal globulaire de Cer-XLF (1-127) :

Un alignement de séquences a été effectué (Figure 52) afin de s'assurer que la protéine Cer-XLF que nous possédons physiquement correspond parfaitement à la protéine sur laquelle nous lançons le programme de docking (PDB : 3SR2), et plus particulièrement au niveau de la zone d'intérêt qui concerne la tête globulaire (résidus 1-127) :

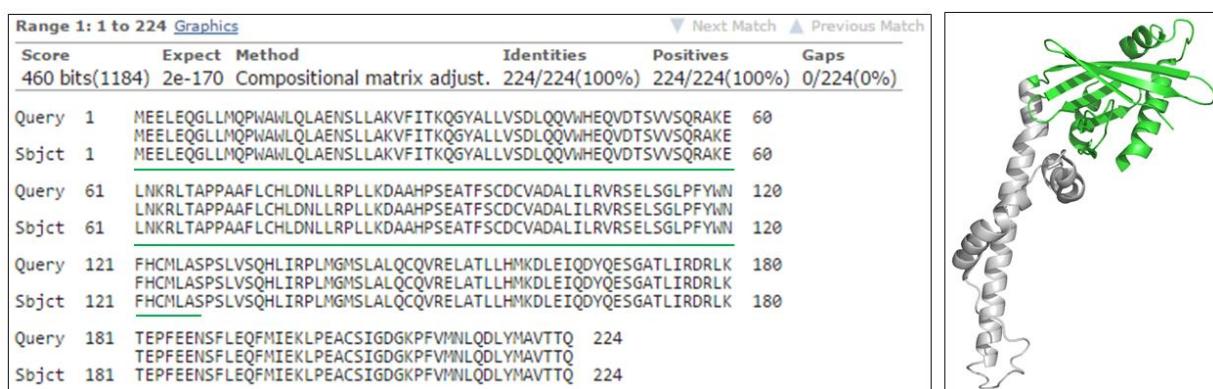


Figure 52: Alignement de séquence par le programme BLAST pour la protéine Cer-XLF.

La séquence protéique (monomère) de Cer-XLF utilisée expérimentalement au cours de ma thèse est comparée à la séquence de la protéine cristallisée 3SR2. La séquence 1-127 (en vert) représente la zone considérée pour le criblage. Aucune mutation n'est présente dans la zone ciblée.

L'alignement de séquence a été effectué sur un seul des monomères de Cer-XLF. La séquence « Subject » correspond à l'insert Cer-XLF contenu dans le plasmide pDEST-14 que nous possédons (voir partie Matériels et Méthodes). La séquence « Query » est celle de la protéine Cer-XLF dans le co-cristal 3SR2. La séquence d'intérêt (1-127 représentée en vert) est parfaitement identique dans les deux cas.

- Ciblage virtuel du site 1 : (Structure cristallographique 3SR2)

Les détails structuraux de la structure 3SR2 de la protéine Cer-XLF (1-127) au niveau du site d'intérêt sont représentés dans la figure ci-dessous (figure 53). Les molécules candidates issues du docking devaient, dans une conformation énergétiquement stable, se stabiliser au niveau de l'interface dynamique constitué majoritairement par le penta-peptide R64-P68 et la Leucine 115 et contacter au moins un des résidus critiques R64 ou L115 (en rouge sur la représentation) qui sont des résidus clés dans l'interaction avec XRCC4, définis par les expériences de mutagénèse dirigée. Là encore, l'étape de docking est semi-flexible, permettant la libre rotation des chaînes latérales des résidus cités ci-dessus.

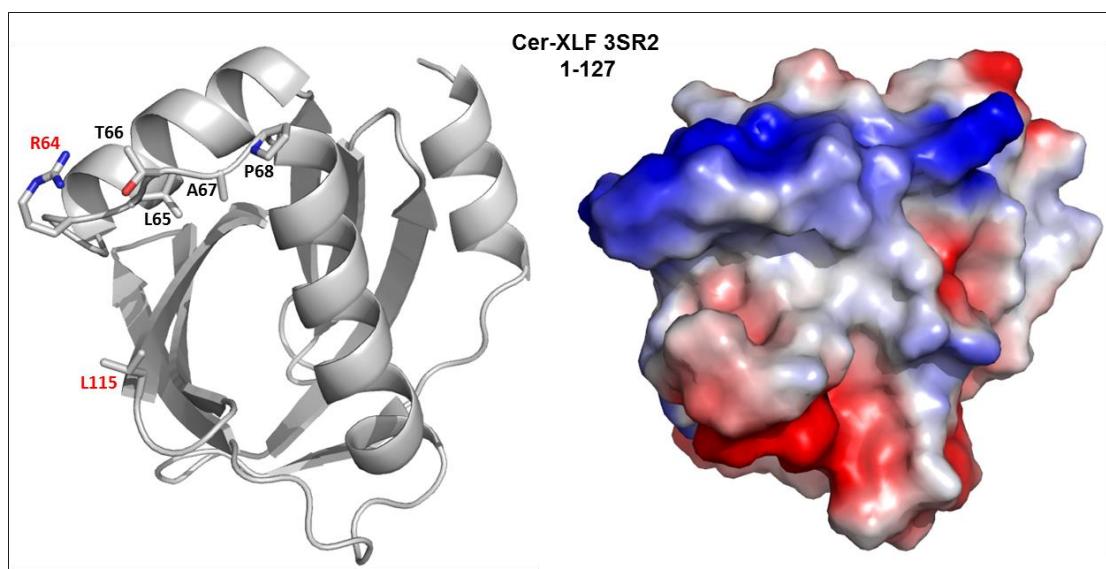


Figure 53: Analyse structurale du domaine N-terminal (1-127) de Cer-XLF (structure cristallographique 3SR2) – détails du site 1.

La protéine 3SR2 (1-127) est représentée en structures secondaires et en surface électrostatique. Les résidus en rouge correspondent aux résidus dont la mutation ponctuelle expérimentale abolie complètement la liaison à XRCC4.

Les démarches de docking et sélection de molécules candidates finales sont les mêmes que précédemment. En fin de calcul, la première conformation de chacune des molécules (considérée comme la plus stable d'après le programme de docking) a été extraite et toutes les molécules se localisant à proximité du centre de gravité du site d'intérêt ont été retenues. Les 200 meilleures molécules en terme de score ont été gardées et l'histogramme de distribution de ces scores montre des énergies

prédites allant de -9.7kcal/mole pour les meilleurs candidats à -7.4kcal/mole (figure 54).

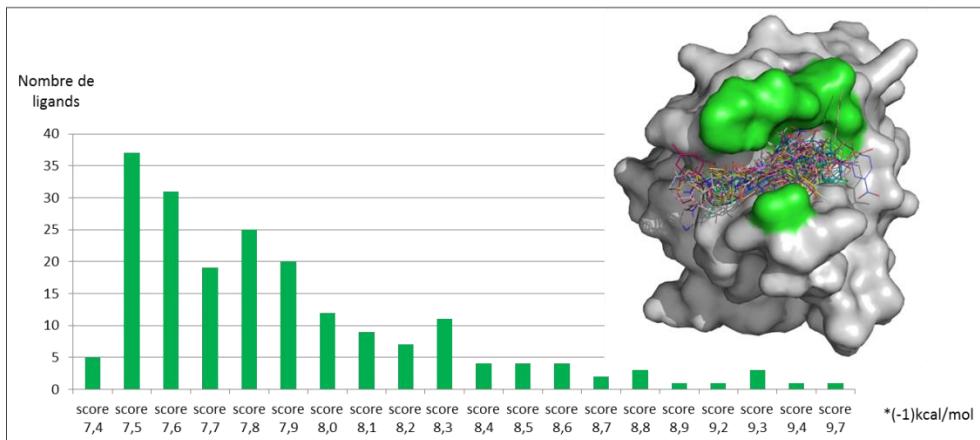


Figure 54: Docking semi-flexible sur le site 1 du domaine Cer-XLF (1-127) – histogramme de distribution des scores de docking et du nombre de ligands correspondant.

La protéine 3SR2 est représentée en surface et les résidus critiques d'interaction à Cer-XLF sont en vert. Les premières poses de docking pour chacune des molécules retenues sont superposées sur la région d'intérêt.

Les molécules qui forment au moins une liaison hydrogène et un contact hydrophobe dans le site d'intérêt et particulièrement avec les résidus critique R64 et L115, ont été conservées. Après élimination des dupliques, des molécules avec un ou des groupements toxiques et des faux-positifs, nous avons obtenu un total de 28 molécules. Une projection (TPSA v.s MW) de ces 28 candidats, obtenue par le programme Screening Assistant 2 est représentée ci-dessous (figure 55).

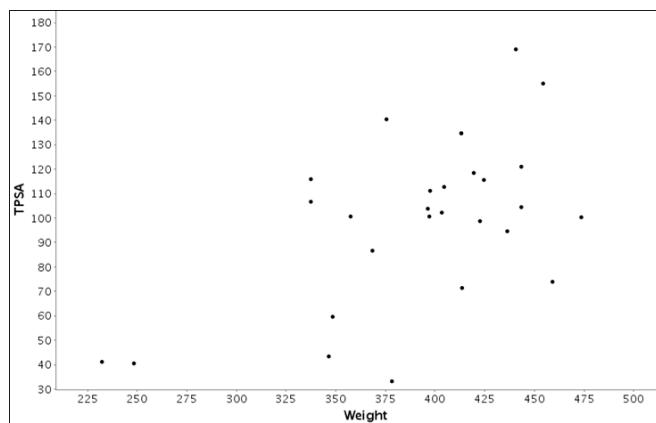


Figure 55: Projection (surface polaire totale v.s poids moléculaire) des 28 candidats retenus à l'issu du docking semi-flexible sur le site 1 du domaine Cer-XLF (1-127), obtenu par le programme « screening Assistant ».

Là encore, cette projection rend compte de la taille et de l'aire totale de la surface polaire de nos molécules, et avec pour quasiment toutes ces molécules une taille $\leq 500\text{Da}$, une TPSA $\leq 140\text{\AA}^2$ et un LogP ≤ 2 , nous réunissons au moins trois des critères de Lipinski en terme de biodisponibilité. Ces 28 molécules ont été également superposées sur le site d'intérêt (en vert, figure 54) et possèdent pour la plupart un scaffold commun s'insérant parfaitement au sein du site 1 autour des résidus critiques. Ce scaffold commun est représenté en figure 56 et apporte un modèle d'interaction prédictive de ces molécules avec Cer-XLF.

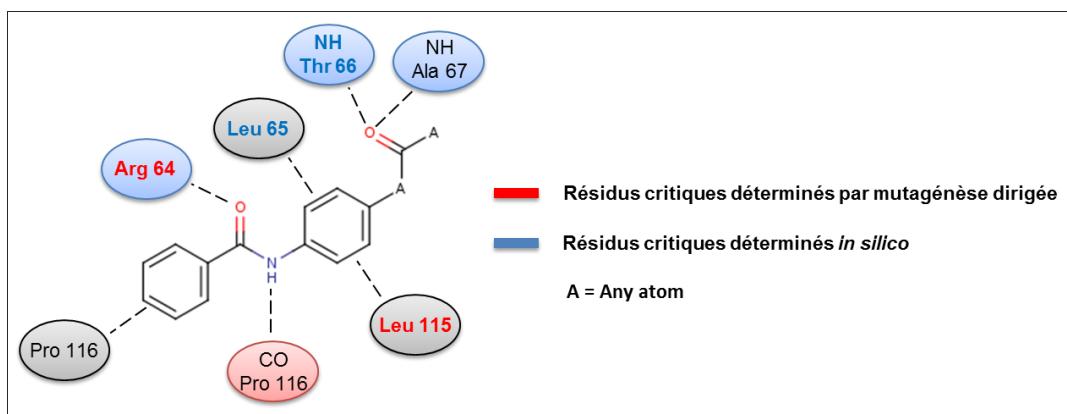


Figure 56: Docking semi-flexible sur le site 1 du domaine Cer-XLF (1-127) – représentation d'un scaffold chimique commun observé pour certains candidats retenus.

La prochaine étape consiste à soumettre chacun des 28 complexes protéine-ligand à une simulation de dynamique moléculaire afin d'évaluer leur stabilité et énergie d'interaction.

II – 2.6 : Analyse par dynamique moléculaire de la stabilité et de l'énergie de liaison des complexes XRCC4 hit et Cer-XLF/hit sélectionnés – Modèles d'interaction

Comme il a été discuté en introduction, l'implémentation du docking par des simulations de dynamiques moléculaires en criblage virtuel basé sur la structure est une méthode parfaitement adaptée à l'introduction du paramètre de flexibilité au cours du processus de découverte de médicaments. La liaison d'une molécule au sein de la protéine est un évènement dynamique. La protéine est en mouvement constant et une molécule peut lier et stabiliser une conformation particulière (sélection de conformation). Cette dernière peut également induire des changements

conformationnel supplémentaires et postérieurs à la liaison (phénomène d'ajustement induit). Ce phénomène peut s'observer lorsque la protéine accommode et s'adapte parfaitement à la liaison d'une molécule dans une conformation énergétiquement stable. Nous verrons plus loin que ce phénomène d'ajustement induit a été observé dans le cas de nos systèmes.

Les avantages de la DM sont multiples. Elle permet notamment de vérifier au cours d'une trajectoire, la stabilité d'un complexe protéine-ligand au sein d'un environnement virtuel solvaté. Elle permet également d'estimer de façon plus précise une énergie de liaison par rapport à un algorithme de scoring issu d'un docking et par conséquent, un meilleur classement des molécules touchées entre elles. Des contacts médiés par l'eau sont également souvent négligés pendant l'étape de docking mais peuvent jouer un rôle crucial lors du processus de liaison. Les différentes étapes post-docking sont schématisées ci-dessous (Figure 57).

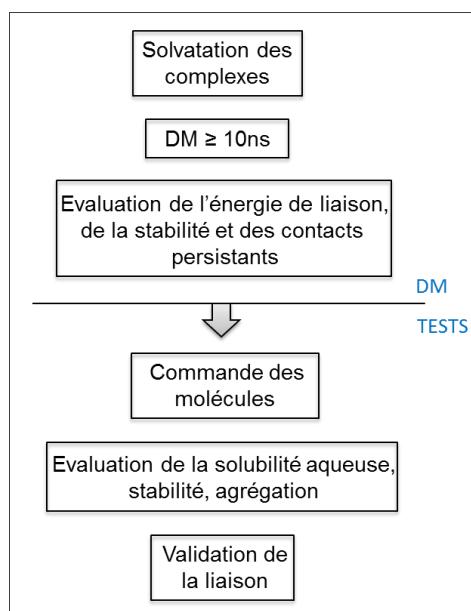


Figure 57: Schéma représentatif des différentes étapes post-docking nécessaires à l'évaluation et à la validation de molécules candidates.

La préparation des complexes et les scripts nécessaires à la préparation et au lancement de la DM sont détaillés en partie Matériels et Méthodes. Les complexes protéines-ligands sélectionnés après l'étape de docking contiennent les molécules dans leur conformation la plus stable, d'après les paramètres de score du

programme de docking. A partir de ce fichier, le système évoluera au cours du temps et dans une boîte solvatée afin de se stabiliser ou bien s'effondrer si le complexe est trop instable. Une dynamique « longue » d'au moins dix nanosecondes est nécessaire afin de pouvoir juger de la stabilité d'un complexe. A l'issue de cette DM, une évaluation de la stabilité, des contacts persistants et de l'énergie d'interaction permettra de classer les touches et sélectionner uniquement les plus prometteuses pour des essais biophysiques et biochimiques.

1. *Simulations de DM sur le domaine N-terminal globulaire de XRCC4 (1-126)*

Nous aborderons ici les simulations faites sur les deux structures XRCC4 utilisées pour le docking semi-flexible : la forme relaxée par DM et la forme cristallographique. Un total de 86 complexes XRCC4/hit potentiel a été soumis à une simulation de DM de dix nanosecondes. A l'issue de ces dynamiques, chaque candidat est d'abord jugé sur sa stabilité au cours de la trajectoire (mesure du RMSD) ainsi que sur ses contacts persistants potentiels avec la cible et plus particulièrement avec ou à proximité des résidus critiques. Un contact est évalué comme persistant s'il atteint de seuil de 90% sur les 8 dernières ns de la trajectoire. L'énergie de liaison théorique est ensuite calculée (MM-PBSA) ce qui permet de classer les molécules entre elles en fonction de cette dernière.

A la suite de cette analyse, nous pouvons dresser un histogramme dont il est facile de visualiser un cluster d'inhibiteurs potentiels, se détachant des autres candidats et constituant un ensemble des meilleures molécules théoriques. Ces meilleures molécules sont commandées et intégrées dans nos tests de validation biophysique et biochimique, dans la mesure de leur disponibilité et de leur solubilité en tampon aqueux. Il a été également jugé utile de commander des molécules en dehors de ces clusters. En effet, ces estimations sont théoriques et nous devons optimiser nos chances de trouver des molécules touches. Ces molécules sont issues d'un enrichissement par criblage virtuel haut-débit et restent toutes des bons candidats potentiels. Ci-dessous est illustré un exemple d'analyse détaillée d'un complexe protéine-ligand stable pour les deux sites ciblés et appartenant au cluster des meilleurs hits théoriques (Figure 58). Les derniers cadres de chacune des trajectoires

stables sont extrait et la structure du complexe correspondant et considérée comme représentative.

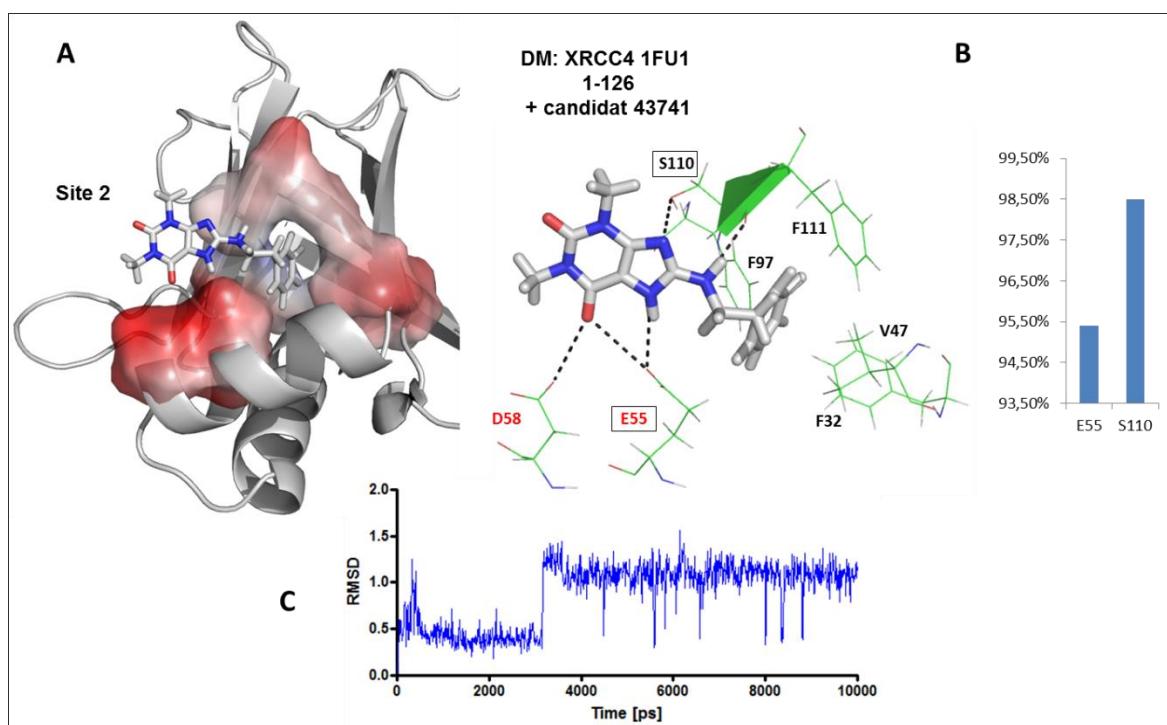
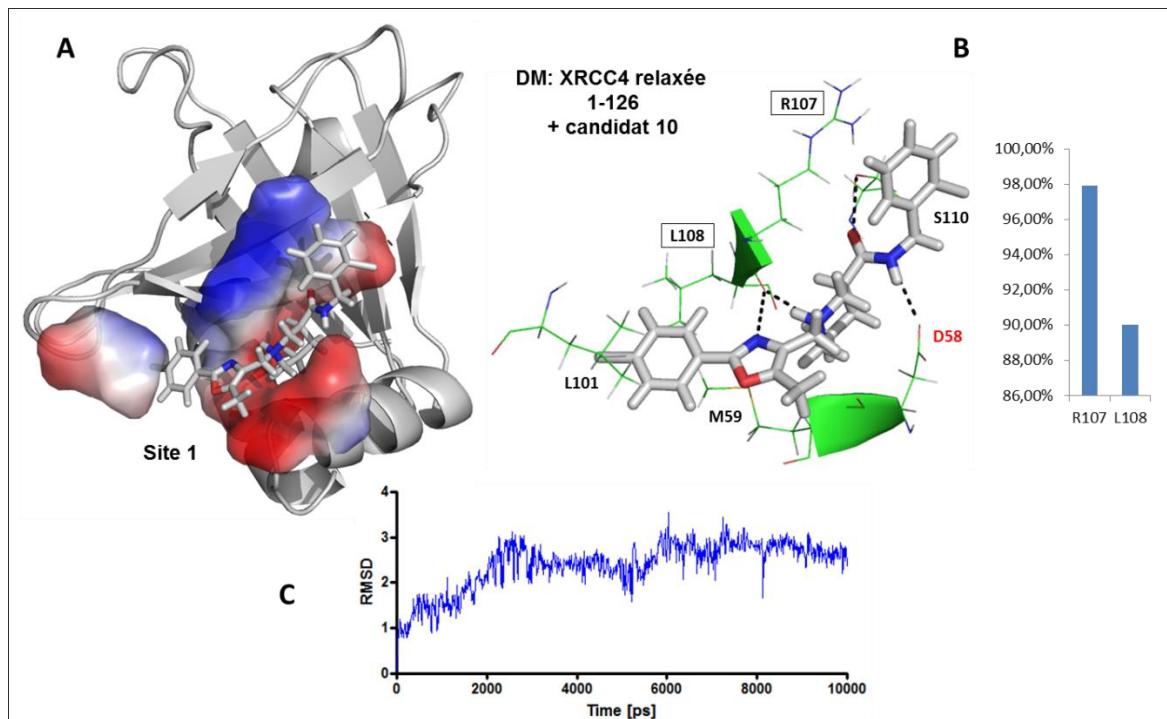


Figure 58: Analyse par DM de la stabilité de deux candidats issus du docking sur les sites 1 et 2 de XRCC4 (1-126).

La stabilité de chacun de ces inhibiteurs est évaluée après analyse du RMSD et des contacts persistants avec la cible protéique. (A) Extraction du dernier cadre de la trajectoire de DM et analyse du mode de liaison au sein des sites 1 et 2 pour ces deux candidats. Les résidus en rouge sont les résidus critiques déterminés par mutagénèse dirigée. Les résidus encadrés en noir sont les résidus avec lesquels le ligand forme des contacts persistants. (B) Histogramme des contacts persistants au-delà de 90% et résidus concernés correspondants. (C) Courbe RMSD du ligand au cours de la trajectoire de dix nanosecondes. Avec des fluctuations $\leq 1\text{\AA}$, les candidats restent très stables et localisés au sein du site concerné.

Cette analyse effectuée sur chacun des complexes XRCC4 hit est résumée dans l'histogramme suivant (Figure 59), reportant l'énergie de liaison et la persistance au site d'intérêt pour chacune des molécules.

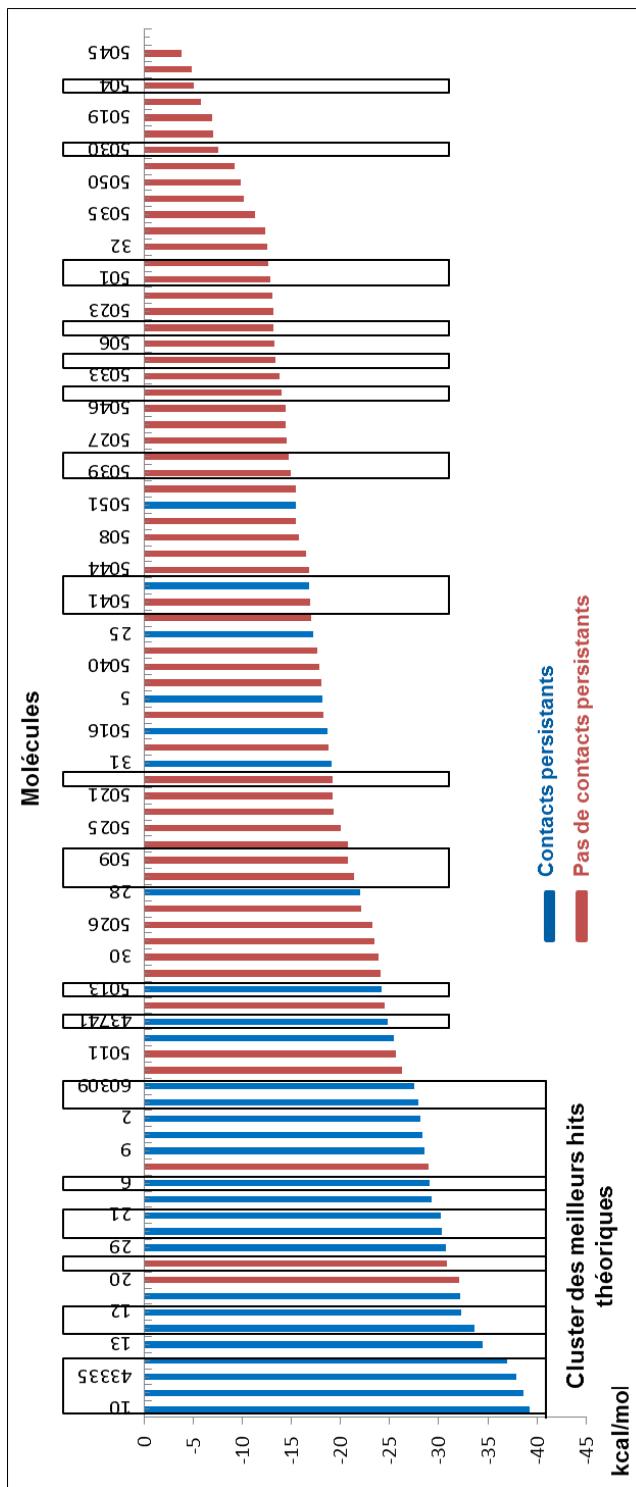


Figure 59: Analyse de l'énergie de liaison théorique après DM pour chaque candidat des sites 1 et 2 de XRCC4 (1-126) – Implémentation par les contacts persistants.

L'énergie de liaison théorique en kcal/mol est évaluée via la trajectoire de DM par le programme MM-PBSA. Les molécules sont classées par ordre croissant d'énergie et un cluster des meilleurs hits théoriques (contacts persistants + faible énergie) peut ainsi être extrait de ces calculs.

De cet histogramme, nous pouvons sortir plusieurs informations. Il est évident que les molécules les plus persistantes (28 au total), ont également en moyenne les meilleures énergies théoriques de liaison. Un cluster d'une vingtaine de molécules regroupe ainsi les meilleurs candidats théoriques pour l'inhibition du complexe XRCC4/Cer-XLF. Le choix de sélectionner également des molécules non persistantes par DM et la non disponibilité commerciale de certaines d'entre elles nous a conduits à commander au total 28 molécules (encadrées en noir sur l'histogramme ci-dessus) afin de poursuivre l'analyse expérimentale. Cette analyse ultérieure comprend la détermination de la solubilité aqueuse de ces ligands, la validation biophysique de leur liaison à XRCC4 et l'analyse de leur effet inhibiteur par des essais biochimiques.

2. *Simulations de DM sur le domaine N-terminal globulaire de Cer-XLF (1-127)*

28 complexes Cer-XLF/hit potentiel, issus des expériences de criblage par docking et identifiés précédemment ont été soumis à une simulation de DM de dix nanosecondes et à une analyse approfondie identique à celle utilisée ci-dessus avec XRCC4. Ci-dessous est illustré un exemple d'analyse détaillée d'un complexe protéine-ligand stable et appartenant au cluster des meilleurs hits théoriques (Figure 60). Les derniers cadres de chacune des trajectoires stables sont extrait et la structure du complexe correspondant et considérée comme représentative.

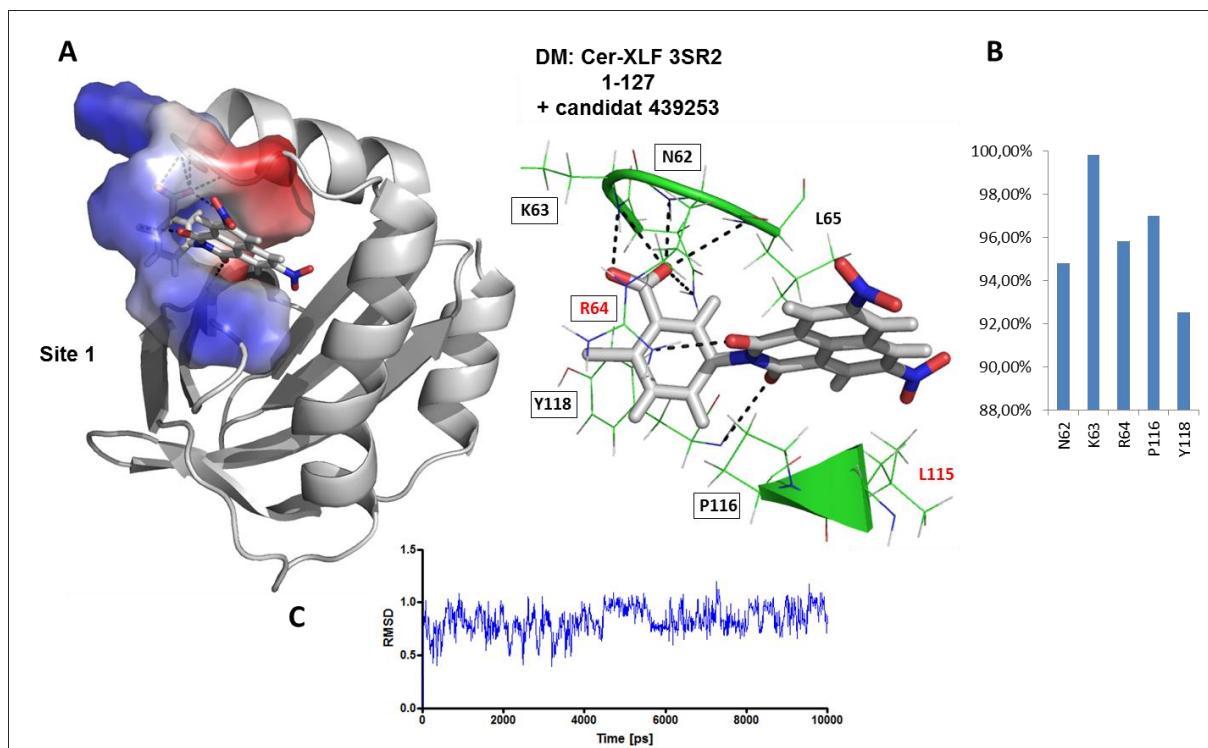


Figure 60: Analyse par DM de la stabilité d'un candidat issu du docking sur le site 1 de Cer-XLF (1-127).

La stabilité de chacun des inhibiteurs est évaluée après analyse du RMSD et des contacts persistants avec la cible protéique. (A) Extraction du dernier cadre de la trajectoire de DM et analyse du mode de liaison au sein du site 1 pour le candidat #439253. Les résidus en rouge sont les résidus critiques déterminés par mutagénèse dirigée. Les résidus encadrés en noir sont les résidus avec lesquels le ligand forme des contacts persistants. (B) Histogramme des contacts persistants au-delà de 90% et résidus concernés correspondants. (C) Courbe RMSD du ligand au cours de la trajectoire de dix nanosecondes. Avec des fluctuations $\leq 1\text{\AA}$, le candidat reste très stable et localisé au sein du site concerné.

L'analyse effectuée sur chacun des complexes Cer-XLF hit est résumée dans l'histogramme suivant (Figure 61), reportant l'énergie de liaison et la persistance au site d'intérêt pour chacune des molécules.

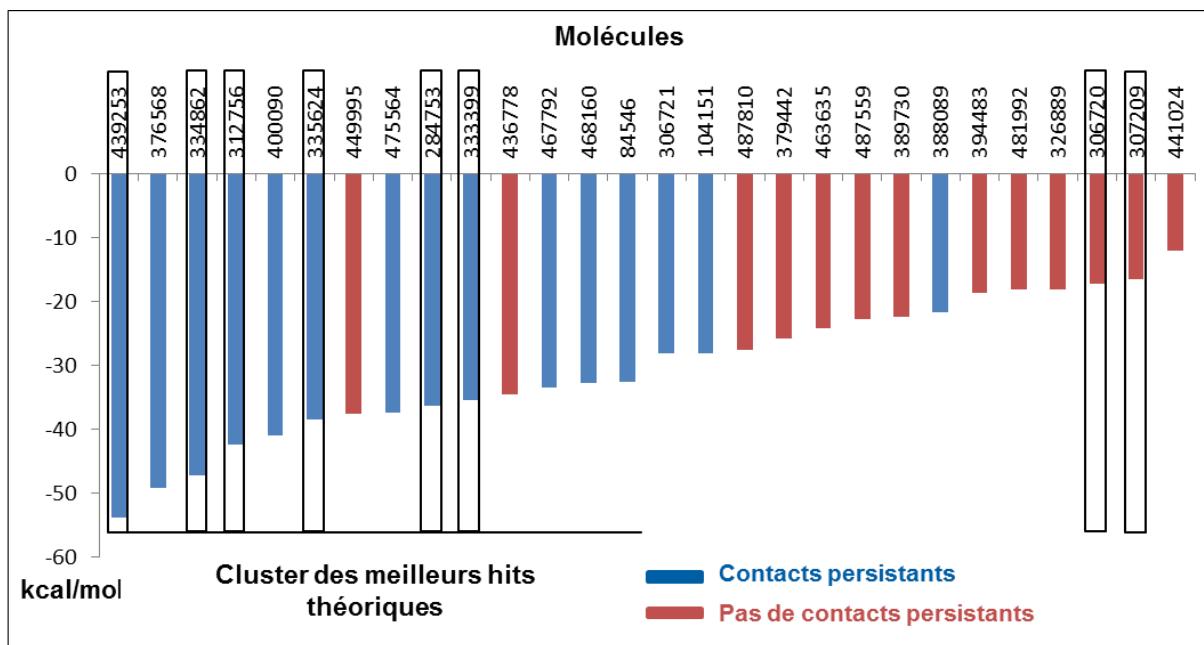


Figure 61: Analyse de l'énergie de liaison théorique après DM pour chaque candidat du site 1 de Cer-XLF (1-127) – Implémentation par les contacts persistants.

L'énergie de liaison théorique en kcal/mol est évaluée via la trajectoire de DM par le programme MM-PBSA. Les molécules sont classées par ordre croissant d'énergie et un cluster des meilleurs hits théoriques (contacts persistants + faible énergie) peut ainsi être extrait de ces calculs.

Là encore, les molécules les plus persistantes (15 au total), ont également en moyenne les meilleures énergies théoriques de liaison. Un cluster d'une dizaine de molécules regroupe ainsi les meilleurs candidats théoriques pour l'inhibition du complexe XRCC4/Cer-XLF. Le choix de sélectionner également des molécules non persistantes par DM et la disponibilité commerciale de chacune d'entre elle nous a permis de commander au total 8 molécules (encadrées en noir sur l'histogramme ci-dessus) pour validation expérimentale.

En conclusion, Un criblage virtuel haut débit par docking et DM nous a permis de constituer une collection de 36 candidats potentiels pour l'inhibition du complexe XRCC4/Cer-XLF. Au cours de ce projet, il a également été question de pouvoir potentiellement alimenter notre criblage virtuel haut-débit par un criblage HTS dont les détails sont explicités dans le paragraphe suivant.

II – 2.7 : Criblage haut-débit : principe, mise en œuvre et résultats obtenus

Les avantages et limites du HTS sont discutés en introduction. Le développement d'un test haut-débit d'interaction entre les protéines XRCC4 et Cer-XLF permet d'identifier rapidement des molécules touchées avec une bonne affinité pour l'une ou l'autre des protéines et, contrairement au criblage virtuel, un effet direct sur l'inhibition de l'interaction entre les deux partenaires est mesuré.

Le test mis en place est un test basé sur une expérience FRET (Förster/Fluorescence Resonance Energy Transfer). Des anticorps ciblant spécifiquement les tags de XRCC4 (anti-MBP) et Cer-XLF (anti-Histidine), sont couplés respectivement à des fluorophores Terbium (D2) et Europium (Eu). Si les deux protéines interagissent, ces fluorophores ont une proximité spatiale inférieure à 50 angströms et l'excitation de l'un d'eux entraîne l'émission de fluorescence du second mesurable à une longueur d'onde spécifique. La mesure de cette énergie émise est la base de notre test et est déterminée après pré-incubation de nos protéines avec des molécules candidates. Ces molécules candidates proviennent d'une chimiothèque n'ayant subi aucun filtrage préalable (Chimiothèque Curie et CNE) ou bien d'une chimiothèque plus ciblée, dont les molécules ont été préalablement filtrées et sélectionnées par des méthodes rationnelles basées sur la structure des protéines (Docking, dynamique moléculaire). Une molécule touche se liant spécifiquement à XRCC4 ou Cer-XLF sur un site essentiel, perturbe l'interaction entre les deux protéines et entraîne donc une perte d'émission de fluorescence (Figure 62). Un contrôle positif était nécessaire pour cette expérience afin de nous apporter un témoin de l'efficacité d'inhibition et un ratio HTRF référent.

Aucune molécule inhibitrice n'est connue pour cette interface. Le témoin positif utilisé a donc été l'un des deux partenaires protéique mais sans étiquette histidine ou MBP.

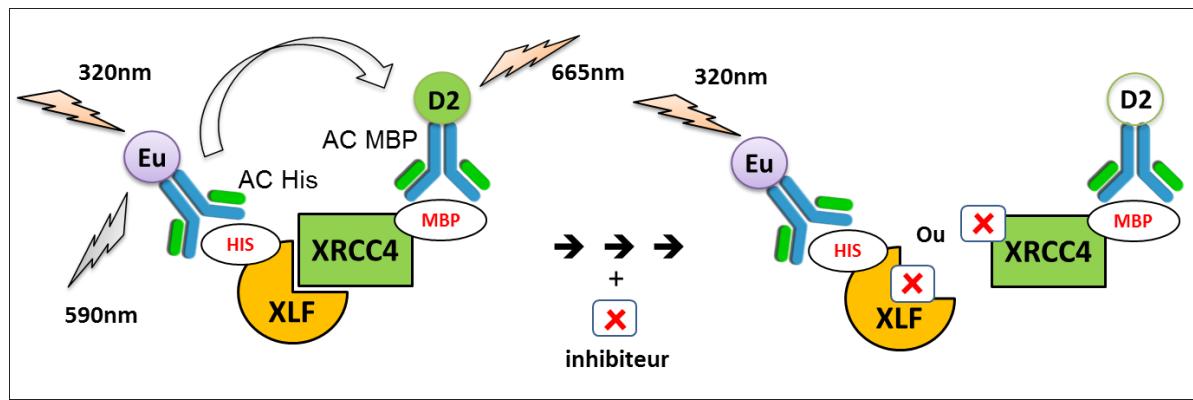


Figure 62: Schéma du test HTS fonctionnel, mis au point pour la découverte d'inhibiteurs de l'interaction XRCC4/Cer-XLF (ETaC-CNRS, Toulouse ; Coll. Frédéric Ausseil).

Ce test fonctionnel est basé sur une expérience de FRET. Le transfert d'énergie s'effectue via le couple terbium (D2)/Europium (Eu), fluorophores couplés à des anticorps spécifiques des tags de chaque protéine. L'inhibition de la fixation des deux partenaires doit se traduire par une diminution du signal à 665nm et pas de variation (voire une légère augmentation) du signal à 590nm. Le pourcentage d'inhibition est calculé avec le ratio HTRF = $((665nm/590nm)*10000)$.

A l'issu de ce test, les molécules touchées devront être confirmées dans un second test indépendant, puis validées comme étant des vrais positifs par des méthodes complémentaires. La RMN nous permet de valider l'interaction des molécules touchées avec nos protéines et des tests d'activités valideront finalement leur effet inhibiteur. Inversement, des molécules touchées trouvées par criblage virtuel et validées par RMN et dans nos tests d'activité biochimiques pourront être intégrées à ce test d'interaction afin de confirmer ou d'inflimer l'activité inhibitrice de ces candidats.

Un résultat positif à l'issu de ce cible HTS nous permet de supposer que la molécule candidate a, d'une part, une affinité micromolaire pour l'un des deux partenaires et peut donc entrer en compétition avec l'autre protéine, et d'autre part, qu'elle se lie au niveau des têtes globulaire des protéines car l'interaction s'effectue via ces domaines. Cependant ces hypothèses doivent être vérifiées expérimentalement et par des méthodes complémentaires. De même, les molécules touchées se lient-elles sur une protéine ou bien les deux ? sur un site spécifique ou bien plusieurs ? se lient-elles à l'interface entre les deux protéines ou ont elles un effet allostérique ? ... Toutes ces questions doivent être traitées afin de valider ou bien d'inflimer l'intérêt

certain d'une touche et avant d'aller plus loin dans la structure des complexes cible-ligand et l'optimisation chimique.

Finalement, 3 molécules touchées ont été déterminées par ce test comme étant des inhibiteurs potentiels de l'interaction XRCC4/Cer-XLF. Mais après une seconde analyse, il s'est avéré que ces trois candidats n'étaient finalement que des faux positifs (molécules interférentes). En conclusion et à l'issue de cette campagne de criblage haut-débit, 10160 molécules ont été testées et aucune molécule touchée n'a été identifiée comme vrai positif, avec un effet direct et confirmé sur l'inhibition de l'interaction entre XRCC4 et Cer-XLF.

II – 2.8 : Analyse par DLS et RMN ^1H de la solubilité et intégrité des molécules retenues

Dans un processus de recherche et mise au point de candidats médicaments, un des problèmes majeurs rencontré est la solubilité des molécules. La solubilité (mesurée en mol/L) est la quantité maximale de matière que l'on peut dissoudre dans 1Litre d'eau. Elle est dite intrinsèque pour un composé neutre et apparente pour une forme ionisable à un certain pH. La conversion solubilité intrinsèque/ionisable s'effectue grâce à la fameuse équation d'Henderson-Hasselbach ($\text{pH} = pK_a + \log [\text{A}^-] / [\text{HA}]$).

Parmi les différentes étapes nécessaires à la découverte d'une nouvelle molécule bioactive, l'étape première de criblage est tout particulièrement concernée par cette notion de solubilité. Les essais ultérieurs de validation des effets *in vitro* sont fortement dépendants de cette dernière. Ce facteur est habituellement inclus lors d'une étape de filtrage d'une base de données virtuelle de molécules avant l'étape de criblage (LogP, SFI...). Cependant, les algorithmes de prédition de la solubilité en milieu aqueux d'une molécule organique sont nombreux et donnent parfois des résultats très différents, voire faux. Il est important alors, de confronter au moins deux filtres afin d'avoir en moyenne une bonne prédition de la solubilité de nos produits. Les molécules chargées sont généralement plus solubles dans l'eau que les molécules neutres (solubilité dépendante du pH).

L'insolubilité et l'agrégation des molécules testées sont des problèmes majeurs car ils causent des faux positifs dans les tests biophysiques de liaison (expérience de WaterLogsy en RMN par exemple) et les tests d'activité biologiques. Ce phénomène

d'agrégation est aussi souvent la raison de faux positifs dans les tests HTS. Pour pallier ce problème, il est impératif d'être extrêmement rigoureux dans les phases initiales de détermination des concentrations solubles maximales et dans des conditions précises de tampon et de température. Ces conditions pourront être alors intégrées aux différents tests d'interaction et d'activité. De plus, déterminer une limite de solubilité d'un composé dans un tampon donné permet par la suite d'éviter des problèmes de précipitation ou d'agrégation lors d'une étape de cocrystallisation, ou de trempage de cristaux de protéines dans une solution d'inhibiteur. Des méthodes de criblage biophysiques, par exemple par RMN ou ITC, permettent rapidement de caractériser un faux d'un vrai positif (titration, obtention de constantes d'affinités), mais ces techniques sont souvent longues et laborieuses et bien qu'indispensables à l'étape de validation des molécules touchées, elles ne permettent que du criblage bas à moyen débit.

Dans notre cas, nous avons effectué un criblage virtuel haut débit. Ce type de recherche est sujet à de nombreux faux-positifs (et également faux-négatifs) mais c'est un filtre rapide et efficace qui nous a permis d'obtenir une chimiothèque de 36 molécules. Ces molécules sont toutes des inhibiteurs potentiels de l'interaction entre XRCC4 et Cer-XLF. Une vérification rigoureuse de la solubilité de chacune a été réalisée par des techniques orthogonales (RMN, DLS) et afin d'éviter les faux positifs dans nos tests d'activité ultérieurs. La DLS nous a permis d'écartier les composés qui étaient susceptibles de former des agrégats (les « promiscuous compounds ») à partir d'une certaine concentration ou même des auto-assemblages volumineux de molécules. Ces entités sont facilement observables sous une binoculaire et ont été confirmés de cette manière. La RMN nous a permis d'évaluer la concentration soluble maximale dans un tampon aqueux pour une concentration donnée calculée ainsi que de vérifier la pureté et l'intégrité des molécules commandées (présence de solvants, produits de dégradation...). A réception des molécules, les différentes étapes d'analyse effectuées ont été schématisées dans la figure ci-dessous (Figure 63).

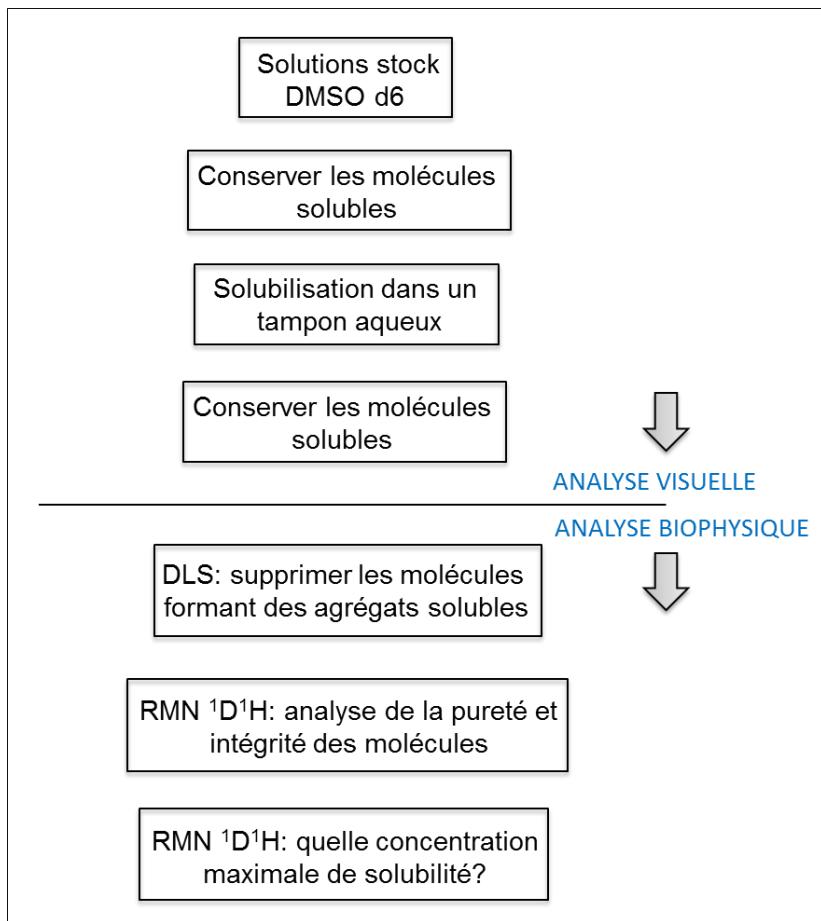


Figure 63: Schéma représentatif des différentes étapes utilisées pour la détermination de la solubilité et intégrité des molécules touchées issues de l'étape de criblage.

Une première analyse visuelle consiste à préparer des solutions stock et concentrées à 100mM de chacune des molécules dans du dimethyl sulfoxyde (DMSO) deutéré afin d'écartier les molécules formant des agrégats très nettement visibles. L'analyse d'un signal STD positif par RMN et les tests d'activité biologiques pour nos molécules se font à une concentration n'excédant pas 1mM (dans le cas d'une titration cependant, nous augmentons la concentration jusqu'à obtention d'un phénomène de saturation). Les molécules restantes sont donc solubilisées à une concentration finale de 1mM dans un tampon aqueux (tampon dit « tampon RMN standard » : 20mM phosphate ; 50mM NaCl ; 100% D2O et pD 6.8 – concentration finale en DMSO : 1%). Après équilibration 12h à température ambiante et sous légère agitation, les molécules formant des agrégats visibles (binoculaire) sont là encore écartées.

Une analyse biophysique ultérieure par DLS puis RMN du proton permet d'écartier les molécules formant des agrégats solubles et invisibles à l'œil nu et de déterminer l'intégrité, l'attribution et la concentration maximale de solubilité de ces dernières pour 1mM calculé, à pH 7.2 (pD 6.8) et à 298K. Cette concentration maximale de solubilité dans 1mM de tampon aqueux est critique puisqu'elle détermine la concentration limite à ne pas dépasser lors de nos tests d'interaction et d'activité afin d'éviter les faux-positifs. Les 36 molécules issues du crible virtuel contre XRCC4 et Cer-XLF et leur état de solubilité sont résumées dans le tableau ci-dessous (Figure 64).

Ultérieurement, la concentration soluble maximale de nos molécules sera déterminée (avec cette fois une gamme de 0.5 à 20mM) dans un tampon de cristallisation et des conditions de pH et température bien précises afin d'anticiper l'étape de cristallisation des complexes protéine-ligand. Le but étant de mettre la protéine en présence d'une solution la plus concentrée possible en ligand avec le moins possible de DMSO, tout en évitant la précipitation de ce dernier.

Molecule	Fournisseur	PM (Da)	SFI	Ag DMSO	Sol pH7.2	DLS	Protéine
5186933	Chembridge	456,37	-0,94	NO	500µM	S	Cer-XLF
5233340	Chembridge	415,36	0,32	YES	-	-	Cer-XLF
7926063	Chembridge	338,32	0,46	NO	500µM	S	Cer-XLF
3S-77617	IBS	283,3	0,97	NO	1mM	S	XRCC4
7949364	Chembridge	441	2,22	NO	1mM	S	XRCC4
6737930	Chembridge	412,47	2,3	YES	-	-	Cer-XLF
6821224	Chembridge	429	2,66	NO	200µM	S	XRCC4
8876638	Chembridge	441,74	3,28	NO	500µM	S	Cer-XLF
7908462	Chembridge	443,53	3,46	NO	NO	-	Cer-XLF
5545659	Chembridge	397,25	3,49	NO	500µM	S	Cer-XLF
5665101	Chembridge	460	3,75	NO	NO	-	XRCC4
5568011	Chembridge	371	3,82	YES	-	-	XRCC4
6435411	Chembridge	468	3,83	NO	1mM	S	XRCC4
4S-62686	IBS	281,3	3,98	NO	1mM	S	XRCC4
6563488	Chembridge	465	4,03	NO	NO	-	XRCC4
3S-91582	IBS	299,3	4,37	YES	-	-	XRCC4
53464	IBS	465,9	4,47	NO	1mM	S	XRCC4
7468661	Chembridge	480	4,5	NO	500µM	S	XRCC4
42527	IBS	433,4	4,55	NO	1mM	S	XRCC4
C660-0154	ChemDiv	459	4,58	YES	-	-	XRCC4
7691166	Chembridge	452	4,69	NO	NO	-	XRCC4
51686	IBS	453,9	4,7	NO	NO	-	XRCC4
7800166	Chembridge	397,39	4,75	NO	NO	-	Cer-XLF
51330	IBS	447,5	4,86	YES	-	-	XRCC4
7681-0150	ChemDiv	434	5,33	YES	-	-	XRCC4
3343-3136	ChemDiv	472	5,79	YES	-	-	XRCC4
C522-3902	ChemDiv	433	5,82	NO	200µM	S	XRCC4
57632	IBS	488,5	5,9	NO	1mM	S	XRCC4
6256994	Chembridge	529	5,97	NO	NO	-	XRCC4
41376	IBS	486,5	6,02	NO	1mM	S	XRCC4
38618	IBS	402,4	6,03	YES	-	-	XRCC4
7691275	Chembridge	465	6,33	NO	NO	-	XRCC4
7413418	Chembridge	436	6,44	NO	1mM	S	XRCC4
C522-5329	ChemDiv	437,5	6,57	NO	NO	-	XRCC4
C522-4944	ChemDiv	437,5	6,58	YES	-	-	XRCC4
C522-3745	ChemDiv	422,5	6,69	NO	NO	-	XRCC4

Figure 64: Tableau représentant les 36 molécules issues du crible virtuel dirigé contre XRCC4 et Cer-XLF, ainsi que leur solubilité théorique et expérimentale respective.

L'analyse visuelle de l'agrégation dans le DMSO est évaluée pour une concentration stock de 100mM. Les mesures RMN de solubilité en tampon aqueux sont effectuées dans un tampon RMN à pH7.2 et à 298K, pour une concentration calculée de 1mM. Les mesures de DLS sont réalisées dans ce même tampon RMN et à la concentration maximale déterminée (S : Soluble ; Ag : Agrégat ; Pr : Précipité).

Le facteur SFI (« Solubility Forecast Index ») d'une molécule donnée (Hill and Young, 2010) est calculé selon la formule : $SFI = \text{clogD}_{\text{pH7.4}} + \text{nombre de cycles aromatiques}$

(avec $c\log D$ étant le coefficient de distribution « calculé » à un pH donné selon la formule : $c\log D = \log P - \log (1+10^{(pH-pK_a)})$). Ce facteur rend compte du caractère lipophile d'un composé et prédit qu'une molécule ayant une valeur inférieure à 5 a des chances raisonnables d'avoir une solubilité aqueuse supérieure ou égale à $200\mu M$, ce qui est favorable à une bonne biodisponibilité. A titre d'exemple, une drogue que l'on administre par voie orale possède un SFI qui se situe en moyenne autour de 2.4.

Cette valeur prédictive nous permet, tout comme le LogP mentionné précédemment, d'anticiper les problèmes de solubilité pour chaque molécule. Il est intéressant de dresser un histogramme de prédiction de solubilité par le SFI pour chacune des molécules comparativement à leur solubilité aqueuse expérimentale (figure 65). Ce comparatif permettra par la suite (lors d'un prochain criblage) de prévenir peut-être de façon plus efficace un nombre trop important de molécules insolubles (en prenant bien sûr en compte le pourcentage d'erreur moyen apporté par ce facteur).

