

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

DÉCOUVERTE ET CARACTÉRISATION MOLÉCULAIRE DE NOUVELLES SUMO E3  
LIGASES ATYPIQUES CHEZ L'HUMAIN

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN BIOCHIMIE

PAR

ALI HARAKE

MAI 2024

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.12-2023). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

## REMERCIEMENTS

À Antoine Bouchard, Julien Plamondon, et Fabian Ortner, je dédie mes plus sincères remerciements. Votre amitié indéfectible et votre assistance constante ont été des piliers sur lesquels je me suis appuyé tout au long de ce parcours. Votre présence a été une source de motivation et d'inspiration inépuisable.

Laurent Cappadocia mérite tout particulièrement ma gratitude. Tes enseignements avisés et ton soutien inébranlable m'ont non seulement aidé à surmonter les défis scientifiques, mais m'ont également poussé à exceller en tant que chercheur. Ton influence a été cruciale dans mon développement professionnel.

Marc Lussier et Zoé Joly-Lopez, vos conseils ont été d'une grande pertinence et ont significativement contribué à l'avancement de mon travail. Votre expertise et votre approche ciblée ont été très appréciées.

Un immense merci à tous les membres de notre laboratoire pour leurs perspectives enrichissantes et leur collaboration précieuse. Chaque réunion a été une occasion d'apprendre et de grandir grâce à votre perspicacité.

J'aimerais aussi souligner mon appréciation pour Valérie Cabana et Amanda Szubinski pour leur amitié inestimable. Votre soutien et votre camaraderie durant les moments les plus intenses de cette aventure académique ont été d'une aide significative.

Une attention spéciale à mes compagnons fidèles, Bella et Lucy, pour leur support moral constant. Leur compagnie réconfortante à mes côtés lors des longues heures de travail a été une source chère de paix et d'énergie.

Enfin, je suis infiniment reconnaissant à Andréa Khouri pour son amour, son soutien sans faille, et sa présence rassurante. Andréa, tu as joué un rôle indispensable dans mon parcours, offrant à la fois appui émotionnel et encouragements.

## DÉDICACE

À ma mère, à mon père et à Andréa

« Parfois, la science est plus un art qu'une science, Morty. Beaucoup de gens ne comprennent pas ça »

— Rick Sanchez

## TABLE DES MATIÈRES

<b>REMERCIEMENTS</b> .....	<b>ii</b>
<b>DÉDICACE</b> .....	<b>iii</b>
<b>LISTE DES FIGURES</b> .....	<b>vii</b>
<b>LISTE DES TABLEAUX</b> .....	<b>ix</b>
<b>RÉSUMÉ</b> .....	<b>x</b>
<b>CHAPITRE 1 INTRODUCTION</b> .....	<b>1</b>
1.1 Les modifications post-traductionnelles des protéines.....	1
1.1.1 La phosphorylation.....	2
1.1.2 L'Ubiquitination.....	3
1.1.3 Conjugaison aux protéines Ubiquitine-like.....	3
1.1.4 La SUMOylation.....	5
1.2 Signalisation SUMO et implications de <i>SUMO Interacting Motifs</i> (SIMs).....	7
1.2.1 Les protéines responsables de la désSUMOylation : les SUMO-Specific Proteases (SENPs) .....	7
1.2.2 SUMO Interacting Motif.....	8
1.2.3 Site consensus et sites de SUMOylations .....	9
1.2.4 Localisation subcellulaire de SUMOylation .....	11
1.2.5 L'enzyme d'activation de SUMO : la E1 .....	13
1.2.6 L'enzyme de conjugaison à SUMO : la E2.....	14
1.2.7 Les SUMO E3 Ligases.....	15
1.2.7.1 Les SUMO E3 Ligases typiques .....	17
1.2.7.2 Les SUMO E3 ligases atypiques.....	18
1.2.7.3 Évolution et conservation des différents types de E3 .....	20
1.2.7.4 Modification des E3 .....	22
1.2.7.5 Spécificité des E3 .....	23
1.2.7.6 Interaction avec SUMO.....	24
1.2.8 Substrats de SUMOylation.....	25

1.2.9	Exploitation d'AlphaFold comme outil de prédiction des interactions SUMO .....	26
1.3	But du projet.....	27
1.3.1	Problématique et hypothèses .....	27
1.3.2	Objectifs de recherche.....	28
<b>CHAPITRE 2</b>	<b>MATÉRIEL ET MÉTHODES.....</b>	<b>30</b>
2.1	BIOINFORMATIQUE.....	30
2.1.1	Identification de protéines candidates SUMO E3 Ligases par une approche bio-informatique.....	30
2.1.2	Génération de modèles d'interactions Alphafold.....	31
2.2	BIOLOGIE MOLÉCULAIRE, EXPRESSION ET PURIFICATION DE PROTÉINES..	31
2.2.1	Biologie moléculaire et clonage de candidats .....	31
2.2.2	Extraction d'ARN total des cellules HEK293T et HeLa .....	37
2.2.3	Synthèse de l'ADNc à partir de l'ARN total.....	38
2.2.4	Réaction de PCR .....	38
2.2.5	Digestion de l'ADN par enzyme de restriction.....	38
2.2.6	Réaction de Gibson .....	38
2.2.7	Préparation de bactéries chimiocompétentes .....	39
2.2.8	Transformation bactérienne.....	39
2.2.9	Extraction de plasmides par minipreps .....	40
2.2.10	Extraction sur gel de bandes d'ADN.....	40
2.2.11	Expression bactérienne de protéine recombinante .....	40
2.2.12	Purification de protéines recombinantes .....	40
2.2.12.1	Chromatographie d'affinité NiNTA.....	40
2.2.12.2	Chromatographie par exclusion stérique.....	41
2.2.12.3	Chromatographie par échange ionique.....	41
2.2.12.4	Clivage des étiquettes des protéines recombinantes .....	42
2.2.12.5	Électrophorèse SDS-PAGE sur gel d'acrylamide.....	42
2.3	Immunobuvardage.....	42
<b>CHAPITRE 3</b>	<b>RÉSULTATS.....</b>	<b>44</b>
3.1	Criblage bio-informatique .....	44
3.1.1	Identification de protéines possédant des caractéristiques associées aux SUMO E3 ligases .....	44

3.1.2	Analyse bio-informatique des candidats retenus.....	50
3.2	Élaboration d'un criblage <i>in bactério</i> pour détecter l'activité SUMO E3 ligase de protéines d'intérêt.....	58
3.2.1	Considérations pour le design du criblage .....	58
3.2.2	Choix d'un vecteur pBAD pour exprimer les candidats .....	60
3.2.3	Clonage des candidats dans le vecteur pBAD.....	61
3.2.4	Détermination de la quantité d'arabinose nécessaire pour une expression optimale.....	62
3.2.5	SUMOylation <i>in bactério</i> des candidats .....	63
3.3	Purification de candidats .....	65
3.4	Essais de SUMOylation <i>in vitro</i> des candidats .....	68
<b>CHAPITRE 4</b>	<b>DISCUSSION .....</b>	<b>74</b>
4.1	Criblage bio-informatique et mécanisme SUMO E3 Ligase de candidats.....	74
4.2	Test <i>in bactério</i> et <i>in vitro</i> des candidats.....	85
	<b>CONCLUSION ET PERSPECTIVES .....</b>	<b>88</b>
	<b>ANNEXE A .....</b>	<b>90</b>
	<b>ANNEXE B .....</b>	<b>92</b>
	<b>ANNEXE C .....</b>	<b>93</b>
	<b>ANNEXE D .....</b>	<b>94</b>
	<b>ANNEXE E .....</b>	<b>96</b>
	<b>ANNEXE F.....</b>	<b>99</b>
	<b>BIBLIOGRAPHIE .....</b>	<b>100</b>

## LISTE DES FIGURES

Figure 1.1: Les différents types de modifications post-traductionnelles pouvant affecter les protéines .....	2
Figure 1.2: Structure tridimensionnelle de l'Ubiquitine ainsi que les Ubls ISG15, ATG8 et SUMO .....	4
Figure 1.3: Mécanismes employés par les E3 pour conjuguer les Ubls à des substrats.....	5
Figure 1.4: Cascade enzymatique de SUMOylation de substrat avec activité de clivage isopeptidique des SENP via maturation de SUMO ou déconjugaison de SUMO .....	8
Figure 1.5: Domaine SP-RING de PIAS1 en interaction avec la E2 et SUMO2/3.....	18
Figure 1.6: Interaction entre les domaines strictement nécessaires pour une activité SUMO E3 Ligase de ZNF451, RANBP2 et NSE2 avec la E2 et SUMO2/3 .....	20
Figure 1.7: Arbre phylogénétique de vraisemblance maximale des SUMO E3 Ligases provenant de Chlamydomonas reinhardtii, Arabidopsis thaliana, Homo sapiens, Saccharomyces cerevisiae. ....	22
Figure 3.1: Pipeline bio-informatique pour la sélection des candidats SUMO E3 Ligases .....	44
Figure 3.2: Représentation graphique de 1521 protéines présentant une localisation nucléaire et présentant au moins un site de SUMOylation selon la base de données de SUMOylation de Hendriks et Vertegaal.....	47
Figure 3.3: Diagramme de Venn illustrant les protéines localisées au noyau, SUMOylées et présentant une interaction avec la protéine E2.....	50
Figure 3.4: Représentation de la structure primaire de la E2 et SUMO2 ainsi que la structure tertiaire de l'enzyme E2 et de SUMO2.....	51
Figure 3.5: Interactions prédites par Alphafold entre les candidats avec SUMO2 et la E2 .....	52
Figure 3.6: Représentation schématique des domaines des candidats retenues et leurs sites de SUMOylation. ....	53
Figure 3.7: Structure tridimensionnelle, prédite au moyen d'Alphafold, des complexes entre les candidats et la machinerie de SUMOylation .....	56
Figure 3.8: Mécanisme SUMO E3 Ligases proposé des candidats en se basant sur les données d'Alphafold .....	57

Figure 3.9: Séquences en nucléotide et en acides aminés de la cassette qui a été insérée dans le site NcoI de pBAD. ....	61
Figure 3.10: Gel d'agarose 1 % avec les amplifications par PCR des différents candidats. Chaque puits correspond à une amplification différente d'un candidat .....	62
Figure 3.11: Gel SDS-Page 15 % montrant l'expression et la solubilité de SUMO1 suivant une induction de bactéries E. coli BL21DE3 (CP) par la L-Arabinose et IPTG pour pSUMO. ....	63
Figure 3.12: Détection de SUMOylation de substrats de bactéries exprimant un candidat en présence de pSUMO1 et pSUMO2/3 par Immunobuvardage par anticorps anti-HA.....	65
Figure 3.13: Séquences en nucléotide et en acides aminés des séquences utilisées pour la purification de candidats qui a été insérée dans le site BamHI d'un vecteur pRSF-TRX .67 .....	66
Figure 3.14: Gels SDS-Page 12 % montrant la purification des candidats HNRPC, TYDP2, IF5A1 et ZNF24. ....	68
Figure 3.15: Essais in vitro pour vérifier l'activité SUMO E3 Ligase des candidats HNRPC, ZNF24 et TYDP2 sur gels SDS-Page à gradient 10-20 % .....	69
Figure 5.1: Superposition de SUMO2/3 avec les domaines SUMO-like de HNRPC, NF2IP et RBMX.....	98

## LISTE DES TABLEAUX

Tableau 1.1: SUMO E3 Ligases connues et leurs caractéristiques cellulaires.....	16
Tableau 2.1: Amorces utilisées lors de PCR pour l'amplification de candidats .....	31
Tableau 3.1: Potentiel des protéines interagissant avec E2 en tant que SUMO E3 Ligases : Résumé des expériences réalisées .....	70
Tableau 3.2 : Potentiel des protéines ayant un SIM interagissant avec E2 en tant que SUMO E3 Ligases : Résumé des expériences réalisées.....	72

## RÉSUMÉ

La SUMOylation est une modification post-traductionnelle des protéines permettant aux cellules de s'adapter rapidement aux changements métaboliques ou environnementaux en modifiant les caractéristiques protéiques comme la localisation ou les interactions inter protéines. Cette modification réversible consiste en l'apposition d'une protéine nommée SUMO sur une protéine substrat. La SUMOylation implique l'action de trois protéines: une E1 pour activer SUMO, une E2 pour conjuguer SUMO et une E3 pour diriger l'enzyme E2 vers des substrats spécifiques. Les E3 se divisent en 2 catégories : typiques, contenant un domaine RING et atypiques, contenant d'autres motifs structuraux. La protéine ZNF451 appartient à cette dernière catégorie et contient deux motifs d'interaction à SUMO (SIMs) qui sont suffisants pour l'apposition de SUMO sur des protéines. Notre hypothèse est qu'il existe d'autres protéines ayant une activité E3 dépendant de motifs SIM. Grâce à une approche bio-informatique, qui inclut l'identification de caractéristiques communes des E3 et la prédiction de structures des protéines en complexes avec des composantes de la machinerie, 50 candidats ont été identifiés. Un candidat, ZNF24, a été identifié comme SUMO E3 ligase, utilisant possiblement un SIM et une lysine SUMOylée pour lier E2-SUMO, facilitant la SUMOylation subséquente de substrats. Ce travail souligne l'importance des motifs SIMs dans la médiation de l'activité SUMO E3 ligase, démontrant que des protéines sans domaine RING peuvent faciliter la SUMOylation. La découverte de ZNF24 enrichit notre compréhension de la diversité des mécanismes SUMO E3 ligase et ouvre de nouvelles avenues pour explorer l'implication de la SUMOylation dans divers processus biologiques.

Mots clés : SUMOylation, SIMs, SUMO E3 Ligases, Alphafold, ZNF451, essais *in vitro*, pSUMO, ZNF24

## CHAPITRE 1 INTRODUCTION

### 1.1 Les modifications post-traductionnelles des protéines

Les modifications post-traductionnelles sont des modifications ayant lieu sur des protéines afin de changer leurs caractéristiques. Celle-ci affectent les protéines après leur conception changeant leurs propriétés par clivage protéolytique ou par l'ajout de groupement (Figure 1.1). Les modifications post-traductionnelles des protéines peuvent comprendre l'addition de petits groupes chimiques, tels que la méthylation, l'acétylation, ou la phosphorylation. Elles peuvent également impliquer l'incorporation de structures moléculaires plus importantes, comme dans le cas de la SUMOylation ou de l'Ubiquitination, qui représentent l'ajout de protéines entières. (Ramazi et al, 2020). Les modifications peuvent être réversibles ou irréversibles (Wang et al, 2014). Les modifications réversibles sont principalement caractérisées par la modification des liaisons covalentes. En revanche, les modifications irréversibles sont associées aux processus de clivage protéique (Blom et al, 2004). Ces modifications affectent un ou plusieurs types d'acides aminés et entraînent des modifications dans les propriétés chimiques des sites modifiés (Huang et al, 2018). Ces processus biologiques affectant le comportement protéique incluent les fonctions et assemblages des enzymes (Ryšlavá et al, 2013), les interactions protéines-protéines (Marshall, 1993), trafic moléculaire, solubilité protéique (Karve et al, 2011) et localisation protéique (Audagnotto et al, 2017). Les modifications post-traductionnelles ont un impact sur la structure et le dynamisme des protéines influençant ainsi plusieurs processus biologiques (Mann et al, 2003; Xu et al, 2016). Ces modifications sont impliquées dans divers processus biologiques tels que la transduction du signal, la régulation de l'expression génique, l'activation des gènes, la réparation de l'ADN et le contrôle du cycle cellulaire (Wang et al, 2015; Strumillo et al, 2015; Wei et al, 2019).

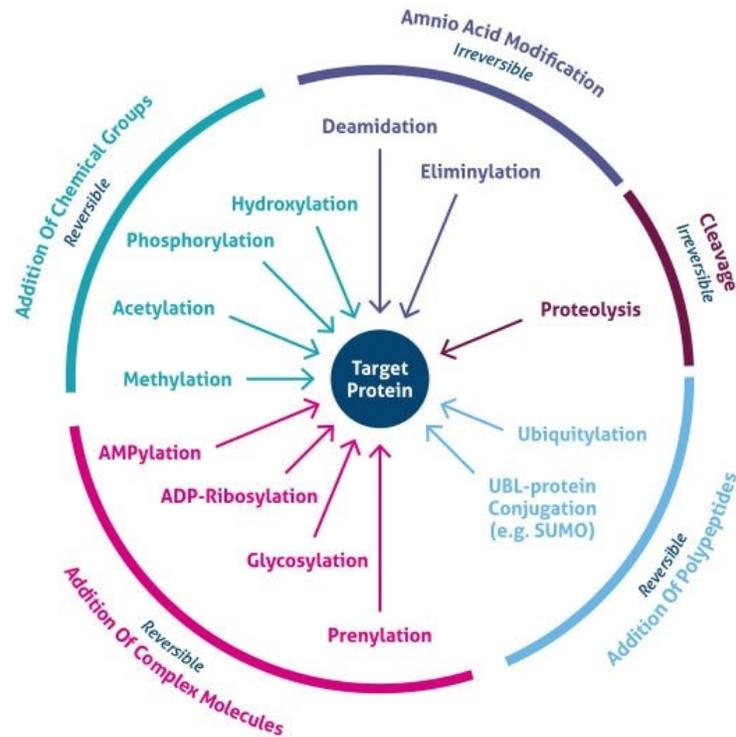


Figure 1.1: Les différents types de modifications post-traductionnelles pouvant affecter les protéines (Ribet et Cossart, 2010)

### 1.1.1 La phosphorylation

La phosphorylation est une modification post-traductionnelle covalente et réversible. Cette modification implique l'ajout d'un petit groupement, soit  $PO_4^-$ , sur une protéine. L'ajout de ce groupement a lieu sur les tyrosines, thréonines et serines pour les eucaryotes (Nestler et al, 1999) ou tyrosines, thréonines, serines, arginines, acide aspartique, histidines et cystéines pour les procaryotes (Shi et al, 2020). Cette réaction permet d'altérer la charge protéique se retrouvant sur la surface de la protéine ce qui pourrait, par la suite, permettre un changement de conformation (Nestler et al, 1999). La phosphorylation de résidus est un mécanisme réversible régulant divers processus dont la régulation de la synthèse protéique, la division cellulaire, la transduction de signal et la croissance cellulaire, la régulation de l'activité enzymatiques et autres (McCance et al, 2014). La réaction de phosphorylation nécessite l'action d'enzymes spécifiques : les kinases, responsables de l'ajout de groupements phosphate aux acides aminés, et les phosphatases, qui retirent ces groupements phosphate. Cette réaction dépend de l'ATP qui agit comme donneur de phosphate (Ardito et al, 2017). Cette modification est la plus prévalente affectant presque 1/3 du protéome humain à tout moment (Vlastaridis et al, 2017) ainsi que 5 à 10 % du protéome bactérien (Pietack

et al, 2010). Pour les sites de phosphorylation humains, au moins 95 % des groupements phosphates se retrouvent sur les serines, 3 à 4 % sur les thréonines et moins de 1 % sur les tyrosines (Nestler et al, 1999).

### 1.1.2 L'Ubiquitination

L'Ubiquitination est une modification covalente post-traductionnelle qui est réversible. Cette modification consiste en l'apposition d'une protéine de 76 acides aminés, soit de ~9 kDa, sur une protéine substrat. Cette modification se produit dans le cytoplasme et dans le noyau (Lingbeck et al., 2003). Cette modification a lieu lors de plusieurs événements cellulaires tels que le changement de compartiment subcellulaire de protéines, l'activation de kinases ou bien lors de dégradation de protéines. Celle-ci requiert l'action de 3 enzymes de manière séquentielle : une E1 qui va venir activer l'Ubiquitine, une E2 qui va venir lier l'ubiquitine activée et une E3 qui va venir faciliter le transfert de l'ubiquitine vers le substrat à travers divers mécanisme. Celle-ci peut se faire avec une seule Ubiquitine soit simple (monoubiquitination) ou multiple (polyubiquitination) via les lysines accessibles. (Swatek et al, 2016). Il existe entre 600 à 700 Ubiquitine E3 Ligases chez l'humain ou proche de 5 % du génome, mais aucune chez la bactérie (Vozandychova et al, 2021; George et al, 2018).

### 1.1.3 Conjugaison aux protéines Ubiquitine-like

Les Ubls ou les protéines apparentées à l'Ubiquitine (Ubiquitine-Like proteins) sont une classe de modifications post traductionnelles de protéine qui ressemblent structurellement à l'Ubiquitine (Figure 1.2) (Vlastaridis et al., 2017; Pietack et al., 2010; Nestler et al., 1999). La plupart des Ubls, soit presque 20, ont une cascade enzymatique similaire à celle de l'ubiquitine et cette cascade permet l'ajout d'un Ubl sur une protéine. La cascade enzymatique est composée d'une E1 qui va activer l'Ubl, une E2 qui va conjuguer l'Ubl et une E3 qui va permettre l'apposition de l'Ubl sur des substrats de façon spécifique. Parmi les Ubls les plus connues et les plus étudiées, figurent SUMO (voir section 1.1.4), Nedd8, ISG15 et ATG8 (Figure 1.2) (Van der veen et al., 2012). Étant donné que la plupart des Ubl ont moins de substrats connus que l'ubiquitine, ces modifications nécessitent un nombre plus restreint d'enzymes E2 ainsi que de E3 pour altérer les protéines (Van der veen et al., 2012). Cependant chaque modification possède sa E1, sa E2 et sa E3 pour modifier

les substrats. De plus, les différentes modifications, par le biais de la E3, possèdent différents mécanismes pour modifier les acides aminés visés.

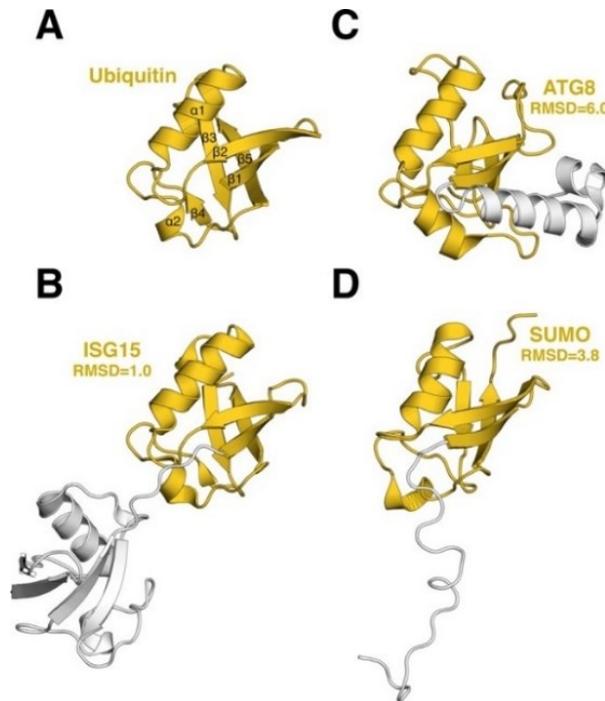


Figure 1.2: Structure tridimensionnelle de l'Ubiquitine (A) ainsi que les Ubls ISG15 (B), ATG8 (C) et SUMO (D) (Cappadocia et al., 2018)

Premièrement, Nedd8 est l'Ubl avec la séquence qui ressemble le plus à l'ubiquitine avec 58 % de similarité de séquence. Cette protéine, cruciale pour le développement de l'organisme, peut être létale en cas de mutation. Nedd8 utilise une seule E1 (NAE), deux E2 (Ubc12 et Ube2F) et certaines E3 (Rabut et al., 2008; Watson et al., 2011). La reconnaissance du substrat pour la neddylation se fait par la E3 grâce à un domaine RING (*Really Interesting New Gene*), Rbx1 ou Rbx2. Les enzymes E3 de la neddylation contiennent un ou plusieurs domaines RING qui est également impliquée dans l'Ubiquitination et SUMOylation (Figure 1.3). Ce domaine de doigts de Zinc ou Zinc fingers, contient certains motifs qui vont lier deux ions de Zincs à travers un domaine  $C_3HC_4$  (Cys-X<sub>2</sub>-Cys-X<sub>[9-39]</sub> — Cys-X<sub>[1-3]</sub> — His-X<sub>[2-3]</sub> — Cys-X<sub>2</sub>-Cys-X<sub>[4-48]</sub> — Cys-X<sub>2</sub>-Cys ou X indique n'importe quel acide amine) et permet la conjugaison du substrat avec l'Ubl (Borden et al., 1996).

Pour ISG15 ou *Interferon Stimulated Gene 15*, celle-ci contient deux domaines ressemblant à l'ubiquitine fusionnée ensemble (domaine Ubiquitine-like) (Jeon et al., 2010). La conjugaison par ISG15 implique une E1 (Ube1L), deux E2 (UbcH6 et UbcH8) et des E3 (Jeon et al., 2010). L'enzyme E3 prédominante pour cette modification est la Herc5 qui est une E3 ligase de type HECT (Homologous to the E6-AP Carboxyle Terminus) (Jeon et al., 2010; Malakhov et al., 2002). Les domaines de type HECT, impliqué également dans l'ubiquitination, est un domaine qui, par le biais d'une cystéine, permet le transfert de l'Ubl de la E2 vers le substrat (Figure 1.3). Ce type de transfert se fait en relais de la E2 vers la E3 suivi du substrat. Contrairement au domaine de type RING qui fait un relais de la E2 directement sur le substrat par la E3 qui permet un rapprochement entre ces éléments (Figure 1.3) (Sluimer and Distel, 2018; Weber et al., 2019).

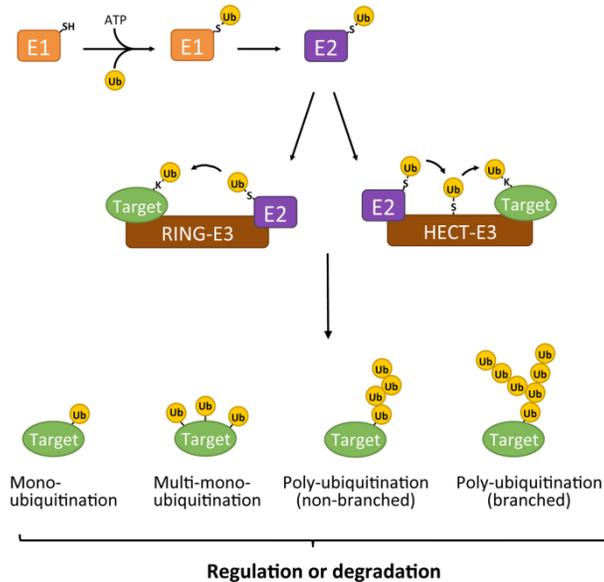


Figure 1.3: Mécanismes employés par les E3 pour conjuguer les Ubls à des substrats. (Sluimer et al., 2018)

#### 1.1.4 La SUMOylation

Durant les années 1990, il a été découvert qu'une protéine tridimensionnelle similaire à l'ubiquitine (Bayer et al., 1998; Mossessova et al., 2000; Bernier-Villamor et al., 2002), soit SUMO ou Small Ubiquitin-like Modifier, se fixe de manière covalente à ses cibles, modifiant leur fonctionnement (Matunis et al., 1996; Mahajan et al., 1997). La découverte de SUMO s'est déroulée en trois étapes. Tout d'abord, le gène SUMO (SMT3) a été identifié chez *Saccharomyces cerevisiae* lors d'une recherche de suppresseurs d'une protéine centromérique (Meluh et al., 1995). Suivant cette étude

initiale, trois études ont révélé que SUMO était un partenaire de liaison pour certaines protéines humaines RAD51, RAD52, (Shen et al., 1996) FAS (Okura et al., 1996) et PML (Boddy et al., 1996) dans des essais de double hybride chez la levure. Finalement, SUMO a été découvert attacher de manière covalente à la protéine RanGAP1 (Matunis et al., 1996; Mahajan et al., 1997). La protéine au sein de cette modification post-traductionnelle, soit SUMO, est une protéine de ~10 kDa. Le processus d'apposition de SUMO, soit la SUMOylation, est une modification covalente et réversible. La SUMOylation est une modification ayant lieu principalement dans le noyau des cellules eucaryotes. Cependant, elle n'est pas exclusivement confinée au noyau. Celle-ci joue des rôles significatifs dans divers processus cellulaires à la fois dans le noyau et le cytoplasme. (Matunis et al., 2016). Cette modification post-traductionnelle est cruciale pour divers processus cellulaires. Les fonctions principales de cette modification sont la régulation de fonction protéique, la transcription génique, la régulation de la localisation subcellulaire, la réparation des cassures de l'ADN, le cycle cellulaire et autres (Maejima et al., 2014). Illustrant l'impact de la SUMOylation sur la dynamique subcellulaire des protéines, la protéine p53 sert d'exemple frappant. Cette protéine, centrale dans les mécanismes de réponse au cancer, subit une modification de sa localisation, passant du noyau au cytoplasme à la suite de sa SUMOylation. Ce phénomène souligne l'importance des modifications post-traductionnelles dans la régulation fonctionnelle des protéines au sein de la cellule (Wu et Cheng, 2009). La conjugaison de SUMO sur le substrat se fait via un lien isopeptidique entre le C-terminal de SUMO et la chaîne latérale d'une lysine (Naik et al, 2017). Un peu comme l'Ubiquitination, la SUMOylation implique l'action séquentielle de 3 enzymes : une E1 pour activer le SUMO, une E2 pour lier SUMO activé et une E3 qui va venir rapprocher la E2 et le substrat afin de SUMOyler ce dernier. (Celen et al, 2020). Contrairement à l'ubiquitination, la SUMOylation présente un site consensus dans lequel la E2 va SUMOylée les protéines soit :  $\Psi KxE$  avec  $\Psi$  étant un acide aminé aliphatique et X étant n'importe quel acide amine (Geiss-Friedlander et al., 2007). Trois isoformes de SUMO, SUMO1, SUMO2, et SUMO3, jouent un rôle majeur dans ce processus, principalement dans le noyau (Imbert et al, 2022). Parmi eux, SUMO2 et SUMO3 présentent une ressemblance de 97 %, tandis que SUMO2 et SUMO1 partagent seulement 50 % de similitude. SUMO4 et SUMO5, exprimés spécifiquement dans certains tissus, ne sont pas considérés ici en raison de leur portée limitée (Celen et al, 2020). Due à leurs similarités, SUMO2 et SUMO3 sont généralement indiscernables, tant au niveau des expériences, qu'au niveau de la littérature. Au niveau de la localisation, SUMO1 se retrouve au

niveau du nucléole, l'enveloppe nucléaire et foyer cytoplasmiques tandis que SUMO2/3 s'accumule au niveau des chromosomes et peut se retrouver dans le cytoplasme (Suk et al., 2023; Hecker et al., 2006). SUMO2/3 a également tendance à former des chaînes (polySUMOylation) sur les substrats (Sun et al., 2012) tandis que SUMO1 a tendance à se retrouver seul (monoSUMOylation) et rarement en chaîne (Yang et al., 2006).

## 1.2 Signalisation SUMO et implications de *SUMO Interacting Motifs* (SIMs)

### 1.2.1 Les protéines responsables de la désSUMOylation : les SUMO-Specific Proteases (SENP)

Chez les mammifères, les SENP ou SUMO-Specific Proteases sont des protéases à base de cystéine ayant une activité isopeptidique au niveau du noyau et du cytoplasme. Les SENPs consistent en une famille de 7 membres avec SENP1, SENP2, SENP3, SENP5, SENP6, SENP7 et SENP8. Cependant, SENP8 est une protéase spécifique à l'Ubl Nedd8 donc celle-ci sera omise pour le restant de ce mémoire (Kunz et al., 2018). Parmi ces isopeptidases, il existe des formes qui existent uniquement au noyau (SENP3, SENP5, SENP6) ou peuvent se retrouver dans le noyau et ainsi que le cytoplasme (SENP1, SENP7, SENP2) (Kunz et al., 2018). L'activité de ces isopeptidases est de faciliter la déconjugaison de SUMO des substrats protéique via un clivage isopeptidique (Figure 1.4) (Tokarz et al., 2021). Ceci permet un équilibre entre la forme conjuguée et non-conjuguée de SUMO à des substrats (Kumar et al., 2015). Cette réaction, principalement nucléaire, est connue sous le nom de désSUMOylation. La SUMOylation est une modification réversible grâce à ces isopeptidases qui peut cliver SUMO se trouvant sur les substrats protéiques (Tokarz et al., 2021). Il est également possible de noter que ces enzymes sont indispensables pour la réaction de SUMOylation puisque les SENP clivent les formes inactives ou formes pros de SUMO au C terminal. Cette activité d'hydrolase permet d'exposer le motif diglycine de SUMO générant ainsi une SUMO mature (Kumar et al., 2015). Ces enzymes sont connues pour réguler des voies de signalisation telles que la réparation des cassures double brins de l'ADN, le cycle cellulaire, néovascularisation et autres processus cellulaires (Tokarz et al., 2021). Il existe 6 isoformes de cette enzyme identifiée à ce jour chez les mammifères soit SENP1, SENP2, SENP3, SENP5, SENP6 et SENP7. Les isoformes 1 et 2 constituent une première famille de protéase spécifique à SUMO et sont spécifiques pour les SUMO1, 2 et 3. Les SENP 3 et 5 forment une seconde famille tandis que SENP 6 et 7 forment une troisième famille qui ont une spécificité pour SUMO2 et 3

(Kumar et al., 2015). Chez les levures, le mécanisme de désSUMOylation est assuré par les protéines Ulp1 et Ulp2. Ulp1 est connue pour avoir une activité d'hydrolase et d'isopeptidase tandis que Ulp2 n'a qu'une activité d'isopeptidase. Les SENP peuvent être englobés en deux familles distinctes permettant de mieux les identifier : la famille Ulp1 (SENP1, SENP2, SENP3 et SENP5) et la famille Ulp2 (SENP6 et SENP7) (Hickey et al., 2012). De plus, Ulp1 est principalement nucléaire tandis que Ulp2 est principalement cytoplasmique.

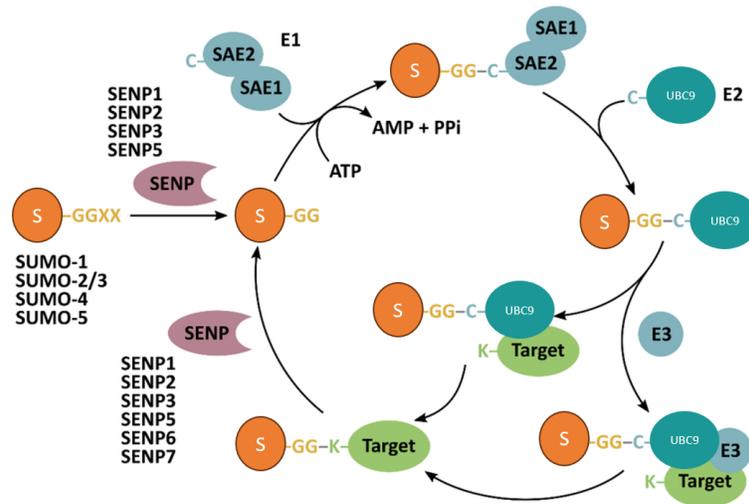


Figure 1.4: Cascade enzymatique de SUMOylation de substrat avec activité de clivage isopeptidique des SENP via maturation de SUMO ou déconjugaison de SUMO. Le cercle orange représente une protéine SUMO. (Figure adaptée de Tokarz et Wozniak, 2021)

### 1.2.2 SUMO Interacting Motif

Les SIMs ou SUMO Interacting Motif sont de petits motifs typiquement composés entre trois et cinq acides aminés hydrophobes, incluant des acides aminés acides situés en amont ou en aval de ceux-ci. Souvent, les SIMs ne contiennent pas de résidus acides mais des résidus phosphorylables. Ces phospho-SIMs pourront mimer une charge négative, lorsque phosphorylée (Lascorz et al., 2022). Ces résidus acides, chargés négativement, interagissent avec des résidus chargés positivement à la surface de SUMO, augmentant l'affinité entre le SIM et SUMO (Lascorz et al., 2022). Les SIMs font des liaisons de complémentarité en orientation parallèle ou anti-parallèle avec un feuillet  $\beta$  de SUMO (Cappadocia et al., 2015). Ils sont principalement sous la forme  $\Psi\Psi_x\Psi/\Psi_x\Psi\Psi/\Psi\Psi\Psi\Psi/x\Psi\Psi\Psi$  ( $\Psi$  sont les acides aminés aliphatiques et x n'importe quel acide aminé), présentant une liaison accrue à SUMO (Geiss-Friedlander et al., 2007). Cependant, il existe

d'autres formes de SIMs telles que  $\Psi D \Psi D \Psi$  qui existent (Beauclair et al., 2015). Les liaisons entre SUMO et SIM se font de manière non covalente (Varejao et al., 2010) et ont généralement des constantes de dissociation (Kd) entre 1 et 100  $\mu\text{M}$  (Song et al., 2005; Hecker et al., 2006). Certaines protéines ont été démontrées comme étant des protéines qui nécessitent les SIM pour leur fonction (comme DAXX et PML), mais également pour leur SUMOylation (comme Daxx et certaines SUMO E3 Ligases PIAS) (Geiss-Friedlander et al., 2007). Ce motif est principalement un motif d'amarrage permettant le recrutement de la E2 chargé avec SUMO ou de protéines étant SUMOylées (Yang et al., 2010). Ce motif particulier est présent chez toutes les SUMO E3 Ligase découvert à ce jour (Shi et al., 2022). Les protéines contenant des SIMs ont parfois des préférences pour un isoforme de SUMO comme ZNF451 ayant une préférence pour SUMO2/3 (Kd de 120-150  $\mu\text{M}$ ) (Cappadocia et al., 2015) ou Daxx ayant une préférence pour SUMO1 (Kd de 55  $\mu\text{M}$ ) (Chang et al., 2011). La cavité hydrophobe dans SUMO1 et SUMO2/3 étant assez conservée, les différences les plus significatives entre les isoformes de SUMO humaines se trouvent dans la zone électropositive qui se lie à la région acide du SIM. Les différences dans les contacts électrostatiques établis par ces résidus dans les paralogues de SUMO affectent l'affinité et la spécificité de chaque SIM ou phospho-SIM particulier (Lascorz et al., 2022). Donc, les conséquences des protéines conjuguées à SUMO dépendent d'interactions protéine-protéine non covalentes directes entre SUMO et les protéines possédant des domaines de liaison à SUMO.

Il est possible, par des algorithmes, de détecter la présence de SIMs présent chez les protéines (Zhao et al., 2014; Beauclair et al., 2015). Les plateformes principales capables de ceci sont GPS-SUMO et JASSA qui peuvent déterminer la présence de SIMs dans les séquences protéiques basées sur des données expérimentales (voir section 1.2.3). Cependant, l'accessibilité de SIMs est difficile à prédire dû au manque de structures expérimentales ; certains motifs hydrophobes dans les domaines de protéines peuvent être mal interprétés comme étant des SIMs.