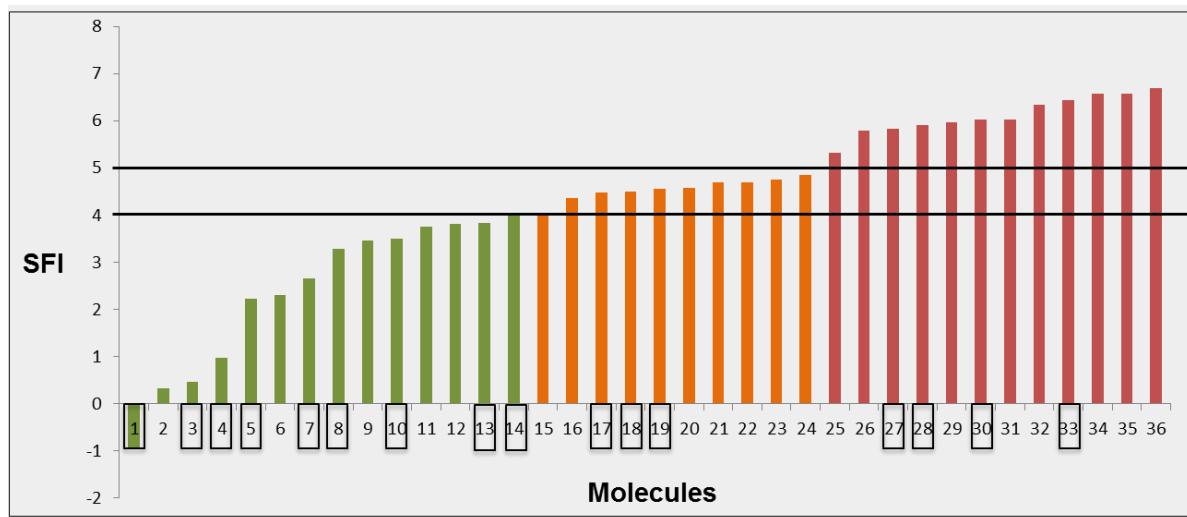


Figure 65: Histogramme de prédiction de solubilité par le SFI pour les 36 molécules issues du crible virtuel dirigé contre XRCC4 et Cer-XLF.

SFI ≤ 4 : en vert, bonne solubilité prédictée ; SFI ≤ 5 : en orange, solubilité moyenne prédictée ; SFI > 5 : en rouge, mauvaise solubilité prédictée. Les molécules solubles expérimentalement en tampon RMN à pH7.2 et 298K (de $200\mu M$ à 1mM pour 1 mM calculée) sont encadrées en noir.

L'histogramme ci-dessus représente parmi nos 36 candidats, les molécules prédictées comme ayant une bonne solubilité (en vert, seuil de SFI fixé à 4), une solubilité moyenne (en orange, seuil de SFI fixé à 5) et une mauvaise solubilité (en rouge, SFI

au-delà de 5). Les molécules encadrées sont celles dont la solubilité en tampon aqueux a été démontrée expérimentalement par RMN du proton avec une concentration soluble calculée allant de 200 μ M à 1mM environ (calcul effectué par intégration de l'aire des pics comparativement à un étalon interne : le DSS). Environ 2/3 des molécules dont le SFI est ≤ 4 sont effectivement solubles (64%). Ce pourcentage tombe à 50% si le seuil de SFI est fixé à 5. On note ici une prédiction de solubilité qui reste raisonnable en moyenne même si cette prédiction est sujette à une part importante de faux-positifs. Néanmoins, il est également important de tester les molécules ayant une mauvaise solubilité théorique car certaines d'entre elles sont expérimentalement solubles d'après ce que l'on peut voir sur cet histogramme. Lors d'une prochaine étape de filtrage d'une base de données virtuelle de molécules, un facteur SFI fixé à 2 serait un bon compromis (voir chapitre sur l'inhibition entre XRCC4 et le C-terminal de la ligase 4).

En conclusion et d'après notre analyse, 12 molécules dirigées contre XRCC4 sont solubles en tampon aqueux dans les conditions expérimentales testées (avec des concentrations allant de 200 μ M à 1mM pour 1mM calculé) ainsi que 4 molécules dirigées contre Cer-XLF (à 500 μ M pour 1mM calculé). Ces 16 molécules sont représentées ci-dessous (Figures 66-69). Ces concentrations seront prises en compte dans les étapes suivantes de validation des interactions cible-ligands et d'activité biologique des composés positifs (hits).

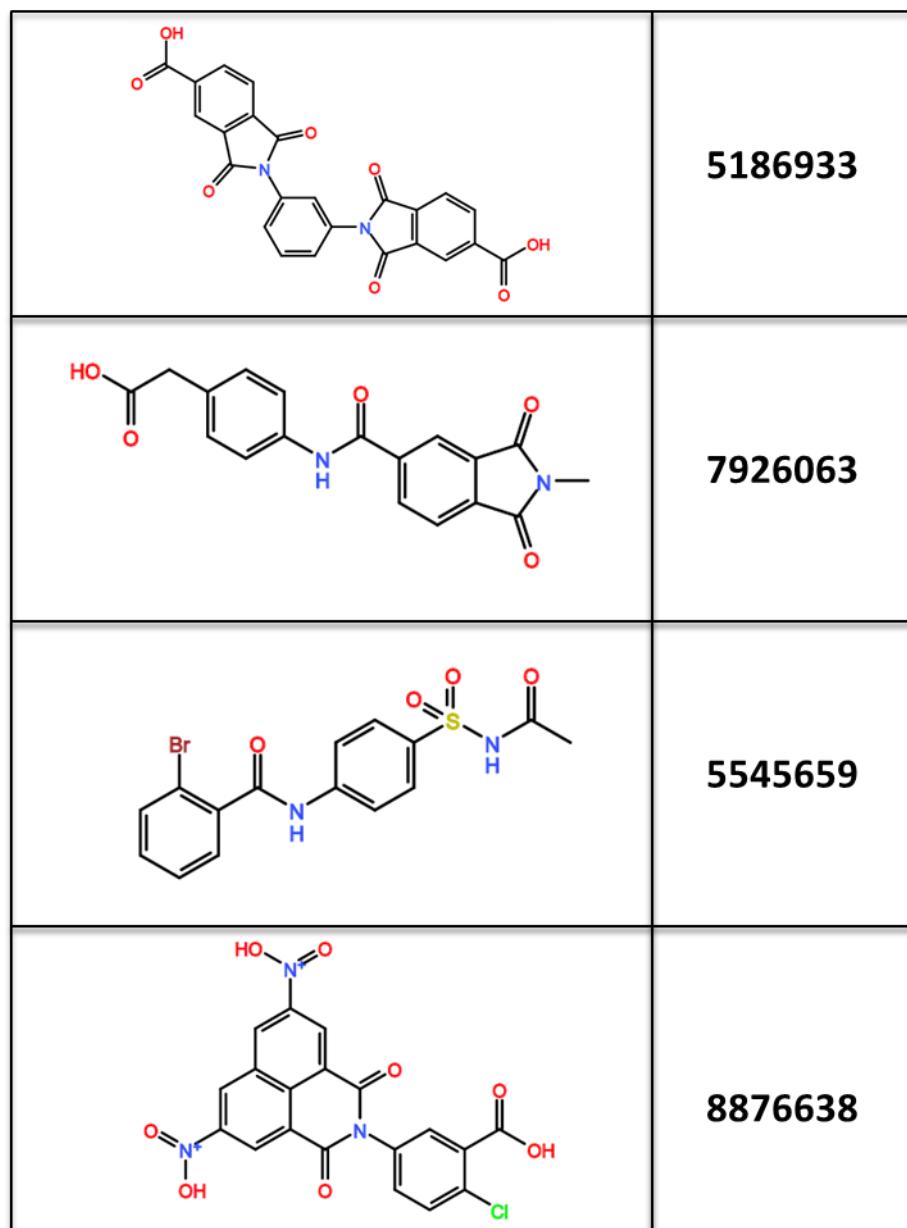


Figure 66: Tableau 1/4 représentant les molécules solubles issues du crible virtuel dirigé contre XRCC4 et Cer-XLF.

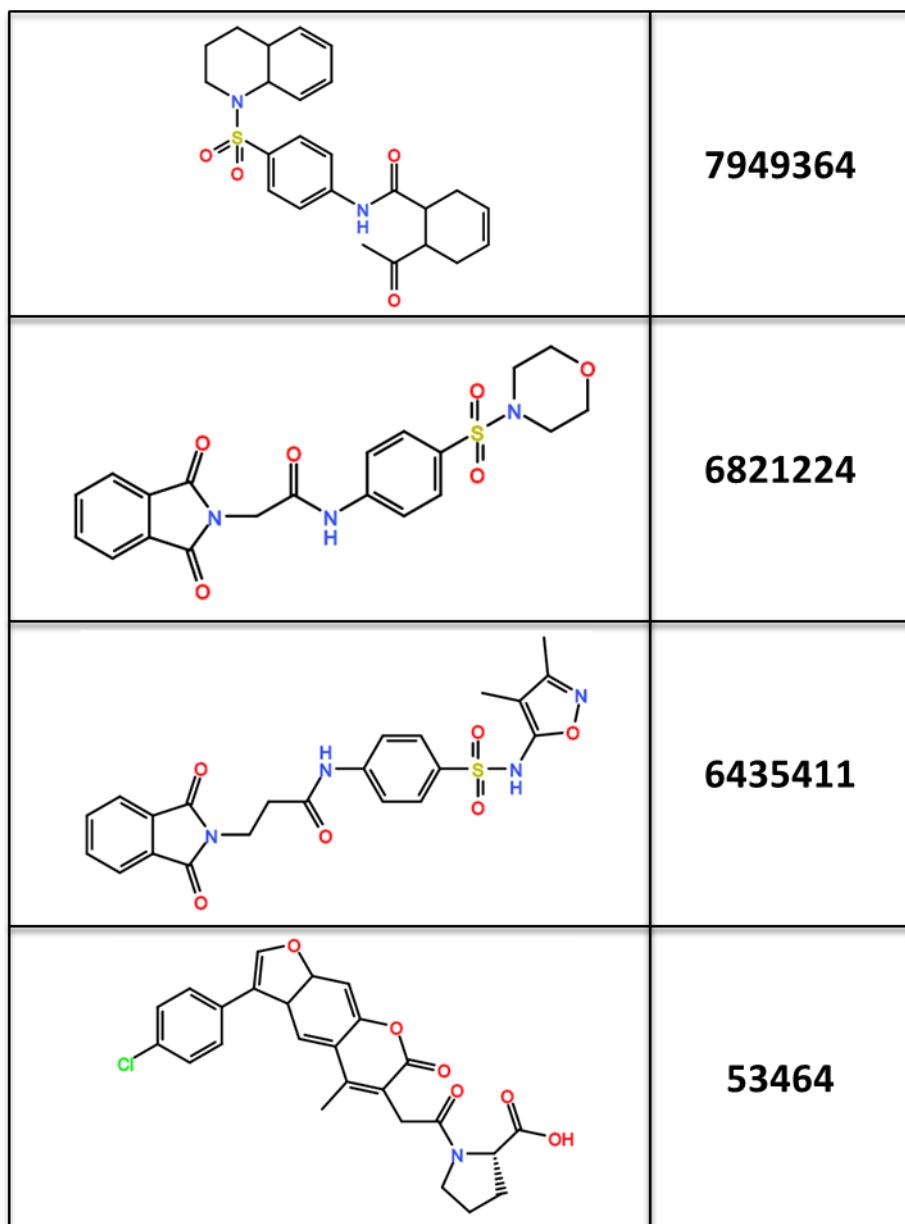


Figure 67: Tableau 2/4 représentant les molécules solubles issues du crible virtuel dirigé contre XRCC4 et Cer-XLF.

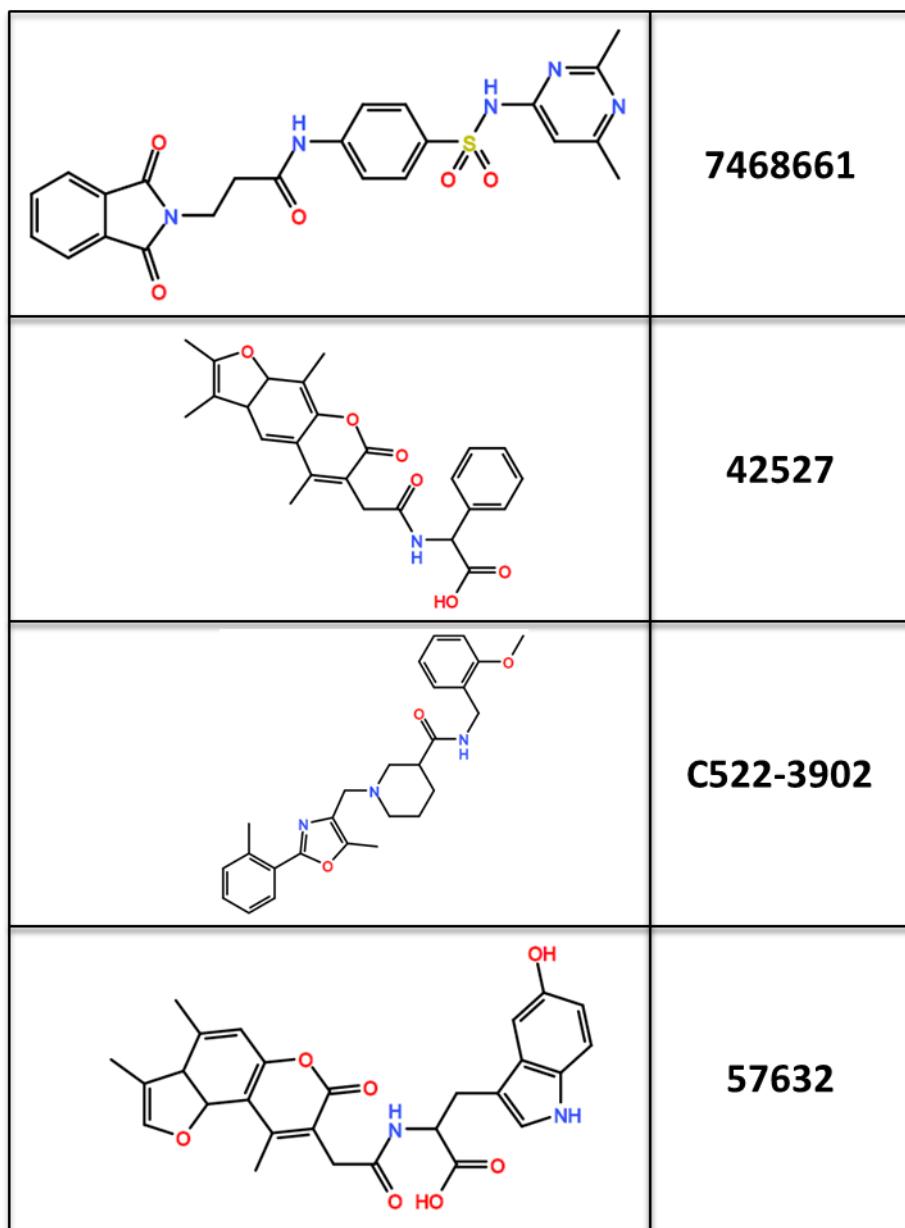


Figure 68: Tableau 3/4 représentant les molécules solubles issues du crible virtuel dirigé contre XRCC4 et Cer-XLF.

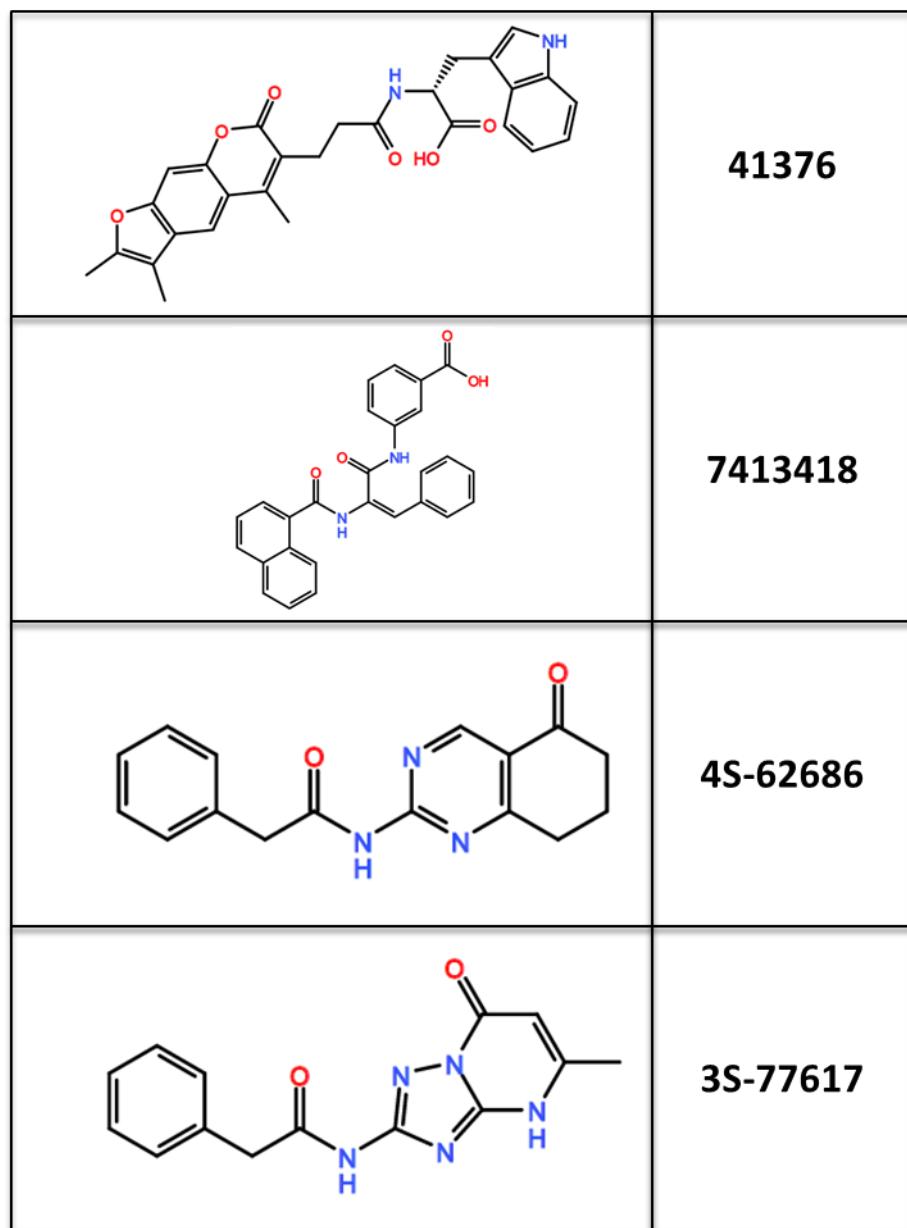


Figure 69: Tableau 4/4 représentant les molécules solubles issues du crible virtuel dirigé contre XRCC4 et Cer-XLF.

II – 2.9 : Validation biophysique des interactions protéine-ligand par DSF et STD-NMR

L'étape de validation biophysique des interactions protéine-ligand est primordiale. Elle permet, en amont d'un test biologique adapté, de déterminer si les composés sélectionnés interagissent de façon spécifique avec leur cible protéique et si tel est le cas, avec quelle affinité (K_d), quel mode de liaison (liaison multiple, coopérative...), quelles parties de la molécule («Group Epitope Mapping » ou GEM) (Mayer and Meyer, 2001) et à quel endroit sur la protéine (Cala et al., 2014). La localisation du ligand sur la protéine est notamment possible dans une stratégie de marquage et attribution RMN de ladite cible protéique par exemple (Petros et al., 2006). Si le marquage n'est pas réalisable ou que la taille de la cible est trop importante, une localisation indirecte par ajout d'un compétiteur de haute affinité pour le site ciblé est encore possible (Wang et al., 2004b). La molécule compétitrice, de plus forte affinité, déplacera notre molécule candidate du site de liaison si celui-ci est le même. Malheureusement, aucune molécule inhibitrice n'est connue pour l'interaction entre XRCC4 et Cer-XLF. Si tel était le cas, une stratégie de criblage basée sur le ligand aurait également été envisagée (pharmacophore, relation structure-activité).

Dans notre cas, un marquage ou la résolution cristallographique de la structure des complexes devra être envisagé afin de localiser et optimiser chimiquement les molécules inhibitrices retenues. L'étape de validation biophysique de l'interaction permet également de déterminer à l'issu du test HTS quelle protéine XRCC4 ou bien Cer-XLF est impliquée de façon spécifique dans la liaison de la molécule inhibitrice. En terme de sensibilité, les tests de liaison basés sur la fluorescence sont particulièrement bien adaptés (expérience d'anisotropie de fluorescence par exemple) mais nécessite la dérivation chimique ou la présence d'un groupement ou d'une molécule intrinsèquement fluorescente. Cependant, dans le cas d'une modification, cela pourrait alors entraîner une diminution de l'affinité pour la cible. De façon récurrente, la DSF (« Differential Scanning Fluorimetry ») est utilisée en tant qu'outil de première ligne pour une validation moyen-débit d'interactants moléculaires (Pfaff et al., 2015) mais cette technique est indirecte et les résultats obtenus sont parfois difficilement interprétables. La RMN est quant à elle, une méthode robuste et parfaitement adaptée pour la caractérisation des interactions protéine-ligand, bien

que moins sensible que des techniques basées sur la fluorescence. En conditions de fort excès en ligand par rapport à la protéine, une expérience RMN basée sur l'observation du signal de ces ligands (STD) a été utilisée à cet effet. Les principes, avantages et inconvénients de cette technique sont détaillés en introduction et ne seront pas rappelés ici. Afin de valider l'interaction de nos inhibiteurs potentiels dirigés contre XRCC4 ou Cer-XLF, la démarche expérimentale utilisée a été la suivante :

- Une expérience DSF est d'abord réalisée à un ratio protéine : ligand de 1 : 400 et avec une concentration en ligand de 2mM.
- Pour les molécules positives en DSF, une expérience STD est réalisée pour confirmation et à un ratio protéine : ligand de 1 :100, avec une concentration en ligand de 500 μ M dans un tampon RMN décrit précédemment et à 298K (les protéines ont été préalablement dialysées dans ce tampon RMN).
- La même expérience STD est reproduite en contrôle, dans les mêmes conditions mais sans protéine (Vérification de l'effet induit par une liaison à la cible, suppression du signal background).
- Si une réponse STD est positive les ligands retenus sont titrés, toujours par STD, afin de déterminer si la liaison à la protéine est spécifique et d'estimer une affinité de liaison pour cette dernière (Kd). Lors de cette titration, on peut supposer que la solubilité du ligand est augmentée par l'effet de la liaison à la protéine. Nous tentons ainsi de déterminer un phénomène de saturation et de trouver l'équation adaptée à cette courbe afin d'en extraire le Kd.

Les molécules issues d'un crible virtuel basé sur la structure comme celui effectué précédemment, sont sujettes à de nombreux faux-positifs. Les vrais-positifs quant à eux, ont une affinité de l'ordre micro à millimolaire. Cette gamme d'affinité (échange rapide) est parfaitement adaptée à la technique STD utilisée. Ainsi, une molécule possédant une affinité trop forte (ordre du nanomolaire) ou bien trop faible (dizaine de millimolaire) pour sa cible ne sera pas détectée. Un autre avantage de cette technique est qu'elle ne nécessite aucun marquage de la protéine et permet de travailler en condition de faible quantité de cette dernière (autour de 5 μ M dans un volume de 170 μ L).

Au cours de l'expérience de DSF (protocole détaillé en partie Matériels et Méthodes), le Tm de la protéine XRCC4 est évalué en présence de 2mM de chacune des molécules solubles (ou d'une concentration équivalente en DMSO pour la condition contrôle). La protéine XRCC4 possède une température de fusion autour de 57°C dans son tampon de conservation à l'issue de la purification et en présence de DMSO. On observe à l'issue de l'expérience que sur les dix molécules testées, trois molécules (#53464, #42527 et #41376) sont des molécules interférentes (rendant ainsi la courbe de fluorescence du SYPRO inexploitable, deux molécules n'ont pas d'effet significatif sur le Tm de XRCC4 (#6435411 et #7468661), deux molécules ont un effet déstabilisateur trop important sur le Tm de XRCC4 (#7949364 et #7413418), signe d'une précipitation potentielle de la protéine induite par le ligand et enfin trois molécules (#3S-77617, #4S-62686 et #57632) induisent un shift de -0,5 à -1.5°C par rapport au contrôle DMSO (Figure 70). L'analyse de ces résultats nous permet de prédire que les trois candidats dernièrement cités sont potentiellement les plus intéressants et spécifiques pour la cible.

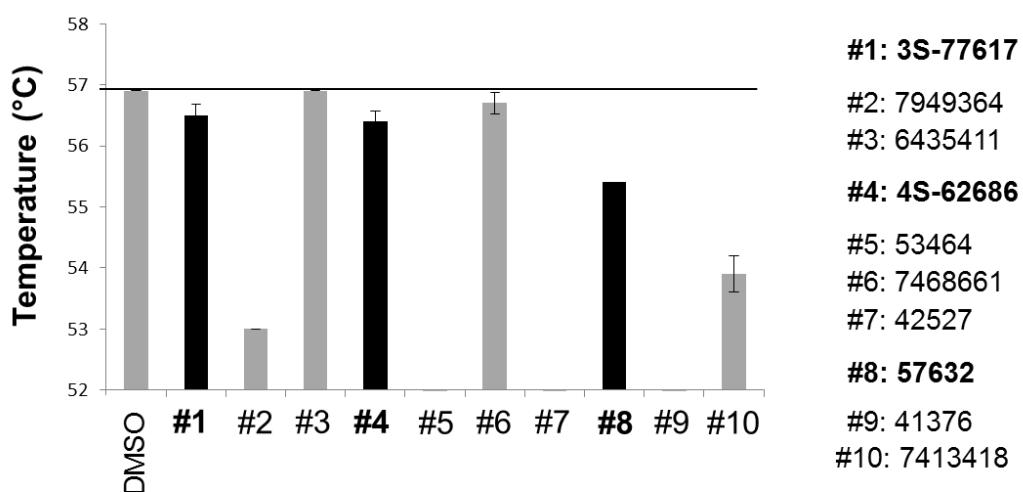


Figure 70: Analyse biophysique de la stabilité thermique par DSF de XRCC4 seul et en présence des candidats inhibiteurs sélectionnés. La dénaturation est suivie grâce au SYPRO Orange (10X) et entre 20 et 89°C. Les valeurs de Tm ont été estimées comme la moyenne d'une expérience réalisée en triplicat (Moy. \pm DS, n=3). Les bâtons colorés en noirs représentent les molécules induisant une variation de Tm significative.

Afin de confirmer ces résultats, une expérience de STD-NMR a été réalisée avec ces molécules. Au cours de cette expérience, les fréquences de on et off-résonance ont été fixées respectivement à -1 et 30 ppm. Un pulse Gaussien de 50ms et 86Hz a été

utilisé. Ci-dessous, les spectres STD des deux molécules positives et spécifiques pour la liaison à XRCC4 sont représentés (Figures 71-72). Un exemple de titration montrant une liaison non spécifique à la cible est également illustré (Figure 73).

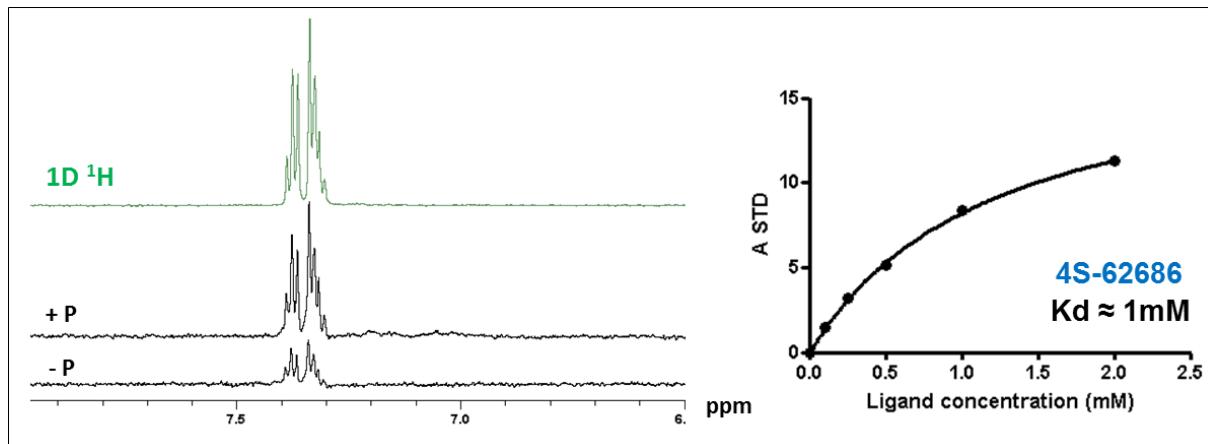


Figure 71: Analyse d'interaction et titration STD du composé #4S-62686 sur la protéine XRCC4. Le spectre en vert est le spectre de la molécule seule à $500\mu\text{M}$, réalisée dans un tampon RMN. Les spectres en noir notés « +P » et « -P » sont respectivement les expériences STD sur le mélange protéine-molécule avec et sans protéine XRCC4 à $5\mu\text{M}$.

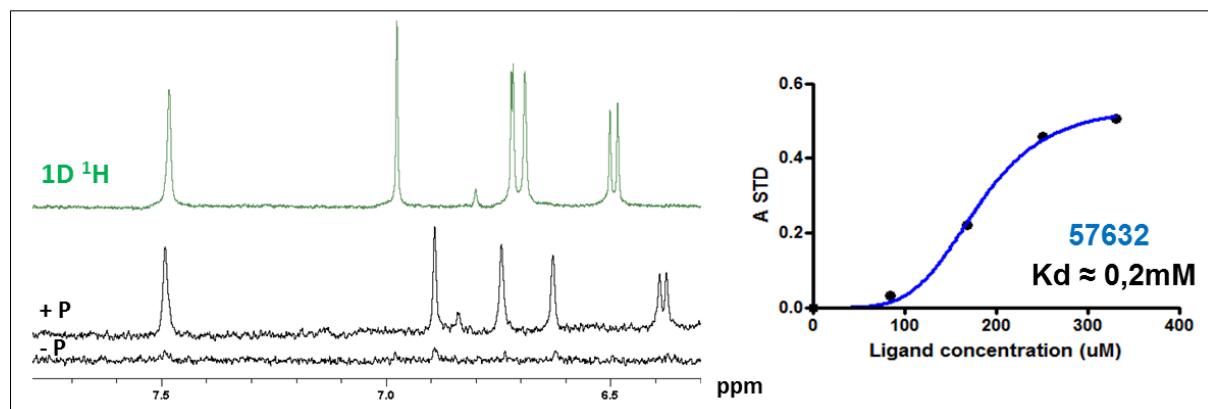


Figure 72: Analyse d'interaction et titration STD du composé #57632 sur la protéine XRCC4. Le spectre en vert est le spectre de la molécule seule à $500\mu\text{M}$, réalisée dans un tampon RMN. Les spectres en noir notés « +P » et « -P » sont respectivement les expériences STD sur le mélange protéine-molécule avec et sans protéine XRCC4 à $5\mu\text{M}$. Les déplacements chimiques observés sont dus à des changements dans le tampon.

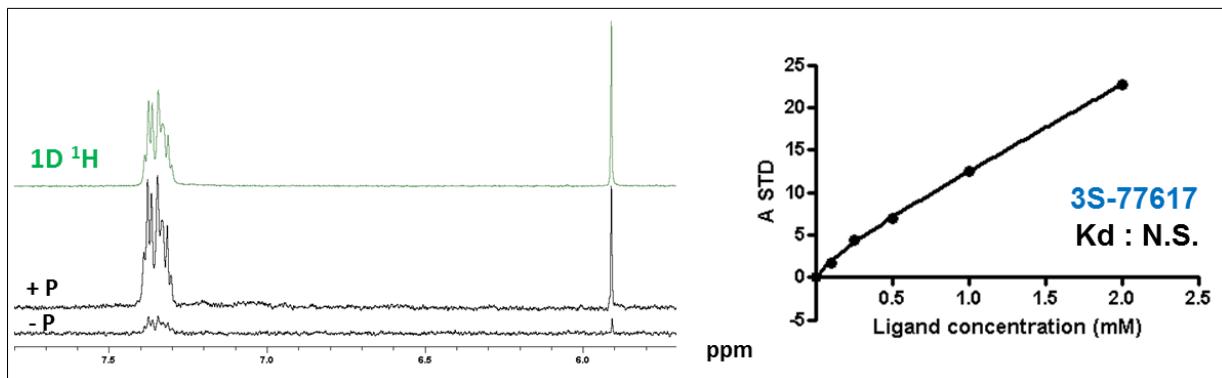


Figure 73: Analyse d'interaction et titration STD du composé #3S-77617 sur la protéine XRCC4. Le spectre en vert est le spectre de la molécule seule à $500\mu\text{M}$, réalisée dans un tampon RMN. Les spectres en noir notés « +P » et « -P » sont respectivement les expériences STD sur le mélange protéine-molécule avec et sans protéine XRCC4 à $5\mu\text{M}$.

La plupart des molécules testées ont montré un signal STD positif mais avec une liaison non spécifique à la cible (courbe linéaire – exemple ci-dessus avec la molécule #3S-77617). Seules deux molécules : #4S-62686 et #57632 ont finalement été retenues à la suite de cette analyse comme étant des molécules touchées spécifiques de la protéine XRCC4 et avec des K_d respectifs autour de 1 et 0.2mM . Ce qui est intéressant, c'est que les trois molécules interférentes en DSF n'ont donné aucun signal STD. A l'inverse toutes les autres molécules ont donné un signal STD positif mais seules deux molécules des trois prédictes comme intéressante pour la cible sont spécifique après titration (Figure 74).

Aucune des molécules dirigées contre Cer-XLF n'a montré de signal STD positif. L'hypothèse que nous avons émise selon laquelle la dynamique de la tête de Cer-XLF (observée en présence de XRCC4) permettrait d'accommoder une molécule organique de faible à moyenne affinité est remise en question. Ainsi, la poche criblée précédemment (site1 Figure 41) n'a pas permis d'isoler des inhibiteurs dirigés contre Cer-XLF.

Un tableau récapitulatif est finalement dressé (Figure 74) afin de résumer les candidats qui seront intégrés aux tests de validation biochimiques d'activité.

Molecule	Protéine	STD	Specificité	Kd (mM)	Temp (K)	pH
3S-77617	XRCC4	YES	NO	-	298	7,2
7949364	XRCC4	YES	NO	-	298	7,2
6821224	XRCC4	ND	ND	ND	298	7,2
6435411	XRCC4	YES	NO	-	298	7,2
4S-62686	XRCC4	YES	YES	1	298	7,2
53464	XRCC4	NO	-	-	298	7,2
7468661	XRCC4	YES	NO	-	298	7,2
42527	XRCC4	NO	-	-	298	7,2
C522-3902	XRCC4	ND	ND	ND	298	7,2
57632	XRCC4	YES	YES	0,2	298	7,2
41376	XRCC4	NO	-	-	298	7,2
7413418	XRCC4	YES	NO	-	298	7,2
5186933	Cer-XLF	NO	-	-	298	7,2
7926063	Cer-XLF	NO	-	-	298	7,2
8876638	Cer-XLF	NO	-	-	298	7,2
5545659	Cer-XLF	NO	-	-	298	7,2

Figure 74: Tableau final de l'analyse biophysique d'interaction des composés solubles issus du cible virtuel contre XRCC4 et Cer-XLF.

Bien que deux candidats seulement soient positifs en terme de liaison spécifique à la protéine XRCC4, toutes les molécules montrant un signal positif en STD seront intégrées aux tests d'activité biologique afin de les valider ou de les infirmer de façon définitive. Le profil de titration du candidat #57632 n'est pas une hyperbole classique mais une courbe sigmoïdale. Cette sigmoïde pourrait indiquer un mode de liaison coopératif à la protéine XRCC4. Ce mécanisme reste cependant à confirmer par une méthode complémentaire (ITC par exemple).

II – 2.10 : Approche peptidique : caractérisation et analyse d'interaction par RMN des peptides XRCC4 et Cer-XLF

Deux types d'interaction protéine-protéine peuvent être dressés : les interactions domaine-domaine et les interactions médiées par des peptides (Hou et al., 2012). Précédemment, lors de l'analyse des contacts persistants entre XRCC4 et Cer-XLF au cours d'une trajectoire de DM, nous avons mentionné l'existence de « patchs » de trois ou quatre résidus consécutifs pour chacune des protéines et qui établissaient des contacts très persistants (Figure 37). Ces patchs contiennent des résidus critiques pour l'interaction entre les deux partenaires et ils pourraient, en théorie, se lier de façon spécifique à leur cible protéique s'ils étaient synthétisés.

Ainsi, une approche peptidique a été également envisagée dans la stratégie d'inhibition du complexe XRCC4/Cer-XLF. Cette approche, rationnelle, est complémentaire de la recherche de molécules organiques et les stratégies employées sont les mêmes que précédemment.

Dans un projet de caractérisation d'inhibiteurs de protéines ou de modulateurs d'interaction protéine-protéine, il semble évident de penser également à une approche peptidique. D'ailleurs, un grand nombre de drogues efficaces et utilisées en clinique sont des peptides dérivés de la nature (exemple des antibiotiques actinomycine D, vancomycine, daptomycine...). En dépit d'un nombre non négligeable d'inconvénients liés au développement d'inhibiteurs basés sur des peptides (faible biodisponibilité, faible résistance aux protéases/peptidases, faible pénétration cellulaire, haute flexibilité), une interaction peptide-protéine a surtout les avantages d'être hautement spécifique et non immunogène. Bien sûr, on parle ici de problèmes rencontrés dans un stade assez avancé de développement d'un candidat médicament potentiel. Lors de la caractérisation d'un hit peptidique, des modifications peuvent permettre de pallier à un certain nombre de ces problèmes. Par exemple, nous savons que presque tous les peptides naturels et bioactifs possèdent une structure non canonique avec le plus fréquemment un motif en macrocycle (ref). Ces peptides rigides (peptides cycliques, macrocycles) sont structuralement contraints ce qui augmente la résistance protéolytique et la perméabilité membranaire, diminue l'immunogénicité et permet d'obtenir des affinités proches de celles obtenues par l'utilisation d'anticorps. L'utilité de ces peptides contraints est surtout évidente lorsque le site de liaison à la protéine n'est pas linéaire mais conformationnel. D'autres types de modifications peuvent être l'utilisation d'acides aminés non standards ou de stéréochimie D, la suppression des sites de clivage, la N-méthylation du squelette peptidique...

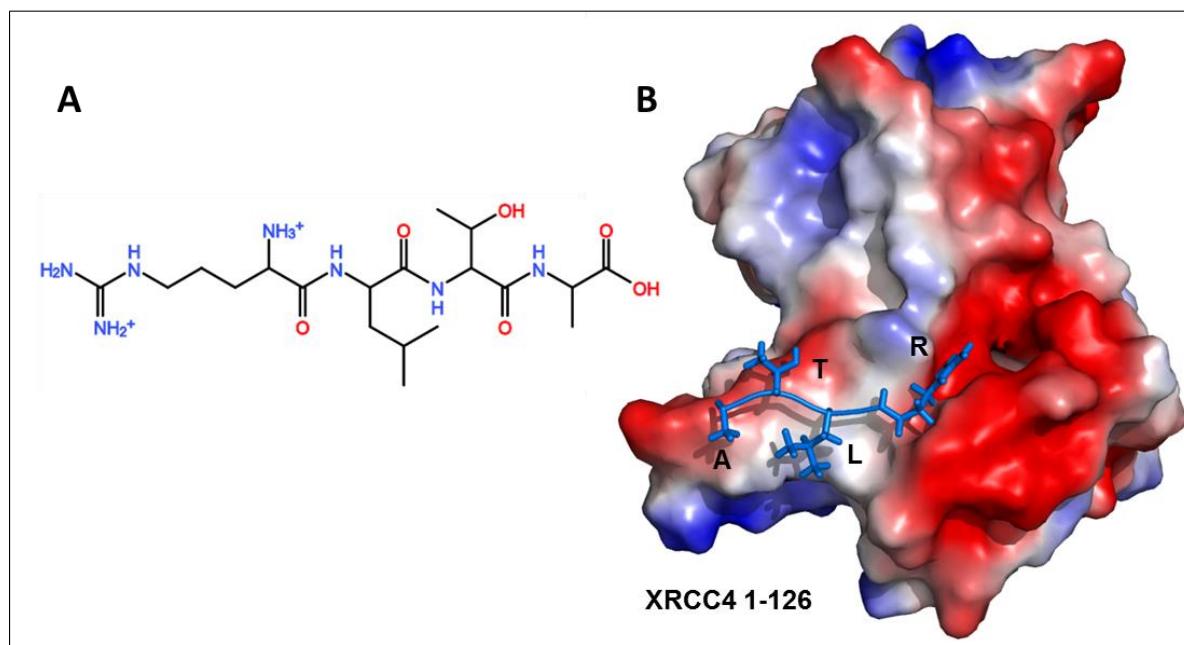
L'interaction entre XRCC4 et Cer-XLF est médiée majoritairement par des structures non secondaires (des loops), et les résidus critiques d'interaction pour chacune de ces protéines, se situent sur des structures différentes et éloignées (voir chapitre introductif). Ces interfaces médiées par des loops font partie des 50% d'interfaces protéine-protéine environ, dont l'interaction est médiées par des structures non

régulières (Gavenonis et al., 2014). La littérature a pu démontrer que ces types d'interface pouvaient être sujets notamment à des inhibitions très efficaces et médiées par des peptides contraints ou des macrocycles (Gavenonis et al., 2014). En sachant que la mutation d'un seul de ces résidus critiques est suffisante pour rompre le complexe, nous nous sommes demandés si des peptides linéaires représentant une partie seulement des épitopes d'interaction de XRCC4 et Cer-XLF pouvaient interagir avec leur cible protéique respective. Si un des peptides se révélait positif à l'interaction et avec une forte affinité, il pourrait servir par exemple de compétiteur afin de déplacer d'autres candidats potentiels et donner une preuve indirecte de la localisation de ces derniers sur la cible. Il pourrait alors également être à la base d'un inhibiteur potentiel, que l'on dériverait afin de le rendre plus résistant et spécifique (augmentation de la taille, cyclisation...).

Deux peptides linéaires ont été synthétisés :

- Un peptide Cer-XLF : NH2-R64-L-T-A67-COOH pour l'interaction sur XRCC4.
- Un peptide XRCC4 : NH2-V104-S-F-R107-COOH pour l'interaction sur Cer-XLF.

La figure ci-dessous représente ces deux peptides ainsi que leur site d'interaction respectif (Figure 75).



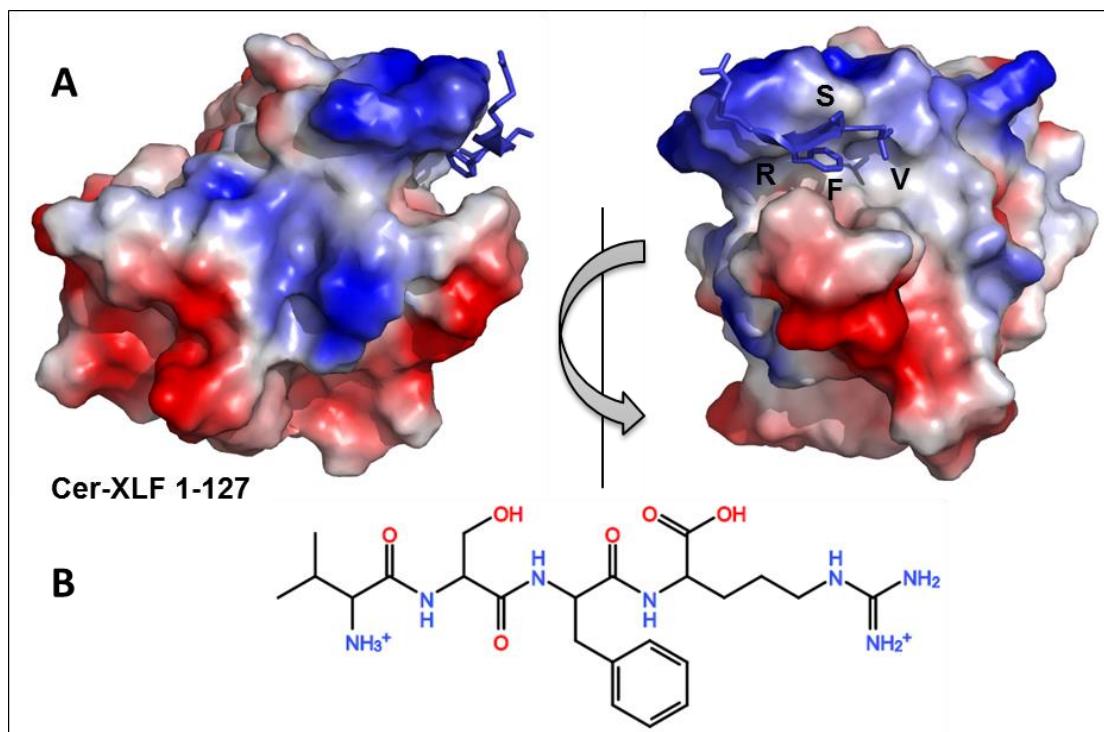


Figure 75: Représentations semi-développées des structures des peptides RLTA et VSFR, dirigés respectivement contre les protéines XRCC4 et Cer-XLF.

La stratégie employée a été la suivante : Une fois synthétisés, les peptides ont été caractérisés par RMN mono et bidimensionnelle afin d'évaluer leur pureté, intégrité et effectuer leur attribution. Une expérience de STD est réalisée avec les protéines XRCC4 et Cer-XLF et si l'expérience est positive, un contrôle sans protéine et une titration et intégration aux tests d'activité biologiques sont envisagées.

1. Caractérisation des peptides synthétisés.

Une fois synthétisés, les peptides sont purifiés par chromatographie HPLC. Les peptides RLTA et VSFR ont été jugés pur à 95% et une attribution par RMN du proton a permis de caractériser avec certitude l'intégrité de ces derniers (Figure 76). Des spectres 1D-¹H ainsi que 2D-TOCSY et NOESY ont été réalisés dans les conditions suivantes :

- Tampon RMN pH 7.0 – 10% D20
- Peptide 2mM dans un volume final de 600µL
- Spectromètre 500MHz – 298K
- Temps de mélange TOCSY : 60ms / NOESY : 500ms.

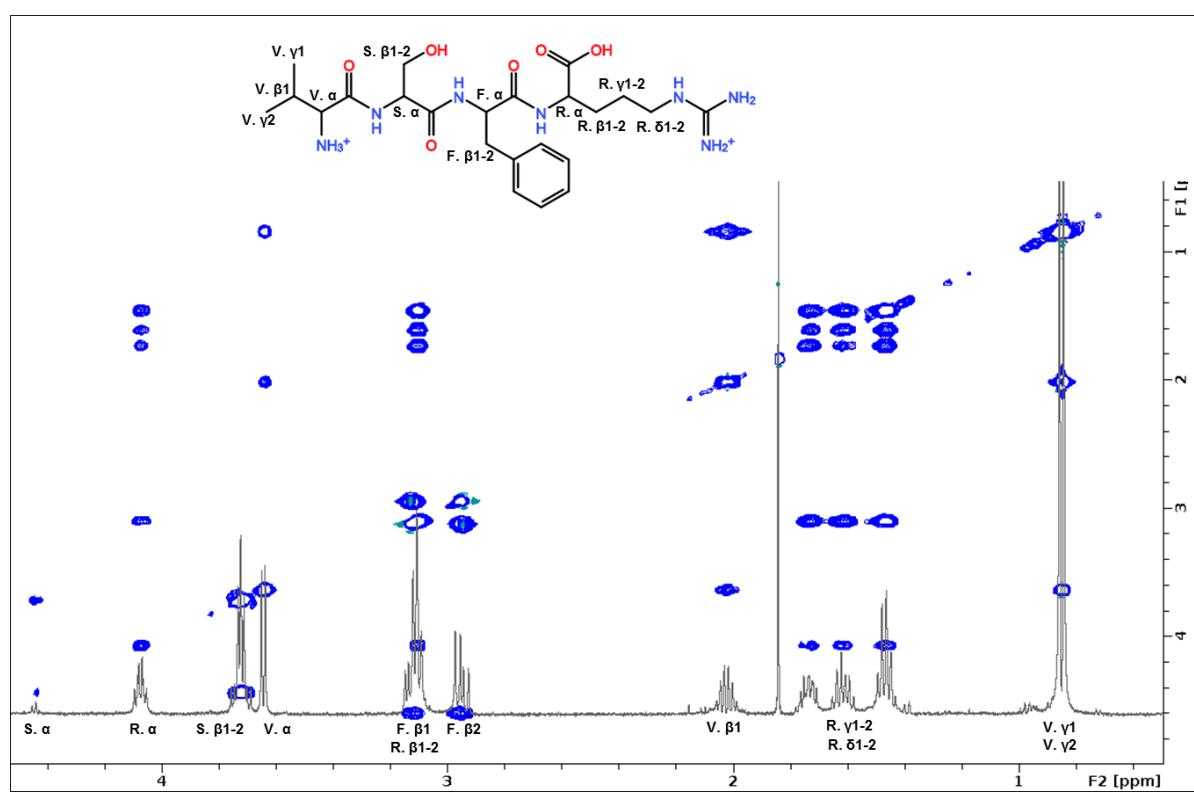
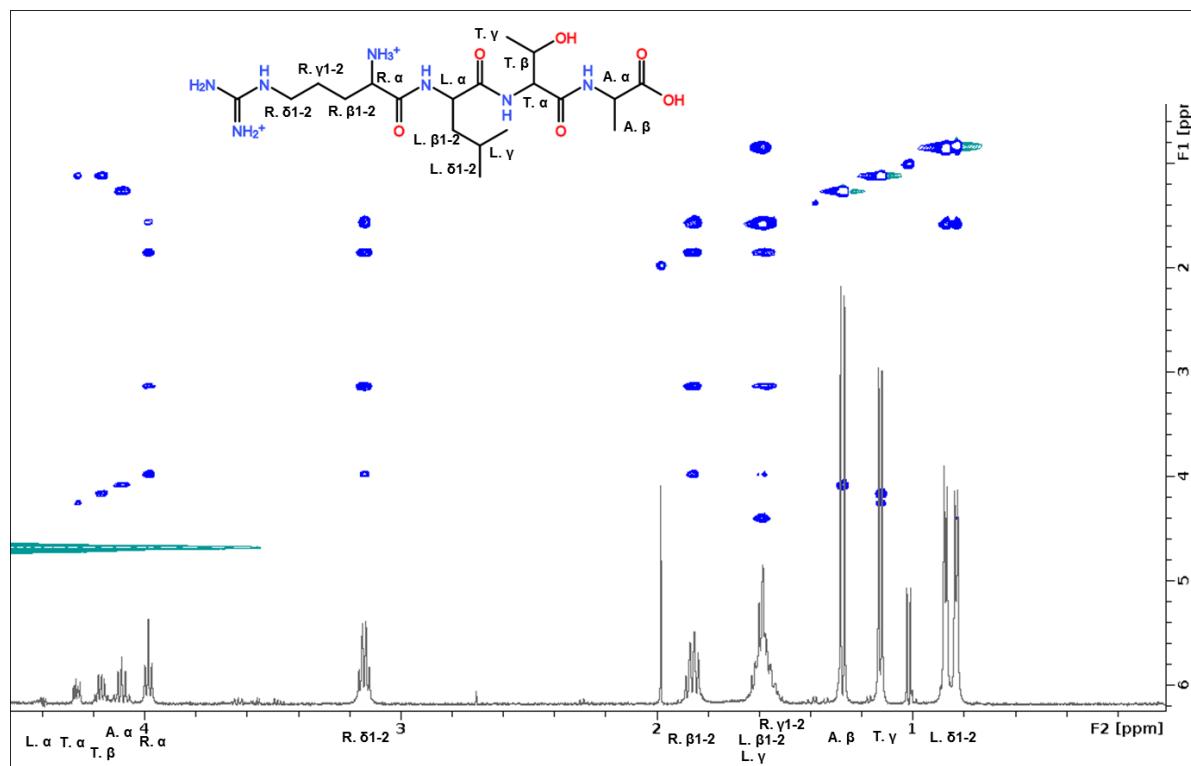


Figure 76: Spectres 2D-NOESY ^1H - ^1H et attribution des peptides RLTA (haut) et VSFR (bas).

Cette attribution permet non seulement de vérifier la bonne intégrité de nos peptides et dans les conditions expérimentales utilisées, mais aussi d'avoir une idée précise des groupements qui seront concernés par un signal STD ultérieur éventuel et d'effectuer de la cartographie d'épitope en localisant les résidus plus ou moins en interaction avec la cible protéique correspondante.

2. Analyse par STD des interactions protéine-peptide.

Les expériences de STD sont réalisées à un ratio protéine : peptide de 1 :400 et avec une concentration en peptide de 2mM. L'expérience contrôle (sans protéine) est réalisée dans un tampon qui correspond au tampon de conservation de la protéine d'intérêt. La température au cours de l'expérience est fixée à 283K afin de ralentir le phénomène d'échange des protons amides et de pouvoir observer ces derniers sur le spectre. Les fréquences de on et off-résonance ont été fixées respectivement à -1 et 30 ppm. Un pulse Gaussien de 50ms et 86Hz est utilisé et la séquence STD contient également une suppression du signal de l'eau en Watergate et un spin lock de 30ms à 4960Hz afin de supprimer les signaux résiduels de la protéine (Figure 77).

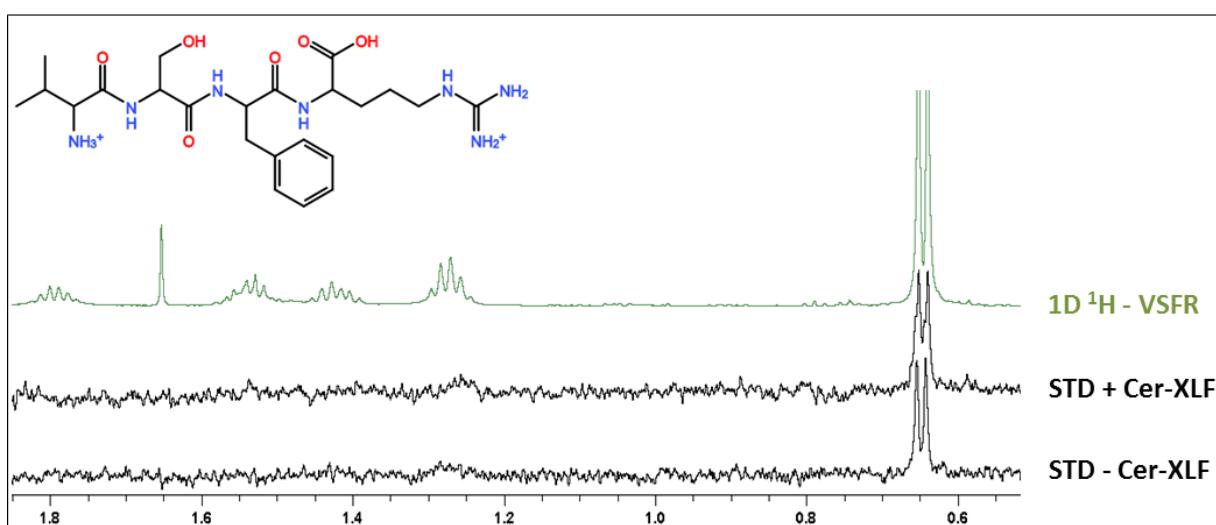
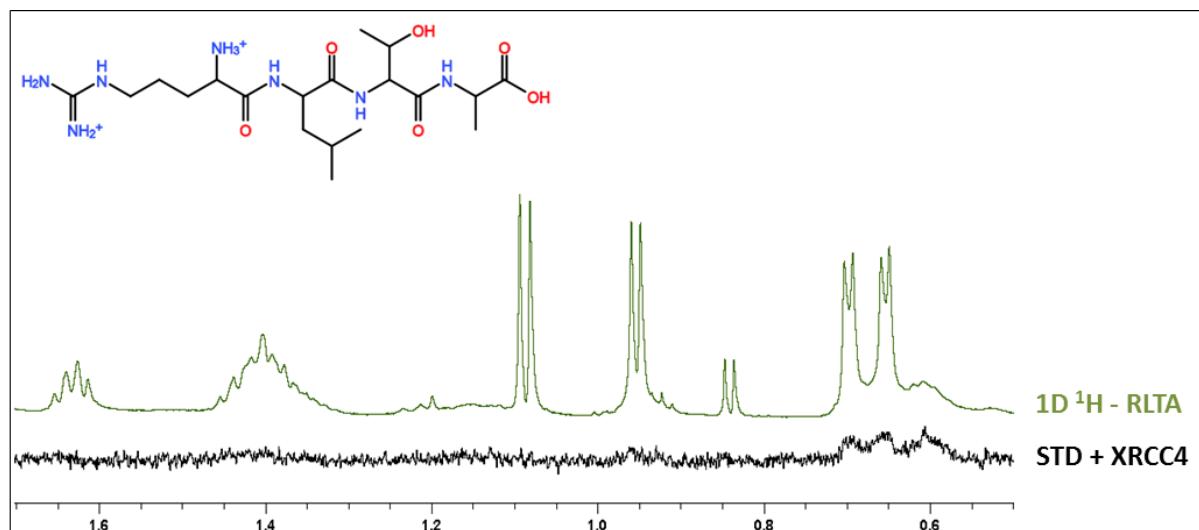


Figure 77: Spectres 1D- ^1H et 1D- ^1H STD des peptides RLTA (haut) et VSFR (bas). Les spectres ont été réalisés dans les conditions fixées ci-dessus et en présence ou absence (contrôle de spécificité d'irradiation si signal STD positif) de protéine (Cer-XLF pour le peptide VSFR et XRCC4 pour le peptide RLTA) pour l'expérience de STD.

Si une réponse STD est positive, le contrôle sans protéine est réalisé en suivant et une titration permet de déterminer si la liaison est spécifique ainsi que le Kd. Dans notre cas, le peptide RLTA n'a donné aucun signal STD sur XRCC4. On note seulement un signal parasite quasiment confondu avec le signal du bruit. Le peptide VSFR quant à lui donne un signal STD positif (correspondant au groupement dimethyl de la Valine 1 – Figures 76-77), mais ce signal est un signal non spécifique car la même expérience réalisée sans protéine Cer-XLF donne le même résultat. Au final, aucun des deux peptides n'interagit avec sa cible respective dans les conditions expérimentales utilisées. Une seconde expérience biophysique

complémentaire pourrait être réalisée afin de confirmer ce résultat (ITC par exemple). Cependant, cela semble confirmer que chaque résidu impliqué dans d'interaction entre XRCC4 et Cer-XLF a son importance (épitope conformationnel) et que réduire la zone d'interaction à un peptide linéaire de quelques résidus critiques est une stratégie qui, ici, n'a pas abouti. Lorsque l'on reprend l'exemple de Cer-XLF, la mutation ponctuelle de sa Leucine 115 prévient l'interaction avec XRCC4. L'ancrage de cette leucine au sein d'une poche hydrophobe du domaine globulaire de XRCC4 est, à priori, primordial pour la stabilisation des interactions entre les autres paires de résidus (Figure 40). Nos peptides étant aussi potentiellement très flexibles (car non contraints structuralement), cela pourrait également expliquer l'absence d'une interaction stable avec la cible.

Ainsi, l'approche peptidique pour l'inhibition de l'interaction XRCC4/Cer-XLF a été explorée mais les premiers résultats sont négatifs. L'utilisation de peptides linéaires contenant seulement quelques-uns des résidus critiques semble être une technique inadaptée dans ce cas précis. Des expériences complémentaires avec des banques de peptides mais également des peptides cycliques (plus adaptés aux épitopes conformationnels et aux interactions médiées par des structures non canoniques), pourraient cependant être testées.

II – 2.11 : Validation des inhibiteurs potentiels par des tests de stimulation de la ligase T4

Durant cette campagne de criblage, plusieurs stratégies ont été explorées en parallèle afin de découvrir des inhibiteurs potentiels de l'interaction entre XRCC4 et Cer-XLF (HTS, criblage virtuel, approche peptidique). La campagne HTS ainsi que l'approche par les peptides se sont avérées infructueuses. Seule la campagne virtuelle a permis d'aboutir à un petit jeu de molécules, solubles, positives à l'interaction avec XRCC4, et donc potentiellement inhibitrices. La zone d'interaction ainsi que le mode de liaison de ces molécules sont des facteurs encore inconnus à ce stade. Ils devront cependant être explorés à l'issus des tests de validation d'activité biologique si une inhibition se confirme.

Dans le chapitre II – 1.3, nous avons développé le principe et l'optimisation du test indirect d'activité du complexe XRCC4/Cer-XLF, basé sur une stimulation de l'activité

de ligation intermoléculaire de la ligase T4. Les conditions étant mises au point, il nous faut tester les molécules potentielles afin d'évaluer si la stimulation de la ligase s'en retrouve atténuée. Cela témoignerait d'un effet préventif de l'association entre XRCC4 et Cer-XLF et marquerait une première étape dans la preuve de concept d'une inhibition de la réparation de l'ADN par l'intermédiaire d'une petite molécule, inhibitrice de l'interaction XRCC4/Cer-XLF.

Les candidats testés sont les molécules ayant donné une réponse STD positive dans les essais de validation biophysique précédents. Ces 7 molécules sont :

#3S-77617, #4S-62686, #7949364, #6435411, #7468661, #57632 et #7413418.

Parmi ces dernières, seules les molécules #4S-62686 et #57632 ont une activité spécifique de liaison à XRCC4 déterminée par titration STD. Mais toutes les molécules seront intégrées au test sans à priori, même si l'on s'attendrait à un effet provenant de ces deux en particulier.

La liaison à XRCC4 est spécifique pour ces deux molécules mais l'effet inhibiteur *in vitro* pourrait être inefficace cependant et ce pour plusieurs raisons :

- La première raison est que nous n'avons aucune preuve expérimentale à ce stade d'une liaison au niveau de la zone d'interaction avec Cer-XLF. Les molécules en question pourraient très bien se loger ailleurs que dans la zone ciblée virtuellement (dans une cavité cryptique par exemple, sur une structure qui n'est pas visible en cristallographie ou par DM) sans perturber la liaison à Cer-XLF.
- De plus, les molécules ci-dessus ont une affinité de liaison faible (estimé de 0.2 à 1mM). Ce Kd est très éloigné du Kd entre les protéines XRCC4 et Cer-XLF (autour de 7 μ M), et pourrait ne pas suffire pour permettre une compétition efficace avec Cer-XLF sans optimisation chimique.
- Finalement, une liaison non spécifique des molécules sur la ligase T4 ou sur l'ADN pourrait engendrer une perte significative d'efficacité de ces dernières.

La stratégie adoptée au cours de ces expériences a donc été de réaliser la même expérience de stimulation de la ligase T4 en incubant soit directement le complexe XRCC4/Cer-XLF dans le milieu réactionnel, soit en pré-incubant d'abord à température ambiante la protéine XRCC4 avec l'une des molécules, puis en rajoutant

au milieu réactionnel les complexes XRCC4/molécule préformés, suivi de la protéine Cer-XLF. Le ratio protéine : ligand utilisé a été de 1 : 500 et 1 : 1000, se plaçant ainsi en condition de fort excès en ligand par rapport à la cible et également par rapport à Cer-XLF (tout comme dans les conditions du STD). De plus, chaque molécule a également été testée pour un éventuel effet non spécifique sur la ligase T4 et éviter ainsi de biaiser notre résultat. Le résultat de l'effet inhibiteur des candidats sélectionnés est représenté dans la figure ci-dessous (Figure 78).

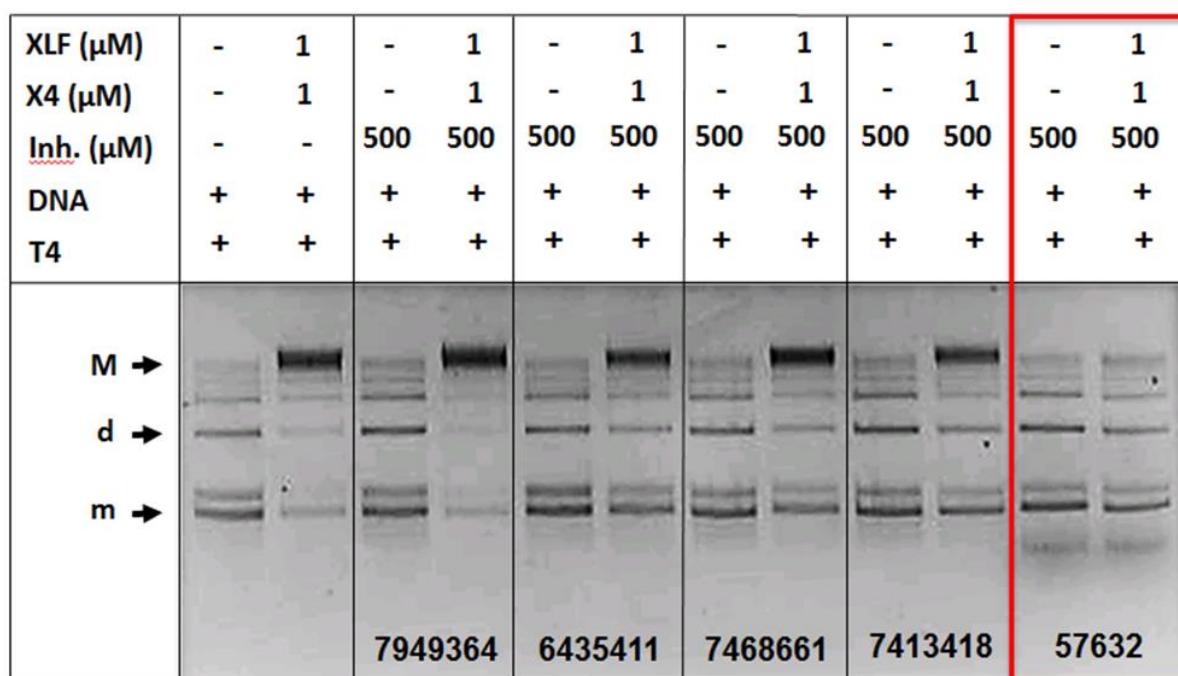


Figure 78: Expérience de stimulation de la ligase T4 par le complexe XRCC4/Cer-XLF en présence des molécules touches sélectionnées. Inh. = Inhibiteur ; M = Multimères ; d = dimères ; m = monomères. Chaque inhibiteur a été testé en présence ou absence de protéines afin de démontrer l'absence d'un effet non spécifique éventuel sur la ligase T4. La molécule #57632 (encadré en rouge) est la seule ici à inhiber de manière très significative l'effet du complexe XRCC4/Cer-XLF sans affecter l'activité basale de la ligase T4.

Ces résultats nous indiquent que les molécules testées n'induisent aucune inhibition spécifique de l'activité basale de la ligase T4. La stimulation observée est de forte intensité et la concentration ultérieure utilisée en protéines a été de 0.5μM au lieu de 1μM afin d'atténuer légèrement cet effet.

Nous pouvons facilement conclure que seul l'inhibiteur #57632 possède une activité inhibitrice claire sur la stimulation de la ligase par le complexe XRCC4/Cer-XLF après pré-incubation de ce dernier sur XRCC4. L'activité de la ligase T4 retourne à

un état basal après traitement par ce dernier. Cette molécule (candidat favoris à l'issu des tests de validation biophysique) est par conséquent un candidat sérieux pour des étapes ultérieures de caractérisation structurale et d'optimisations chimiques.

Il reste toutefois à confirmer son effet lors d'un test NHEJ *in vitro*. Ce test permet de valider définitivement l'effet inhibiteur et spécifique de notre molécule car il est réalisé en présence d'extraits cellulaires (et non uniquement des protéines XRCC4 et CernXL purifiées). Un comparatif est également dressé afin de comparer ses effets spécifiques sur le mécanisme de NHEJ classique dans des extraits proficients ou bien déficients en ligase 4 humaine. Un effet dose est mesuré et éventuellement, des molécules dites « decoy » (molécules structuralement très proches des molécules inhibitrices mais sans activité de liaison à XRCC4) pourraient être intégrées. On observe également au cours de l'expérience ci-dessus, l'apparition d'une bande sous l'espèce monomérique de l'ADN. Cette bande n'est pas présente au cours de la ligation mais résulte de l'ajout de l'inhibiteur #57632. Seul cet inhibiteur possède cet effet. Nous pourrions émettre l'hypothèse que l'inhibiteur affecte l'activité de ligation de la ligase T4, induisant un mécanisme de religation intramoléculaire du plasmide linéarisé et donc, un faux-positif dans le test d'inhibition de la stimulation par le complexe protéique. Cette nouvelle bande pourrait également être la molécule elle-même, liant ainsi le BET ou bien encore une forme d'ADN complexée à la molécule.

Afin de tester ces hypothèses, une même expérience a été réalisée en utilisant comme ci-dessus, un plasmide linéarisé par une seule enzyme de restriction (pouvant donc permettre de la religation intramoléculaire), mais également un plasmide coupé par deux enzymes différentes, empêchant ainsi cette religation éventuelle. Le résultat est présenté dans la figure ci-dessous (Figure 79).

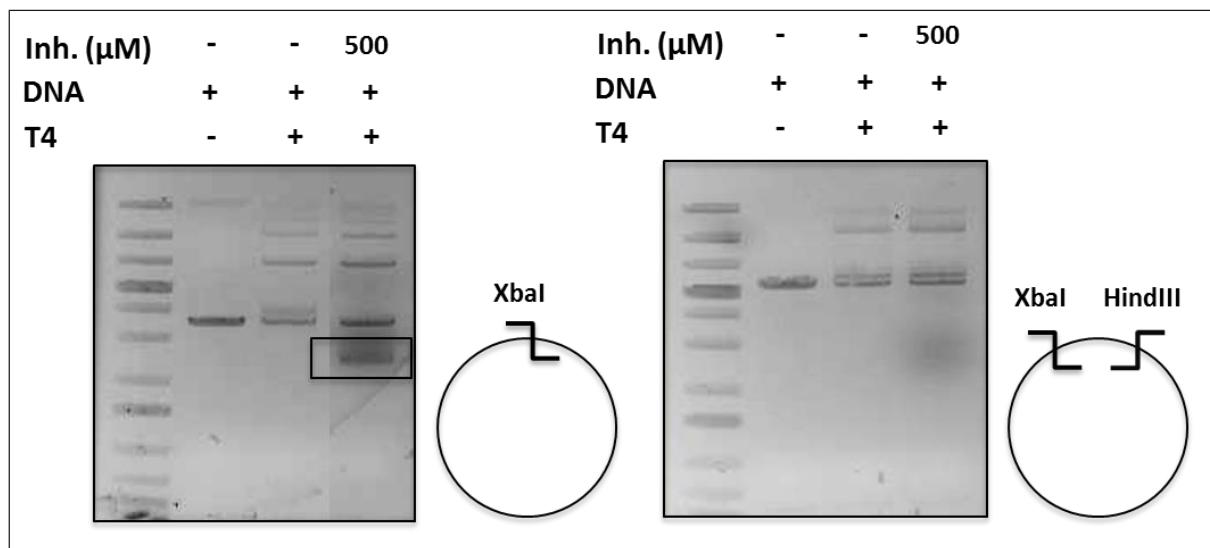


Figure 79: Expériences de religation intermoléculaire par la ligase T4 en présence d'un plasmide simplement (XbaI) ou doublement (XbaI + HindIII) coupé. Inh. = Inhibiteur. Par rapport à l'expérience utilisant un plasmide doublement digéré, l'expérience utilisant un plasmide simplement digéré présente en fin de réaction une forme d'ADN correspondant certainement à une religation intramoléculaire (encadrée en noir).

Au vu de cette expérience, il est clair que la bande observée précédemment était un produit de religation intramoléculaire de l'ADN linéaire. Le plasmide doublement coupé a donc été utilisé pour la suite des expériences. Dans la figure ci-dessous (Figure 80), on démontre que cet effet de religation non spécifique une fois supprimé, on conserve l'activité inhibitrice du composé #57632 sur le complexe XRCC4/Cer-XLF. Il n'y a pas d'effet inhibiteur possible du DMSO sur les protéines XRCC4 et Cer-XLF car celui-ci est ajouté en contrôle dans les pistes ne contenant pas la molécule inhibitrice. L'effet inhibiteur de #57632 est reproductible et visible à une concentration en molécule de 500μM (Avec un ratio protéine : ligand final utilisé ici de 1 : 1000).

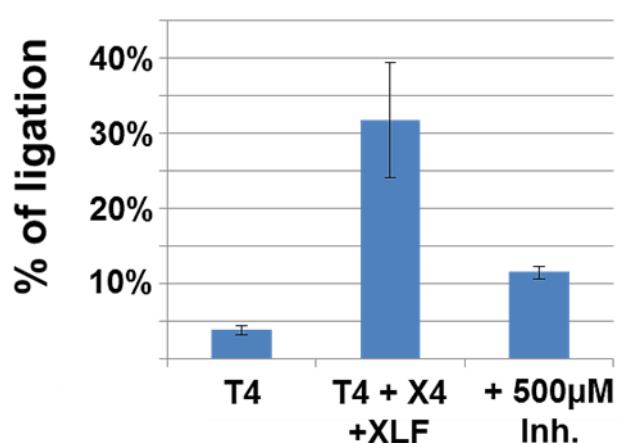
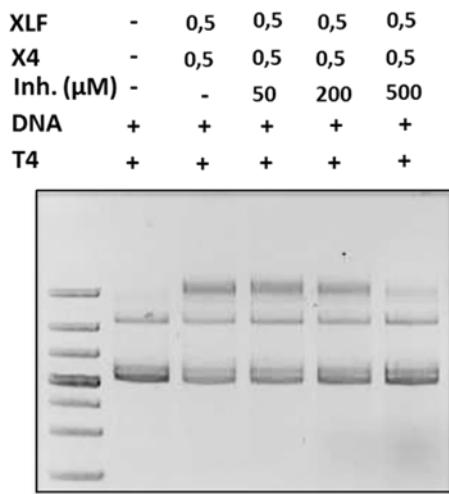


Figure 80: Expériences de stimulation de la ligase T4 par le complexe XRCC4/Cer-XLF et en présence de l'inhibiteur #57632. *Inh.* = Inhibiteur. Lors de cette expérience, l'activité de stimulation est presque complètement inhibée à partir d'une concentration en molécule de 500µM. Cet effet est significatif avec un retour à une activité basale, et reproductible (à droite, histogramme montrant le % de ligation de la ligase T4 seule et en présence du complexe XRCC4/Cer-XLF ± inhibiteur #57632. La quantification est réalisée selon le rapport : bande des multimères / ADN total * 100 et via le programme « ImageJ »).

Nous avons également définitivement écarté la possibilité d'une intercalation possible de notre composé avec l'ADN par une expérience de gel shift en utilisant la doxorubicine comme témoin positif d'intercalation et dans le même tampon que celui du test d'activité précédent (Figure 81).

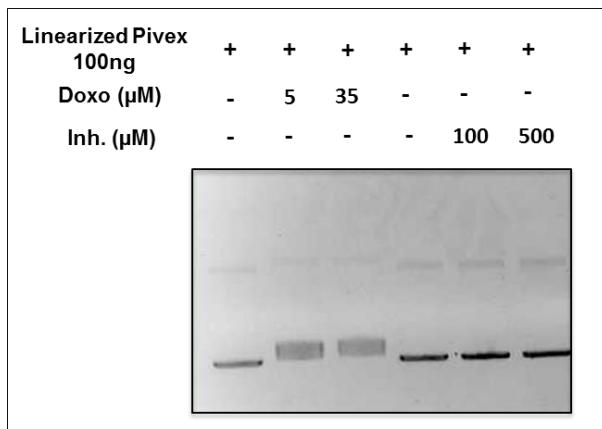


Figure 81: Expérience d'intercalation à l'ADN. *Inh.* = Inhibiteur ; *Linearized Pivex* = plasmide Pivex linéarisé ; *Doxo* = Doxorubicine. La molécule de doxorubicine s'intercale à l'ADN et un shift consécutif à cette intercalation est présent sur le gel d'agarose. Aucun phénomène d'intercalation n'est visible ici avec l'inhibiteur #57632.

Le candidat #57632 étant favori à l'issue de ces premiers tests de validation biochimique et biophysique, une analyse de ce composé via le serveur web « FAF-Drugs3 » (Lagorce et al., 2015) a été effectuée et nous permet de proposer ce dernier comme un candidat « Drug-Like » intermédiaire, qui ne viole aucune des règles de Lipinski, avec une bonne biodisponibilité théorique ainsi qu'une prédition comme pouvant être potentiellement un inhibiteur d'interaction protéine-protéine. Cela nous encourage donc par la suite à s'intéresser plus particulièrement à ce composé.

II – 2.12 : Validation des inhibiteurs potentiels sur des essais NHEJ avec extraits cellulaires

Afin d'aller plus loin dans la caractérisation de l'effet inhibiteur de la molécule #57632 et également prouver sa spécificité au sein du complexe XRCC4/Cer-XLF impliqué dans la voie NHEJ classique, des tests NHEJ *in vitro* ont été réalisés avec des extraits cellulaires (voir partie Matériels et Méthodes). Ces extraits cellulaires proviennent de cellules HCT116 sauvages (voie NHEJ classique proficiente) ou déficientes en ADN Ligase IV humaine (voie NHEJ classique déficiente). L'extrait déficient en Ligase IV est activé pour une ligature dépendante de la Ligase III par ajout d'un oligonucléotide de 32 bases (Yuan et al., soumis pour publication). Ce test nous a permis de confirmer d'une part, l'effet inhibiteur spécifique de la molécule #57632 au sein d'un système plus complexe non composé uniquement de protéines purifiées, mais également de tester la spécificité envers le complexe XRCC4/Cer-XLF ciblé (car les cellules déficientes en ADN Ligase IV font alors appel à une voie de réparation alternative (Alt-NHEJ), dépendante de l'ADN Ligase III). Le test réalisé est présenté ci-dessous (Figure 82).

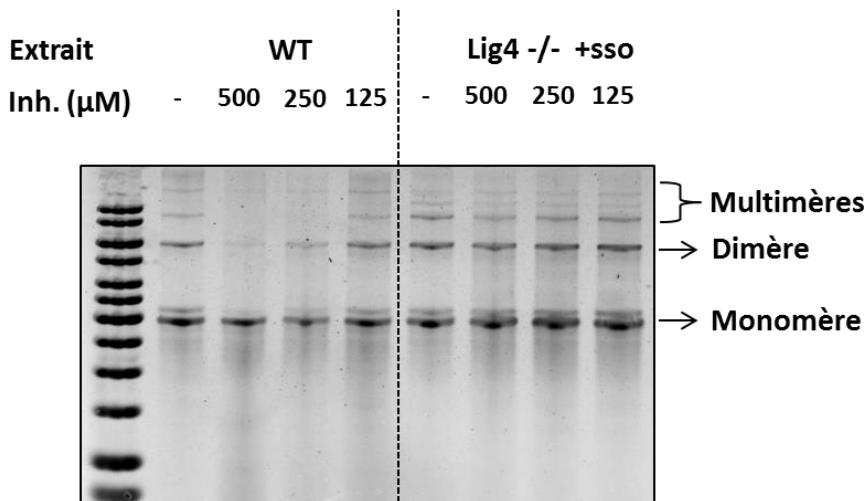


Figure 82: Essais NHEJ *in vitro* avec extraits cellulaires proficients (WT) ou déficient en ADN Ligase IV humaine (Lig4 -/-) et en présence d'une dose croissante d'inhibiteur #57632. Inh. = Inhibiteur ; Extrait = Extrait cellulaire. Dans ces conditions, l'inhibition par le composé #57632 du mécanisme NHEJ est effectif à partir d'une concentration de 250µM. Cet effet est encore plus important à 500µM. L'effet inhibiteur n'est pas observé pour les extraits cellulaires provenant de cellules déficientes en ADN Ligase IV.

Le résultat de ce test nous indique que l'inhibiteur est également efficace dans un milieu plus complexe tel qu'un extrait cellulaire. La concentration effective est du même ordre de grandeur (250-500µM) que pour le test de stimulation de la ligase T4 présenté ci-dessus. Finalement, cet effet semble spécifique du mécanisme NHEJ classique puisque l'inhibition n'est pas observée dans un extrait Ligase 4 -/. Ce résultat apporte une preuve supplémentaire sur la spécificité d'action du candidat, très probablement localisé au niveau de la protéine XRCC4. Cette localisation reste toutefois à confirmer expérimentalement.

II – 2.13 : Clonage et production de la partie N-terminale de XRCC4 (1-112 ; 1-119)

La caractérisation structurale des complexes protéine-inhibiteurs est une étape clé dans l'élaboration d'un candidat médicament de haute affinité. Pouvoir déterminer de façon précise la localisation ainsi que le mode de liaison du ligand sur sa cible nous permet d'entrer dans une phase d'optimisation chimique de ce dernier (stratégie « Hit-to-Lead »), et espérer obtenir ainsi une molécule plus affine et spécifique de la protéine à inhiber. Les molécules #57632 ainsi que #4S-62686, ont une activité spécifique de liaison à XRCC4. Leur localisation et mode de liaison sur XRCC4 sont

inconnus mais l'hypothèse est qu'elles se fixeraient au niveau de la tête globulaire de XRCC4, non loin de la zone critique d'interaction à Cer-XLF. La preuve expérimentale de l'activité inhibitrice du composé #57632 dans le test de stimulation de la ligase T4, pourrait apporter un premier élément de confirmation indirect de sa localisation sur la cible mais cela reste incertain car il pourrait agir en tant qu'inhibiteur allostérique, au niveau d'un site éloigné de celui ciblé de façon rationnelle. L'affinité de ces deux composés favoris est assez faible au vu de l'affinité entre XRCC4 et Cer-XLF et une optimisation chimique permettrait de modifier de manière favorable ce Kd de liaison à la cible, et en particulier pour le #57632 qui est un ligand jusque-là prometteur. Sans cela, l'étape de validation de l'effet potentiellement radiosensibilisant de nos molécules n'est pas encore possible au niveau cellulaire (ou en tout cas, risque d'être inefficace).

Les deux techniques biophysiques clés utilisées dans la détermination structurale des complexes protéine-ligand sont la RMN et la cristallographie et diffraction des rayons X. Nous ne détaillerons pas ici les avantages et inconvénients de ces deux techniques biophysiques complémentaires mais nous exposerons les raisons et stratégies employées afin de déterminer la localisation des deux composés ci-dessus au sein de la protéine XRCC4.

Afin d'optimiser nos chances d'obtenir cette information, les deux techniques (cristallographie et RMN) ont été utilisées en parallèle. Je détaillerai dans cette partie le travail mené pour la caractérisation structurale des complexes par RMN. Cette méthode nécessite, tout comme pour la cristallographie, la production d'une protéine en grande quantité (plusieurs milligrammes par millilitre), homogène et très pure. Mais surtout, la protéine doit être produite avec un double marquage N¹⁵/C¹³, afin de pouvoir conduire les expériences RMN adaptées à la caractérisation structurale des complexes (après attribution du squelette de la protéine).

Nous avons choisi d'effectuer une première étude de faisabilité, en produisant des formes tronquées de la protéine XRCC4 disponible dans notre laboratoire. Ces formes tronquées (correspondant à la tête globulaire en N-terminal de XRCC4) possèdent une taille plus adaptée à la résolution structurale par RMN (inférieure à 30kDa). De plus, la forme sauvage 1-336 de XRCC4 que nous possédons, permet de produire une protéine soluble mais que l'on ne peut concentrer au-delà de 30µM.

Cette limite de concentration est un frein à la résolution de structure par RMN. Nous devons par conséquent tenter de produire un domaine protéique stable dont la concentration pourra être nettement supérieure (au-delà de 100 μ M). Donc, en plus d'apporter un système d'étude plus simple et adapté à la RMN, ces domaines tronqués et privés de leur partie C-terminale pourraient permettre une plus grande concentration. De plus, si les têtes globulaires produites ne subissent pas de changement conformationnels par rapport à la forme sauvage non tronquée, elles permettraient d'apporter un élément de réponse quant à la localisation effective de nos molécules au niveau des têtes globulaires et non du coiled-coil par exemple, ou de la partie C-terminale non structurée de XRCC4.

Ci-dessous, les différentes étapes nécessaires à notre étude sont schématisées dans l'ordre (Figure 83). Si l'une de ces étapes n'est pas validée, la suite des expériences n'est pas possible.

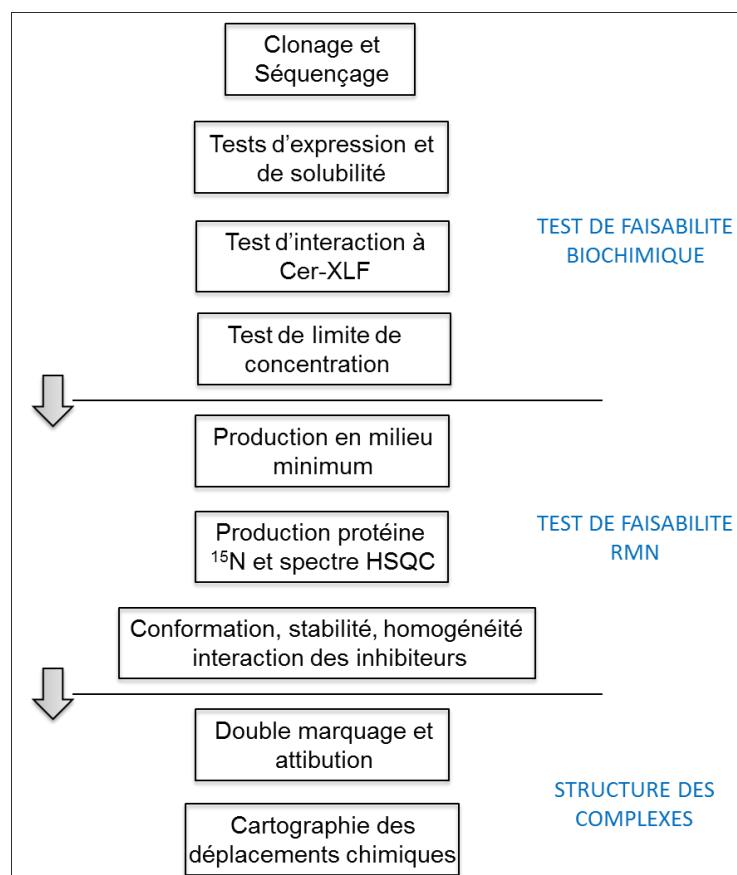


Figure 83: Schéma représentatif des différentes étapes utilisées pour la détermination par RMN de la zone d'interaction de l'inhibiteur #57632 au sein de la tête globulaire de XRCC4.

Les séquences ainsi que les détails de la stratégie de clonage des domaines N-terminaux tronqués de XRCC4, sont détaillés en partie Matériels et Méthodes. Ci-dessous, sont représentées les formes clonées de XRCC4 (Figure 82). Ces formes tronquées ont été choisi selon quelques critères (notamment : domaine complet, non tronqué au niveau d'une structure secondaire). Cependant, nous devions émettre l'hypothèse que ces domaines pouvaient ne pas être solubles au vu d'une stabilisation par une partie du coiled-coil que l'on peut observer sur la structure cristallographique 1FU1 (Figure 84).

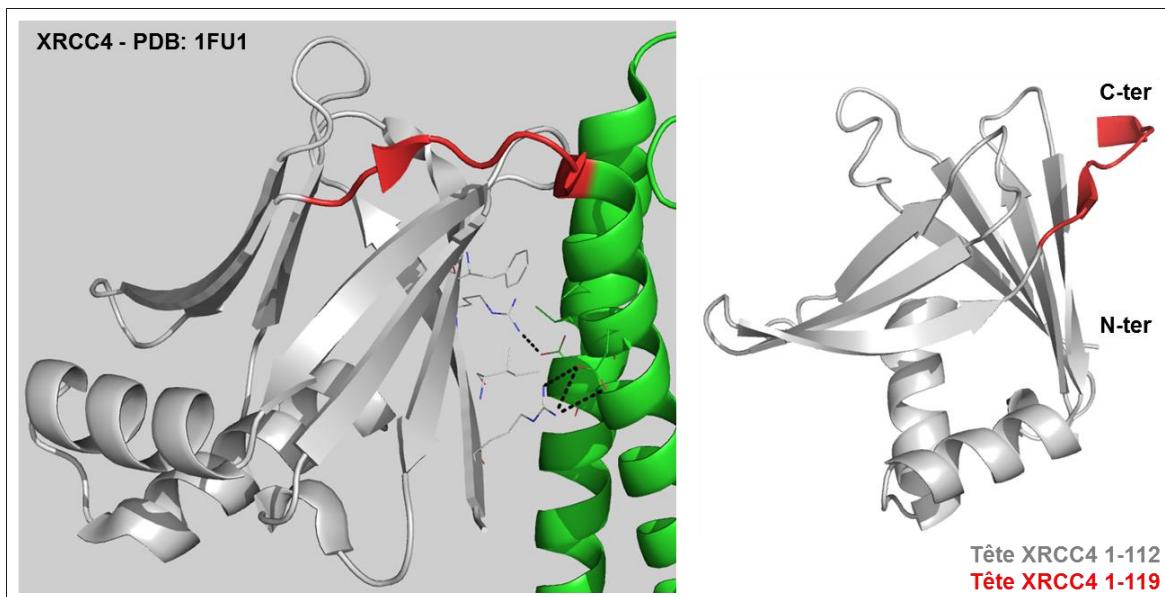


Figure 84: Structure cristallographique de la protéine XRCC4 (PDB code : 1FU1) centrée sur les domaines N-terminaux 1-112 et 1-119. Le domaine tronqué gris représente le domaine cloné 1-112. Le domaine tronqué gris + rouge représente le domaine cloné 1-119. La stabilisation de ce domaine N-terminal au sein de la structure native se fait grâce à un réseau de liaisons hydrogènes (pointillés noir) avec des résidus du coiled-coil.

Une fois les domaines clonés et séquencés, des tests d'expression et de solubilité sont effectués dans une souche bactérienne BL21STAR. Les domaines produits doivent conserver leur activité de liaison à la protéine Cer-XLF (preuve indirecte de la conservation de leur forme native par rapport à la protéine sauvage). Eventuellement, une expérience de dichroïsme circulaire (DC) pourrait permettre à ce stade d'évaluer le repliement de la protéine sans toutefois d'indication sur l'état natif de ce dernier. Les tests d'interaction peuvent être effectués selon la méthode

des gels retards, exposés en chapitre II – 1.2. Si les domaines sont solubles et conservent leur activité de liaison à Cer-XLF, un test de limite de concentration devra être mené pour évaluer la faisabilité des expériences RMN. Une fois ces tests validés, la production des domaines tronqués en milieu minimum sera testée afin, là encore, d'établir la faisabilité d'une production de protéines solubles, actives et en quantité suffisante. Des premiers spectres HSQC permettront de déterminer de manière plus précise l'état conformationnel, la stabilité au cours du temps ainsi que l'homogénéité des échantillons produits. Finalement, une étape critique est celle du test d'interaction de nos molécules candidates sur ces domaines. En supposant que la structure des domaines est conservée et que ces ligands interagissent bien au niveau de la zone N-terminal ciblée, ces derniers doivent être testés afin de démontrer qu'ils induisent bien un shift de déplacement chimique ou un changement d'intensité ou de conformation au niveau des taches de corrélation sur les spectres HSQC. De plus, un mouvement de quelques tâches isolées pourrait être une preuve de spécificité du composé pour la cible. Si tel est le cas, la stratégie de cartographie des déplacements chimiques et caractérisation de la zone de liaison de nos molécules est possible. Une titration HSQC sera également effectuée afin de confirmer la valeur de K_d trouvée précédemment par STD.

Les clonages des domaines 1-112 et 1-119 de XRCC4 ont été effectués comme décrit dans la partie Matériels et Méthodes. Après vérification par séquençage, les premiers tests d'expressions et de solubilité ont été réalisés avec la forme tronquée 1-112 de XRCC4 (Figure 85). Ces productions en petits volumes ont été réalisées en milieu LB riche ou minimum auto-inductible (MMAI). Ce dernier aurait notamment permis une production lente (30h à 20°C) et graduelle de la protéine, en évitant la formation de corps d'inclusion et en favorisant une forme soluble. Pour le milieu MMAI, la consommation d'arabinose va permettre la production d'allolactose, qui est un analogue structural de l'IPTG. En milieu riche, deux conditions standard : 4h à 37°C et 24h à 20°C ont été testées. Après expression, une lyse bactérienne (par sonication) a permis de séparer le culot du surnageant et chacune de ces fractions a été analysées sur gel de polyacrylamide. Finalement, aucune des trois conditions testées n'a pu permettre de produire cette construction sous forme soluble.

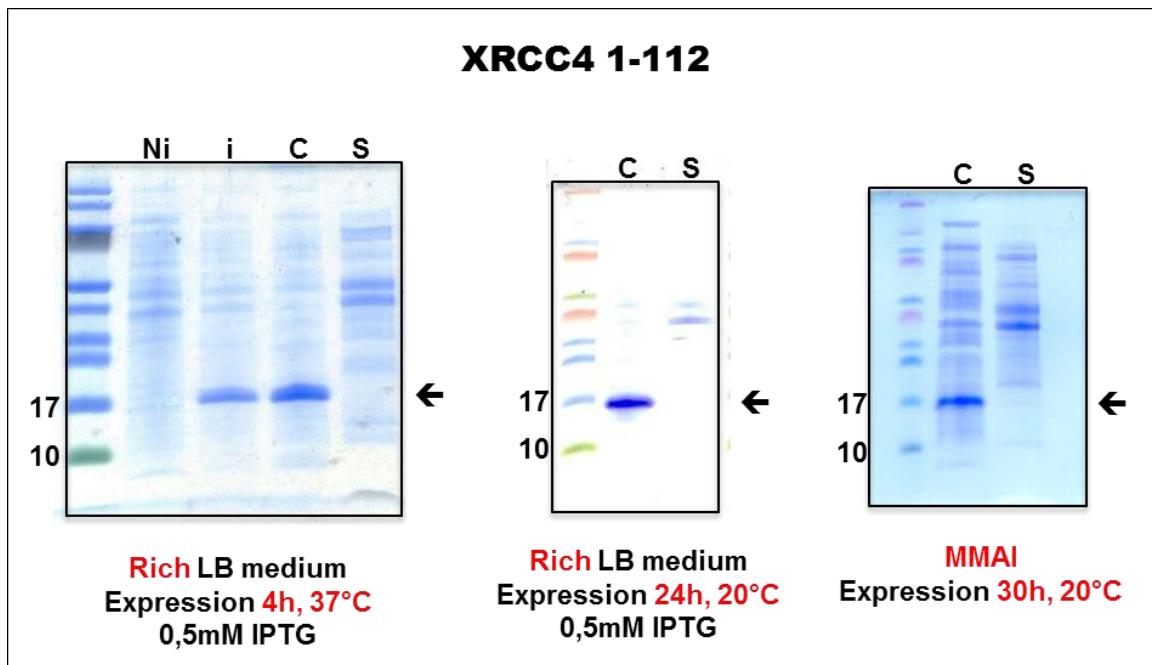


Figure 85: Tests d'expression et de solubilité de la forme XRCC4 tronquée 1-112. Ni = Non-induit ; i = induit ; C = Culot ; S = Surnageant. La bande correspondant au domaine surexprimé se situe autour de 17kDa et est précisée par une flèche noire.

Les seconds tests d'expressions et de solubilité ont été réalisés avec la forme tronquée 1-119 de XRCC4 (Figure 86). Ces productions en petits volumes ont été réalisées en milieu LB riche durant 4h à 37°C ou bien 24h à 20°C. Pour ces tests, seule la condition : milieu LB riche, expression 24h à 20°C avec 0.5mM IPTG a permis de produire une quantité faible de domaine soluble.

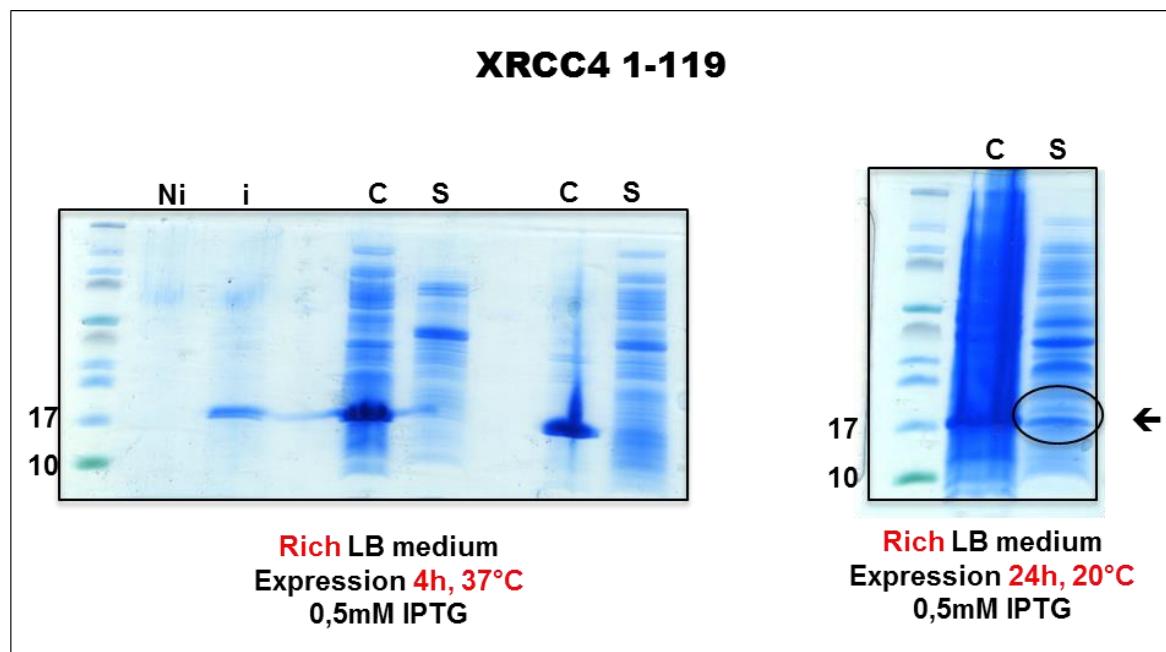


Figure 86: Tests d'expression et de solubilité de la forme XRCC4 tronquée 1-119. *Ni = Non-induit ; i = induit ; C = Culot ; S = Surnageant. La bande correspondant au domaine surexprimé se situe autour de 17kDa et est précisée par une flèche noire. La forme soluble produite est entourée d'un cercle noir.*

En conclusion pour cette partie clonage de la tête XRCC4, il semble que ce domaine ce comporte majoritairement comme une protéine insoluble et dans la plupart des conditions. Un faible rendement de protéine soluble a pu cependant être obtenu (conditions ci-dessus) mais ne permet pas d'envisager les études structurales par RMN précédemment citées. Ce faible rendement, néanmoins, peut-être exploité via une production large volume et le domaine soluble produit permettrait de réaliser les expériences STD précédente avec l'inhibiteur #57632 favori et la forme non tronquée de XRCC4. Si positives, ces expériences STD confirmeraient alors la liaison spécifique du composé au niveau de la tête globulaire de XRCC4. Toutefois, ceci est à la condition que le domaine produit conserve une structuration identique à la forme non tronquée (et cela peut être éventuellement vérifié en gel natif par interaction avec Cer-XLF).

II – 2.14 : Cristallisation de la protéine XRCC4 – Trempage de cristaux

La découverte de molécules touchées avec une affinité milli ou sub-millimolaire pour la protéine XRCC4 fournit un point de départ dans une stratégie d'optimisation chimique de ces dernières afin d'obtenir des molécules plus spécifiques et avec une

activité cellulaire significative. La localisation ainsi que le mode de liaison de ces premiers chemotypes au sein de la protéine XRCC4 permettrait une optimisation rationnelle, basée sur la structure. L'expression d'un domaine XRCC4 soluble, centré sur la tête globulaire n'étant pas assez significative, la stratégie d'optimisation par RMN n'a pas été possible ici. Une approche par cristallographie a donc été envisagée, en sachant la structure cristallographique de XRCC4 et les conditions de cristallisation disponibles. Nous possédons des informations précises et issues de la littérature permettant de reproduire la cristallisation de la protéine XRCC4 seule ou bien en complexe avec le C-ter de la ligase 4 et dans des conditions de tampon optimales de stabilisation.

Deux approches seront envisagées dans cette thèse pour l'obtention d'une structure XRCC4/inhibiteur : le trempage de cristal et la co-cristallographie avec la molécule active (#57632, Kd 0.2mM). Eventuellement, la seconde molécule spécifique mais inactive *in vitro* (#4S-62686, Kd 1mM) pourra également être intégrée à cette étude structurale en vue d'une optimisation ultérieure si la localisation au sein de la protéine est adéquate. L'approche utilisant le ligand directement lors du processus de purification de la protéine n'est pas envisagée ici au vu de la faible affinité de ces derniers pour la cible.

Le travail a tout d'abord consisté à reproduire des cristaux de la protéine XRCC4 (aa 1-203) et d'exploiter ces cristaux pour un trempage avec l'inhibiteur #57632. Dans ce cas de figure, Plusieurs paramètres sont à considérer :

- quelle est la limite de concentration (et donc de solubilité dans le tampon de cryo-protection de la protéine) de la molécule que l'on peut utiliser en sachant que nous chercherons à effectuer un trempage dans une solution de ligand la plus concentrée possible ?
- Quelle sera par conséquent la concentration finale utilisée en DMSO et est-ce-que cela perturbera notre cristal (instabilité de la protéine) ?
- Quel sera le temps de trempage ?
- Le ligand étant large, cela peut-il perturber l'agencement cristallin ?
- Le packing cristallin permet-il une accessibilité des têtes à la molécule ?

Au vue de l'analyse du packing cristallin, la cristallisation de la construction XRCC4 seule semble avoir une accessibilité réduite pour notre inhibiteur mais les conditions de cristallisation ne nécessitent pas de muter cette construction (à l'inverse du complexe XRCC4/C-ter Ligase 4 dont l'arrangement spatial au sein du cristal nécessite de muter la construction afin de favoriser une liaison ionique venant stabiliser les protéines entre elles). Nous avons donc choisi de nous orienter en premier lieu sur XRCC4 seule et sauvage et de tenter tout de même l'expérience de trempage.

Les conditions d'expression et de purification de la construction XRCC4 pleine taille (aa 1-342) que nous possédons ne nous permettant pas d'atteindre une concentration de cette dernière au-delà de 1mg/ml, la stratégie est de co-exprimer et co-purifier la forme tronquée (aa 1-203) en présence du C-ter de la ligase 4 (voir partie Matériels et Méthodes). Le complexe étant très stable, le rendement ainsi que la concentration en complexe à l'issu du procédé de purification nous permet alors d'obtenir une protéine dans des conditions plus adéquates pour la cristallisation. Cependant, et dans le cas de la cristallisation de XRCC4 seule, la solution de cristallisation devra contenir du cacodylate de sodium. Ce sel permet de modifier de manière covalente les résidus cystéines présent dans notre complexe et de rompre l'interaction entre XRCC4 et le C-ter de la ligase 4 (Figure 87).

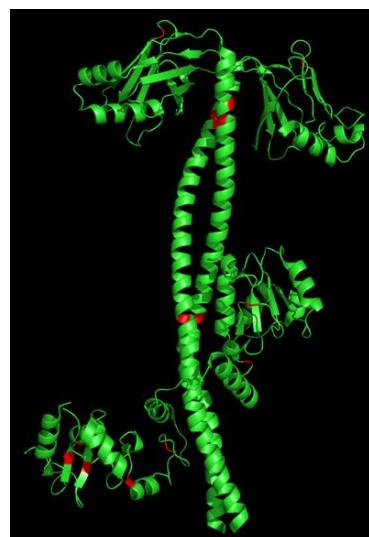


Figure 87: Structure cristallographique du complexe XRCC4/C-ter Ligase 4 (PDB code : 3II6). Les résidus cystéines modifiés par le cacodylate sont précisés en rouge.

Le C-ter ligase 4 étant instable lorsqu'il n'est plus en complexe (et particulièrement à une concentration élevée), celui-ci va alors précipiter et la protéine XRCC4 sera amenée à cristalliser. Les étapes de co-expression et co-purification ainsi que le rendement d'expression pour ce complexe protéique et les étapes et conditions de cristallisation sont explicités en parties II – 1.1 et partie Matériels et Méthodes. Ces conditions ont été reproduites (en plaques 24 puits) en faisant varier le pH ainsi que la concentration en agent précipitant, autour de la condition initiale connue. La méthode de diffusion en goutte suspendue a été utilisée. Après 72h, des premiers cristaux apparaissent et atteignent une dimension de 100 à 200µm de longueur après 48h supplémentaires (Figure 88).

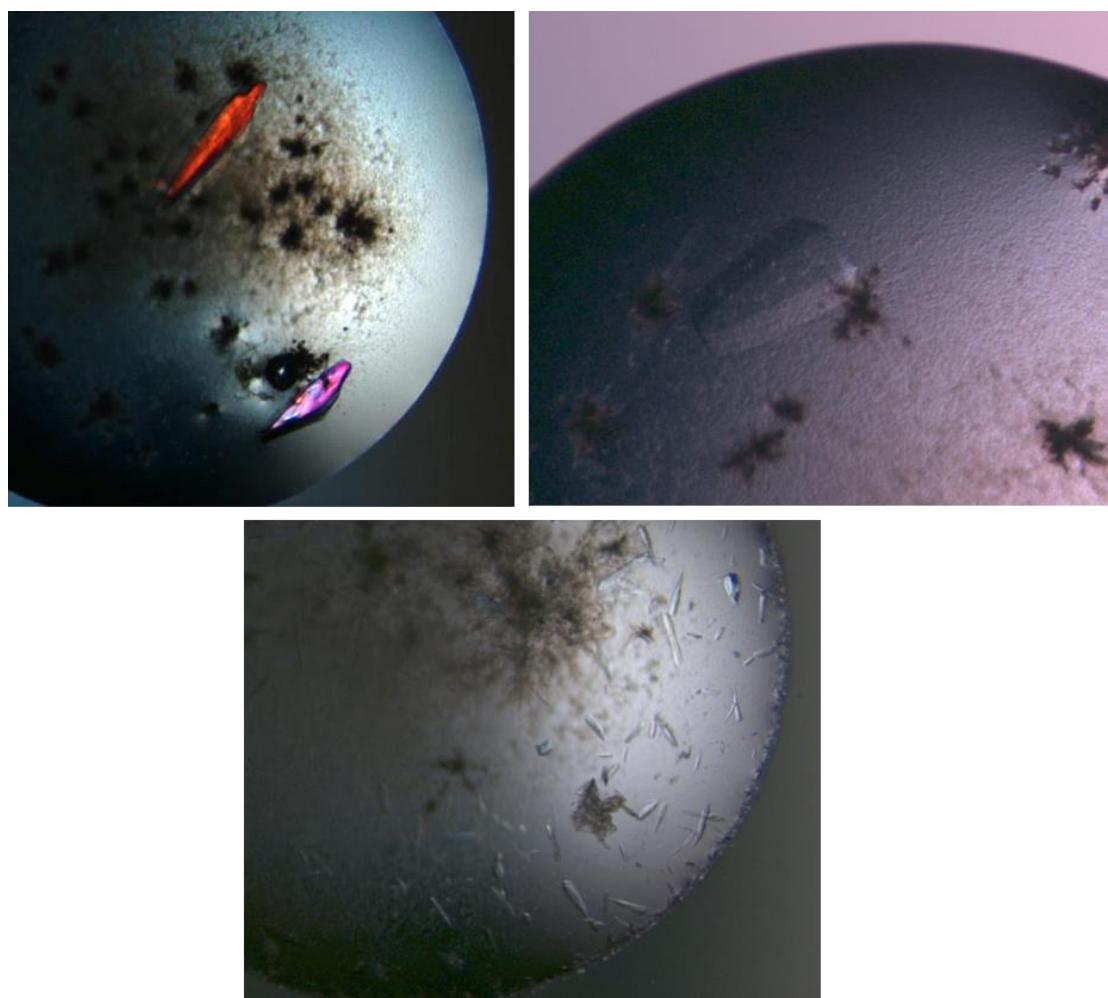


Figure 88: Cristaux de la protéine XRCC4 sous forme apo.

Les profils cristallins sont différents et les cristaux les plus gros et réguliers ont été choisis pour la diffraction de la forme XRCC4 apo et après trempage. De plus, Les

conditions permettant l'obtention de cristaux de taille importante et réguliers, ont été reproduites avec succès. Le cristal dont la taille était la plus importante a été utilisée pour plusieurs trempages successifs dans une solution de ligand #57632 (à une concentration de 1mM finale dans la solution de cryo-protectant). Cependant, malgré la faible concentration en ligand, le cristal a montré des fractures et une grande fragilité après seulement quelques secondes de trempage. (Ainsi, même une approche par trempage successif semblait compromise). A ce stade, il nous est impossible de déterminer avec précision si cette fragilité était due au DMSO lui-même ou bien au ligand. On aurait pu également tester le trempage dans une solution de stabilisation avant le trempage du ligand.

Des cristaux de la protéine seule et sans trempage ont tout de même été transférés dans une solution de cryo-protectant pour les faire diffracter et s'assurer que les conditions de purification, concentration et cryo-protection étaient suffisantes pour l'obtention de clichés de diffraction de moyenne à haute résolution. Une bonne résolution est effectivement indispensable afin de pouvoir visualiser sans ambiguïtés la présence ou non d'un ligand au sein de la protéine (tolérée jusqu'à environ 2.5 à 3Å pour ce type d'étude). Les données de diffraction, une fois collectée au synchrotron, ont été traitées de la sorte :

- Le programme XDS a permis d'indexer, intégrer et mettre à l'échelle les taches de réflexion.
- Le fichier .hkl produit a été converti en fichier .mtz via l'utilitaire XDSConv
- Ce fichier .mtz a ensuite été traité via le programme CCP4. Les différentes étapes ont été de confirmer le nombre de molécules par unité asymétrique (2 ici), d'effectuer une étape de remplacement moléculaire via l'utilitaire Phaser et en utilisant la structure PDB de XRCC4 1-203 (PDB code : 1FU1), et finalement, d'effectuer un affinement en corps rigide sur la solution unique trouvée par Phaser grâce à l'utilitaire Refmac 5 suivi de plusieurs cycles d'affinement sous contraintes.
- Après chaque cycle, les valeurs de Rfree et Rfactor diminuaient et convergeaient.

- Finalement, la structure obtenue à une résolution de 3.5Å a été manuellement optimisée via le programme Coot.

La résolution finale obtenue étant faible (3.5 Å), il fut difficile d'affiner chacune des chaînes latérales dont la majorité des densités étaient absentes. A cette résolution, il semble difficile d'envisager la localisation d'un ligand. Seule une information sur l'enveloppe générale de la protéine XRCC4 était présente au niveau de la carte de densité électronique (Figure 89). Cette information a pu nous permettre de confirmer l'agencement de la protéine au sein du cristal (groupe d'espace C 1 2 1) par rapport à la littérature.

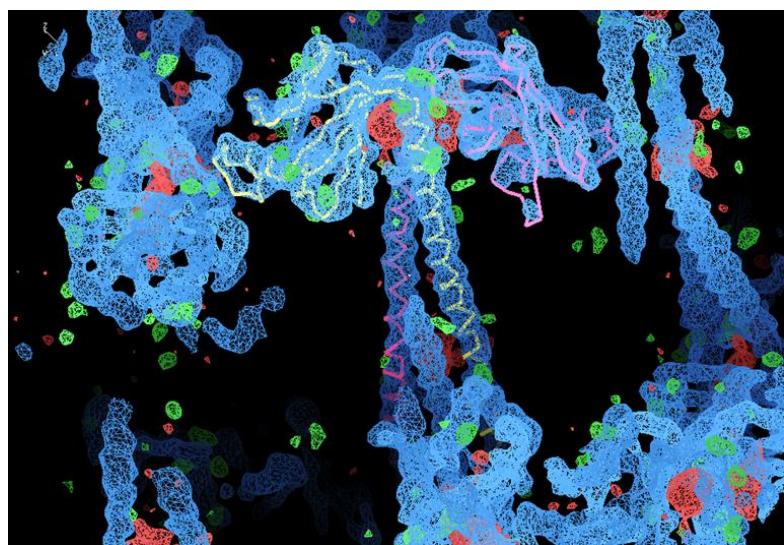


Figure 89: Carte de densité électronique de la protéine XRCC4 obtenue après remplacement moléculaire par le programme Phaser. La résolution obtenue est de 3.5Å.