### 1.2.3 Site consensus et sites de SUMOylations

Bien que l'on ait initialement postulé que la SUMOylation intervenait essentiellement sur des sites spécifiques, tels que  $\Psi K x E$ , ou inversement,  $[E/D]-x-K-\Psi$ , lorsque la structure le permettait (Rodriguez et al., 2001), la découverte de la SUMOylation de la protéine Daxx en 2006 a remis en question cette norme. Daxx est devenu un exemple précurseur, illustrant que la SUMOylation peut

également se produire sur des sites non-consensus, ouvrant ainsi la voie à une nouvelle compréhension de la flexibilité et de la diversité des cibles de SUMOylation (Lin et al., 2006). Le mécanisme découlant de cette SUMOylation non-consensus de Daxx est le recrutement de la E2 chargée avec SUMO via un motif d'interaction à SUMO (SIM) qui lie SUMO. Cette interaction forcerait la E2 de relarguer SUMO sur une lysine non-consensus sur la protéine contenant le SIM ou une protéine avoisinante. Une liaison SUMO-SIM pourrait également changer la conformation de la protéine contenant le SIM permettant une auto-SUMOylation (Lin et al., 2006). Suite à cette découverte, d'autres protéines contenant des SIMs ont également été identifiées avec BLM et HIPK2 étant SUMOylée à des sites non-consensus (Zhu et al., 2008). Aujourd'hui, il est possible d'observer que proche de 50 % des sites de SUMOylation ne sont pas des sites consensus ou consensus inverse (Hendriks et al., 2014; Lamoliatte et al., 2014). De plus, il existe des variantes du motif de SUMOylation soit les NDSM (*negatively charged amino acid-dependent SUMOylation motif*) et les PDSM (*phosphorylation dependent SUMOylation motifs*). Le motif PDSM correspond à une séquence avec un acide aminé phosphorylable en aval du site de SUMOylation tandis que NDSM correspond à des acides aminés négatifs en aval du site consensus (Beauclair et al., 2015; Yang et al., 2006b). Ceci se voit de la façon suivante :  $\psi$ -K-x-E-x-x-S-P avec S étant une serine phosphoryler et P étant une proline pour le PDSM et  $\psi$ -K-x-E-x-x-(D/E)-(D/E)-(D/E)-(D/E)-x-x-x-x pour le NDSM (Beauclair et al., 2015; Hendriks et al., 2014; Mohideen et al., 2009; Yang et al., 2006b). Cependant, seulement ~75 % des protéines SUMOylées possèdent un motif consensus de SUMOylation. Presque ~25 % des lysines ne se trouvant pas dans un motif consensus sont SUMOylées donc même si ce motif est un bon indicateur de SUMOylation, il n'est pas absolu (Wilkinson et al., 2010).

Il existe aujourd'hui des algorithmes capables de prédire la SUMOylation de sites consensus, consensus inverse, PDSM, NDSM et non consensus. Parmi les algorithmes, il existe plusieurs approches pour la détection de sites de SUMOylation potentiels comme SUMOsp et GPS-SUMO qui utilise un système de prédiction basée sur des groupes (GPS) ou Jassa qui utilise une approche de matrice de fréquence de position (FPM) basée sur un alignement de séquence avec des données validées expérimentalement (Chang et al., 2018).

Pour JASSA, les acides aminés se retrouvant dans l'environnement des lysines sont accordés un score grâce à un algorithme. L'algorithme employé utilise une matrice de fréquence de position pour évaluer chaque acide aminé, afin de déterminer si les lysines se situent dans des environnements susceptibles d'être SUMOylés. Par la suite, l'algorithme va décider, via un arbre de décision, d'inclure, ou non, la lysine pouvant se faire SUMOylée en se basant sur le score de chaque lysine. Pour les SIMs, une approche différente est utilisée ; les acides aminés des protéines sont comparés à une liste de SIMs établie dans la littérature. À chaque groupe d'acides aminés est attribué un score, et comme mentionné précédemment, ceux surpassant un certain seuil seront présentée à l'utilisateur comme étant un motif d'interaction avec SUMO. Ce score augmente lorsque des acides aminés phosphorylables ou acidiqes se retrouvent en amont ou en aval (D, E, S) du noyau hydrophobe (Beauclair et al., 2015).

Pour DEEPSUMO, une base de données ne présentant pas ou peu de redondances pour des lysines SUMOylées et SIMs ont été établies par recherche manuelle de littérature. Grâce à un algorithme GPS de quatrième génération (Group-based Prediction System), un score est accordé à chaque séquence. Un score positif est accordé aux sites de SUMOylation ressemblant ceux qui ont été identifié empiriquement. Pour raffiner l'algorithme, 4 étapes de formation ont été suivies : le clustering k-means, la sélection de la longueur du motif, l'entraînement des poids, et la mutation matricielle. Ces formations permettent davantage de filtrer les sites de SUMOylation pour refuser les faux positifs. Basé sur un retour continu, l'algorithme subit des mises à jour de façon périodique pouvant améliorer ses capacités prédictives (Zhao et al., 2014).

Cependant, les deux méthodes, soit DEEPSUMO et Jassa, utilisent des sites de SUMOylation expérimentalement identifiés pour constituer une base de données, qui sert ensuite à comparer chaque nouvelle séquence. En bref, les algorithmes de prédiction reposent sur l'entraînement d'un modèle par de nombreuses données de SUMOylation, permettant d'extraire des tendances spécifiques. Ce processus aboutit à des modèles simplifiés, utilisés ensuite pour détecter les sites de SUMOylation (Zhao et al., 2014; Beauclair et al., 2015).

#### 1.2.4 Localisation subcellulaire de SUMOylation

La SUMOylation est un processus qui requiert l'activité de plusieurs enzymes. Parmi celles impliquées dans la SUMOylation, une grande majorité possède une séquence de localisation

nucléaire (NLS) (Truong et al., 2012). La SUMOylation requiert l'action séquentielle de 3 enzymes pour être rajoutée sur des protéines donc la localisation de ces 3 enzymes va affecter le niveau de SUMOylation des protéines. Ces enzymes sont principalement localisées dans le noyau, toutefois, elles peuvent également se trouver dans le cytoplasme afin de réguler certaines protéines, telles que les canaux potassiques (Mayo, 2012; Naidu et al., 2012; Wasik et al., 2014). La protéine SUMO, par une détection par immunobuvardage et immunofluorescence, semble localiser au niveau du nucléoplasme, la membrane externe de la mitochondrie, le réticulum endoplasmique, membrane plasmique et pore nucléaire (Geiss-Friedlander et al., 2007).

Cependant, les enzymes responsables de la SUMOylation sont fortement concentrées dans le noyau, ce qui indique que la SUMOylation est principalement une réaction nucléaire (Wasik et al., 2014). Parmi les enzymes responsables de la SUMOylation, la E1, enzyme d'activation de SUMO, est une protéine principalement nucléaire pouvant se retrouver au cytoplasme. Sa localisation subcellulaire cytoplasmique dépend de son niveau de SUMOylation avec une localisation cytoplasmique lorsque faiblement SUMOylée (Truong et al., 2012). La deuxième enzyme impliquée dans la SUMOylation, soit la E2, est une enzyme qui vient conjuguer SUMO. Cette protéine a été démontrée comme ayant à la fois une activité au niveau du cytoplasme et au niveau du noyau avec une localisation cytoplasmique au niveau du pore nucléaire (Zhang et al., 2002). Donc, la E1 et la E2, pour le processus de SUMOylation, se retrouvent majoritairement dans le noyau avec un potentiel de localisation au cytoplasme. Les fonctions nucléaires de SUMO sont variées allant du maintien des structures nucléaires (corps PML et corps Polycomb) jusqu'à la modulation d'interactions protéines-protéines. SUMO a un effet sur divers structures nucléaires tels que l'hétérochromatine, les télomères et les centromères. Un exemple de protéine importante qui doit se faire SUMOylée est la Topoisomérase II (TOP2). La SUMOylation de TOP2 permet à cette protéine de se localiser aux centromères et effectuer sa fonction de décaténer l'ADN entrelacé avant l'anaphase. Un autre exemple de protéine importante est Orc2 qui, lorsque SUMOylée, permet la production d'ARN non codant permettant la maintenance des hétérochromatines (Zhao, 2018). En bref, la SUMOylation est un processus essentiel à la cellule qui permet le maintien de l'homéostasie. Bien que SUMO agisse principalement sur le noyau, celui-ci influence la cellule en entier. Les fonctions cytoplasmiques de SUMO, bien que les données de littérature ne soient pas exhaustives, indiqueraient que la SUMOylation joue un rôle au niveau des activités de canaux

potassiques et calciques, de signalisation de protéine G, régule certaines enzymes dont certaines kinases (*Calcium/calmodulin-dependent serine protein kinase*, *Focal adhesion kinase*, *Protein-tyrosine phosphatase 1B*), joue un rôle au niveau du cytosquelette et agit sur la fragmentation des mitochondries (Wasik et al., 2014). Cependant, il est difficile de distinguer l'impact de la SUMOylation de protéines nucléaires sur des phénomènes cytoplasmiques et l'effet de la SUMOylation de protéines cytoplasmiques sur ces mêmes phénomènes.

En ce qui concerne les 11 SUMO E3 caractérisés à ce jour, 8 protéines parmi les 11 protéines présentent une localisation cellulaire nucléaire avec peu ou aucune évidence suggérant une activité cytoplasmique (Tableau 1.1). Parmi les 3 protéines restantes qui n'exercent pas une fonction strictement nucléaire, 1 protéine exerce une fonction majoritairement nucléaire avec des évidences d'activité au niveau du cytoplasme (Tableau 1.1). Parmi les 11 protéines, il semblerait que seulement 1 protéine semble être majoritairement cytoplasmique (PIAS3) jouant une activité au niveau des canaux voltage dépendant au niveau de la membrane plasmique (Geiss-Friedlander et al., 2013). Considérant les fonctions de la SUMOylation, soit la régulation de la transcription, la régulation de structure de chromatine et réparation de l'ADN, cette localisation cellulaire semble être adéquate et permet l'homéostasie cellulaire.

### 1.2.5 L'enzyme d'activation de SUMO : la E1

La cascade signalétique, que ce soit pour l'ubiquitine ou pour les Ubl comme SUMO, requiert l'action séquentielle de plusieurs enzymes afin de se conjuguer à des protéines (voir Figure 1.4). Pour la E1, elle initie la conjugaison de SUMO à travers l'activation de la protéine SUMO clivée préalablement par les SENP. Pour activer SUMO via la E1, de l'ATP est nécessaire pour initier la réaction. Une seule enzyme E1 existe à ce jour pour la SUMOylation chez l'humain. Cette enzyme est un hétérodimère de ~110 kDa composé de 2 sous-unités, soit SAE1 et SAE2 (SUMO Activating enzyme 1 et 2). L'hétérodimère E1 est présent de façon prédominante dans le noyau, mais peut se retrouver dans le cytoplasme (Zhang et al., 2002; Melchior et al., 2003). La première sous-unité SAE2, soit celle d'adénylation (AD ou adenylation domain) peut exister sous sa forme active ou inactive et existe pour reconnaître SUMO ainsi que la fixation de l'ATP et  $Mg^{2+}$ . Cette adénylation a lieu sur le C terminal de SUMO durant les premières étapes d'activation de SUMO. Le domaine catalytique cystéique est divisé en 2 sous-domaines avec FCCH (first catalytic cysteine half-

domain) et SCCH (second catalytic cystein half-domain) (Lv et al., 2018). Ce second sous-domaine est celui qui lie SUMO via la cystéine catalytique formant ainsi le lien thioester. L'adénylation et la formation du lien thioester sont toutes deux catalysées en un seul site sur l'enzyme. Ces réactions ont lieu grâce à plusieurs changements de conformation complémentaire. Le transfert de SUMO de la cystéine catalytique de la E1 a lieu sur la cystéine catalytique de la E2. Toutes les E1 des Ubl ont un mécanisme catalytique séquentiel similaire à celui-ci avec deux étapes, soit adénylation et formation de lien thioester (Lv et al., 2018).

#### 1.2.6 L'enzyme de conjugaison à SUMO : la E2

La seule SUMO E2 connue chez l'humain est UBC9 et celle-ci est strictement nécessaire pour la conjugaison de SUMO et son largage sur différents substrats (Gareau et al., 2010). Cette protéine se retrouve principalement dans le noyau (Pan et al., 2006; Kuo et al., 2009), mais se retrouve également dans le cytoplasme (Bennett et al., 2012). La SUMO E2 contient une cystéine conservée dans son site actif, permettant la formation du thioester E2-SUMO et utilise l'énergie emmagasinée dans le lien thioester pour le transfert de SUMO vers la lysine du substrat (Varejao et al., 2021). Grâce à des données de calorimétrie par titration isothermique, il a été démontré précédemment que l'interaction E2-SUMO possède une constante de dissociation ( $K_d$ ) de  $2.1 \pm 0.3 \mu\text{M}$  pour SUMO (Cappadocia et al., 2015). Pour la SUMOylation, les substrats peuvent interagir directement avec la E2 pour se faire SUMOyler (Figure 1.4) (Brackett et al., 2020; Villamor et al., 2002). Ces sites SUMOylables semblent jouer un rôle critique dans la régulation entre l'enzyme E2 et le substrat. Les motifs consensus adoptent une conformation étendue avec la lysine acceptatrice qui rentre dans une poche hydrophobe (Yau et al., 2021). Suivant cette insertion, des interactions électrostatiques et des ponts hydrogènes entre la E2 et le substrat, de part et d'autre de la lysine, facilitent la reconnaissance du motif consensus (Gareau et al., 2010). De plus, les résidus se retrouvant sur la E2 positionnent la lysine acceptrice, mais réprime également le  $pK_a$  de cette lysine afin d'augmenter le taux de catalyse. Le motif consensus de SUMOylation adopte une conformation étendue avec l'enzyme E2. Ces interactions vont permettre la SUMOylation de la région intrinsèquement désordonnée du substrat en dehors de son pli globulaire (Gareau et al., 2010). La E2 est capable de reconnaître des substrats possédant des PDSM dues à l'interaction de ces phospho-serine interagissent avec des résidus basiques. Également, les NDSMs ont un mécanisme similaire que les PDSM : les NDSM contiennent des charges négatives en aval de la

lysine de SUMOylation. Ceci permettrait de lier des lysines se retrouvant sur la E2 permettant ainsi d'avoir des lysines dans un site au-delà du site consensus d'être SUMOylé (Yang et al., 2006). Un exemple de lysine qui ne se trouve pas dans un site consensus est la lysine 14 (K14) de la E2, dont la SUMOylation est facilitée par l'acide aspartique 127 (D127). Ce deuxième acide amine est proposé de promouvoir la déprotonation de la lysine du substrat, augmentant ainsi sa nucléophilie permettant sa SUMOylation (Goffinont et al., 2023). Il est certainement possible d'avoir uniquement la E1 et la E2 pour SUMOyler des substrats, mais la E3, même si non nécessaire, permet d'accélérer le transfert de SUMO de la E2, spécifiquement vers le substrat (Goffinont et al., 2023). En ce qui concerne la SUMOylation de substrat, il existe une panoplie de protéines pouvant se faire SUMOyler par la E2 avec différentes constantes d'affinités ayant des  $K_d$  autour de 1-25  $\mu\text{M}$  (Francois-Moutal et al., 2018; Knipscheer et al., 2008).

### 1.2.7 Les SUMO E3 Ligases

Les ligases E3 sont des protéines qui relient le processus de SUMOylation à ses substrats, et elles jouent un rôle crucial dans ce processus. Leur importance est due à leur spécificité : la SUMOylation est spécifique à certains substrats durant certains événements cellulaires grâce aux SUMO E3 Ligases (Shi, et al., 2022). Considérant qu'il n'existe qu'une seule E2, celle-ci est incapable de spécifiquement choisir les substrats nucléaire et cytoplasmique qui doivent se faire SUMOyler parmi des milliers de protéines. Pour ces raisons, les SUMO E3 Ligases sont importantes et cruciales pour un fonctionnement adéquat de la cellule et de l'organisme (Gareau et al., 2010). Le rôle des SUMO E3 Ligases est la décharge de SUMO de la E2 vers le substrat à partir de divers mécanismes. Plus précisément, les SUMO E3 Ligases stimulent la décharge de SUMO ~ E2 (~ démontre un lien thioesther) en positionnant SUMO dans une conformation fermée et active (Cappadocia et al., 2015; Varejao et al., 2020). Ceci est autant vrai pour les SUMO E3 Ligases typique que pour les atypiques. A ce jour, il existe seulement 11 SUMO E3 Ligases qui ont été identifiées et caractérisées: PIAS1, PIAS2, PIAS3, PIAS4, CBX4, ZBED1, NSE2, KIAA1586, EGR2, RANBP2 et ZNF451.

Tableau 1.1: SUMO E3 Ligases connues et leurs caractéristiques cellulaires

Nom de protéine	Localisation cellulaire	Pourcentage de lysines SUMOylées	Nombres de SIMs	Preuve d'interactions avec la E2
PIAS1	Noyau, Cytosquelette	28.6 %	7	Immuno-Buvarbage (Driscoll et al., 2010), protéines purifiées <i>in vitro</i> (Robinson et al., 2021), Expression de gène rapporteur (Lee et al., 2003)
PIAS2	Noyau	31.4 %	5	Immuno-Buvarbage (Saul et al., 2015), protéines purifiées <i>in vitro</i> (Schmidt and Müller, 2002), Expression de gène rapporteur (Lee et al., 2003)
PIAS3	Noyau, Cytoplasme	24.2 %	6	Expression de gène rapporteur (Bandyopadhyay et al., 2010)
PIAS4	Noyau	30.3 %	3	Protéines purifiées <i>in vitro</i> (Ryu and Azuma, 2010), Expression de gène rapporteur (Wang et al., 2011)
CBX4	Noyau	26.4 %	3	N/A
ZBED1	Noyau	N/A	6	Protéines purifiées <i>in vitro</i> (Yamashita et al., 2016), Expression de gène rapporteur (Luck et al., 2020)
NSE2	Noyau	26.9 %	2	N/A
KIAA1586	N/A	4.2 %	7	N/A
EGR2	Noyau	N/A	1	Protéines purifiées <i>in vitro</i> (García-Gutiérrez et al., 2011), Expression de gène rapporteur (García-Gutiérrez et al., 2011)
RANBP2	Noyau	8.4 %	18	Identification par MS/MS (Huttlin et al., 2021), Immuno-Buvarbage (Swaminathan et al., 2004), Co-cristallisation (Tatham et al., 2004), protéines purifiées <i>in vitro</i> (Werner, Flotho and Melchior, 2012), Expression de gène rapporteur (Bandyopadhyay et al., 2010)

ZNF451	Noyau	73.8 %	4	Co-cristallisation (Cappadocia et al, 2015), Expression de gène rapporteur (Karvonen et al., 2008)
--------	-------	--------	---	--

Tableau indiquant les caractéristiques connues des SUMO E3 Ligases ayant été caractérisées. La localisation cellulaire a été extraite de UNIPROT (<https://www.uniprot.org/>). Le pourcentage de lysines SUMOylées a été déterminé grâce aux données provenant de Hendricks et al., 2015. Le nombre de SIMs provient de DEEPSUMO ou JASSA et les preuves d'interactions proviennent de BIOGRID (<https://thebiogrid.org/>).

### 1.2.7.1 Les SUMO E3 Ligases typiques

Les SUMO E3 Ligases typiques sont principalement des protéines nucléaires avec quelques protéines cytoplasmiques (Wang et al., 2017). Ces protéines ont un mécanisme d'action qui implique un domaine SP-RING (Siz PIAS-RING) liant la E2 chargée avec SUMO de façon non covalente. Ces protéines agissent donc comme un échafaud (Hochstrasser, 2001). Les protéines comportant un domaine SP-RING incluent principalement les protéines PIAS chez les humains soit PIAS1, PIAS3, PIASx/a/b et PIASy et MMS21. Ces protéines, découvertes durant les années 1990 et fortement conservées à travers les espèces, contiennent un domaine SP-RING ressemblant au domaine RING de l'Ubiquitine (Liu et al., 2015; Yunus et al., 2009). Chez l'ubiquitine, ce domaine lie une paire d'atomes de Zinc en coordonnant principalement avec des résidus cystéines et histidines créant ainsi un domaine globulaire liant la E2. Cependant, le domaine SP-RING ne contient pas les 2 cystéines qui sont conservées chez les protéines RING de E3 (Hochstrasser, 2001). De plus, ce domaine ressemble structurellement aux U-box trouvés dans le système Ubiquitine. Le domaine SP-RING présent chez les protéines PIAS lie la E2 permettant le positionnement de SUMO pour son transfert vers le substrat (Figure 1.5) (Shi et al., 2022; Yunus et al., 2009). Au-delà du domaine SP-RING, les protéines PIAS présentent un SIM au milieu de leur séquence, soit SIM1, qui est conservé. Ce SIM permet une liaison plus forte de la protéine PIAS à SUMO (Minty et al., 2000). Il existe également un SIM 2 chez certains membres de la famille PIAS (PIAS1-4) se situant en C-terminal (PIAS1-3) ou entre le domaine SP-RING et le SIM 1 (PIAS4). Pour PIAS1 et 2, ce deuxième SIM permet la liaison à SUMO1 jouant un rôle dans la formation du complexe E3-E2-SUMO. Pour PIAS4 et y, ce deuxième SIM se lie de façon préférentielle à SUMO3 permettant une meilleure SUMOylation des substrats protéiques (Minty et al., 2000; Lussier-Price et al., 2020; Kaur et al., 2017). La liaison du domaine SP-RING de

PIAS4 avec SUMO3 indique une constante de dissociation de  $81 \mu\text{M} \pm 4.9 \mu\text{M}$  (Kaur et al., 2017). Cette valeur devrait être similaire pour tous les domaines SP-RING considérant leur conservation.



Figure 1.5: Domaine SP-RING de PIAS1 (vert) en interaction avec la E2 (cyan) et SUMO2/3 (orange). La structure a été générée par Alphafold.

### 1.2.7.2 Les SUMO E3 ligases atypiques

Dans le système ubiquitine, les E3 atypiques sont des protéines n'ayant pas un mécanisme d'action qui utilise un domaine RING ou SP-RING. Dans le cas de l'ubiquitine, il existe de nombreuses E3 atypiques dont RNF31, UBR4, ZFP91 (Barnsby-Greer, et al., 2023; Liu et al., 2022; Chen et al., 2023). Ces protéines n'utilisent pas le domaine RING typiquement observé chez les E3, mais plutôt un mécanisme différent : par exemple, RNF31 utilise un domaine RING-between-RING (RBR) avec le domaine RING1 qui transfère l'ubiquitine vers une cystéine sur RING2 pour un transfert subséquent sur un substrat (Smit et al., 2014; Chen et al., 2023).

Parallèlement, les SUMO E3 Ligases atypiques peuvent diminuer la constante de liaison du substrat tout en augmentant le taux de transfert, entraînant ainsi une augmentation de la constante de spécificité (constante de vitesse/constante de liaison). Le relargage de SUMO, ayant lieu grâce

à la E2 se retrouvant dans une conformation active, a été démontré pour la première fois pour la SUMO E3 ligase RANBP2. RANBP2 appartient à une deuxième classe de SUMO E3 ligase qui coordonnent la E2 chargée en utilisant un motif IR1-M-IR2 (Figure 1.6B). Chaque IR constitue un module catalytique qui comprend un SIM qui lie SUMO dans le contexte du SUMO chargé en thioester avec la E2. Cela est suivi d'éléments structurels qui engagent l'interface entre SUMO et E2 avant de s'enrouler autour de la E2. D'autres ligases SUMO E3 ont été proposées, mais leur mécanisme d'action reste méconnu. Certains d'entre eux, tels que PC2 et SLX4, possèdent plusieurs SIM et semblent stimuler la conjugaison SUMO de manière dépendante des SIMs (Guervilly et al., 2015; Yang et Sharrocks, 2010). En plus d'interagir avec SUMO, la SUMO E2 peut interagir avec une deuxième molécule de SUMO via des interactions non covalentes sur la surface dorsale de la E2. Cela permet de former un complexe SUMO — E2 ~ SUMO (le signe — démontre une interaction non covalente). Les interactions E2 – SUMO dans la voie SUMO sont analogues à celles observées pour les complexes E2 – Ubl ou E2 – Ub. L'interaction non covalente E2 – SUMO est importante pour augmenter le taux de formation de chaînes. ZNF451 comprend deux SIM N-terminal prédits, suivis de douze domaines à doigts de zinc C2H2 et d'un motif interagissant avec l'ubiquitine. Cette protéine interagit avec SUMO et les enzymes de conjugaison SUMO et se localise dans les corps nucléaires PML de manière dépendante de SUMO. ZNF451 est ciblé par la modification SUMO sur plusieurs résidus de lysine et il a été démontré que la modification de ZNF451 se produisait *in vitro* de manière dépendante aux SIMs. Un domaine de la protéine ZNF451 en N terminal englobe les SIM en tandem et un motif PLRP intermédiaire catalyse la SUMO E3 ligase avec une activité préférentielle pour SUMO2/3. Des études structurales ont révélé que le SIM1 interagit avec un SUMO situé derrière la E2, tandis que le SIM2 interagit avec le SUMO présent dans le site actif de la E2, ainsi qu'avec le motif PLRP qui se loge à l'interface entre SUMO – E2 ~ SUMO pour établir des contacts avec la E2. Semblable aux autres SUMO E3, le ZNF451 augmente le taux de catalyse en coordonnant la E2 chargée dans une configuration « fermée » activée. Ceci augmenterait l'affinité de liaison pour les substrats (Figure 1.6A). Il est également possible de noter que ZNF451 peut s'automodifier pouvant donc contourner l'exigence de SUMO exogène (Cappadocia et al., 2015). Les constantes de dissociation pour ZNF451 avec la E2 chargée de SUMO2/3 dans son site actif est de  $2.9 \pm 0.2 \mu\text{M}$  tandis que l'interaction de ZNF451 avec la E2 ayant SUMO2/3 dans son site dorsal est de  $1.9 \pm 0.2 \mu\text{M}$ . De plus, l'interaction établit entre SUMO2/3 seule avec ZNF451 présenté une constante de

dissociation de 120-150  $\mu\text{M}$  (Cappadocia et al., 2015). En ce qui concerne RANBP2, cette protéine possède une affinité accrue pour SUMO1 plutôt que SUMO2/3 avec une constante de dissociation de son module M-IR2 avec SUMO1 étant de 1.8  $\mu\text{M}$  et 22  $\mu\text{M}$  avec SUMO2/3 (Namanja et al., 2012).

Pour la protéine SUMO E3 Ligase NSE2, l'activité provient d'une interface entre des motifs SIMs-like et SUMO ainsi qu'un domaine SP-RING qui interagit avec la E2-SUMO similairement au domaine RING (Figure 1.6C). Le contributeur principal pour ces interactions est principalement His202 qui, grâce au Zinc, permet une liaison à SUMO. Ce type d'interaction, qui n'est pas observée chez SIZ1, rend NSE2 une protéine atypique (Varejao et al., 2021). De manière similaire à ZNF451, un SIM situé au C-terminal de NSE2 emploie un mécanisme qui implique l'utilisation de la poche dorsale de la E2, grâce à une liaison avec un SUMO (Varejao et al., 2021).

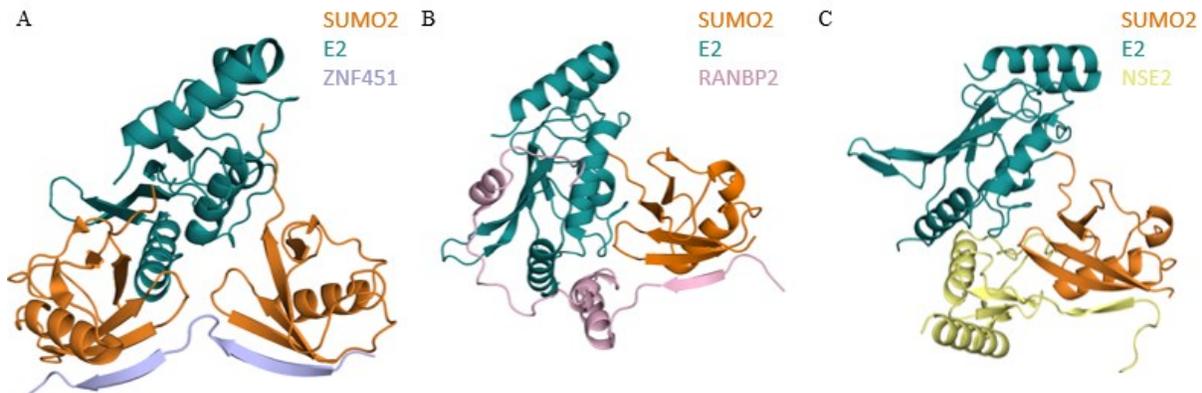


Figure 1.6: Interaction entre les domaines strictement nécessaires pour une activité SUMO E3 Ligase de ZNF451 (A) (PDB 5D2M), RANBP2 (B) (PDB 3UIN) et NSE2 (C) avec la E2 (en cyan) et SUMO2/3 (en orange)

### 1.2.7.3 Évolution et conservation des différents types de E3

En ce qui concerne l'évolution de la machinerie de SUMOylation, cette modification post-traductionnelle se retrouve à la fois chez les levures et les plantes (Park et al., 2011) avec aucune évidence de présence chez les bactéries. La levure, *Saccharomyces cerevisiae*, ainsi que plusieurs organismes tels que *Caenorhabditis elegans* et *Drosophila melanogaster* présentent un seul gène SUMO contrairement aux humains qui en présentent 5 (Celen et al, 2020; Park et al., 2011). Pour les plantes, *Arabidopsis thaliana* est une espèce qui présente 8 isoformes (AtSUM1-8) avec une

version additionnelle sous forme de pseudogène (AtSUMO9) avec principalement les versions 1, 2, 3 et 5 qui sont détectées à présent (Kurepa et al., 2003).

La fonction de toutes les SUMO E3 Ligases est le transfert de SUMO activé provenant de la E2 vers un substrat. Les protéines E3 possédant un domaine SP-RING sont conservées au fil de l'évolution étant présent à la fois chez les plantes, levures, humains et animaux avec une bonne conservation (Garcia-Dominguez et al., 2008; Johnson et al., 2001; Kahyo et al., 2001; Miura et al., 2005; Novatchkova et al., 2012; Park et al., 2010; Strunnikov et al., 2001; Suzuki et al., 2009; Takahashi et al., 2001). Contrairement à la spécificité de l'ubiquitination, les SUMO E3 de plantes ainsi que levures présentent au moins une centaine de substrats (Rytz et al., 2018; Reindle et al., 2006).

Il existe des exceptions au niveau de l'évolution avec la SUMO E3 de *Chlamydomonas* MMS21 qui présente peu de conservation chez les résidus de cystéine et d'histidine coordonnant le Zinc (Figure 1.7). Ces résidus semblent cependant être conservés chez les plantes, les humains et la levure. De plus, aucune ligase E3 de type PIAS n'a été identifiée dans le génome de *Chlamydomonas* tandis que cette protéine est présente à la fois chez les plantes, animaux, les levures et les humains (Lin et al., 2020).

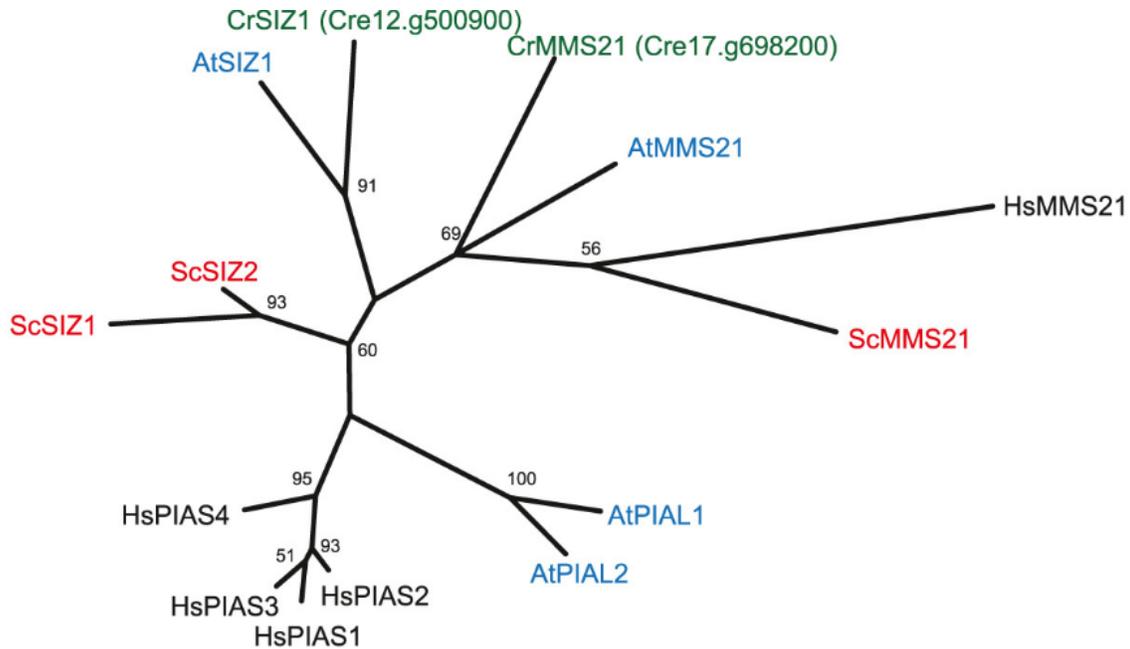


Figure 1.7: Arbre phylogénétique de vraisemblance maximale des SUMO E3 Ligases provenant de *Chlamydomonas reinhardtii* (Cr), *Arabidopsis thaliana* (At), *Homo sapiens* (Hs), *Saccharomyces cerevisiae* (Sc). Le pourcentage de conservation est présent sur chaque branche.

#### 1.2.7.4 Modification des E3

Les SUMO E3 Ligases sont sujettes à de nombreuses modifications dont la SUMOylation, phosphorylation, ubiquitination acétylation et autres MPT (Gareau et al., 2010). Pour commencer, la phosphorylation est une modification étant reliée à la SUMOylation tant au niveau des enzymes de SUMOylation que les substrats (Gareau et al., 2010). Il existe également des SIMs qui, seulement lorsque SUMOylée peut contribuer à la liaison de manière adéquate à SUMO. Ces SIMs sont connus sous le nom de phopsho-SIMs (voir section 1.2.2) (Gareau et al., 2010). Une protéine SUMO E3 Ligase, soit Pc2, semble régulé par la phosphorylation avec une localisation subcellulaire dépendante de sa phosphorylation par HIPK2 (Roscic et al., 2006). De plus, l'activité SUMO E3 Ligase de Pc2 semble fortement relié à sa phosphorylation par HIPK2 (Roscic et al., 2006). Ce phénomène a déjà été reporté auparavant pour l'ubiquitination (Gao et al., 2005).

En ce qui concerne l'autoSUMOylation, il semble que les ligases SUMO E3 actives s'autoSUMOylent. Cette automodification intrinsèque est encore un phénomène peu compris et il existe peu de littérature sur ce sujet. Cependant, l'hypothèse existante est que, comme l'ubiquitine, l'automodification des SUMO E3 Ligases permettrait de moduler son activité (Ihara et al., 2005;

Ivanov et al., 2007; Yohan et al., 2013; Ismaïl et al., 2012). Compte tenu de la préférence de RNF4 pour les chaînes de SUMO les ligases SUMO E3 accumulent des chaînes de SUMO par autoSUMOylation. Alternativement, les ligases SUMO E3 pourraient s'autoSUMOyler en plusieurs sites comme indiquer par ces données : les lysines 40, 46, 56 et 58 de PIAS1 ; les lysines 430, 443, 452, 464 et 489 de PIAS2 ; les lysines 46 et 56 de PIAS3 ; les lysines 490, 500, 508, 522, 532, 537 et plusieurs autres de ZNF451 ; les lysines 30, 41, 47, 65 et 70 de NSMCE2 ; ainsi que les lysines 59 et 69, 128 et 135 de PIAS4 ; et les lysines 1596 et 1605, 2513 et 2531, et 2571 et 2592 de RanBP2 (Kumar et al., 2017).

#### 1.2.7.5 Spécificité des E3

Les SUMO E3 Ligases permettent le rapprochement entre la E2~SUMO et le substrat qui doit se faire SUMOyler. La SUMOylation des substrats se fait en dirigeant la E2~SUMO vers le substrat par des interactions protéines-protéines (Gareau et al., 2010). Parmi les SUMO E3 Ligases découvertes à ce jour, certaines ont fait l'objet d'un criblage visant à identifier leurs substrats respectifs.

Parmi ces ligases avec des substrats connus, les protéines PIAS SUMOylerent une panoplie de protéines tels que les protéines STATs (Shaheen et al., 2018), la protéine PML, MYC, p53, MDM2, AKT (Rabellino et al., 2017), les protéines SMADs (Lee et al., 2003 c; Liang et al., 2004; Lin et al., 2003a; Lin et al., 2003b; Long et al., 2004; Ohshima et al., 2003) ainsi que les récepteurs aux androgènes (Nishida et al., 2002) et MBD1 (Lyst et al., 2006). La SUMOylation se fait par le recrutement de substrats par le motif PINIT (Proline-Isoleucine-Asparagine-Isoleucine-Threonine) permettant la SUMOylation de ceux-ci par la E2-SUMO recruter par le domaine SP-RING (Kaur et al., 2017). Pour ce qui en est d'autres protéines SUMO E3 Ligases, RANBP2 permet la SUMOylation de SP100 (Kirsh et al., 2002), de HDAC4 (Kirsh et al., 2002), MDM2 (Miyachi et al., 2002). Il existe peu d'informations sur les substrats de ZNF451, mais une étude a montré que celui-ci favorise la SUMOylation de TWIST2 (Zeng et al., 2021). Pour NSE2, il semble que celui-ci présente peu de substrat avec une SUMOylation principalement de MCD1 (McAleenan et al., 2012). Pour CBX4, membre du groupe protéique polycomb, celui-ci SUMOyle plusieurs protéines, soit BMI1 (Ismaïl et al., 2012), CtIP (Soria-Bretones et al., 2017), HIF-1a (Li et al., 2014), la télomérase hTERT (Sanyal et al., 2020), Prdm16 (Chen et al., 2018) et EZH2 (Wu et al., 2022) et

plusieurs autres. Son activité provient principalement d'une région qui, lorsque liée à l'ADN, induit un changement de conformation qui augmente son activité SUMO E3 Ligase (Varejão et al., 2018). Pour les protéines EGR2 et ZBED1, il semblerait que ceux-ci SUMOylent principalement les protéines NAB1/NAB2 (Garcia-Gutierrez et al., 2011) et le facteur CHD3/Mi2-a lié à l'ADN (Yamashita et al., 2016).

Grâce à cette liste non exhaustive de substrats protéiques susceptibles d'être SUMOylés par les PIAS, CBX4, entre autres, il est évident que de nombreuses protéines subissent la SUMOylation. Cependant, on estime à près de 4000 le nombre de substrats protéiques qui peuvent être SUMOylés (Hendriks et al., 2016) suggérant qu'il existe possiblement d'autres SUMO E3 Ligases.

#### 1.2.7.6 Interaction avec SUMO

Parmi les régions de protéines pouvant interagir avec SUMO, le motif principal d'interaction est le SIM. Ce motif est présent dans toutes les SUMO E3 ligases identifiées jusqu'à présent (Pichler et al., 2017), et représente la classe principale de liaisons non covalentes de SUMO (Ouyang et al., 2009). Cet élément clé pour la liaison à SUMO est idéal pour la stabilisation de la conformation fermée de la E2 chargée de SUMO dans son site actif permettant d'augmenter la réactivité de SUMO pour son transfert vers un substrat (Varejão et al., 2021).

Il est possible que les différences dans le noyau hydrophobe soient suffisantes pour déterminer si la liaison se produit dans une orientation antiparallèle ou parallèle. Par exemple, un SIM de type  $\psi\text{-}\psi\text{-}x\text{-}\psi$  réagit différemment d'un SIM  $\psi\text{-}\psi\text{-}\psi\text{-}\psi$ . La liaison parallèle par le SIM  $\psi\text{-}\psi\text{-}x\text{-}\psi$  est préférée, car l'orientation parallèle optimise les contacts non polaires entre le noyau SIM et SUMO. Comme mentionné précédemment, il existe des évidences suggérant que le noyau se lie avec une affinité similaire dans les deux orientations, parallèle et antiparallèle. Il semble que dans le complexe antiparallèle, les résidus encadrant le noyau SIM sont largement non structurés, tandis que dans le complexe parallèle, les résidus de part et d'autre de ce motif se retrouvent dans des interactions favorables avec des résidus de SUMO (Yau et al., 2021). Cela implique que l'affinité de liaison et l'orientation de liaison sont fortement influencées par des interactions impliquant des résidus en dehors du noyau hydrophobe du SIM (Yau et al., 2021). D'autres études soulignent l'importance de la distribution des résidus chargés de part et d'autre dans la détermination de

l'orientation de la liaison. Les SIMs qui se lient à SUMO dans une orientation parallèle contiennent souvent une région négativement chargée en C-terminal, tandis que les SIMs qui se lient dans une orientation antiparallèle contiennent souvent une région négativement chargée en N terminal. Dans chaque cas, l'orientation préférée place généralement la région chargée négativement du SIM au-dessus d'une zone conservée de résidus chargés positivement dans SUMO (Gareau et al., 2010; Jardin et al., 2015).

Un exemple de phosphoSIM contribuant à une liaison accrue entre une protéine et SUMO est DAXX qui grâce aux serines phosphorylées autour de son SIM, permet de lier SUMO1 contribuant ainsi à sa fonction proapoptotique. Il semblerait que le phosphoSIM aurait une affinité 30 fois plus forte que celle du SIM (Chang et al., 2011) grâce à des interactions électrostatiques entre le phosphoSIM et un site d'acide aminé positif se retrouvant uniquement chez SUMO1. L'interaction entre le SIM de DAXX et SUMO1 a été mesurée autour de  $8.1 \pm 1.7 \mu\text{M}$  tandis que l'interaction SIM-SUMO2/3 a été reportée de  $8.7 \pm 0.9 \mu\text{M}$  (Escobar-Cabrera et al., 2011). Certaines modifications, comme l'acétylation de lysine de SUMO, diminuent l'affinité SIM-SUMO (Lussier-Price et al., 2020). Ceci est apparent avec les protéines PIAS, qui, par le biais de leurs 2 SIMs ayant des charges négatives en C-terminal des SIMs, ne peuvent lier SUMO ayant des lysines acétylées (Yau et al., 2021). Il a été démontré que certains SIM ont une préférence pour un isoforme de SUMO plus qu'une autre. Ceci n'est pas étonnant considérant que SUMO1 et SUMO2/3 ne partagent que 44 % d'identité de séquence. La possibilité de SIMs spécifique aux isoformes de SUMO a été confirmée par une étude dans laquelle un criblage par phages a été utilisé pour identifier des séquences ayant une affinité différentielle pour SUMO1 et SUMO2 (Hughes et al., 2017).

### 1.2.8 Substrats de SUMOylation

La SUMOylation est principalement un processus nucléaire qui affecte une vaste gamme de protéines et de fonctions. Près de 4000 protéines subissent la SUMOylation. Parmi celles-ci, 66 %, soit les deux tiers du protéome SUMOylé, sont localisées dans le noyau, comparativement à 32 % de l'ensemble du protéome. En parallèle, presque 52 % des protéines ayant été annotées comme étant SUMOylées contiennent un motif consensus de SUMOylation. Lors d'un stress, il est possible de voir seulement ~23 % de sites de SUMOylation ayant un motif consensus exclusif au stress. Il

est également notable que seulement ~14 % des sites soient également ubiquitinés ce qui indiquerait soit la prévalence de la SUMOylation chez les protéines SUMOylées, soit l'utilisation d'un site consensus distinct pour l'ubiquitination (Hendricks et al., 2016).

Afin de pouvoir étudier les protéines SUMO E3 Ligases, il faut étudier l'enzyme et le substrat de cette SUMO E3 Ligase. De plus, idéalement, il faut déterminer le substrat de chaque SUMO E3 Ligase afin de permettre une compréhension optimale des événements cellulaires. Pour ce faire, des méthodes doivent être établies pour permettre l'identification de substrats et SUMO E3 Ligases. Parmi les méthodes, il existe une méthode surnommer SATTs ou SUMO-activated target traps. Cette méthode consiste en la fusion d'une queue SUMO à une protéine E3. Lors de certains événements cellulaires, la protéine recombinante va exercer sa fonction de Ligase et va SUMOyler une protéine avec la queue SUMO permettant l'identification de la protéine substrat par spectrométrie de masse. Grâce à cette méthode, une étude menée par le groupe Gonzalez-Prieto a permis l'identification de plus de 900 protéines substrats de SUMO E3 Ligases (Salas-Lloret et al., 2023).

### 1.2.9 Exploitation d'AlphaFold comme outil de prédiction des interactions SUMO

Grâce à un effort expérimental, les structures d'environ 100 000 protéines uniques ont été déterminées. Cependant, cela représente une petite fraction des milliards de séquences protéiques connues. La biologie structurale est limitée par les années d'effort minutieux nécessaires pour déterminer la structure d'une seule protéine. Prédire la structure tridimensionnelle qu'une protéine adoptera en se basant uniquement sur sa séquence d'acides a été un problème de recherche ouvert important depuis plus de 50 ans. Malgré les progrès récents, les méthodes existantes sont loin d'atteindre une précision atomique, en particulier lorsqu'aucune structure homologue n'est disponible. Cependant, AlphaFold est la première méthode informatique capable de prédire les structures protéiques avec une précision atomique, même dans les cas où aucune structure similaire n'est connue. La dernière version d'AlphaFold repose sur une nouvelle approche d'apprentissage automatique qui intègre des connaissances physiques et biologiques sur la structure des protéines, tirant parti des alignements multiséquences, dans la conception de l'algorithme d'apprentissage en profondeur (Jumper et al., 2021). AlphaFold permet de résoudre une multitude de problèmes complexes, notamment ceux liés à des techniques chronophages qui requièrent beaucoup de temps

et d'expertise, telles que la cristallographie aux rayons X, la spectroscopie RMN (Résonance Magnétique Nucléaire), et la cryomicroscopie électronique. Alphafold permet également de visualiser des régions problématiques qui sont enclines au mouvement tel que les SIMs qui sont difficiles à obtenir par cristallographie aux rayons X. Certaines protéines, en particulier celles de grande taille, sont difficiles à analyser avec les méthodes expérimentales traditionnelles. De plus, ce programme permet de comprendre comment les protéines interagissent entre elles de façon simple sans manipulations qui requiert de l'expertise (Goulet et Cambillau, 2022). Dans ce cas, Alphafold a permis de générer des structures de protéine avec la E2 chargée de SUMO (E2~SUMO) dans son site actif ainsi que la E2 chargée avec SUMO et une SUMO dans son site dorsal (SUMO-E2~SUMO). Ces prédictions de structures permettent de générer plusieurs choses : la structure de protéine n'ayant pas été résolue expérimentalement et les interactions de cette protéine avec la E2 chargée de SUMO. Les modèles de prédiction obtenue, lorsque non couplées à des techniques expérimentales, ne sont pas idéales et peuvent mener à de mauvaises hypothèses. Pour cela, cette technique bio-informatique, lorsque couplée avec des techniques expérimentales de validation, peut être un outil sans équivoque (Diaz-Rovira et al., 2023).

### 1.3 But du projet

#### 1.3.1 Problématique et hypothèses

La SUMOylation, une modification post-traductionnelle cruciale, joue un rôle majeur dans la régulation de divers processus cellulaires. Récemment, une avancée significative a été faite avec la découverte de deux types de SUMO E3 Ligases. La première catégorie fonctionne via un domaine SP-RING, tandis que la seconde opère par des mécanismes différents, pas toujours pleinement caractérisés. Grâce aux outils bio-informatiques courants tels que BLAST, l'analyse des domaines SP-RING est désormais rapide et efficace, facilitant grandement la recherche de protéine présentant ce domaine. Cependant, l'identification des SUMO E3 Ligases qui n'utilisent pas le domaine SP-RING s'avère plus complexe. Le manque de conservation dans leur mécanisme de SUMOylation rend difficile l'observation de tendances claires. Sur les 11 SUMO E3 Ligases identifiés jusqu'à présent, seulement 3 présentent un mécanisme de SUMOylation atypique. Parmi ces 3 protéines atypiques, aucune ne partage le même mécanisme. Ces mécanismes sont l'utilisation de domaine IR1-M-IR2 pour RANBP2, l'utilisation de domaine RING-like pour NSE2 et l'utilisation de 2 motifs SIMs pour ZNF451. ZNF451 se distingue par ses motifs d'interaction SUMO en

tandem (SIM) relié par un motif Pro-Leu-Arg-Pro (PLRP), qui interagit avec une E2~SUMO chargée et une deuxième molécule SUMO. Le mécanisme de cette SUMOylation s'éloigne des ligases traditionnelles basées sur des domaines SP-RING. Malgré l'identification de seulement 11 SUMO E3 Ligases pour environ 4000 substrats humains, il existe probablement d'autres protéines avec une activité E3 Ligase qui n'ont pas encore été découvertes. Ces protéines pourraient utiliser de petits motifs structuraux, comme les SIMs, pour promouvoir la SUMOylation de leurs substrats. Le but de ce projet est donc d'identifier des protéines SUMO E3 Ligases de type ZNF451 ou de type atypique qui pourrait améliorer notre compréhension des protéines SUMO E3 Ligases identifiées ainsi que la SUMOylation. L'hypothèse émise dans cette étude suggère qu'il existe des protéines encore non découvertes jusqu'à présent, possédant une activité de Ligases et utilisant de petits motifs catalytiques pour promouvoir la SUMOylation, ce qui constituerait un mécanisme atypique.

### 1.3.2 Objectifs de recherche

Le projet de recherche présent vise à identifier et caractériser des protéines présentant des caractéristiques atypiques de SUMO E3 Ligases dans le protéome humain. Pour cela, un criblage bio-informatique approfondi a été réalisé, en se basant sur des critères tels que la localisation nucléaire, la présence de multiples sites de SUMOylation, la taille de la protéine, et la présence de séquences SIM. Une fois les candidats identifiés, une analyse bio-informatique détaillée a été menée pour examiner leurs interactions avec l'enzyme E2 et SUMO. De plus, des méthodes telles que Alphafold ont été employées pour prédire les interactions entre les protéines candidates, E2 et SUMO, et pour évaluer la prévalence et possibilité de ces interactions. Un criblage *in bacterio* a été élaboré pour tester la capacité des protéines candidates à s'auto-SUMOyler, utilisant des vecteurs d'expression bactériens spécifiques et des tests d'immunobuvardage. Ce criblage a été complété par des tests de SUMOylation *in vitro*, avec l'utilisation de protéines recombinantes purifiées et de substrats spécifiques, pour confirmer l'activité SUMO E3 Ligase des protéines candidates. Enfin, les protéines candidates seront purifiées et caractérisées biochimiquement, permettant une analyse de leurs mécanismes d'action.

L'objectif global de ce projet est donc de mieux comprendre les mécanismes de SUMOylation et d'identifier de nouvelles protéines potentiellement impliquées dans ce processus complexe et crucial.

## CHAPITRE 2 MATÉRIEL ET MÉTHODES

### 2.1 BIOINFORMATIQUE

#### 2.1.1 Identification de protéines candidates SUMO E3 Ligases par une approche bio-informatique

Pour la partie bio-informatique, la plateforme Jupyter Notebook version 6.4.8 a été utilisée via Anaconda Navigator 2.4.0. Afin de choisir les protéines qui ont une localisation nucléaire, le code1 (Annexe A) a été utilisé pour générer un dictionnaire de localisation de protéines avec *input file* qui contient la localisation du fichier texte du protéome humain d'Uniprot (<https://www.uniprot.org/proteomes/UP000005640>). Par la suite, le code2 (Annexe B) a permis de choisir les protéines basées sur leur localisation et créer une liste pour les protéines nucléaires pour leur utilisation subséquente. Pour la sélection des protéines SUMOylées, les données de protéines ayant des lysines SUMOylées ont été extraites de Hendriks et al. (Hendriks et al., 2016) et ont été stockées dans la variable *Lysinesprotlist*. Grâce à ces données, un croisement a pu être fait entre les protéines nucléaires et les protéines SUMOylées (Code3, Annexe C). Une particularité des données *lysinesprotlist* est le fait que chaque protéine contient autant d'entrées dans la liste qu'elle ne contient de lysines SUMOylée. Cette information peut être utilisée afin de répertorier chaque protéine par rapport au nombre de lysines SUMOylée que celle-ci possède. Premièrement, par une fonction *Counter* de python, le nombre de lysines SUMOylées pour chaque protéine est compté. En second lieu, une recherche sur Uniprot pour la séquence de chaque protéine a lieu pour, à la fois déterminer la taille de chaque protéine et pour déterminer le nombre de lysines total se retrouvant dans chaque séquence. Grâce à ces données, le ratio lysine SUMOylées/lysine totale est déterminée et les protéines sont triées par rapport au ratio ainsi que la taille (Code4, Annexe D). Pour les 161 protéines résultantes, elles sont manuellement recherchées dans les bases de données JASSA et DeepSUMO afin de retrouver les SIMs. Par la suite, une liste *SIMnucprotlist* a été mise au point avec les protéines ayant répondu à toutes ces caractéristiques. Afin de raffiner davantage la recherche, les interacteurs de la E2 (UBC9) ont été générés grâce à la banque de donnée de Biogrid (Code5, Annexe E) et un croisement entre *SIMnucprotlist* et les interacteurs a pu avoir lieu. Pour l'utilisation de ce code, un fichier Biogrid nommée BIOGRID-ORGANISM-Homo\_sapiens-4.4.219.tab3.txt (<https://downloads.thebiogrid.org/File/BioGRID/Release-Archive/BIOGRID-4.4.219/BIOGRID-ORGANISM-4.4.219.tab3.zip>) a été utilisé.

### 2.1.2 Génération de modèles d'interactions Alphafold

Afin de générer tous les modèles tridimensionnels par Alphafold, il est possible d'utiliser une version disponible sur un serveur Google soit colabfold (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>).

Ce site contient la dernière version d'Alphafold (version 1.5.5 en Avril 2024) qui permet de générer des modèles de protéines proches de 1200 acides aminés. Cette version d'Alphafold ne requiert que la séquence à être modélisé et Alphafold devrait être capable de générer les fichiers requis pour la modélisation, soit les fichiers d'alignement de séquences. La séquence de la protéine qui veut être modélisée est insérée dans « query\_sequence » avec les différents domaines séparés par un « : ». Dans le cas de ce mémoire, SUMO2/3 et la E2 sont insérés en premier suivie du candidat qui doit être modéliser. Les autres variables restent intouchées (template\_mode=none, pair\_mode=unpaired\_paired, model\_type=auto, num\_recycles=3, recycle\_early\_stop\_tolerance=auto, relax\_max\_iterations=200, pairing\_strategy=greedy). Une fois les prédictions lancées et terminées, il est possible de télécharger les fichiers grâce à l'onglet « Package and download results » qui permet de télécharger les résultats obtenus d'Alphafold.

## 2.2 BIOLOGIE MOLÉCULAIRE, EXPRESSION ET PURIFICATION DE PROTÉINES

### 2.2.1 Biologie moléculaire et clonage de candidats

Tableau 2.1: Amorces utilisées lors de PCR pour l'amplification de candidats

Gène	Vecteur	Amorce Sens	Amorce Antisens
		5'-	5'-
ZNF24	pBAD	CATCATCATCATC	CCCTGGAAATAC
		ACAGCAGCGGTA	AAGTTTTTCGGTAC
		CCATGAGCGCAC	CCACCTTCACAAC
		AGAGCGTTGAAG	ATTCAGCAGTTTC
		AAG -3'	-3'
		5'-	5'-
	pRSF-TRX	GGCCGGTACCGA	CAAGCTTGTCGAC
		AAACT	GGAGCTCG

		TGTATTTCCAG GGATCC ATGAGCGCACAG AGCGT TGAAGAAG -3'	AATTCGGATCCTC ACACCTTCA CAACATTCAGCA GTTTC -3'
	pBAD	5'- CATCATCATCATC ACAGCAGCGGTA CCATGGCCAGCA ACGTTACCAACA AGACAG -3'	5'- CCCTGGAAATAC AAGTTTTTCGGTAC CAGAGTCATCCTC GCCATTGGCGCTG TCTC -3'
HNRPC	pRSF-TRX	5'- CCGAAAACCTTGT ATTTCCAGGGATC CATGGCCAGCAA CGTTACCAAC AAG -3'	5'- CAAGCTTGTCGAC GGAGCTCGAATT CGGATCCTCAAG AGTCATCCTCGCC ATTG -3'
	pBAD	5'- CATCATCATCATC ACAGCAGCGGTA CCATGGAGATCG GCACCGAGATC AGCCGCAAGATC -3'	5'- CCCTGGAAATAC AAGTTTTTCGGTAC CTTCGCTGGTTTG AGGTCGAATCCA TTTCAAGG -3'
ZC3HE	pRSF-TRX	5'- CCGAAAACCTTGT ATTTCCAGGGATC CATGGAGATCGG CACCGAGATC AGCCGCAAGATC -3'	5'- CAAGCTTGTCGAC GGAGCTCGAATT CGGATCCTCATTC GCTGGTTTGAGGT CGAAT -3'

		5'-	5'-
		CATCATCATCATC	CCCTGGAAATAC
		ACAGCAGCGGTA	AAGTTTTTC
BLM	pBAD	CCATGGCT	GGTACCTGAGAA
		GCTGTTCTCAAA	TGCATATGAAGG
		ATAATCTACAGG -	CTTAAGAAACGG
		3'	-3'
		5'-	5'-
		CATCATCATCATC	CCCTGGAAATAC
		ACAGCAGCGGTA	AAGTTTTTCGGTAC
PML	pBAD	CCATGGAGCCTG	CAATTAGAAAGG
		CACCCGCCCGATC	GGTGGGGGTAGC
		TCCGAGG -3'	CCCAGGAG -3'
		5'-	5'-
		CATCATCATCATC	CCCTGGAAATAC
		ACAGCAGCGGTA	AAGTTTTTCGGTAC
	pBAD	CCATGGTTGAAG	CGTATCTGCTTCT
		CAGATCGCC	GCCTCCCCCTCTA
		CAGGAAAGCTC -	TCAGATC -3'
RBMX		3'	
		5'-	5'-
		CCGAAAACCTTGT	CAAGCTTGTCGAC
	pRSF-TRX	ATTTCCAGGGATC	GGAGCTCGAATT
		CATGGTTGAAGC	CGGATCCTCAGTA
		AGATCGCCCAGG	TCTGCTTCTGCCT
		AAAGCTC -3'	CCCCCTC -3'
		5'-	5'-
		CATCATCATCATC	CCCTGGAAATAC
		ACAGCAGCGGTA	AAGTTTTTC
IF5A1	pBAD	CCATGGCAG	GGTACCTTTTGCC
		ATGACTTGGACTT	

		CGAGACAGGAG -	ATGGCCTTGATTG
		3'	CAACAGC -3'
		5'-	5'-
	pRSF-TRX	CCGAAAACCTTGT	CAAGCTTGTCGAC
		ATTTCCAGGGATC	GGAGCTCGAATT
		CATGGCAGATGA	CGGATCCTCATTT
		CTTGGACTTCGAG	TGCCATGGCCTTG
		ACAGGAG -3'	ATTGC -3'
		5'-	5'-
	pBAD	CATCATCATCATC	CCCTGGAAATAC
		ACAGCAGCGGTA	AAGTTTTTC
		CCATGGAGTTGG	GGTACCCAATATT
		GGAGTTGCCTGG	ATATCTAAGTTGC
		AGGGCGGG -3'	ACAGAAGACC -3'
TYDP2		5'-	5'-
	pRSF-TRX	CCGAAAACCTTGT	CAAGCTTGTCGAC
		ATTTCCAGGGATC	GGAGCTCGAATT
		CATGGAGTTGGG	CGGATCCTCACA
		GAGTTGCCTGGA	ATATTATATCTAA
		GGG -3'	GTTGCACAG -3'
		5'-	5'-
	pBAD	CATCATCATCATC	CCCTGGAAATAC
		ACAGCAGCGGTA	AAGTTTTTC
		CCATGGC	GGTACC
		GGAGCCTGTGGG	GCCCCAGACCTC
		GAAGCGGGGCCG	AATGAGGTCCCC
		CTGG -3'	AGATTCC -3'
		5'-	5'-
RanBP2 (2631-2771)	pBAD	CATCATCATCATC	CCCTGGAAATAC
		ACAGCAGCGGTA	AAGTTTTTCGGTAC
		CCGATGTTCTCAT	CTATTTCTTCTGT

		TGTATATGAACTA ACTCC -3'	CTGAGATTTTTGA GC -3'
		5'-	5'-
	pRSF-TRX	CCGAAAACCTTGT ATTTCCAGGGATC CGATGTTCTCATT GTATATGAACTA ACTCC -3'	CAAGCTTGTCGAC GGAGCTCGAATT CGGATCCTCATAT TTCTTCTGTCTGA GATTTTT -3'
ZNF451 (1-525)	pBAD	5'- CATCATCATCATC ACAGCAGCGGTA CCATGGG AGACCCGGGGTC GGAGATAATAGA ATC -3'	5'- CCCTGGAAATAC AAGTTTTTC GGTACCATGTGCC CCTCCGTGAATCC GGCTCATGTG -3'
		5'-	5'-
	pBAD	CATCATCATCATC ACAGCAGCGGTA CCGATGAAAATG AAGACGACATTC AG -3'	CCCTGGAAATAC AAGTTTTTC GGTACCAGGCTCT TCATCATCACTGC TGACCAGATC-3'
ZNF451 (25-56)		5'-	5'-
	pRSF-TRX	CCGAAAACCTTGT ATTTCCAGGGATC CGATGAAAATGA AGACGACATTCA G -3'	CAAGCTTGTCGAC GGAGCTCGAATT CGGATCCTCAAG GCTTTCATCATC ACTGCTG -3'
		5'-	5'-
ZNF451 (1-56)	pBAD	CATCATCATCATC ACAGCAGCGGTA CCATGGG	CCCTGGAAATAC AAGTTTTTC GGTACCAGGCTCT

		AGACCCGGGGTC GGAGATAATAGA ATC -3'	TCATCATCACTGC TGACCAGATC -3'
	pRSF-TRX	5'- CCGAAAACCTTGT ATTTCCAGGGATC CATGGG AGACCCGGGGTC GGAG -3'	5'- CAAGCTTGTTCGAC GGAGCTCGAATT CGGATCCTCAAG GCTCTTCATCATC ACTGCTG -3'
SLX4	pBAD	5'- CATCATCATCATC ACAGCAGCGGTA CCATGTTTTCTT CAGGTGCTTGGA TGAAGC -3'	5'- CCCTGGAAATAC AAGTTTTTCGGTAC CGTTCCGCTCCAC CTTCTTCTTGCCC CGAGG -3'
DAXX	pBAD	5'- CATCATCATCATC ACAGCAGCGGTA CCATGGCCACCG CTAACAGCATCAT CGTGCTGG -3'	5'- CCCTGGAAATAC AAGTTTTTC GGTACCATCAGA GTCTGAGAGCAC GATGATCTCTTCT G -3'
UQ	pBAD	5'- CATCATCATCATC ACAGCAGCGGTA CCATGCAGATCTT CGTGAAGAC -3'	5'- CCCTGGAAATAC AAGTTTTTC GGTACCTCACCCA CCTCTGAGACGG AG -3'
SUMO1	pBAD	5'- CATCATCATCATC ACAGCAGCGGTA	5'- CCCTGGAAATAC AAGTTTTTC

		CCATGTCTGAC	GGTACCCTAAACT
		CAGGAGGC -3'	GTTGAATGACCCC
			CC -3'
		5'-	5'-
		GCTAACAGGAGG	CCGCCAAAACAG
Substrat de	pBAD	AATTAACCATGG	CC AAGCTTTTAA
SUMOylation		GCAGCAGCCATC	CCTCCTTCGGCCT
		ATCATCATCATCA	TAATGCCGC -3'
		CAGC -3'	

Séquence en nucléotide du substrat de SUMOylation produite par Integrated DNA Technologies (IDT) dans le but d'insérer dans pBAD : 5'-CCATGGGCAGCAGCCATCATCATCATCACAGCAGCGGTACCGAAAACCTTGTATTCCAGGGATCCTATCCGTATGATGTGCCGATTATGCGGGCGAATTCATGGGCGCGGATGGCGGCCCGATTAAGGCAGAGGGTGGCATTAAAGCGGAAGGCGGCATTAAGGCCGAAGGAGGTAAAAGCTTGAATTCACCCTCAAGAGTGAAGACCTGCTGAAAAGCGA GGATTCACTGAAGAGTGAGGATTGAAAGCTT -3'

Les amorces de séquençage Sanger ont été choisies en fonction des vecteurs et gènes :

- pour le plasmide pBAD, 5' -TGCCATAGCATT TTTTATCC-3' (sens) et 5' -GATTTAATCTGTATCAGG-3' (antisens)
- pour le plasmide pRSF-TRX, 5' - TAATACGACTCACTATAGGG-3' (sens) et 5' -GGATCTCGACGCTCTCCCT-3' (antisens)
- pour le gène ZC3HE, les amorces n'ont pas été capable de séquencer l'entièreté de la séquence donc l'amorce 5' — CAGAGGTGCTTAACAGCTTAGA-3' a été utilisée pour pRSF-TRX et pBAD

## 2.2.2 Extraction d'ARN total des cellules HEK293T et HeLa

L'extraction de l'ARN total a été faite avec le réactif de TRIzol (Invitrogen # 15596026). L'extraction a été faite en utilisant 8 x 10<sup>6</sup> cellules HEK293T ou HeLa en suivant les procédures

établies par le manufacturier pour les cellules humaines. Le culot d'ARN résultants de l'extraction a été solubilisé avec de l'eau sans RNase.

### 2.2.3 Synthèse de l'ADNc à partir de l'ARN total

La synthèse d'ADNc a été faite grâce aux amorces oligo dT sans ancrés. Le kit qui a été utilisé est iScript select cDNA Synthesis Kit (BioRad # 1708896). Pour le kit iScript, 500 ng d'ARN ont été utilisés pour une réaction de 90 minutes à 42 °C suivi de 5 minutes à 85 °C. Pour les réactions de PCR subséquentes, l'ADNc est dilué 10 fois avant son utilisation.

### 2.2.4 Réaction de PCR

Les PCR ont été effectuées avec une première étape de dénaturation à 95 °C pendant 5 minutes suivi de 35 cycles de dénaturation à 95 °C pendant 30 secondes, d'appariement à 55 °C pendant 30 secondes et une extension à 72 °C de 1 minute par 1 kb de taille d'amplicon attendues. Par la suite, une étape d'extension finale a lieu à 72°C pendant 10 minutes. La PCR se fait en rajoutant 2 µM d'amorce sens et 2 µM antisens ainsi que 1 µL de matrice dans un volume final de 50 µL avec une concentration de 200 µM de dNTPs, 1µg/mL de polymérase SSB purifiées. Cette réaction a lieu dans un tampon composé de 50 mM TRIS-HCl pH 8.0, 20 mM KCl, 15 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.2 mM MgSO<sub>4</sub> et 0.14 mg/mL de BSA.

### 2.2.5 Digestion de l'ADN par enzyme de restriction

La digestion se fait dans un volume final de 100 µL avec 50 µL d'ADN (amplicons ou plasmides à digérer), 10 µL de tampon CutSmart (NEB #B6004S), 1 µL de chaque enzyme de restriction et compléter avec de l'eau nanopure. La réaction a lieu à 37 °C pendant au moins 1 heure. Suite à la digestion, une migration a lieu sur gel d'agarose 1 % contenant du SYBR-Safe (Invitrogen # S33102) pendant 20 minutes à 120 V. La révélation du gel se fait dans le chemidoc (BioRad) et une excision peut avoir lieu pour extraire les bandes voulues.

### 2.2.6 Réaction de Gibson

Le volume final de réaction de Gibson est de 10µL et se fait à 50°C pendant 1 h. L'amplicon intact (provenant de la PCR) est incubé avec un plasmide digéré par une ou des enzymes de restriction afin d'avoir une stœchiométrie de 3 amplicons pour 1 plasmide dans un volume de 5µL. Par la

suite, 5 $\mu$ L d'un mix de Gibson 2X constitué de 0.04 U T5 exonucléase (NEB #M0363S), 0.25 U Phusion High-Fidelity DNA Polymerase (ThermoFisher # F-530L), 10 U Taq DNA Ligase (NEB # M0208S) dans 0.2 M Tris-HCl pH 7.5, 20 mM MgCl<sub>2</sub>, 10 % w/v PEG-8000, 20mM DTT, 0.132 % w/v NAD, 0.1 mM dNTPs. Suivant cette incubation, une transformation a lieu avec le produit de Gibson.

### 2.2.7 Préparation de bactéries chimiocompétentes

Les bactéries *E. coli* de type DH5a ou BL21 (DE3) CP ont été utilisées pour toutes les manipulations subséquentes. Les bactéries sontensemencées dans du LB (10 g/L de NaCl, 5 g/L d'extraits de levures et 10 g/L de peptone) a 37 °C avec agitation jusqu'à l'obtention d'une D.O 600 de 0.6, et ensuite centrifuger à vitesse maximale pendant 15 minutes à 4 °C. Les bactéries sont ensuite resuspendues dans du TFB I froid (30 mM KOAc, 50 mM MnCl<sub>2</sub>, 100 mM RbCl, 10 mM CaCl<sub>2</sub>, 15% [v/v] de glycérol). Par la suite, les bactéries sont centrifugées à vitesse maximale pendant 8 minutes à 4°C pour ensuite être resuspendues dans du TFB II froid (10 mM Na MOPS pH 7.0, 75 mM CaCl<sub>2</sub>, 10 mM RbCl, 15% [v/v] glycérol) et aliquoter. L'entreposage des bactéries rendues compétentes se fait à -80 °C.

### 2.2.8 Transformation bactérienne

Les bactéries *E. coli* DH5a ont été utilisées pour la production de plasmide. Les bactéries *E. coli* BL21 (DE3) Codon Plus ont été utilisées pour l'expression de protéines. Pour chaque 10  $\mu$ L de bactéries compétentes qui ont été utilisées, 1  $\mu$ L d'ADN à concentration 60ng/ $\mu$ L a été utilisé. En premier lieu, les bactéries sont incubées avec les plasmides sur glace 15 minutes, suivies d'un choc à la chaleur de 45 secondes à 42°C et une incubation de 1 minute sur glace. 300  $\mu$ L de milieu LB (10 g/L de NaCl, 5 g/L d'extraits de levures et 10 g/L de peptone) ont été ajoutés en condition stérile et les bactéries ont été incubées à 37 °C pendant 30 minutes. Par la suite, l'entièreté du volume de LB est rajoutée à un pétri contenant du LB-Agar-Antibiotique (10 g/L de NaCl, 5 g/L d'extraits de levures, 10 g/L de peptone et 15 g/L Agar). Pour les antibiotiques, la streptomycine (100 $\mu$ g/mL), la kanamycine (50 $\mu$ g/mL), le chloramphénicol (34 $\mu$ g/mL) et l'ampicilline (100 $\mu$ g/mL) ont été utilisés dépendamment de la résistance du plasmide. Les pétris ont été incubés à 37°C pendant 16 h et utilisés par la suite pour les expériences qui le requiert.

### 2.2.9 Extraction de plasmides par minipreps

Une colonie de bactéries *E. coli* DH5a transformées a été prélevée et cultivée dans 3 mL de milieu de culture LB-Antibiotique pendant 16 h à 200 rpm à 37 °C. Le kit de minipreps (Promega # A1465) a été utilisé pour la purification des plasmides des bactéries en suivant les directives du fabricant. La quantification des plasmides a été faite à l'aide d'un NanoDrop One (ThermoFisher).

### 2.2.10 Extraction sur gel de bandes d'ADN

Pour l'extraction sur gel d'agarose des bandes, à l'aide d'une lame de rasoir, les bandes ont été excisées pour une purification par le kit EZ-10 Spin Column DNA Gel Extraction Kit de BioBasic (BS354) en suivant les directives du fabricant.

### 2.2.11 Expression bactérienne de protéine recombinante

Une colonie de bactéries BL21 (DE3) CP contenant le plasmide d'intérêt est introduite dans 100 mL de LB (10 g/L de NaCl, 5 g/L d'extraits de levures et 10 g/L de peptone) avec les antibiotiques appropriés. Ce milieu est incubé pendant 16 h à 28 °C. Par la suite, 30 mL de ce milieu contenant les bactéries est transféré dans 1 L de milieu SB (5 g/L de NaCl, 20 g/L d'extraits de levures et 32 g/L de peptone à pH 7.5) contenant les antibiotiques appropriés. Le milieu SB contenant les bactéries est incubé à 37 °C jusqu'à l'obtention d'une D.O.<sub>600</sub>=1.0. Par la suite, l'induction a lieu (300 µM IPTG ou 0.2 % L-Arabinose) pendant 16 h à 18 °C. Finalement, le milieu SB contenant les bactéries induites est centrifugé à 10 000 g pendant 40 minutes. Le culot est ensuite solubilisé dans du sucrose 20 % (w/v) avec 1 % (w/v) de lysozyme et congelé à l'aide de l'azote liquide et entreposé à -20 °C jusqu'à son utilisation.

### 2.2.12 Purification de protéines recombinantes

#### 2.2.12.1 Chromatographie d'affinité NiNTA

Les culots bactériens dans le sucrose 20 % sont utilisés pour cette partie. Lors de la décongélation, 500 mM de NaCl, 20 mM d'imidazole, 20 mM de Tris-HCl pH 8.0 et 5 mM de β-Mercaptoéthanol sont ajoutés au culot afin d'avoir un milieu adéquat pour la solubilité protéique lors de la lyse. La lyse se fait par sonication (50 % de cycle de service avec une puissance de 5) 3 x 3 minutes avec

3 minutes de repos entre chaque cycle. Suite à cela, le produit de lysat est centrifugé à 33 000 g pendant 40 minutes. Le surnageant provenant de cette centrifugation contient les protéines solubles et est ce qui sera utilisé pour la chromatographie d'affinité. Le surnageant est passé sur une colonne de 5 mL HisTrap FF Crude (Cytiva #11000458) sur un FPLC Akta Start (Cytiva # 29022094). Une fois le surnageant passé sur la colonne, celle-ci est lavée par le tampon de lavage (500 mM de NaCl, 20 mM d'imidazole, 20 mM de Tris-HCl pH 8.0 et 5 mM de  $\beta$ -Mercaptoethanol) pour 10-15 volumes de colonne. Une fois le lavage achevé, les protéines liées à la colonne sont éluées avec le tampon d'éluat (500 mM de NaCl, 500 mM d'imidazole, 20 mM de Tris-HCl pH 8.0 et 5 mM de  $\beta$ -Mercaptoethanol) et la récolte de l'éluat se fait en incrément de 1 mL.

#### 2.2.12.2 Chromatographie par exclusion stérique

Pour la purification par exclusion stérique, les colonnes HiLoad 16/600 Superdex 200 pg (Cytiva # 28989335) et HiLoad 16/600 Superdex 75 pg (Cytiva # 28989333) sont utilisées sur un FPLC Akta pur (Cytiva # 29018224). Le tampon utilisé pour la chromatographie par exclusion stérique est composé de 150 mM de NaCl, 20 mM de Tris-HCl pH 8.0 et 5 mM de  $\beta$ -Mercaptoethanol. Les colonnes sont choisies en fonction de la taille des protéines à purifier ainsi que la taille des contaminants. Les protéines qui sont injectées dans le système ne dépassent pas 5 % du volume de colonne (entre 1 et 5 mL). La récolte d'échantillons sortants de colonne se fait en incrément de 5 mL.

#### 2.2.12.3 Chromatographie par échange ionique

Pour la purification de protéine par échange ionique, les colonnes HiTrap SP HP Cation Exchange (Cytiva # 17115101) et HiTrap Q HP Anion Exchange (Cytiva #17115301) ont été utilisés sur FPLC Akta start (Cytiva #29022094). Pour avoir une adhésion de protéine sur cette colonne, un tampon à faible force ionique (20 mM de NaCl, 20 mM de Tris-HCl pH 8.0 et 5 mM de  $\beta$ -Mercaptoethanol) est initialement utilisé. Une fois que les protéines se retrouvent liées à la colonne, une augmentation graduelle a lieu avec un tampon à haute force ionique (1 M de NaCl, 20 mM de Tris-HCl pH 8.0 et 5 mM de  $\beta$ -Mercaptoethanol). Cette augmentation se fait en incrément de 2 % de tampon de haute force ionique par mL. La récolte d'échantillons sortants de colonne se fait en incrément de 1 mL.