Dans ces conditions, un paramètre à optimiser afin d'améliorer cette résolution peut être les conditions de cryo-protection. Ci-dessous, une liste dressant les statistiques de collecte et d'affinement des données pour la structure XRCC4 obtenue sous forme apo (Figure 90):

Data collection

Space group C 1 2 1
Cell dimensions a,b,c (Å) 164.4, 74.1, 88.5
Resolution (Å) 42.3 - 3.5
 R_{meas} (%) 8.7 (87.5)
 $C_{\text{c}_{\text{half}}}$ 99.8 (70.4)
 $I/\sigma(I)$ 9.06 (1.58)
Completeness (%) 99.2 (98.9)

Refinement

Resolution (Å) 42.3 - 3.5
No. of unique reflections 12077
 $R_{\text{work}}/R_{\text{free}}$ (%) 19/25
RMS bondLength 0.012
RMS bondAngle 1.92
Ramachandran statistics (%)
Favored regions 96%
Allowed regions 2.8%
Outliers 1.2%

Figure 90: Tableau de statistiques de collecte et d'affinement de données pour la structure cristallographique de XRCC4 apo obtenue après diffraction et résolution de structure par remplacement moléculaire.

II – 2.15 : Expériences de cocristallisation

Dans un second temps, des expériences de cocristallisation XRCC4/inhibiteur et complexe XRCC4/C-ter Ligase 4/inhibiteur ont été menées. Au-delà du fait que par cette méthode on pourrait s'affranchir du problème d'accessibilité du site de liaison (packing cristallin différent), la cocristallisation nous aurait également permis de visualiser des changements conformationnels potentiellement induits par l'inhibiteur, et que l'on n'aurait pas pu voir par une expérience de trempage.

Dans le cas de la cocristallisation XRCC4/inhibiteur, les conditions testées ont été les mêmes que les précédentes en plaque 24 puits. Pour cette expérience, un mélange complexe XRCC4/C-ter Ligase 4/inhibiteur est incubé au préalable (à une concentration finale en inhibiteur de 5mM) avant dissociation du complexe et cristallisation de la protéine XRCC4 (détails de la procédure en partie Matériels et Méthodes).

Dans le cas de la cocrystallisation XRCC4/C-ter Ligase 4/inhibiteur, cinq cibles de 96 conditions ont été testées (Qiagen : JCSG IV ; JCSG Plus ; PACT ; PEG Suite et PEG II). Pour cette expérience, un mélange complexe XRCC4/C-ter Ligase 4/inhibiteur est incubé (à une concentration finale en inhibiteur de 5mM) et ce complexe tripartite est directement intégré aux cibles de cristallisation, sans dissociation préalable entre XRCC4 et le C-ter Ligase 4. Un avantage d'utiliser ce complexe est que l'accessibilité au niveau des têtes de XRCC4 est moins réduite lors du packing cristallin (PDB code: 3II6). Cependant, les auteurs de cette structure ont obtenu des cristaux bien ordonnés et de diffraction satisfaisante en réalisant une mutation ponctuelle sur XRCC4 (A60E) favorisant le packing cristallin (Wu et al., 2009). Ne possédant pas cette forme mutée, nous avons donc commencé par cribler de nouvelles conditions de cristallisation afin de pouvoir obtenir des cristaux du complexe XRCC4/C-ter Ligase 4 sauvage sous forme apo et en complexe avec l'inhibiteur.

A ce jour, Aucun cristal provenant des deux stratégies de cocrystallisation n'a pu être obtenu. Il est cependant possible de réitérer ces expériences en testant d'autres cibles de cristallisation et/ou en mutant la protéine XRCC4 (Wu et al., 2009). De même, l'obtention de conditions robustes de cristallisation de ce complexe XRCC4/C-ter Ligase 4 sauvage permettrait aussi d'envisager des expériences de trempage avec des molécules fragments. Cette approche par fragment serait une stratégie rationnelle basée sur de l'optimisation par des méthodes comme le « growing » ou le « linking » et grâce à l'information de positionnement au sein de la structure.

II – 2.16 : Conclusions et perspectives – Vers la structure d'un complexe XRCC4/inhibiteur

Par une approche rationnelle, basée sur l'utilisation combinée d'un criblage virtuel (soutenu et validé par des simulations de dynamiques moléculaires) et de techniques biophysiques et biochimiques, nous avons caractérisé à l'issu de cette étude le premier inhibiteur de bas poids moléculaire de l'interaction protéine-protéine entre XRCC4 et Cer-XLF (l'approche basée sur des peptides ayant été ici infructueuse). Ce complexe représente pourtant un réel challenge au vu de l'interface concernée,

de l'absence d'inhibiteurs déjà connus et d'une quelconque activité enzymatique. Toutefois, la disponibilité des structures cristallographiques pour chacun des partenaires, la connaissance précise des résidus hot-spot impliqués dans cette interaction, l'affinité relativement faible entre les deux protéines (à l'inverse du complexe XRCC4/C-ter ligase 4) et l'implémentation des méthodes de criblage conventionnelles par des analyses de dynamiques moléculaires, sont tous des éléments rendant complètement envisageable l'inhibition de cette interaction par des approches rationnelles et via des inhibiteurs de bas poids moléculaires.

La molécule caractérisée (#57632), interagit de manière spécifique avec la protéine XRCC4 et inhibe la formation de ce complexe *in vitro* dans des expériences de stimulation de la ligase T4 et de NHEJ avec des extraits cellulaires. Elle possède une activité sub-millimolaire spécifique de ce mécanisme NHEJ et pourrait être par conséquent une base solide pour l'optimisation rationnelle par chimie médicinale d'un inhibiteur de plus haute affinité, basée sur la structure. Notre étude a donc permis d'apporter plusieurs éléments :

- Démontrer la faisabilité de l'inhibition de ce complexe de réparation de l'ADN, encore non explorée aujourd'hui.
- Valider ce complexe (et particulièrement la protéine XRCC4) comme toute nouvelle cible pharmacologique potentielle et qualifier cette protéine comme étant « druggable ».
- Fournir un premier composé, d'affinité modérée, pouvant maintenant servir de point de départ à l'élaboration de molécules plus affines et spécifiques (pour des tests cellulaires par exemple).

Aucune information structurale n'est à ce jour disponible concernant la localisation précise du composé #57632 au sein de la protéine XRCC4. Cette information essentielle a été néanmoins explorée au travers d'études structurales menées par des expériences de cristallographie. Les diverses approches explorées ont pu conduire à l'obtention de cristaux de la protéine XRCC4 sous forme apo mais toujours pas de structure de complexe XRCC4/inhibiteur #57632. Il reste encore plusieurs voies et conditions cristallographiques à explorer afin d'obtenir cette dernière.

Au-delà d'une résolution structurale, des approches, moins précises mais tout de même précieuses pourraient également être explorées. Par exemple, réaliser des expériences STD sur le domaine N-terminal isolé de XRCC4 ou en situation de compétition avec la protéine Cer-XLF apporterait une preuve de la localisation au sein de ce domaine. Des expériences de mutagénèse dirigée, permettraient également de déterminer un effet potentiel sur la liaison de la molécule et donc apporter indirectement une preuve de localisation.

D'autres pistes restent cependant à explorer. Des banques de molécules récemment créées et dédiées aux interactions protéine-protéine sont disponibles et pourraient être utilisées pour la recherche de modulateurs de l'interaction XRCC4/Cer-XLF. De même, l'approche par fragments, semble être une piste prometteuse dans le cas de cette cible prometteuse mais difficile.

II – 3 : Inhibition de l’interaction XRCC4/C-terminal de la Ligase 4

Résumé :

La très haute affinité entre les partenaires XRCC4 et C-ter ligase 4 ainsi que l’interface protéine-protéine flexible et très étendue de la ligase rend le criblage d’inhibiteurs potentiels pour cette interaction particulièrement difficile. Nous avons réalisé une étude pionnière visant à prévenir l’interaction entre ces deux partenaires. Une analyse quantitative de dynamique moléculaire sur la structure cristalline du complexe corrélée à une étude biophysique existante nous a permis de définir la région clamp de la ligase 4 comme cible privilégiée pour un criblage virtuel par docking haut-débit à partir d’une chimiothèque filtrée de 95.000 molécules. Ce criblage basé sur la structure du clamp et consolidé par des études de stabilité par dynamique moléculaire nous a conduits à la sélection de plusieurs molécules touches dont la validation par deux méthodes biophysiques orthogonales (RMN et DSF) a permis de confirmer ou d’infirmer la liaison de ces derniers sur le domaine C-terminal de la ligase 4. Suite à cette étude, nous avons pu isoler la première molécule capable d’interférer avec la formation *in vitro* du complexe XRCC4/C-ter ligase 4 et avec un IC50 < 1mM. Un modèle issu de la dynamique moléculaire permet d’entrevoir que ce composé d’affinité modérée se lie de manière stable au sein de la structure tripartite du domaine clamp (Hélice1-Loop-Hélice2) de la ligase 4 et au niveau de résidus critiques d’interaction avec XRCC4. Ce premier chemotype fournit un point de départ pour le design rationnel d’inhibiteurs analogues et de plus haute affinité. Cette étude démontre également la faisabilité de la prévention par une petite molécule, d’une interaction extrêmement affine entre XRCC4 et le C-ter de la ligase 4 et valide le domaine clamp comme domaine minimal de la ligase 4 pour un criblage *in silico*. La structure du complexe ligase4/inhibiteur est essentielle à la modification rationnelle de l’inhibiteur. Dans cette optique, et sachant le domaine C-terminal de la ligase 4 instable *in vitro*, nous avons donc tenté d’exprimer un domaine plus stable de cette dernière dont la taille, la stabilité et la limite de solubilité seraient parfaitement adaptés à une étude de marquage et de structure par RMN.

II – 3.1 : Simulations de dynamique moléculaire sur le complexe XRCC4/C-ter Ligase 4

De la même manière que pour l'interface entre XRCC4 et Cer-XLF, la stratégie pour l'inhibition de l'interaction entre XRCC4 et la partie C-terminale de la ligase 4 a été d'effectuer un criblage rationnel basé sur la structure de cette ligase. L'interface protéine-protéine entre XRCC4 et le C-ter de la ligase s'étend sur environ 2900Å² (Wu et al., 2009) et cette interaction résiste à de fortes concentrations en sels ou urée (Critchlow et al., 1997) (Grawunder et al., 1997) voire même, à des solvants organiques (McFadden et al., 2014). Cependant, le chapitre introductif évoque la preuve de concept de l'inhibition de ce complexe et apporte la preuve expérimentale d'une augmentation de la radiosensibilisation des cellules cancéreuses suite à cette inhibition. Ainsi, et au vue de la forte affinité entre les deux protéines (Kd nanomolaire), isoler des molécules capables de se lier spécifiquement sur la ligase 4 pourrait prévenir l'association de la protéine XRCC4 et conduirait à un mécanisme de radiosensibilisation cellulaire. La preuve de concept d'un effet thérapeutique (radiosensibilisation) par l'inhibition cellulaire de la ligase 4, couplé à la caractérisation de molécules « Drug-Like » inhibitrices démontreraient pour la première fois, la druggabilité de cette interface.

Afin d'effectuer une analyse quantitative des interactions dynamiques entre les résidus impliqués dans l'interaction XRCC4/C-terminal de la ligase 4, une simulation de dynamique moléculaire (DM) de dix nanosecondes a été effectuée sur ce complexe dont la structure cristallographique a été extraite de la PDB (PDB code 3II6). A l'issue de cette simulation, nous avons analysé l'évolution du RMSD (Root Mean Square Deviation) pour chacun des partenaires ainsi que les contacts intermoléculaires persistants. L'analyse des 7 dernières nanosecondes de cette simulation (après équilibration du système) a permis de dresser un histogramme des pourcentages de contacts intermoléculaires pour des résidus de la partie C-terminale de la ligase 4 (XIR + BRCT2) avec les chaines A et B de XRCC4 (Figure 94), et à partir d'un programme maison. Tout comme pour le complexe XRCC4/Cer-XLF, un seuil fixé à 90% a permis d'établir quels étaient les résidus à contact très persistant au cours des 7 dernières nanosecondes de la trajectoire et donc, dans cette interaction.

Cette analyse *in silico* souligne l'importance accrue de certains de ces résidus (pouvant être des résidus « Hot-spots », les plus énergétiquement favorables) dans la stabilité de ce complexe. Il serait intéressant cependant, de pouvoir confirmer cette analyse par des expériences de type « Alanine-scanning mutagenesis » (mutagénèse dirigée sur ces résidus et ITC comparative).

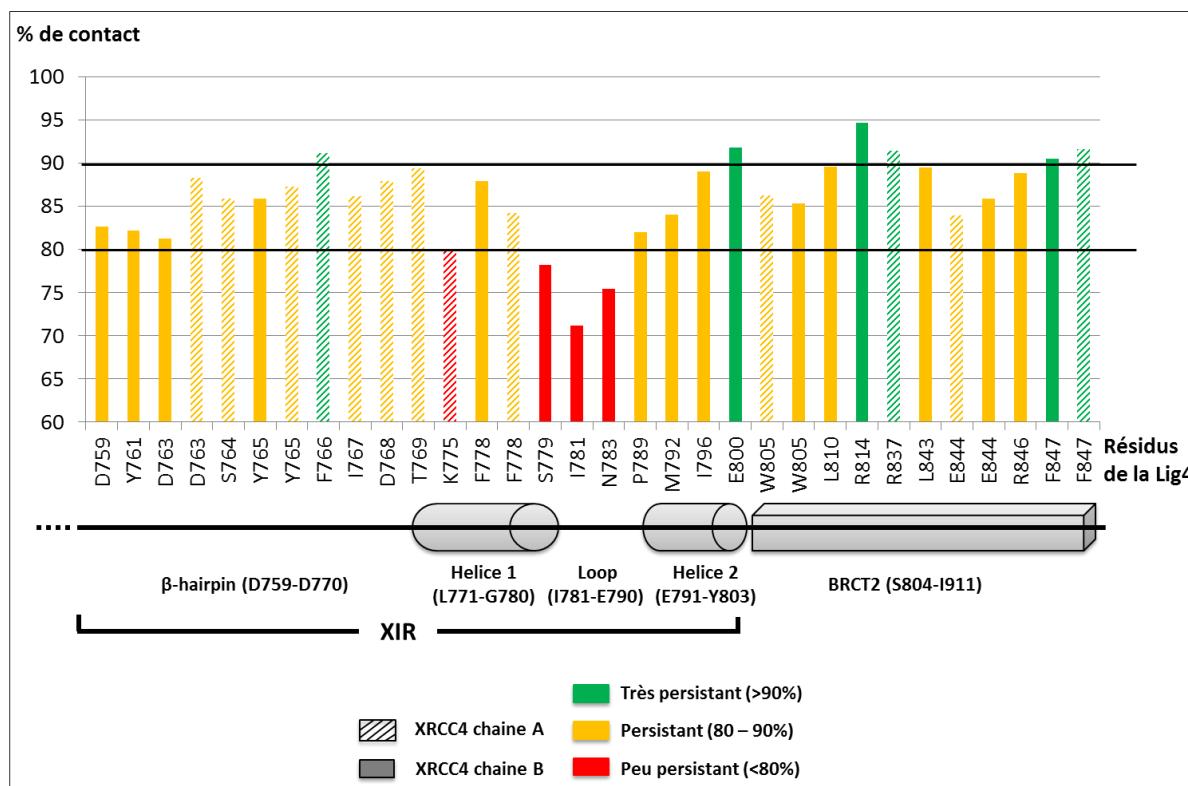


Figure 94: Analyse des contacts persistants entre les chaînes A et B de XRCC4 et la partie C-terminale de la ligase 4 (654-911) au cours d'une trajectoire de DM de dix nanosecondes.

Sur cet histogramme sont représentés uniquement les résidus de la ligase 4 qui interagissent avec XRCC4. Les résidus très persistants (qui atteignent un seuil de 90%), sont représentés en vert sur l'histogramme. Les résidus moins (80-90%) et peu (<80%) persistants sont représentés en jaune et rouge respectivement. Le domaine clamp (771-803) interagit presque exclusivement avec la chaîne B de XRCC4 et de façon plus persistante au niveau de l'hélice 2 (791-803). La majorité des contacts très persistants se localise au niveau du domaine BRCT2 indispensable à la stabilisation *in vivo* de la ligase sur XRCC4.

Sur cet histogramme, un code couleur permet de discriminer les contacts très persistants (en vert) des contacts moins persistants (en jaune) et peu persistants (en rouge). Après analyse des résultats obtenus, les résidus Phe766 (β-hairpin), Glu800 (Hélice2), Arg814, Arg837 et Phe847 (domaine BRCT2) représentent les contacts les

plus persistants avec XRCC4. Ces contacts prédisés comme les plus énergétiquement favorables dans l'association XRCC4/C-ter ligase 4 sont surtout présents sur le domaine BRCT2 (S804-I911). Ce résultat appuie l'hypothèse selon laquelle le domaine BRCT2 est critique pour la stabilisation du domaine C-terminal de la ligase 4 sur XRCC4 (Wu et al., 2009). Cependant, des expériences cellulaires consistant en une expression ectopique de fragments de la ligase 4 ont permis de montrer que le domaine BRCT2 était cependant incapable, seul, de prévenir la formation du complexe endogène XRCC4/Ligase 4 (Wu et al., 2009). De plus, une étude biophysique récente avec des fragments de la Région d'Interaction à XRCC4 (XIR), montre qu'un déplacement compétitif du XIR (D759-Y803) sur XRCC4 n'est significatif qu'avec l'hélice 2 du clamp ou des fragments contenant cette hélice 2 (Figure 95) (McFadden et al., 2014).

Notre simulation confirme cette étude en montrant qu'au sein du domaine clamp de la ligase (hélice1 – boucle – hélice2 ; L771-Y803), les contacts les plus persistants se localisent surtout au niveau de l'hélice 2 (4 résidus contre un seul sur l'hélice 1). Nous validons ainsi notre approche virtuelle par comparaison avec une étude biophysique menée sur ce complexe et qui démontre l'importance du clamp de la ligase 4. Par ailleurs, et au vu d'une structure plus adaptée au docking que les autres parties interagissant avec XRCC4, ces résultats nous orientent donc vers un criblage virtuel centré sur ce domaine clamp de la ligase 4.

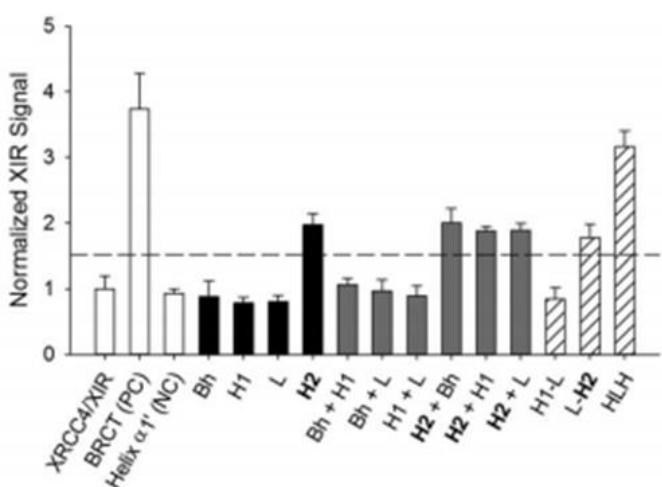


Figure 95: Essai de déplacement compétitif du domaine XIR sur XRCC4 par des fragments du XIR (d'après McFadden et al., 2014).

II – 3.2 : Dynamique moléculaire et échantillonnage conformationnel de la structure clamp en solution

Ci-dessous est représentée la structure du domaine clamp (771-803) de la ligase 4 (Figure 96) et la persistance de certains de ses résidus identifiée sur l'histogramme précédent (issu de la structure PDB : 3II6, en complexe avec XRCC4).

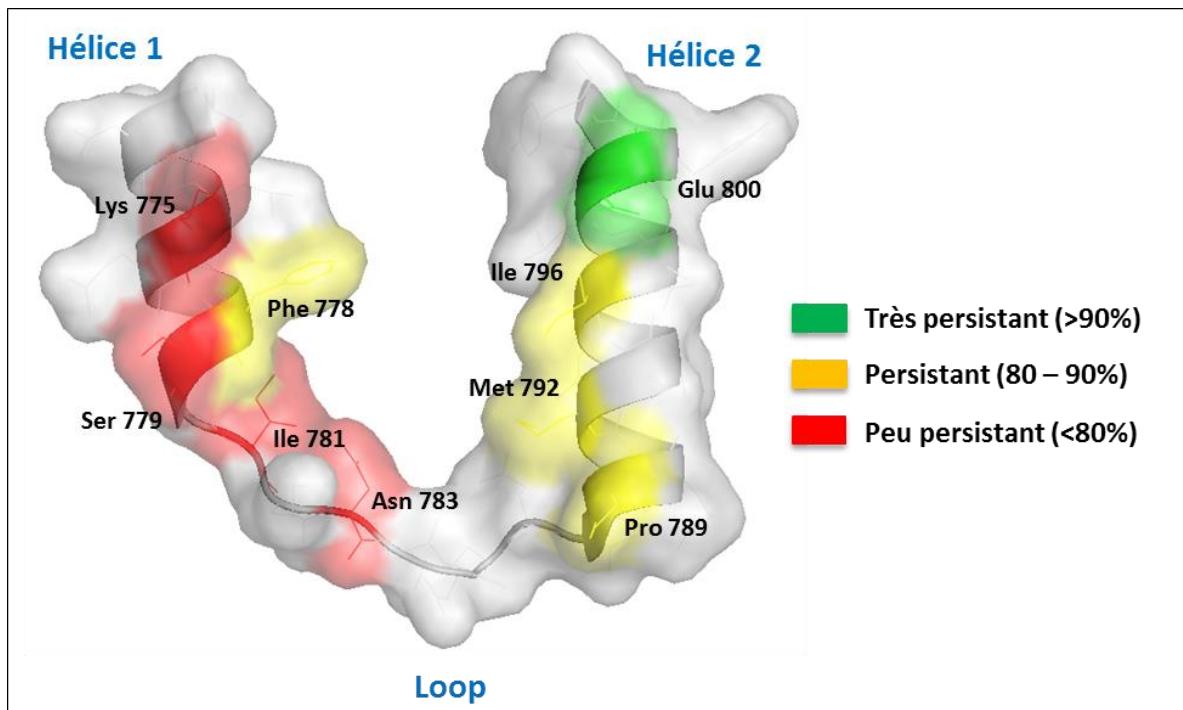


Figure 96: Domaine clamp (771-803) de la ligase 4 représenté en surface avec la persistance des résidus d'interaction à XRCC4 déterminée par DM.

Ce domaine clamp est formé par deux hélices amphipatiques reliée par un linker (« loop ») majoritairement hydrophile. On devine qu'un criblage orienté sur cette interface n'est pas simple au vu de la surface large, plane et exposée au solvant de ce clamp. D'autant plus qu'une structure de ce domaine non complexé à XRCC4 n'est pas disponible. Néanmoins, il est possible d'envisager que des molécules de taille moyenne à importante (au-delà de 300Da, limite définissant la taille moyenne maximale d'un fragment dans la « règle des trois »), se stabilisent au sein de cette structure en venant contacter des résidus hydrophobes sur chacune des hélices et en formant des contacts polaires avec le linker. En particulier, les résultats de la simulation précédente ainsi que ceux de Mac Fadden (McFadden et al., 2014) nous conduisent à favoriser ultérieurement des molécules formant des contacts persistants et plus particulièrement avec des résidus de l'hélice 2 du clamp.

Au-delà de la fonction de score, une limitation majeure rencontrée lors du processus de docking est que nous ne possédons pas toutes les conformations du récepteur. Un faible échantillonnage conformationnel est d'ailleurs l'une des raisons de l'obtention d'un taux très faible de vrais positifs (molécules touchées expérimentalement validées). De plus, en sachant que la structure du C-terminal de la ligase 4 a été obtenue par co-cristallographie avec son partenaire XRCC4, nous devions évaluer le comportement de ce domaine isolé de XRCC4 en termes de flexibilité et dans un environnement solvaté. Nous avons donc choisi d'utiliser la méthode RCS (« Relaxed Complex Scheme »). Cette méthode permet de réaliser plusieurs étapes de docking sur un ensemble de structures représentatives du récepteur, extraites d'une simulation de DM (Lin et al., 2003).

Cet échantillonnage conformationnel a donc été réalisé sur le domaine C-terminal par une DM de 15 nanosecondes (Figure 97) afin d'en extraire différentes structures représentatives du clamp et d'effectuer un docking sur chacune de ces structures. La DM a été réalisée dans une boîte périodique solvatée et nous avons conservé les domaines BRCT1 et BRCT2 pour tenir compte d'un effet éventuel de contraintes structurales sur le clamp.

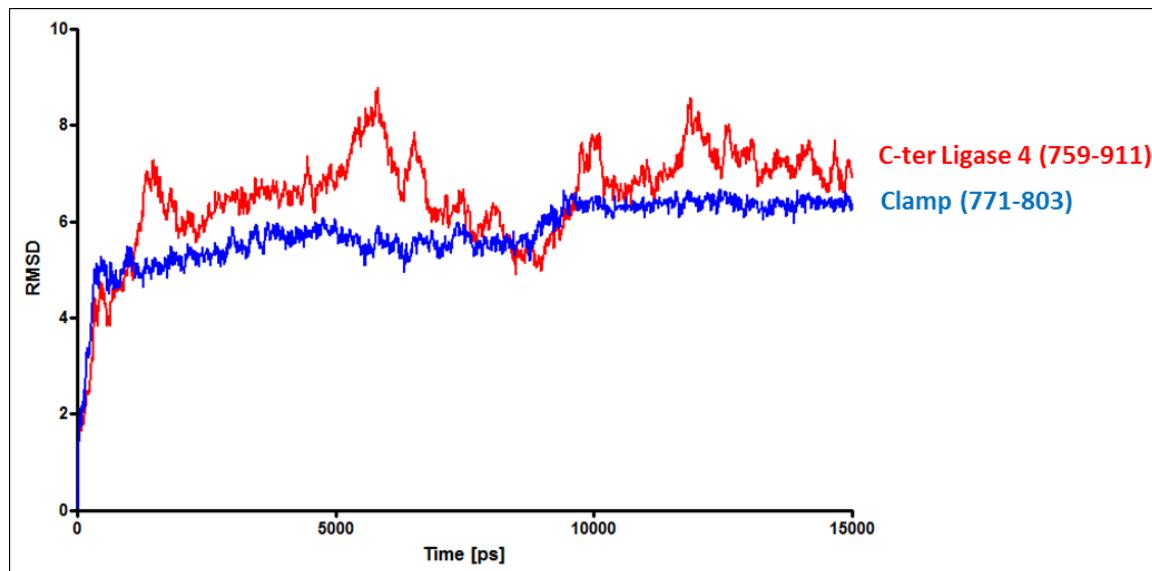


Figure 97: Simulation de DM de 15ns sur le C-terminal de la ligase 4 – Analyse du RMSD pour le C-terminal pleine taille et pour le domaine clamp.

L'observation du RMSD en fonction du temps montre un profil très fluctuant pour le C-terminal complet par rapport au clamp seul, dû à la mobilité des deux domaines BRCT. On observe un RMSD très stable au cours du temps pour le domaine clamp qui se stabilise très rapidement autour de 5Å. Celui-ci subit néanmoins un changement conformationnel important autour de 10ns et se stabilise à nouveau dans cette nouvelle conformation autour de 6Å. Les 700 premiers cadres de la trajectoire (correspondant aux 7 premières nanosecondes) ont été analysés afin de clustérer et extraire les structures représentatives du clamp. Avec une valeur seuil pour le clustering fixée à 2Å (c'est-à-dire que l'on extrait un cluster de structures indépendantes lorsqu'une déviation sur les carbones alphas de 2Å est introduite par rapport au cluster précédent), nous avons pu extraire au total 11 centroïdes ainsi que leur structure représentative respective (Figure 98).

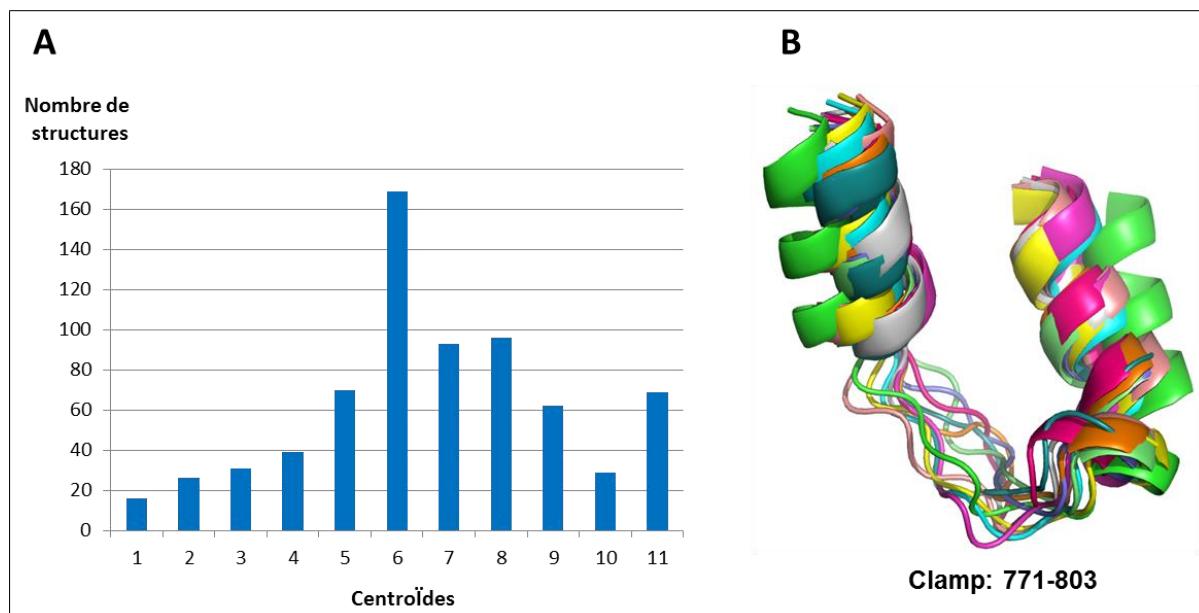


Figure 98: Simulation de DM sur le C-terminal de la ligase 4 (654-911) – Analyse des centroïdes pour le domaine clamp (771-803).

(A) Analyses des 11 centroïdes obtenus et leur population de conformères. Le centroïde 6 est largement majoritaire mais l'ensemble des centroïdes sera considéré. (B) Superposition des 11 structures représentatives du domaine clamp de la ligase 4 et qui seront utilisés pour l'étape de docking rigide. On note un rapprochement progressif entre les deux hélices.

Au cours de cette simulation, et la ligase n'étant plus en complexe avec la protéine XRCC4, on observe un rapprochement prédictible entre les deux hélices via des interactions hydrophobes entre les résidus F778 de l'hélice 1 et L799 de l'hélice 2. Ce rapprochement crée un nouveau lien permettant la formation d'une cavité qui

pourrait être parfaitement adaptée pour accommoder des petites molécules inhibitrices (Figures 98 et 101). Ces molécules pourraient alors stabiliser cette conformation et prévenir l'interaction avec XRCC4. Ce rapprochement atteint néanmoins un seuil (autour de 10ns sur la figure précédente) après lequel le clamp effectue un léger changement conformationnel au niveau de l'hélice 2. Suite à ce changement, le domaine clamp est moins accessible pour une petite molécule. C'est pour cette raison que les 700 premiers cadres uniquement ont été considérés pour cette étude de clustering. Les chaines principales des 11 structures représentatives issues de chacun de ces centroïdes ont été superposées sur la figure ci-dessus et l'étape de pré-criblage virtuel par docking sera effectuée sur chacune d'entre elles. Elles représentent les conformations stables du C-terminal de la ligase 4 les plus probables lorsque XRCC4 n'est pas associée.

II – 3.3 : Préparation et filtrage d'une base de données virtuelle de molécules

Tout comme pour le chapitre précédent, les critères de sélection (molécules disponibles commercialement) et préparation des molécules virtuelles sont identiques. Une des exceptions pour ce criblage est que nous avons choisi de travailler sur une sous-banque de la chimiothèque ZINC (« ZINC-clean-Drug-like Library ») (Irwin and Shoichet, 2005). Cette sous-banque contient plusieurs millions de molécules de 150 à 500Da environ et avec une grande diversité chimique. Nous avons également inclus au sein de ce jeu de données virtuelles, des fichiers correspondant à des molécules appartenant à notre chimiothèque propre. Cette chimiothèque maison permet d'enrichir la sous-banque ZINC utilisée et contient des molécules dont la synthèse est parfaitement maîtrisée et dont des dérivés peuvent être facilement obtenus. Le filtre utilisé par la suite sur l'ensemble de ces molécules contient notamment le paramètre SFI (Solubility Forecast Index) (Hill and Young, 2010). Les critères de ce filtre sont précisés dans le tableau ci-dessous (Figure 99).

0	450	# Molecular weight limits
-3	1	# logP
-3	1	# Solubility Forecast Index (logD + #Ar)
0	110	# Polar Surface Area
0	5	# Hydrogen Bond Donors
0	5	# Hydrogen Bond Acceptors
0	6	# No. of rotatable bonds
0	2	# No. of ring systems
0	2	# No. of rings
0	5	# No. of "N" & "O" atoms
-1	1	# Charges

Figure 99: Tableau représentatif des différents critères utilisés pour le filtrage de notre base de données virtuelle de molécules, destinée au criblage par docking sur le clamp de la ligase 4.

Là encore, le programme Screening Assistant 2 (Guilloux et al., 2012) a été utilisé dans le but de supprimer les molécules doublons, toxiques et les faux-positifs connus. La taille des molécules a été tolérée jusqu'à 500Da afin de permettre à ces dernières le contact simultané des différentes parties du clamp (hélice 1-loop-hélice 2). Après utilisation de ce filtre ainsi que du programme Screening Assistant, un total d'environ 95.000 molécules a été obtenu. Ces molécules ont été utilisées pour l'étape de pré-criblage par un docking rigide sur les 11 structures représentatives du clamp grâce à AutoDock Vina (Trott and Olson, 2010).

II – 3.4 : Docking sur les différentes structures du clamp – Tri et analyse des hits potentiels

Un alignement de séquences a été effectué (Figure 100) afin de s'assurer que le C-terminal de la ligase 4 que nous possédons physiquement correspond parfaitement à la protéine sur laquelle nous lançons le programme de docking (PDB : 3II6 – C-terminal co-cristallisé avec XRCC4) et plus particulièrement au niveau de la zone d'intérêt qui concerne le domaine clamp (résidus 771-803) :

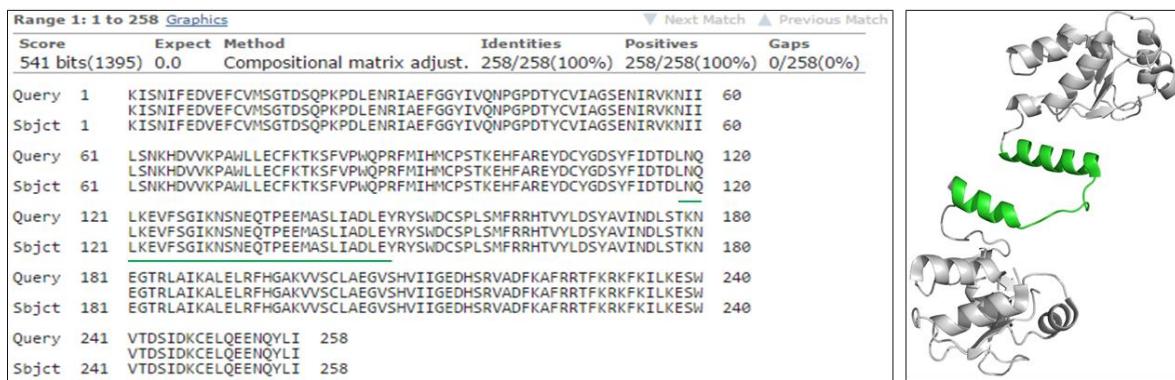


Figure 100: Alignement de séquence par le programme BLAST pour la protéine C-ter Ligase 4.

La séquence protéique de la ligase utilisée expérimentalement au cours de ma thèse est comparée à la séquence de la protéine co-cristallisée 3II6. La séquence 771-803 (en vert) représente la zone considérée pour le criblage. Aucune mutation n'est présente dans la zone ciblée.

Ci-dessus, la séquence « Subject » correspond à l'insert C-terminal de la ligase 4 contenu dans le plasmide PWY1119 que nous possédons (voir partie Matériels et Méthodes). La séquence « Query » est celle de la protéine dans le co-cristal 3II6. La séquence d'intérêt (771-803 représentée en vert) ainsi que le reste de la séquence est parfaitement identique dans les deux cas.

La DM du domaine clamp de la ligase a été réalisée en prenant en compte les domaines BRCT1 et 2. Cela permet de conserver des effets éventuels de contraintes par ces domaines sur la structure clamp. Ces domaines ont été cependant enlevés durant l'étape de docking rigide centré uniquement sur ce domaine clamp. Ils ont été à nouveau conservés lors de l'étape ultérieure de validation des molécules touchées par DM pour ne pas biaiser le résultat et montrer la stabilité et la spécificité de nos candidats potentiels au sein de la structure clamp. Durant l'étape de docking, le récepteur a été considéré comme rigide et les ligands complètement flexibles. Cela a permis notamment de prendre en compte la sélection d'un conformère particulier par le ligand. La flexibilité des complexes protéine/ligand a ensuite été réintroduite durant l'étape ultérieure de DM.

A l'issu de ce docking et pour chacune des 11 structures, les molécules ont été classées selon leur score attribué par le programme Autodock Vina (Trott and Olson, 2010). Une valeur maximale de -7kcal/mole a été fixée. L'étape de sélection suivante

a consisté à analyser visuellement chacune de ces molécules sur leur structure clamp respective (environ 200 molécules). Deux critères ont été choisis : nous avons conservé uniquement les molécules formant au moins un contact polaire et un contact hydrophobe avec le clamp (et particulièrement avec les résidus critiques identifiés précédemment par DM - Figure), et formant un seul cluster de positions au sein de ce dernier (Figure 101).

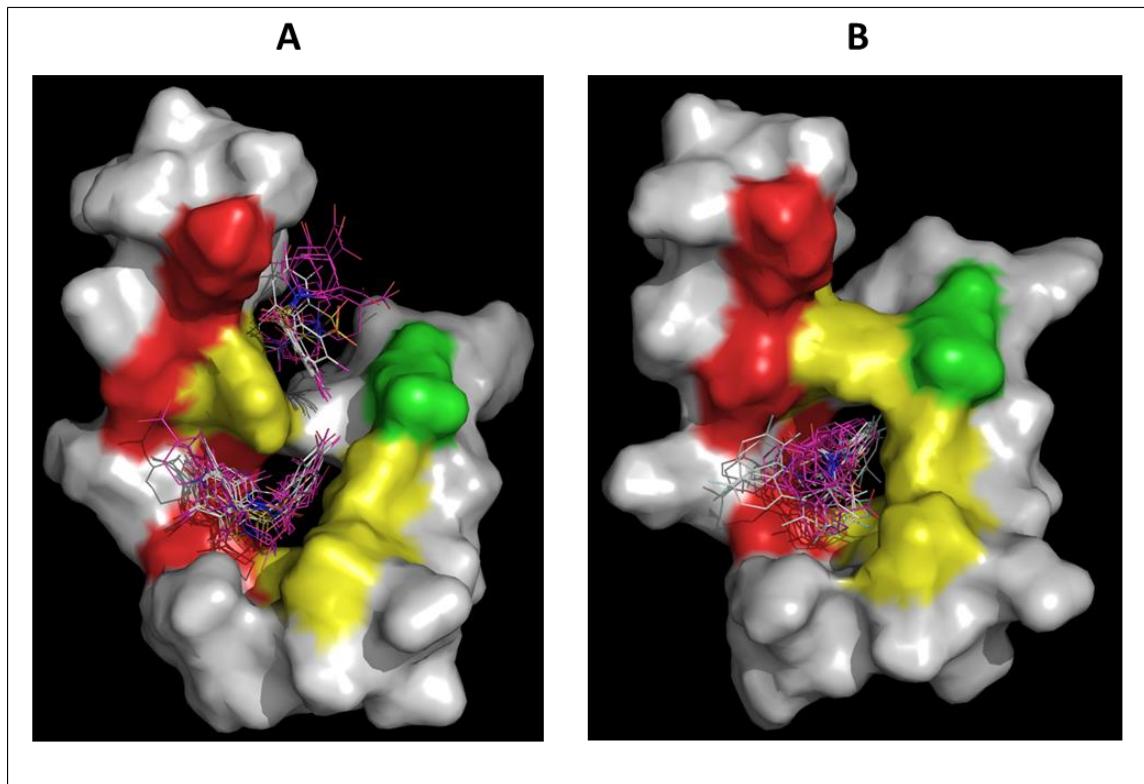


Figure 101: Docking rigide et clustering pour deux molécules au sein de leur structure représentative du domaine clamp (771-803) – Superposition des différents conformères stables à l'issu du docking.

Le domaine clamp est représenté en surface avec le code couleur de la persistance des résidus critiques définis précédemment. (A) Le ligand possède deux clusters distincts au sein du clamp. (B) Le ligand se loge dans une cavité issue du rapprochement des deux hélices du clamp et ne forme qu'un seul cluster au niveau de cette dernière et des résidus critiques d'interaction à XRCC4.

La figure ci-dessus montre le clustering positionnel pour deux exemples de ligands à l'issus du docking rigide sur deux conformations différentes du clamp (Ligands A et B). Pour les deux ligands, 10 à 20 conformères ont été superposés. Le ligand B (molécule #93291), a atteint un score d'au moins -7kcal/mole pour sa conformation la plus stable et forme un seul cluster localisé et formant des contacts notamment avec

des résidus critiques d'interaction avec XRCC4 (en jaune sur la structure). La molécule B sera donc favorisée par rapport à la molécule A.

Au total, 29 candidats ont été retenus comme étant des molécules potentielles d'interaction avec le domaine clamp de la ligase 4. La conformation considérée comme représentative (car étant la plus énergétiquement stable d'après le score de docking) pour chacune de ces 29 molécules a été extraite et le complexe protéine-ligand correspondant a été analysé ultérieurement par DM (stabilité au cours du temps, contacts persistants, énergie de liaison). Ces 29 inhibiteurs potentiels de l'interaction entre XRCC4 et la ligase 4 se localisent tous au sein du clamp, qui forme une cavité due au rapprochement des deux hélices par contacts hydrophobes. Le détail de la meilleure pose pour l'une de ces molécules est donné dans la figure suivante (Figure 102) et explique le score observé durant l'étape de docking.

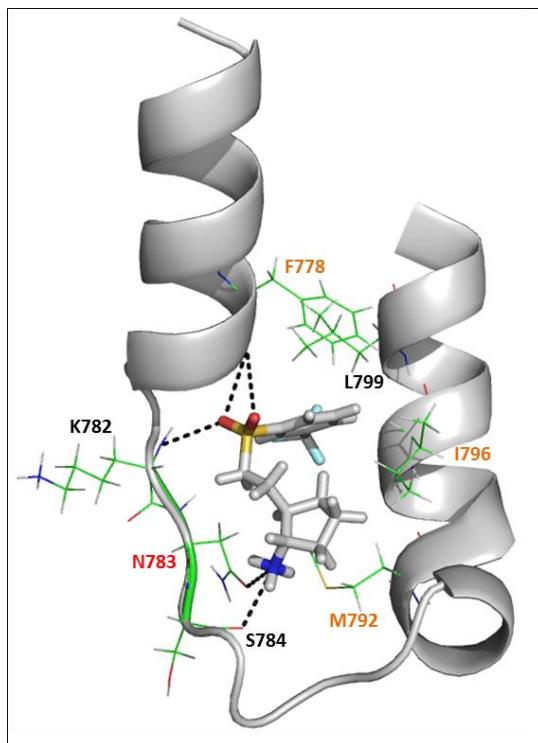


Figure 102: Docking rigide sur une des structures représentatives du domaine clamp (771-803) obtenu par DM – conformation stable et détails des interactions du candidat #231 avec sa cible.

Parmi les résidus contactés par le candidat, F778, M792 et I796 (en orange), sont des résidus critiques d'interaction avec XRCC4. Les liaisons hydrogènes sont représentés en pointillés noir.

Le candidat ci-dessus (molécule #231) est un des 29 candidats sélectionnés à l'issu de l'étape de docking et possède un score au-delà de -7kcal/mole s'expliquant par de nombreuses interactions. Notamment, un réseau important de liaisons hydrogène s'établie ici entre le dioxyde de soufre du ligand et l'amide de la chaîne principale du résidu K782 ainsi qu'entre le groupement NH3 du ligand et les chaînes latérale et principale des résidus respectifs N783 et S784. De plus, le ligand se loge parfaitement dans une cavité hydrophobe formée par les chaînes latérales des résidus M792, I796, F778 et L799. Finalement, l'interaction hydrophobe entre F778 et L799 « referme » le clamp sur lui-même, créant ainsi une cavité dans laquelle le ligand se stabilise. Les résidus identifiés précédemment comme critiques dans l'interaction avec XRCC4 (notamment F778, I796, M792 ci-dessus en orange) interviennent dans cette stabilisation du ligand. La plupart de ces caractéristiques ont été retrouvées pour les molécules sélectionnées et ont servi à établir la liste des 29 candidats potentiels.

La prochaine étape a consisté à évaluer par DM la stabilité et l'énergie de liaison théorique pour chacun des complexes C-ter Ligase 4/molécule. On rappelle que cette évaluation ultérieure est critique car elle permet de réintroduire une dynamique au sein du complexe et dans un environnement solvaté. Le docking est en effet un outil puissant qui permet d'émettre une hypothèse quant au positionnement d'une molécule (d'un « hit ») au sein d'une cavité mais ne peut pas être utilisé pour une détermination exacte des énergies de liaison et de la stabilité des complexes. De plus, il ne permet pas d'observer par exemple, une adaptation de la protéine suite à la liaison du ligand (phénomène d'ajustement induit).

II – 3.5 : Analyse par dynamique moléculaire de la stabilité et de l'énergie de liaison des complexes C-ter Ligase 4 hit sélectionnés – Nouveaux classements et modèles d'interaction

Rappelons qu'un échantillonnage conformationnel par DM a permis de docker notre chimiothèque filtrée de 95.000 molécules environ sur différentes structures représentatives du clamp (méthode du Relaxed Complex Scheme), augmentant ainsi théoriquement nos chances de trouver une molécule touche. Pour ne pas biaiser le résultat, l'échantillonnage de la structure clamp a été fait en conservant les deux

domaines BRCT (conservant ainsi l'influence qu'ils pourraient avoir en termes de contraintes structurales sur le clamp).

Le docking a été centré sur ces 11 structures représentatives (clamp sans les domaines BRCT) et les 29 complexes clamp hit retenus ont été soumis à une autre étape de sélection par des DM de 10 nanosecondes. Ces simulations ont, à nouveau, été lancées sur le C-terminal entier de la ligase 4 afin de se rapprocher le plus possible des complexe protéine/hit en solution. Là encore, à l'issue de ces dynamiques, chaque candidat est jugé sur sa stabilité au cours de la trajectoire (mesure du RMSD au cours du temps), son mode de liaison, son énergie théorique de liaison en kcal/mole (programme MM-PBSA, explicité en partie Matériels et Méthodes) et ses contacts persistants avec des résidus clés du clamp. Un contact a été évalué comme persistant s'il a atteint un seuil de 90% sur les 2 dernières nanosecondes de la trajectoire. Les évènements d'ajustements induits sont également analysés. Les meilleures molécules (RMSD stable, bonne énergie de liaison, présence de contacts persistants avec des résidus critiques d'interaction avec XRCC4) sont commandées et intégrées dans nos tests de validation biophysique et biochimiques, dans la mesure de leur disponibilité et de leur solubilité en tampon aqueux.

A l'issu de la trajectoire de DM pour chacun des complexes, les derniers cadres de chacune de ces trajectoires sont extrait et la structure du complexe correspondant et considérée comme représentative. Ci-dessous est illustré un exemple d'analyse détaillée pour un complexe C-ter Ligase 4/hit stable (molécule #231) et appartenant au cluster des meilleurs hits théoriques à l'issu de l'analyse par DM (Figure 103).

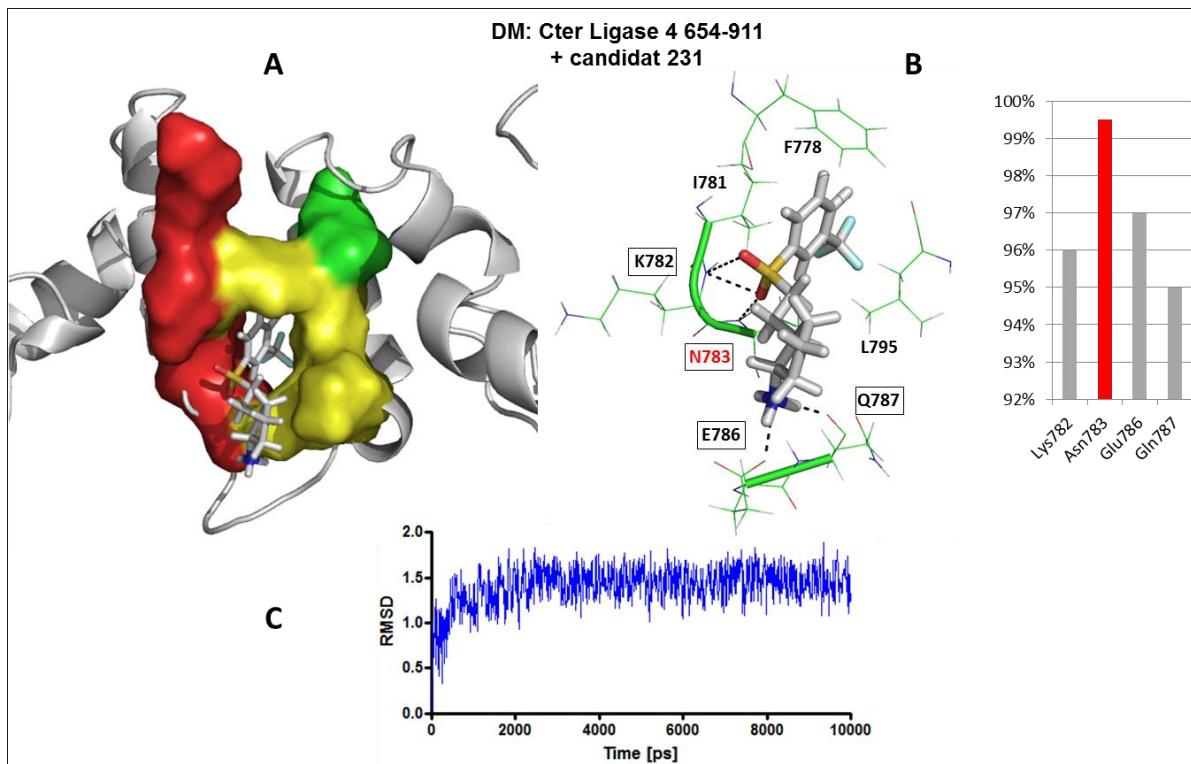


Figure 103: Analyse par DM de la stabilité du candidat #231 issus du docking sur le clamp de la ligase 4 (771-803).

La stabilité de chacun des inhibiteurs est évaluée après analyse du RMSD et des contacts persistants avec la cible protéique. (A) Extraction du dernier cadre de la trajectoire de DM et analyse du mode de liaison au sein du clamp pour le candidat #231. Le résidu en rouge N783 est un des résidus d'interaction avec la protéine XRCC4. Les résidus encadrés en noir sont les résidus avec lesquels le ligand forme des contacts persistants. (B) Histogramme des contacts persistants au-delà de 90% et résidus concernés correspondants. (C) Courbe RMSD du ligand au cours de la trajectoire de dix nanosecondes. Avec des fluctuations $\leq 1\text{\AA}$, ce candidat reste très stable et localisé au sein du site concerné.

Cette analyse effectuée pour chacun des 29 complexes C-ter Ligase4 hit est résumée dans l'histogramme suivant (Figure 104), reportant l'énergie de liaison et la persistance au site d'intérêt pour chacune des molécules. Une molécule est classée dans les contacts persistants (en bleu sur l'histogramme) si au moins un contact persistant au sein du site d'intérêt est présent durant toute la trajectoire avec un résidu critique d'interaction avec XRCC4. Cet histogramme permet de reclasser les 29 candidats via leur énergie théorique de liaison et de façon plus pertinente que d'après le score obtenu par docking avec Vina.

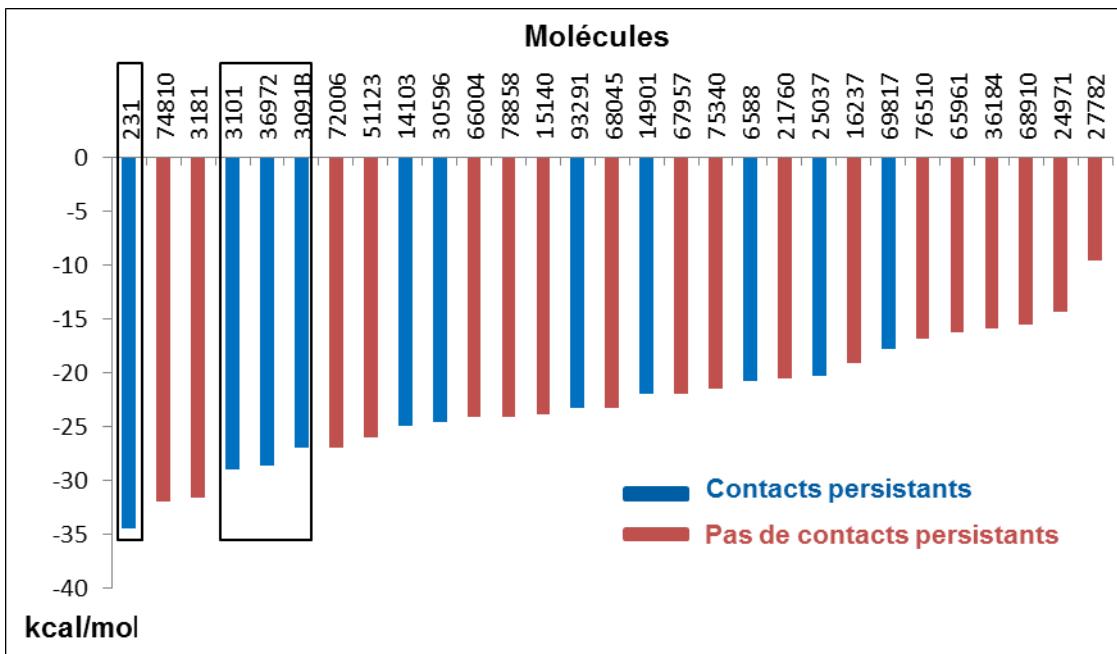


Figure 104: Analyse de l'énergie de liaison théorique après DM pour chaque candidat du clamp de la ligase 4 (771-803) – Implémentation par les contacts persistants.

L'énergie de liaison théorique en kcal/mol est évaluée via la trajectoire de DM par le programme MM-PBSA. Les molécules sont classées par ordre croissant d'énergie. Les molécules encadrées en noir ont été physiquement disponibles et ont été intégrées aux tests de validation expérimentale.

Cependant, les molécules les plus persistantes ici (11 sur 29, en bleu) n'ont pas forcément en moyenne les meilleures énergies théoriques de liaison. C'est pour cela qu'il est important de ne pas considérer uniquement cette énergie théorique mais surtout favoriser les molécules formant des contacts persistants avec la protéine cible et ayant un RMSD stable au cours du temps. Afin d'évaluer cette stabilité des hits au cours du temps au sein du domaine clamp, nous avons attribué un score purement qualitatif pour chaque RMSD (Figure 106). Le score est de 3 pour une molécule très stable (fluctuations $\leq 1 \text{ \AA}$) ; un score de 2 est pour un RMSD plus fluctuant (mais conservation par le hit d'un même mode de liaison) ; Et finalement un score est de 1 pour un RMSD très fluctuant (et une molécule possédant plusieurs modes, voire sites de liaison). Un exemple est donné ci-dessous (Figure 105).

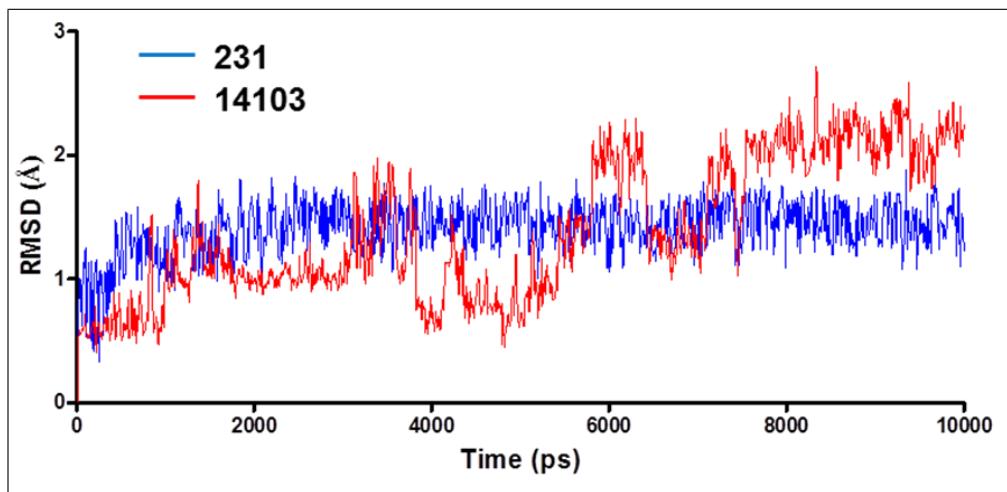


Figure 105: Exemples de suivi de RMSD au cours du temps lors d'une trajectoire de DM avec deux candidats sélectionnés pour l'interaction avec le clamp de la ligase 4.

Ces courbes montrent l'évolution du RMSD pour une molécule stable au sein du domaine clamp (#231, en bleu. Score RMSD = 3) et pour une molécule moins stable (#14103, en rouge. Score RMSD = 1) lors d'une DM de dix nanosecondes avec le domaine C-terminal de la ligase 4.

Finalement, un jeu de données virtuelles contenant 11 candidats potentiels pour l'interaction avec le C-ter Ligase 4 a été établi. Ces candidats sont toutes les molécules de l'histogramme ci-dessus (en bleu), établissant des contacts persistants avec le domaine clamp. Un tableau représentatif de ces 11 molécules avec leur structure, leur classement (basé sur l'énergie théorique de liaison) et leur score RMSD est dressé ci-dessous (Figure 106).

Mol.	Rank	RMSD	Structure	Mol.	Rank	RMSD	Structure
231	1	3		93291	7	2	
3101	2	3		14901	8	1	
36972	3	1		6588	9	3	
3091B	4	1		25037	10	1	
14103	5	1		69817	11	2	
30596	6	1					

Figure 106: Tableau des 11 candidats sélectionnés pour l'interaction avec le clamp de la ligase 4, et à l'issu d'une analyse par simulation de DM.

Les 11 molécules touchées représentées forment toutes des contacts persistants au sein du domaine clamp de la ligase au cours d'une trajectoire de DM de 10ns. Ces molécules ont été classées d'après leur énergie de liaison prédictive.

D'après ce tableau, on pourrait prédire par exemple que les deux premiers hits : #231 et #3101 sont les molécules prédictées comme les plus intéressantes en termes d'énergie de liaison et de RMSD. Il est à noter également que parmi ces molécules, certains de ces scaffolds sont complètement différents (par exemple entre les molécules #3101 et #25037), bien que se liant au niveau du même site d'intérêt d'après l'analyse virtuelle.

Malheureusement, la plupart des 11 molécules sélectionnées n'étaient plus disponibles, étaient insolubles ou étaient sujets à une synthèse très couteuse. Cela nous a conduits à intégrer finalement les 4 molécules: #231, #3101, #36972 et

#3091B (encadrés dans l'histogramme ci-dessus) à nos essais biochimiques et biophysique de validation expérimentale. Là encore, un criblage virtuel haut-débit par docking et DM nous a permis de constituer un petit jeu de 4 candidats potentiels pour la prévention de l'interaction entre XRCC4 et le C-terminal de la ligase 4 (Figure 107).

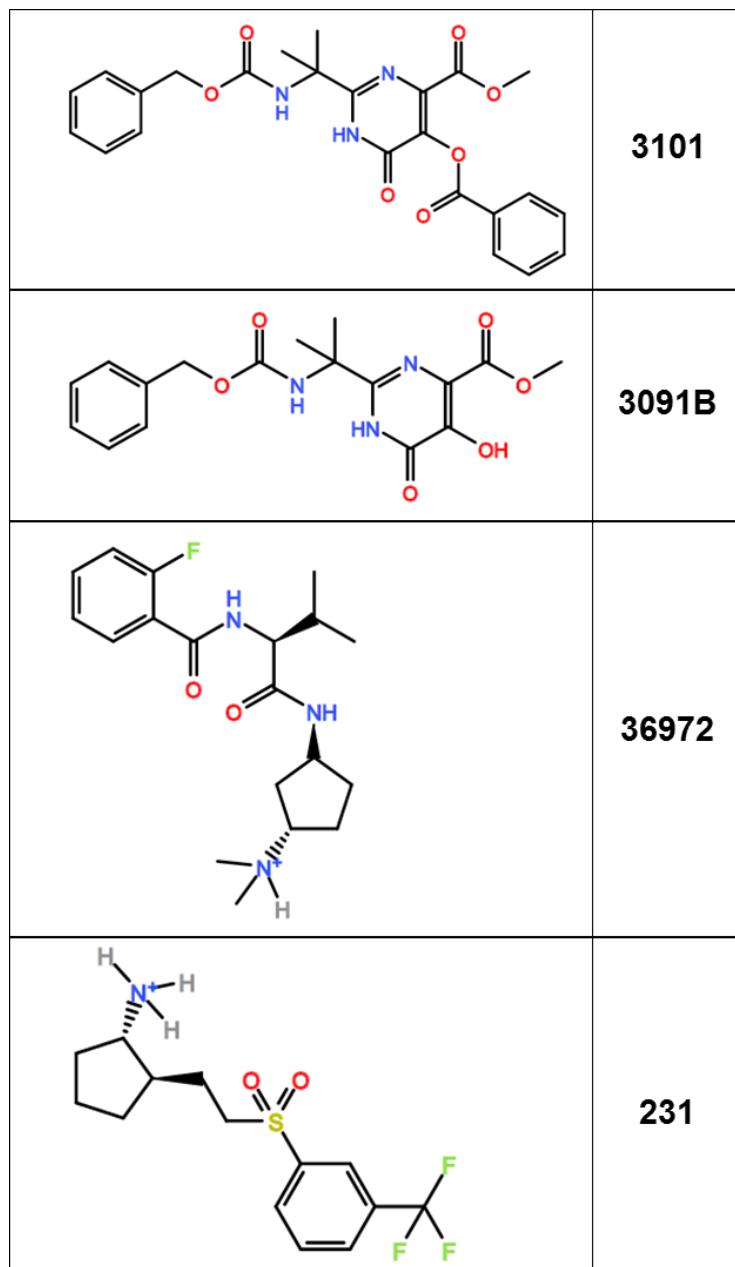


Figure 107: Inhibiteurs potentiels de l'interface XRCC4/C-terminal ligase 4 issus du criblage virtuel centré sur le domaine clamp de la ligase.

On peut toutefois observer la présence d'un noyau similaire aux trois molécules : #3101, #3091B et #36972, formé par un cycle aromatique, une liaison amide et un

groupement dimethyl. Ce noyau, si les trois molécules sont positives à l'issu des essais de validation expérimentale, pourrait être à l'origine d'un même mécanisme de liaison de ces trois inhibiteurs au sein du clamp (sauf si la partie de la molécule qui n'est pas commune aux autres interfère dans l'interaction). Cela reste toutefois à confirmer par des analyses structurales ou des expériences de relations structure-activité. Si l'importance de ce noyau commun est avérée, il serait tout à fait envisageable de trouver des molécules dérivées afin d'entrer dans une phase d'optimisation chimique.

II – 3.6 : Analyse par DLS et RMN ^1H de la solubilité et intégrité des molécules virtuelles retenues

Tout comme dans le chapitre précédent, une vérification rigoureuse de la solubilité de chacune des quatre molécules obtenues a été réalisée par des techniques orthogonales (RMN, DLS) et afin d'éviter les faux positifs dans nos tests d'activité ultérieurs. La DLS nous a permis d'écartier les composés qui étaient susceptibles de former des agrégats (les « promiscuous compounds ») à partir d'une certaine concentration ou même des auto-assemblages volumineux de molécules. Ces entités sont facilement observables sous une binoculaire et ont été confirmés de cette manière. La RMN nous a permis d'évaluer la concentration soluble maximale dans un tampon aqueux pour une concentration donnée calculée (jusqu'à 3mM pour nos tests biochimiques) ainsi que de vérifier la pureté et l'intégrité des molécules (présence de solvants, produits de dégradation...). Les quatre molécules ci-dessus sont solubles dans un tampon RMN (voir chapitre précédent) mais également dans chacun des tampons respectifs contenant les protéines C-terminal de la ligase 4 et XRCC4, et jusqu'à une concentration de 3mM environ. Cette concentration limite a donc été fixée afin de ne pas produire de faux-positifs induits par un phénomène d'agrégation.

II – 3.7 : Tests d'interaction C-ter ligase 4-ligand par DSF et STD-NMR

L'étape suivante a consisté à valider biophysiquement chacune des interactions C-ter ligase 4/molécule candidate par les deux méthodes orthogonales : DSF et RMN. Nous ne re-détaillerons pas ici les avantages et inconvénients de ces deux techniques. Pour cela, l'expression et la purification des constructions XRCC4 (1-

336) et C-ter ligase 4 (654-911) ont été réalisés comme décrits dans le chapitre 2-1. La pureté chimique et la fonctionnalité *in vitro* de ces protéines recombinantes ont été évaluées et jugées satisfaisante pour l'intégration aux tests d'interaction. Ci-dessous, le résultat de l'expérience DSF est présenté pour le C-ter ligase 4 en présence de DMSO ou des molécules candidates (Figure 108).

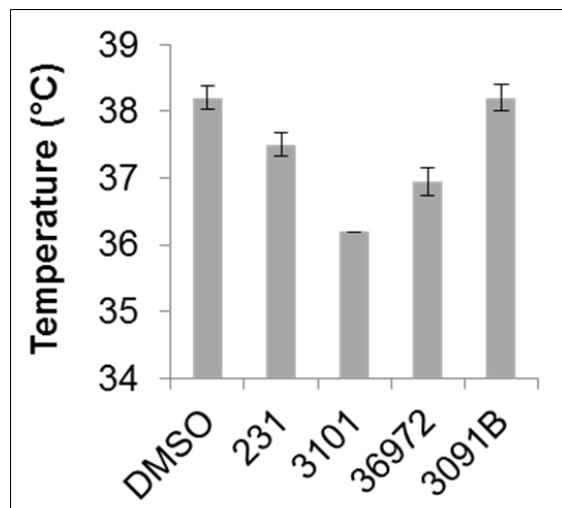


Figure 108: Analyse biophysique de la stabilité thermique par DSF du domaine C-terminal de la ligase 4 seul et en présence des candidats inhibiteurs sélectionnés. La dénaturation est suivie grâce au SYPRO Orange (10X) et entre 20 et 89°C. Les valeurs de T_m ont été estimées comme la moyenne d'une expérience réalisée en triplicat (Moy. \pm DS, $n=3$).

Le domaine recombinant C-ter ligase 4 possède une température de fusion de 38.2°C dans les conditions du test (et en présence de DMSO). Cette température est faible en comparaison notamment avec la température de fusion de XRCC4 qui est de 57°C (Figure 30 – Chapitre II – 1.2). Cette information est en accord avec les données précédentes et confirme l'instabilité intrinsèque de ce domaine lorsqu'il n'est pas stabilisé par son partenaire XRCC4. A l'exception de la molécule #3091B, les trois molécules #231, #3101 et #36972 diminuent significativement la température de fusion du domaine C-ter ligase 4 par rapport au contrôle DMSO (jusqu'à -2°C pour le #3101). Cette diminution de stabilité thermique indique d'une part une liaison de ces molécules sur ce domaine, et d'autre part que cette liaison déstabilise la protéine.

L'effet le plus important est observé pour le candidat #3101. Toutefois, la raison pour laquelle cette déstabilisation a lieu nous est inconnue. La technique de STD-NMR a été également utilisée en complément afin de valider cette liaison au sein du

domaine C-ter ligase 4 recombinant. Sur les quatre composés, seules les molécules #3101 et #231 avaient un signal STD significatif, à en juger par l'augmentation de l'intensité des résonances dans la région aromatique de ces molécules et en comparaison avec la même expérience réalisée en contrôle, sans protéine (Figure 109).

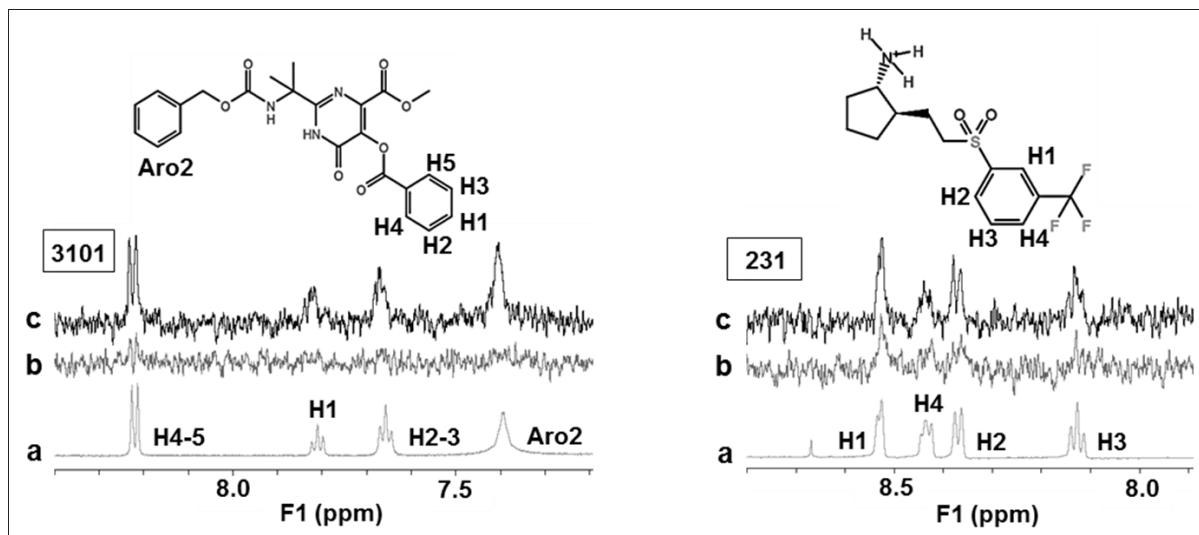


Figure 109: Validation biophysique de l'interaction de deux candidats sélectionnés sur le domaine C-terminal de la ligase 4 par STD-NMR.

Les spectres ^1H (a) et ^1H STD-NMR des deux molécules #3101 et #231 seules (b) ou en présence du domaine C-terminal (c) ont été réalisés. Seules les régions aromatiques des molécules sont représentées ici. L'attribution de ces protons est présentée sur les molécules et reportée sur le spectre ^1H correspondant.

Cette expérience RMN témoigne d'un transfert efficace de saturation entre la protéine et ces inhibiteurs potentiels. Il est à noter que ces deux candidats étaient en tête de liste d'après nos analyses issues des données virtuelles et ont répondu positivement aux deux méthodes de validation biophysique. Egalement, cela souligne l'importante d'intégrer des méthodes orthogonales pour la validation de ces molécules. En effet, le candidat #36972, bien qu'induisant un shift de température significatif, n'a donné aucune réponse STD. Les deux molécules #3101 et #231 ont été validé biophysiquement pour l'interaction avec le C-ter ligase 4 mais nous ne possédons pas, cependant, d'information concernant la spécificité de cette interaction, ainsi que l'effet inhibiteur de ces dernières. Ces deux informations nous ont été fournies après intégration des candidats dans un test de validation biochimique.

II – 3.8 : Test de l'effet préventif des molécules sur l'association XRCC4/C-ter Ligase 4

Rappelons que les protéines recombinantes XRCC4 et C-ter ligase 4 produites ont été validées dans un test d'interaction en condition native par gel d'électrophorèse (chapitre II – 1.2). Ne possédant pas d'activité enzymatique propre pour ce complexe, un test biochimique d'interaction reste un bon moyen de démontrer l'effet d'un inhibiteur potentiel sur une interaction protéine-protéine.

Ce test a donc été reproduit en pré-incubant au préalable le C-ter ligase 4 avec chacun des composés précédents et à des doses croissantes. Cela nous a permis d'établir s'il y avait un effet préventif ou non de ces molécules sur l'association avec le partenaire XRCC4, et si oui, si cet effet était spécifique (effet dose-réponse). Les quatre molécules #231, #3101, #3091B et #36972 ont été intégrées à ce test (Figure 110). Trois concentrations en inhibiteur ont été utilisées pour la pré-incubation avec la ligase : 0.3 ; 1 et 3mM. Après migration du gel, l'effet inhibiteur a été quantifié par le pourcentage de complexe XRCC4/C-ter ligase4 restant en fonction de la concentration en inhibiteur. La molécule #231 possède un effet inhibiteur sur l'association du complexe mais cet effet n'est pas dose-réponse. Les ligands #36972 et #3091B possèdent un effet inhibiteur faible, à partir de 3mM.

Finalement, seul le candidat #3101 possède un effet préventif et dose-réponse sur l'association XRCC4/C-ter Ligase 4, et cet effet est reproductible. Cet effet dose est illustré en (figure 111) et la courbe correspondante montre un IC_{50} autour de 500 μ M pour cet inhibiteur.

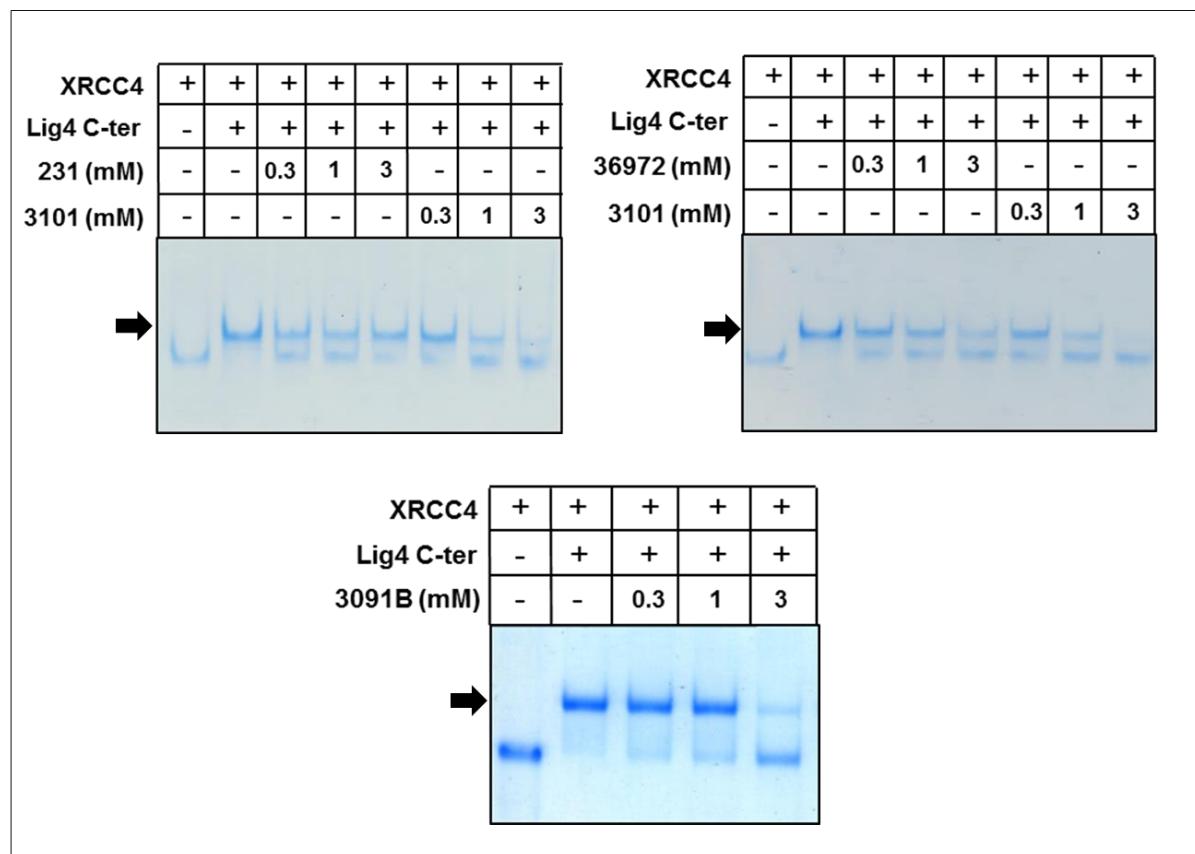


Figure 110: Analyse fonctionnelle des molécules candidates sélectionnés

L'analyse biochimique en gel natif de l'effet préventif des molécules #3101, #231, #36972 et #3091B sur l'association XRCC4/C-terminal de la ligase 4 a été réalisée dans les concentrations et conditions indiquées. Les flèches noires indiquent la présence du complexe protéique.

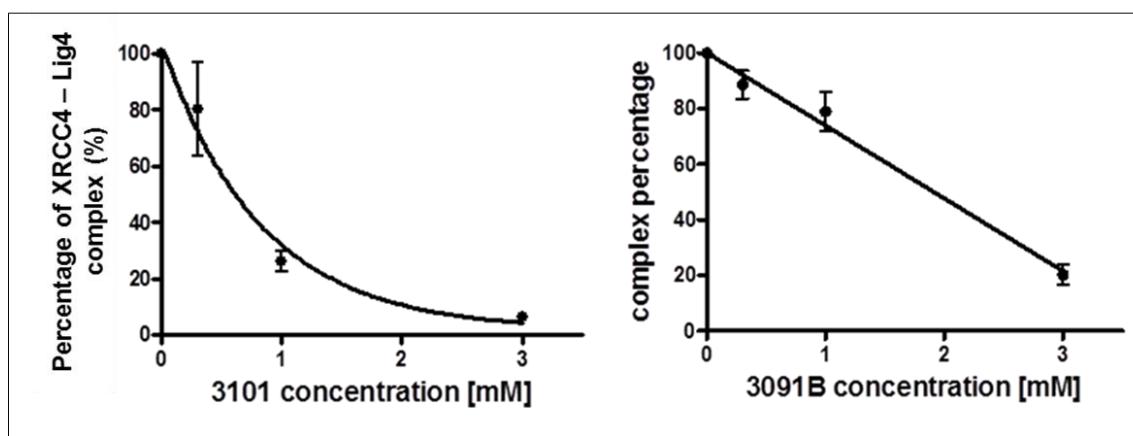


Figure 111: Quantifications pour les expériences d'analyse fonctionnelle des molécules #3101 et #3091B.

Ces quantifications ont été représentées comme la moyenne d'une expérience réalisée en triplicat (Moy. \pm DS, $n=3$). Ces courbes permettent de mettre en évidence particulièrement l'effet dose-réponse du candidat #3101 et avec un $IC_{50} \approx 500\mu M$.

De façon intéressante, le composé #3091B est un précurseur de synthèse de l'inhibiteur #3101 (schéma de synthèse en Figure 112). Par rapport à ce dernier, le #3091B est négatif dans les tests de validation biophysique et possède un faible effet inhibiteur ($IC_{50} > 1\text{mM}$) dans le test d'interaction protéine-protéine par gel natif. Cela souligne l'importance du groupe benzoyle terminal dans la molécule #3101 et nous conforte dans l'idée qu'une molécule plus large est donc nécessaire car plus adaptée à la liaison de la large surface du clamp.

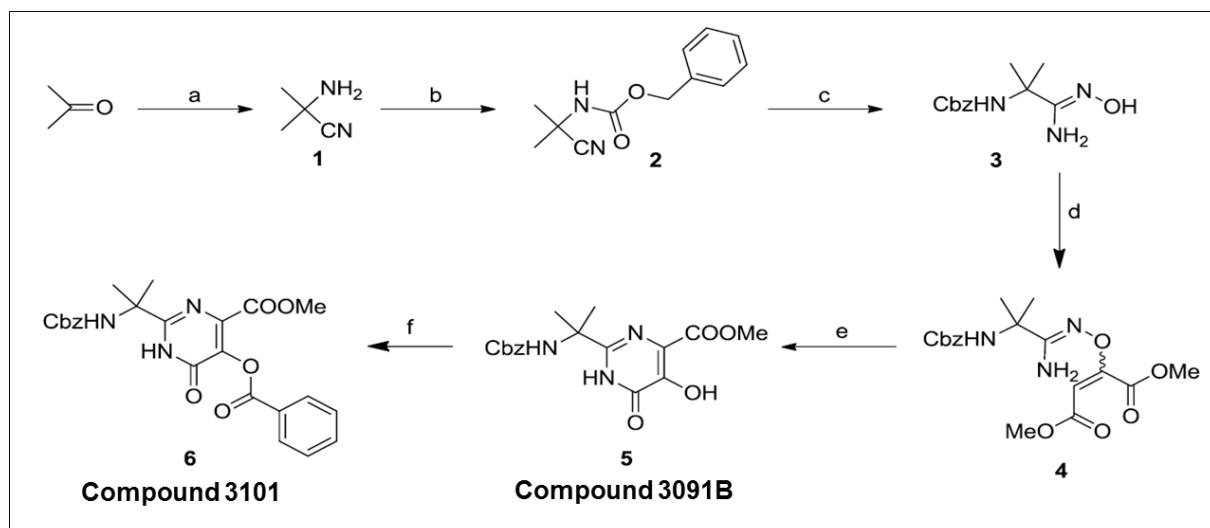


Figure 112: Schéma de synthèse du composé #3101

(a) NH_4Cl , 28% NH_3 , $NaCN$, $EtOH$, H_2O , 25-50°C, 4 days; (b) $Cbz-Cl$, Na_2CO_3 , H_2O , 0°C-rt, 5 h; (c) 50% $NH_2OH - H_2O$, $iPrOH$, 60°C, 5 h; (d) $DMAD$, $MeOH$, rt, 4 h; (e) $Xylene$, 150°C, 48 h; (f) $(PhCO)_2O$, Py , rt, 12-15 h.

Finalement, d'autres molécules dérivées du composé #3101 ont été synthétisées afin de pouvoir effectuer une étude de relation structure-activité et commencer à discriminer (sans a priori sur la structure du complexe protéine/inhibiteur), les zones importantes de la molécule dans la liaison au C-ter ligase 4. Cela aurait pu également nous permettre de trouver des molécules inhibitrices plus efficaces (Figure 113).

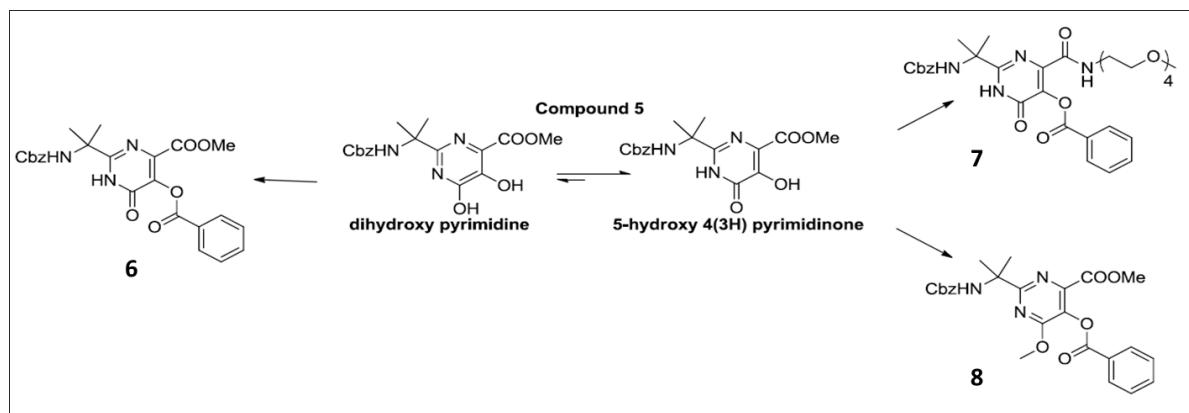


Figure 113: Schéma de molécules dérivées du composé #3101 (6)

D'après nos observations, ces molécules dérivées (7 et 8 sur la Figure 113) étaient moins actives que la molécule #3101 dans notre test biochimique. Ainsi, des modifications dans la zone methoxy ou sur le carbone du groupement pyrimidine n'ont pas amélioré l'activité inhibitrice de ce composé. Des efforts doivent être maintenant centrés sur des modifications encore non explorées (comme l'azote du groupement pyrimidine ou le remplacement de la partie carboxamide).

II – 3.9 : Modèle d'interaction entre le C-ter ligase 4 et l'inhibiteur #3101

Afin d'essayer de comprendre l'interaction possible entre l'inhibiteur #3101 et le domaine clamp de la ligase 4, un modèle issu de la DM du complexe a été proposé. Le dernier cadre de la simulation de DM de 10ns pour le complexe C-ter ligase4/#3101 a donc été extrait (Figure 114). Tout au long de cette simulation, et malgré la présence des deux domaines BRCT, la molécule est restée localisée au sein du clamp, confirmant ainsi sa spécificité pour ce domaine.

Le modèle permet d'entrevoir que cette molécule possède un mode de liaison unique et contacte simultanément et de façon persistante les trois parties du domaine clamp (hélice1-loop-hélice2). Elle établit un pont hydrogène persistant entre son groupe carbonyle et l'amine de la chaîne latérale du résidu N783 (loop) ainsi que des interactions hydrophobes entre ses groupes diméthyl et methoxy et les chaînes latérales des résidus critiques F778 (hélice 1) et I796 (hélice2).

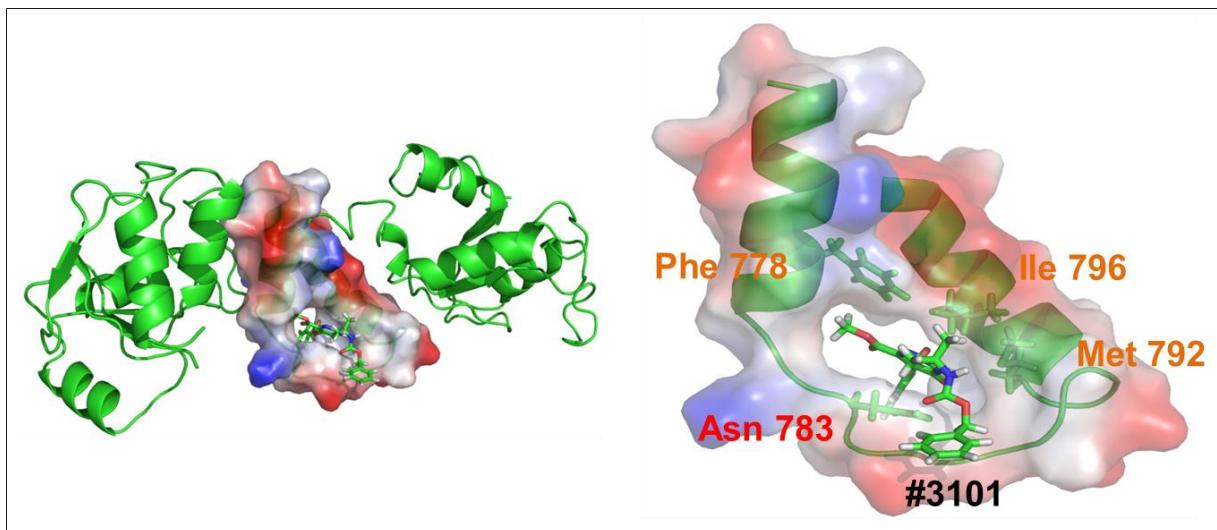


Figure 114: Modèle d'interaction entre le C-ter ligase 4 et le composé #3101, issu d'une simulation de DM de 10ns.

A gauche, le domaine C-ter (BRCT1-clamp-BRCT2) est représenté avec le domaine clamp en surface et coloré d'après le potentiel électrostatique. A droite, un agrandissement du composé #3101 au sein du domaine clamp. Les résidus formant des contacts persistants avec la molécule sont représentés avec le même code couleur que dans la figure 92.

Ce mode de liaison est un premier modèle de ligand interagissant avec le domaine clamp de la ligase 4. Ce modèle établie une base, comme un point de départ possible dans une stratégie de design rationnel d'inhibiteurs de plus haute affinité pour ce domaine.

II – 3.10 : Conclusion et perspectives – Vers la structure d'un complexe C-ter Ligase4/inhibiteur

A l'issue de cette étude, nous avons pu isoler la première molécule capable de prévenir *in vitro*, l'interaction extrêmement forte entre les protéines XRCC4 et C-terminal de la ligase 4. Au vu de leur forte affinité, de l'absence de cavités ou sillons adaptés au criblage *in silico* de petites molécules, ainsi que de l'absence d'une structure tridimensionnelle du C-ter non complexé à XRCC4, d'une quelconque activité enzymatique ou d'inhibiteurs déjà connus (pour des méthodes basées sur le ligand ou des expériences de compétition pour démontrer indirectement l'affinité d'un hit pour le site d'intérêt), inhiber cette interface protéine-protéine est un réel challenge. Notre étude a permis toutefois d'apporter plusieurs éléments :

- Démontrer la faisabilité d'une prévention de l'interaction pourtant extrêmement affine entre XRCC4 et le domaine C-terminal de la ligase 4 par une petite molécule.
- Valider le domaine clamp de la ligase 4 comme cible minimale dans une stratégie de criblage virtuel.
- Fournir un modèle d'interaction avec un premier composé d'affinité modérée, pouvant maintenant servir à l'élaboration rationnelle de molécules dérivées dans une stratégie d'optimisation par chimie médicinale. Ces modifications auraient pour but de trouver des molécules de plus haute affinité et spécificité pour le domaine clamp afin d'entamer par exemple, des essais cellulaires.

Dans une stratégie ultérieure d'optimisations chimiques ou de recherche d'inhibiteurs de plus haute affinité, il semblerait judicieux de poursuivre les modifications chimiques entamées sur le composé #3101 actif en s'aidant, par exemple, du modèle issu de la DM proposé et d'établir des relations structure-activité. D'autant plus que ce composé ouvre de nombreuses voies de modification possibles. D'autres composés pourraient être recherchés (dans d'autres bases de données par exemple), en s'appuyant sur une recherche de molécules similaires en termes de scaffold, conformation et potentiel électrostatique (par exemple, programme OpenEye). Si une série de molécules analogues et actives dans nos essais biochimiques est obtenue, il est tout à fait envisageable d'établir un pharmacophore commun à ces inhibiteurs. Ce pharmacophore servirait alors de base pour le filtrage de nouveaux hits virtuels potentiels. Ces méthodes, bien qu'elles permettent probablement de trouver des molécules analogues efficaces, sont basées sur l'analyse du ligand mais ne fournissent aucunes preuves directes de la localisation de ces inhibiteurs au sein du site ciblé.

Cette preuve, tout-à-fait indispensable, pourrait être apportée de trois façons différentes :

- Par STD en situation de compétition. En effet, le clamp de la ligase 4 est le domaine ici ciblé. En bloquant l'accès à ce dernier par formation du complexe avec XRCC4, une expérience STD comme réalisée précédemment serait

alors négative si le clamp est bien le site d'interaction spécifique de notre composé.

- Par mutagénèse dirigée. Un modèle issu de la DM établi les contacts persistants entre le clamp et l'inhibiteur #3101. En mutant un ou plusieurs de ces résidus clés, nous devrions perdre cette interaction.
- En résolvant la structure tridimensionnelle du complexe protéine/inhibiteur de manière expérimentale, par RMN ou cristallographie et diffraction des rayons X.

Cette dernière méthode permettrait par ailleurs, de modifier de manière rationnelle notre inhibiteur afin de le rendre potentiellement plus efficace (stratégie basée sur la structure). Sans cela, les meilleures preuves indirectes de spécificité d'interaction au site d'intérêt que nous possédons est donnée par l'activité dose-réponse de notre inhibiteur #3101 dans les essais biochimiques et par le fait que l'expérience inverse dans ces essais (pré-incubation de la drogue avec XRCC4) ne prévient pas la formation du complexe.

Afin de confirmer de manière biophysique la liaison spécifique du composé #3101 au sein du domaine C-terminal de la ligase 4, une expérience RMN complémentaire de trNOESY a été réalisée pour le ligand seul (à 1mM) et en présence de 14 μ M des domaines C-terminal ligase 4 et BRCT2 purifiés (Figure 115). Lorsqu'elle est seule en solution, cette molécule génère des crosspeaks dus à des NOEs positifs, montrant ainsi une complète solubilité dans ces conditions de tampon (identiques au tampon STD) et donc, pas de phénomène d'agrégation. Des crosspeaks dus à des NOEs négatifs sont générés lorsque ce ligand est mis en présence du domaine C-terminal de la ligase 4 (ce qui n'est pas le cas avec le domaine BRCT2). Ainsi, cette molécule n'est pas un composé pouvant se lier de manière non spécifique et particulièrement au niveau du domaine BRCT2 non ciblé dans cette stratégie.

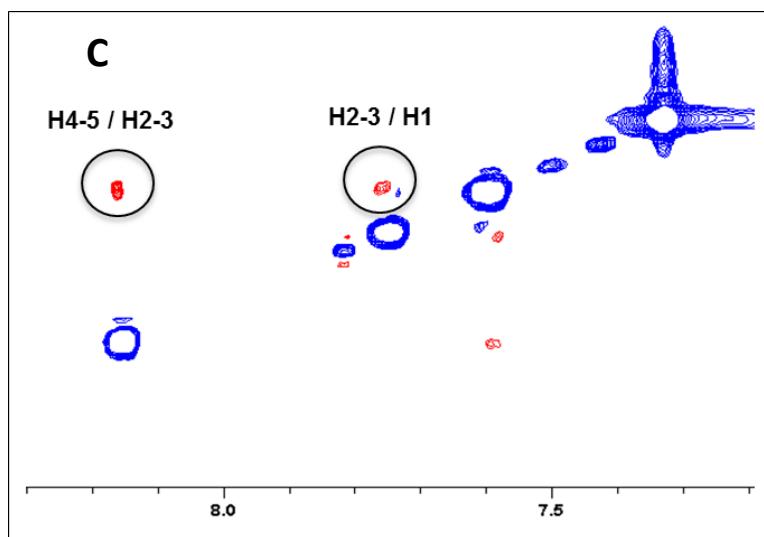
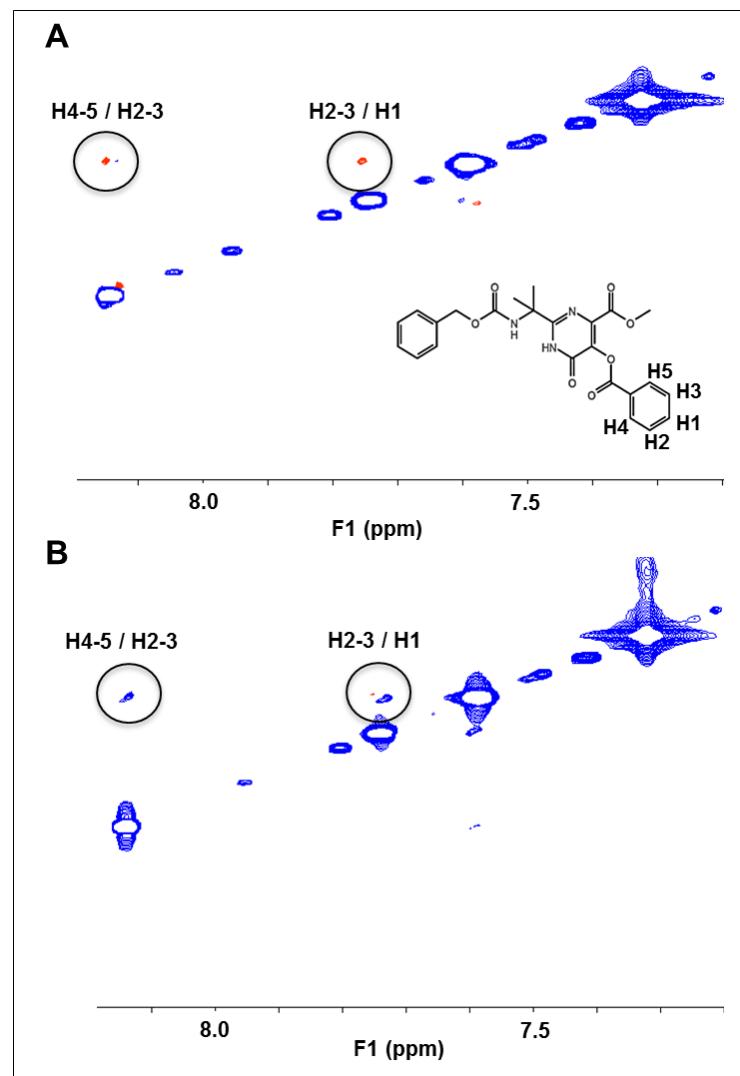


Figure 115: Analyse de l'interaction du composé #3101 avec les domaines C-terminal de la Ligase 4 et BRCT2 par trNOESY.

(A) Spectre *trNOESY* (temps de mélange 200ms) du composé #3101 (concentration 1mM) seul dans un tampon RMN. Les crosspeaks (cercles noir) montrant la proximité spatiale entre H4-5 et H2-3 et entre H2-3 et H1 sont dus à des NOEs positifs (en rouge, de signe opposé à la diagonale). (B) Spectre *trNOESY* (temps de mélange 200ms) du composé #3101 (concentration 1mM) en présence de 14 μ M de C-ter Ligase 4. Les crosspeaks (cercles noir) montrant la proximité spatiale entre H4-5 et H2-3 et entre H2-3 et H1 sont dus à des NOEs négatifs (en bleu, de même signe que la diagonale). (C) Spectre *trNOESY* (temps de mélange 200ms) du composé #3101 (concentration 1mM) en présence de 14 μ M de BRCT2. Les crosspeaks (cercles noir) montrant la proximité spatiale entre H4-5 et H2-3 et entre H2-3 et H1 sont dus à des NOEs positifs (en rouge, de signe opposé à la diagonale).

Envisager une structure du complexe par cristallographie :

Une étude par cristallographie nécessite d'avoir une protéine largement concentrée (plusieurs mg/ml), pure, monodisperse et très stable dans le temps. Le domaine C-ter de la ligase 4 étant peu stable (cocristallisation nécessaire avec son partenaire XRCC4, bloquant alors l'accès au clamp), obtenue dans des quantités faibles (inférieures au mg/ml) lors de nos essais et également s'organisant en trois domaines distincts (BRCT1-clamp-BRCT2) et donc probablement très flexible en solution, il serait par conséquent difficile d'envisager des expériences de cristallisation sur ce dernier. Notre construction (domaine C-ter ligase 4 : 654-911) est tagguée 6*His et une solution face au faible rendement serait de changer de tag (GST par exemple). La stabilisation de ce domaine pourrait être induite par l'inhibiteur lui-même. En effet, si ce modèle de cavité formée par les deux hélices du clamp est juste, un ligand pourrait stabiliser ce dernier. Finalement, il pourrait être envisagé de travailler sur un domaine tronqué, à condition que le clamp soit toujours fonctionnel *in vitro*, plus stable et avec un meilleur rendement une fois l'un ou les deux domaines BRCT enlevés.

Envisager une structure du complexe par RMN :

Une étude par RMN nécessite également d'avoir une protéine largement concentrée, pure, et très stable dans le temps. Le problème de la flexibilité intrinsèque de la protéine n'est pas important ici, mais une étude structurale par RMN nécessite cependant de marquer la protéine. Une tentative de marquage ^{15}N du domaine C-terminal de la ligase 4 en milieu minimum M9 a été menée mais a conduit à une large production de protéine insoluble. Là encore, il pourrait être envisagé de travailler sur un domaine tronqué, à condition que le clamp soit toujours soluble,

fonctionnel *in vitro*, plus stable et avec un meilleur rendement une fois l'un ou les deux domaines BRCT enlevés.

Afin d'augmenter le rendement de production et la stabilité du système tout en travaillant sur un domaine plus simple pour la RMN, les constructions : β hairpin-clamp-GST (aa 759-803) et BRCT1- β hairpin-clamp-GST (aa 654-803) du domaine C-ter de la ligase 4 ont été produites et testées en pour leur interaction avec XRCC4 en gel natif (Figure 116). Les domaines tronqués produits ont été jugé purs et sous forme monomérique en conditions natives et dans les conditions de tampon utilisées.

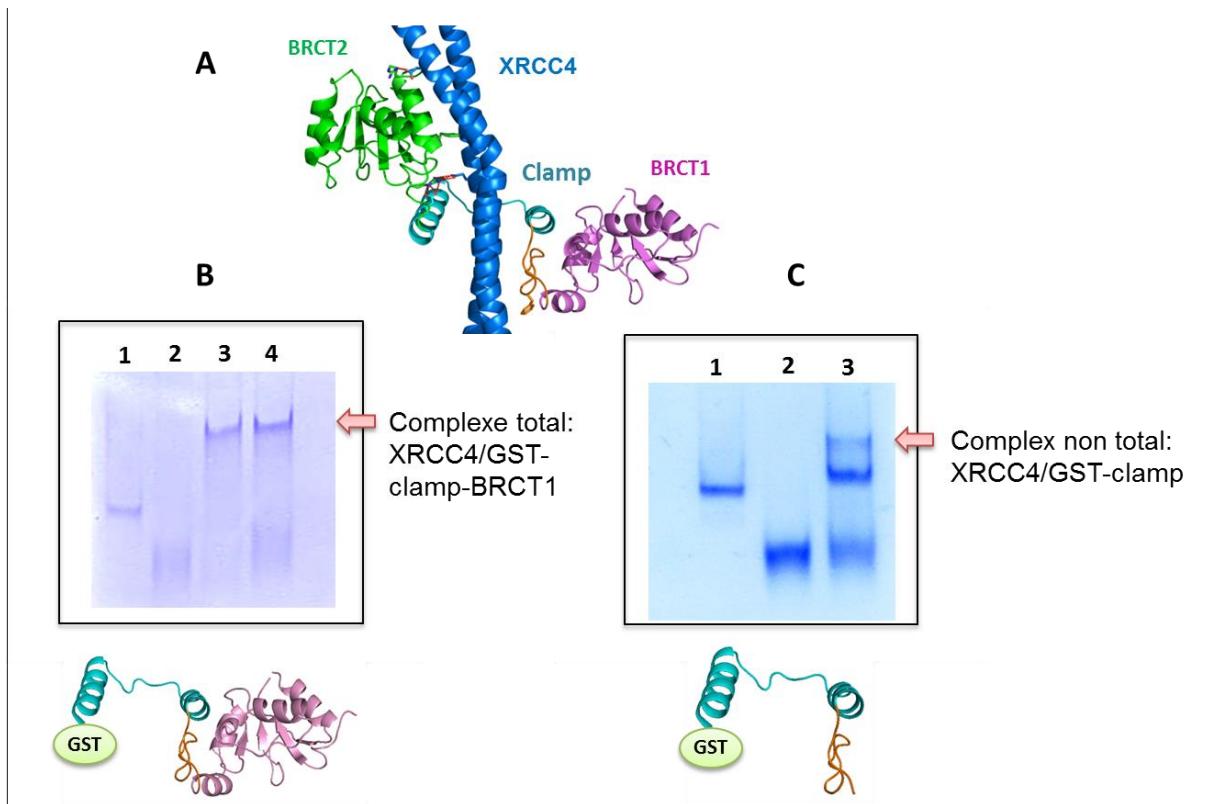


Figure 116: Analyse fonctionnelle par gel retard des domaines GST-clamp-BRCT1 et GST-clamp

(A) Structure du domaine C-terminal de la ligase 4, en complexe avec XRCC4. (B) gel natif pour le suivi de l'interaction entre le domaine tronqué GST-clamp-BRCT1 et XRCC4. (C) gel natif pour le suivi de l'interaction entre le domaine tronqué GST-clamp et XRCC4.

Cette expérience nous a permis de montrer que le C-terminal de la ligase 4 privé de son domaine BRCT2 est toujours capable d'interagir totalement avec XRCC4 *in vitro*. Cela a été également confirmé par des expériences de pull-down avec deux constructions : XRCC4-His et XRCC4-eGFP-His. Le domaine GST-clamp-BRCT1

forme un complexe total avec XRCC4 en condition équimolaire. Dans le cas du domaine GST-clamp, cette liaison n'est pas totale, et ce même en présence d'un large excès de ce dernier. L'hypothèse d'un encombrement stérique par le domaine GST a été vérifiée en clivant ce tag et en reproduisant la même expérience d'interaction en gel natif avec XRCC4 (Figure 117).

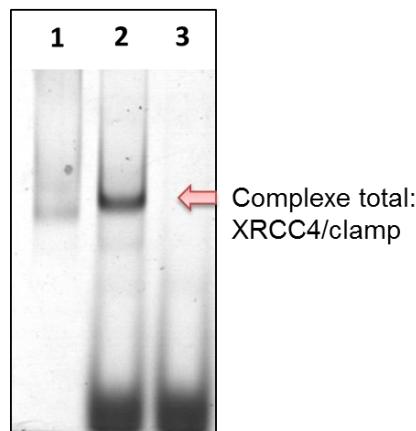


Figure 117: Analyse fonctionnelle par gel retard du domaine clamp détagué

Cela a permis de démontrer que ce clamp détagué conserve *in vitro* son activité d'interaction totale avec XRCC4 et sans aucun domaine BRCT. Cela va dans le sens de l'étude structurale menée par Sibanda *et al.* en 2001, montrant la résolution cristallographique d'une structure XRCC4 en complexe avec un fragment de la ligase 4 de 37 résidus, contenant le β -hairpin et une partie de l'hélice 1 (Figure 118).

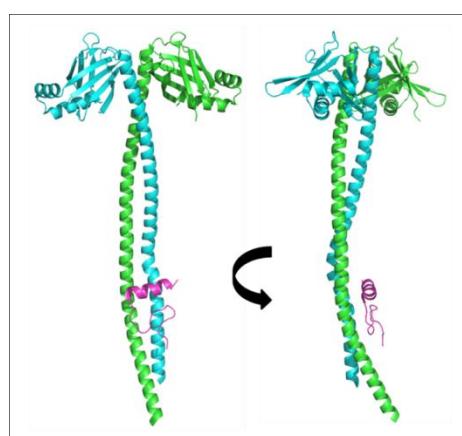


Figure 118: Structure cristallographique d'un fragment du clamp (« XIR ») en interaction avec XRCC4 (D'après Sibanda *et al*, 2001).

Le domaine clamp détagué de 71 résidus produit précédemment a pu être concentré au-delà de 1mg/ml et testé pour l'interaction avec l'inhibiteur #3101 caractérisé précédemment par une expérience STD. Par rapport à l'expérience concluante de STD réalisée sur le domaine C-terminal pleine taille de la ligase 4, aucun signal STD significatif n'a pu être enregistré avec ce domaine clamp isolé.

Ainsi, il a été émis l'hypothèse que ce domaine était majoritairement non structuré en solution lorsqu'il était isolé de ses deux domaines BRCT et qu'il se structurait au contact de la protéine XRCC4. Une expérience de dichroïsme circulaire pourrait confirmer cette absence de structure secondaire. Sachant cela, aucune interaction spécifique du composé #3101 n'est possible si ce dernier n'est pas structuré. Si tel est le cas, il semblerait qu'il faille conserver les deux domaines BRCT pour avoir une reconnaissance du domaine clamp en solution par l'inhibiteur et ainsi caractériser la structure du complexe par RMN. Dans ce cas, d'autres méthodes adaptées pourraient être envisagées afin de déterminer sans ambiguïté le site de liaison de notre inhibiteur ainsi que la conformation bioactive de ce dernier sur le domaine C-terminal complet. Ces techniques, comme celles basées sur les NOEs (« NOE matching » par exemple) et la méthode SOS-NMR, utilisent peu de protéine et aucun marquage spécifique (Dias and Ciulli, 2014).

Pour conclure, Des approches très diverses sont encore à exploiter dans cette étude afin de pouvoir entamer une campagne de chimie médicinale en vue de modifier rationnellement notre inhibiteur (#3101) grâce à la structure. Néanmoins, cet inhibiteur est un point de départ et apporte par-dessus tout une preuve de la « druggabilité » du domaine C-terminal de la ligase 4, encore jusque-là non exploité car probablement considéré comme non « druggable ». L'approche basée sur l'utilisation d'une petite molécule pour inhiber des interfaces protéine-protéine est efficace et a déjà fait ses preuves. Mais une approche peptido-mimétique pourrait être également considérée. En effet, le meilleur compétiteur possible pour le domaine C-terminal de la ligase 4 est très probablement le domaine clamp lui-même. La construction produite durant cette thèse a été efficace dans une stratégie de prévention de la formation du complexe XRCC4/C-ter ligase 4 en pré-incubant XRCC4 avec cette dernière dans une expérience de gel natif. De plus, nous savons

qu'une construction plus courte de ce domaine peut interagir spécifiquement avec XRCC4 (Sibanda et al., 2001). Il est alors tout à fait envisageable de produire une partie du domaine clamp (ou bien même un peptide mimétique ou modifié chimiquement pour augmenter la stabilité) pour la prévention *in cellulo* de la formation de ce complexe de réparation de l'ADN, et donc dans une approche thérapeutique. Evidemment, cette approche pourrait être efficace dans la mesure où la taille et les propriétés physico-chimiques de ce dérivé peptidique autorisent une pénétration cellulaire efficace.

L'approche peptidique a été explorée lors de cette thèse mais par la production et la tentative de formation d'une partie du coiled-coil de XRCC4, en interaction avec le domaine clamp de la ligase. Malheureusement, cette approche n'a pas fonctionné ici car ce peptide n'était pas structuré en solution et n'était pas soluble dans des conditions physiologiques.

Ce travail sur la partie XRCC4/C-terminal de la Ligase 4 a fait l'objet d'un premier article soumis pour publication, et présenté à la fin de ce manuscrit de thèse.

Conclusion générale et Perspectives

La protéine Ligase 4 humaine est une cible de choix dans le but d'inhiber le mécanisme de réparation des cassures double-brins de l'ADN. Elle est l'unique ligase associée au mécanisme de réparation par NHEJ et n'intervient dans aucune autre voie métabolique. Cependant, étant donnée la conservation du domaine catalytique dans les ligases humaines, une plus grande spécificité de ciblage est nécessaire afin d'éviter des effets cellulaires hors-cible. A ce jour, deux stratégies d'inhibition spécifique de cette ligase 4 peuvent être envisagées et n'ont encore jamais été exploitées : une inhibition directe entre le domaine C-terminal de la ligase et la protéine XRCC4, ou bien indirecte en inhibant la formation du complexe XRCC4/Cer-XLF nécessaire à son activité. D'ores et déjà, des preuves de concept de l'effet de l'inhibition *in cellulo* de ces interactions sur la radiosensibilisation des cellules cancéreuses sont disponibles et ont été détaillés dans cette thèse.