#### 2.2.12.4 Clivage des étiquettes des protéines recombinantes

La plupart des protéines recombinantes ont été exprimées avec une étiquette se retrouvant en N terminal. Pour se débarrasser de cette étiquette, un clivage a lieu par des enzymes (protéase TEV) qui reconnaissent des séquences consensus à la fin des étiquettes. Pour le clivage, les protéines sont incubées avec ~0.2 mg/mL d'enzyme de clivage pendant 16 h à 4 °C. Pour se débarrasser de l'étiquette et de l'enzyme de clivage, une chromatographie devrait suivre le clivage.

#### 2.2.12.5 Électrophorèse SDS-PAGE sur gel d'acrylamide

La migration des protéines s'est faite à courant constant de 180 V pour des gels à épaisseur 1 mm pendant 55 minutes pour des gels de 10-20 %, 12 % ou 15 %. Pour la détection des protéines, la coloration a été faite par bleu de Coomassie, mais un ajout de TCE (1, 1, 1 — Trichloroéthane) dans les gels est possible pour la visualisation des protéines grâce à une exposition aux rayons UV. La visualisation des protéines sur gel a été faite à l'aide du ChemiDoc (BioRad).

### 2.3 Immunobuvardage

Les anticorps utilisés sont :

- Anticorps primaires: Biolegend IgG1 Anti-HA # 901503
- Anticorps secondaires: Cell Signaling Technology Anti-mouse IgG, HRP-linked Antibody # 7076

Les protéines ont été transférées du gel SDS-PAGE vers une membrane PVDF à l'aide du Trans-Blot Turbo (BioRad) selon les procédures du fabricant. Le transfert a été fait à 2.5 A avec un maximum de 25 V pendant 7 minutes. Le transfert des protéines sur la membrane a été vérifié par détection du 2, 2, 2-trichloroéthanol à l'aide du ChemiDoc (BioRad). Les membranes ont été bloquées grâce à une solution 5 % de lait dans du TBS-T (20 mM Tris-HCl pH 7.5, 140 mM NaCl et 0.3 % Tween-20) pendant 1 h à température pièce. Par la suite, la membrane est lavée 3x5 minutes dans du TBS-T et une incubation a lieu pendant 1 h à température pièce avec l'anticorps primaire Anti-HA dilué 1/1000 dans du TBS-T contenant 0.5 % NaN<sub>3</sub>. Suivant les 3 lavages, la membrane est incubée pendant 1 h à température pièce avec l'anticorps secondaire dilué 1/10000

dans du TBS-T. La révélation des protéines se fait à l'aide du kit Clarity Western ECL Substrate (BioRad # 1705061) et à l'aide du ChemiDoc (BioRad).

## CHAPITRE 3 RÉSULTATS

### 3.1 Criblage bio-informatique

#### 3.1.1 Identification de protéines possédant des caractéristiques associées aux SUMO E3 ligases

Grâce à un criblage bio-informatique, il est possible d'identifier des protéines ayant des caractéristiques associées aux SUMO E3 Ligases. En partant du protéome humain complet tel que retrouvé sur la base de données Uniprot ([www.uniprot.org](http://www.uniprot.org)), nous avons ainsi appliqué plusieurs filtres successifs. La figure 3.1 résume les différents filtres utilisés ainsi que le nombre de candidats retenus à chaque étape.

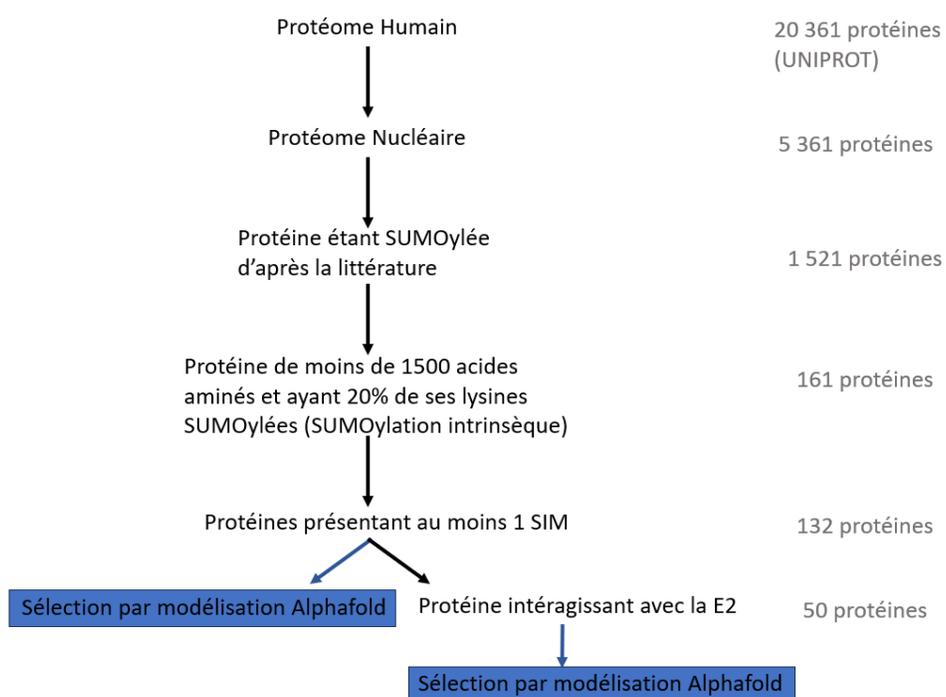


Figure 3.1: Pipeline bio-informatique pour la sélection des candidats SUMO E3 Ligases. Pour chaque critère de sélection, le nombre de protéines présentant la caractéristique respective ainsi que celles satisfaisant aux critères antérieurs est inscrit sur le côté droit de la figure. Chaque flèche sur le schéma symbolise une phase de sélection des candidats. Les carrés bleus désignent les candidats ayant fait l'objet d'une évaluation par le biais de la méthode Alphafold et qui manifestent un potentiel en tant que SUMO E3 Ligases.

Tout d'abord, en considérant que la SUMOylation est un processus nucléaire, nous nous sommes concentrés sur les protéines présentant une localisation nucléaire. Ainsi, partant du protéome humain composé de 20 361 protéines (nombre de protéines répertoriées sur Uniprot au moment de

la recherche, soit février 2022), il a été observé que 26 % de celles-ci se présentent une annotation *nucléaire*, ce qui représente 5361 protéines (Figure 3.1). Il est important de noter ici que les protéines qui ont été choisies comme ayant une annotation *nucléaire* ne sont pas exclusivement nucléaires : les protéines peuvent être à la fois cytoplasmiques et nucléaires et tout de même avoir une annotation *nucléaire*. Les 5 363 protéines identifiées à cette étape correspondent ainsi à des protéines qui présentent minimalement une localisation nucléaire partielle.

Par la suite, étant donné que les SUMO E3 Ligases ont une capacité d'autoSUMOylation élevée (toutes les SUMO E3 Ligases découverts à ce jour sont fortement SUMOylées) (voir 1.2.7.4. Modifications des SUMO E3 Ligases), nous avons identifié les protéines présentant de nombreux sites de SUMOylation à partir d'une banque de données générée par Hendriks et Vertegaal (Hendriks et al., 2016) et qui répertorie l'ensemble des sites de SUMOylation identifiés par spectrométrie de masse à l'échelle du protéome humain. Pour ce critère, nous nous sommes concentrés sur les protéines où au moins 20 % des lysines totales sont SUMOylées à un moment ou à un autre (Figure 3.1). Par exemple, si une protéine possède un total de dix lysines et que trois lysines différentes ont été identifiées comme des sites de SUMOylation, cette protéine serait considérée dans notre analyse comme fortement SUMOylée. Le critère de 20 % de SUMOylation a été établi arbitrairement à la suite d'une visualisation des 1 521 protéines SUMOylées. Parmi les protéines étant SUMOylées, le graphique à la Figure 3.2 semble suggérer deux populations : la première population étant les protéines SUMOylées présentant moins de 20 % de leurs lysines SUMOylées (concentration de plusieurs points), et une deuxième population de protéines présentant au moins 20 % de lysines SUMOylées (points plus dispersés). Un seuil plus élevé de SUMOylation aurait pu être établi, tel que 30 %, mais ceci aurait éliminé certains contrôles internes tels que PIAS1. Pour cette forte SUMOylation, notre hypothèse est en effet que des résidus lysines des SUMO E3 ligases entrent fréquemment en contact avec l'enzyme E2 activée (E2~SUMO) par simple effet de proximité, ce qui entraîne la SUMOylation de nombreux sites au sein de la protéine. Un nombre de sites de SUMOylation élevé constitue ainsi une *signature* d'une activité SUMO E3 ligase. Toutefois, pour faciliter les étapes subséquentes de validation *in cellulo* et *in vitro*, nous n'avons retenu que les protéines dont la taille est inférieure à 1500 acides aminés. Les protéines dont la taille dépasse 1500 acides aminés sont en effet plus à même de présenter des repliements complexes et sont notamment plus difficiles à exprimer de manière recombinante dans des

systèmes procaryotes (Canaves et al., 2004; Goh et al., 2004; Gråslund et al., 2008a). Ainsi, sur un total de 1 521 protéines présentant au moins un site de SUMOylation selon la base de données de Hendriks et Vertegaal, 161 protéines présentent un taux de SUMOylation d'au moins 20 % tout en ayant une taille de moins de 1500 acides aminés ont été identifiées (Figure 3.1). Grâce à une représentation graphique des candidats (Figure 3.2), il est possible d'observer un enrichissement significatif de plusieurs protéines connues pour leur implication dans le processus de SUMOylation. On y retrouve notamment les protéines SUMO1, SUMO2, SUMO3 ainsi que l'enzyme E2 (UBC9). De manière importante, des SUMO E3 ligases connues telles que PIAS1, PIAS2, PIAS4 et ZNF451 se retrouvent parmi les 161 protéines fortement SUMOylées et sont visible sur le graphique du bas de la Figure 3.2.

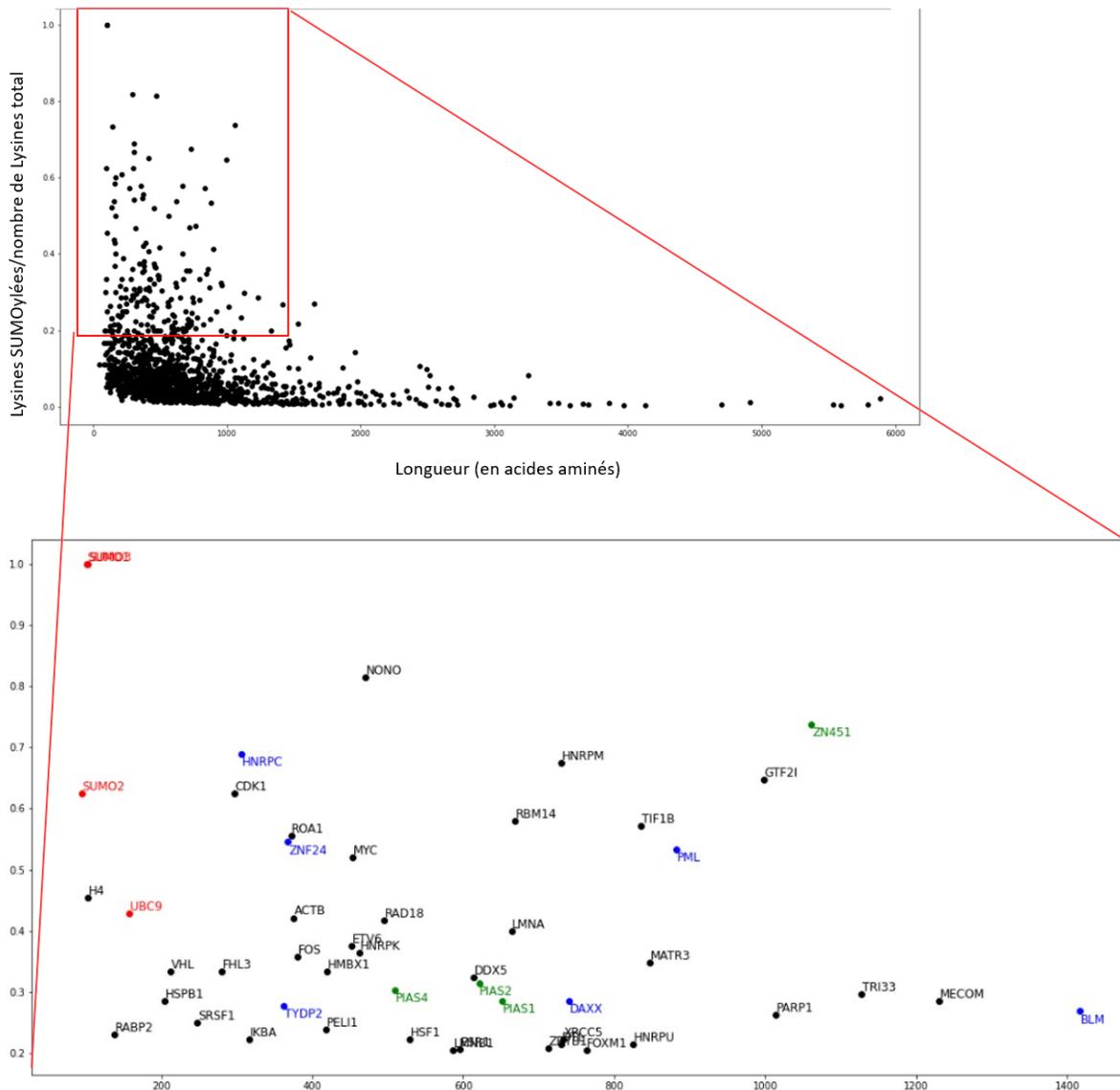


Figure 3.2: Représentation graphique de 1521 protéines présentant une localisation nucléaire et présentant au moins un site de SUMOylation selon la base de données de SUMOylation de Hendriks et Vertegaal (Hendriks et al., 2016). La partie basse met en évidence les protéines de moins de 1500 acides aminés ayant une SUMOylation d’au moins 20 % de leurs résidus lysine. En rouge sont les protéines de la cascade signalétique de SUMOylation, en vert sont les E3 connues et identifiées à ce jour et en bleu sont les candidats choisis pour cette étude.

En considérant que des séquences SIMs sont retrouvées chez toutes les SUMO E3 ligases connues (Tableau 1.1) (Shi et al., 2022) et que plusieurs SUMO E3 ligases en possèdent même plusieurs (Cappadocia et al., 2018), nous nous sommes concentrés sur les protéines présentant au moins une

séquence SIM. Notre hypothèse est que ce type de séquence est cruciale pour permettre à la protéine d'orienter la protéine SUMO au sein du complexe E3-E2~SUMO dans une conformation *active* qui est apte à la catalyse. Pour la sélection de protéines contenant des SIMs, deux programmes de prédiction ont été utilisés, soit JASSA et DEEPSUMO. Ces deux programmes présentent des différences de fonctionnement et d'efficacité, mais se basent tous les deux sur la détection de motifs consensus avec une détection de 295 et 251 SIM identifiés respectivement chez les protéines candidates ayant été retenues jusqu'à ce stade (soit les protéines humaines, nucléaires, étant SUMOylées d'après la littérature et présentant moins de 1500 a.a. et plus de 20 % de lysines SUMOylées). Ces motifs ont été identifiés par les deux programmes après avoir répertorié, aligné et classifié une centaine de motifs SIMs identifiés expérimentalement. Plus précisément, JASSA se base sur 102 motifs non redondants validés expérimentalement provenant de protéines obtenues grâce à la recherche de mots clés « SIM », « SBD » et « SBM » de NCBI allant jusqu'à janvier 2014 (Beauclair et al, 2015). DEEPSUMO se base quant à lui sur une liste non redondante de 137 SIMs auxquels les protéines sont comparées (Zhao et al., 2014). Cependant, les deux programmes sont incapables de choisir précisément les SIMs étant accessibles et ne peuvent faire une discrimination que par la séquence fournie de protéine. Pour la sélection basée sur les SIMs, les paramètres de JASSA sont « All types w/o H&M » qui permet la recherche de 10 motifs consensus courants tout en excluant deux motifs retrouvés plus rarement et « Only interesting results » pour ne retourner que les sites dont le score dépasse un certain seuil. Pour DEEPSUMO, un seuil de 0.08375 a été recommandé par les auteurs du programme pour permettre de retrouver uniquement les résultats les plus pertinents. Les résultats obtenus grâce à ces deux programmes ont été compilés et ont été utilisés pour la suite. Ce critère de sélection a permis de restreindre la sélection à 132 protéines candidates.

Afin de s'assurer que les protéines candidates peuvent transférer SUMO à partir du site actif de l'enzyme E2 vers un substrat, nous nous sommes concentrés sur les protéines présentant des interactions avec l'enzyme E2 ou avec E2~SUMO (Figure 3.1). Nous avons, pour ce faire, utilisé la base de données BioGrid répertoriant les interactions protéines-protéines afin de découvrir lesquelles des 132 protéines candidates peuvent interagir avec l'enzyme E2. À la suite de cette analyse, 50 protéines, dont 42 candidats (et 8 contrôles internes soit les composantes de la machinerie de SUMOylation, SUMO1, SUMO2, SUMO3, UBC9, ainsi que des SUMO E3 ligases

connues, PIAS1, PIAS2, PIAS3, ZNF451), ont été retenues incluant les protéines BLM, DAXX, HNRPC, PML, TYDP2 et ZNF24 (Tableau 3.1, Figure 3.1, 3.2, 3.3, 5.1). Pour chacune des 132 protéines candidates, un criblage d'interaction avec une enzyme E2 chargée avec SUMO2/3 (E2~SUMO2/3) a également été effectué au moyen d'Alphafold. Dans le cadre de cette recherche, nous considérons qu'il y a interaction entre une protéine candidate et la E2 ou SUMO2/3 lorsqu'une valeur du *Predicted Alignment Error (PAE)* de 15 ou moins est obtenu entre la protéine candidate et la E2 ou SUMO2/3. Le PAE permet en effet d'estimer la capacité de deux régions à se stabiliser mutuellement et donc à interagir et un seuil de 10 est généralement admis comme indicateur d'une interaction (Elfmann et al., 2023). Le paramètre PAE de 15 a été utilisé afin de s'assurer de choisir la totalité des protéines interagissant avec la E2 ou SUMO2/3. Le seuil établi à 15 permettrait ainsi de choisir toutes les interactions potentielles entre le candidat et la E2~SUMO.

Quelques interactions identifiées par Alphafold ont été observées durant cette analyse : BLM interagit avec SUMO2/3 (Figure 3.7 B), PML interagit avec la E2 et SUMO2/3 (Figure 3.7 G, H, I), TYDP2 interagit avec la E2 et SUMO2/3 (Figure 3.7 J, K), DAXX interagit avec SUMO2/3 (Figure 3.7 L), HNRPC interagit avec la E2 et SUMO2/3 (Figure 3.7 M, N) et ZNF24 interagit avec SUMO2/3 (Figure 3.7 O, P). La modélisation Alphafold a permis d'identifier plusieurs candidats additionnels parmi les 132 protéines, dû à leur interaction probable et similaire aux candidats proposés ci-dessus avec la E2. Parmi ces 132 protéines, 4 semblent avoir des interactions très probables avec la E2 avec des mécanismes d'action atypiques. Ces protéines sont RBMX, IF5A1, NF2IP et ZC3HE et ne se retrouvaient pas dans la liste des interacteurs de l'enzyme E2 retrouvé sur BioGRID (Tableau 3.2, Figure 3.1, 3.2, 3.3, 5.1). Les 4 candidats qui ont été choisis, soit RBMX, IF5A1, NF2IP et ZC3HE sont parmi une petite liste de protéine SUMOylé qui semble, d'après Alphafold, interagir avec la E2, mais ne se trouvant pas dans les données de BioGrid. Finalement, une dernière protéine a été ajoutée à la liste, soit SLX4. Cette protéine, qui possède plusieurs SIMs, semble être un bon candidat puisqu'une activité SUMO E3 Ligase a été rapportée pour un complexe protéique contenant cette protéine (Guervilly et al., 2015). Cela dit, l'activité SUMO E3 ligase de la protéine seule n'a jamais été confirmée. En ce qui concerne la modélisation par Alphafold de SLX4, cette protéine de plus de 1500 acides aminés n'a pas pu être modélisée en raison de restrictions informatiques, et par conséquent, les données d'interaction d'Alphafold ne sont pas incluses.

Enfin, sur un total de 20 361 protéines composant le protéome humain, notre analyse a permis de retenir 11 candidats. Le diagramme de Venn de la Figure 3.3 illustre les différents critères utilisés ainsi que le nombre de protéines répondant à chacun des critères ou à leur combinaison. Chaque cercle correspond à une caractéristique de protéines et les intersections correspondent aux protéines partageant plusieurs de ces caractéristiques.

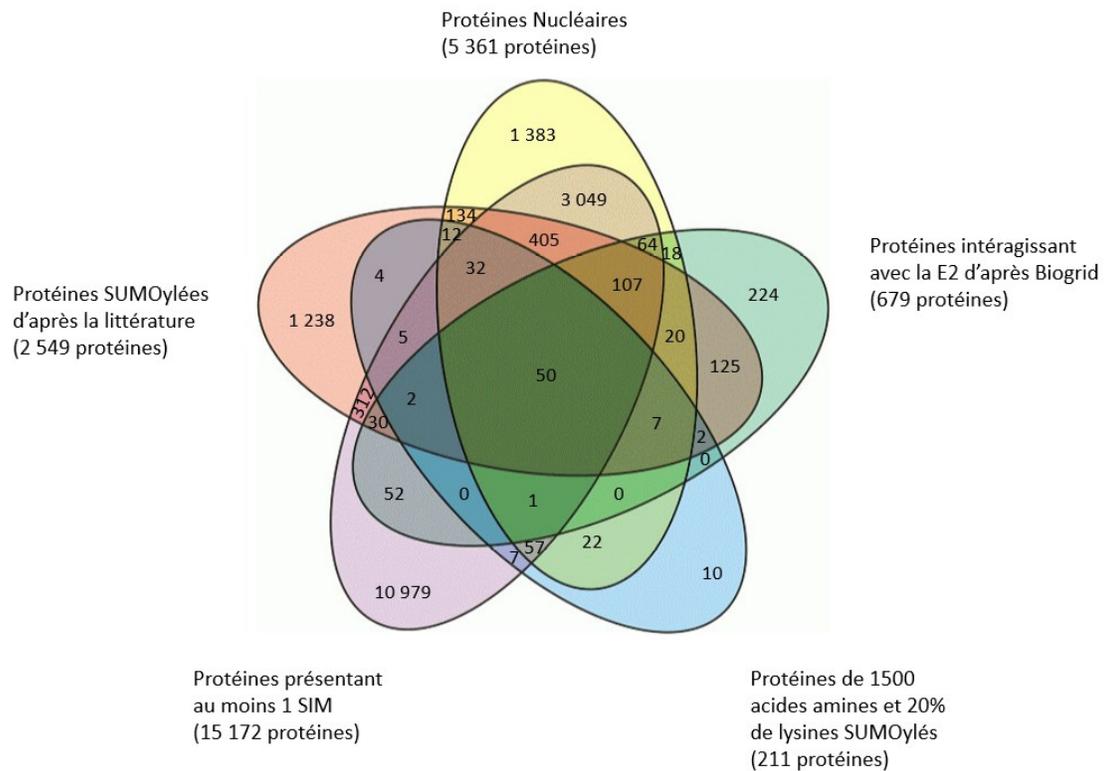


Figure 3.3: Diagramme de Venn illustrant les protéines localisées au noyau, SUMOylées et présentant une interaction avec la protéine E2.

### 3.1.2 Analyse bio-informatique des candidats retenus

De surcroît, notre criblage AlphaFold permet de mettre en lumière le mode d'association des candidats à l'enzyme E2 ainsi qu'à SUMO. La figure 3.4 met ainsi en lumière les sites d'interaction des protéines candidates au niveau de l'enzyme E2. De manière générale, il ne semble pas émerger de tendance claire au niveau de la surface reconnue par la E2. On constate, en effet, que les différentes protéines candidates se lient à différentes régions de l'Enzyme E2 qui sont distantes

dans la structure primaire ainsi que dans la structure tertiaire. Cela dit, un certain enrichissement est observé au niveau de la région N-terminal de la E2 où se lient les protéines NF2IP, BLM, IF5A1 et RBMX. Ceci est intéressant puisque les SUMO E3 ligases de type SP-RING se lient à l'enzyme E2 notamment via l'hélice entre les résidus 1 à 18. Une liaison est aussi observée en aval de la cystéine catalytique dont par les protéines BLM, HNRPC, PML, RBMX et TYDP2.

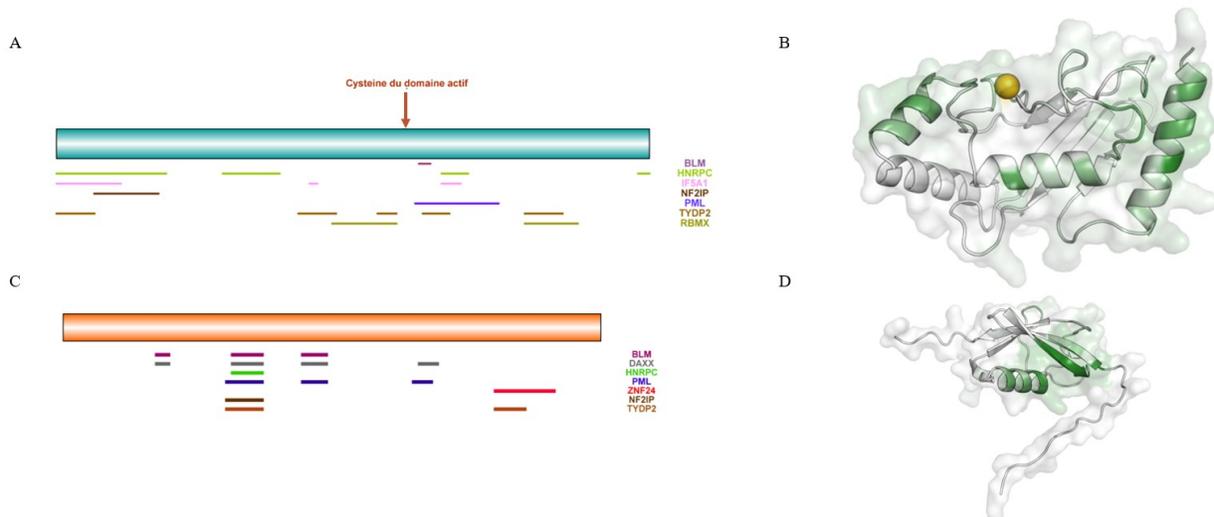


Figure 3.4: Représentation de la structure primaire de la E2 (A) et SUMO2 (C) ainsi que la structure tertiaire (AFDB : AF-P63279-F1) (B) de l'enzyme E2 et de SUMO2 (AFDB : AF-AFP63279F1) (D). L'enzyme E2 est présentée en turquoise tandis que SUMO2/3 est illustré en orange. La couleur verte sur les représentations tridimensionnelle illustre les sites d'interactions observés entre l'enzyme E2 ou SUMO2 et différents candidats lors de notre criblage AlphaFold. La cystéine du site actif de la E2 est représentée avec une sphère jaune sur la représentation tridimensionnelle (B).

La figure 3.5 met en évidence l'interaction des protéines candidates avec les protéines SUMO2/3 et E2 d'après Alphafold. On constate une liaison préférentielle d'une région en N terminal de SUMO2/3 avec 6 candidats parmi les 7 (Figure 3.4 A). Cette interaction entre SUMO2/3 et les candidats est due à un feuillet  $\beta$  de SUMO qui interagit avec les SIMs des candidats (Figure 3.5 A) dans une orientation feuillet  $\beta$  parallèle. Il existe également d'autres régions de SUMO2/3 qui interagit avec les candidats tels que ZNF24 qui vient lier un segment en C terminal de SUMO2/3 (Figure 3.5 A). Pour les interactions entre les candidats et la E2, les interactions les plus notables sont celle représentée à la figure 3.5 B avec la contribution de 2 hélices  $\alpha$  de la E2. Ces interactions entre candidats et E2 ne semblent indiquer aucun mécanisme particulier de la part des candidats envers la machinerie. De plus, il semblerait que les candidats présentent davantage des interactions

à SUMO que la E2. Les principales interactions entre les candidats et SUMO sont des interactions SIM-SUMO.

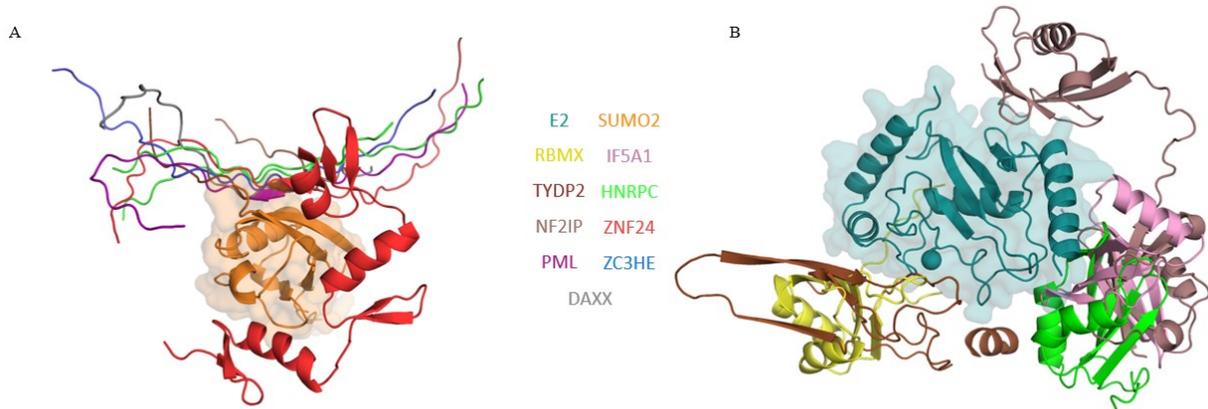


Figure 3.5: Interactions prédites par Alphafold entre les candidats avec SUMO2 (A) et la E2 (B). Les protéines sont colorées avec la SUMO2 étant en orange, la E2 en turquoise, HNRPC en vert, IF5A1 en rose, PML en mauve, TYDP2 en brun foncé, RBMX en jaune, ZNF24 en rouge, ZC3HE en bleu, NF2IP en brun clair et DAXX en gris. La cystéine du site actif de la E2 est représentée par une sphère.

Parallèlement, la structure primaire des candidats retenus est présentée à la figure 3.6 afin de visualiser l'emplacement de leurs différents domaines/régions/motifs, de leurs sites de SUMOylation, ainsi que l'emplacement des motifs SIMs. Cette figure présente également les sites d'interaction avec la machinerie de SUMOylation (soit la E2, soit SUMO) tels qu'identifiés par notre criblage AlphaFold.

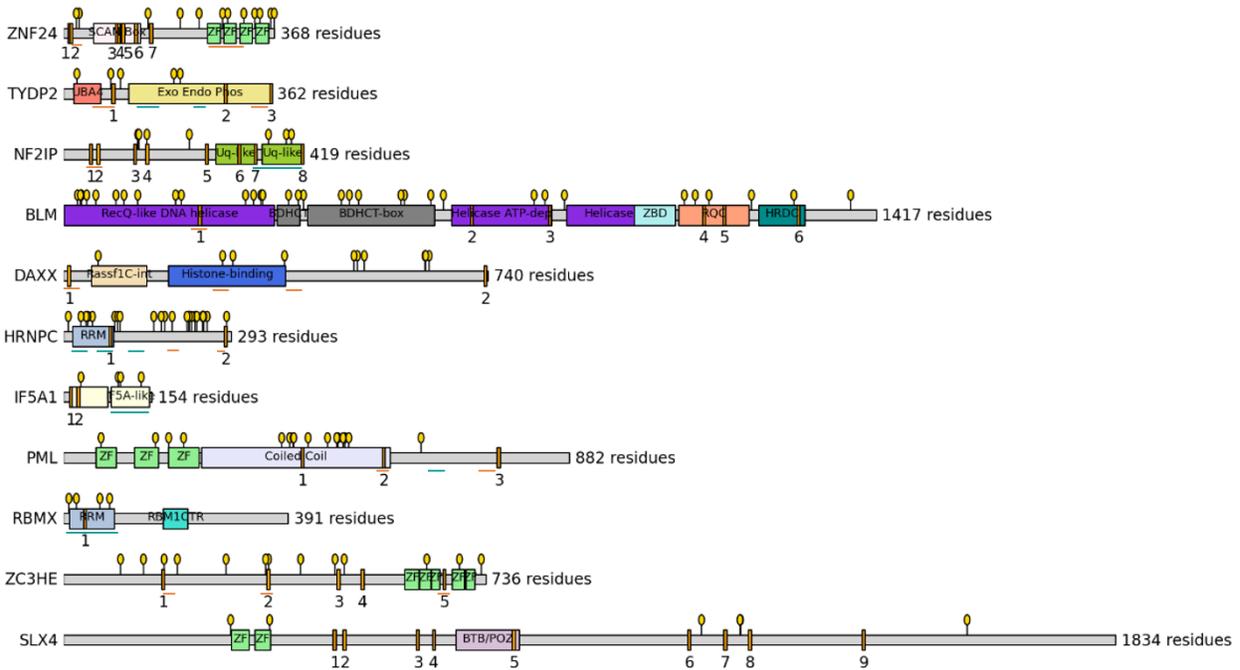
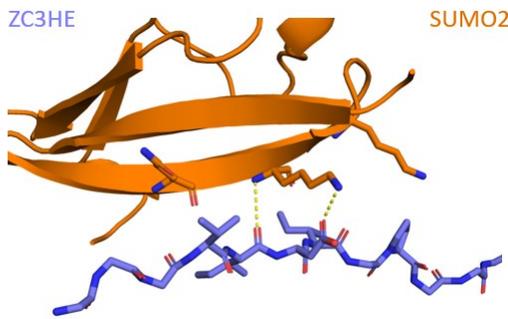
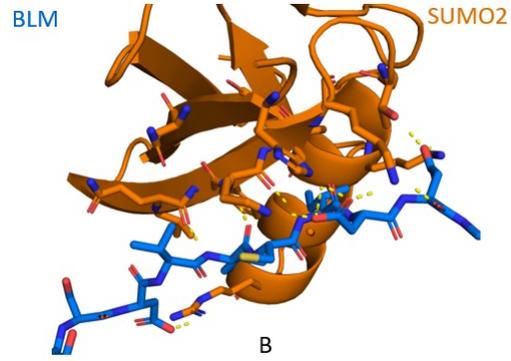


Figure 3.6: Représentation schématique des domaines des candidats retenues et leurs sites de SUMOylation. Les lysines SUMOylées sont représentés grâce à un cercle jaune, les motifs SIMs sont représentés par des rectangles orange numérotés et les domaines protéique sont représentés par des rectangles d'autres couleurs. Une ligne orange sous chacune des protéines illustre les interactions avec SUMO et une ligne de couleur cyan illustre les interactions avec la E2.

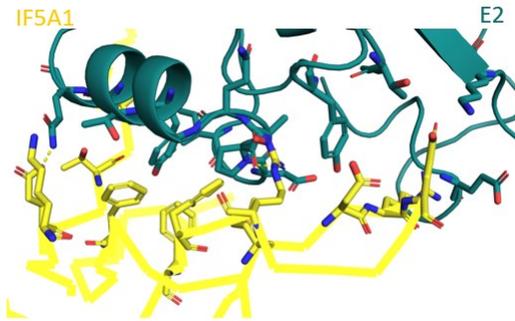
Les figures 3.4, 3.5, 3.6, 3.7 et 3.8 illustrent les protéines ZNF24, PML, NF2IP, DAXX et BLM présentant des régions d'interactions avec SUMO. Les motifs contribuant aux interactions candidats-SUMO se manifestent principalement en tant que SIM (PML, NF2IP, DAXX et BLM). Parmi les interactions de candidats avec SUMO, il semblerait que les domaines Zinc Fingers contribuent à ces interactions telles que visible avec ZNF24 et ZC3HE. Ces protéines contiennent des domaines qui lient SUMO grâce à des motifs hydrophobes se retrouvant dans les domaines Zinc Fingers. Pour l'interaction de TYDP2 et DAXX avec SUMO, cette interaction se retrouve dans un domaine d'interaction avec l'ADN. Ces exemples soulignent l'importance de la SUMOylation et l'intersection de ce processus avec les domaines d'interactions avec l'ADN.