Notre projet était basé sur l'inhibition spécifique d'interactions protéine-protéine (PPI). La question développée au cours de cette thèse a été de savoir s'il était possible d'interférer avec les complexes XRCC4/C-terminal Ligase 4 et XRCC4/Cer-XLF grâce au criblage virtuel de petites molécules inhibitrices. La stratégie globale a été de tirer profit des structures cristallographiques existantes pour chacune de ces protéines et de réaliser par étapes :

- Un criblage virtuel par docking haut débit sur des structures relaxées par dynamiques moléculaires des protéines XRCC4, Cer-XLF et le domaine C-terminal de la Ligase 4, et au niveau de résidus critiques d'interaction.
- Une validation *in silico* des molécules touchées virtuelles par des simulations de dynamiques moléculaires des complexes protéine/ligand sélectionnés.
- La production des trois protéines recombinantes en système bactérien hétérologue et la mise en place de tests d'activité biochimiques.
- La validation biophysique de la liaison des molécules par des techniques complémentaires de Fluorimétrie à balayage différentiel (DSF) et de Différence de Transfert de Saturation (STD-NMR).
- La validation biochimique de l'effet inhibiteur des molécules touchées sur les interactions XRCC4/domaine C-terminal Ligase 4 et XRCC4/Cer-XLF (tests

d'interaction en gel natif, tests de stimulation de la ligase T4, test NHEJ *in vitro* avec extraits cellulaires).

- Enfin, la caractérisation structurale des inhibiteurs sélectionnés.

Le résultat majeur de ce travail a été de pouvoir fournir en fin de thèse et pour chacune des deux interfaces ciblées, une molécule inhibitrice spécifique avec une activité sub-millimolaire (molécule #3101 pour l'interaction avec le C-terminal de la Ligase 4 et molécule #57632 pour l'interaction avec XRCC4). Cela démontre la possibilité d'inhiber ces PPI par des petites molécules et valide la stratégie employée.

Une des difficultés majeures rencontrées au cours de ce travail était l'absence d'une structure isolée du domaine C-terminal de la Ligase 4. Ce domaine a pu néanmoins être obtenu par modélisation moléculaire sous forme séparée de XRCC4, permettant ainsi de révéler une surface d'interaction accessible au docking de petites molécules. Nous n'avons cependant aucune preuve expérimentale de l'existence des conformations obtenues. De même, le nombre de candidats validés à l'issue de nos tests est limité à une molécule par interface, soulignant aussi les limites de l'approche virtuelle dans des programmes de découverte de molécules inhibitrices. Cette approche a pourtant été la seule à pouvoir fournir des composés fonctionnels (à l'inverse par exemple de la stratégie HTS également employée). Finalement, les chimiothèques utilisées auraient pu être enrichies en molécules mieux adaptées aux PPI. De nombreuses banques de molécules « PPI-Friendly » sont aujourd'hui disponibles et pourraient permettre d'augmenter le nombre de molécules touchées lors d'un criblage virtuel ultérieur.

Ces résultats ouvrent de nombreuses perspectives quant à la suite de ce projet. Les inhibiteurs caractérisés sont des points de départ pour l'élaboration de molécules de plus haute affinité. Ainsi, à court-terme, des expériences de caractérisation structurale des complexes C-terminal Ligase 4/inhibiteur #3101 et XRCC4/inhibiteur #57632 sont indispensables. D'après nos résultats, la molécule #57632 inhibitrice de l'interaction entre XRCC4 et Cer-XLF semble la plus prometteuse (meilleur Kd, validation de l'activité inhibitrice à partir de milieux complexes tels que des extraits cellulaires). La cristallographie des complexes XRCC4/composé #57632 doit être poursuivie. Des expériences de trempage de cristaux d'un complexe XRCC4/C-

terminal de la Ligase 4 ou de co-cristallisation doivent encore être explorées. Le mode de liaison de cet inhibiteur sur la protéine XRCC4 apporterait des informations précieuses afin d'anticiper, de manière rationnelle, des modifications de ce dernier par chimie médicinale. Il apporterait également des éléments de réponses qui restent aujourd'hui en suspens. Par exemple, nous n'avons pas d'explication claire sur le mécanisme de liaison de cet inhibiteur conduisant à une courbe sigmoïdale lors d'une expérience de titration par STD-NMR sur la protéine XRCC4. Cette courbe pourrait traduire un évènement de coopérativité ou bien même une liaison sur deux sites distincts et d'affinités différentes pour la molécule.

A plus long terme, une approche nouvelle par fragment pourrait également être envisagée. En effet, il serait intéressant d'isoler des fragments issus de la décomposition de notre molécule #57632 inhibitrice initiale (Figure 119). Deux fragments (A et B) pourraient être obtenus. S'ils se lient toujours à la protéine cible, ils apporteraient un squelette de base plus simple, dont des modifications chimiques par « growing » pourraient mener à une molécule différente et de meilleure affinité.

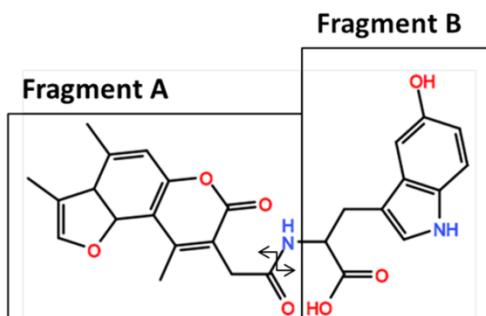


Figure 119: Décomposition en deux fragments A et B de l'inhibiteur #57632 de l'interaction entre XRCC4 et Cer-XLF. La double flèche représente le site de coupure.

Les modifications chimiques devront être menées tout en prenant en compte le profil pharmacocinétique des composés néo-synthétisés et ce afin de pouvoir, dès les premières étapes, satisfaire des critères d'entrée dans des programmes d'optimisation de composés « lead » et de développement pré-clinique.

De même, des molécules analogues pourraient être recherchées tout en conservant certains critères de l'inhibiteur #57632 (squelette, conformation, potentiel électrostatique). Si une série de molécules analogues et actives est obtenue, il serait alors possible de dresser un pharmacophore commun à ces inhibiteurs. Ce

pharmacophore servirait alors de base pour le filtrage de nouvelles touches potentielles.

Article:

**“Structure-Based Virtual Ligand Screening on the
XRCC4/DNA Ligase IV Interface”**

(Soumis à « Scientific Reports »

le 15.07.15)

Auteurs :

Grégory Menchon^{1,2,3}, Oriane Bombarde^{1,2,3}, Mansi Trivedi^{1,2}, Aurélie Négrel^{5,6,7,8}, Cyril Inard⁴, Brigitte Guidetti⁴, Michel Baltas⁴, Alain Milon^{1,2}, Mauro Modesti^{5,6,7,8}, Georges Czaplicki^{1,2}, Patrick Calsou^{1,2,3}

¹ CNRS; IPBS (Institut de Pharmacologie et de Biologie Structurale); BP 64182, 205 route de Narbonne, F-31077 Toulouse, Cedex4, France

² Université de Toulouse; UPS; IPBS; F-31077 Toulouse, France

³ Equipe labellisée Ligue Nationale Contre le Cancer

⁴ CNRS - UMR 5068 - Synthèse et physico-chimie de molécules d'intérêt biologique (SPCMIB) Université de Toulouse 118 route de Narbonne F-31062 Toulouse, France

⁵ Centre de Recherche en Cancérologie de Marseille, CNRS, UMR7258, Marseille, F-13009, France;

⁶ INSERM, U1068, Marseille, F-13009, France;

⁷ Institut Paoli-Calmettes, Marseille, F-13009, France;

⁸ Aix-Marseille Université, Marseille, F-13284, France;

Matériaux et Méthodes

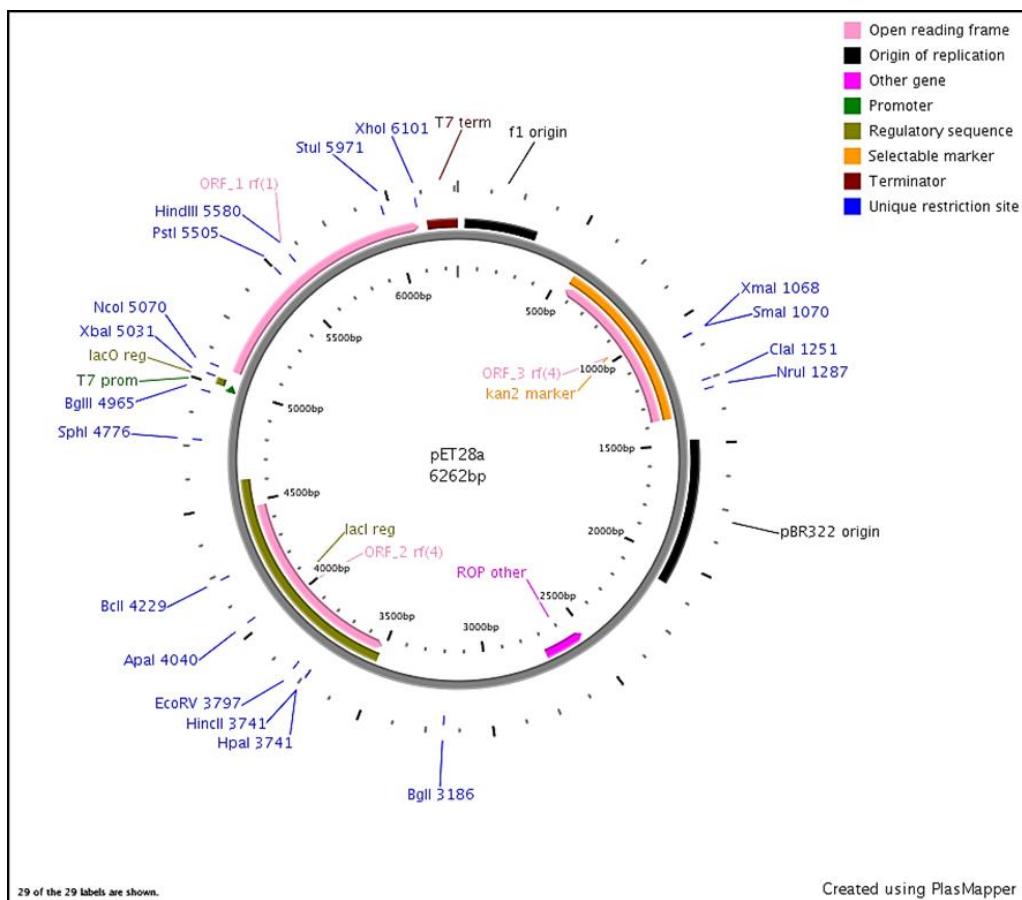
- **Biologie moléculaire : carte des vecteurs et séquences protéiques**
- **Protocole de clonage – Application aux domaines tronqués 1-112 et 1-119 de XRCC4**
- **Protocole de transformation bactérienne**
- **Expression des protéines du complexe de ligature**
- **Purification des protéines du complexe de ligature**
- **Protocole de purification de protéines recombinantes sur système AKTA**
- **Protocole de Fluorimétrie à balayage différentiel (DSF)**
- **Protocole de Différence de Transfert de Saturation (STD-NMR)**
- **Gels d'électrophorèse : Tricine-SDS-PAGE et gels natifs**
- **Protocole de stimulation de la ligase T4 par le complexe XRCC4/Cer-XLF humain**
- **Test d'intercalation à l'ADN**
- **Test NHEJ *in vitro* avec extraits cellulaires – Test des inhibiteurs de l'interaction XRCC4/Cer-XLF**
- **Cristallisation**
- **Docking et dynamique moléculaire**

Biologie moléculaire : carte des vecteurs et séquences protéiques

Protéine XRCC4 (pET28a)

Résistance : kanamycine

Tag : 6*Histidines en C-terminal



Sequence protéique:

10	20	30	40	50	60
MERKISRIHL VSEPSITHFL QVSWEKTLES GFVITLTDGH SAWTGTVSSES EISQEADDMA					
70	80	90	100	110	120
MEKGKYVGEL RKALLSGAGP ADVYTFNFSK ESCYFFFKEKN LKDVSFRLGS FNLEKVENPA					
130	140	150	160	170	180
EVIRELICYC LDTIAENQAK NEHLQKENER LLRDWNDVQG RFEKCVSAKE ALETDLYKRF					
190	200	210	220	230	240
ILVLNEKKTK IRSILHNKLLN AAQEREKDIK QEGETAICSE MTADRDPVYD ESTDEESENQ					
250	260	270	280	290	300
TDLGLASAA VSKDDSISS LDVTIAPSRR KRRQRMQRNL GTEPKMAPQE NQLQEKENSR					
310	320	330	340		
PDSSLPETSK KEHISAENMS LETLRNSSPE DLFDEIHHHH HH					

Paramètres protéiques:

- Nombre d'acides aminés: 342
- Poids moléculaire: 39109.5
- Point isoélectrique théorique: 5.15

Nombre total de résidus chargés négativement (Asp + Glu): 64

Nombre total de résidus chargés positivement (Arg + Lys): 45

- Les coefficients d'extinction molaire sont en $M^{-1} \text{ cm}^{-1}$, à 280 nm dans l'eau.

Ext. coefficient 25690

Abs 0.1% (=1 g/l) 0.657 (cas où tous les résidus Cys forment des cystines)

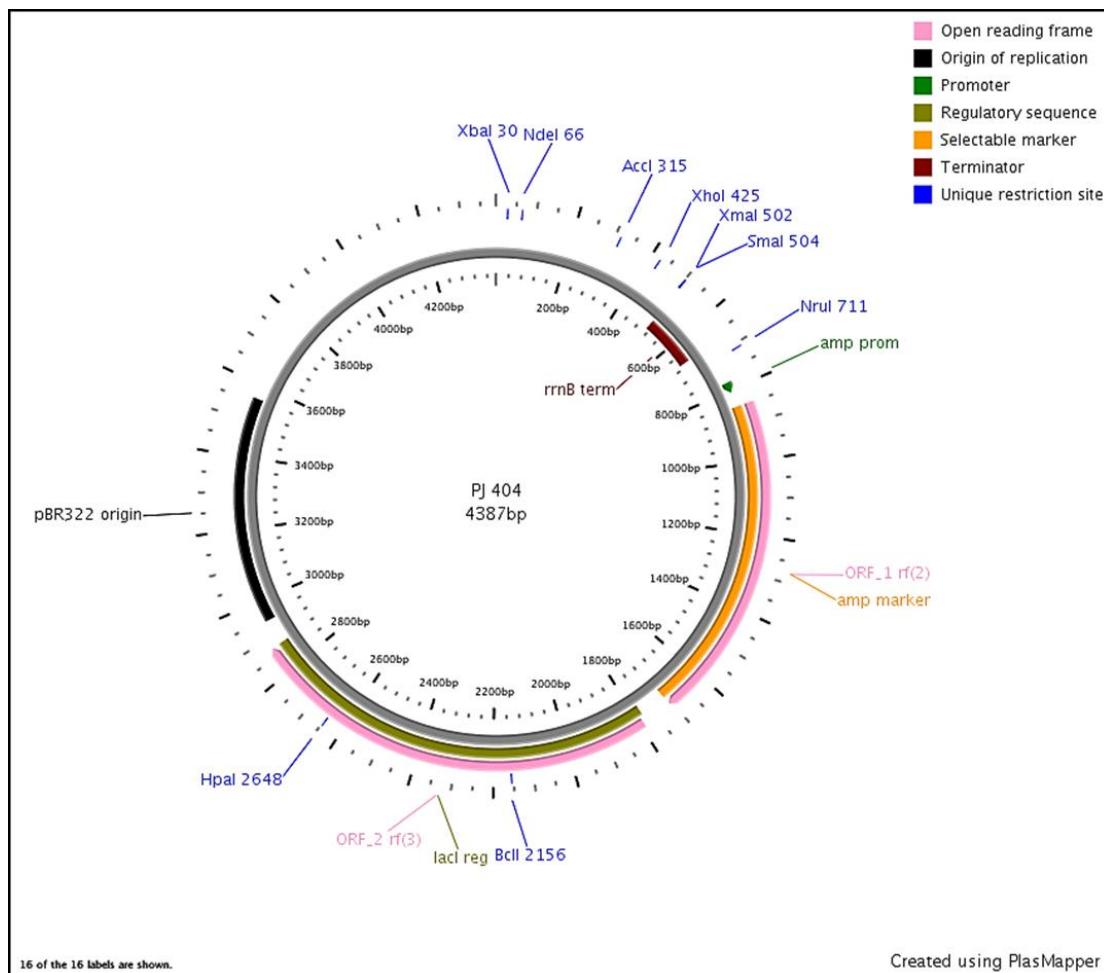
Ext. coefficient 25440

Abs 0.1% (=1 g/l) 0.650 (cas où tous les résidus Cys sont réduits).

Tête globulaire de la protéine XRCC4 (PJ 404)

Résistance : ampicilline

Tag : 6*Histidines en C-terminal



Sequence protéique pour la forme longue 1-119:

10	20	30	40	50	60
MERKISRIHL VSEPSITHFL QVSWEKTL ES FVITLTDGH SAWTGTSES EISQEADDMA					
70	80	90	100	110	120
MEKGKYVGEL RKALLSGAGP ADVYTFNFSK ESCYFFFKN LKDVSFRLGS FNLEKVENPL					
130					
ENLYFQGH ₁₁ HHHHH					

Paramètres protéiques:

- Nombre d'acides aminés: 135
- Poids moléculaire: 15511.3
- Point isoélectrique théorique: 5.96

Nombre total de résidus chargés négativement (Asp + Glu): 19

Nombre total de résidus chargés positivement (Arg + Lys): 13

- Les coefficients d'extinction molaire sont en $M^{-1} \text{ cm}^{-1}$, à 280 nm dans l'eau.

Ext. coefficient 16960

Abs 0.1% (=1 g/l) 1.093 (cas où tous les résidus Cys forment des cystines)

Ext. coefficient 16960

Abs 0.1% (=1 g/l) 1.093 (cas où tous les résidus Cys sont réduits).

Sequence protéique pour la forme courte 1-112:

10	20	30	40	50	60
MERKISRIHL	VSEPSITHFL	QVSWEKTLES	GFVITLTDGH	SAWTGTVSES	EISQEADDMA
70	80	90	100	110	120
MEKGKYVGEI	RKALLSGAGP	ADVYTFNFSK	ESCYFFFEKN	LKDVSFRLGS	FNLENLYFQG

HHHHHHHH

Paramètres protéiques:

- Nombre d'acides aminés: 128
- Poids moléculaire: 14701.4
- Point isoélectrique théorique: 6.09

Nombre total de résidus chargés négativement (Asp + Glu): 17

Nombre total de résidus chargés positivement (Arg + Lys): 12

- Les coefficients d'extinction molaire sont en $M^{-1} \text{ cm}^{-1}$, à 280 nm dans l'eau.

Ext. coefficient 16960

Abs 0.1% (=1 g/l) 1.154 (cas où tous les résidus Cys forment des cystines)

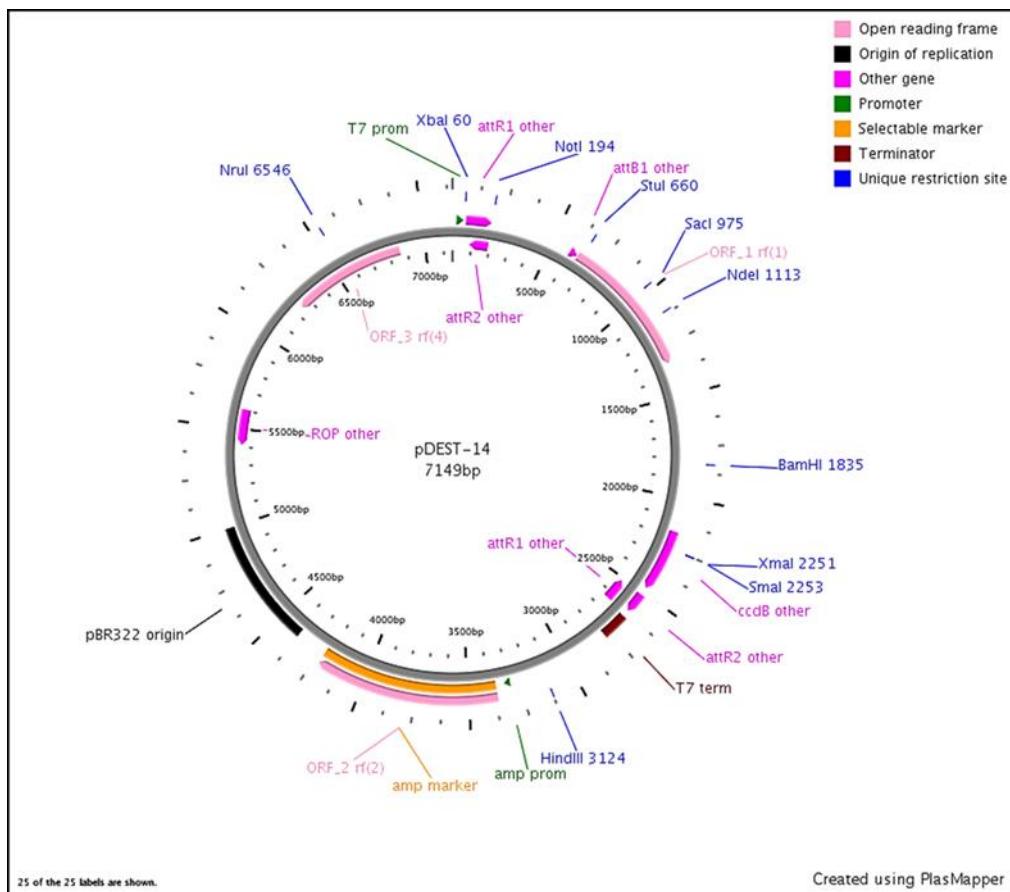
Ext. coefficient 16960

Abs 0.1% (=1 g/l) 1.154 (cas où tous les résidus Cys sont réduits).

Protéine Cer-XLF (pDEST-14)

Résistance : ampicilline

Tag : 6*Histidines en C-terminal



Sequence protéique:

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MEELEQGLLM	QPWAHLQLAE	NSLLAKVFIT	KQGYALLVSD	LQQVNHEQVD	TSVVSQRRAKE
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
LNKRLTAPPA	AFLCHLDNLL	RPLLKDAAHP	SEATFSCDCV	ADALILRVR	ELSGLPFYWN
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
FHCMLASPSL	VSQHLIRPLM	GMSLALQCQV	RELATLLHMK	DLEIQDYQES	GATLIRDRLK
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	
TEPFEENSFL	EQFMIEKLPE	ACSIGDGKPF	VMNLQDLYMA	VTTQHHHHHH	

Paramètres protéiques:

- Nombre d'acides aminés: 230
- Poids moléculaire: 26256.3
- Point isoélectrique théorique: 5.58

Nombre total de résidus chargés négativement (Asp + Glu): 28

Nombre total de résidus chargés positivement (Arg + Lys): 18

- Les coefficients d'extinction molaire sont en $M^{-1} \text{ cm}^{-1}$, à 280 nm dans l'eau.

Ext. coefficient 28335

Abs 0.1% (=1 g/l) 1.079 (cas où tous les résidus Cys forment des cystines)

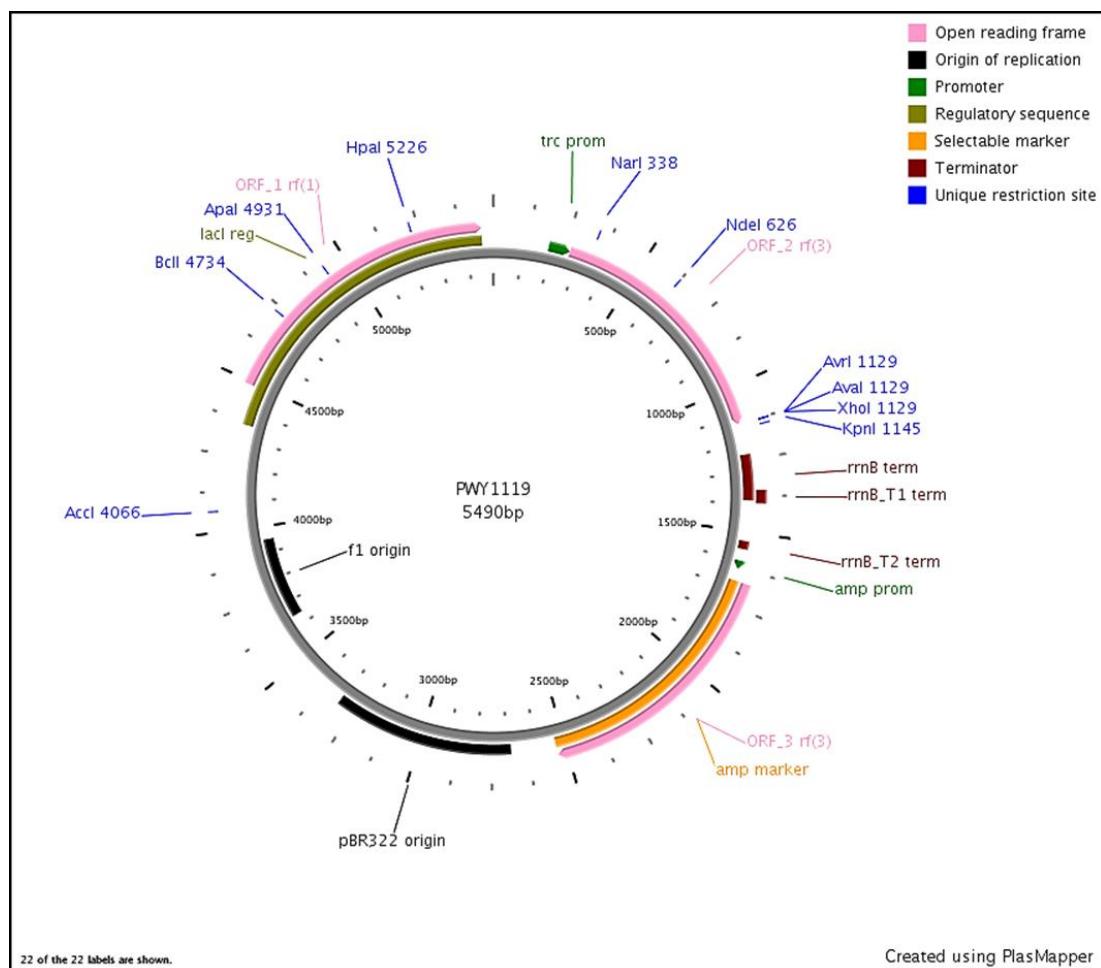
Ext. coefficient 27960

Abs 0.1% (=1 g/l) 1.065 (cas où tous les résidus Cys sont réduits).

Domaine C-terminal de la Ligase IV Cer-XLF (PWY1119)

Résistance : ampicilline

Tag : 6*Histidines en N-terminal



Sequence:

10	20	30	40	50	60
MSYYHHHHHH	DYDIPTTENL	YFQGAMGSKI	SNIFEDVEFC	VMSGTDSQPK	PDLENRIAEG
70	80	90	100	110	120
GGYIVQNP GP	DTYCVIAGSE	NIRVKNIILS	NKHDVVKPAW	LLECFKTKSF	VPWQPRFM IH
130	140	150	160	170	180
MCPSTKEHFA	REYDCYGD SY	FIDTDLNQLK	EVFSGIKNSN	EQTPEEMASL	IADLEYRYSW
190	200	210	220	230	240
DCSPLSMFRR	HTVYLD SYAV	INDLSTKNEG	TRLAIKALEL	RFHGAKVVSC	LAEGVSHVII
250	260	270	280		
GEDHSRVADF	KAFRRTFKRK	FKILKESWVT	DSIDKCELQE	ENQYLI	

Paramètres protéiques:

- Nombre d'acides aminés: 286
- Poids moléculaire: 33150.5
- Point isoélectrique théorique: 5.80

Nombre total de résidus chargés négativement (Asp + Glu): 41

Nombre total de résidus chargés positivement (Arg + Lys): 32

- Les coefficients d'extinction molaire sont en $M^{-1} \text{ cm}^{-1}$, à 280 nm dans l'eau.

Ext. coefficient 43360

Abs 0.1% (=1 g/l) 1.308 (cas où tous les résidus Cys forment des cystines)

Ext. coefficient 42860

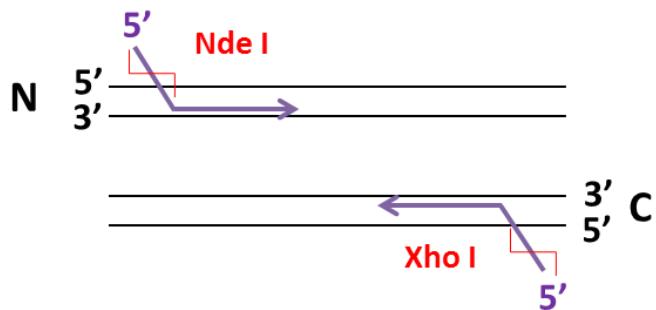
Abs 0.1% (=1 g/l) 1.293 (cas où tous les résidus Cys sont réduits).

Protocole de clonage – Application aux domaines tronqués 1-112 et 1-119 de XRCC4

→ Ce protocole de clonage nécessite une amplification PCR des fragments XRCC4 tronqués à partir de la matrice plasmidique (pET28a), et une intégration de ces domaines dans un nouveau vecteur d'expression (PJ 404). Ceci sera suivi du protocole de transformation bactérienne.

Amplification PCR sur la matrice pET28a

La concentration finale en primer dans le volume réactionnel durant la PCR ne doit pas excéder 0.5µM (une plus haute concentration pourrait inhiber la réaction de PCR). Pour une amplification extrêmement fidèle, prévoir la PCR avec l'ADN polymérase « Phusion ». Attention également à la concentration en matrice (une concentration trop élevée conduira à une contamination après purification et contaminer la réaction ultérieure de ligation, et ainsi produire des clones faux-positifs). Stratégie d'amorces à queues flottantes :



- Matrice initiale : environ 50ng/µl. Primers utilisés :
Oligonucléotide 5' Nde I (site de coupure souligné ; début de la séquence en bleu)

5' GAGATATCATATGGAGAGAGAAAATAAGCAG 3'

Oligonucléotides 3' Xho I (site de coupure souligné ; début de la séquence en bleu ; nucléotide rouge : mutation insérée afin de respecter l'usage des codons par E. Coli)

Forme courte 1-112 : 5' TCAACTTCTCGAGGTTGAAGGAAC 3'

Forme longue 1-119 : 5' TAATGACCTCGAGCGGGTTTCAACTTTC 3'

5041 TTTTGTAACTTAAGAAGGAGATATAccatggagagaaaaataaggcagaatccaccc 5100
AAAACAAATTGAAATTCTCCTCTATATGGTACCTCTCTTTTATTCTAGGTGGAA

5101 gtttctgaacccagtataactcatttctacaagtatctggagaaaaacactggaatct 5160
CAAAGACTTGGGTATATTGAGTAAAAGATGTTCATAGAACCCCTTTTGACCTTAGA

5161 gttttgttattacacttactgtatggtcattcagcatggactggacagtttctgaatca 5220
CCAAAACAATAATGTGAATGACTACCAGTAAGTCGTACCTGACCTGTCAAAGACTTAGT

5221 gagatttccaaagaagctgtatgacatggcaatggaaaaaggaaatatgttggtaactg 5280
CTCTAAAGGGTTCTCGACTACTGTACCGTTACCTTTCCCTTATACAACCACTTGAC

5281 agaaaaagcattgttgcaggagcaggaccagctgtatgtatacagttattttctaaa 5340
TCTTTCTGTAACAAACAGTCCTCGCTGGTCGACTACATATGTGCAAATTAAAAAGATT

5341 gagttcttgcattttctttggaaaaacctgtaaagatgtctcattcagacttggttcc 5400
CTCAGAACAAATAAGAAGAAACTCTTTGGACTTACAGAGTAAGTCTGAACCAAGG

5401 ttcaaccttagagaaaaagtgtaaaaaccctgtaaaggacttatttgc 5460
AAGTTGGATCTCTTCAACTTTGGTCGACTTCAGTAATCTCTGAATAAACAAATAACG

Le trait bleu représente la forme courte de la construction clonée. Le trait bleu + rouge représente la forme longue.

Programme PCR utilisé :

30s – 98°C
10s – 98°C
30s – 55°C
30s – 72°C
7min – 72°C
20°C forever

Le stock de primers est à 100µM. Un stock fille à 20µM est réalisé. Mix PCR : 34.5µl Milli-Q, 1ul de matrice, 10µl de tampon phusion, 1µl de dNTPs (d'un mix à 10mM), 1.5µl primer1 (à 20µM), 1.5µl primer2 (à 20µM), 0.5µl de polymérase phusion. Effectuer une PCR de contrôle, sans primers.

Après amplification, le produit est purifié (Kit de purification PCR de Qiagen). La concentration des amplicons est mesurée au nanodrop (ainsi que le ratio 260/280) et un aliquot est vérifié sur gel d'agarose 1% (migration 45 minutes à 50V).

Digestion des amplicons et de la matrice réceptrice (vecteur PJ 404) :

Mix digestion de la matrice : 10µl de plasmide PJ404 (environ 300ng), 10µl Milli-Q, 1µl Ndel, 1µl Xhol, 1µl Pst-1 (pour être sûr de digérer la matrice restante), 2.5µl de tampon 10X NEB 4.

Mix digestion des amplicons : 10µl amplicons, 10µl Milli-Q, 1µl Ndel, 1µl Xhol, 2.5µl de tampon 10X NEB 4.

La digestion s'effectue en deux temps pour un meilleur rendement de coupure (deux fois 30 minutes à 37°C avec 2*1µl d'enzyme de restriction). Inactivation des enzymes 10 minutes à 65°C.

Précipitation du plasmide digéré :

Ajout dans le produit de digestion précédent de

- 1/10 du volume de la réaction en 3M NaOAc pH5.2
- 3 volumes d'éthanol absolu froid (-20°C)
- Vortex et laisser 20 minutes à -80°C pour la précipitation
- Centrifuger 20 minutes à 12.000rpm et à 4°C
- Nettoyer délicatement le culot d'ADN avec de l'éthanol froid à 70%
- Centrifuger 10 minutes à 12.000rpm et à 4°C
- Sécher le culot d'ADN 5 minutes à l'air libre (attention de ne pas laisser trop sécher ou la resolubilisation sera difficile)
- Resuspendre le culot dans 15µl de tampon d'élution (EB). Diluer avec 15µl d'eau Milli-Q
- Mesurer la concentration au nanodrop.

Ligation des produits d'amplification :

Effectuer également lors de cette étape, un contrôle de ligation sans inserts.

Pour la ligation, utiliser un ratio de 1plasmide pour 4 inserts.

Mix réactionnel : 10µl d'amplicons (clivés), 5µl de plasmide (clivé), 2µl d'eau Milli-Q, 2µl de tampon T4 10X de NEB, 1µl d'ADN ligase T4. Laisser la réaction sans agitation durant 45 minutes à 16°C. Inactiver ensuite la ligase durant 10 minutes à 65°C. Pour le contrôle, utiliser de l'eau Milli-Q à la place des amplicons. Contrôler la ligation sur un gel d'agarose 1%. Si la ligation est bonne, passer à l'étape de chimio-transformation bactérienne avec le produit de ligation.

Protocole de transformation bactérienne

→ Protocole général, utilisé pour la transformation des vecteurs d'expression des protéines recombinantes XRCC4 (+ domaines tronqués), Cer-XLF et domaine C-terminal de la ligase 4.

Transformation d'un produit de ligation (à l'issu d'un clonage).

Chimio-transformation de bactéries TOP10 F' compétentes avec un produit de ligation :

- Préparer des boites LB agar + antibiotique (30 ou 50µg/ml) en nombre suffisant (en comptant les contrôles : bactéries sans plasmide, produit de ligation sans insert) ainsi qu'un bain marie chauffé à 42°C.
- Laisser sur glace 100µL de cellules durant 30 minutes dans 10µL de produit de ligation (protocole précédent).
- Chauffer les cellules dans le bain à 42°C durant 1 minute 30 (heat shock) et mettre immédiatement sur glace pendant 15 minutes.
- Ajouter 900µL de milieu LB stérile et laisser 1h à 37°C sous agitation légère.
- Etaler sur boite et laisser à 37°C sur la nuit.

Amplification plasmidique et séquençage :

- Sélectionner 2 ou 3 clones par boite et ensemencer dans 5mL de milieu LB + antibiotique pour chaque clone, sur la nuit.
- Le lendemain, culotter les cellules par centrifugation (5000 rpm) et purifier les plasmides de chaque clones via le kit de purification de Qiagen. Mesurer la concentration de chaque plasmide purifié par nanodrop.
- Vérifier éventuellement la présence de l'insert par une carte de restriction et envoyer pour séquençage.

Transformation d'un vecteur d'expression (positif à l'issu du séquençage).

Electro-transformation de bactéries (BL21 DE3 ou STAR) compétentes avec un plasmide pur, concentré et validé par séquençage :

- Préparer des boites LB agar + antibiotique (30 ou 50µg/ml) en nombre suffisant (en comptant les contrôles : bactéries sans plasmide).
- Sur glace et en conditions stériles, incuber 50µL de bactéries compétentes + 1-2µL de vecteur d'expression à 0.1 gamma.
- Electrotransfert : 2.5kV ; 4.4 secondes ; 25µF ; 200Ω. Mettre sur glace.
- Resuspendre doucement les bactéries dans 950µL de milieu LB stérile et laisser 1h à 37°C sous agitation légère pour la néo-synthèse.
- Etaler sur boite et laisser à 37°C sur la nuit.
- Le lendemain, effectuer des pré-cultures pour chaque clone obtenu. Lors de la culture, prélevé un échantillon de bactéries en phase exponentielle de croissance pour un stock glycérol à -80°C.
- Effectuer un test d'expression afin de valider ces clones.

Composition du milieu LB (Luria-Bertani) pour 1L.

- Bacto Peptone: 10g
- Bacto Yeast Extract: 5g
- NaCl: 10g
- Ajuster si besoin à pH7.0 et autoclaver pour stérilisation.

Expression des protéines du complexe de ligature

Domaine C-terminal de la ligase 4

Cette séquence protéique a été clonée au sein du plasmide PWY1119 et transformée par électroporation dans une souche *Escherichia coli* BL21 (DE3).

1. Pré-culture :

Ensemencement d'une pré-culture de 50mL en milieu riche LB en présence de 100µg/mL d'ampicilline. Laisser sur la nuit et sous agitation à 37°C.

2. Culture :

Phase de croissance : Ensemencement d'une culture de 1L en milieu riche LB en présence de 100µg/mL d'ampicilline et avec une DO₆₀₀ initiale à 0.02. Laisser sous agitation à 37°C jusqu'à une DO₆₀₀ de 0.4.

Acclimatation des bactéries 1h à 15°C sous agitation.

Phase d'induction : induction en phase exponentielle de croissance à une DO₆₀₀ proche de 0.8, par ajout de 0.5mM d'IPTG durant 16h et sous agitation à 15°C.

La culture en fin d'induction est centrifugée et le culot bactérien est repris dans un tampon PBS 1X et laissé à -80°C.

Protéines XRCC4 et Cer-XLF

Ces séquences protéiques ont été clonées au sein des plasmides Pet28a (XRCC4) et pDEST-14 (Cer-XLF) et transformée par électroporation dans les souches respectives *Escherichia coli* BL21 (DE3) et *Escherichia coli* BL21 STAR.

1. Pré-culture :

Ensemencement d'une pré-culture de 50mL en milieu riche LB en présence de 100µg/mL d'ampicilline. Laisser sur la nuit et sous agitation à 37°C.

2. Culture :

Phase de croissance : Ensemencement d'une culture de 2*2L en milieu riche LB en présence de 100µg/mL d'ampicilline et avec une DO₆₀₀ initiale à 0.02. Laisser sous agitation à 37°C jusqu'à une DO₆₀₀ de 0.4.

Phase d'induction : induction en phase exponentielle de croissance à une DO₆₀₀ proche de 0.8, par ajout de 1mM d'IPTG durant 5h et sous agitation à 37°C.

La culture en fin d'induction est centrifugée et le culot bactérien est repris dans un tampon PBS 1X et laissé à -80°C.

Complexe XRCC4/domaine C-terminal de la ligase 4

Ces séquences protéiques ont été clonées au sein des plasmides PWY1190 (XRCC4 forme courte 1-203 ; gène de résistance au chloramphénicol) et PWY1119 (C-terminal forme BRCT1-XIR-BRCT2 654-911 ; gène de résistance à l'ampicilline) et ces vecteurs ont été co-transformés par électroporation dans une souche *Escherichia coli* BL21 STAR.

1. Pré-culture :

Ensemencement d'une pré-culture de 50mL en milieu riche LB en présence de 100 μ g/mL d'ampicilline et 170 μ g/mL de chloramphénicol. Laisser sur la nuit et sous agitation à 37°C.

2. Culture :

Phase de croissance : Ensemencement d'une culture de 2L en milieu riche LB en présence de 100 μ g/mL d'ampicilline + 170 μ g/mL de chloramphénicol et avec une DO₆₀₀ initiale à 0.02. Laisser sous agitation à 37°C jusqu'à une DO₆₀₀ de 0.4.

Acclimatation des bactéries 20 minutes à 20°C sous agitation.

Phase d'induction : induction en phase exponentielle de croissance à une DO₆₀₀ proche de 0.8, par ajout de 0.5mM d'IPTG durant 3h et sous agitation à 20°C.

La culture en fin d'induction est centrifugée et le culot bactérien est laissé à -80°C.

Purification des protéines du complexe de ligature

Domaine C-terminal de la ligase 4

La lyse bactérienne est assurée par l'action combinée de la décongélation, du lysozyme, du NP-40 (déturgent) et par plusieurs cycles de sonication sur glace. Volume du lysat total : 30mL. La fraction soluble est séparée par centrifugation à 17.000 rcf durant 45 minutes et filtrée sur un filtre 0.22µm pour obtenir un lysat clair.

1. Tampon de lyse bactérienne :

Ce tampon est un tampon Ni-A (décrit ci-dessous), complémenté avec

- 1mg/mL lysozyme
- 1mM PMSF
- 0.5% NP-40
- Benzonase : 5µl/g de cellules
- Cocktail d'inhibiteurs de protéases (Roche)

2. Chromatographie d'affinité au nickel (colonne HisTrap 5mL):

Tampon Ni-A :

- 30mM tris pH8.0
- 750mM KCl
- 10% glycérol
- 10mM imidazole
- 0.2mM β-mercaptopropanoïde

Tampon Ni-B : Ni-A + 500mM imidazole.

Les fractions collectées à l'issu de cette première étape de purification sont rassemblées et diluée pour une concentration finale en KCl de 0.1M (avant la seconde étape de purification). Cette dilution se fera en tampon Q-A, décrit ci-dessous.

3. Chromatographie échangeuse d'anions (colonne HiTrap Q HP 1 ou 5mL) :

Tampon Q-A :

- 20mM HEPES pH7.5
- 50mM KCl
- 10% glycérol
- 2mM EDTA
- 1mM DTT

Tampon Q-B : Q-A + 500mM KCl.

Les fractions collectées à l'issu de cette seconde étape de purification sont rassemblées et 60mM de NDSB-195 sont ajouté extemporanément.

Protéines XRCC4 et Cer-XLF

La lyse bactérienne est assurée par l'action combinée de la décongélation, du lysozyme, du NP-40 (déturgent) et par plusieurs cycles de sonication sur glace. Volume du lysat total : 30mL. La fraction soluble est séparée par centrifugation à 17.000 rcf durant 45 minutes et filtrée sur un filtre 0.22µm pour obtenir un lysat clair.

1. Tampon de lyse bactérienne :

Ce tampon est un tampon Ni-A (décris ci-dessous), complémenté avec

- 1mg/mL lysozyme
- 1mM PMSF
- 0.5% NP-40
- Benzonase : 5µl/g de cellules
- Cocktail d'inhibiteurs de protéases (Roche)

2. Chromatographie d'affinité au nickel (colonne HisTrap 5mL):

Tampon Ni-A :

- 20mM tris pH8.0
- 100mM KCl
- 10% glycérol
- 1mM β-mercaptopropanoïde

Tampon Ni-B : Ni-A + 500mM imidazole.

Les fractions collectées à l'issu de cette première étape de purification sont rassemblées et concentrées (Amicon ultra 3K) jusqu'à un volume final de 2mL. Ce volume final est ensuite injecté dans une colonne de gel filtration.

3. Chromatographie d'exclusion (colonne Superdex 75 16/600) :

Tampon d'équilibration :

- 10mM tris pH8.0
- 150mM KCl
- 5mM DTT
- 1mM EDTA

Les fractions collectées à l'issu de cette seconde étape de purification sont rassemblées et concentrées (Amicon ultra 3K).

Complexe XRCC4/domaine C-terminal de la ligase 4

La lyse bactérienne est assurée par l'action combinée de la décongélation, du lysozyme, du LDAO (déturgent) et par 3 passages dans une presse de French. Volume du lysat total : 30mL. La fraction soluble est séparée par centrifugation à 17.000 rcf durant 45 minutes et filtrée sur un filtre 0.22µm pour obtenir un lysat clair.

1. Tampon de lyse bactérienne :

Ce tampon est un tampon Ni-A (décrit ci-dessous), complémenté avec

- 1mg/mL lysozyme
- 1mM PMSF
- Benzonase : 5µl/g de cellules
- Cocktail d'inhibiteurs de protéases (Roche)

2. Chromatographie d'affinité au nickel (colonne HisTrap 5mL):

Tampon Ni-A :

- 20mM tris pH8.5
- 500mM KCl
- 0.1% LDAO
- 10% glycérol
- 10mM imidazole
- 2mM β -mercaptoethanol

Tampon Ni-B : Ni-A + 300mM imidazole.

Les fractions collectées à l'issu de cette première étape de purification sont rassemblées, diluées pour une concentration finale en KCl de 150mM (via le tampon Q-A décrit ci-dessous) et injectées dans une colonne échangeuse d'anion.

3. Chromatographie échangeuse d'anion (colonne HiTrap Q HP 1 ou 5mL) :

Tampon Q-A :

- 20mM tris pH8.0
- 10mM DTT
- 10% glycérol

Tampon Q-B : Q-A + 500mM KCl.

Les fractions collectées à l'issu de cette seconde étape de purification sont rassemblées, dialysées dans un tampon de cristallisation et immédiatement concentrées (Amicon ultra 3K) au-delà de 7mg/ml pour les tests de cristallisation (voir partie cristallisation).

Tampon de cristallisation :

- 20mM Bis tris propane pH6.9
- 150mM KCl
- 1mM EDTA
- 10mM DTT (frais !)

Protocole de purification de protéines recombinantes sur système AKTA

Les indications entre « () » correspondent aux commandes machine.

Lavage et préparation de la colonne :

- Commande manuelle : lavage des pompes A et B (flow : 5mL/min - gradient : 100%B – execute). Attendre la stabilisation de la conductivité.
- (gradient : 0%B). Voir la stabilisation de la conductivité. (Si l'espèce chimique saline et la concentration sont les mêmes dans les tampons A et B, la conductivité au niveau du plateau doit être la même).
- (pause). Installer ou paquer la colonne sur l'appareil. Rincer ou paquer la résine avec le tampon de lavage (1mL/min) et attendre la stabilisation de la conductivité.
- Commande (end). Laver la pompe échantillon P960 dans le tampon de lavage (Sampleflow 960 : 2mL/min). Commande (end).
- Mettre la pompe échantillon dans le surnageant pour chargement.

Chargement de l'échantillon sur la colonne :

- Commande manuelle (flow 1mL/min – gradient : 0%B – execute). Attendre la stabilisation de l'UV280 et de la conductivité et lancer une commande (autozero).
- (Flow: 0mL/min – sampleflow 960: 1mL/min – methodbase 960: samplepump – injection valve : inject – execute).
- Collecter le FT dans un tube falcon Durant le chargement de l'échantillon sur la résine.

Lavage de la résine et élution de l'échantillon :

- Lorsque tout le surnageant provenant du lysat bactérien est injecté, laver abondamment la résine avec du tampon A de lavage (flow : 1mL/min – sampleflow 960 : 0mL/min – methodbase 960 : systempump – injection valve : load – execute).
- Attendre que l'UV280 soit proche de la ligne de base et démarrer le gradient (linéaire ou par pallier en fonction du protocole choisi) (gradient : x% B – x minutes).
- Au même moment, démarrer la collecte de l'échantillon éluté (fractionation 960 : x mL) en choisissant le mode de collecte ainsi que le volume.

Colonnes de gel filtration (S200 (ref))

Les indications entre « () » correspondent aux commandes machine.

Lavage et conditionnement de la colonne :

- Commande manuelle : (sans installer la colonne) lavage de la pompe A avec le tampon de conditionnement (tampon de gel filtration) (flow : 5mL/min – execute). Attendre la stabilisation de la conductivité.

- Descendre le débit à 1mL/min et le laisser durant l'installation de la colonne. Pendant cette installation, commencer par connecter le haut de la colonne à la machine tout en laissant le piston pour maintenir la colonne sous pression.
- Connecter finalement le bas de la colonne à la machine et équilibrer la résine avec un faible débit (flow : 0.2mL/min).

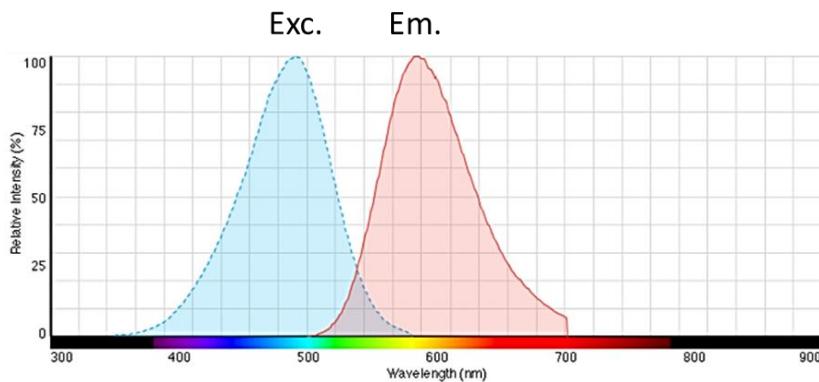
Chargement de l'échantillon :

- Après équilibration, commande (end).
- Laver avec une seringue la boucle échantillon (de volume en adéquation avec le volume d'échantillon à injecter) via le tampon d'équilibration.
- Machine en position (load). Connecter cette boucle échantillon sur la machine.
- Commande manuelle : (flow : 1mL/min – autozero – execute).
- Injecter l'échantillon dans la boucle échantillon à l'aide d'une seringue.
- Commande manuelle : (injection valve : inject).
- Au même moment, démarrer la collecte de l'échantillon élué (fractionation 960 : x mL) en choisissant le mode de collecte ainsi que le volume.

Protocole de Fluorimétrie à balayage différentiel (DSF)

Equipement et solutions :

- Plaque 96 puits (white/clear. Bio-Rad) et bande d'étanchéité (Bio-Rad).
- Thermocycleur (CFX96 – real-time PCR system. Bio-Rad). Filtres d'excitation (480/500nm) et d'émission (560/580nm).
- Sonde fluorescente : SYPRO Orange 5000X (Invitrogen) – Spectre d'excitation/émission du SYPRO ci-dessous :



Mélange réactionnel :

Le mélange s'effectue dans un volume final de 20 μ L. La protéine (dont la pureté a été au préalable vérifiée sur gel SDS-PAGE) est présente à 5 μ M. Dans le cas d'une interaction protéine-ligand, le ligand est présent à 2mM (ou l'équivalent en DMSO final dans la condition contrôle). Le SYPRO Orange est à une concentration finale de 10X. Le volume est ajusté avec le tampon de conservation de la protéine. Dans le cas d'une interaction (protéine-protéine ou protéine-ligand), une pré-incubation de 10 minutes à température ambiante sera effectuée avant d'ajuster le volume. Une fois scellée, la plaque sera centrifugée à basse vitesse afin d'enlever les bulles d'air éventuelles.

Acquisition et traitement des données :

- Logiciel CFX Manager (Bio-Rad Laboratories, USA).
- L'échantillon est chauffé de 20 à 89°C par paliers de 0.3°C/min.
- Le T_m sera calculé comme étant la température au point d'infexion de la transition observée sur la courbe de dénaturation (courbe sigmoïdale $RFU=f(t)$). Cette valeur est caractéristique de la température de demi-dénaturation de la protéine.

Protocole de Différence de Transfert de Saturation (STD-NMR)

Equipement et solutions :

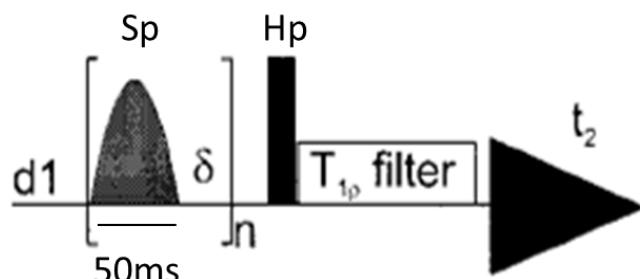
- D2O (pour le lock)
- Solution de protéine pure et concentrée
- Solution stock de ligand concentré dans 100% DMSO d6
- Tube capillaire 3mm (volume final 165µL)
- Spectromètre Bruker Avance 600MHz avec cryosonde TXI
- Logiciel RMN Topspin 3.2 (Bruker).

Mélange réactionnel :

Le mélange s'effectue dans un volume final de 160µL. La protéine (dont la pureté a été au préalable vérifiée sur gel SDS-PAGE) est présente à 5µM final. Le ligand (dont la pureté et la concentration soluble maximale a été au préalable vérifiée dans le tampon de la protéine) sera initialement testé à 500µM final. Le volume est ajusté avec le tampon de conservation de la protéine + 10%D2O final. La température standard utilisée est de 298K (ou 283K pour les protéines instables comme le domaine C-terminal de la ligase 4).

Acquisition et traitement des données :

Séquence d'impulsion :



d_1 : délai de relaxation : 100ms

Sp : Shape pulse : pulse gaussien d'irradiation sélective de 50ms à 86Hz

On-résonance : -600.13Hz

Off-résonance : 18000Hz

Hp: Hard pulse 90°

Δ : 1ms

Filtre T1 rho: spin-lock pour suppression du signal de la protéine

Etapes :

- Attendre la stabilisation de la température au niveau de l'échantillon.
- Lock ; Wobb ; Shim (charger une matrice de shim, puis commande topshim).
- Calibrer l'impulsion (hard pulse).
- Régler les paramètres d'acquisition des données.
- Régler le RG.

- Lancer la manip (en multizg avec manips ISTD + Io)

Démarche expérimentale :

- Recherche des concentrations en protéine et ligand adéquates afin d'avoir un bon signal/bruit (en général, ratio protéine : ligand 1 :100 avec 5µM en protéine). Régler la température en fonction de la stabilité intrinsèque du récepteur. Optimiser le temps de saturation.
- Si signal STD : effectuer un contrôle sans protéine afin de déterminer la proportion de signal parasite non spécifique.
- Le signal STD est-il spécifique ? : (a) un effet « epitope mapping » peut-être effectué en utilisant par exemple un contrôle BSA (pour un mode de liaison spécifique et unique, on distinguera les protons exposés au solvant des protons en contact avec la protéine). (b) une titration permettra également de définir la spécificité d'interaction.
- Extraction du Kd. Courbe hyperbolique dose-réponse $A_{STD} = f([ligand])$, à ajuster à l'équation de Langmuir :

$$A_{STD} = (\alpha_{STD} * [L]) / ([L] + Kd)$$

Avec A_{STD} le facteur d'amplification du STD (calculé comme étant : excès de ligand * (Io-Isat)/Io)) ; α_{STD} , la valeur A_{STD} maximale au plateau de la courbe et

[L] la concentration en ligand.