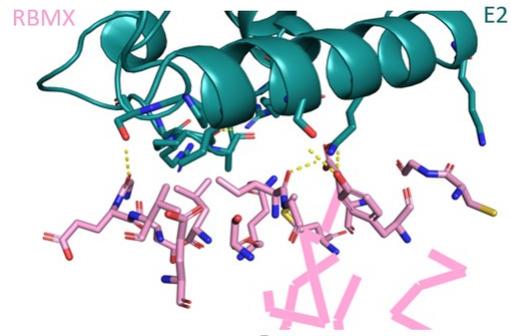
A



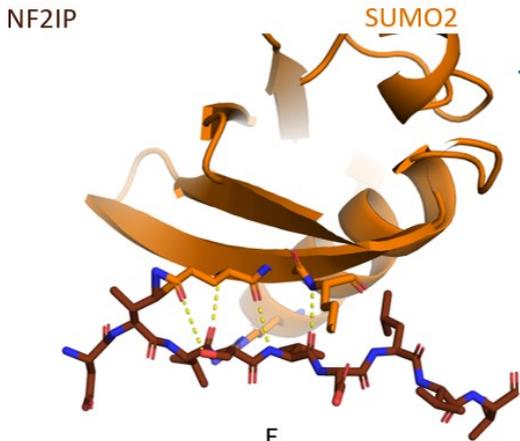
B



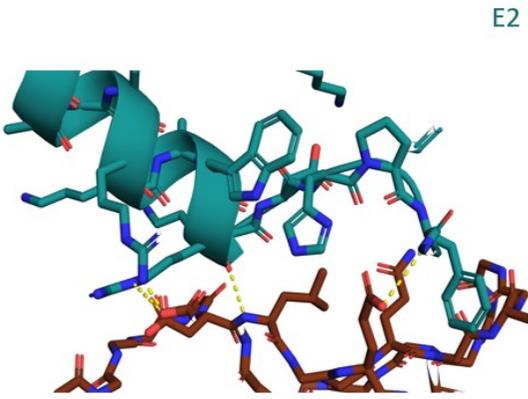
C



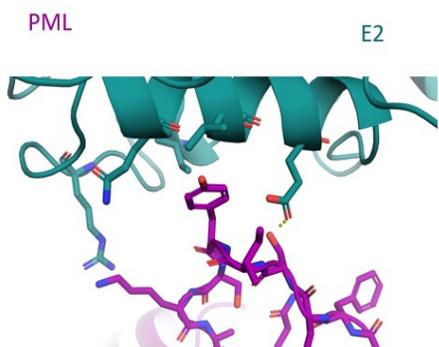
D



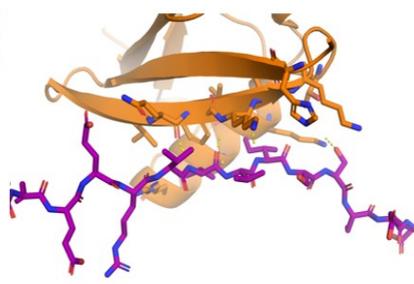
E



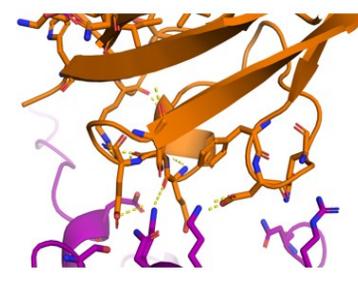
F



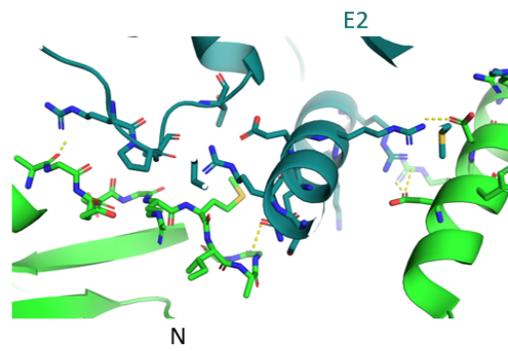
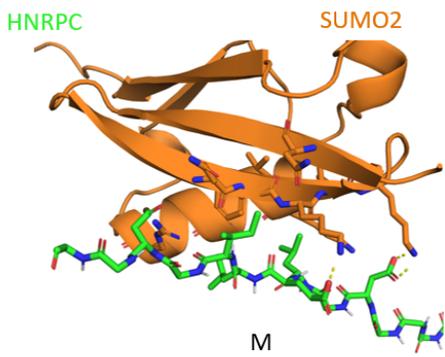
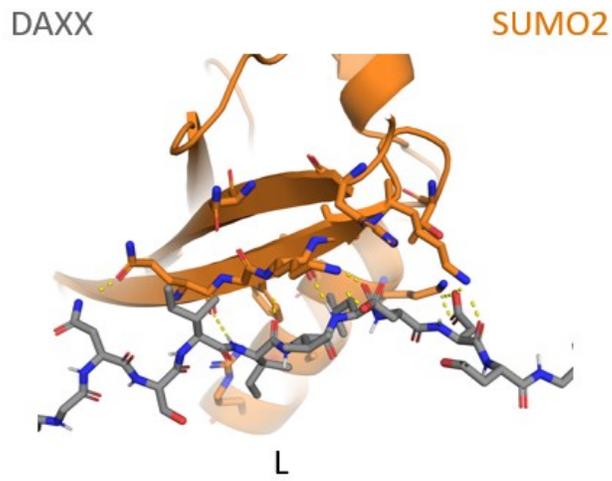
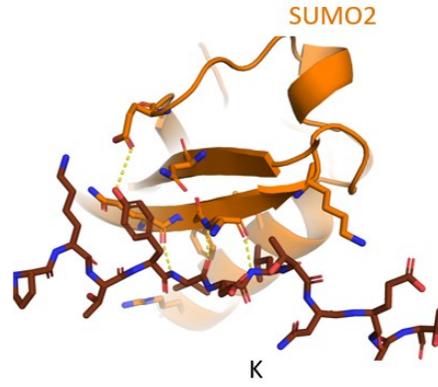
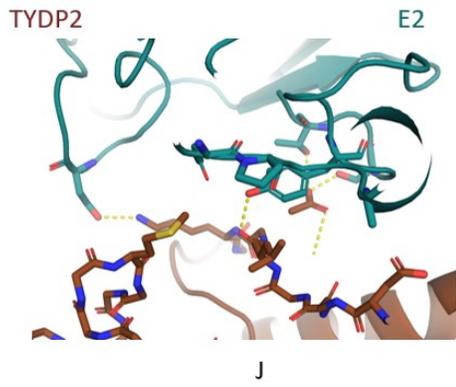
G



H



I



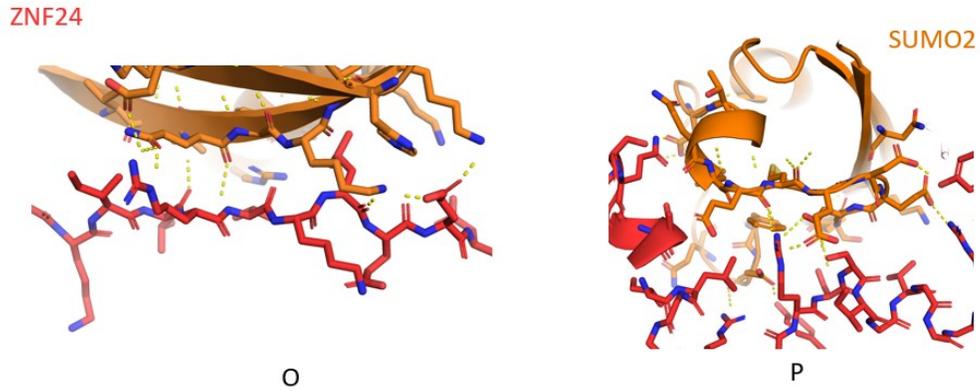
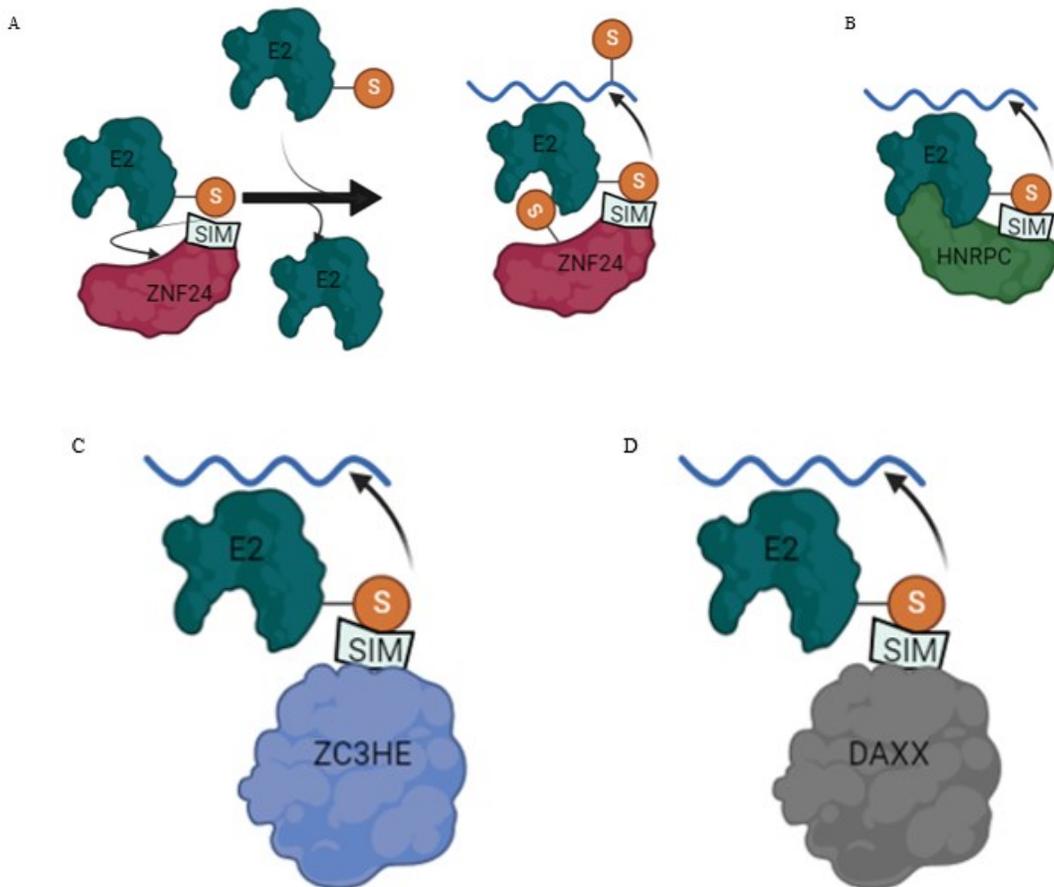


Figure 3.7: Structure tridimensionnelle, prédite au moyen d'AlphaFold, des complexes entre les candidats et la machinerie de SUMOylation. SUMO2/3 est en orange et la E2 est en cyan. Les images représentent toutes les interactions susceptibles de contribuer à une potentielle activité de SUMO E3 Ligase entre le candidat et la E2 ou SUMO2/3. Les ponts hydrogènes sont représentés par des pointillées jaunes tandis que les autres types d'interactions ne sont pas représentés directement sur les figures.



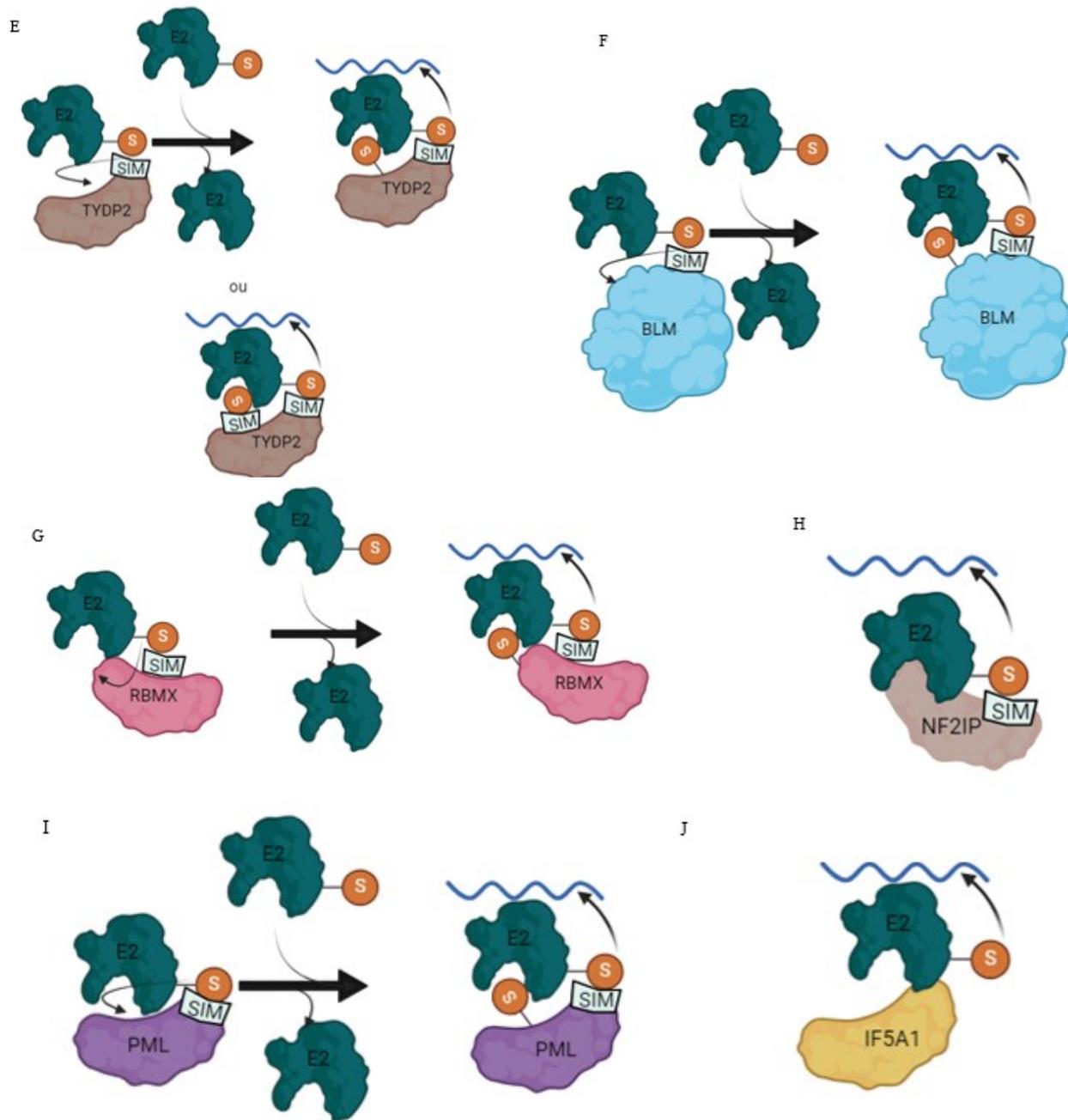


Figure 3.8: Mécanisme SUMO E3 Ligases proposé des candidats en se basant sur les données d'Alphafold. Les protéines se retrouvant dans le schéma sont la E2 en cyan, SUMO en orange, ZNF24 en rouge, HNRPC en vert, ZC3HE en bleu foncée, DAXX en gris, TYDP2 en brun, BLM en bleu clair, RBMX en rose, NF2IP en marron et PML en mauve. Le motif SIM des candidats contribuant à une activité est en blanc. Le substrat susceptible de subir la SUMOylation est représenté par un ruban bleu.

Pour les protéines RBMX, NF2IP et HNRPC, les interactions avec la E2, d'après Alphafold, ont lieu grâce à un domaine ressemblant à un SUMO ou SUMO-like domain (Annexe E Figure 5.1).

Ce domaine, en mimant SUMO, permet de lier la E2 en N terminal. Ceci permet d'outrepasser le besoin d'un SIM en se liant directement à la E2.

En conclusion, notre criblage bio-informatique a permis d'identifier 10 candidats ainsi que SLX4 (tous les candidats étant nucléaire, fortement SUMOylés, étant plus petits que 1500 a.a. avec au moins 1 SIM. Parmi ces candidats, 6 d'entre eux ont un critère additionnel : une interaction avec la E2, figure 3.1) qui présentent des signatures associées aux SUMO E3 ligases. Notre criblage présente toutefois un certain nombre de limitations (voir section 4.1) et bien que les protéines isolées lors de notre criblage soient plus susceptibles de posséder une activité SUMO E3 ligase, il demeure essentiel de confirmer ces résultats au moyen d'essais *in cellulo* et *in vitro*. Les prochaines sections traiteront de l'élaboration d'un essai de SUMOylation *in cellulo* (section 3.2) ainsi que d'une validation plus classique au moyen de protéines recombinantes purifiées et d'essais de SUMOylation (section 3.3).

### 3.2 Élaboration d'un criblage *in bactério* pour détecter l'activité SUMO E3 ligase de protéines d'intérêt.

#### 3.2.1 Considérations pour le design du criblage

Afin de déterminer l'activité SUMO E3 Ligases des candidats, ceux-ci doivent être soumis à des tests permettant de tester leur activité. L'utilisation de protéines recombinantes et d'essais de SUMOylation *in vitro* constituent le moyen le mieux établi pour caractériser l'activité SUMO E3 ligases des protéines. Le test *in vitro* consiste en la purification de la machinerie de SUMOylation ainsi que du candidat. À la suite de l'incubation de toutes ces composantes avec un substrat, dans un milieu adéquat, une SUMOylation devrait se produire si le candidat possède une activité de Ligase. Toutefois, afin de faciliter et d'accélérer la validation d'un grand nombre de candidats, nous avons décidé d'élaborer une procédure de criblage *in bactério*. Ce criblage détectera la capacité des protéines candidates à s'auto-SUMOyler. Le principe de ce test est l'inoculation et l'expression de deux vecteurs, soit un contenant la machinerie de SUMOylation et un contenant le candidat fusionné à des lysines SUMOylable, dans une bactérie. Notre hypothèse est que si la protéine candidate possède une activité SUMO E3 Ligase, elle aura tendance à s'auto-SUMOyler. Toutefois, une forte SUMOylation de la protéine ne constitue pas une preuve en soi. Étant donné qu'une protéine possédant davantage de résidus lysines pourrait être davantage SUMOylée par la

E2 qu'une protéine possédant moins de résidus lysines. De plus, la séquence jouxtant les résidus lysines influence la SUMOylation et les lysines qui font partie d'un consensus  $\Psi$ -K-x- $\alpha$  ou  $\Psi$  est un résidu hydrophobe, K la lysine SUMOylable, et  $\alpha$  un acide aminé chargé négativement constitue de meilleurs substrats pour l'enzyme E2. Il est donc nécessaire d'établir que le criblage ne prenne pas en compte les différences intrinsèques entre les protéines candidates, notamment en termes d'abondance et de localisation de leurs résidus de lysine. Pour ce faire, la séquence d'ADN codant pour la protéine d'intérêt sera insérée dans une cassette comportant dans l'ordre : (i) une étiquette His6 pour une purification éventuelle (ii) le gène d'intérêt (iii) un site de clivage à la protéase TEV (iv) une étiquette HA (séquence de l'étiquette HA : YPYDVPDYA) pour la détection de la protéine par immunoblot et (v) trois sites consensus de SUMOylation (MGADGGPIKAEGGIKAEGGIKAEGG ou les lysines SUMOylables sont en gras tandis que le motif consensus de SUMOylation est souligné) (Figure 3.9). Ainsi, une fois la protéine exprimée *in bacterio*, il est possible de l'exposer à la protéase TEV qui libérera un fragment rapporteur constitué uniquement de l'étiquette HA et des trois sites consensus de SUMOylation SUMOylée ou non. Ce motif SUMOylable sera le même pour toutes les protéines, mais la SUMOylation de ce motif dépendra de la capacité de la protéine candidate à s'autoSUMOyler.

Le choix d'un système bactérien s'explique quant à lui par trois considérations : (i) les bactéries ne possèdent pas de machinerie de SUMOylation donc les résultats de l'essai ne peuvent pas être faussés par l'interaction de la protéine candidate avec d'autres SUMO E3 ligases. Pour permettre toutefois une SUMOylation des protéines d'intérêts, les bactéries seront co-transformées avec les vecteurs pSUMO1 ou pSUMO2/3 qui permettent l'expression concomitante de SUMO1 (ou SUMO2/3), de l'enzyme E1 et de l'enzyme E2 (Weber et al., 2014). (ii) Les bactéries comme *E. coli* ne possèdent pas de SUMO protéases ce qui simplifiera l'analyse des données. (iii) Ce criblage présentera également l'avantage de faciliter l'identification des protéines s'exprimant bien et présentant une bonne solubilité. En effet, le niveau d'expression de chaque candidat est suivi grâce à l'étiquette HA fusionnée à la protéine d'intérêt. Nous prédisons aussi que les protéines insolubles (qui sont dirigées vers les corps d'inclusion de la bactérie) ne seront pas facilement SUMOylées. Bien que ce dernier point puisse résulter en une plus grande proportion de faux négatifs, il permettra de prioriser le criblage de protéines solubles ce qui facilitera ultimement la validation ultérieure des candidats par des essais *in vitro*.

### 3.2.2 Choix d'un vecteur pBAD pour exprimer les candidats

Tel que brièvement mentionné ultérieurement, la bactérie contiendra ainsi 2 vecteurs dont 1 vecteur, pSUMO, qui contient la E1, la E2 et SUMO et 1 vecteur contenant la cassette du candidat E3. En considérant que le vecteur pSUMO est inductible à l'IPTG, il est important de choisir pour les candidats un vecteur d'expression qui est inductible à autre chose pour permettre l'induction différentielle de pSUMO et du candidat. À ces fins, nous avons retenu pBAD, un vecteur d'expression inductible à l'arabinose. L'induction à l'arabinose permet de surcroît de moduler le niveau de la protéine et potentiellement de normaliser le niveau d'expression de toutes les protéines. Ceci permet également de diminuer les niveaux d'inductions via une diminution de concentration en arabinose pour éviter une toxicité associée à une trop grande concentration de la protéine.

L'élaboration d'une procédure de criblage nécessite l'utilisation de contrôle adéquat. Comme contrôles positifs nous avons retenu les modules strictement nécessaires pour une activité SUMO E3 ligases des protéines ZNF451, soit les acides aminés 25-56 ainsi que de RanBP2 soit les acides aminés 2631 à 2771. Bien que les protéines pleine longueur aurait pu être choisi pour ce criblage, les modules E3 Ligases sont plus facile à obtenir puisqu'ils sont plus petits que les protéines pleine longueur (1 061 acide aminé pour ZNF451 et 3 224 acide aminé pour RanBP2).

Ces deux protéines ont été choisies puisque leur mécanisme est déjà élucidé et possèdent une robuste activité SUMO E3 ligase. ZNF451 contient les deux SIMs, ainsi que les acides aminés acidiques ou phosphorylable de part et d'autre des SIMs avec les acides aminés 25-56 étant strictement nécessaire pour son activité de ligases tandis que RANBP2 contient les domaines IR1 et IR2 qui sont dotés d'activité SUMO E3 ligases. Comme contrôles négatifs, les protéines Ubiquitine et SUMO1 ont été choisies afin d'évaluer l'activité de SUMOylation de la E2 sur les trois sites de SUMOylation en absence de E3 tout en gardant une protéine en N-terminal du substrat. Considérant que ces deux protéines peuvent se faire SUMOyler et sont bien reconnues par la E2, il est possible de voir l'effet de la E2 sur le substrat. Parallèlement, un vecteur pBAD ne contenant pas de E3, donc ne contenant que le substrat sera utilisé avec pSUMO afin de s'assurer qu'aucune réaction adverse n'ait lieu.

### 3.2.3 Clonage des candidats dans le vecteur pBAD

La Figure 3.9 illustre la cassette qui a été commandée chez Integrated DNA Technologies et qui a été insérée au niveau du site NcoI d'un vecteur pBAD pour générer le vecteur pBAD-cassette\_vider :



Figure 3.9: Séquences en nucléotide (A) et en acides aminés (B) de la cassette (avant le clonage des candidats) qui a été insérée dans le site NcoI de pBAD.

En parallèle, nous avons procédé à l'amplification de la séquence codante pour les candidats d'intérêt. Grâce à l'ARN extrait par Trizol de cellules HEK293T et HeLa et convertis en ADNc, une amplification des candidats a été effectuée. En premier lieu, les gènes HNRPC, ZNF451, BLM, IF5A1, TYDP2 et DAXX ont été amplifiés en utilisant l'ADNc HEK293T tandis que les gènes RanBP2, RBMX, NF2IP, ZNF24 et ZC3HE ont été amplifiés en utilisant l'ADNc HeLa. Le gène SLX4 n'a pas pu être amplifié. Les gènes qui ont été obtenus par amplification PCR sont les gènes ZNF451 (1-56 et 25-56), ZNF24, IF5A1, RanBP2, RBMX, HNRPC, TYDP2 et ZC3HE, soit ceux se retrouvant encadrés à la Figure 3.10. Les gènes qui n'ont pas été obtenus sont BLM, NF2IP et DAXX en dépit de nos tentatives d'optimisation (e.g. changement de température d'hybridation, utilisation de 5 % de DMSO, utilisation de techniques de PCR alternative comme le touch-down PCR) (Korbie et Mattick, 2008). Les gènes Ubiquitine et SUMO1 n'ont pas été montrés sur gel d'agarose, car ceux-ci proviennent de source externe (c'est à dire ne proviennent pas de cellules, mais de vecteurs préalablement présentés au laboratoire) et ont été inséré dans pBAD.

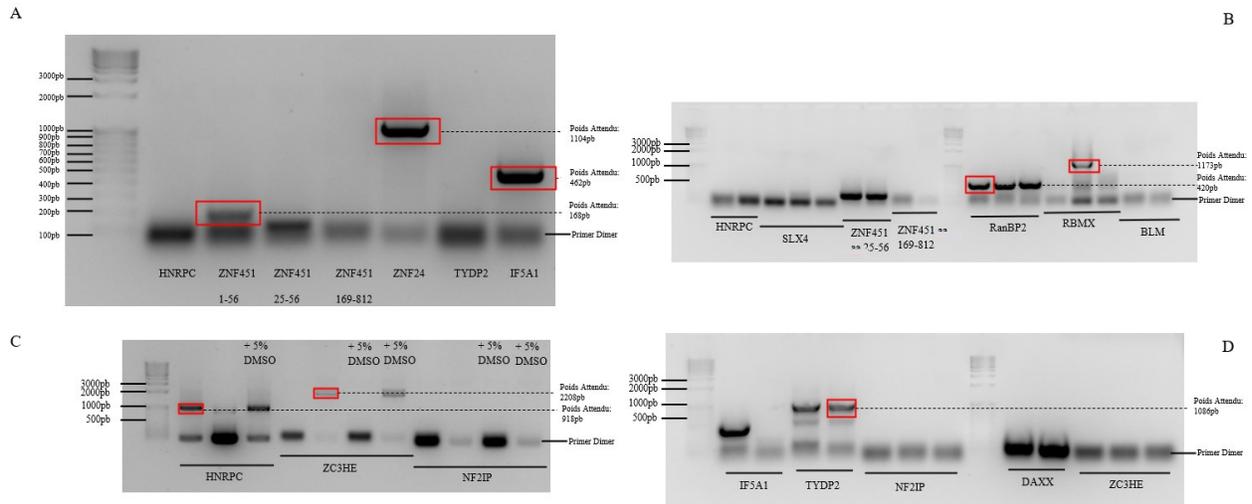


Figure 3.10: Gel d'agarose 1 % avec les amplifications par PCR des différents candidats. Chaque puits correspond à une amplification différente d'un candidat. Les carrés rouges correspondent aux bandes excisées qui ont été utilisées pour la suite des expériences. A) Amplification de HNRPC, ZNF451, ZNF24, TYDP2 et IF5A1. B et D) Amplification de HNRPC, IF5A1, SLX4, ZNF451, TYDP2, NF2IP, DAXX, ZC3HE, RBMX et BLM ainsi que certains contrôles dont ZNF451, 25-56, RanBP2 et SLX4 C) Exemple d'optimisation pour des amplifications de gènes de HNRPC, ZC3HE et NF2IP qui n'ont pas été obtenues lors d'A, B et C.

Les candidats ayant pu faire l'objet d'amplifications ont ensuite été insérés via une réaction de Gibson au niveau du site KpnI du plasmide pBAD-cassette\_vides. En aval du candidat inséré, plusieurs lysines sont présentes dans des sites consensus de SUMOylation (voir figure 3.9 B). Ces sites de lysines peuvent se faire SUMOylés par la protéine candidate se trouvant en amont, soit par auto-SUMOylation. Le plasmide résultant est inductible de façon concentration dépendante à la L-Arabinose et contient le candidat E3 à tester et a été utilisé en conjonction avec les plasmides pSUMO1 ou pSUMO2/3 (inductibles à l'IPTG) afin de recréer la machinerie de SUMOylation chez la bactérie.

### 3.2.4 Détermination de la quantité d'arabinose nécessaire pour une expression optimale

Un test d'expression *in bacterio* a été fait avec le contrôle pBAD-SUMO1 avec ou sans les vecteurs pSUMO1 ou pSUMO2) afin de s'assurer de la concentration en L-Arabinose à utiliser pour les expressions des candidats. A la Figure 3.11, il est possible d'observer la différence entre les bactéries non induites (NI), soit les bactéries n'ayant pas reçu de L-Arabinose, et les bactéries induites avec l'arabinose. Les différences principales se retrouvent au niveau de 25 et 30 kb comme montrés à la figure 3.11. Il est possible de voir que ces bandes sont présentes chez l'échantillon de

SUMO1 purifiée par NiNTA. La condition qui semble présenter l'apparition de bande la plus prononcée, soit les protéines qui semblent être la plus abondante, sont les bactéries induites avec 0.2 % d'arabinose avec la bande de 25 kDa qui semble être plus présentes. Ceci semble être le cas pour les 3 conditions, soit en présence et absence de pSUMO. Ceci indiquerait que SUMO1 est abondant davantage a 0.2 % d'induction a l'arabinose que 0.02 %.

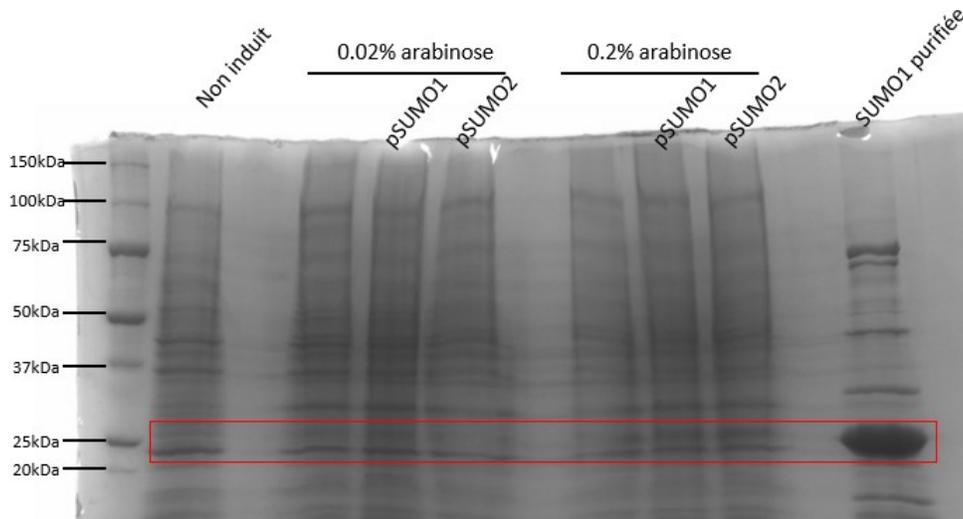


Figure 3.11: Gel SDS-Page 15 % montrant l'expression et la solubilité de SUMO1 suivant une induction de bactéries *E. coli* BL21DE3 (CP) par la L-Arabinose et IPTG pour pSUMO. L'induction a été réalisée en présence de 0.02 % ou 0.2 % de L-Arabinose. Les bactéries expriment le vecteur pBAD-SUMO1, avec ou sans les vecteurs pSUMO1 et pSUMO2. La SUMO1 purifiée équivaut à la production obtenue à partir d'un litre de culture bactérienne de pBAD-SUMO1, ayant été induite avec 0.2 % de L-Arabinose, puis purifiée par chromatographie d'affinité NiNTA.

### 3.2.5 SUMOylation *in bacterio* des candidats

Le test *in bacterio* consiste en l'expression, pour chaque candidat, de son système rapporteur et de pSUMO1 et 2 suivis d'une détection par immunobuvardage avec des anticorps anti-HA. Sachant que SUMO1, par le biais du système d'expression pBAD, est abondamment exprimé avec 0.2 % d'arabinose, un candidat, ZNF24 est testée ainsi que le contrôle positif ZNF451 et le système rapporteur sans candidats (sans E3) (Figure 3.12). L'immunobuvardage des bactéries n'exprimant pas de candidats, soit le système rapporteur seul sans E3, ne présentant aucune bande (aucune détection de bande anti-HA), celui-ci fut exclu de la partie résultats. Par la suite, ZNF451, le contrôle positif de l'expérience a présenté plusieurs bandes pour l'induction a 2 % et 0.2 % d'arabinose, avant ou après la lyse, en presence de pSUMO1 ou pSUMO2. De plus, suite à un

traitement à la protéase TEV, il est possible de voir 4 bandes. Etant donnée que la TEV clive entre le candidat et le substrat étiqueté avec HA, ces bandes correspondent à multiples stades de SUMOylation du substrat. En parallèle, un traitement par la protéase ULP1 (avec ou sans TEV), qui permet la désSUMOylation, résulte en la disparition de toutes les bandes. Pour ZNF24, il est possible de voir 2 bandes distinctes pour l'induction à 2 % et 0.2 % d'arabinose autant pour pSUMO1 que pSUMO2. Cependant, le signal étant saturé, il est possible que d'autres bandes soient présentes, mais avec un signal moindre ce qui les rend indistinguables. Suite à la lyse, les mêmes bandes sont présentées pour 0.2 % d'arabinose. Pour les bandes d'induction à 2 %, celle-ci sont difficile à identifier due à la proximité au signal saturé des premiers puits. Par la suite, suite à un traitement à la protéase TEV, il est possible de voir 4 bandes. De même, ces bandes correspondent aux substrats présentant différents stades de SUMOylation. En parallèle, un traitement par la protéase ULP1 fait disparaître toutes les bandes sauf pour une bande prédominante coïncidant avec la bande inférieure pour le lysat et les inductions indiquant que cette bande correspond à la forme basal de ZNF24. De plus, un traitement ULP1+TEV, résulte en la disparition de toutes les bandes. Pour IF5A1 et ZC3HE, outre les bandes au niveau de l'extraction brut, il ne semble pas y avoir de bande indiquant une SUMOylation du candidat donc ceux-ci seront considérés comme étant des candidats négatifs. En ce qui concerne HNRPC, celui-ci présente de nombreuses bandes ayant une intensité faible au niveau de l'extrait brut avec 2 % arabinose. De plus, suivant un clivage à la TEV, plusieurs stades de SUMOylation apparaissent ce qui pourrait indiquer un potentiel activité SUMO E3 Ligase. Afin de tester pour une activité SUMO E3 Ligase, HNRPC doit être assujéti à des tests *in vitro*. Pour les protéines RBMX et TYDP2, en dépit du fait que leur séquence ont été obtenues par RT-PCR, ceux-ci ont eu des problèmes d'expression avec le système bactérien. Pour RBMX, il n'a pas été possible d'obtenir des résultats tant au niveau des tests *in bacterio* qu'au niveau de la purification des protéines, tandis que TYDP2 n'a été exprimé que pour la purification des protéines. Ces deux candidats seront donc omis pour les tests *in vitro*. Des résultats similaires ont été observés pour toutes les conditions pour les 2 répliquats ayant eu lieu.

En concluant les tests *in bacterio*, ZNF24 semble présenter une activité de SUMO E3 Ligase, tandis que ZC3H1 et IF5A1 ne semblent pas présenter d'activité de SUMO E3 Ligase. Pour HNRPC, considérant qu'aucune conclusion sur son activité de SUMO E3 Ligase ne peut être émise en se basant sur les résultats *in bacterio*, des tests *in vitro* ont été effectués.

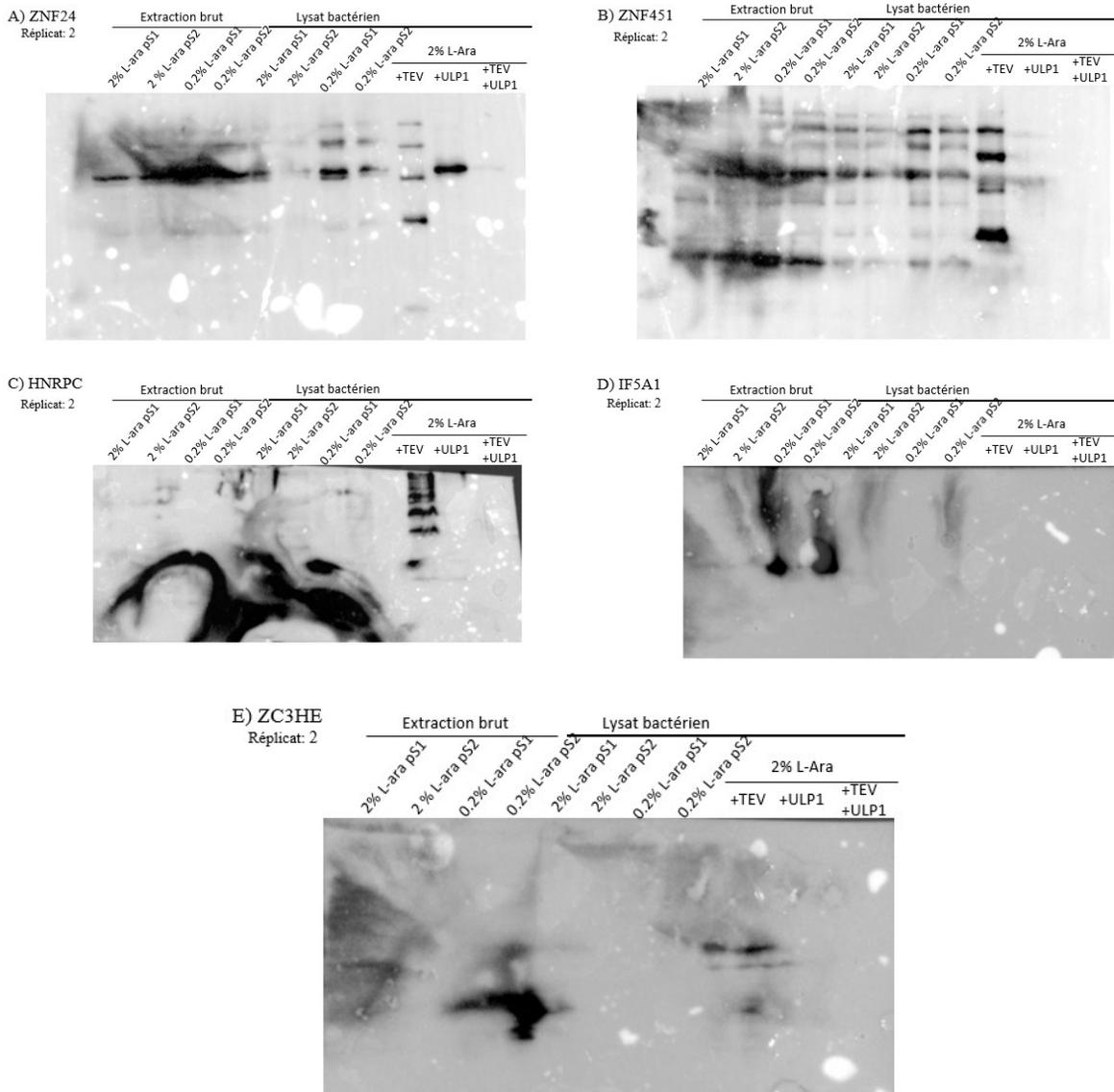


Figure 3.12: Détection de SUMOylation de substrats de bactéries exprimant un candidat en présence de pSUMO1 et pSUMO2/3 par Immunobuvardage par anticorps anti-HA. Une induction a été réalisée en présence de 2 % ou 0.2 % de L-Arabinose. L'extraction brut correspond aux bactéries suivant une induction tandis que le lysat bactérien correspond aux bactéries induites ayant subi une sonication et une clarification par centrifugation. Les gels d'immunobuvardage présenté ne comporte pas de marqueurs de poids moléculaire. PS1=pSUMO1, PS2=pSUMO2, L-ara=L-arabinose. L'expérience, via des essais indépendants, a été reproduite 2 fois pour chaque condition présentée ci-dessus.

### 3.3 Purification de candidats

Pour donner suite aux tests effectués *in bactério*, des tests *in vitro* doivent être effectués pour les candidats HNRPC, ZNF24. Pour ce faire, les protéines ont dû être purifiées. Afin de pouvoir

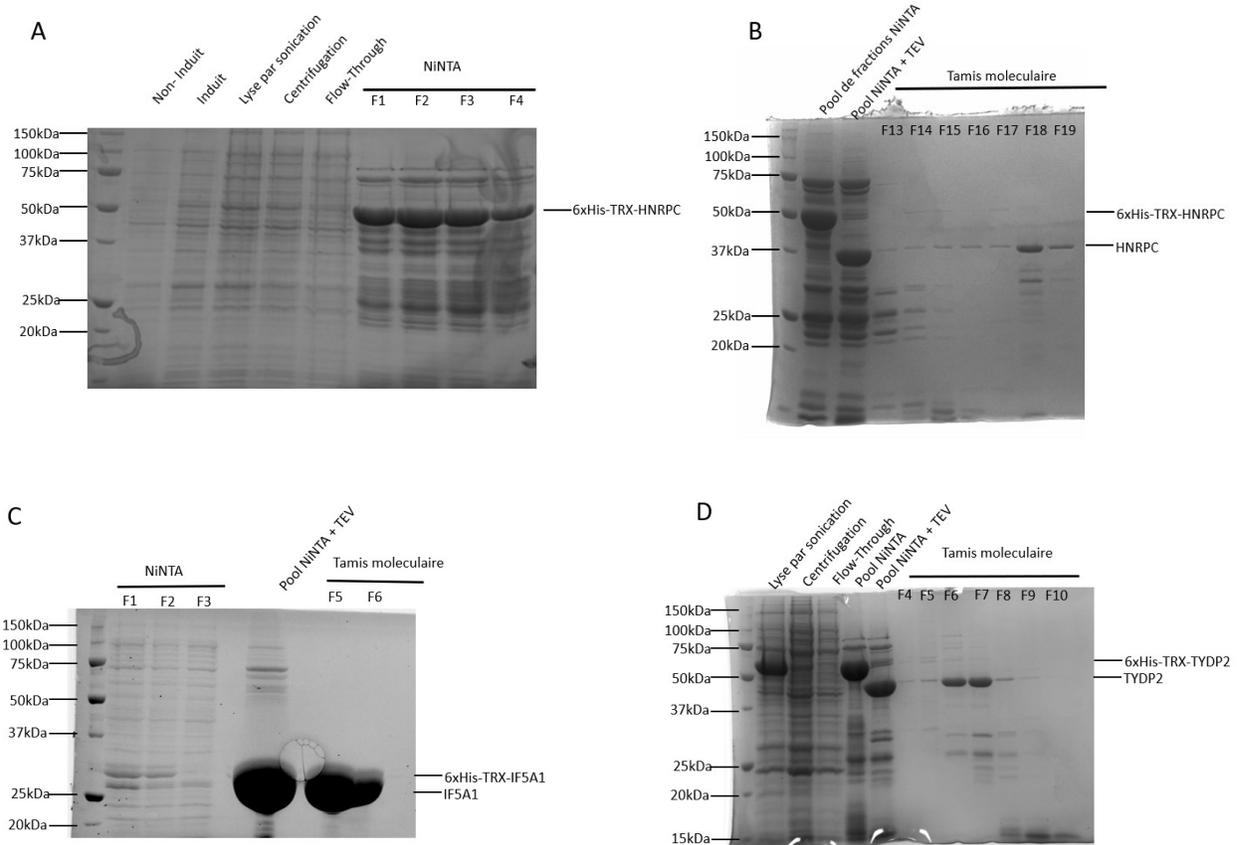
purifier les candidats, nous avons cloné les candidats indiqués ci-haut dans un vecteur pRSF-TRX. Ce vecteur est idéal pour la purification de protéines, car celui-ci contient une étiquette His6 et une étiquette de solubilité, soit la thioredoxine (TRX). Cette étiquette TRX, hautement soluble, est fusionné au candidat et permet d'augmenter la solubilité du candidat lors de la production de protéine en bactérie. De plus, grâce au site de clivage à la TEV, il est possible d'éliminer les étiquettes afin de conserver uniquement la protéine d'intérêt, soit le candidat. Nous avons, grâce aux séquences préalablement amplifiées provenant de cellules (Figure 3.10), cloné les protéines dans le vecteur pRSF-TRX par une réaction de Gibson grâce au site BamHI de ce vecteur (Figure 3.13).



Figure 3.13: Séquences en nucléotide (A) et en acides aminés (B) des séquences utilisées pour la purification de candidats (avant le clonage des candidats) qui a été insérée dans le site BamHI d'un vecteur pRSF-TRX.

Pour la purification des candidats, les protéines sont purifiées par chromatographie d'affinité NiNTA et permettent d'obtenir des fractions de 1 mL (Figure 3.14, A et E). Une fois la chromatographie d'affinité effectuée, les 5 fractions contenant le plus de protéine sont combinées pour donner une fraction combinée. Après avoir combiné les fractions, la protéase TEV est rajoutée (250 µM) afin de cliver les étiquettes His6 et TRX. Par la suite, nous avons utilisé un tamis moléculaire S75 ou S200 pour purifier davantage les protéines par exploitation de leur masse

moléculaire. Cette étape permet, entre autres, de séparer la protéase TEV (~27 kDa) de la protéine d'intérêt qui ont, tous les deux, des poids différents. Également, le tamis moléculaire permet de purifier un mélange hétérogène de protéine pour obtenir la protéine d'intérêt de façon homogène. Ceci est visible à la figure 3.14 B, C, D et F pour les protéines HNRPC, ZNF24, TYDP2 et IF5A1 ou celles-ci ont pu être purifiées par tamis moléculaire avec la fraction combinée NiNTA+ TEV étant l'échantillon soumis au tamis. Pour le tamis moléculaire S200, les échantillons F18 et F19 de HNRPC. Les échantillons F7 et F8 de ZNF24 ont été conservés et utilisés pour les tests enzymatiques *in vitro*. Les échantillons F6 et F7 de TYDP2 provenant d'une purification par S200 et F5 et F6 de IF5A1 provenant de S75 ont été préservés.



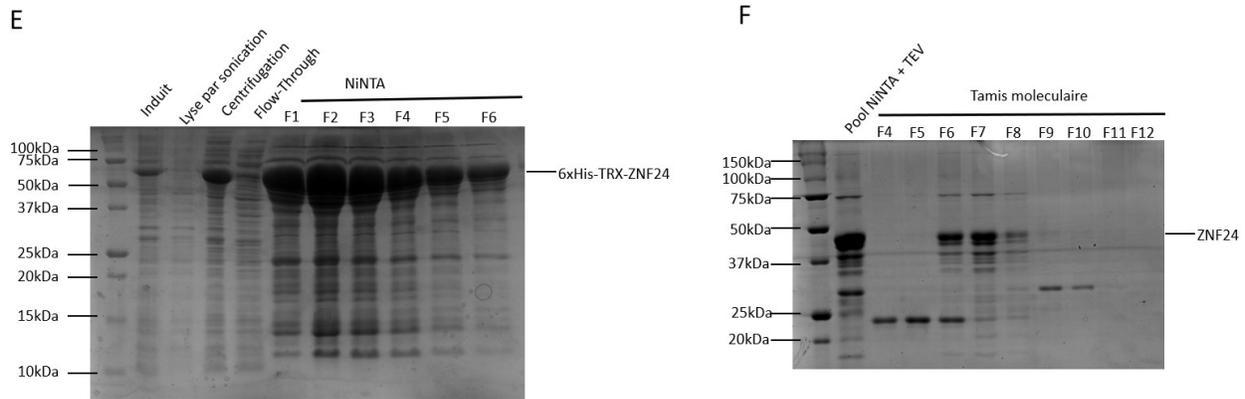


Figure 3.14: Gels SDS-Page 12 % montrant la purification des candidats HNRPC, TYDP2, IF5A1 et ZNF24. Pour la purification de HNRPC, celle-ci a subi une purification par affinité NiNTA (A) et par Tamis moléculaire S200 (B). Pour la purification de IF5A1, celle-ci a subi une purification par NiNTA et Tamis moléculaire S75 (C). Pour TYDP2, la protéine a subi une purification par affinité et tamis moléculaire S200 (D). Pour ZNF24, celle-ci subi une purification par affinité NiNTA et un tamis moléculaire S75 (F).

### 3.4 Essais de SUMOylation *in vitro* des candidats

Pour donner suite aux tests effectués *in bactério*, des tests *in vitro* ont été effectués pour les protéines HNRPC et ZNF24. Le contrôle positif de cette expérience est le module SUMO E3 Ligase de ZNF451. Les tests *in vitro* permettent l'identification et la caractérisation de l'activité SUMO E3 ligase des protéines (Werner et al, 2009). Cette méthode permet, par le biais de protéines recombinantes purifiées, d'évaluer adéquatement l'activité E3 de protéines testées. Pour cela, les protéines ayant été soupçonnées de posséder une activité SUMO E3 Ligase aux sections précédentes, ont été testées via des tests *in vitro*. Ces tests *in vitro* permettent, à la fois de valider le test *in bactério* mais également de permettre une quantification adéquate de la cinétique enzymatique. Nous utiliserons pour ce faire deux indicateurs : (i) l'autoSUMOylation de la protéine candidate et (ii) l'augmentation de SUMOylation d'une protéine substrat. Dans ce dernier cas, il s'agit d'un fragment de la protéine p53 marqué par la fluorescéine qui possède un site consensus de SUMOylation et qui a déjà été utilisé pour caractériser des SUMO E3 ligases (Cappadocia et al., 2015; Pichler et al., 2002).

Pour les tests *in vitro*, une condition modèle (Werner et al, 2009) et une condition avec 10 fois plus de E1 et E2 ont été utilisées soit 100 nM de E1, 100 nM de E2 ou 1  $\mu$ M de E1 et 1  $\mu$ M de E2. Il est possible de voir que peu de changement a eu lieu pour les réactions à 100 nM pour la totalité

des essais avec l'apparition de très peu de bandes après 180 minutes. Cependant, des changements drastiques sont présents pour les réactions à 1  $\mu$ M pour ZNF24 et ZNF451 en comparant avec ceux de la E2 et ce, visible après 60 minutes (Figure 3.15). Il est possible d'observer l'apparition de plusieurs bandes à poids moléculaires élevées (supérieur à  $\sim$ 75 kDa) à partir de 30 minutes pour ZNF451 ou 60 minutes pour ZNF24. Ceci indiquerait qu'en présence de ces protéines, une SUMOylation importante a lieu. Si l'on compare avec HNRPC, le gel de l'essai *in vitro* ne présente pas de différence significative avec le gel sans candidats, soit le contrôle négatif. Toutes ces observations ont été observable pour les 3 réplication de l'expérience.

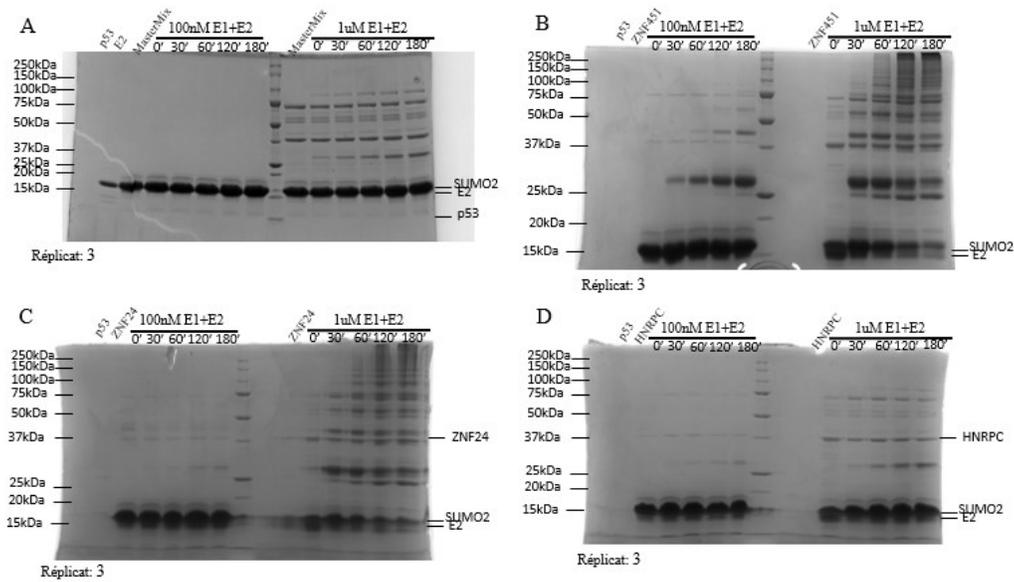


Figure 3.15: Essais *in vitro* pour vérifier l'activité SUMO E3 Ligase des candidats HNRPC, ZNF24 et TYDP2 sur gels SDS-Page à gradient 10-20 %. Les expérimentations *in vitro* ont été conduites en utilisant des protéines purifiées. En employant la protéine E1 (à des concentrations de 100 nM ou 1  $\mu$ M), la protéine E2 (à des concentrations de 100 nM ou 1  $\mu$ M), SUMO2/3 (à une concentration de 50  $\mu$ M), ainsi que le substrat p53 (à une concentration de 1  $\mu$ M), un candidat (à une concentration de 1  $\mu$ M) a été évalué à divers moments, notamment 0, 30, 60, 120 et 180 minutes. Cette réaction été amorcées par l'ajout de 2 mM d'ATP. Chaque essais a été effectué 3 fois (essais indépendant avec des lots différents de protéines à chaque fois). A) Absence de candidats, B) ZNF451, C) ZNF24, D) HNRPC

Tableau 3.1: Potentiel des protéines interagissant avec E2 en tant que SUMO E3 Ligases : Résumé des expériences réalisées

Gène	ZNF24	TYDP2	HNRPC	DAXX	PML	BLM
Raisons	Interaction avec la E2 d'après la littérature (Gocke, Yu et Kang, 2005), SIMs en N-terminal accessibles pour lier SUMO, protéine hautement SUMOylées	Interaction avec la E2 d'après la littérature (Lee et al., 2003), forte interaction avec la E2 d'après Alphafold, Régions désordonnée qui pourraient contribuer a à une activité SUMO E3 ligase	Interaction avec la E2 d'après la littérature (Rolland et al., 2014), SUMO-like domain en N terminal qui interagit avec la E2, hautement SUMOylées	Interaction avec la E2 d'après la littérature (Hsieh et al., 2013), protéine hautement SUMOylées, SIMs accessible pour lier SUMO	Interaction avec la E2 d'après la littérature (Bregnard et al., 2022; McManus et al., 2018; Hembram et al., 2020; Jo et al., 2016), protéine hautement SUMOylées, SIMs accessible pour lier SUMO, Régions désordonnée qui pourraient contribuer a à une activité SUMO E3 ligase	Interaction avec la E2 d'après la littérature (Eladad et al., 2005), protéine hautement SUMOylées, SIMs accessible pour lier SUMO, Régions désordonnée qui pourraient contribuer a à une activité SUMO E3 ligase
Amplification	Oui	Oui	Oui	Non	Non	Non
Isoforme obtenu	1	1	1	N/A	N/A	N/A
Expression en bactérie	Exprimable et Soluble	Exprimable et Soluble	Exprimable et Soluble	N/A	N/A	N/A
Tests <i>in Bactério</i>	Fig 3.12	N/A	N/A	N/A	N/A	N/A
Tests <i>in vitro</i>	Fig 3.15	N/A	Fig 3.15	N/A	N/A	N/A

Taille de protéine (en Da)	42 155	40 930	32 338	81 373	97 551	159 000
Nombres de sites SUMOyles	12	5	24	10	16	33
Nombres totaux de lysines	22	18	29	35	30	123
Nombres de SIMs	5	3	2	2	3	6
Cellules utilisées pour l'amplification	N/A	HEK293T	HEK293T	HEK293T	HeLa	HEK293T
Abondance d'ARN (en nTPM)	N/A	145	963	N/A	145	128
Abondance de la protéine (en ppm)	N/A	145 (Geiger et al., 2012)	98.2 (Geiger et al., 2012)	N/A (Geiger et al., 2012)	130 (Geiger et al., 2012)	128 (Geiger et al., 2012)
Activité SUMO E3 Ligase	Oui	N/A	Non	N/A	N/A	N/A

Le tableau présente une liste de candidats potentiels pour des SUMO E3 Ligases, leurs caractéristiques associées, et les résultats des expériences bio-informatiques et biochimiques. La catégorie gène correspond aux gènes candidats, raison correspond aux caractéristiques bio-informatiques suggérant une activité SUMO E3 ligase au gène, amplification correspond à l'obtention ou non du gène par PCR à partir d'ADNc isolés de lignées cellulaires, le numéro de l'isoforme amplifié selon la numérotation utilisée par Uniprot, expression en bactérie correspond à l'expression soluble et distinguable en bactérie en comparant les bactéries non induites et induites, les tests *in bacterio* et *in vitro* correspondent aux figures dans lequel se retrouve les résultats de ces tests, taille de protéine correspond à la taille en Da de la protéine à l'étude, le nombre de sites SUMOyles correspond au nombre de lysines étant SUMOyles d'après les données de Hendriks et Vertegaal (2016), le nombre de SIMs correspond au nombre de SIMs criblée de façon bio-informatique par JASSA ou DEEPSUMO, les cellules utilisées correspondent à la lignée cellulaire utilisée pour l'essai d'obtention du gène en question, abondance d'ARN correspond à l'abondance de l'ARN dans la lignée cellulaire choisie ([www.proteinatlas.org](http://www.proteinatlas.org)), Abondance de la protéine correspond à l'abondance de la protéine dans la lignée cellulaire en question en ppm et l'activité E3 Ligase correspond à la détermination du gène comme ayant une activité SUMO E3 Ligase ou non. "N/A", signifie que les données spécifiques ne sont pas disponibles pour cette catégorie.

Tableau 3.2 : Potentiel des protéines ayant un SIM interagissant avec E2 en tant que SUMO E3 Ligases : Résumé des expériences réalisées

Gène	RBMX	IF5A1	NF2IP	ZC3HE
Raisons	SUMO-like domain en N terminal interagissant avec la E2, SIM accessible pour lier SUMO, Régions désordonnée qui pourraient contribuer à une activité SUMO E3 Ligase	SIMs accessible pour lier SUMO, protéine hautement SUMOylées	2 Uq-like domain en C-terminal qui interagissent avec la E2, présence de plusieurs SIMs, Régions désordonnée qui pourraient contribuer à une activité SUMO E3 Ligase	SIMs accessible pour lier SUMO, protéine hautement SUMOylées, Régions désordonnée qui pourraient contribuer à une activité SUMO E3 Ligase
Amplification	Oui	Oui	Non	Oui
Isoforme obtenu	H0Y6E7	1	N/A	2
Expression en bactérie	Exprimable et Soluble	Exprimable et Soluble	N/A	Exprimable et Soluble
Tests <i>in bactério</i>	N/A	N/A	N/A	N/A
Tests <i>in vitro</i>	N/A	N/A	N/A	N/A
Taille de protéine (en Da)	42 332	16 832	45 817	82 876
Nombres de sites SUMOyles	4	7	7	13
Nombres totaux de lysines	13	13	23	53
Nombres de SIMs	2	4	8	5
Cellules utilisées pour l'amplification	HeLa	HEK293T	HeLa	HeLa
Abondance d'ARN (en nTPM)	231.7	925.9	56.2	43
Abondance de la protéine (en ppm)	218 (Geiger et al., 2012)	68.5 (Geiger et al., 2012)	153 (Geiger et al., 2012)	159 (Geiger et al., 2012)

Le tableau présente une liste de candidats potentiels pour des SUMO E3 Ligases, leurs caractéristiques associées, et les résultats des expériences bio-informatiques et biochimiques. La catégorie gène correspond aux gènes candidats, raison correspond aux caractéristiques bio-informatiques suggérant une activité SUMO E3 ligase au gène, amplification correspond à l'obtention ou non du gène par PCR à partir d'ADNc isolés de lignées cellulaires, le numéro de l'isoforme amplifié selon la numérotation utilisée par Uniprot, expression en bactérie correspond à l'expression soluble et distinguable en bactérie en comparant les bactéries non induites et induites, les tests *in bacterio* et *in vitro* correspondent aux figures dans lequel se retrouve les résultats de ces tests, taille de protéine correspond à la taille en Da de la protéine à l'étude, le nombre de sites SUMOyles correspond au nombre de lysines étant SUMOyles d'après les données de Hendriks et Vertegaal (2016), le nombre de SIMs correspond au nombre de SIMs criblée de façon bio-informatique par JASSA ou DEEPSUMO, les cellules utilisées correspond à la lignée cellulaire utilisée pour l'essai d'obtention du gène en question, abondance d'ARN correspond à l'abondance de l'ARN dans la lignée cellulaire choisie ([www.proteinatlas.org](http://www.proteinatlas.org)), Abondance de la protéine correspond à l'abondance de la protéine dans la lignée cellulaire en question en ppm et l'activité E3 Ligase correspond à la détermination du gène comme ayant une activité SUMO E3 Ligase ou non. "N/A", signifie que les données spécifiques ne sont pas disponibles pour cette catégorie.

## CHAPITRE 4 DISCUSSION

### 4.1 Criblage bio-informatique et mécanisme SUMO E3 Ligase de candidats

Afin de découvrir de nouvelles SUMO E3 Ligases ayant un mécanisme d'action atypique, un criblage bio-informatique a permis d'identifier 50 protéines. Parmi ces protéines, 11 candidats ont été retenus pour des tests d'activité SUMO E3 Ligases.

La sélection bio-informatique de candidats a consisté en l'isolation de protéines présentant certaines caractéristiques retrouvées principalement chez les SUMO E3 Ligases. Un premier critère de sélection est une annotation nucléaire (GO :0005634 sur Uniprot). L'annotation *nucéaire* permet de choisir les protéines qui se retrouvent au niveau du noyau à un moment ou à un autre durant leur vie. Ce critère nucléaire a été choisi, car la SUMOylation, bien qu'elle puisse avoir lieu dans le cytoplasme (Mayo, 2012; Naidu et al., 2012), est effectuée principalement dans le noyau dû à l'abondance des enzymes E1 et E2 dans ce compartiment (Wasik et al, 2014). Cette sélection inclut les protéines qui ont une activité strictement nucléaire, mais également les protéines qui ont une activité au-delà du noyau. Par exemple, l'histone H2B (Uniprot : Q93079) est une protéine qui a pour fonction la compaction de l'ADN en chromatine donc son rôle est entièrement nucléaire (Bonenfant et al, 2006). Cependant, sa traduction a lieu dans le cytoplasme donc son annotation Uniprot comporte la localisation subcellulaire nucléaire et cytoplasmique.

Concernant la localisation, comme mentionné précédemment, la SUMOylation soit majoritairement un processus nucléaire, il existe des évidences de SUMOylation au niveau du cytoplasme (Wasik et al., 2014). Afin de raffiner la recherche de candidats, d'autres critères sont requis pour découvrir de nouvelles SUMO E3 Ligases. Une des caractéristiques qui est importante à retenir est la forte SUMOylation intrinsèque des SUMO E3 Ligases. Les composantes de la machinerie de SUMOylation, soit la E2 et E3 ont été rapportée présentant une autoSUMOylation. La E2, par le biais d'automodification par SUMOylation, s'autoinhibe (Ho et al., 2011). Pour les SUMO E3 connues, elles ont tendance à s'autoSUMOyler (Kumar et al., 2017; Schellenberg et al., 2017; Schmidt et al, 2002; Hendriks et al., 2014). Ce phénomène est également observé chez les Ubiquitine E3 ligases avec un haut niveau d'Ubiquitination. Pour l'ubiquitination,

l'automodification joue un rôle d'autorégulation de l'activité de la protéine allant souvent vers l'inhibition (de Bie et al., 2011; Amemiya et al., 2008; Yong et al., 2023; Peterson et al., 2019).

Grâce aux données extraites de la revue de littérature d'Hendriks et al., il est possible d'identifier les protéines ayant au moins un site de SUMOylation, soit 1521 protéines. Ces données exhaustives contiennent plus de sites de SUMOylation identifiés que ceux répertoriés dans plusieurs bases de données, telles que Uniprot et PhosphoSitePlus, qui enregistrent respectivement 58 et 39 sites de SUMOylation pour ZNF451. Pour les données issues de la revue de littérature, celle-ci indique que ZNF451 présentent 62 sites de SUMOylation, soit 4 de plus que ceux émis par Uniprot et 23 que ceux de PhosphoSitePlus. Concernant les sites de SUMOylation, il faut noter que certaines banques de données que nous avons utilisées présentent également certains biais de représentation. Par exemple, la banque de données de sites de SUMOylation se base sur des études protéomiques. Or, certaines protéines peu abondantes peuvent ne pas se retrouver dans cette banque de données. Également, certains sites de SUMOylation peuvent avoir été omis dû à la difficulté à identifier sans équivoque les peptides résultants. En ce qui concerne les isoformes, peu d'études issues de la revue indiquent l'isoforme étudié donc celui-ci sera omis ; l'isoforme principal, fourni par Uniprot, a été utilisé pour l'ensemble de la recherche bio-informatique. Ce critère, bien que permettant de faciliter la recherche de candidats, pourrait être considéré comme une limitation de l'étude. Faute d'information présentée sur les isoformes des protéines à l'étude, un isoforme non principal d'une protéine pourrait avoir une activité SUMO E3 Ligase et donc pourrait passer inaperçu. Malgré cette considération, il est convaincant que ce choix méthodologique permette d'obtenir des résultats significatifs et informatifs pour l'analyse de la SUMOylation des protéines étudiées. Parmi les protéines qui ont été retenues, un critère de taille a été rajouté afin de permettre une co-expression efficace dans *E. coli* : certaines études indiquent une limite de production de protéine recombinante de 60 kDa chez les bactéries (Canaves et al., 2004; Goh et al., 2004; Gråslund et al., 2008a). Des protéines recombinantes allant de 16 kDa (IF5A1) à 159 kDa (BLM) ont été choisies pour cette étude. Les différentes protéines ont été choisies pour leur potentiel de SUMO E3 Ligases qui, parmi les 50 protéines finales retenues, exprimait le meilleur potentiel. Ce potentiel est jugé sur la capacité à lier la machinerie de SUMOylation par des interactions observables et possibles observer via AlphaFold. Afin de pouvoir tester la capacité E3 des candidats, ceux-ci, ainsi que la machinerie de SUMOylation, doivent être exprimés et solubles. La restriction de taille protéique est un biais

de l'étude, car des protéines ayant plus que 1500 acides aminés pourraient présenter une activité SUMO E3 Ligases. Cependant, due à l'organisme de production de protéine utilisée, soit *E. coli*, la limite de 1500 acides aminés est imposée. La solubilité des protéines est un critère important, car une protéine insoluble pourrait être détectée par SDS-PAGE et immunobuvardage, mais ne présenterait pas d'activité SUMO E3 Ligase donnant des bandes similaires aux contrôles négatifs (les contrôles sont SUMO1 et pSUMO sans E3) puisque la machinerie de SUMOylation se retrouverait dans le cytoplasme tandis que le candidat serait dans les corps d'inclusions. Pour la détection de protéines par immunobuvardage, le manque d'activité du candidat due à l'insolubilité ne serait pas distinguable dans les conditions établies ce qui pourrait être un biais du système rapporteur. L'activité de ce candidat ne pouvant être évaluée par le système, celui-ci apparaîtrait comme un candidat négatif donc ne présentant aucune SUMOylation sur le substrat au même niveau que les contrôles négatifs. Le critère de taille, soit un maximum de 1500 acides aminés, correspond également au clonage de gènes d'une taille maximum de 4500 paires de bases. Cette limite est également bénéfique pour faciliter le clonage des candidats à partir d'une banque d'ADNc issus d'ARN. L'ARN est en effet une molécule instable et facilement dégradée par une panoplie de ribonucléase, l'extraction d'ARN est une tâche difficile (Dastgheib et al, 2014; Hernandez et al, 2009; Chrigwin et al, 1979; Potenza et al, 2006). Pour cela, plus l'ARN est petit et plus les chances d'obtenir le candidat sont grandes. Pour extraire l'ARN, la méthode d'extraction par TRIzol a été la méthode la plus efficace et ayant permis l'obtention de la plupart des candidats. Cette méthode, comparée aux kits commerciaux d'extraction d'ARN, permet d'obtenir une plus grande quantité d'ARN, tout en sacrifiant la qualité de celui-ci (Tesena et al., 2017). Pour compenser pour ce niveau plus élevé de matériel non voulu, un nettoyage des produits de PCR peut être faite avec le matériel amplifié permettant ainsi de rectifier cette basse qualité de matériel. Pour les candidats qui sont intéressants mais qui n'ont pas pu être obtenus par RT-PCR, d'autres lignées cellulaires exprimant ces candidats auraient pu être utilisées ou ceux-ci auraient pu être commandés via des plateformes tels que Integrated DNA Technologies, GenScript, OriGene ou Twist Bioscience. Concernant ces candidats, la tentative d'obtenir uniquement une portion du gène n'était pas envisageable. Bien qu'AlphaFold puisse suggérer une région ayant une activité de Ligase, cette indication peut être erronée. Par conséquent, la portion du gène identifiée pourrait représenter seulement une petite partie du domaine responsable de l'activité de Ligase, et cela pourrait ne pas être signalé par AlphaFold. Ainsi, l'activité Ligase de ce candidat pourrait ne pas être détectée lors

des essais *in vitro* et *in bactério*. Pour ce qui est des autres candidats ayant été obtenus par extraction d'ARN, seulement RBMX n'a pas pu être purifiée. Cependant, cette protéine a été purifiée par un groupe auparavant sans précision sur les méthodes employées (Matsunaga et al., 2012) donc cet article n'aurait pas pu être utilisé pour la purification de RBMX. Pour les autres candidats, aucune mention de littérature n'a lieu de groupe de recherche ayant purifié ceux-ci. Pour obtenir ces candidats solubles et purifiés, il faudrait extraire les protéines d'un lysat cellulaire afin d'avoir la protéine avec les modifications post-traductionnelles eucaryote. Par la suite, des tests *in vitro* pourraient suivre afin de confirmer la présence ou absence d'activité SUMO E3 ligase. En raison de contraintes budgétaires, l'achat des gènes nécessaires et protéines préalablement purifié n'a pas été effectué.

En ce qui concerne l'autoSUMOylation des SUMO E3 Ligases, il existe peu de littérature sur la raison de cette automodification intrinsèque. Cependant, l'hypothèse existante est que, comme l'ubiquitine, l'automodification des SUMO E3 Ligases permettrait d'automoduler son activité (Ihara et al., 2005; Ivanov et al., 2007; Yohan et al., 2013). Les protéines SUMO E3 Ligase sont donc des protéines qui s'autoSUMOylent. Parallèlement, il est possible d'observer 4 SUMO E3 Ligases connus (figure 3.2) ayant une forte SUMOylation (supérieur à 20 % de lysines SUMOylées par rapport au nombre de lysines totales). Donc, les probabilités que ces protéines soient des protéines hautement SUMOylées sont également hautes. De plus, les lysines se retrouvant dans un site consensus ne sont pas les seuls pouvant se faire SUMOylées. Toutes les lysines, si accessibles, pourraient se faire SUMOylées en cas de stress comme est le cas de certaines études répertoriée par Hendricks et Vertegaal qui impliquait l'utilisation d'un inhibiteur de protéasome (MG132) ou l'induction de stress thermique pour induire une SUMOylation. (Hendriks et al., 2016; Cubenas-Potts et al., 2015; Hendriks et al., 2014; Henley et al., 2014; Johnson et al, 1997; Hendriks et al., 2017; Psakhye et al. 2012) Pour diminuer le nombre de protéines retenues à un chiffre plus raisonnable, une limite d'au moins 20 % des lysines SUMOylées a été imposée. Tel que mentionné précédemment, ce critère arbitraire a été imposé afin de s'assurer qu'une discrimination ait lieu entre les deux populations de protéines, soit les protéines SUMOylées présentant moins de 20 % de leur lysines SUMOylées (concentration de plusieurs points) et une deuxième population de protéines présentant au moins 20 % de lysines SUMOylées (points plus réparties). Ce critère permet de diminuer la liste de candidats potentiels à 161 protéines. Cette limite établie (Lysine

SUMOylées/lysines totales) permet notamment de différencier les protéines n'ayant probablement pas d'activité SUMO E3 Ligase (moins de 20 % de lysines SUMOylées/lysines totales) contre les protéines ayant probablement une activité SUMO E3 Ligase (plus de 20 % de lysine SUMOylées/lysines totales) (Figure 3.2). Les protéines ayant un rapport inférieur à 20 % sont principalement les protéines simplement SUMOylées qui auraient une autre fonction. Ces changements de fonction se manifestent par un changement de partenaire d'interaction ou un changement de compartiment cellulaire à la suite de la SUMOylation (Tirard et al., 2012). La pertinence du critère « ratio de lysines SUMOylées/lysines total » est confirmée par la présence de 7 SUMO E3 Ligases parmi les 10 connues qui présentent au moins 20 % de leurs lysines SUMOylées soit PIAS1, PIAS2, PIAS3, PIAS4, CBX4, NSE2 et ZNF451 (Kagey et al., 2003; Kahyo et al., 2001; Sachdev, 2001; Potts et al., 2005; Cappadocia et al., 2015; Hendricks et al., 2016). Par la suite, les protéines possédant au moins un SIM ont été choisies parmi les protéines potentielles pour 2 raisons : tous les SUMO E3 ligases ont des SIMs en dépit de leur mécanisme de SUMOylation (Cappadocia et al., 2017) et le but de cette recherche est de découvrir des SUMO E3 Ligases atypiques qui fonctionnent de manière conceptuellement similaire à ZNF451. À partir des 132 protéines restantes, 2 criblages ont été faits : un premier par prédiction d'interaction avec la E2 chargée avec une SUMO2/3 ont été faits grâce à la plateforme de prédiction de structure AlphaFold2 (Jumper et al., 2021) et un deuxième en filtrant les protéines ayant une interaction avec la E2 d'après la banque de données BioGrid suivie d'un criblage AlphaFold pour le choix de candidats. Ce dernier critère de sélection est important, car la E3 est une protéine qui agit en aval de la E2 et qui va rapprocher le substrat et la E2 pour effectuer sa fonction de ligase. Concernant la détection des SIMs, il est important de noter que les séquences identifiées pourraient se retrouver dans des régions de la protéine qui ne sont pas accessibles. À cet égard, il aurait été judicieux de tester l'accessibilité au solvant de ces régions, par exemple en utilisant des outils de prédiction de structure tels que AlphaFold. Cela dit, l'évaluation de l'accessibilité des SIMs est complexe. Par exemple, certaines protéines possèdent plusieurs conformations, et certaines conformations peuvent présenter un SIM dans une conformation accessible tandis que le SIM ne sera pas accessible dans une autre conformation. Au niveau des interactions, il faut noter que toutes les protéines qui interagissent avec l'enzyme E2 ne sont pas nécessairement des SUMO E3 ligases. De plus, les interactions répertoriées sur BioGRID peuvent être indirectes. Par exemple, les

protéines détectées au moyen d'expériences d'immunoprécipitation peuvent ne pas interagir directement entre elles, mais simplement appartenir à un même complexe protéique.

Il est important de mentionner que notre criblage bio-informatique présente d'emblée un certain nombre de limitations : pour chacune des protéines, nous ne nous sommes concentrés que sur l'isoforme principal rapporté par Uniprot. Certaines protéines présentent de nombreux isoformes (par exemple, la protéine PML), mais la présence de ces formes additionnelles aurait compliqué les étapes subséquentes (notamment le clonage des gènes à partir de bibliothèques d'ADN complémentaires). Nous n'avons pas non plus tenu compte des protéines alternatives étant donné que les informations disponibles sur ces protéines sont malheureusement trop rares.