- Confirmer ces données par des expériences biophysiques orthogonales (DSF, ITC...).

Gels d'electrophorèse : Tricine-SDS-PAGE et gels natifs

Tricine-SDS-PAGE :

Composition (calculée pour un gel, avec un espace inter-plaques de 0.75mm - BioRad) :

→Running gel (10%)

- 30% Acryl/Bis (37.5 :1) : 2.0ml
- 3M tris pH 8.4 : 2.0ml
- 50% glycérol : 2.0ml
- 10% SDS : 60.0µl
- 10% APS : 75.0µl
- TEMED : 7.5µl

→Stacking gel (7.5%)

- 30% Acryl/Bis (37.5 :1) : 0.75ml
- 3M tris pH 8.4 : 0.75ml
- H2O : 1.50ml
- 10% SDS : 30.0µl
- 10% APS : 30.0µl
- TEMED : 3µl

Préparation de l'échantillon protéique :

- 1M DTT : 1µl
- Echantillon : 10µl
- Tampon SDS de chargement (2X): 10µl

Migration :

- Tampon anode (10X) : 1M tris-HCl, pH 8.9 (utilisé à 1X)
- Tampon cathode (10X) : 1M tris, 1M tricine, 1% SDS, pH 8.25 (utilisé à 1X)
- Voltage pour le stacking gel : 50V
- Voltage pour le running gel : 175V

Fixation et coloration des protéines :

- 50% méthanol + 10% acide acétique
- Coloration : 0.025% Coomassie dans 40% méthanol + 10% acide acétique (30 minutes)
- Décoloration : 40% méthanol + 10% acide acétique (60 minutes)
- Autre méthode de coloration possible : Instant-Blue (Expedeon), sans décoloration ultérieure.

Gels natifs (suivi d'interactions protéine-protéine):

Composition (calculée pour un gel, avec un espace inter-plaques de 0.75mm - BioRad) :

→Running gel (6%) – pas de stacking gel.

- 30% Acryl/Bis (37.5 :1) : 2.0ml
- Tampon TBE pH 8.0 (10X) : 1.0ml
- H2O : 7.0ml

- 10% APS : 75.0µl
- TEMED : 7.5µl

Composition du tampon TBE (pour 1L, concentré 10X):

- tris base : 108g
- Acide borique : 55g
- Na₂EDTA : 9.3g

Milieu réactionnel :

Le volume final de la réaction est de 10µL. Le tampon réactionnel est un tampon contenant 5% de glycérol (20mM tris pH8.0, 50mM KCl, 0.1mM DTT, 5% glycérol).

-La réaction démarre en incubant 10pmoles de chacun des partenaires protéiques durant 15 minutes et à température ambiante. Puis, le volume est ajusté à 10µL avec du tampon réactionnel.

-Dans le cas de l'utilisation d'un inhibiteur : pré-incubation préalable entre l'inhibiteur (ou contrôle DMSO) avec la protéine durant 15 minutes et à température ambiante. Puis ajout du second partenaire protéique et incubation 15 minutes à température ambiante. Finalement, le volume est ajusté à 10µl avec du tampon réactionnel.

Migration :

- Tampon TBE pH8.0 1X (utilisé à la fois en tant que tampon anode et tampon cathode)
- Pré-migration du gel, sans chargement des protéines : 30 minutes à 150V et à 4°C
- Migration après chargement : 1h30 à 150V et à 4°C

Fixation et coloration des protéines : même protocole.

Protocole de stimulation de la ligase T4 par le complexe XRCC4/Cer-XLF humain

Conditions et solutions:

- 1µM final en protéines XRCC4 et Cer-XLF (stocks -80°C à 10µM).
- 1µL final en ligase T4 NEB (d'une dilution 1/10 soit 40U)
- 100ng final d'ADN plasmidique linéarisé (double digestion XbaI - HindIII)
- 1% final de solution de sarcosyl
- 2mg/ml final de pronase
- 1mM final d'ATP
- 2mM final de MgCl₂
- Tampon de conservation des protéines (identique pour XRCC4 et Cer-XLF) – Pour le contenu de ces tampons, voir la partie « purification des protéines ».
- Gel d'agarose 0.8%
- Tampon de migration : TAE 1X final
- Tampon de charge ADN 1X final
- Solution de BET (révélation par trempage du gel)

Double digestion de la matrice ADN :

Pour un volume final réactionnel de 100µl :

- 20µl de plasmide pET28a à 700ng/µl
- 2µl d'enzyme XbaI + 2µl d'enzyme HindIII
- 10µl de tampon « fast digest »
- 68µl d'eau distillée.

La réaction s'effectue en deux étapes (pour un meilleur rendement de coupure) : deux fois 30 minutes à 37°C (avec à chaque fois 1µl des enzymes respectives).

Une fois la réaction de coupure terminée, purification de l'ADN linéaire via le kit Qiagen : »PCR Purification Kit ». Tester l'efficacité et le rendement de coupure sur un gel d'agarose 0.8%.

Composition du tampon TAE (solution 1X) :

- 40mM tris
- 20mM acide acétique
- 1mM EDTA

Démarche expérimentale :

Le volume final réactionnel est de 10µL. Chaque aliquot (protéique ou d'ATP) est utilisé une seule fois puis jeté. La réaction de ligation par la ligase T4 s'effectue à 16°C.

➔ Stimulation sans inhibiteur :

1 : Mélanger l'ATP, le MgCl₂, l'ADN et le tampon des protéines (faire un mégamix et répartir un volume égal de ce mélange dans chaque tube).

2 : Ajouter les protéines XRCC4 et Cer-XLF (sauf dans le contrôle T4 sans stimulation).

3 : Ajouter la Ligase T4 et laisser la réaction à 16°C, sans agitation durant 30 minutes.

4 : Ajouter la pronase et le sarcosyl et incuber 30 minutes à 55°C.

5 : Centrifuger les tubes pour avoir un volume constant.

6 : Ajouter le tampon de charge ADN.

7 : Déposer 10µl de la réaction sur un gel d'agarose 0.8% et faire migrer 1h15 à 50V.

➔ Stimulation en présence d'un inhibiteur de XRCC4 :

1 : Mélanger l'ATP, le MgCl₂, l'ADN et le tampon des protéines (faire un mégamix et répartir un volume égal de ce mélange dans chaque tube).

2 : Ajouter la protéine XRCC4 (sauf dans le contrôle T4 sans stimulation).

3 : Pré-incuber l'inhibiteur (ou le contrôle DMSO) à la concentration désirée, durant 15 minutes à température ambiante.

4 : Ajouter la protéine Cer-XLF et laisser incuber durant 5 minutes à température ambiante.

5 : Ajouter la Ligase T4 et laisser la réaction à 16°C, sans agitation durant 30 minutes.

6 : Ajouter la pronase et le sarcosyl et incuber 30 minutes à 55°C.

7 : Centrifuger les tubes pour avoir un volume constant.

8 : Ajouter le tampon de charge ADN.

9 : Déposer 10µl de la réaction sur un gel d'agarose 0.8% et faire migrer 1h15 à 50V.

Révélation :

Trempeage 5 à 10 minutes dans une solution fraiche de BET et révéler sous UV.

Test d'intercalation à l'ADN

Conditions et solutions:

- 100ng d'ADN plasmidique linéarisé (pET28a)
- Solution stock de Doxorubicine à 3.5mM (contrôle positif)
- Solution stock de l'inhibiteur à tester (#57632)
- Eau Milli-Q
- Gel d'agarose 1%
- Tampon de migration : TAE 1X final
- Tampon de charge ADN 1X final
- Solution de BET (révélation par trempage du gel).

Démarche et conditions expérimentales :

Le volume final réactionnel est de 10µL. 100ng d'ADN linéarisé sont incubé en présence de concentrations croissantes en doxorubicine (intercalant, témoin positif) ou de l'inhibiteur à tester durant 15 minutes à 37°C sans agitation. Le volume est ajusté avec de l'eau milli-Q. Une fois la réaction terminée, le tampon de charge est ajouté et le gel est mis à migrer durant 1h à 50V.

Révélation :

Trempage 5 à 10 minutes dans une solution fraîche de BET et révéler sous UV.

Tests NHEJ *in vitro* avec extraits cellulaires

Test des inhibiteurs de l'interaction XRCC4/Cer-XLF

Préparation des extraits cellulaires :

Les extraits cellulaires utilisés proviennent de cellules HCT116 sauvages et déficientes en ADN Ligase IV humaine. Ils ont été préparés d'après le protocole précédemment décrit (Buck et al, 2006).

1. Des cellules en phase exponentielle de croissance sont lysées dans un tampon de lyse (25 mM tris-HCl pH 7.5, 333 mM KCl, 1.5 mM EDTA pH 8.0, 4 mM DTT, cocktail d'inhibiteurs de protéases-phosphatases (Pierce, Thermo Scientific)).
2. Les lysats sont laissés durant 20 minutes à 4°C puis clarifiés par centrifugation.
3. Le surnageant est dialysé dans un tampon : 20 mM tris-HCl pH 8.0, 100 mM acétate de potassium, 20% glycérol, 0.5 mM EDTA pH 8.0 et 1 mM DTT.
4. La concentration en protéine totale est mesurée par la méthode de Bradford (Bio-Rad), et les extraits sont stockés à -80°C.

Essais NHEJ *in vitro* avec les extraits cellulaires HCT116 :

Le volume réactionnel total est de 10 μ l

1. Pré-incubation de 10 minutes à 4°C de 40 μ g d'extrait cellulaire (sauvage ou mutant) précédemment dialysé avec l'inhibiteur
2. Incubation de l'extrait cellulaire avec 10ng de plasmide linéarisé durant 1h à 25°C et dans un tampon EJ (50mM Triethanolamine pH 8.0, 0.5mM acétate de magnésium, 1mM DTT, 0.1mg/ml BSA, 60mM acétate de potassium).
3. La réaction est initiée après ajout de 1mM d'ATP (Sigma-Aldrich).
4. Traitement des échantillons avec 100 μ g/ml de RNase A à 37°C et déprotéinisation
5. Analyse des produits de ligation sur gel d'agarose 0.7%.

Révélation :

Le gel d'agarose est trempé dans une solution de SYBR-Green (Invitrogen). La fluorescence est détectée et analysée via un appareil Typhoon fluorimager (Molecular Dynamics).

Cristallisation

Cristallisation de la protéine XRCC4 (aa 1-203):

1. Production du complexe XRCC4/domaine C-terminal de la ligase 4 (voir protocole ci-dessus).
2. Les fractions collectées à l'issu de la seconde étape de purification du complexe sont rassemblées et dialysées dans un tampon de cristallisation (une première fois 2h à 4°C dans 2L puis une seconde fois sur la nuit à 4°C dans 2L).
3. Le complexe protéique dialysé est centrifugé (4°C – 16.000 rcf) afin d'enlever de potentiels agrégats ou précipités et est immédiatement concentré (Amicon ultra 3K) au-delà de 7 à 8 mg/ml.

Tampon de cristallisation (de dialyse):

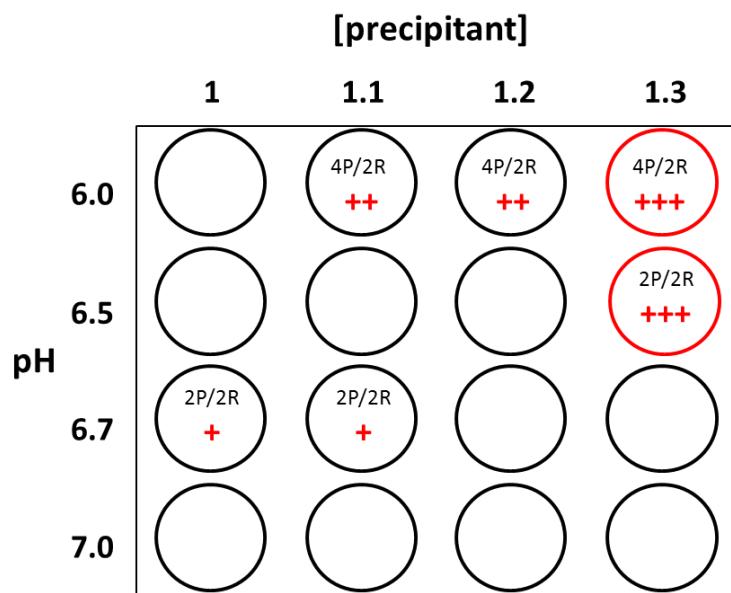
- 20mM Bis tris propane pH6.9
- 150mM KCl
- 1mM EDTA
- 10mM DTT (frais !)

4. Le complexe est mélangé à une solution de précipitation dont la composition est la suivante :
 - 100mM Cacodylate de sodium pH6.5
 - 10mM DTT
 - 1.2M sulfate d'ammonium
 - 10mM acétate de magnésium

La méthode de la goutte suspendue est utilisée lors de cette cristallisation. L'agent précipitant est ici le sulfate d'ammonium. Cette solution de précipitation a été optimisée et utilisée au niveau du réservoir. La température est de 293K.

Le Cacodylate de sodium, couplé au DTT va réagir de manière covalente avec les cystéines présentes au sein de ce complexe et ainsi dissocier ce dernier. Le domaine C-terminal de la ligase 4, instable, va précipiter et la protéine XRCC4 cristallise seule.

Lors de cette étude, plusieurs conditions de pH et de concentrations en sulfate d'ammonium ont été testées (en plaques 24 trous) et sont représentés ci-dessous :



Les ratios P/R correspondent aux volumes en μ l de protéine/solution de réservoir mélangés. Les cercles rouges représentent les conditions ayant donné les meilleurs cristaux (taille, régularité) en 72h.

Solution de cryo-protection utilisée :

- Solution de réservoir
- 20% Ethylène glycol

Expérience de trempage des cristaux XRCC4 (cas de l'inhibiteur #57632):

1. Préparation d'une solution de cryo-protection identique à celle utilisée ci-dessus
2. Détermination de la limite de solubilité du composé à tester, dans cette solution de cryo-protection (voir l'absence de précipité sous la binoculaire)
3. Déterminer sur un premier cristal, la résistance au DMSO (à la concentration maximale qui sera par conséquent utilisée)
4. Préparation d'un mélange : solution de cryo-protection/composé à tester à la concentration maximale
5. Trempage d'un second cristal de quelques secondes à plusieurs minutes dans ce mélange, en fonction de la résistance de ce dernier (plusieurs trempages successifs dans différentes solutions préparées à l'avance sont possibles)
6. Congélation dans l'azote liquide avant exposition au rayonnement synchrotron.

Expérience de cocrystallisation avec XRCC4 (cas de l'inhibiteur #57632):

Lors de cette thèse, l'expérience de cocrystallisation a été testée entre la protéine XRCC4, dissociée du complexe XRCC4/C-terminal ligase 4 et l'inhibiteur #57632. Les conditions de cocrystallisation testées ont été les mêmes que celles présentées ci-dessus, en plaques 24 puits. Les différentes étapes ont été :

1. Préparation d'un stock frais de complexe XRCC4/domaine C-terminal de la ligase 4 concentré
2. Détermination de la limite de solubilité du composé à tester (#57632), dans le tampon de cristallisation (de dialyse) du complexe (voir l'absence de précipité sous la binoculaire).
3. Incubation à température ambiante et sur roue durant une heure, d'un mélange complexe/inhibiteur (concentration finale en inhibiteur utilisée : 5mM).
4. Centrifugation à basse vitesse du mélange afin d'enlever toute trace de précipité potentiel
5. Mélange de cette solution de protéine/inhibiteur avec les différentes conditions et solutions de précipitation citées ci-dessus (la protéine sans inhibiteur sera testée également en parallèle en contrôle).

Expérience de cocrystallisation avec le complexe XRCC4/domaine C-terminal de la ligase 4 (cas de l'inhibiteur #57632):

1. Préparation d'un stock frais de complexe XRCC4/domaine C-terminal de la ligase 4 concentré
2. Détermination de la limite de solubilité du composé à tester (#57632), dans le tampon de cristallisation (de dialyse) du complexe (voir l'absence de précipité sous la binoculaire).
3. Incubation à température ambiante et sous agitation durant une heure, d'un mélange complexe/inhibiteur (concentration finale en inhibiteur utilisée : 5mM concentration finale en DMSO : 2.5%).
4. Centrifugation à basse vitesse du mélange afin d'enlever toute trace de précipité potentiel
5. Crible en plaques 96 puits de différentes conditions de cristallisation : complexe non dissocié/inhibiteur (avec en parallèle le complexe sans inhibiteur).

Les 5 cribles : JCSG IV ; JCSG Plus ; PACT ; PEG Suite et PEG II de Qiagen ont été testé pour les deux conditions complexe/inhibiteur et complexe seul.

Préparation des fichiers d'entrée du docking

A/ Préparation de la chimiothèque (à partir d'un format sdf initial).

Les molécules sont converties en format mol2, nettoyées, converties en coordonnées 3D et finalement converties en format pdbqt et séparées en fichiers individuels. Commande :

```
1 « obabel –isdf name.sdf –omol2 –O name.mol2 –r –h –p7 --gen3d »
```

```
2 « obabel –imol2 name.mol2 –opdbqt –O name.pdbqt –m »
```

B/ Préparation des fichiers de configuration de la protéine (fichier Vina_conf.dat)

1. pour le docking rigide :

receptor = name.pdbqt

center_x = value

center_y = value

center_z = value

size_x = value

size_y = value

size_z = value

out = out.pdbqt

log = log.txt

cpu = 8

exhaustiveness = 80

num_mode = 20

```
energy_range = 2
```

2. pour le docking flexible:

```
receptor = name.pdbqt
```

```
flex = name.pdbqt
```

```
center_x = value
```

```
center_y = value
```

```
center_z = value
```

```
size_x = value
```

```
size_y = value
```

```
size_z = value
```

```
out = out.pdbqt
```

```
log = log.txt
```

```
cpu = 8
```

```
exhaustiveness = 80
```

```
num_mode = 20
```

```
energy_range = 2
```

Docking haut débit

Les campagnes de criblage par docking haut-débit sont réalisées via AutoDockTools v1.5.4 et Autodock Vina. Les différentes étapes sont développées ci-dessous :

1. Les structures des protéines cibles sont extraites de la banque PDB et leurs coordonnées pdbqt sont préparées via AutoDockTools v1.5.4. Durant cette préparation des protéines, les hydrogènes polaires sont ajoutés.
2. Les dimensions de la grille sont définies avec un spacing de 1 Angstrom. Le centre de la grille est placé au niveau du centre de gravité de la protéine. Ces coordonnées sont notées et précisées dans le fichier « Vinaconf » de configuration du récepteur.
3. Les molécules issues de la chimiothèque que l'on souhaite cribler, sont converties en fichiers pdbqt via le programme Open Babel 2.2.3 avec addition des hydrogènes, calcul de l'état de charge à pH7.2 et calcul des coordonnées 3D. Ces molécules ont également été au préalable filtrées d'après des critères physico-chimiques prédéfinis ci-dessous (programme maison « Sfilter ») :

Molecular weight limits
logP
Solubility Forecast Index (logD + #Ar)
Hydrogen Bond Donors
Hydrogen Bond Acceptors
No. of rotatable bonds
No. of ring systems
No. of rings
No. of "N" & "O" atoms
Charges

4. Au cours de l'étape de docking, les chaines latérales des résidus sont conservées dans un état rigide ou certaines ont été choisies comme flexible (fichier flex = name.pdbqt ajouté dans le fichier de configuration de la protéine). L'étape de docking est réalisée sur 8 CPUs/nœud (25 nœuds au total) avec une valeur « exhaustiveness » de 80 et un maximum de 20 conformères par molécule en sortie de docking.
5. Les ligands issus du docking sont classés par leur énergie d'interaction théorique définie par le score de Vina (en kcal/mol). Les ligands retenus après inspection visuelle sont intégrés dans une étape de simulation de dynamique moléculaire du complexe protéine/ligand correspondant.

Traitement des fichiers de sortie du docking

1. En fin de docking, les fichiers de sortie out.pdbqt (avec l'ensemble des poses) sont regroupés dans un répertoire Total_out
2. Lancement d'un programme maison en fortran « extract_model » afin d'extraire seulement les 1ères poses de chaque fichier (1^{er} conformères) ; Ces nouveaux fichiers sont ensuite convertis en format pdb et stockés dans un nouveau répertoire lig_01
3. Un fichier filelist.dat est créé et contient dans une colonne le nom de chaque fichier
4. Lancement d'un programme maison « script_grep » afin d'enlever dans tous ces fichiers les lignes contenant la partie résidu flexible. Stocker les nouveaux fichiers dans un répertoire de sortie out_grep
5. Créer une liglist et une reclist nécessaires au lancement du programme maison « ligdist », qui génère les centres de gravité de chaque ligand par rapport au centre de gravité des résidus de la protéine sélectionnés dans la reclist.
6. Ne conserver que les ligands proches du site d'intérêt (zone criblée) d'après la valeur seuil fixée. Lancement du programme « vina_sort.py » afin de ne garder que les x premiers ligands (d'après le score de Vina).
7. Inspection visuelle des contacts protéine-ligand pour chaque candidat et sélection des molécules touchées les plus intéressantes (en terme de nombre de contacts polaires et hydrophobes).
8. Lancement des simulations de dynamique moléculaire pour chaque couple protéine-ligand sélectionné.

Préparation des fichiers d'entrée de dynamique moléculaire

Création des fichiers d'entrée (topologie et coordonnées) via AMBER :

1. Protonation du ligand, sauvegarde en fichier MOL2. Commande : « babel –ipdb name.pdb –omol2 –O name.mol2 –h »
2. Utilisation d'antechamber pour créer les fichiers d'entrée pour tleap. Commande : « antechamber –nc 0 –rn LIG –i name.mol2 –fi mol2 –o name.prep –fo prepc –c bcc –s 2 » (nc = net charge : s'assurer au préalable de l'état de charge de la molécule ; rn = residue name ; c = charge du modèle ; s = niveau de verbosité).
3. Vérifier que tous les paramètres sont disponibles. Commande : « parmchk –name.prep –o name.frcmod »
4. Utilisation de tleap pour créer les fichiers de topologie et de coordonnées. Commandes séquentielles :

```
« tleap »  
« source leaprc.ff03 »  
« source leaprc.gaff »  
« loadamberprep name.prep »  
« list » (s'assurer que LIG est bien dans la liste)  
« loadamberparams name.frcmod »  
« check LIG »  
« saveamberparm LIG name.prmtop name.inpcrd »  
« savepdb LIG nameNEW.pdb »
```

5. Vérification du fichier protéine.pdb (Prendre une protéine déprotonée : tleap ajoute lui-même les hydrogènes explicites). Commandes séquentielles :

```
« tleap »  
« source leaprc.ff03 »  
« mol = loadpdb protein.pdb »  
« list » (s'assurer que mol est bien dans la liste)  
« check mol »  
« quit » (si l'unité est OK).
```

6. Combinaison des fichiers pdb de la protéine et du ligand. Sauvegarder e tant que « protall.pdb »
7. Utilisation de tleap pour combiner toutes les données en un seul set de fichiers moléculaires. Commandes séquentielles :

```
« tleap »  
« source leaprc.ff03 »  
« source leaprc.gaff »  
« loadamberprep name.prep »  
« loadamberparams name.frcmod »  
« MOL = loadpdb protall.pdb »  
« check MOL »  
« solvatebox MOL TIP3PBOX 10 »  
« charge MOL »  
« addions MOL Na+ 0 »  
« saveamberparm MOL protallsol.prmtop protallsol.inpcrd »  
« savepdb MOL protallsol.pdb »  
« quit »
```

8. Création d'un fichier protallsol0.rst à partir du fichier protallsol.inpcrd

Simulations de dynamique moléculaire

Les simulations de dynamique moléculaire ont été réalisées :

- Sur les complexes protéiques XRCC4/C-ter Lig4 (PDB : 3II6) et XRCC4/Cer-XLF (PDB : 3SR2) durant 10ns ;
- Sur la protéine C-ter Lig4 (PDB : 3II6) séparée de XRCC4 durant 15ns ;
- Sur les 29 complexes C-ter Lig4/molécule durant 10ns ;
- Sur les 86 complexes XRCC4 (1-126; PDB : 1FU1)/molécule durant 10ns ;
- Sur les 28 complexes Cer-XLF (1-127)/molécule durant 10ns ;
- sur la protéine XRCC4(res.1-126 ; PDB : 1FU1) durant 10ns.

Ces simulations ont été réalisées grâce au programme Amber 12 incluant les champs de force ff03 et GAFF respectivement pour les protéines et les ligands.

1. Les systèmes sont protonés et centrés au sein d'une boîte périodique solvatée de type TIP3P, contenant des ions positifs Na⁺ ou négatifs Cl⁻ pour la neutralité du système.
2. L'équilibration du système s'effectue en plusieurs étapes. Tout d'abord l'énergie du système est minimisée via 1000 cycles de l'algorithme SD, avec le soluté gardé fixe, et par contrainte de ses coordonnées cartésiennes en utilisant un potentiel harmonique avec la constante de force k égale à 100 kcal/mol/Å².
3. Une seconde étape consiste à minimiser l'énergie via 500 cycles de SD et 1500 cycles de l'algorithme de gradient conjugué (CG), avec le soluté sous contraintes faibles ($k = 10$ kcal/mol/Å²). Une courte simulation de DM de 20ps est réalisée sur le soluté sous contraintes faibles et avec une variation linéaire de la température de 0 à 300 K. Le contrôle de la température est réalisé grâce à la dynamique de Langevin, avec le paramètre de fréquence de collision γ égal à 1.0 ps⁻¹. Le pas d'intégration au cours de ce calcul est de 1fs. Une valeur seuil de 12 Å au cours du calcul est fixée pour les interactions électrostatiques.
4. La simulation de DM continue durant 100ps à une température constante de 300 K, sans contraintes et avec un pas d'intégration de 2fs.
5. Finalement un calcul de 10 à 15ns est lancé, avec une pression constante de 1 bar et des coordonnées atomiques sauvegardées chaque 10ps. La dynamique de Langevin est utilisée pour contrôler la température, avec $\gamma = 1.0$ ps⁻¹ et la pression est contrôlée via le barostat anisotropique de Berendsen

avec un temps de relaxation de pression $\tau_p = 2$ ps. Les liaisons impliquant des hydrogènes sont contraintes via l'algorithme SHAKE.

Les trajectoires obtenues à la fin de ces simulations de dynamique moléculaire sont analysées via les programmes de la suite AmberTools (scripts « **ptraj** »), ainsi que des programmes écrits au sein du laboratoire (« **trajan** »). Les énergies d'interactions intermoléculaires protéine-ligand sont déterminées via la méthode **MMPBSA**.

Analyse « ptraj »

Le fichier « protallsol5.mdcrd » est obtenu à la fin du calcul de la dynamique moléculaire et contient l'ensemble de la trajectoire.

Analyse du RMSD :

```
trajin protallsol5.mdcrd 1 9999
center :1-X mass origin
image origin center familiar
strip : X-Y
rms first mass out RMSD.txt time 10 :X
```

Analyse de la distance entre deux atomes :

```
trajin protallsol5.mdcrd 1 9999
center :1-X mass origin
image origin center familiar
distance proteinZ :xx@XX :xx@XX out proteinZ_dist.dat
```

« Trajan »

Pour déterminer les contacts persistants, toutes les distances intermoléculaires sont suivies au cours d'une trajectoire afin de déterminer celles qui restent au-dessous d'un seuil prédéfini dans la quasi-totalité des cadres.

« MMPBSA »

Deux approches ont été utilisées : la méthode de Poisson-Boltzman et la méthode généralisée de Born.

Echantillonnage conformationnel

Pour regrouper les conformères d'une protéine, le programme « kclust » du MMTSB Tool Set est utilisé, avec le rayon de 2Å. Les structures retenues sont celles qui sont les plus proches des « centroides » de chaque « cluster ».

Minirevue:

“Looking for human DNA ligases inhibitors for combination anticancer therapy”

(En cours de finalisation – Journal visé :)

Auteurs :

Grégory Menchon^{1,2,3}, Alain Milon^{1,2}, Georges Czaplicki^{1,2}, Patrick Calsou^{1,2,3}

¹ CNRS; IPBS (Institut de Pharmacologie et de Biologie Structurale); BP 64182, 205 route de Narbonne, F-31077 Toulouse, Cedex4, France

² Université de Toulouse; UPS; IPBS; F-31077 Toulouse, France

³ Equipe labellisée Ligue Nationale Contre le Cancer

Looking for human DNA ligases inhibitors for combination anticancer therapy

Combined Radio- and Chemo-therapies aim at damaging cancer cells by disrupting the DNA backbone molecule and mainly leading to apoptosis. Human DNA ligases are ATP-dependent ligases and participate in numerous pathways of the DNA Damage Response. The search for specific human ligases activity inhibitors is growing and concomitant with the discovery of ligases 3D structures. This inhibition is considered as one of the strategic weapons in combination anticancer therapies.

To fight against tumor development, several weapons, alone or in combination, are used. Radiotherapy is the major one and is used today to treat most of cancers. To face the intrinsic or acquired radiation resistance mechanism a combined treatment with chemotherapeutic molecules is used. This treatment aims at increasing tumor sensitization and decreasing healthy cells cytotoxicity (1). Combination of an oxidative stress with ionizing radiations causes massive DNA damages as single and double strand breaks (2). The latter is the most toxic effect for the cell and need to be repaired for correct cell living and proliferation by maintaining genome integrity (3). In addition to multidrug resistance pathways, tumors can develop resistance to specific genotoxic injuries through several mechanisms. One of these mechanisms is the over-expression of some DNA repair proteins (1). Several reviews aimed at describing the advantages to inhibit DNA repair pathways to increase the efficacy of antineoplastic drugs in cancer cells (1) (4). The strategy to compromise DNA repair enzymes was validated for several target in numerous in vitro and in vivo experiments by modulating their expression level (1) (4). From these proteins, human DNA ligase seems to be a prime target.

Human ligases (encoded by the mammalian LIG1, LIG3 and LIG4 genes) are numerous and depending on specific protein-protein interactions, they are committed to different repair pathways associated to specific protein partners either with a structural or catalytic role (5). These enzymes have a conserved catalytic domain and differ by their N- and C-terminal parts (5). These ligases use ATP as enzymatic cofactor and the universal ATP-dependent DNA ligation mechanism is depicted in Figure1. In mammalian cells, DNA ligase 1 (hLig1) is involved in DNA replication (with the ligation of Okazaki fragments) and repair (6) (7). This protein is quite ubiquitous in repair pathways as it is involved in Base Excision Repair (BER) (8), Nucleotide Excision Repair (NER) (9), Mismatch Repair (MMR) (10) and Microhomology-Mediated-End-Joining (MHEJ) pathway (11). DNA ligase 3 (hLig3) is restricted to vertebrates (7). It is involved in BER and Back-up non-homologous end joining (BNHEJ). DNA ligase 4 (hLig4) is restricted to NHEJ pathway and tightly bound to XRCC4 protein which stimulates its re-adenylation (12) and protects it from degradation because hLig4 is highly unstable alone (13).

Because DNA ligase is involved at the last step of DNA repair pathways linking broken strands (the 5'-phosphate and 3'-hydroxy groups at the nick of duplex DNA), inhibition of its action would greatly affect cell proliferation and would sensitize tumor cells to ionizing radiations as inhibitors of DNA repair increase the efficiency of DNA-damaging anticancer molecules (1) (4). This makes human DNA ligases crucial targets for drug candidates with clinical potential in combination anticancer therapies (14).

Papers have been dealt with the mechanistic and the detailed analysis of DNA ligases structures. This and the inhibition of NAD⁺-dependent DNA ligases will not be

covered here. This minireview underlines the clinical potential of human ATP-dependent DNA ligases as pharmacological targets in anticancer therapies and collect the ligase inhibitors and their associated discovery techniques that have been promising during the two last decades.

Human DNA Ligases as targets in anticancer therapies.

Modulating the DNA damage response (DDR) to increase ionizing radiation-induced killing of tumor cells is a well-known strategy that has been developed and shown efficiency in some cases, especially when a tumor radioresistance is conferred by an up-regulation of DNA repair proteins (15) (4) (1). Other important components of the DDR are the cell-cycle checkpoints whose inhibitors also increase the radiosensitivity of cancer cells (1) but they are not treated here. Obviously the selectivity is always a parameter to consider as we want to limit injuries to surrounding normal tissues but the hyper-proliferative nature of cancer cells with a higher rate of replication and expression of DNA repair enzymes after injuries make them more selectively sensitive to a potential drug acting on a DNA repair protein. This is also enhanced with modern targeted radiotherapy (1). To consider DNA ligases (encoded by the three human LIG genes) as potential pharmacological targets their activities alteration should be correlated with a variation in cell survival to radio and chemotherapy. Several examples from literature emphasize the effect of different human DNA ligases deficiencies in cells and patients radio-sensitization (5) (16) (14). Moreover, human DNA ligases are major components in DNA replication, recombination and repair (7) (17). The extensive research for human ligases inhibitors emerged with three independent and really useful discoveries. These are the expression and purification of recombinant human DNA ligases from

heterologous systems, the increasing number of high-throughput screening methods for the development of large chemical libraries and last but not least the crystallographic structure of the ligases catalytic domains that allowed inhibitors discovery by a more rational screening. Since the early 1990's people were interested in inhibiting the human DNA ligase 1 activity (18) (19). DNA ligase 1 is the major DNA replicative activity in proliferating cells and is involved in different aspects of DNA metabolism (19) (14). In these proliferating cells the DNA ligase 1 has a higher level of expression than in non-dividing or resting cells. Besides, in 2001 Sun D. et al. have shown an up-regulation of this protein compared to normal tissues and determined that it was essential for survival of tumor cells (17). DNA ligase 1 deficiencies seem to cause immuno-deficiencies and cellular hypersensitivity to DNA-damaging agents (5). All these clues led to the conclusion that the inhibition of this ligase may be a promising pharmacological target for fighting against cancer.

The first report of a systematic search for hLig1 activity inhibitors has been fulfilled in 1996 (19). Ghee T. TAN et al. adapted a procedure previously described (20) to a modified protocol. This DNA ligase assay was done with a recombinant hLig1 from a baculovirus expression and in 96-well plates for inhibitor screening. Briefly, a 5'-labeled DNA duplex is incubated with an unlabeled DNA substrate and the recombinant hLig1. After a treatment by alkaline phosphatase enzyme scintillation counting allows to discriminate between 5'-labeled phosphomonoesters and alkaline-phosphatase-resistant diesters. From a library of different pure compounds and plant extracts the authors found several products that were able to inhibit hLig1 activity. Nevertheless no one could be characterized for its inhibition mode, kinetic activity, specificity and cell toxicity.

In 2002, Sun D. et al developed a non-electrophoretic assay method for high-throughput screening of hLig1 activity inhibitors (17) (See Figure 1). In this assay a 5'-biotinylated nicked DNA duplex is submitted to a DNA ligase reaction with a purified hLig1 in 96-well streptavidin-coated plates. The 5'-extremity of the nick is labeled. If no ligation does occur, removal of labeled strand after denaturation leads to no detectable radioactivity. This absence of radioactivity is the base for a positive ligase inhibitor. This protocol allowed them to characterize two potent inhibitors from a 2000 compounds library (so a hit rate of 0.1%). These inhibitors had a dose-dependent inhibition of DNA ligase activity and moderate toxicity against MCF-7 breast carcinoma cells (as determined by MTT method).

In 2006, Xi Chen et al described a fluorescence-based ligation assay for high-throughput screening of chemical libraries against human DNA ligases activity (21). Insisting on the fact that gel-based assays with radioactive labeled substrates were not suitable for a large number of assays, they defined a new protocol, in a 96-well format, based on fluorescence transfer and quenching mechanisms (see Figure 2). A nicked DNA duplex has a fluorophore close to the 3'-extremity of the nick and a quencher molecule close to the 5'-one. When ligation occurs, both components are still too far from each other to promote fluorescence quenching. However, after extensive wash with the same but not modified strand, the single-derivatized strand adopts a hairpin structure leading to the close proximity of both components and a fluorescence quenching. In this plan, an efficient inhibitor can prevent the ligation, thus promoting the release of two individual strands and without possible hairpin formation nor fluorescence quenching. Their contribution by the improvement of ligase inhibitors screening as well as the development of new protocols to

overexpress and purify recombinant DNA ligase 1, 3beta and 4 (in complex with XRCC4 for stability (12)) gave a further advance in this quest.

From, 2004 an important work revealed the first atomic resolution structure of the human DNA ligase 1 (residues 233 to 919) in complex with a nicked adenylated DNA molecule by X-ray crystallography (22). This advance allowed to better understand the unique feature of mammalian ligases: the presence of a DNA binding domain (DBD; residues 262 to 535) that encircle and unwind DNA positioning the nick in close contact with the enzyme catalytic core (Figure 2). A major particularity of the ligase is its switch between an open-extended conformation onto a compact one right after DNA binding. Apart from a better understanding of the structural machinery of ligases this important discovery opened now the door to the rationalized in silico screening of chemical libraries based on structure.

This rational design of ligase inhibitors gave rise to a first report in 2008 by Xi Chen et al. (23) In this article, the authors have chosen to target the DBD as it is the predominant DNA-binding activity within hlig1 and the two other domains (OBD and Add) could undergo conformational changes during enzymatic reaction (22). The targeted binding pocket was defined as Gly448-Arg451-Ala455 and was subjected to a virtual screening on more than 1.5 million molecules. Selected putative inhibitors were assayed by an in vitro fluorescence-based joining assay previously described (21). Binding assays further confirmed the inhibitory effect of three compounds with different mechanisms of action. IC50 were between 4 and 12uM. L189 was characterized as a step-2-dependent competitive inhibitor with respect to the nicked DNA substrate. L82 was a step-3-dependent uncompetitive inhibitor of hLig1 and L67 was a step-3-dependent competitive inhibitor. As DBD is a conserved structure

among human DNA ligases, non-desired binding was however also observed for Lig3 and Lig4. Despite this lack of hLig1 specificity, the inhibitory effect in the micromolar range was well detectable in vitro and in addition, these compounds inhibited cell proliferation and potentiated the killing of human cancer cells by DNA-damaging agents (DNA4-methylating agent, methyl methanesulfonate, ionizing radiation) but not in normal cells. This observed specificity favors the use of DNA ligases inhibitors in a therapeutic strategy to potentiate cancer cells cytotoxicity of DNA damaging agents (14).

During the same year, authors still decided to target the DBD domain by virtual screening of database compounds (24) but defining the putative binding pocket as: His337-Arg449-Gly453. Compared to their previous work, the docking step of around 50.000 molecules was done on a set of several hLig1 structures generated by a 5ns molecular dynamic (MD) simulation to account for protein flexibility. MD simulations take today an important place in drug discovery as it can for example identify potentially druggable sites whose X-ray or NMR structures can obscure. These binding sites are often called “cryptic binding sites”. Docking of libraries into multiple MD-extracted protein conformations also gives a higher spectrum of scores to allow a better ranking of the hits and increases to chance to find positive hits by the sampling of different stable target conformations (25). After clustering, selected putative inhibitors were assayed by an in vitro fluorescence-based joining assay previously described (21) and further confirmed by radioactive-gel-based DNA ligation assays with purified recombinant enzymes. This in silico screening yielded at least ten active compounds over 233 selected (a hit rate of about 4-5%) against hLig1 activity and specific for the DBD as they did not alter T4 DNA ligase activity (isolated from bacteriophage T4) which does not have this domain. This study among others

underlines the strong advantage for drug screening to have an available protein target structure and also that implication of MD simulations into the screening step can be of great help to discover positive hits. The IC₅₀ values of the 10 hLig1 identified inhibitors ranged from 0.6 to 25uM. Electrophoretic mobility shift assays clearly show that four inhibitors from the screening competitively inhibited DNA binding to hLig1. Moreover several of the active compounds also inhibited hLig3 and/or hLig4. Finally, some of them were able to inhibit cell proliferation either in normal or cancer cells.

Among the DNA repair pathways to disturb, NHEJ is probably the most interesting one and the pathway to prioritize. It's indeed the major pathway in DNA double strand breaks repair whose effects are key lesions of the lethal effect of ionizing radiations and radiomimetic drugs (26). This pathway is involved in the sealing of two DNA ends from a double strand break and without the need for sequence homology between them (27). Compared to homologous recombination, NHEJ is active in all cell cycle stages (27) (28) (15). It has also been shown than NHEJ provoked cancer cells resistance (15). It has already been shown that NHEJ inhibition could impair cancer cells growth and enhance the sensitivity to radiations (29). Besides, DNA-PK complex, a major component of the NHEJ pathway was studied and reviewed as a potential pharmacological target in cancer chemo- and radio-therapy (30).The unique ligase associated with this pathway is the Ligase 4. As this ligase does not play any other role in other pathways, its targeting becomes more specific because centered on the up-regulated NHEJ pathway centered on DNA double strand breaks. Some LIG4 syndrome phenotypes are characterized by a severe radio-sensitivity (5) (14). Moreover, ectopic expression of hLig4 deletion mutants competes with normal full-length hLig4 for its XRCC4 binding site leading to LigIV down-regulation and a

defective NHEJ activity in cell extract as well as radio-sensitized cell clones (31). All these arguments make the ligase 4 another promising target in combined anticancer therapy.

In 2012, Hui-Min Tseng et al designed an in vitro enzymatic assay measuring AMP transfer by hLig4 to a double strand DNA substrate and adapted it to a high throughput scintillation proximity-based assay for the screening of large chemical libraries (26) (see Figure 3). This new screening method has been already successfully used by the same group for a CDC7 kinase screening project (32). Clearly, this assay based on the loss of radioactivity from the labeled AMP-ligase 4 covalent complex was designed to target the AMP transfer step from the adenylation domain. Their screening was done with a purified XRCC4/Lig4 complex from insect sf9 cells. XRCC4/hLig4 was further treated by an adenylation reaction to have an isotope-labeled complex. The assay workflow consists in the pre-incubation of labeled-adenylated-His-tagged-Lig4 with a test compound followed by the incubation of a double strand DNA. If a deadenylation reaction occurs, the added His tag PS imaging beads in the mixture will be excited by beta particles emission and emit a high measurable signal. If there is an inhibition of the deadenylation process, it will only be a low signal emission from the beads due to proximity effects. This technique allowed the identification of four compounds from a 5280 compounds library with IC₅₀ values ranging from 1 to 30uM and further confirmed by in vitro DNA ligation assays. These compounds were found to inhibit the transfer step of the ligation by hLig4 and the two most active ones (rabeprazole and U73122) did not alter the activity of hLig1 and hLig3 testifying of a good specificity. Finally, these two compounds were able to slow down the DNA repair process in HeLa cells after ionizing radiation treatment.

The last attempt in the search for hLig4 inhibition was centered on its DBD and was also based on a rational study of ligase structures (33). It was the first time that the ligase inhibitor discovery was driven from the docking step to the animal model of human cancer. A homology-based 3D model of ligase 4 (1-241 aa) was built based on the DBD structure of other ligases knowing that they share large sequence similarities in the catalytic domain (22) (7) (5) (14) and that the DBD of hLig4 was not crystallized yet. A docking step of a DNA-containing DSB was done in the putative DBD of hLig4 defining critical residues that could contact the DNA molecule. Based on this binding pocket and on the previous identified non-specific inhibitor L189 (23), in silico modifications of this ligand was carried out to produce a new highly specific and energetically-favorable inhibitor called SCR7. A very complete study was done to characterize the inhibitory activity of this molecule. Conclusions were that SCR7 specifically interferes with hLig4 activity within NHEJ pathway from 10uM and in a concentration-dependent manner, binding its DBD. NHEJ inhibition was monitored with cell-free extracts. Within cells, SCR7 induces an increasing amount of unpaired DSBs and slows down tumor progression in different animal models of cancer. This DSB accumulation led to a p53-mediated apoptosis mechanism. Importantly, authors proved that their drug could be considered in a combination anticancer therapy by demonstrating the sensitivity enhancement of cancer cells to radio- and chemotherapeutic agents upon treatment with SCR7.

Conclusion and Future directions.

It is now well accepted that human DNA ligases targeting is a strategic weapon in cancer cells radiosensitization (14). Through several examples people have shown the feasibility of inhibiting the ligases activity, interfering with normal DNA repair

pathways and sensitizing cancer cells to DNA damaging treatments. High-Throughput Screening techniques have been developed the two last decades facilitating the inhibitor discovery process. Nevertheless, this screening really assumed full meaning with structural information given by the first crystal structure of hLig1 protein. With this structure, the introduction of virtual screening, 3D structure modeling and molecular dynamic simulations particularly introduced the rational part to help in designing putative inhibitor molecules. However, a large problem remains and is focused on the specific action of these ligase inhibitors. We saw through examples that a lack of specificity from inhibitor targeting the catalytic function of ligase can lead to the binding to other human ligases due to the conserved DBD domain of ligases (23) and hence, to the alteration of another metabolic pathway. To overcome this lack of specificity, a combination of structure-based and fragment-based drug design with help of powerful biophysical screening techniques and structural informations should be considered. Virtual and High-Throughput Screening techniques are key steps in filtering huge molecule databases (34) (35). But in the quest for a highly-specific inhibitor, a deeper view on the mode of binding, the binding location on protein's key residues as well as the binding energetics (binding entropy and enthalpy) and the affinity is necessary. In the strategy to optimize promising molecules that have been already found one should consider for example protein-inhibitor co-crystallization assays or the expression of labeled DBD of ligase to chemical-shift-map the exact binding location of an interacting drug in NMR experiments. NMR or crystal structure of enzyme-inhibitor complexes, In addition with Isothermal Titration Calorimetry (ITC) data, would be a start in the optimization process to get highly potent and specific inhibitors. Recently, the atomic structure resolution of the catalytic domain of hLig4 (36) has been solved and will help a lot in

this promising adventure. With this last structure, we now have the three DBDs structures of human ligases in detail.

In this review, we reported several author studies on the catalytic domain inhibition of human ligases. As this catalytic region is concerned, an efficient decrease in activity is easy to monitor but this catalytic activity mechanism and the domain structure are conserved among ligases. Then, it would be interesting to consider the challenging and more specific protein-protein interactions between ligases and their bound partner (as for XRCC4/hLig4 or XRCC1/hLig3 interfaces). In accordance with this, an inhibition would aim at directly preventing or disturbing enough the specific association between ligases and their regulatory protein partners to alter the ligase activity. This type of protein-protein interface targeting was studied by virtual screening with ERCC1 and XPF, two major components of the NER pathway and led to compounds which increased cytotoxicity after ionizing radiations (37). As this protein-protein association is very tight and specific, a synthetic peptide approach coupled with a HTS technique would be one strategy among others to face this problem (38). Going to this way, it would be interesting to take benefit from the C-terminal part of hLig4 than have been co-crystallized in interaction with its XRCC4 partner and whose structure was resolved (31). This structure shows two main XRCC4- interacting domains (called the “XIR” and “BRCT2” domains) that could be potential targets. In this paper, authors show the critical effect of disrupting this interaction by dominant negative experiments. Indeed, competition assays with the ectopic expression of truncated hLig4 constructs led to endogenous hLig4 degradation and a subsequent tumor cells radio-sensitization. The XIR domain is of first importance as it makes the majority of contacts between the ligase and XRCC4. Recently, Mc Fadden et al, got a more precise insight onto the key interfaces in the

XIR domain of XRCC4/hLig4 interaction and delimited the minimal interacting region to disrupt in a targeted therapy against NHEJ pathway. This region is a single helix that patches XRCC4 coiled-coil structure (39). The search for indirect hLig4 inhibition was also done targeting XRCC4 by a virtual screening process and led to several in silico potential hits (Tung-Ti Chang 2011). These putative hits were submitted to molecular dynamic simulations analysis of trajectory and persistent contacts and were found stables in the target. However, these hits still need to be tested for specific binding and inhibition properties.

In the strategy to indirectly interfere with regulators of hLig4 activity, a last promising interface remains. XRCC4 and XLF proteins are important hLig4 partners in the ligation reaction within NHEJ pathway (12) (40). These proteins make a critical head-to-head interaction (41). This interaction is weaker (K_d around 4 to 7 micromolar as determined by Biacore and ITC methods) than the XRCC4/hLig4 association or the respective homodimers as a single punctual mutation is enough to dissociate these proteins (42) (43). Despite this weaker association, this complex is not less essential as preventing its formation leads to NHEJ deficits and increases cells radiosensitivity (44). As it was reviewed in 2013 (45), this head-to head interaction forms filaments of variable lengths with alternating XRCC4 and XLF homodimers, leading to left-handed super-helical filaments with predicted flexible and tube-like structures (42) (46). There is no direct evidence for the existence of these filaments *in vivo* but it is suggested that its structural role as a huge molecular scaffold helps for DNA ends bridging, alignment, protection and synapsis and hence, could enhance the end-joining activity of hLig4. Moreover, previous studies have shown these proteins as necessary to promote hLig4 re-adenylation and obviously without this re-adenylation process, hLig4 is unable to enter a new cycle of ligation.

These arguments make these proteins other clue targets in the indirect hlig4 activity disturbance. Again, a rational drug design can be centered on this interface as the structure of XRCC4/XLF complex was resolved and critical interacting residues defined (44). These critical “hotspot” residues were further confirmed by mutational analysis (46) (43) and should be considered first in a rational targeting.

Finally, it has been shown that hLig3 had a role in back-up NHEJ pathway. Usually, this pathway takes place when the normal NHEJ pathway is compromised (5) (47). In some cases, cancer cells (as in ER-/PR- breast cancer cell lines) show an up-regulation of this alternative pathway with higher levels of Ligase III alpha and PARP1 proteins while the predominant DNA-PK-dependent NHEJ pathway is down-regulated (48). This study demonstrated an hypersensitivity of the cells after a combined treatment with PARP and DNA ligase III inhibitors and then, that this pathway could be a novel therapeutic target in breast cancers. In other cases, It would be another track to specifically target this ligase in a combined therapy as non-operant rescue pathways would make the cells again more sensitive to radiotherapy. Again, the atomic structure of the catalytic domain of hLig3 has been solved (49) and can be exploited for a rational drug design strategy.

FIGURE 1

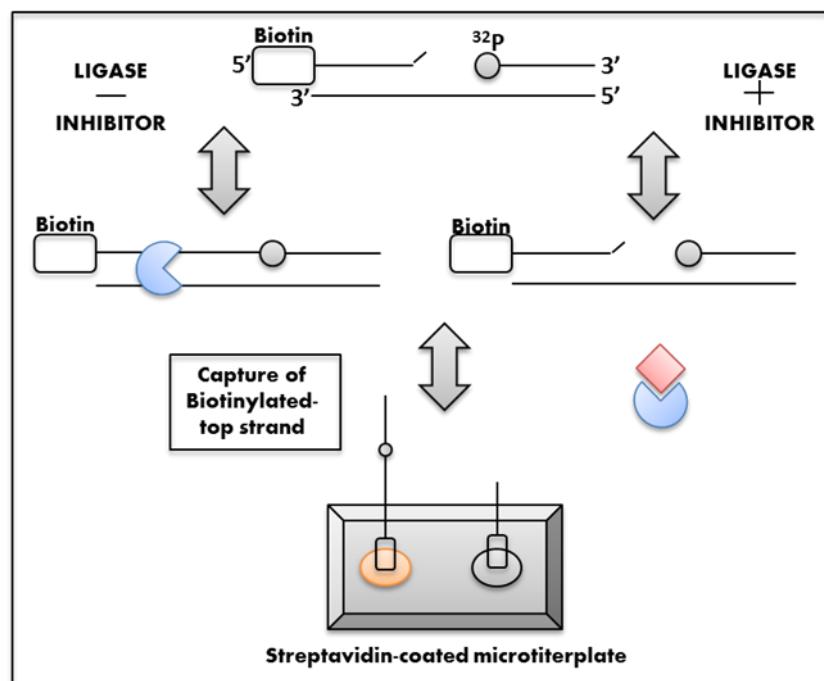


FIGURE 2

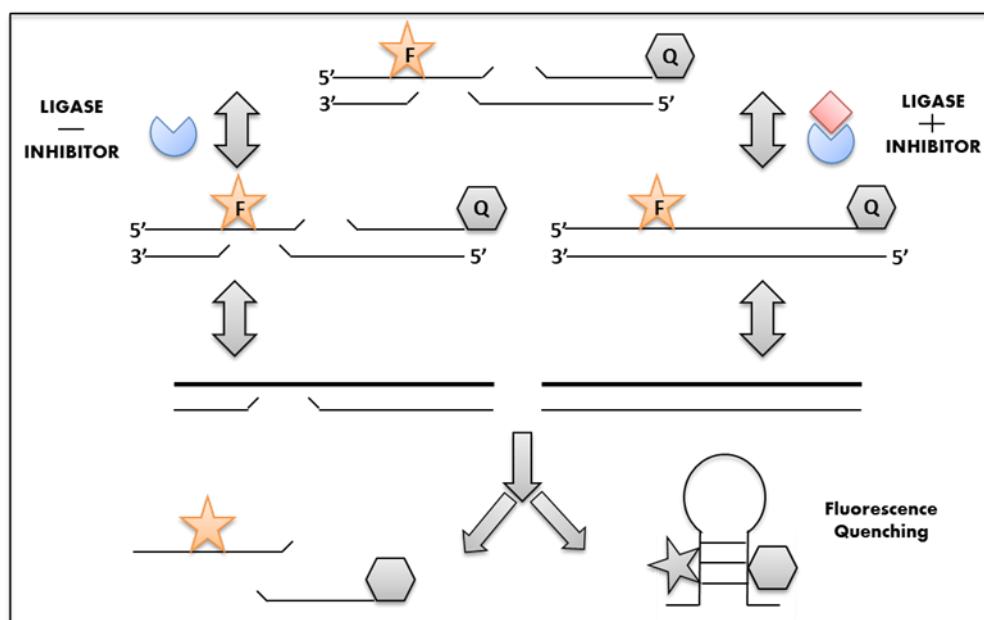
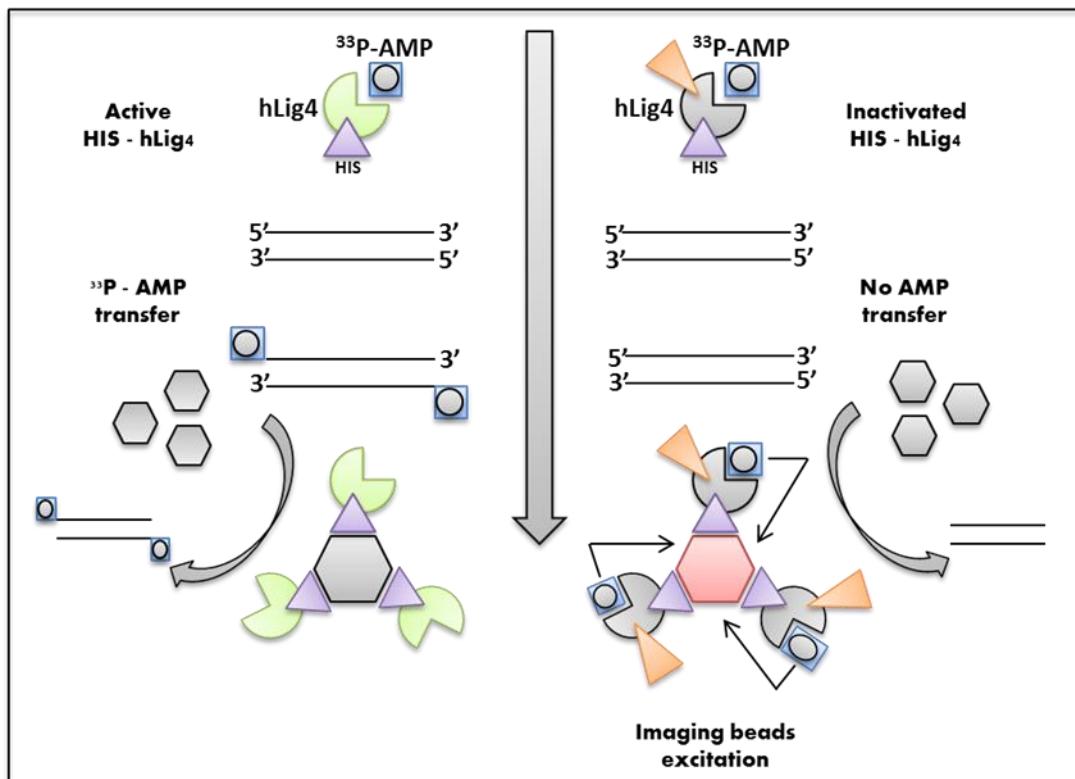


FIGURE 3



REFERENCES

1. Begg AC, Stewart FA, & Vens C (2011) Strategies to improve radiotherapy with targeted drugs. *Nature Reviews Cancer* 11(4):239-253.
2. Azzam EI, Jay-Gerin J-P, & Pain D (2012) Ionizing radiation-induced metabolic oxidative stress and prolonged cell injury. *Cancer Letters* 327(1-2):48-60.
3. Lieber MR (1998) Pathological and Physiological Double-Strand Breaks Roles in Cancer, Aging, and the Immune System.
4. Srinivasan A & Gold B (2012) Small-molecule inhibitors of DNA damage-repair pathways: an approach to overcome tumor resistance to alkylating anticancer drugs. *Future Medicinal Chemistry* 4(9):1093-1111.
5. Ellenberger T & Tomkinson AE (2008) Eukaryotic DNA Ligases: Structural and Functional Insights. *Annual Review of Biochemistry* 77(1):313-338.
6. DAVID S. LEVIN* WB, NINA YAO†‡, MICHAEL O'DONNELL,‡ AND ALAN E. TOMKINSON*§ (1997) An interaction between DNA ligase I and proliferating cell nuclear antigen: Implications for Okazaki fragment synthesis and joining.
7. Shuman S (2009) DNA Ligases: Progress and Prospects. *Journal of Biological Chemistry* 284(26):17365-17369.
8. Hoeijmakers JHJ (2001) Genome maintenance mechanisms for preventing cancer.
9. Sarah C. Shuck, Emily A. Short, & Turchi JJ (2008) Eukaryotic nucleotide excision repair, from understanding mechanisms to influencing biology.