Par ailleurs, le criblage AlphaFold est fait afin de s'assurer de l'utilisation de mécanismes probables de la protéine candidate avec la E2-SUMO. Pour la protéine ZNF24, certains modèles de prédiction provenant d'AlphaFold suggèrent des interactions principalement entre ZNF24 et SUMO2/3 avec peu d'interaction avec la E2. Il semblerait que ZNF24 aurait 2 mécanismes possibles pour la SUMOylation de substrats. Le premier mécanisme potentiel étant l'utilisation du SIM 2 pour une interaction avec SUMO2/3 (Figure 3.7 O) permettant de rapprocher SUMO sur les lysines accessibles étant sur le substrat ou celle sur le candidat directement. Cependant, seulement le SIM 2 semble impliquer dans cette interaction d'après AlphaFold. Cependant, la présence du SIM 2 n'est pas suffisante pour avoir une activité de SUMOylation, car un SIM, bien que présent chez tous les candidats, n'est pas suffisant pour avoir une activité de SUMO E3 Ligases atypique ressemblant à ZNF451 (tels que visible avec les candidats négatifs). Deux lysines accessibles se retrouvant en amont du SIM, K22 et K27, lorsque SUMOylée, pourraient venir mimer un second SIM par une interaction avec la E2 similaire à une interaction SUMO-SIM (Cappadocia et al., 2015). Le groupement SUMO lié à la lysine permettrait de se lier à la région dorsale de E2~SUMO tandis que SIM1 maintiendrait E2~SUMO dans une conformation activée. Cette interaction permettrait donc de recruter plus facilement une E2, facilitant la SUMOylation de substrats se retrouvant à proximité (Figure 3.8 A). Un second mécanisme moins fiable proposé par AlphaFold serait une stabilisation de SUMO2/3 par des interactions transitoires avec les régions de Zinc Fingers permettant ainsi la SUMOylation de lysines proches et accessibles (Figure 3.7 P). Avec les informations fournies par AlphaFold, l'hypothèse est que l'interaction avec SUMO2/3 se retrouvant

dans le site actif de la E2 lie le SIM2 exposé au solvant en N-terminal et une lysine SUMOylées en aval de ce SIM (K22 ou K27) mimerait un second SIM en liant le site dorsal de la E2 donnant ainsi une activité de SUMO E3 Ligase similaire à ZNF451, mais avec un mécanisme d'action légèrement différent. Afin de confirmer le mécanisme d'action de ZNF24, des mutations devraient être introduites au site d'interaction avec la E2. Pour le candidat IF5A1, les interactions prédites par Alphafold sont sur une hélice alpha se retrouvant en N terminal de la E2 avec le SIM 2 du candidat. Ces interactions prédites se retrouvent à côté de la cystéine du site actif de la E2 ce qui met IF5A1 à proximité du site actif permettant une SUMOylation de substrats ou de la protéine elle-même (Figure 3.8 J). La protéine candidate a été choisie pour sa composition en structure secondaire enrichie en feuillet  $\beta$  permettant une interaction avec la E2 ainsi que l'interaction du SIM 2 avec la E2 (Figure 3.7 C). De plus, la protéine étant composée principalement d'acides aminés acides et hydrophobes, la protéine était prédite d'avoir un fort potentiel SUMO E3 Ligase (Figure 3.7 C). Pour la protéine HNRPC, le domaine de RNA Recognition Motif prédit par Alphafold ressemble structurellement à SUMO2/3 avec un RMSD de 2.8 indiquant une moyenne de déviation de 2.8 Ångström entre SUMO2/3 et le domaine de HNRPC (plus le nombre est bas et plus les structures se ressemblent) (Figure 5.1). De plus, les prédictions entre le domaine SUMO de HNRPC et la E2 semblent indiquer des interactions presque identiques aux interactions de la partie dorsal de SUMO2/3 avec la E2 (Figure 3.7 N). Également, le SIM 2 du candidat aurait également une interaction feuillet  $\beta$  anti-parallèle avec SUMO2/3 se retrouvant dans le site actif de la E2 (Figure 3.7 M). Ces évidences bio-informatiques indiqueraient un mécanisme d'action de SUMOylation similaire à celui de ZNF451 avec le SIM 2 qui lie SUMO stabilisant SUMO du site actif de la E2 ainsi qu'un domaine ressemblant à SUMO qui lie la région dorsale de la E2 (Figure 3.8 B). Pour le candidat ZC3HE, peu de modèles d'Alphafold semblent indiquer une interaction entre le candidat et la machinerie, mais le SIM 1 semble interagir par des feuillets bêta anti parallèle avec SUMO2/3 (Figure 3.7 A). De plus, la protéine n'ayant pas domaine intrinsèquement ordonné, ce candidat a été ajouté à la liste de candidat. Ces domaines, qui sont prédits intrinsèquement désordonnés, pourraient être assez malléables permettant l'interaction des SIMs avec la machinerie de SUMOylation. Ces interactions, par le biais de SIMs, comme ZNF451, pourrait stabiliser SUMO se retrouvant dans le site actif de la E2 et permettre d'apposer SUMO sur le substrat (Figure 3.8 C). Pour le candidat DAXX, les interactions entre son SIM 2 et SUMO sont bien répertoriés dans la littérature (Lin et al., 2006; Chang et al., 2011). Cependant, aucune étude ne considère les

SIMs comme potentielle activité SUMO E3 Ligase ressemblante celle de ZNF451. De plus, l'interaction de feuillet  $\beta$  antiparallèle du SIM 1 de DAXX avec SUMO (Figure 3.7 L) ainsi que des lysines se retrouvant en amont du SIM 1, pourraient conférer à DAXX une activité de SUMO E3 Ligase (Figure 3.8 D). Pour la protéine TYDP2, deux mécanismes probables peuvent être mis avec le premier étant le SIM 1 qui interagit avec SUMO par le biais de feuillet  $\beta$  anti parallèle qui positionne SUMO proche d'une lysine de la protéine (Figure 3.7 K) avec l'utilisation d'une amorce qui viendrait stabiliser la E2-SUMO par des interactions électrostatiques (Figure 3.7 J). Par la suite, comme ZNF24, un SIM et une lysine SUMOylée (K82 ou K99) engendreraient une activité SUMO E3 Ligase. Ceci par le groupement SUMO lié à la lysine permettrait de se lier à la région dorsale de E2~SUMO tandis que SIM1 maintiendrait E2~SUMO dans une conformation activée. Le second mécanisme possible est le recrutement de SUMO par le candidat via son SIM 1 ainsi que le recrutement d'une deuxième SUMO grâce au SIM 2 du candidat permettant une interaction avec la région dorsale de la E2 de façon similaire au module catalytique de ZNF451 (Cappadocia et al., 2015) (Figure 3.8 E). Pour la protéine BLM, certaines évidences (Zhu et al., 2008; Böhm et al., 2014) suggèrent que BLM contient des SIMs qui sont importants pour un recrutement aux corps PML SUMOylées des cellules ou BLM accomplit ses fonctions (de Thé et al., 2012). Certaines évidences suggèrent que BLM se fait SUMOylé par la SUMO E3 Ligase NSE2 pour jouer un rôle au niveau des fourches de réplication durant certains stress (Pond et al., 2019) expliquant son haut taux de SUMOylation. La protéine BLM fut tout de même choisie comme candidat pour son SIM 1 qui interagit, d'après Alphafold, avec SUMO par des brins  $\beta$  parallèles (Figure 3.7 B). Cette interaction, ainsi que les lysines accessibles K225, K259, K271, K272, K273 pourraient également se faire SUMOyler mimant un deuxième SIM facilitant ainsi le recrutement de E2~SUMO et son maintien dans une conformation activée. Ce mécanisme, comme énoncé auparavant, viendrait mimer le mécanisme de ZNF24, permettant une activité SUMO E3 Ligase (Figure 3.8 F). Pour la protéine RBMX, un domaine en N terminal ressemblant à SUMO, soit le domaine RNA Recognition Motif, ressemble structurellement à SUMO2/3. Cette structure, en N-terminale du candidat et présentant des structures secondaires similaire à SUMO2 et SUMO3, présente un RMSD avec SUMO2 de 5.6 Ångström (Figure 5.1). Le domaine de RBMX viendrait interagir avec une partie en C terminal de la E2 qui est différente de l'interaction entre ZNF451 et la E2 (Figure 3.7 D). Cette interaction permettrait de recruter la E2 par le biais d'interaction dans la région dorsale de celle-ci permettant l'autoSUMOylation du candidat. Cette protéine est

intéressante pour cette étude, non pas pour son interaction potentielle avec la E2, mais pour son domaine RRM qui ressemble à SUMO et qui permet une interaction avec la région dorsale de la E2 (Figure 3.8 G). Pour NF2IP, cette protéine contient, en C terminal, 2 motifs qui ressemblent à l'ubiquitine qui permet de lier la région dorsale de la E2, similairement à ZNF451 (Figure 3.7 F). Le candidat, grâce à ces deux motifs particuliers, pourrait présenter une activité de SUMO E3 Ligase atypique, reposant sur un mécanisme qui n'a pas encore été caractérisé chez les SUMO E3 Ligases. De plus, le SIM 1 de ce candidat semble également recruter SUMO se retrouvant dans le site actif de la E2 donnant ainsi le mécanisme suivant : le SIM 1 lie SUMO se retrouvant dans le site actif de la E2 et le domaine SUMO-like, en C-terminal, lie la région dorsale de la E2 et force la E2 à relarguer le SUMO se retrouvant dans son site catalytique sur un substrat (Figure 3.7 F, Figure 3.8 H). Pour le candidat PML, le SIM 2 est celui qui lie SUMO se retrouvant dans le site actif de la E2 (Figure 3.7 H). Comme précédemment énoncé, le mode d'action potentiel est l'utilisation d'une lysine se retrouvant proche dans l'espace du SIM, soit la lysine K651, qui lie la E2 dans la région dorsale tandis que SUMO se retrouvant dans le site catalytique de la E2 est stabilisée par le SIM permettant l'apposition de SUMO sur un substrat. Par la suite, la protéine contribue à la stabilisation de la E2 et SUMO par des interactions électrostatiques (Figure 3.7 G, I) et pourrait contribuer au relargage de SUMO lorsque celle-ci se retrouve dans une conformation activée.

Le criblage AlphaFold contient cependant de nombreuses limitations dont, l'absence de modifications post-traductionnels ou cofacteurs essentiels lors de la prédiction, la difficulté à prédire, l'incapacité à modéliser les régions flexibles et la difficulté de prédire des protéines ayant peu d'homologues protéiques (Bertoline et al., 2023). De plus, AlphaFold est capable de prédire plusieurs modèles pouvant aller de 1 à plusieurs. Dans notre cas, 5 modèles ont été générés et ceux-ci appartiennent à une seule conformation de protéine ne montrant pas la motilité des résidus de la protéine et les différentes conformations adoptées par ces protéines (Bertoline et al., 2023). Les contraintes indiquées ci-haut affectent le criblage de SUMO E3 Ligases, car les SIMs, dans les cellules, se retrouvent principalement dans des régions intrinsèquement désordonnées. Il est également possible que l'interaction prédite soit simplement une réaction de SUMOylation de la protéine candidates par la E2 et non une activité SUMO E3 ligases. Étant donné qu'AlphaFold utilise des protéines résolues expérimentalement, il se base sur les protéines ayant été

cartographiées pour prédire celles qui ne l'ont pas été (Jumper et al., 2021). Considérant que seulement 11 SUMO E3 Ligases ont été découverts à ce jour et que seulement 3 ont un mécanisme d'action atypique, il se peut qu'Alphafold ait de la difficulté à déterminer des protéines ayant un mécanisme atypique comme celui de ZNF451 et RANBP2. Un mécanisme plus prévalent, soit la SUMOylation de protéine par la E2, est plus présent dans la littérature. Pour cela, il est plus facile de prédire une SUMOylation de protéine par la E2 qu'un mécanisme de SUMOylation par le complexe candidat-E2-SUMO. Pour pallier à ces contraintes, des tests d'activité ont été effectués par le biais de tests *in bactério* et ensuite validés par tests *in vitro*.

Une limitation de notre criblage bio-informatique est notamment l'utilisation de données de spectrométrie de masse qui dépend de l'abondance de protéine pour une identification efficace (de Godoy et al., 2006). L'abondance des ARNs des protéines candidates ainsi que les lignées cellulaires les exprimant sont prises en compte lors du choix de candidat. Les lignées cellulaires qui nous étaient accessibles sont les cellules HEK293T ainsi que les cellules HeLa. Pour ceci, la plateforme The Human Protein Atlas ([www.proteinatlas.org](http://www.proteinatlas.org)) a été choisie pour comparer les abondances de transcripts entre ces deux lignées donnant les données les plus abondantes suivantes : 231.7 nTPM (normalized transcripts per million) pour RBMX avec les cellules HeLa, 43.0 nTPM pour ZC3HE avec les cellules HeLa, 56.2 nTPM pour NF2IP avec les cellules HeLa, 49.9 nTPM pour TYDP2 avec les cellules HEK293, 19.2 nTPM pour BLM avec les cellules HEK293, 23.6 nTPM pour PML avec les cellules HEK293, 37.1 nTPM pour DAXX avec les cellules HEK293, 862.5 nTPM pour HNRPC avec les cellules HEK293 et 936.7 nTPM pour IF5A1 avec les cellules HeLa. Pour les échantillons de protéines, les échantillons de lysat de cellules humaines utilisées pour la détection de SUMOylation dans la revue de littérature (Hendriks et al., 2016) ont été enrichis en grâce à des purifications préalables (pulldown, immunoprécipitation ou purification par étiquette His6 de SUMO recombinant), il reste possible que certains gènes ou sites SUMOylés soient tout de même non détectables (Garbis et al., 2005). SILAC, qui est une des principales techniques utilisées pour les études protéomiques répertoriant les protéines SUMOylées, est une technique d'ajout d'acide aminé isotopique stable dans une culture cellulaire pour son incorporation dans les protéines. Cette technique est la méthode la plus adaptée pour les modifications post traductionnels (Lai et al., 2014). En comparant avec les autres méthodes de protéomiques avec étiquettes, tel que iTRAQ, ICAT ou même l'introduction d'amine

isotopique 15, SILAC est plus sensible et permet une meilleure couverture du protéome à l'étude. (Chen et al., 2015)

Parmi les 3 protéines RBMX, NF2IP et HNRPC contenant un domaine similaire à SUMO, seulement NF2IP a été sujet d'une étude pour caractériser son domaine SUMO-like (Novatchkova et al., 2005; Hertz et al., 2023). Outre NF2IP, très peu de littérature existe pour les protéines possédant des domaines SUMO-like. Il est fort probable que ce domaine joue un rôle biologique au-delà d'une activité potentielle de ligase comme son homologue : le domaine Ubiquitin-like. Ce domaine Ubiquitin-like, dans le cas de RAD23 et PARK2, joue un rôle de moduler le niveau d'ubiquitination des autres protéines (Su et al., 2009; Gurzau et al., 2021; Gladkova et al., 2018). Il est possible que le domaine SUMO-like joue un rôle similaire pour les protéines : ce domaine permettrait de moduler l'activité de SUMOylation des protéines ayant ce domaine (autoSUMOylation) par le recrutement de la E2 (Prudden et al., 2009) et permettrait davantage la SUMOylation de substrat. Pour HNRPC, le domaine SUMO-like (le domaine RRM) prédit se superpose avec SUMO2/3 ayant un RMSD de 2.8 Å ce qui veut dire que 2.8 Ångströms est la distance moyenne entre les atomes des domaines SUMO-like et SUMO2/3 (Figure 5.1). Pour RBMX, cette valeur se retrouve à 5.6 Å pour le domaine SUMO-like (Figure 5.1). Certes, ces valeurs pour tous les atomes des structures sont élevées, mais ceci pourrait être dû à la prédiction AlphaFold qui ne prend en considération qu'une seule conformation du domaine des protéines.

Parmi les candidats, nombreux sont ceux avec des domaines zinc fingers soit PML, ZNF24 et ZC3HE. Ce domaine d'interactions à l'ADN semble jouer un rôle dans la contribution à l'interaction avec la E2-chargee de SUMO2/3. Par ailleurs, 7 des 11 SUMO E3 ligases contiennent des domaines liant le zinc, soit Zinc finger ou domaine SP-Ring, (ZFN451, EGR2, PIAS1/2/3/4, NSE2) ce qui pourrait indiquer que la fonction de ligase pourrait être facilitée grâce à ces domaines ou que les ligases sont requises autour de l'ADN (Hochstrasser, 2001). En effet, la SUMOylation étant requise lors de réponses au dommage à l'ADN (Jackson et al., 2013; Ulrich, 2014), les SUMO E3 Ligases auraient, par besoin, une localisation proche de l'ADN.

## 4.2 Test *in bactério* et *in vitro* des candidats

Afin de tester l'activité des candidats potentiels, des tests ont eu lieu qui permettent de cribler l'activité SUMO E3 Ligases de façon rapide et efficace. Un de ces tests consiste en l'utilisation d'un système rapporteur couplé au système pSUMO1/2 qui reconstitue la cascade enzymatique (E1, E2) avec SUMO1 ou SUMO2/3 *in bactério* (Weber et al., 2014). Les vecteurs pSUMO1 et pSUMO2/3 ont été créés par le laboratoire Schaer pour faciliter l'obtention de protéine SUMOylées (Weber et al., 2014). Dans notre cas, nous avons adapté ce système pour tester l'activité SUMO E3 ligase de candidats protéiques. Les vecteurs pSUMO1 et pSUMO2 sont essentiels pour ce criblage, car les procaryotes ne possèdent pas les enzymes nécessaires à la cascade signalétique pour la SUMOylation (Iribarren et al., 2015). Ceci permet de tester l'activité du candidat sans interférence de SUMOylation endogène par la bactérie. Ce criblage n'aurait pas été possible dans les cellules humaines telles que HEK293 ou HeLa considérant que ceux-ci sont eucaryotes et possèdent à la fois les enzymes nécessaires pour SUMOyler (E1, E2 et E3 endogène), mais aussi déSUMOyler (SUMO protéase SENP) donc une interférence due à la SUMOylation et déSUMOylation endogène aurait eu lieu. Une fois induites, les bactéries qui présentent des bandes de haut poids moléculaire suggèrent une SUMOylation, confirmation qui est renforcée par l'activité de la SUMO protéase ULP1, laquelle déSUMOyle le substrat. À ce type de test, il nous a été possible de constater que ZNF24 semble avoir une activité SUMO E3 ligase comparable à celle de ZNF451. Les autres candidats ne semblent pas présenter d'activité notable durant les tests. Cependant, pour les protéines n'ayant pas d'activité SUMO E3 Ligase aperçu pour le criblage *in bactério*, il se peut que ce soit dû au manque de modifications post-traductionnelles présentes chez l'humain que la bactérie est incapable de reproduire (Shi et al., 2020). D'autres limitations pouvant influencer le système rapporteur *in bactério* sont également les niveaux d'expressions de protéines, la solubilité des protéines, etc. Pour la phosphorylation, la phosphorylation de serine, thréonine et tyrosine autour des SIMs peut se révéler primordiale pour une bonne liaison des SIMs avec SUMO. Les bactéries possèdent des enzymes capables de phosphoryler les arginines, acides aspartiques, histidine et cystéine pouvant influencer différents acides aminés environnant les SIMs. Cette modification a lieu différemment chez les procaryotes que les eucaryotes ce qui pourrait exercer une influence sur une potentielle activité SUMO E3 ligase (Shi et al., 2020). Pour l'expression de protéines recombinantes dans la bactérie, de nombreux problèmes peuvent survenir dont la solubilité des candidats et leur expression. Considérant le système d'expression, soit le système

pBad, l'expression de protéine par induction a l'arabinose n'est pas optimale due à son catabolisme ou déplétion du milieu par les bactéries ce qui peut causer une diminution de l'expression protéique recombinante réguler par l'arabinose (Crozier et al., 2021). De plus, peu de contrôle a lieu lors de ce type d'induction : une diffusion d'arabinose dans une bactérie peut être plus importante que dans une autre ce qui pourrait causer une expression plus élevée dans celle ayant plus d'arabinose. Une expression plus élevée de protéines recombinantes peut conduire à leur agrégation (Schramm et al., 2020). Dans le cadre des techniques de gel SDS-PAGE et d'immunobuvardage, il a été observé que les agrégats protéiques correspondent aux poids moléculaires anticipés (Bondos et al., 2003). Cette observation pourrait compromettre l'efficacité des tests *in bacterio*, car les protéines agrégées peuvent se faire SUMOylé par la E2. Par conséquent, cela pourrait entraîner des résultats faussement négatifs. Afin d'essayer de pallier ces limitations, il faut confirmer l'activité de candidats par des tests *in vitro*. Ces tests nécessitent l'utilisation d'enzymes purifiées de la cascade enzymatique de la SUMOylation, soit SUMO2/3, la E1, la E2 et le candidat (Werner et al., 2009; Flotho et al., 2012). Grâce à cette méthode, il a été possible de confirmer davantage l'activité de ZNF24 qui possède une activité SUMO E3 Ligase modérée comparable à celle du SUMO E3 Ligase atypique ZNF451. ZNF24 est un facteur de transcription liant l'ADN pour la répression de la transcription du gène VEGF par une liaison du promoteur proximal de celui-ci (Jia et al., 2013) et de PDGFR (Li et al., 2010) et stimule la production de  $\beta$ -caténine (Liu et al., 2012). Il se peut que ZNF24 permette la répression de certains gènes par la SUMOylation de certaines protéines dans les régions proches de sa fonction, soit la SUMOylation de protéine proche des promoteurs de VEGF, PDGFR et des  $\beta$ -caténines permettant sa fonction. ZNF24 semble avoir, d'après BioGrid et les données de SUMOylation (Hendricks et al., 2015), plusieurs interacteurs pouvant être SUMOylés tels que ZSCAN32, ZNF446, PGBD1, ZNF174, ZNF444, ZKSCAN4, ZNF397, ZSCAN20, ZSCAN21, JMJD6, MZF1, PARP1, ZKSCAN1, ZKSCAN8, ZNF263. La majorité de ces substrats ont une fonction effectrice sur l'ADN et le substrat présentant le taux de SUMOylation le plus élevée parmi les interacteurs de ZNF24 est PARP1 qui a été étudiée de façon exhaustive pour son rôle dans le dommage à l'ADN (Pascal, 2018; Xie et al., 2020). Il est possible de noter que la SUMOylation de PARP1 a été attribuée à la SUMO E3 Ligase PIASy dans des conditions de choc thermique (Martin et al., 2009). Cependant, il est possible que pour autres évènements cellulaire, tel que des dommages double brin de l'ADN, un relais ait lieu entre PIASy et ZNF24. Ce dernier pourrait SUMOyler les protéines situées à proximité de son site d'action en tant que

facteur de transcription, c'est-à-dire lié à l'ADN, et ainsi SUMOyler les protéines à doigt de zinc, ainsi que PARP1. ZNF24 est une protéine ne possédant aucun orthologue en dehors de la souris, rat et certaines espèces de singes. L'absence de conservation de ZNF24 au sein de nombreuses espèces suggérant un rôle biologique unique au sein de ces organismes. La même observation s'applique à la majorité des SUMO E3 ligases atypiques, telles que ZNF451 et RANBP2, qui semblent également être limitées à un petit nombre d'organismes. Ceci suggère que les ligases SUMO E3 atypiques pourraient très probablement représenter une caractéristique nouvellement acquise grâce à l'évolution. Cependant, chez les espèces exprimant ZNF24, cette protéine est exprimée dans la plupart des tissus.

## CONCLUSION ET PERSPECTIVES

Dans cette étude, nous avons entrepris un criblage bio-informatique rigoureux pour identifier de nouvelles SUMO E3 Ligases atypiques, aboutissant à la découverte de 50 protéines candidates. Après un processus de sélection basé sur des critères spécifiques tels que l'annotation nucléaire, la présence de sites de SUMOylation et la capacité à interagir avec la machinerie de SUMOylation, 11 candidats ont été retenus pour des tests d'activité SUMO E3 Ligases *in vitro* et *in bactério*. Cette analyse a permis de mettre en évidence un candidat prometteur, notamment ZNF24, qui montre une activité SUMO E3 Ligase modérée. Avec cette étude, il est maintenant possible d'envisager une activité principalement avec 1 SIM et 1 lysine SUMOylée. Notre étude a également révélé les défis liés au criblage de protéines SUMO E3 Ligases, incluant les limitations des données bio-informatiques et les spécificités de l'expression protéique dans les systèmes bactériens. Malgré ces obstacles, l'approche adoptée a permis d'identifier une méthode pour valider une activité de Ligase et identifier une protéine avec une activité SUMO E3 Ligase comparable à celle de ZNF451, ouvrant la voie à de futures recherches sur la fonction biologique et mécanisme d'action de ZNF24. Le candidat nouvellement identifié, ZNF24, employant un mécanisme atypique, similaire à celui de ZNF451, n'est probablement pas la seule protéine avec ce type de mécanisme. Ce mécanisme particulier pourrait également être régulé par autres MPT comme l'ubiquitination des lysines SUMOylées requises pour la SUMOylation. Pour les candidats ne présentant pas d'activité SUMO E3 ligase, il faudrait extraire les protéines d'un lysat cellulaire via pulldown afin d'avoir les protéines avec les modifications post-traductionnelles eucaryote. Par la suite, des tests *in vitro* pourraient permettre de confirmer la présence ou absence d'activité SUMO E3 ligase. Il serait pertinent d'inclure les protéines alternatives dans un futur travail de criblage de candidats en incluant les séquences de chromosomes dans ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>) et modéliser les protéines par Alphafold avec la machinerie de SUMOylation. Considérant le peu d'informations existantes sur les protéines alternatives, il serait difficile de retrouver leur niveau de SUMOylation par protéomique. Cependant, ceci permettrait de retrouver davantage de SUMO E3 Ligases ayant des mécanismes atypiques.

Finalement, notre étude contribue à la compréhension des mécanismes de SUMOylation et à l'identification de nouvelles SUMO E3 Ligases, enrichissant ainsi notre connaissance des processus

cellulaires régulés par la SUMOylation. Les découvertes, notamment l'activité de SUMO E3 Ligase de ZNF24, soulignent l'importance de la SUMOylation dans la régulation transcriptionnelle et la réponse au dommage de l'ADN, et ouvrent de nouvelles pistes pour la recherche sur le rôle des SUMO E3 Ligases dans divers contextes biologiques et pathologiques.

## ANNEXE A

CODE1 :

```
function = "C"
input_file="Localisation fichier "
with open(input_file, "r") as f:
    rawdata = f.read()
count=0
protein_localization_dict={}
localization_protein_dict={}
protein_sequence_dict={}
ID_sequence_dict={}
accession_to_ID_dict={}
for line in rawdata.split("\n"):
    if line.startswith("ID"):
        ID=line[5:].split(" ")[0]
        count+=1
        sequence=""
    if line.startswith("AC"):
        AC=line [5:].replace(" ", "").split(";")[0:-1]
        for singleAC in AC:
            accession_to_ID_dict[singleAC]=ID
    if line.startswith("DR  GO;"):
        if line[21]==function:
            GO=line [23:].split(";")[0]
            if ID in protein_localization_dict:
                protein_localization_dict[ID].append(GO)
            else:
                protein_localization_dict[ID]=[GO]
```

```
if GO in localization_protein_dict:
    localization_protein_dict[GO].append(ID) #ID
else:
    localization_protein_dict[GO] = [ID] #ID
if line.startswith(" "):
    sequence+=line
if line.startswith("//"):
    for myAC in AC :
        protein_sequence_dict[myAC] = sequence.replace(" ", "")
    ID_sequence_dict[ID]=sequence.replace(" ", "")
```

## ANNEXE B

CODE2:

```
target_localizations = ['nucleus']
localizations = []
nuclearprotslist=[]
for localization in target_localizations:
    localizations+= localization_protein_dict[localization]
    print("Il y a % i protéines possédant la localisation \"% s\" %
(len(localization_protein_dict[localization]),localization))
proteins=sorted(set(localizations))
print("Sans doublons, % i protéines possédant les localisations demandees\n\" % (len (proteins)))
for protein in proteins:
    name=\"% s.fasta\" % (protein)
    sequence=ID_sequence_dict[protein]
    elem=">% s\n%s » % (protein,ID_sequence_dict[protein])
    nuclearprotslist.append(name)
print("Ecriture du fichier \"% s'. La longueur de la séquence est de % s residus\n » %
[name,len(sequence)])
print(elem+"\n")
```

## ANNEXE C

CODE3 :

```
nucleasumoylatedlist=[]  
for protAC in Lysinesprotlist :  
    if protAC in nuclearprotlist:  
        nucleasumoylatedlist.append(protAC)
```

## ANNEXE D

CODE4:

```
from collections import Counter
cnt=Counter(nucleasumoylatedlist)
ACprotcounted=cnt.most_common()
print(ACprotcounted)
proteinnames=[]
seqlengths=[]
allfracs=[]
for pprot in v:

    try:
        contents = urllib.request.urlopen("https://www.uniprot.org/uniprot/" + pprot +
        ".fasta").read()
        #print(contents)
        seq="" .join((str (contents). split (« \n ») [1 :]))[0:-1]
        header=str(contents).split(« \n »)[0][2:]
        protein_name = (((str (contents)).split(« \n »)[0]).split("|")[2]).split(" ")[0]
        lysinecount=seq.count("K")
        sequence_length= len(seq)
        for ppprot in ACprotcounted:
            if pprot==ppprot[0]:
                SUMOlys=ppprot[1]
                fractionlysines=float(SUMOlys/lysinecount)
                proteinnames.append(protein_name)
                allfracs.append(fractionlysines)
                sequence_length= len(seq)
                seqlengths.append(sequence_length)
```

```
except IndexError:
    print("This protein "+pprot+" has a problem\n")
cleanSUMOnucprots=[]
for i in range(len(proteinnames)):
    x=seqlengths[i]
    y=allfracs[i]
    name=proteinnames[i][0:-6]
    if x<1500 and y>0.2:
        cleanSUMOnucprots.append(name)
```

## ANNEXE E

CODE5 :

```
listOfGenes=["UBC9"]
fastaLocation=Espace pour recevoir les fichiers fasta
filterByThroughput = False
minimalNumberOfObservations = 1
import requests as r
from Bio import SeqIO
from io import StringIO
import os
from collections import Counter
import pandas as pd
import time
start_time = time.time()
print("Debut de l'analyse")
def getFASTA(cID,fastaLocation):
    baseUrl="http://www.uniprot.org/uniprot/"
    currentUrl=baseUrl+cID+". fasta »
    response = r. post (currentUrl)
    cData= ". join (response.text)
    fname=fastaLocation+cData.split (« ») [0]. split (« | »)[— 1]+".fasta"
    #print(fname)
    #fname=fastaLocation+"%s.fasta" % (cID)
    with open(fname, "w") as f:
        f.write(cData)
    Seq=StringIO(cData)
    pSeq=list(SeqIO.parse(Seq,'fasta'))
    print("Equivalent protein %s has %i residues\n" % (cData.split(" ")[0].split("|")[-1],
len(pSeq[0].seq)))
```

```

    print(cData)
#THIS IS THE TABLE FOR CONVERTING THE BIOGRID IDs INTO UNIPROT IDs
decoder = pd.read_csv("/BIOGRID-ORGANISM-Homo_sapiens-
4.4.219.tab3.txt",sep="\t",skiprows=28)
#CAPUTURING ONLY THE UNIPROT IDs
FilterIdentifier = decoder["IDENTIFIER_TYPE"] == "SWISS-PROT"
SP = decoder [FilterIdentifier]

#MAKING A FRESH INDEX-RESETTED TABLE
#NOTE THAT THIS COULD BE REMOVED IN FUTURE VERSIONS TO MAKE THE
CODE MORE UNIVERSAL
OrganismIdentifier = SP["ORGANISM_OFFICIAL_NAME"] == "Homo sapiens"
SPbis=SP[OrganismIdentifier].reset_index(drop=True)

#MAKING A DICTIONNARY TO CONVERT BIOGRID IDs INTO UNIPROT IDs
HomoSapiensDict={}
SPbis=SP.reset_index(drop=True)
for i in range(len(SPbis["BIOGRID_ID"])):
    HomoSapiensDict[SPbis["BIOGRID_ID"][i]]=SPbis["IDENTIFIER_VALUE"][i]
dictTiming = time.time()
print("Dictionnaire de conversion créé ! Le processus a pris : % s secondes » % [dictTiming -
start_time])
#THIS IS THE TABLE WITH INTERACTIONS
lolo = pd.read_csv("/BIOGRID-ORGANISM-Homo_sapiens-4.4.219.tab3.txt",sep="\t")
#Filter BioGRID text file for query proteins
mydata = lolo[lolo["Official Symbol Interactor A"].isin(listOfGenes) | lolo["Official Symbol
Interactor B"].isin(listOfGenes)]

#Select full dataset or low-throughput
if filterByThroughput == True:
    print("Only considering low-throughput interactions")

```

```

cond = mydata['Throughput'] == "Low Throughput"
mydata = mydata[cond]

#Combine BAIT and HIT
mydataA = mydata["BioGRID ID Interactor A"] #Possible to use "Systematic Name Interactor
A"
mydataB = mydata["BioGRID ID Interactor B"] #Possible to use "Systematic Name Interactor
B"

#Make a single list with BAIT and HIT
total =[]
total.extend(mydataA.tolist())
total.extend(mydataB.tolist())

print("Nombre d'interactions identifiées : %i\n" % int(len(total)/2))
mostCommons= Counter(total).most_common()

interactTiming = time.time()
print("Le processus a pris : % s secondes" % (interactTiming - dictTiming))

for elem in mostCommons:
    if elem[1]>=minimalNumberOfObservations:
        print("La protéine % s a % i interactions" % (elem[0],elem[1]))
        if elem[0] in HomoSapiensDict :
            print(HomoSapiensDict[elem[0]])
            getFASTA(HomoSapiensDict[elem[0]],fastaLocation)
        else:
            print("PROBLEME")
print("Terminé ! Le processus complet a pris : % s secondes » % [time.time() - start_time])

```

## ANNEXE F

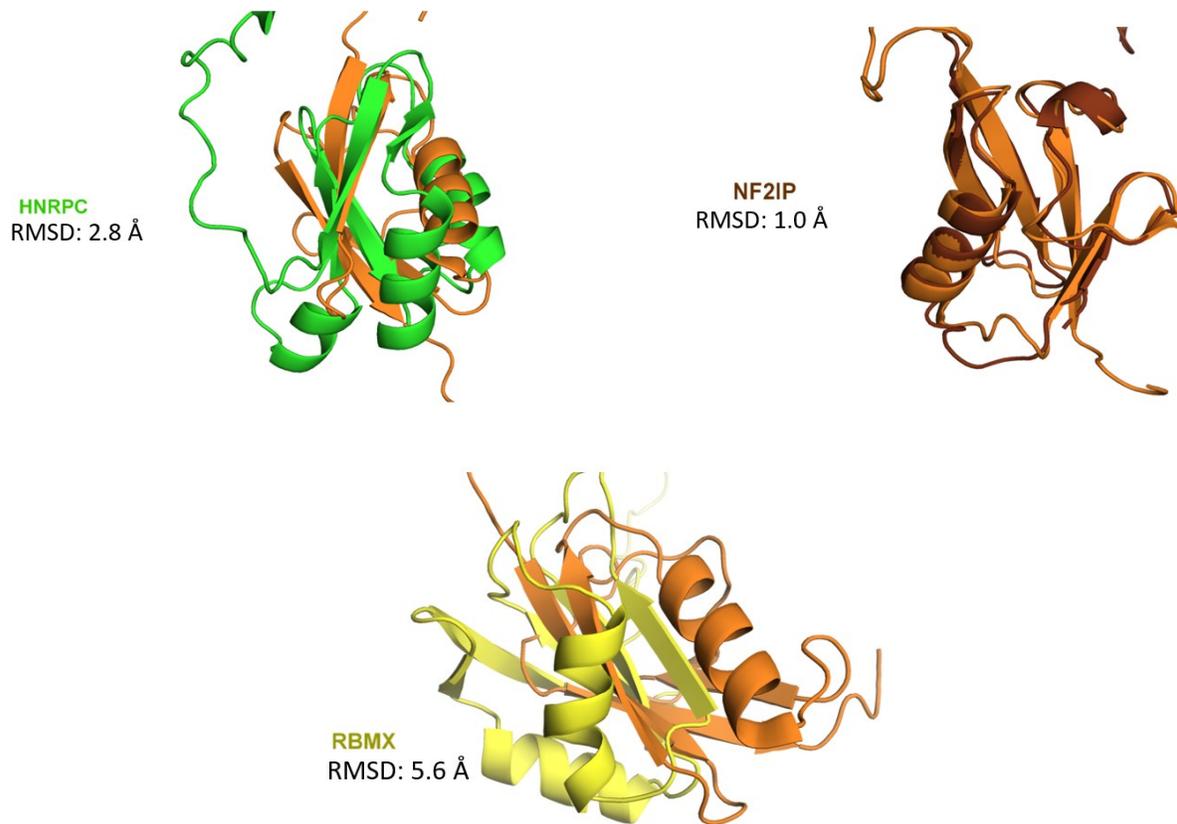


Figure 5.1: Superposition de SUMO2/3 (orange) (PDB 2N1W) avec les domaines SUMO-like de HNRPC (vert), NF2IP (brun) et RBMX (jaune).

## BIBLIOGRAPHIE

- Amemiya, Y., Azmi, P. & Seth, A. (2008). Autoubiquitination Of BCA2 RING E3 Ligase Regulates Its Own Stability And Affects Cell Migration. *Molecular Cancer Research*, 6(9), Pp.1385–1396. Doi:<https://doi.org/10.1158/1541-7786.Mcr-08-0094>.
- Audagnotto, M., & Dal Peraro, M. (2017). Protein Post-Translational Modifications: In Silico Prediction Tools And Molecular Modeling. *Computational And Structural Biotechnology Journal*, 15, 307–319. <https://doi.org/10.1016/j.csbj.2017.03.004>
- Baneyx, F., & Mujacic, M. (2004). Recombinant Protein Folding And Misfolding In Escherichia Coli. *Nature Biotechnology*, 22(11), 1399–1408. Doi:10.1038/Nbt1029
- Barnsby-Greer, L., Mabbitt, P.D., Marc-André Déry, Squair, D.R., Wood, N.T., Lange, S.M. & Satpal Virdee (2023). An Atypical E3 Ligase Module In UBR4 Mediates Destabilization Of N-Degron Substrates. *Biorxiv (Cold Spring Harbor Laboratory)*. Doi:<https://doi.org/10.1101/2023.05.08.539884>.
- Bayer, P., Arndt, A., Metzger, S., Mahajan, R., Melchior, F., Jaenicke, R. & Becker, J. (1998). Structure Determination Of The Small Ubiquitin-Related Modifier SUMO-1. *Journal Of Molecular Biology*, 280(2), Pp.275–286. Doi:<https://doi.org/10.1006/jmbi.1998.1839>.
- Beauclair, G., Bridier-Nahmias, A., Zagury, J. — F., Saïb, A. & Zamborlini, A. (2015). JASSA : A Comprehensive Tool For Prediction Of Sumoylation Sites And Sims. *Bioinformatics (Oxford, England)*, [Online] 31(21), Pp.3483–3491. Doi:<https://doi.org/10.1093/bioinformatics/btv403>.
- Bennett, R., Pan, Y., Christian, J., Hui, T. & May Jr., W.S. (2012). The RAX/PACT-PKR Stress Response Pathway Promotes P53 Sumoylation And Activation, Leading To G<sub>1</sub>arrest. *Cell Cycle*, 11(2), Pp.407–417. Doi:<https://doi.org/10.4161/cc.11.2.18999>.
- Bernier-Villamor, V., Sampson, D.A., Matunis, M.J. & Lima, C.D. (2002). Structural Basis For E2-Mediated SUMO Conjugation Revealed By A Complex Between Ubiquitin-Conjugating Enzyme Ubc9 And Rangap1. *Cell*, 108(3), Pp.345–356. Doi:[https://doi.org/10.1016/S0092-8674\(02\)00630-X](https://doi.org/10.1016/S0092-8674(02)00630-X).
- Bertoline, L.M.F., Lima, A. N., Krieger, J.E. & Teixeira, S.K. (2023). Before And After AlphaFold2: An Overview Of Protein Structure Prediction. *Frontiers In Bioinformatics*, 3. Doi:<https://doi.org/10.3389/fbinf.2023.1120370>.
- Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., & Brunak, S. (2004). Prediction Of Post-Translational Glycosylation And Phosphorylation Of Proteins From The Amino Acid Sequence. *Proteomics*, 4(6), 1633–1649. <https://doi.org/10.1002/pmic.200300771>
- Boddy, M.N., Howe, K., Etkin, L.D., Solomon, E. & Freemont, P.S. (1996). PIC 1, A Novel Ubiquitin-Like Protein Which Interacts With The PML Component Of A Multiprotein Complex That Is Disrupted In Acute Promyelocytic Leukaemia. *Oncogene*, [Online] 13(5), Pp.971–982. Available At: <https://pubmed.ncbi.nlm.nih.gov/8806687/>.