10. Li G-M (2008) Mechanisms and functions of DNA mismatch repair. *Cell Research* 18(1):85-98.
11. Liang L, et al. (2008) Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. *Nucleic Acids Research* 36(10):3297-3310.
12. Mauro Modesti, Joanne E. Hesse, & Gellert M (1999) DNA binding of Xrcc4 protein is associated with V(D)J recombination but not with stimulation of DNA ligase IV activity.
13. Foster RE, Nnakwe C, Woo L, & Frank KM (2006) Monoubiquitination of the nonhomologous end joining protein XRCC4. *Biochemical and Biophysical Research Communications* 341(1):175-183.
14. Alan E. Tomkinson, Timothy R.L. Howes, & Wiest NE (2013) DNA ligases as therapeutic targets.
15. Helleday T, Petermann E, Lundin C, Hodgson B, & Sharma RA (2008) DNA repair pathways as targets for cancer therapy. *Nature Reviews Cancer* 8(3):193-204.
16. Wu PY, et al. (2007) Interplay between Cernunnos-XLF and Nonhomologous End-joining Proteins at DNA Ends in the Cell. *Journal of Biological Chemistry* 282(44):31937-31943.
17. Daekyu Sun & Urrabaz R (2002) Development of non-electrophoretic assay method for DNA ligases and its application to screening of chemical inhibitors of DNA ligase I.

18. Shu-Wei Yangz, Peng Huangj, William Plunkettj, FrederickF . Beckerz, & ChanST JH (1992) Dual Mode of Inhibition of Purified DNA Ligase I from Human Cells by 9-&~-Arabinofuranosyl-2-fluoroadenine Triphosphate*.
19. Ghee T. TAN, et al. (1996) Natural-product inhibitors of human DNA ligase I.
20. Alan E. Tomkinson, Emma Roberts, Graham Daly, Nicholas F. Tottyl, & Lindahl T (1991) Three Distinct DNA Ligases in Mammalian Cells.
21. Chen X, et al. (2006) Human DNA Ligases I, III, and IV—Purification and New Specific Assays for These Enzymes. 409:39-52.
22. John M. Pascal, Patrick J. O'Brien, Alan E. Tomkinson, & Ellenberger T (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA.
23. Chen X, et al. (2008) Rational Design of Human DNA Ligase Inhibitors that Target Cellular DNA Replication and Repair. Cancer Research 68(9):3169-3177.
24. Zhong S, et al. (2008) Identification and Validation of Human DNA Ligase Inhibitors Using Computer-Aided Drug Design. Journal of Medicinal Chemistry 51(15):4553-4562.
25. Durrant JD & McCammon aJA (2011) Molecular dynamics simulations and drug discovery.
26. Tseng H-M, et al. (2012) A High-Throughput Scintillation Proximity-Based Assay for Human DNA Ligase IV. ASSAY and Drug Development Technologies 10(3):235-249.
27. Lieber MR, Gu J, Lu H, Shimazaki N, & Tsai AG (2010) Nonhomologous DNA End Joining (NHEJ) and Chromosomal Translocations in Humans. 50:279-296.

28. Lieber MR (2010) The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review of Biochemistry* 79(1):181-211.
29. Schneider G, et al. (2012) Inhibition of Non-Homologous End Joining Repair Impairs Pancreatic Cancer Growth and Enhances Radiation Response. *PLoS ONE* 7(6):e39588.
30. Salles B, Calsou P, Frit P, & Muller C (2006) The DNA repair complex DNA-PK, a pharmacological target in cancer chemotherapy and radiotherapy. *Pathologie Biologie* 54(4):185-193.
31. Wu PY, et al. (2009) Structural and Functional Interaction between the Human DNA Repair Proteins DNA Ligase IV and XRCC4. *Molecular and Cellular Biology* 29(11):3163-3172.
32. Toshimitsu Takagi, et al. (2011) Comparison of Luminescence ADP Production Assay and Radiometric Scintillation Proximity Assay for Cdc7 Kinase.
33. Srivastava M, et al. (2012) An Inhibitor of Nonhomologous End-Joining Abrogates Double-Strand Break Repair and Impedes Cancer Progression. *Cell* 151(7):1474-1487.
34. Bleicher KH, Böhm H-J, Müller K, & Alanine AI (2003) A guide to drug discovery: Hit and lead generation: beyond high-throughput screening. *Nature Reviews Drug Discovery* 2(5):369-378.
35. Vyas V (2008) Virtual Screening: A Fast Tool for Drug Design. *Scientia Pharmaceutica* 76(3):333-360.

36. Ochi T, Gu X, & Blundell Tom L (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.
37. Jordheim LP, et al. (2013) Small Molecule Inhibitors of ERCC1-XPF Protein-Protein Interaction Synergize Alkylating Agents in Cancer Cells. *Molecular Pharmacology* 84(1):12-24.
38. Wu B, et al. (2013) HTS by NMR of Combinatorial Libraries: A Fragment-Based Approach to Ligand Discovery. *Chemistry & Biology* 20(1):19-33.
39. McFadden MJ, Lee WKY, Brennan JD, & Junop MS (2013) Delineation of key XRCC4/LigaseIV interfaces for targeted disruption of non-homologous end joining DNA repair. *Proteins: Structure, Function, and Bioinformatics*:n/a-n/a.
40. Yano K-i, Morotomi-Yano K, & Akiyama H (2009) Cernunnos/XLF: A new player in DNA double-strand break repair. *The International Journal of Biochemistry & Cell Biology* 41(6):1237-1240.
41. 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.
42. Andres SN, et al. (2012) A human XRCC4-XLF complex bridges DNA. *Nucleic Acids Research* 40(4):1868-1878.
43. Ropars V & al. e (2011) Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining.

44. Malivert L, et al. (2010) Delineation of the Xrcc4-interacting Region in the Globular Head Domain of Cernunnos/XLF. *Journal of Biological Chemistry* 285(34):26475-26483.
45. Mahaney BL, Hammel M, Meek K, Tainer JA, & Lees-Miller SP (2013) XRCC4 and XLF form long helical protein filaments suitable for DNA end protection and alignment to facilitate DNA double strand break repair1. *Biochemistry and Cell Biology* 91(1):31-41.
46. Hammel M, et al. (2011) XRCC4 Protein Interactions with XRCC4-like Factor (XLF) Create an Extended Grooved Scaffold for DNA Ligation and Double Strand Break Repair. *Journal of Biological Chemistry* 286(37):32638-32650.
47. Huichen Wang, Bustanur Rosidi, & Ronel Perrault ea (2005) DNA Ligase III as a Candidate Component of Backup Pathways of Nonhomologous End Joining.
48. Tobin LA, et al. (2011) Targeting Abnormal DNA Repair in Therapy-Resistant Breast Cancers. *Molecular Cancer Research* 10(1):96-107.
49. Cotner-Gohara E, et al. (2010) Human DNA Ligase III Recognizes DNA Ends by Dynamic Switching between Two DNA-Bound States. *Biochemistry* 49(29):6165-6176.

Bibliographie

A

- Ader, I., C. Muller, J. Bonnet, G. Favre, E. Cohen-Jonathan, B. Salles, and C. Toulas. 2002. The radioprotective effect of the 24 kDa FGF-2 isoform in HeLa cells is related to an increased expression and activity of the DNA dependent protein kinase (DNA-PK) catalytic subunit. *Oncogene*. 21:6471-6479.
- Ahnesorg, P., P. Smith, and S.P. Jackson. 2006. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell*. 124:301-313.
- Akopiants, K., R.Z. Zhou, S. Mohapatra, K. Valerie, S.P. Lees-Miller, K.J. Lee, D.J. Chen, P. Revy, J.P. de Villartay, and L.F. Povirk. 2009. Requirement for XLF/Cernunnos in alignment-based gap filling by DNA polymerases lambda and mu for nonhomologous end joining in human whole-cell extracts. *Nucleic Acids Res*. 37:4055-4062.
- Allen, C., J. Halbrook, and J.A. Nickoloff. 2003. Interactive competition between homologous recombination and non-homologous end joining. *Mol Cancer Res*. 1:913-920.
- Andres, S.N., M. Modesti, C.J. Tsai, G. Chu, and M.S. Junop. 2007. Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. *Mol Cell*. 28:1093-1101.
- Andres, S.N., M.J. Schellenberg, B.D. Wallace, P. Tumbale, and R.S. Williams. 2015. Recognition and repair of chemically heterogeneous structures at DNA ends. *Environ Mol Mutagen*. 56:1-21.
- Andres, S.N., A. Vergnes, D. Ristic, C. Wyman, M. Modesti, and M. Junop. 2012. A human XRCC4-XLF complex bridges DNA. *Nucleic Acids Res*. 40:1868-1878.
- Archakov, A.I., V.M. Govorun, A.V. Dubanov, Y.D. Ivanov, A.V. Veselovsky, P. Lewi, and P. Janssen. 2003. Protein-protein interactions as a target for drugs in proteomics. *Proteomics*. 3:380-391.
- Arkin, M.R., M. Randal, W.L. DeLano, J. Hyde, T.N. Luong, J.D. Oslob, D.R. Raphael, L. Taylor, J. Wang, R.S. McDowell, J.A. Wells, and A.C. Braisted. 2003. Binding of small molecules to an adaptive protein-protein interface. *Proc Natl Acad Sci U S A*. 100:1603-1608.

B

- Bahadur, R.P., and M. Zacharias. 2008. The interface of protein-protein complexes: analysis of contacts and prediction of interactions. *Cell Mol Life Sci*. 65:1059-1072.
- Barakat, K.H., L.P. Jordheim, R. Perez-Pineiro, D. Wishart, C. Dumontet, and J.A. Tuszyński. 2012. Virtual screening and biological evaluation of inhibitors targeting the XPA-ERCC1 interaction. *PLoS One*. 7:e51329.
- Barnett, G.C., C.M. West, A.M. Dunning, R.M. Elliott, C.E. Coles, P.D. Pharoah, and N.G. Burnet. 2009. Normal tissue reactions to radiotherapy: towards tailoring treatment dose by genotype. *Nat Rev Cancer*. 9:134-142.
- Bartkova, J., Z. Horejsi, K. Koed, A. Kramer, F. Tort, K. Zieger, P. Guldberg, M. Sehested, J.M. Nesland, C. Lukas, T. Orntoft, J. Lukas, and J. Bartek. 2005. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature*. 434:864-870.

- Begg, A.C., F.A. Stewart, and C. Vens. 2011. Strategies to improve radiotherapy with targeted drugs. *Nat Rev Cancer.* 11:239-253.
- Berg, E., M.O. Christensen, I. Dalla Rosa, E. Wannagat, R.U. Janicke, L.M. Rosner, W.G. Dirks, F. Boege, and C. Mielke. 2011. XRCC4 controls nuclear import and distribution of Ligase IV and exchanges faster at damaged DNA in complex with Ligase IV. *DNA Repair (Amst).* 10:1232-1242.
- Beskow, C., J. Skikuniene, A. Holgersson, B. Nilsson, R. Lewensohn, L. Kanter, and K. Viktorsson. 2009. Radioresistant cervical cancer shows upregulation of the NHEJ proteins DNA-PKcs, Ku70 and Ku86. *Br J Cancer.* 101:816-821.
- Bettermier, M., P. Bertrand, and B.S. Lopez. 2014. Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet.* 10:e1004086.
- Blocher, D., and W. Pohlit. 1982. DNA double strand breaks in Ehrlich ascites tumour cells at low doses of x-rays. II. Can cell death be attributed to double strand breaks? *Int J Radiat Biol Relat Stud Phys Chem Med.* 42:329-338.
- Bork, P., K. Hofmann, P. Bucher, A.F. Neuwald, S.F. Altschul, and E.V. Koonin. 1997. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J.* 11:68-76.
- Bryans, M., M.C. Valenzano, and T.D. Stamato. 1999. Absence of DNA ligase IV protein in XR-1 cells: evidence for stabilization by XRCC4. *Mutat Res.* 433:53-58.
- Buck, D., L. Malivert, R. de Chasseval, A. Barraud, M.C. Fondaneche, O. Sanal, A. Plebani, J.L. Stephan, M. Hufnagel, F. le Deist, A. Fischer, A. Durandy, J.P. de Villartay, and P. Revy. 2006. Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell.* 124:287-299.

C

- Cala, O., F. Guilliere, and I. Krimm. 2014. NMR-based analysis of protein-ligand interactions. *Anal Bioanal Chem.* 406:943-956.
- Chen, B.P., D.W. Chan, J. Kobayashi, S. Burma, A. Asaithamby, K. Morotomi-Yano, E. Botvinick, J. Qin, and D.J. Chen. 2005. Cell cycle dependence of DNA-dependent protein kinase phosphorylation in response to DNA double strand breaks. *J Biol Chem.* 280:14709-14715.
- Chen, X., S. Zhong, X. Zhu, B. Dziegielewska, T. Ellenberger, G.M. Wilson, A.D. MacKerell, Jr., and A.E. Tomkinson. 2008. Rational design of human DNA ligase inhibitors that target cellular DNA replication and repair. *Cancer Res.* 68:3169-3177.
- Chene, P. 2006. Drugs targeting protein-protein interactions. *ChemMedChem.* 1:400-411.
- Cheng, Q., N. Barboule, P. Frit, D. Gomez, O. Bombarde, B. Couderc, G.S. Ren, B. Salles, and P. Calsou. 2011. Ku counteracts mobilization of PARP1 and MRN in chromatin damaged with DNA double-strand breaks. *Nucleic Acids Res.* 39:9605-9619.
- Ciccia, A., and S.J. Elledge. 2010. The DNA damage response: making it safe to play with knives. *Mol Cell.* 40:179-204.
- Clackson, T., and J.A. Wells. 1995. A hot spot of binding energy in a hormone-receptor interface. *Science.* 267:383-386.
- Colas, P. 2008. High-throughput screening assays to discover small-molecule inhibitors of protein interactions. *Curr Drug Discov Technol.* 5:190-199.
- Collis, S.J., T.L. DeWeese, P.A. Jeggo, and A.R. Parker. 2005. The life and death of DNA-PK. *Oncogene.* 24:949-961.

- Congreve, M., C.W. Murray, and T.L. Blundell. 2005. Structural biology and drug discovery. *Drug Discov Today*. 10:895-907.
- Connell, P.P., K. Jayathilaka, D.J. Haraf, R.R. Weichselbaum, E.E. Vokes, and M.W. Lingen. 2006. Pilot study examining tumor expression of RAD51 and clinical outcomes in human head cancers. *Int J Oncol*. 28:1113-1119.
- Connell, P.P., S.J. Kron, and R.R. Weichselbaum. 2004. Relevance and irrelevance of DNA damage response to radiotherapy. *DNA Repair (Amst)*. 3:1245-1251.
- Costantini, S., L. Woodbine, L. Andreoli, P.A. Jeggo, and A. Vindigni. 2007. Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair (Amst)*. 6:712-722.
- Cottarel, J., P. Frit, O. Bombarde, B. Salles, A. Negrel, S. Bernard, P.A. Jeggo, M.R. Lieber, M. Modesti, and P. Calsou. 2013. A noncatalytic function of the ligation complex during nonhomologous end joining. *J Cell Biol*. 200:173-186.
- Cox, D.B., R.J. Platt, and F. Zhang. 2015. Therapeutic genome editing: prospects and challenges. *Nat Med*. 21:121-131.
- Cozzini, P., G.E. Kellogg, F. Spyarakis, D.J. Abraham, G. Costantino, A. Emerson, F. Fanelli, H. Gohlke, L.A. Kuhn, G.M. Morris, M. Orozco, T.A. Pertinhez, M. Rizzi, and C.A. Sottriffer. 2008. Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem*. 51:6237-6255.
- Critchlow, S.E., R.P. Bowater, and S.P. Jackson. 1997. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr Biol*. 7:588-598.

D

- Damm-Ganamet, K.L., R.D. Smith, J.B. Dunbar, Jr., J.A. Stuckey, and H.A. Carlson. 2013. CSAR benchmark exercise 2011-2012: evaluation of results from docking and relative ranking of blinded congeneric series. *J Chem Inf Model*. 53:1853-1870.
- Davidson, D., L. Amrein, L. Panasci, and R. Aloyz. 2013. Small Molecules, Inhibitors of DNA-PK, Targeting DNA Repair, and Beyond. *Front Pharmacol*. 4:5.
- DeFazio, L.G., R.M. Stansel, J.D. Griffith, and G. Chu. 2002. Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J*. 21:3192-3200.
- Delaney, G., S. Jacob, C. Featherstone, and M. Barton. 2005. The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines. *Cancer*. 104:1129-1137.
- Delarue, M., J.B. Boule, J. Lescar, N. Expert-Bezancon, N. Jourdan, N. Sukumar, F. Rougeon, and C. Papanicolaou. 2002. Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. *EMBO J*. 21:427-439.
- Deriano, L., and D.B. Roth. 2013. Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu Rev Genet*. 47:433-455.
- Deshpande, R.A., and T.E. Wilson. 2007. Modes of interaction among yeast Nej1, Lif1 and Dnl4 proteins and comparison to human XLF, XRCC4 and Lig4. *DNA Repair (Amst)*. 6:1507-1516.
- Dias, D.M., and A. Ciulli. 2014. NMR approaches in structure-based lead discovery: recent developments and new frontiers for targeting multi-protein complexes. *Prog Biophys Mol Biol*. 116:101-112.
- Dietlein, F., L. Thelen, and H.C. Reinhardt. 2014. Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches. *Trends Genet*. 30:326-339.

- Durante, M., J.S. Bedford, D.J. Chen, S. Conrad, M.N. Cornforth, A.T. Natarajan, D.C. van Gent, and G. Obe. 2013. From DNA damage to chromosome aberrations: joining the break. *Mutat Res.* 756:5-13.
- Durrant, J.D., and J.A. McCammon. 2010. Potential drug-like inhibitors of Group 1 influenza neuraminidase identified through computer-aided drug design. *Comput Biol Chem.* 34:97-105.

E

- Ellenberger, T., and A.E. Tomkinson. 2008. Eukaryotic DNA ligases: structural and functional insights. *Annu Rev Biochem.* 77:313-338.
- Eyrisch, S., and V. Helms. 2009. What induces pocket openings on protein surface patches involved in protein-protein interactions? *J Comput Aided Mol Des.* 23:73-86.

F

- Feng, W., S. Huang, H. Wu, and M. Zhang. 2007. Molecular basis of Bcl-xL's target recognition versatility revealed by the structure of Bcl-xL in complex with the BH3 domain of Beclin-1. *J Mol Biol.* 372:223-235.
- Frit, P., N. Barboule, Y. Yuan, D. Gomez, and P. Calsou. 2014. Alternative end-joining pathway(s): bricolage at DNA breaks. *DNA Repair (Amst).* 17:81-97.
- Fry, D.C. 2006. Protein-protein interactions as targets for small molecule drug discovery. *Biopolymers.* 84:535-552.

G

- Gao, Y., Y. Sun, K.M. Frank, P. Dikkes, Y. Fujiwara, K.J. Seidl, J.M. Sekiguchi, G.A. Rathbun, W. Swat, J. Wang, R.T. Bronson, B.A. Malynn, M. Bryans, C. Zhu, J. Chaudhuri, L. Davidson, R. Ferrini, T. Stamato, S.H. Orkin, M.E. Greenberg, and F.W. Alt. 1998. A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell.* 95:891-902.
- Garces, F., L.H. Pearl, and A.W. Oliver. 2011. The structural basis for substrate recognition by mammalian polynucleotide kinase 3' phosphatase. *Mol Cell.* 44:385-396.
- Garcia-Diaz, M., K. Bebenek, J.M. Krahn, L.C. Pedersen, and T.A. Kunkel. 2007. Role of the catalytic metal during polymerization by DNA polymerase lambda. *DNA Repair (Amst).* 6:1333-1340.
- Gavenonis, J., B.A. Sheneman, T.R. Siegert, M.R. Eshelman, and J.A. Kritzer. 2014. Comprehensive analysis of loops at protein-protein interfaces for macrocycle design. *Nat Chem Biol.* 10:716-722.
- Goglia, A.G., R. Delsite, A.N. Luz, D. Shahbazian, A.F. Salem, R.K. Sundaram, J. Chiaravalli, P.J. Hendrikx, J.A. Wilshire, M. Jasin, H.M. Kluger, J.F. Glickman, S.N. Powell, and R.S. Bindra. 2015. Identification of novel radiosensitizers in a high-throughput, cell-based screen for DSB repair inhibitors. *Mol Cancer Ther.* 14:326-342.
- Gonzalez-Ruiz, D., and H. Gohlke. 2006. Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr Med Chem.* 13:2607-2625.

- Goodarzi, A.A., and P.A. Jeggo. 2013. The repair and signaling responses to DNA double-strand breaks. *Adv Genet.* 82:1-45.
- Gorgoulis, V.G., L.V. Vassiliou, P. Karakaidos, P. Zacharatos, A. Kotsinas, T. Liloglou, M. Venere, R.A. Dittullio, Jr., N.G. Kastrinakis, B. Levy, D. Kletsas, A. Yoneta, M. Herlyn, C. Kittas, and T.D. Halazonetis. 2005. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature.* 434:907-913.
- Grawunder, U., M. Wilm, X. Wu, P. Kulesza, T.E. Wilson, M. Mann, and M.R. Lieber. 1997. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature.* 388:492-495.
- Grawunder, U., D. Zimmer, and M.R. Lieber. 1998. DNA ligase IV binds to XRCC4 via a motif located between rather than within its BRCT domains. *Curr Biol.* 8:873-876.
- Gu, J., H. Lu, A.G. Tsai, K. Schwarz, and M.R. Lieber. 2007. Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic Acids Res.* 35:5755-5762.
- Guilloux, V.L., A. Arrault, L. Colliandre, S. Bourg, P. Vayer, and L. Morin-Allory. 2012. Mining collections of compounds with Screening Assistant 2. *J Cheminform.* 4:20.
- Guirouilh-Barbat, J., S. Huck, P. Bertrand, L. Pirzio, C. Desmaze, L. Sabatier, and B.S. Lopez. 2004. Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. *Mol Cell.* 14:611-623.
- Guo, Z., Y. Shu, H. Zhou, W. Zhang, and H. Wang. 2015. Radiogenomics helps to achieve personalized therapy by evaluating patient responses to radiation treatment. *Carcinogenesis.* 36:307-317.

H

- Hammel, M., M. Rey, Y. Yu, R.S. Mani, S. Classen, M. Liu, M.E. Pique, S. Fang, B.L. Mahaney, M. Weinfeld, D.C. Schriemer, S.P. Lees-Miller, and J.A. Tainer. 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.* 286:32638-32650.
- He, M.M., A.S. Smith, J.D. Oslob, W.M. Flanagan, A.C. Braisted, A. Whitty, M.T. Cancilla, J. Wang, A.A. Lugovskoy, J.C. Yoburn, A.D. Fung, G. Farrington, J.K. Eldredge, E.S. Day, L.A. Cruz, T.G. Cachero, S.K. Miller, J.E. Friedman, I.C. Choong, and B.C. Cunningham. 2005. Small-molecule inhibition of TNF-alpha. *Science.* 310:1022-1025.
- Hentges, P., P. Ahnesorg, R.S. Pitcher, C.K. Bruce, B. Kysela, A.J. Green, J. Bianchi, T.E. Wilson, S.P. Jackson, and A.J. Doherty. 2006. Evolutionary and functional conservation of the DNA non-homologous end-joining protein, XLF/Cernunnos. *J Biol Chem.* 281:37517-37526.
- Higuero, A.P., H. Jubb, and T.L. Blundell. 2013. TIMBAL v2: update of a database holding small molecules modulating protein-protein interactions. *Database (Oxford).* 2013:bat039.
- Hill, A.P., and R.J. Young. 2010. Getting physical in drug discovery: a contemporary perspective on solubility and hydrophobicity. *Drug Discov Today.* 15:648-655.
- Hopkins, A.L., and C.R. Groom. 2002. The druggable genome. *Nat Rev Drug Discov.* 1:727-730.
- Hou, T., N. Li, Y. Li, and W. Wang. 2012. Characterization of domain-peptide interaction interface: prediction of SH3 domain-mediated protein-protein interaction network in yeast by generic structure-based models. *J Proteome Res.* 11:2982-2995.

Hsu, F.M., S. Zhang, and B.P. Chen. 2012. Role of DNA-dependent protein kinase catalytic subunit in cancer development and treatment. *Transl Cancer Res.* 1:22-34.

Huang, N., A. Nagarsekar, G. Xia, J. Hayashi, and A.D. MacKerell, Jr. 2004. Identification of non-phosphate-containing small molecular weight inhibitors of the tyrosine kinase p56 Lck SH2 domain via in silico screening against the pY + 3 binding site. *J Med Chem.* 47:3502-3511.

I

Irwin, J.J., and B.K. Shoichet. 2005. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model.* 45:177-182.

J

Jackson, S.P., and J. Bartek. 2009. The DNA-damage response in human biology and disease. *Nature.* 461:1071-1078.

Janin, J., R.P. Bahadur, and P. Chakrabarti. 2008. Protein-protein interaction and quaternary structure. *Q Rev Biophys.* 41:133-180.

Jeggo, P., and M.F. Lavin. 2009. Cellular radiosensitivity: how much better do we understand it? *Int J Radiat Biol.* 85:1061-1081.

Jekimovs, C., E. Bolderson, A. Suraweera, M. Adams, K.J. O'Byrne, and D.J. Richard. 2014. Chemotherapeutic compounds targeting the DNA double-strand break repair pathways: the good, the bad, and the promising. *Front Oncol.* 4:86.

Jette, N., and S.P. Lees-Miller. 2015. The DNA-dependent protein kinase: A multifunctional protein kinase with roles in DNA double strand break repair and mitosis. *Prog Biophys Mol Biol.* 117:194-205.

Jones, K.R., D.A. Gewirtz, S.M. Yannone, S. Zhou, D.G. Schatz, K. Valerie, and L.F. Povirk. 2005. Radiosensitization of MDA-MB-231 breast tumor cells by adenovirus-mediated overexpression of a fragment of the XRCC4 protein. *Mol Cancer Ther.* 4:1541-1547.

Jones, S., and J.M. Thornton. 1996. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A.* 93:13-20.

Jubb, H., A.P. Higuero, A. Winter, and T.L. Blundell. 2012. Structural biology and drug discovery for protein-protein interactions. *Trends Pharmacol Sci.* 33:241-248.

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

K

Kelly, W., and M. Stumpf. 2008. Protein-protein interactions: from global to local analyses. *Curr Opin Biotechnol.* 19:396-403.

Koch, C.A., R. Agyei, S. Galicia, P. Metalnikov, P. O'Donnell, A. Starostine, M. Weinfeld, and D. Durocher. 2004. Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *EMBO J.* 23:3874-3885.

- Koes, D.R., and C.J. Camacho. 2012. PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure. *Nucleic Acids Res.* 40:W387-392.
- Kruszewski, M., M. Wojewodzka, T. Iwanenko, I. Szumiel, and A. Okuyama. 1998. Differential inhibitory effect of OK-1035 on DNA repair in L5178Y murine lymphoma sublines with functional or defective repair of double strand breaks. *Mutat Res.* 409:31-36.
- Kuenemann, M.A., O. Sperandio, C.M. Labbe, D. Lagorce, M.A. Miteva, and B.O. Villoutreix. 2015. In silico design of low molecular weight protein-protein interaction inhibitors: Overall concept and recent advances. *Prog Biophys Mol Biol.*

L

- Labbe, C.M., J. Rey, D. Lagorce, M. Vavrusa, J. Becot, O. Sperandio, B.O. Villoutreix, P. Tuffery, and M.A. Miteva. 2015. MTiOpenScreen: a web server for structure-based virtual screening. *Nucleic Acids Res.* 43:W448-454.
- Lagorce, D., O. Sperandio, J.B. Baell, M.A. Miteva, and B.O. Villoutreix. 2015. FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic Acids Res.* 43:W200-207.
- Larrieu, D., S. Britton, M. Demir, R. Rodriguez, and S.P. Jackson. 2014. Chemical inhibition of NAT10 corrects defects of laminopathic cells. *Science.* 344:527-532.
- Leach, A.R., V.J. Gillet, R.A. Lewis, and R. Taylor. 2010. Three-dimensional pharmacophore methods in drug discovery. *J Med Chem.* 53:539-558.
- Leber, R., T.W. Wise, R. Mizuta, and K. Meek. 1998. The XRCC4 gene product is a target for and interacts with the DNA-dependent protein kinase. *J Biol Chem.* 273:1794-1801.
- Lee, J.W., S.M. Yannone, D.J. Chen, and L.F. Povirk. 2003. Requirement for XRCC4 and DNA ligase IV in alignment-based gap filling for nonhomologous DNA end joining in vitro. *Cancer Res.* 63:22-24.
- Li, G., C. Nelsen, and E.A. Hendrickson. 2002. Ku86 is essential in human somatic cells. *Proc Natl Acad Sci U S A.* 99:832-837.
- Li, Y., D.Y. Chirgadze, V.M. Bolanos-Garcia, B.L. Sibanda, O.R. Davies, P. Ahnesorg, S.P. Jackson, and T.L. Blundell. 2008. Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ. *EMBO J.* 27:290-300.
- Li, Y.H., X. Wang, Y. Pan, D.H. Lee, D. Chowdhury, and A.C. Kimmelman. 2012. Inhibition of non-homologous end joining repair impairs pancreatic cancer growth and enhances radiation response. *PLoS One.* 7:e39588.
- Lieber, M.R. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem.* 79:181-211.
- Lin, J.H., A.L. Perryman, J.R. Schames, and J.A. McCammon. 2003. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers.* 68:47-62.
- Lindahl, T., C. Prigent, D.E. Barnes, A.R. Lehmann, M.S. Satoh, E. Roberts, R.A. Nash, P. Robins, and G. Daly. 1993. DNA joining in mammalian cells. *Cold Spring Harb Symp Quant Biol.* 58:619-624.
- Lipinski, C.A., F. Lombardo, B.W. Dominy, and P.J. Feeney. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 46:3-26.
- Lu, H., U. Pannicke, K. Schwarz, and M.R. Lieber. 2007. Length-dependent binding of human XLF to DNA and stimulation of XRCC4.DNA ligase IV activity. *J Biol Chem.* 282:11155-11162.

Luchini, A., V. Espina, and L.A. Liotta. 2014. Protein painting reveals solvent-excluded drug targets hidden within native protein-protein interfaces. *Nat Commun.* 5:4413.

M

- Macrae, C.J., R.D. McCulloch, J. Ylanko, D. Durocher, and C.A. Koch. 2008. APLF (C2orf13) facilitates nonhomologous end-joining and undergoes ATM-dependent hyperphosphorylation following ionizing radiation. *DNA Repair (Amst).* 7:292-302.
- Mahaney, B.L., K. Meek, and S.P. Lees-Miller. 2009. Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J.* 417:639-650.
- Malivert, L., V. Ropars, M. Nunez, P. Drevet, S. Miron, G. Faure, R. Guerois, J.P. Mornon, P. Revy, J.B. Charbonnier, I. Callebaut, and J.P. de Villartay. 2010. Delineation of the Xrcc4-interacting region in the globular head domain of cernunnos/XLF. *J Biol Chem.* 285:26475-26483.
- Marchetti, C., S.A. Walker, F. Odreman, A. Vindigni, A.J. Doherty, and P. Jeggo. 2006. Identification of a novel motif in DNA ligases exemplified by DNA ligase IV. *DNA Repair (Amst).* 5:788-798.
- Maruyama, T., S.K. Dougan, M.C. Truttmann, A.M. Bilate, J.R. Ingram, and H.L. Ploegh. 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol.* 33:538-542.
- Mayer, M., and B. Meyer. 2001. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J Am Chem Soc.* 123:6108-6117.
- McFadden, M.J., W.K. Lee, J.D. Brennan, and M.S. Junop. 2014. Delineation of key XRCC4/Ligase IV interfaces for targeted disruption of non-homologous end joining DNA repair. *Proteins.* 82:187-194.
- Meek, K., V. Dang, and S.P. Lees-Miller. 2008. DNA-PK: the means to justify the ends? *Adv Immunol.* 99:33-58.
- Modesti, M., J.E. Hesse, and M. Gellert. 1999. DNA binding of Xrcc4 protein is associated with V(D)J recombination but not with stimulation of DNA ligase IV activity. *EMBO J.* 18:2008-2018.
- Modesti, M., M.S. Junop, R. Ghirlando, M. van de Rakt, M. Gellert, W. Yang, and R. Kanaar. 2003. Tetramerization and DNA ligase IV interaction of the DNA double-strand break repair protein XRCC4 are mutually exclusive. *J Mol Biol.* 334:215-228.
- Moon, A.F., J.M. Pryor, D.A. Ramsden, T.A. Kunkel, K. Bebenek, and L.C. Pedersen. 2014. Sustained active site rigidity during synthesis by human DNA polymerase mu. *Nat Struct Mol Biol.* 21:253-260.
- Morelli, X., R. Bourgeas, and P. Roche. 2011. Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr Opin Chem Biol.* 15:475-481.
- Morris, A.L., M.W. MacArthur, E.G. Hutchinson, and J.M. Thornton. 1992. Stereochemical quality of protein structure coordinates. *Proteins.* 12:345-364.
- Mousseau, G., Q. Raffy, O.P. Thomas, M. Agez, R. Thai, J.P. Renault, S. Pin, F. Ochsenbein, J.C. Cintrat, and B. Rousseau. 2010. Footprinting of protein interactions by tritium labeling. *Biochemistry.* 49:4297-4299.
- Murray, J.E., M. van der Burg, I.J. H, P. Carroll, Q. Wu, T. Ochi, A. Leitch, E.S. Miller, B. Kysela, A. Jawad, A. Bottani, F. Brancati, M. Cappa, V. Cormier-Daire, C. Deshpande, E.A. Fafeih, G.E. Graham, E. Ranza, T.L. Blundell, A.P. Jackson, G.S. Stewart, and L.S. Bicknell.

2015. Mutations in the NHEJ component XRCC4 cause primordial dwarfism. *Am J Hum Genet.* 96:412-424.

N

- Neduva, V., and R.B. Russell. 2006. Peptides mediating interaction networks: new leads at last. *Curr Opin Biotechnol.* 17:465-471.
- Nero, T.L., C.J. Morton, J.K. Holien, J. Wielens, and M.W. Parker. 2014. Oncogenic protein interfaces: small molecules, big challenges. *Nat Rev Cancer.* 14:248-262.
- Neugebauer, A., R.W. Hartmann, and C.D. Klein. 2007. Prediction of protein-protein interaction inhibitors by chemoinformatics and machine learning methods. *J Med Chem.* 50:4665-4668.
- Nikolovska-Coleska, Z., L. Xu, Z. Hu, Y. Tomita, P. Li, P.P. Roller, R. Wang, X. Fang, R. Guo, M. Zhang, M.E. Lippman, D. Yang, and S. Wang. 2004. Discovery of embelin as a cell-permeable, small-molecular weight inhibitor of XIAP through structure-based computational screening of a traditional herbal medicine three-dimensional structure database. *J Med Chem.* 47:2430-2440.

O

- Ochi, T., A.N. Blackford, J. Coates, S. Jhujh, S. Mehmood, N. Tamura, J. Travers, Q. Wu, V.M. Draviam, C.V. Robinson, T.L. Blundell, and S.P. Jackson. 2015. DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science.* 347:185-188.
- Ochi, T., X. Gu, and T.L. Blundell. 2013. Structure of the catalytic region of DNA ligase IV in complex with an Artemis fragment sheds light on double-strand break repair. *Structure.* 21:672-679.
- Ochi, T., B.L. Sibanda, Q. Wu, D.Y. Chirgadze, V.M. Bolanos-Garcia, and T.L. Blundell. 2010. Structural biology of DNA repair: spatial organisation of the multicomponent complexes of nonhomologous end joining. *J Nucleic Acids.* 2010.
- Ochi, T., Q. Wu, and T.L. Blundell. 2014. The spatial organization of non-homologous end joining: from bridging to end joining. *DNA Repair (Amst).* 17:98-109.
- Okada, H., and T.W. Mak. 2004. Pathways of apoptotic and non-apoptotic death in tumour cells. *Nat Rev Cancer.* 4:592-603.

P

- Pagliaro, L., J. Felding, K. Audouze, S.J. Nielsen, R.B. Terry, C. Krog-Jensen, and S. Butcher. 2004. Emerging classes of protein-protein interaction inhibitors and new tools for their development. *Curr Opin Chem Biol.* 8:442-449.
- Pearl, L.H., A.C. Schierz, S.E. Ward, B. Al-Lazikani, and F.M. Pearl. 2015. Therapeutic opportunities within the DNA damage response. *Nat Rev Cancer.* 15:166-180.
- Peddi, P., C.W. Loftin, J.S. Dickey, J.M. Hair, K.J. Burns, K. Aziz, D.C. Francisco, M.I. Panayiotidis, O.A. Sedelnikova, W.M. Bonner, T.A. Winters, and A.G. Georgakilas. 2010. DNA-PKcs deficiency leads to

- persistence of oxidatively induced clustered DNA lesions in human tumor cells. *Free Radic Biol Med.* 48:1435-1443.
- Petros, A.M., J. Dinges, D.J. Augeri, S.A. Baumeister, D.A. Betebenner, M.G. Bures, S.W. Elmore, P.J. Hajduk, M.K. Joseph, S.K. Landis, D.G. Nettlesheim, S.H. Rosenberg, W. Shen, S. Thomas, X. Wang, I. Zanze, H. Zhang, and S.W. Fesik. 2006. Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis. *J Med Chem.* 49:656-663.
- Pfaff, S.J., M.S. Chimenti, M.J. Kelly, and M.R. Arkin. 2015. Biophysical methods for identifying fragment-based inhibitors of protein-protein interactions. *Methods Mol Biol.* 1278:587-613.
- Povirk, L.F. 1996. DNA damage and mutagenesis by radiomimetic DNA-cleaving agents: bleomycin, neocarzinostatin and other enediynes. *Mutat Res.* 355:71-89.
- Przewloka, M.R., P.E. Pardington, S.M. Yannone, D.J. Chen, and R.B. Cary. 2003. In vitro and in vivo interactions of DNA ligase IV with a subunit of the condensin complex. *Mol Biol Cell.* 14:685-697.

R

- Rassool, F.V., and A.E. Tomkinson. 2010. Targeting abnormal DNA double strand break repair in cancer. *Cell Mol Life Sci.* 67:3699-3710.
- Recuero-Checa, M.A., A.S. Dore, E. Arias-Palomo, A. Rivera-Calzada, S.H. Scheres, J.D. Maman, L.H. Pearl, and O. Llorca. 2009. Electron microscopy of Xrcc4 and the DNA ligase IV-Xrcc4 DNA repair complex. *DNA Repair (Amst).* 8:1380-1389.
- Reichmann, D., O. Rahat, M. Cohen, H. Neuvirth, and G. Schreiber. 2007. The molecular architecture of protein-protein binding sites. *Curr Opin Struct Biol.* 17:67-76.
- Reid, D.A., S. Keegan, A. Leo-Macias, G. Watanabe, N.T. Strande, H.H. Chang, B.A. Oksuz, D. Fenyö, M.R. Lieber, D.A. Ramsden, and E. Rothenberg. 2015. Organization and dynamics of the nonhomologous end-joining machinery during DNA double-strand break repair. *Proc Natl Acad Sci U S A.* 112:E2575-2584.
- Riballo, E., L. Woodbine, T. Stiff, S.A. Walker, A.A. Goodarzi, and P.A. Jeggo. 2009. XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucleic Acids Res.* 37:482-492.
- Robins, P., and T. Lindahl. 1996. DNA ligase IV from HeLa cell nuclei. *J Biol Chem.* 271:24257-24261.
- Rooney, S., J. Chaudhuri, and F.W. Alt. 2004. The role of the non-homologous end-joining pathway in lymphocyte development. *Immunol Rev.* 200:115-131.
- Ropars, V., P. Drevet, P. Legrand, S. Baconnais, J. Amram, G. Faure, J.A. Marquez, O. Pietrement, R. Guerois, I. Callebaut, E. Le Cam, P. Revy, J.P. de Villartay, and J.B. Charbonnier. 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.* 108:12663-12668.
- Roy, S., S.N. Andres, A. Vergnes, J.A. Neal, Y. Xu, Y. Yu, S.P. Lees-Miller, M. Junop, M. Modesti, and K. Meek. 2012. XRCC4's interaction with XLF is required for coding (but not signal) end joining. *Nucleic Acids Res.* 40:1684-1694.
- Roy, S., A.J. de Melo, Y. Xu, S.K. Tadi, A. Negrel, E. Hendrickson, M. Modesti, and K. Meek. 2015. XRCC4/XLF interaction is variably required for DNA repair, and is not required for Ligase IV stimulation. *Mol Cell Biol.*

S

- Schames, J.R., R.H. Henchman, J.S. Siegel, C.A. Sottriffer, H. Ni, and J.A. McCammon. 2004. Discovery of a novel binding trench in HIV integrase. *J Med Chem.* 47:1879-1881.
- Shahar, O.D., A. Kalousi, L. Eini, B. Fisher, A. Weiss, J. Darr, O. Mazina, S. Bramson, M. Kupiec, A. Eden, E. Meshorer, A.V. Mazin, L. Brino, M. Goldberg, and E. Soutoglou. 2014. A high-throughput chemical screen with FDA approved drugs reveals that the antihypertensive drug Spironolactone impairs cancer cell survival by inhibiting homology directed repair. *Nucleic Acids Res.* 42:5689-5701.
- Shiloh, Y., and Y. Ziv. 2013. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol.* 14:197-210.
- Shrivastav, M., L.P. De Haro, and J.A. Nickoloff. 2008. Regulation of DNA double-strand break repair pathway choice. *Cell Res.* 18:134-147.
- Shuman, S., and C.D. Lima. 2004. The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Curr Opin Struct Biol.* 14:757-764.
- Sibanda, B.L., D.Y. Chirgadze, and T.L. Blundell. 2010. Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. *Nature.* 463:118-121.
- Sibanda, B.L., S.E. Critchlow, J. Begun, X.Y. Pei, S.P. Jackson, T.L. Blundell, and L. Pellegrini. 2001. Crystal structure of an Xrcc4-DNA ligase IV complex. *Nat Struct Biol.* 8:1015-1019.
- Siddiqi, M.A., and E. Bothe. 1987. Single- and double-strand break formation in DNA irradiated in aqueous solution: dependence on dose and OH radical scavenger concentration. *Radiat Res.* 112:449-463.
- Sperandio, O., C.H. Reynes, A.C. Camproux, and B.O. Villoutreix. 2010. Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug Discov Today.* 15:220-229.
- Srinivasan, A., and B. Gold. 2012. Small-molecule inhibitors of DNA damage-repair pathways: an approach to overcome tumor resistance to alkylating anticancer drugs. *Future Med Chem.* 4:1093-1111.
- Srivastava, M., M. Nambiar, S. Sharma, S.S. Karki, G. Goldsmith, M. Hegde, S. Kumar, M. Pandey, R.K. Singh, P. Ray, R. Natarajan, M. Kelkar, A. De, B. Choudhary, and S.C. Raghavan. 2012. An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell.* 151:1474-1487.
- Stjernschantz, E., and C. Oostenbrink. 2010. Improved ligand-protein binding affinity predictions using multiple binding modes. *Biophys J.* 98:2682-2691.
- Sugaya, N. 2009. [Integrative in silico approach to assess the druggability of protein-protein interactions]. *Tanpakushitsu Kakusan Koso.* 54:1641-1646.
- Sun, M.F., H.Y. Chen, F.J. Tsai, S.H. Lui, and C.Y. Chen. 2011. Search for novel remedies to augment radiation resistance of inhabitants of Fukushima and Chernobyl disasters: identifying DNA repair protein XRCC4 inhibitors. *J Biomol Struct Dyn.* 29:325-337.
- Surova, O., and B. Zhivotovsky. 2013. Various modes of cell death induced by DNA damage. *Oncogene.* 32:3789-3797.

T

- Thiel, P., M. Kaiser, and C. Ottmann. 2012. Small-molecule stabilization of protein-protein interactions: an underestimated concept in drug discovery? *Angew Chem Int Ed Engl.* 51:2012-2018.
- Thompson, A.D., A. Dugan, J.E. Gestwicki, and A.K. Mapp. 2012. Fine-tuning multiprotein complexes using small molecules. *ACS Chem Biol.* 7:1311-1320.
- Thompson, L.H. 2012. Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: the molecular choreography. *Mutat Res.* 751:158-246.
- Tobin, L.A., C. Robert, P. Nagaria, S. Chumsri, W. Twaddell, O.B. Ioffe, G.E. Greco, A.H. Brodie, A.E. Tomkinson, and F.V. Rassool. 2012. Targeting abnormal DNA repair in therapy-resistant breast cancers. *Mol Cancer Res.* 10:96-107.
- Tomkinson, A.E., T.R. Howes, and N.E. Wiest. 2013. DNA ligases as therapeutic targets. *Transl Cancer Res.* 2.
- Tomkinson, A.E., S. Vijayakumar, J.M. Pascal, and T. Ellenberger. 2006. DNA ligases: structure, reaction mechanism, and function. *Chem Rev.* 106:687-699.
- Trott, O., and A.J. Olson. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 31:455-461.
- Tsai, C.J., S.A. Kim, and G. Chu. 2007. Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc Natl Acad Sci U S A.* 104:7851-7856.

V

- Van Der Schans, G.P. 1978. Gamma-ray induced double-strand breaks in DNA resulting from randomly-inflicted single-strand breaks: temporal local denaturation, a new radiation phenomenon? *Int J Radiat Biol Relat Stud Phys Chem Med.* 33:105-120.
- Vassilev, L.T. 2004. Small-molecule antagonists of p53-MDM2 binding: research tools and potential therapeutics. *Cell Cycle.* 3:419-421.
- Villoutreix, B.O., M.A. Kuenemann, J.L. Poyet, H. Bruzzoni-Giovanelli, C. Labbe, D. Lagorce, O. Sperandio, and M.A. Miteva. 2014. Drug-Like Protein-Protein Interaction Modulators: Challenges and Opportunities for Drug Discovery and Chemical Biology. *Mol Inform.* 33:414-437.

W

- Walensky, L.D., A.L. Kung, I. Escher, T.J. Malia, S. Barbuto, R.D. Wright, G. Wagner, G.L. Verdine, and S.J. Korsmeyer. 2004. Activation of apoptosis in vivo by a hydrocarbon-stapled BH3 helix. *Science.* 305:1466-1470.
- Wang, C., and S.P. Lees-Miller. 2013. Detection and repair of ionizing radiation-induced DNA double strand breaks: new developments in nonhomologous end joining. *Int J Radiat Oncol Biol Phys.* 86:440-449.
- Wang, Y.G., C. Nnakwe, W.S. Lane, M. Modesti, and K.M. Frank. 2004a. Phosphorylation and regulation of DNA ligase IV stability by DNA-dependent protein kinase. *J Biol Chem.* 279:37282-37290.
- Wang, Y.S., D. Liu, and D.F. Wyss. 2004b. Competition STD NMR for the detection of high-affinity ligands and NMR-based screening. *Magn Reson Chem.* 42:485-489.

- Waters, C.A., N.T. Strande, D.W. Wyatt, J.M. Pryor, and D.A. Ramsden. 2014. Nonhomologous end joining: a good solution for bad ends. *DNA Repair (Amst)*. 17:39-51.
- Wells, J.A. 1996. Binding in the growth hormone receptor complex. *Proc Natl Acad Sci U S A*. 93:1-6.
- Wells, J.A., and C.L. McClendon. 2007. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*. 450:1001-1009.
- Willmore, E., S. de Caux, N.J. Sunter, M.J. Tilby, G.H. Jackson, C.A. Austin, and B.W. Durkacz. 2004. A novel DNA-dependent protein kinase inhibitor, NU7026, potentiates the cytotoxicity of topoisomerase II poisons used in the treatment of leukemia. *Blood*. 103:4659-4665.
- Woodbine, L., A.R. Gennery, and P.A. Jeggo. 2014. The clinical impact of deficiency in DNA non-homologous end-joining. *DNA Repair (Amst)*. 16:84-96.
- Wu, P.Y., P. Frit, S. Meesala, S. Dauvillier, M. Modesti, S.N. Andres, Y. Huang, J. Sekiguchi, P. Calsou, B. Salles, and M.S. Junop. 2009. Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4. *Mol Cell Biol*. 29:3163-3172.

X

- Xing, M., M. Yang, W. Huo, F. Feng, L. Wei, W. Jiang, S. Ning, Z. Yan, W. Li, Q. Wang, M. Hou, C. Dong, R. Guo, G. Gao, J. Ji, S. Zha, L. Lan, H. Liang, and D. Xu. 2015. Interactome analysis identifies a new paralogue of XRCC4 in non-homologous end joining DNA repair pathway. *Nat Commun*. 6:6233.
- Xue, X., J.L. Wei, L.L. Xu, M.Y. Xi, X.L. Xu, F. Liu, X.K. Guo, L. Wang, X.J. Zhang, M.Y. Zhang, M.C. Lu, H.P. Sun, and Q.D. You. 2013. Effective screening strategy using ensembled pharmacophore models combined with cascade docking: application to p53-MDM2 interaction inhibitors. *J Chem Inf Model*. 53:2715-2729.

Y

- Yano, K., K. Morotomi-Yano, N. Adachi, and H. Akiyama. 2009. Molecular mechanism of protein assembly on DNA double-strand breaks in the non-homologous end-joining pathway. *J Radiat Res*. 50:97-108.
- Yu, Y., B.L. Mahaney, K. Yano, R. Ye, S. Fang, P. Douglas, D.J. Chen, and S.P. Lees-Miller. 2008. DNA-PK and ATM phosphorylation sites in XLF/Cernunnos are not required for repair of DNA double strand breaks. *DNA Repair (Amst)*. 7:1680-1692.
- Yurchenko, V., Z. Xue, and M.J. Sadosky. 2006. SUMO modification of human XRCC4 regulates its localization and function in DNA double-strand break repair. *Mol Cell Biol*. 26:1786-1794.

Z

- Zhang, X., S. Morera, P.A. Bates, P.C. Whitehead, A.I. Coffer, K. Hainbucher, R.A. Nash, M.J. Sternberg, T. Lindahl, and P.S. Freemont. 1998. Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. *EMBO J*. 17:6404-6411.

Nom : Grégory Menchon

Directeurs de thèse : Pr Georges Czaplicki et Dr Patrick Calsou

Date : 28 septembre 2015

Lieu : IPBS, 205 route de Narbonne, 31400 Toulouse

Titre : Criblage virtuel et fonctionnel sur le complexe XRCC4/ADN Ligase IV/Cer-XLF de ligature des cassures double-brin de l'ADN ; application en radiosensibilisation tumorale.

En cancérologie, la radiothérapie est une des armes essentielles pour éradiquer les cellules tumorales. Les cassures des deux brins de l'ADN dites « double-brin » qu'elle induit sont particulièrement toxiques et constituent la principale cause de mort cellulaire. La NHEJ (Jonction d'Extrémités Non-Homologues) est la voie métabolique majeure de réparation de ces cassures double-brin de l'ADN et par ce mécanisme, les cellules humaines adoptent une résistance à la radiothérapie. Ce mécanisme de réparation constitue donc une cible de choix pour un traitement anticancéreux combiné en vue d'augmenter la sensibilité des cellules cancéreuses aux rayons ionisants (radiosensibilisation). Au cours du mécanisme NHEJ, la ligature finale des extrémités d'ADN est assurée par le complexe protéique tripartite: XRCC4/ADN Ligase IV/Cernunnos-XLF. Les interfaces protéiques concernées représentent toutes des cibles potentielles dans une stratégie rationnelle d'isolement de molécules inhibitrices, guidée par les structures tridimensionnelles de chaque protéine. A travers des expériences de criblage virtuel et de validation à la fois biophysique et biochimique, nous avons isolé les premières molécules capable de prévenir *in vitro* les interactions protéine-protéine pour les complexes XRCC4/Lig4 et XRCC4/Cer-XLF, respectivement. Ces composés sont des points de départ pour l'élaboration d'inhibiteurs potentiels de plus haute affinité grâce à l'apport de la biologie structurale, en vue d'un effet radiosensibilisant cellulaire.

Title : Virtual and functional screening on the XRCC4/DNA Ligase IV/Cernunnos-XLF DNA repair protein complex; application to tumor radiosensitization.

Radiotherapy is a major weapon used against cancer. Radio-induced DNA double strand breaks (DSB) are the main lesions responsible for cell death. Non-homologous end-joining (NHEJ) is a predominant DSB repair mechanism which contributes to cancer cells resistance to radiotherapy. NHEJ is thus a good target for strategies which aim at increasing the radio-sensitivity of tumors. Through *in silico* screening and biophysical and biochemical assays, our objective was to find specific ligands for the XRCC4/Lig4 and XRCC4/Cer-XLF protein-protein interactions involved in NHEJ. Here, we isolated the first compounds able to prevent their interaction *in vitro*. These early stage inhibitors are promising tools for cancer therapy with the hope to develop more specific compounds for cellular assays through the 3D structure of the protein/inhibitor complexes.

Mots clés : Cassures double-brin de l'ADN, NHEJ, criblage virtuel, criblage biophysique, XRCC4/ADN Ligase IV/Cernunnos-XLF, inhibiteur, radiosensibilisation tumorale.