- Böhm, S. & Bernstein, K.A. (2014). The Role Of Post-Translational Modifications In Fine-Tuning BLM Helicase Function During DNA Repair. *DNA Repair*, 22, Pp.123–132. Doi:<https://doi.org/10.1016/j.dnarep.2014.07.007>.
- Bondos, S.E. & Bicknell, A. (2003). Detection And Prevention Of Protein Aggregation Before, During, And After Purification. *Analytical Biochemistry*, [Online] 316(2), Pp.223–231. Doi:[https://doi.org/10.1016/S0003-2697\(03\)00059-9](https://doi.org/10.1016/S0003-2697(03)00059-9).
- Bonenfant, D., Coulot, M., Towbin, H., Schindler, P., & Van Oostrum, J. (2006). Characterization Of Histone H2A And H2B Variants And Their Post-Translational Modifications By Mass Spectrometry. *Molecular & Cellular Proteomics : MCP*, 5(3), 541–552. <https://doi.org/10.1074/Mcp.M500288-MCP200>
- Borden, K.L. & Freemont, P.S. (1996). The RING Finger Domain: A Recent Example Of A Sequence—Structure Family. *Current Opinion In Structural Biology*, 6(3), Pp.395–401. Doi:[https://doi.org/10.1016/S0959-440x\(96\)80060-1](https://doi.org/10.1016/S0959-440x(96)80060-1).
- Brackett, C.M. & Blagg, B.S.J. (2020). Current Status Of Sumoylation Inhibitors. *Current Medicinal Chemistry*, 27. Doi:<https://doi.org/10.2174/0929867327666200810135039>.
- Canaves, J.M., Page, R., Wilson, I.A. & Stevens, R.C. (2004). Protein Biophysical Properties That Correlate With Crystallization Success In *Thermotoga Maritima*: Maximum Clustering Strategy For Structural Genomics. *Journal Of Molecular Biology*, [Online] 344(4), Pp.977–991. Doi:<https://doi.org/10.1016/j.jmb.2004.09.076>.
- Cappadocia, L. & Lima, C.D. (2017). Ubiquitin-Like Protein Conjugation: Structures, Chemistry, And Mechanism. *Chemical Reviews*, 118(3), Pp.889–918. Doi:<https://doi.org/10.1021/acs.chemrev.6b00737>.
- Cappadocia, L., Pichler, A. & Lima, C.D. (2015). Structural Basis For Catalytic Activation By The Human ZNF451 SUMO E3 Ligase. *Nature Structural & Molecular Biology*, 22(12), Pp.968–975. Doi:<https://doi.org/10.1038/nsmb.3116>.
- Celen, A.B. & Sahin, U. (2020). Sumoylation On Its 25th Anniversary: Mechanisms, Pathology, And Emerging Concepts. *The FEBS Journal*, 287(15), Pp.3110–3140. Doi:<https://doi.org/10.1111/febs.15319>.
- Chang, C.-C., Naik, Mandar T., Huang, Y.-S., Jeng, J.-C., Liao, P.-H., Kuo, H.-Y., Ho, C.-C., Hsieh, Y.-L., Lin, C.-H., Huang, N.-J., Naik, Nandita M., Kung, Camy C.-H. , Lin, S.-Y., Chen, R.-H., Chang, K.-S., Huang, T.-H. & Shih, H.-M. (2011). Structural And Functional Roles Of Daxx SIM Phosphorylation In SUMO Paralog-Selective Binding And Apoptosis Modulation. *Molecular Cell*, 42(1), Pp.62–74. Doi:<https://doi.org/10.1016/j.molcel.2011.02.022>.
- Chang, C.-C., Tung, C.-H., Chen, C.-W., Tu, C.-H. & Chu, Y.-W. (2018). Sumogo: Prediction Of Sumoylation Sites On Lysines By Motif Screening Models And The Effects Of Various Post-Translational Modifications. *Scientific Reports*, 8(1). Doi:<https://doi.org/10.1038/s41598-018-33951-5>.

- Chen, Q., Huang, L., Pan, D., Zhu, L.J. & Wang, Y.-X. (2018). Cbx4 Sumoylates Prdm16 To Regulate Adipose Tissue Thermogenesis. *Cell Reports*, 22(11), Pp.2860–2872. Doi: <https://doi.org/10.1016/j.celrep.2018.02.057>.
- Chen, X., Wei, S., Ji, Y., Guo, X. & Yang, F. (2015). Quantitative Proteomics Using SILAC: Principles, Applications, And Developments. *PROTEOMICS*, 15(18), Pp.3175–3192. Doi: <https://doi.org/10.1002/pmic.201500108>.
- Chen, Z., Ren, D., Lv, J., Xu, Y., Xie, M., He, X., Shi, W., Qian, Q., Jing, A., Ma, X., Qin, J., Ding, Y., Geng, T., Ma, J., Liu, W., Liu, S. & Ji, J. (2023). The Atypical Ubiquitin Ligase RNF31 Stabilizes C-Myc Via Epigenetic Inactivation Of FBXO32 And Promotes Cancer Development. *Cellular Signalling*, [Online] 107, P.110677. Doi: <https://doi.org/10.1016/j.cellsig.2023.110677>.
- Chirgwin, J.M., Przybyla, A.E., Macdonald, R.J. & Rutter, W.J. (1979). Isolation Of Biologically Active Ribonucleic Acid From Sources Enriched In Ribonuclease. *Biochemistry*, 18(24), Pp.5294–5299. Doi: <https://doi.org/10.1021/bi00591a005>.
- Crozier, L., Marshall, J., Holmes, A., Wright, K.M., Rossez, Y., Merget, B., Humphris, S., Toth, I., Jackson, R.W. & Holden, N.J. (2021). The Role Of L-Arabinose Metabolism For Escherichia Coli O157:H7 In Edible Plants. *Microbiology*, [Online] 167(7), P.001070. Doi: <https://doi.org/10.1099/mic.0.001070>.
- Cubeñas-Potts, C., Srikumar, T., Lee, C., Osula, O., Subramonian, D., Zhang, X.-D., Cotter, R.J., Raught, B. & Matunis, M.J. (2015). Identification Of SUMO-2/3-Modified Proteins Associated With Mitotic Chromosomes. *Proteomics*, [Online] 15(4), Pp.763–772. Doi: <https://doi.org/10.1002/pmic.201400400>.
- Dastgheib, S., Irajie, C., Assaei, R., Koohpeima, F. & Mokarram, P. (2014). Optimization Of RNA Extraction From Rat Pancreatic Tissue. *Iranian Journal Of Medical Sciences*, [Online] 39(3), Pp.282–8.
- De Bie, P. & Ciechanover, A. (2011). Ubiquitination Of E3 Ligases: Self-Regulation Of The Ubiquitin System Via Proteolytic And Non-Proteolytic Mechanisms. *Cell Death & Differentiation*, 18(9), Pp.1393–1402. Doi: <https://doi.org/10.1038/cdd.2011.16>.
- De Godoy, L.M., Olsen, J.V., De Souza, G.A., Li, G., Mortensen, P. & Mann, M. (2006). Status Of Complete Proteome Analysis By Mass Spectrometry: SILAC Labeled Yeast As A Model System. *Genome Biology*, 7(6), P.R50. Doi: <https://doi.org/10.1186/gb-2006-7-6-R50>.
- De Thé, H., Le Bras, M. & Lallemand-Breitenbach, V. (2012). Acute Promyelocytic Leukemia, Arsenic, And PML Bodies. *The Journal Of Cell Biology*, 198(1), Pp.11–21. Doi: <https://doi.org/10.1083/jcb.201112044>.
- Díaz-Rovira, A.M., Martín, H., Beuming, T., Díaz, L., Guallarv. & Ray, S.S. (2023). Are Deep Learning Structural Models Sufficiently Accurate For Virtual Screening? Application Of Docking Algorithms To Alphafold2 Predicted Structures. *Journal Of Chemical Information And Modeling*, 63(6), Pp.1668–1674. Doi: <https://doi.org/10.1021/acs.jcim.2c01270>.

- Elfmann C. & Stülke J. (2023). PAE Viewer: A Webserver For The Interactive Visualization Of The Predicted Aligned Error For Multimer Structure Predictions And Crosslinks. *Nucleic Acids Research*. Doi:<https://doi.org/10.1093/Nar/Gkad350>.
- Escobar-Cabrera, E., Okon, M., Lau, D., Dart, C.F., Bonvin, A. & McIntosh, L.P. (2011). Characterizing The N- And C-Terminal Small Ubiquitin-Like Modifier (SUMO)-Interacting Motifs Of The Scaffold Protein DAXX. *Journal Of Biological Chemistry*, 286(22), Pp.19816–19829. Doi:<https://doi.org/10.1074/Jbc.M111.231647>.
- Flotho, A., Werner, A., Winter, T., Frank, A. S., Ehret, H., & Melchior, F. (2012). Recombinant Reconstitution Of Sumoylation Reactions In Vitro. Ubiquitin Family Modifiers And The Proteasome, 93–110. Doi:10.1007/978-1-61779-474-2\_5
- François-Moutal, L., Dustrude, E.T., Wang, Y., Brustovetsky, T., Dorame, A., Ju, W., Moutal, A., Perez-Miller, S., Brustovetsky, N., Gokhale, V., Khanna, M. & Khanna, R. (2018). Inhibition Of The Ubc9 E2 SUMO Conjugating Enzyme–CRMP2 Interaction Decreases Nav1.7 Currents And Reverses Experimental Neuropathic Pain. *Pain*, [Online] 159(10), Pp.2115–2127. Doi:<https://doi.org/10.1097/J.Pain.0000000000001294>.
- Garbis, S., Lubec, G., & Fountoulakis, M. (2005). Limitations Of Current Proteomics Technologies. *Journal Of Chromatography A*, 1077(1), 1–18. Doi:10.1016/J.Chroma.2005.04.059
- García-Domínguez, M., March-Díaz, R. & Reyes, J.L. (2008). The PHD Domain Of Plant PIAS Proteins Mediates Sumoylation Of Bromodomain GTE Proteins. *Journal Of Biological Chemistry*, 283(31), Pp.21469–21477. Doi:<https://doi.org/10.1074/Jbc.M708176200>.
- Gareau, J.R. & Lima, C.D. (2010). The SUMO Pathway: Emerging Mechanisms That Shape Specificity, Conjugation And Recognition. *Nature Reviews Molecular Cell Biology*, 11(12), Pp.861–871. Doi:<https://doi.org/10.1038/Nrm3011>.
- Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. (2012). Comparative Proteomic Analysis Of Eleven Common Cell Lines Reveals Ubiquitous But Varying Expression Of Most Proteins. *Molecular & Cellular Proteomics*, 11(3), P.M111.014050. Doi:<https://doi.org/10.1074/Mcp.M111.014050>.
- Geiss-Friedlander, R. & Melchior, F. (2007). Concepts In Sumoylation: A Decade On. *Nature Reviews Molecular Cell Biology*, 8(12), Pp.947–956. Doi:<https://doi.org/10.1038/Nrm2293>.
- George, A.J., Hoffiz, Y.C., Charles, A.J., Zhu, Y. & Mabb, A.M. (2018). A Comprehensive Atlas Of E3 Ubiquitin Ligase Mutations In Neurological Disorders. *Frontiers In Genetics*, 9. Doi:<https://doi.org/10.3389/Fgene.2018.00029>.
- Ghosh, A., Stewart, D., & Matlashewski, G. (2004). Regulation Of Human P53 Activity And Cell Localization By Alternative Splicing. *Molecular And Cellular Biology*, 24(18), 7987–7997. <https://doi.org/10.1128/MCB.24.18.7987-7997.2004>
- Gill, G. (2004). SUMO And Ubiquitin In The Nucleus: Different Functions, Similar Mechanisms? *Genes & Development*, 18(17), Pp.2046–2059. Doi:<https://doi.org/10.1101/Gad.1214604>.

- Gladkova, C., Maslen, S.L., Skehel, J.M. & Komander, D. (2018). Mechanism Of Parkin Activation By PINK1. *Nature*, 559(7714), Pp.410–414.  
Doi:<https://doi.org/10.1038/S41586-018-0224-X>.
- Goh, C.-S., Lan, N., Douglas, S.M., Wu, B., Echols, N., Smith, A., Milburn, D., Montelione, G.T., Zhao, H. & Gerstein, M. (2004). Mining The Structural Genomics Pipeline: Identification Of Protein Properties That Affect High-Throughput Experimental Analysis. *Journal Of Molecular Biology*, 336(1), Pp.115–130.  
Doi:<https://doi.org/10.1016/J.Jmb.2003.11.053>.
- Goulet, A. & Cambillau, C. (2022). Present Impact Of Alphafold2 Revolution On Structural Biology, And An Illustration With The Structure Prediction Of The Bacteriophage J-1 Host Adhesion Device. *Frontiers In Molecular Biosciences*, 9.  
Doi:<https://doi.org/10.3389/Fmolb.2022.907452>.
- Gråslund, S. , Nordlund, P. , Weigelt, J. , Hallberg, B.M. , Bray, J. , Gileadi, O. , Knapp, S. , Oppermann, U. , Arrowsmith, C. , Hui, R. , Ming, J. , Dhe-Paganon, S. , Park, H.W. , Savchenko, A. , Yee, A. , Edwards, A. , Vincentelli, R. , Cambillau, C. , Kim, R. , Kim, S.H. , Rao, Z. , Shi, Y. , Terwilliger, T.C. , Kim, C.Y. , Hung, L.W. , Waldo, G.S. , Peleg, Y. , Albeck, S. , Unger, T. , Dym, O. , Prilusky, J. , Sussman, J.L. , Stevens, R.C. , Lesley, S.A. , Wilson, I.A. , Joachimiak, A. , Collart, F. , Dementieva, I. , Donnelly, M.I. , Eschenfeldt, W.H. , Kim, Y. , Stols, L. , Wu, R. , Zhou, M. , Burley, S.K. , Emtage, J.S. , Sauder, J.M. , Thompson, D. , Bain, K. , Luz, J. , Gheyi, T. , Zhang, F. , Atwell, S. , Almo, S.C. , Bonanno, J.B. , Fiser, A. , Swaminathan, S. , Studier, F.W. , Chance, M.R. , Sali, A. , Acton, T.B. , Xiao, R. , Zhao, L. , Ma, L.C. , Hunt, J.F. , Tong, L. , Cunningham, K. , Inouye, M. , Anderson, S. , Janjua, H. , Shastry, R. , Ho, C.K. , Wang, D. , Wang, H. , Jiang, M. , Montelione, G.T. , Stuart, D.I. , Owens, R.J. , Daenke, S. , Schutz, A. , Heinemann, U. , Yokoyama, S. , Bussow, K. , & Gunsalus, K.C. (2008a) Protein Production And Purification. *Nat. Methods* 5:135-146. Doi:<https://doi.org/10.1038/Nmeth.F.202>.
- Hecker, C.-M., Rabiller, M., Haglund, K., Bayer, P. & Dikic, I. (2006). Specification Of SUMO1- And SUMO2-Interacting Motifs. *Journal Of Biological Chemistry*, 281(23), Pp.16117–16127. Doi:<https://doi.org/10.1074/Jbc.M512757200>.
- Hendriks, I.A. & Vertegaal, A.C.O. (2016). A Comprehensive Compilation Of SUMO Proteomics. *Nature Reviews Molecular Cell Biology*, 17(9), Pp.581–595.  
Doi:<https://doi.org/10.1038/Nrm.2016.81>.
- Hendriks, I.A., D'Souza, R.C.J., Yang, B., Verlaan-De Vries, M., Mann, M. & Vertegaal, A.C.O. (2014). Uncovering Global Sumoylation Signaling Networks In A Site-Specific Manner. *Nature Structural & Molecular Biology*, [Online] 21(10), Pp.927–936. Doi:<https://doi.org/10.1038/Nsmb.2890>.
- Hendriks, I.A., Lyon, D., Young, C., Jensen, L.J., Vertegaal, A.C.O. & Nielsen, M.L. (2017). Site-Specific Mapping Of The Human SUMO Proteome Reveals Co-Modification With Phosphorylation. *Nature Structural & Molecular Biology*, [Online] 24(3), Pp.325–336. Doi:<https://doi.org/10.1038/Nsmb.3366>.

- Henley, J.M., Craig, T.J. & Wilkinson, K.A. (2014). Neuronal Sumoylation: Mechanisms, Physiology, And Roles In Neuronal Dysfunction. *Physiological Reviews*, 94(4), Pp.1249–1285. Doi:<https://doi.org/10.1152/Physrev.00008.2014>.
- Hernandez, G.E., Mondala, T.S. & Head, S.R. (2009). Assessing A Novel Room-Temperature RNA Storage Medium For Compatibility In Microarray Gene Expression Analysis. *Biotechniques*, 47(2), Pp.667–670. Doi:<https://doi.org/10.2144/000113209>.
- Hertz, E.P.T., Vega, I.A., Kruse, T., Wang, Y., Hendriks, I.A., Bizard, A.H., Eugui-Anta, A., Hay, R.T., Nielsen, M.L., Nilsson, J., Hickson, I.D. & Mailand, N. (2023). The SUMO–NIP45 Pathway Processes Toxic DNA Catenanes To Prevent Mitotic Failure. *Nature Structural & Molecular Biology*, [Online] 30(9), Pp.1303–1313. Doi:<https://doi.org/10.1038/S41594-023-01045-0>.
- Hickey, C.M., Wilson, N.R. & Hochstrasser, M. (2012). Function And Regulation Of SUMO Proteases. *Nature Reviews Molecular Cell Biology*, 13(12), Pp.755–766. Doi:<https://doi.org/10.1038/Nrm3478>.
- Hietakangas, V., Anckar, J., Blomster, H.A., Fujimoto, M., Palvimo, J.J., Nakai, A. & Sistonen, L. (2006). PDSM, A Motif For Phosphorylation-Dependent SUMO Modification. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, [Online] 103(1), Pp.45–50. Doi:<https://doi.org/10.1073/Pnas.0503698102>.
- Ho, C.-W., Chen, H.-T. & Hwang, J. (2011). UBC9 Autosumoylation Negatively Regulates Sumoylation Of Septins In *Saccharomyces Cerevisiae*. *Journal Of Biological Chemistry*, 286(24), Pp.21826–21834. Doi:<https://doi.org/10.1074/Jbc.M111.234914>.
- Hochstrasser, M. (2001). SP-RING For SUMO: New Functions Bloom For A Ubiquitin-Like Protein. *Cell*, [Online] 107(1), Pp.5–8. Doi:[https://doi.org/10.1016/S0092-8674\(01\)00519-0](https://doi.org/10.1016/S0092-8674(01)00519-0).
- Huang, D.T., Zhuang, M., Ayrault, O. & Schulman, B.A. (2008). Identification Of Conjugation Specificity Determinants Unmasks Vestigial Preference For Ubiquitin Within The NEDD8 E2. *Nature Structural & Molecular Biology*, [Online] 15(3), Pp.280–287. Doi:<https://doi.org/10.1038/Nsmb.1387>.
- Huang, K. Y., Lee, T. Y., Kao, H. J., Ma, C. T., Lee, C. C., Lin, T. H., Chang, W. C., & Huang, H. D. (2019). Dbptm In 2019: Exploring Disease Association And Cross-Talk Of Post-Translational Modifications. *Nucleic Acids Research*, 47(D1), D298–D308. <https://doi.org/10.1093/Nar/Gky1074>
- Huang, R., Huang, Y., Guo, Y., Ji, S., Lu, M. & Li, T.-T. (2018). Systematic Characterization And Prediction Of Post-Translational Modification Cross-Talk Between Proteins. *Bioinformatics*, 35(15), Pp.2626–2633. Doi:<https://doi.org/10.1093/Bioinformatics/Bty1033>.
- Hughes, D.J., Tiede, C., Penswick, N., Tang, A.A.-S., Trinh, C.H., Mandal, U., Zajac, K.Z., Gaule, T., Howell, G., Edwards, T.A., Duan, J., Feyfant, E., Mcpherson, M.J., Tomlinson, D.C. & Whitehouse, A. (2017). Generation Of Specific Inhibitors Of

- SUMO-1– And SUMO-2/3–Mediated Protein-Protein Interactions Using Affimer (Adhiron) Technology. *Science Signaling*, 10(505), P.Eaaj2005.  
Doi:Https://Doi.Org/10.1126/Scisignal.Aaj2005.
- Ihara, M., Yamamoto, H. & Kikuchi, A. (2005). SUMO-1 Modification Of Piasy, An E3 Ligase, Is Necessary For Piasy-Dependent Activation Of Tcf-4. *Molecular And Cellular Biology*, 25(9), Pp.3506–3518.  
Doi:Https://Doi.Org/10.1128/Mcb.25.9.3506-3518.2005.
- Imbert, F., Leavitt, G. & Langford, D. (2022). Sumoylation And Viral Infections Of The Brain. *Pathogens*, [Online] 11(7), P.818.  
Doi:Https://Doi.Org/10.3390/Pathogens11070818.
- Iribarren, P.A., Berazategui, M.A., Cazzulo, J.J. & Alvarez, V.E. (2015). Biosynthesis Of Sumoylated Proteins In Bacteria Using The Trypanosoma Brucei Enzymatic System. *PLOS ONE*, 10(8), P.E0134950. Doi:Https://Doi.Org/10.1371/Journal.Pone.0134950.
- Ismail, I. H., Gagné, J.-P., Caron, M. — C., Mcdonald, D., Xu, Z., Masson, J.-Y., Poirier, G.G. & Hendzel, M.J. (2012). CBX4-Mediated SUMO Modification Regulates BMI1 Recruitment At Sites Of DNA Damage. *Nucleic Acids Research*, 40(12), Pp.5497–5510. Doi:Https://Doi.Org/10.1093/Nar/Gks222.
- Ivanov, A.V., Peng, H., Yurchenko, V., Yap, K.L., Negorev, D.G., Schultz, D.C., Psulkowski, E., Fredericks, W.J., White, D.E., Maul, G.G., Sadofsky, M.J., Zhou, M.-M. & Rauscher, F.J. (2007). PHD Domain-Mediated E3 Ligase Activity Directs Intramolecular Sumoylation Of An Adjacent Bromodomain Which Is Required For Gene Silencing. *Molecular Cell*, [Online] 28(5), Pp.823–837.  
Doi:Https://Doi.Org/10.1016/J.Molcel.2007.11.012.
- Jackson, Stephen P. & Durocher, D. (2013). Regulation Of DNA Damage Responses By Ubiquitin And SUMO. *Molecular Cell*, 49(5), Pp.795–807.  
Doi:Https://Doi.Org/10.1016/J.Molcel.2013.01.017.
- Jardin, C., Horn, A.H.C. & Sticht, H. (2015). Binding Properties Of SUMO-Interacting Motifs (Sims) In Yeast. *Journal Of Molecular Modeling*, 21(3).  
Doi:Https://Doi.Org/10.1007/S00894-015-2597-1.
- Jeon, Y.J., Yoo, H.M. & Chung, C.H. (2010). ISG15 And Immune Diseases. *Biochimica Et Biophysica Acta (BBA) - Molecular Basis Of Disease*, [Online] 1802(5), Pp.485–496.  
Doi:Https://Doi.Org/10.1016/J.Bbadis.2010.02.006.
- Jia, D., Hasso, S.M., Chan, J., Filingeri, D., D'Amore, P. A., Rice, L., Pampo, C., Siemann, D. W., Zurakowski, D., Rodig, S.J. & Moses, M.A. (2013). Transcriptional Repression Of VEGF By ZNF24: Mechanistic Studies And Vascular Consequences In Vivo. *Blood*, [Online] 121(4), Pp.707–715. Doi:Https://Doi.Org/10.1182/Blood-2012-05-433045.
- Johnson, E.S. & Blobel, G. (1997). Ubc9p Is The Conjugating Enzyme For The Ubiquitin-Like Protein Smt3p. *Journal Of Biological Chemistry*, 272(43), Pp.26799–26802.  
Doi:Https://Doi.Org/10.1074/Jbc.272.43.26799.

- Johnson, E.S. & Gupta, A.A. (2001). An E3-Like Factor That Promotes SUMO Conjugation To The Yeast Septins. *Cell*, [Online] 106(6), Pp.735–744.  
Doi:[https://doi.org/10.1016/S0092-8674\(01\)00491-3](https://doi.org/10.1016/S0092-8674(01)00491-3).
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J. & Back, T. (2021). Highly Accurate Protein Structure Prediction With Alphafold. *Nature*, 596(7873), Pp.583–589.
- Kagey, M.H., Melhuish, T.A. & Wotton, D. (2003). The Polycomb Protein Pc2 Is A SUMO E3. *Cell*, 113(1), Pp.127–137. Doi:[https://doi.org/10.1016/S0092-8674\(03\)00159-4](https://doi.org/10.1016/S0092-8674(03)00159-4).
- Kahyo, T., Nishida, T. & Yasuda, H. (2001). Involvement Of PIAS1 In The Sumoylation Of Tumor Suppressor P53. *Molecular Cell*, 8(3), Pp.713–718.  
Doi:[https://doi.org/10.1016/S1097-2765\(01\)00349-5](https://doi.org/10.1016/S1097-2765(01)00349-5).
- Karve, T. M., & Cheema, A. K. (2011). Small Changes Huge Impact: The Role Of Protein Posttranslational Modifications In Cellular Homeostasis And Disease. *Journal Of Amino Acids*, 2011, 207691. <https://doi.org/10.4061/2011/207691>
- Kaur, K., Park, H., Pandey, N., Azuma, Y. & De Guzman, R.N. (2017). Identification Of A New Small Ubiquitin-Like Modifier (SUMO)-Interacting Motif In The E3 Ligase Piasy. *The Journal Of Biological Chemistry*, [Online] 292(24), Pp.10230–10238.  
Doi:<https://doi.org/10.1074/jbc.M117.789982>.
- Kirsh, O., Seeler, Jacob-S., Pichler, A., Gast, A., Muller, S., Miska, E., Mathieu, M., Harel-Bellan, A., Kouzarides, T., Melchior, F. & Dejeans, A. (2002). The SUMO E3 Ligase Ranbp2 Promotes Modification Of The HDAC4 Deacetylase. *The EMBO Journal*, 21(11), Pp.2682–2691. Doi:<https://doi.org/10.1093/emboj/21.11.2682>.
- Knipscheer, P., Flotho, A., Klug, H., Olsen, J.V., Van Dijk, W.J., Fish, A., Johnson, E.S., Mann, M., Sixma, T.K. & Pichler, A. (2008). Ubc9 Sumoylation Regulates SUMO Target Discrimination. *Molecular Cell*, 31(3), Pp.371–382.  
Doi:<https://doi.org/10.1016/j.molcel.2008.05.022>.
- Knipscheer, P., Van Dijk, W.J., Olsen, J.V., Mann, M. & Sixma, T.K. (2007). Noncovalent Interaction Between Ubc9 And SUMO Promotes SUMO Chain Formation. *The EMBO Journal*, 26(11), Pp.2797–2807.  
Doi:<https://doi.org/10.1038/sj.emboj.7601711>.
- Korbie, D.J. & Mattick, J.S. (2008). Touchdown PCR For Increased Specificity And Sensitivity In PCR Amplification. *Nature Protocols*, [Online] 3(9), Pp.1452–1456.  
Doi:<https://doi.org/10.1038/nprot.2008.133>.
- Kumar, A. & Zhang, K. (2015). Advances In The Development Of SUMO Specific Protease (SENp) Inhibitors. [Online] 13, Pp.204–211.  
Doi:<https://doi.org/10.1016/j.csbj.2015.03.001>.
- Kumar, R., González-Prieto, R., Xiao, Z., Verlaan-De Vries, M. & Vertegaal, A.C.O. (2017). The Stubl RNF4 Regulates Protein Group Sumoylation By Targeting The SUMO Conjugation Machinery. *Nature Communications*, 8(1).  
Doi:<https://doi.org/10.1038/s41467-017-01900-x>.

- Kunz, K., Piller, T. & Müller, S. (2018). SUMO-Specific Proteases And Isopeptidases Of The SENP Family At A Glance. *Journal Of Cell Science*, [Online] 131(6), P.Jcs211904. Doi:Https://Doi.Org/10.1242/Jcs.211904.
- Kuo, F.-T., Bentsi-Barnes, I.K., Barlow, G.M., Bae, J. & Pisarska, M.D. (2009). Sumoylation Of Forkhead L2 By Ubc9 Is Required For Its Activity As A Transcriptional Repressor Of The Steroidogenic Acute Regulatory Gene. *Cellular Signalling*, [Online] 21(12), Pp.1935–1944. Doi:Https://Doi.Org/10.1016/J.Cellsig.2009.09.001.
- Kurepa, J., Walker, J.F., Smalle, J.A., Gosink, M., Davis, S.J., Durham, T.L., Sung, D.-Y. & Vierstra, R.D. (2003). The Small Ubiquitin-Like Modifier (SUMO) Protein Modification System In Arabidopsis. *Journal Of Biological Chemistry*, 278(9), Pp.6862–6872. Doi:Https://Doi.Org/10.1074/Jbc.M209694200.
- Lamoliatte, F., Caron, D., Durette, C., Louiza Mahrouche, Mohamed Ali Maroui, Olivier Caron-Lizotte, Éric Bonneil, Chelbi-Alix, M.K. & Thibault, P. (2014). Large-Scale Analysis Of Lysine Sumoylation By SUMO Remnant Immunoaffinity Profiling. *Nature Communications*, 5(1). Doi:Https://Doi.Org/10.1038/Ncomms6409.
- Lascorz, J., Codina-Fabra, J., Reverter, D. & Torres-Rosell, J. (2022). SUMO-SIM Interactions: From Structure To Biological Functions. *Seminars In Cell & Developmental Biology*, [Online] 132, Pp.193–202. Doi:Https://Doi.Org/10.1016/J.Semcdb.2021.11.007.
- Lau, H.-T., Suh, H.W., Golkowski, M. & Ong, S.-E. (2014). Comparing SILAC- and Stable Isotope Dimethyl-Labeling Approaches for Quantitative Proteomics. *Journal of Proteome Research*, 13(9), pp.4164–4174. doi:https://doi.org/10.1021/pr500630a.
- Lee, P.S.W., Chang, C., Liu, D. & Derynck, R. (2003c). Sumoylation Of Smad4, The Common Smad Mediator Of Transforming Growth Factor-B Family Signaling. *Journal Of Biological Chemistry*, 278(30), Pp.27853–27863. Doi:Https://Doi.Org/10.1074/Jbc.M301755200.
- Li, C., Boutet, A., Cristina Mirela Pascariu, Nelson, T., Mathieu Courcelles, Wu, Z., Comtois-Marotte, S., Emery, G. & Thibault, P. (2023a). SUMO Proteomics Analyses Identify Protein Inhibitor Of Activated STAT-Mediated Regulatory Networks Involved In Cell Cycle And Cell Proliferation. *Journal Of Proteome Research*, [Online] 22(3), Pp.812–825. Doi:Https://Doi.Org/10.1021/Acs.Jproteome.2c00557.
- Li, J., Chen, X., Liu, Y., Ding, L., Qiu, L., Hu, Z. & Zhang, J. (2010). The Transcriptional Repression Of Platelet-Derived Growth Factor Receptor-Beta By The Zinc Finger Transcription Factor ZNF24. *Biochemical And Biophysical Research Communications*, [Online] 397(2), Pp.318–322. Doi:Https://Doi.Org/10.1016/J.Bbrc.2010.05.110.
- Li, J., Xu, Y., Long, X.-D., Wang, W., Jiao, H.-K., Mei, Z., Yin, Q., Ma, L., Zhou, A., Wang, L.-S., Yao, M., Xia, Q. & Chen, G.-Q. (2014). Cbx4 Governs HIF-1 $\alpha$  To Potentiate Angiogenesis Of Hepatocellular Carcinoma By Its SUMO E3 Ligase Activity. *Cancer Cell*, 25(1), Pp.118–131. Doi:Https://Doi.Org/10.1016/J.Ccr.2013.12.008.

- Li, K., Xia, Y., He, J., Wang, J., Li, J., Ye, M. & Jin, X. (2023). The Sumoylation And Ubiquitination Crosstalk In Cancer. *Journal Of Cancer Research And Clinical Oncology*, 149(17), Pp.16123–16146. Doi:Https://Doi.Org/10.1007/S00432-023-05310-Z.
- Liang, M., Melchior, F., Feng, X.-H. & Lin, X. (2004). Regulation Of Smad4 Sumoylation And Transforming Growth Factor-B Signaling By Protein Inhibitor Of Activated STAT1. *Journal Of Biological Chemistry*, 279(22), Pp.22857–22865. Doi:Https://Doi.Org/10.1074/Jbc.M401554200.
- Lin, D.-Y., Huang, Y.-S., Jeng, J.-C., Kuo, H.-Y., Chang, C.-C., Chao, T.-T., Ho, C.-C., Chen, Y.-C., Lin, T.-P., Fang, H.-I., Hung, C.-C., Suen, C.-S., Hwang, M.-J., Chang, K.-S., Maul, G.G. & Shih, H.-M. (2006). Role Of SUMO-Interacting Motif In Daxx SUMO Modification, Subnuclear Localization, And Repression Of Sumoylated Transcription Factors. *Molecular Cell*, 24(3), Pp.341–354. Doi:Https://Doi.Org/10.1016/J.Molcel.2006.10.019.
- Lin, X., Liang, M., Liang, Y.-Y., Brunnicardi F. Charles, Melchior, F. & Feng, X.-H. (2003). Activation Of Transforming Growth Factor-B Signaling By SUMO-1 Modification Of Tumor Suppressor Smad4/DPC4. *Journal Of Biological Chemistry*, 278(21), Pp.18714–18719. Doi:Https://Doi.Org/10.1074/Jbc.M302243200.
- Lin, X., Liang, M., Liang, Y.-Y., Brunnicardi, F.Charles. & Feng, X.-H. (2003a). SUMO-1/Ubc9 Promotes Nuclear Accumulation And Metabolic Stability Of Tumor Suppressor Smad4. *Journal Of Biological Chemistry*, 278(33), Pp.31043–31048. Doi:Https://Doi.Org/10.1074/Jbc.C300112200.
- Lin, Y., Chung, C., Huang, P., Chen, C. & Fang, S. (2020). Revised Annotation And Extended Characterizations Of Components Of The *Chlamydomonas Reinhardtii* Sumoylation System. *Plant Direct*, 4(9). Doi:Https://Doi.Org/10.1002/Pld3.266.
- Lingbeck, J.M., Trausch-Azar, J.S., Ciechanover, A. & Schwartz, A.L. (2003). Determinants of Nuclear and Cytoplasmic Ubiquitin-mediated Degradation of MyoD. *Journal of Biological Chemistry*, [online] 278(3), pp.1817–1823. doi:https://doi.org/10.1074/jbc.M208815200.
- Liu, F., Wang, X., Su, M., Yu, M., Zhang, S., Lai, J., Yang, C. & Wang, Y. (2015). Functional Characterization Of Dns1z1, A SIZ/PIAS-Type SUMO E3 Ligase From *Dendrobium*. *BMC Plant Biology*, 15(1). Doi:Https://Doi.Org/10.1186/S12870-015-0613-3.
- Liu, G., Jiang, S., Wang, C., Jiang, W., Liu, Z., Liu, C., Saiyin, H., Yang, X., Shen, S., Jiang, D., Zhou, P., Han, D., Hu, X., Yi, Q. & Yu, L. (2012). Zinc Finger Transcription Factor 191, Directly Binding To B-Catenin Promoter, Promotes Cell Proliferation Of Hepatocellular Carcinoma. *Hepatology*, 55(6), Pp.1830–1839. Doi:Https://Doi.Org/10.1002/Hep.25564.
- Liu, T., Yang, H., Zhuo, F.-F., Yang, Z., Zhao, M., Guo, Q., Liu, Y., Liú, D., Zeng, K. & Tu, P. (2022). Atypical E3 Ligase ZFP91 Promotes Small-Molecule-Induced E2F2 Transcription Factor Degradation For Cancer Therapy. *Ebiomedicine*, 86, Pp.104353–104353. Doi:Https://Doi.Org/10.1016/J.Ebiom.2022.104353.

- Long, J., Wang, G., He, D. & Liu, F. (2004). Repression Of Smad4 Transcriptional Activity By SUMO Modification. *Biochemical Journal*, 379(1), Pp.23–29. Doi:<https://doi.org/10.1042/Bj20031867>.
- Lönn, P., Morén, A., Raja, E., Dahl, M. & Moustakas, A. (2009). Regulating The Stability Of Tgf $\beta$  Receptors And Smads. *Cell Research*, [Online] 19(1), Pp.21–35. Doi:<https://doi.org/10.1038/Cr.2008.308>.
- Lussier-Price, M., Mascle, X.H., Cappadocia, L., Kamada, R., Sakaguchi, K., Wahba, H.M. & Omichinski, J.G. (2020). Characterization Of A C-Terminal SUMO-Interacting Motif Present In Select PIAS-Family Proteins. *Structure*, 28(5), Pp.573-585.E5. Doi:<https://doi.org/10.1016/J.Str.2020.04.002>.
- Lv, Z., Yuan, L., Atkison, J.H., Williams, K.M., Vega, R., Sessions, E.H., Divlianska, D.B., Davies, C., Chen, Y. & Olsen, S.K. (2018). Molecular Mechanism Of A Covalent Allosteric Inhibitor Of SUMO E1 Activating Enzyme. *Nature Communications*, 9(1). Doi:<https://doi.org/10.1038/S41467-018-07015-1>.
- Lyst, M.J., Nan, X. & Stancheva, I. (2006). Regulation Of MBD1-Mediated Transcriptional Repression By SUMO And PIAS Proteins. *The EMBO Journal*, 25(22), Pp.5317–5328. Doi:<https://doi.org/10.1038/Sj.Emboj.7601404>.
- Mahajan, R., Delphin, C., Guan, T., Gerace, L. & Melchior, F. (1997). A Small Ubiquitin-Related Polypeptide Involved In Targeting Rangap1 To Nuclear Pore Complex Protein Ranbp2. *Cell*, 88(1), Pp.97–107. Doi:[https://doi.org/10.1016/S0092-8674\(00\)81862-0](https://doi.org/10.1016/S0092-8674(00)81862-0).
- Malakhov, M.P., Malakhova, O.A., Kim, K.I., Ritchie, K.J. & Zhang, D.-E. (2002). UBP43 (USP18) Specifically Removes ISG15 From Conjugated Proteins. *Journal Of Biological Chemistry*, 277(12), Pp.9976–9981. Doi:<https://doi.org/10.1074/Jbc.M109078200>.
- Mann, M. And Jensen, O.N. (2003) Proteomic Analysis Of Post-Translational Modifications. *Nat. Biotechnol.*, 21, 255–261.
- Marshall C. J. (1993). Protein Prenylation: A Mediator Of Protein-Protein Interactions. *Science* (New York, N.Y.), 259(5103), 1865–1866. <https://doi.org/10.1126/Science.8456312>
- Martin, N., Schwamborn, K., Schreiber, V., Werner, A., Guillier, C., Zhang, X.-D., Bischof, O., Seeler, J.-S. & Dejean, A. (2009). PARP-1 Transcriptional Activity Is Regulated By Sumoylation Upon Heat Shock. *The EMBO Journal*, 28(22), Pp.3534–3548. Doi:<https://doi.org/10.1038/Emboj.2009.279>.
- Matsunaga, S., Takata, H., Morimoto, A., Kayoko H., Higashi, T., Kouhei Akatsuchi, M., E., Yamakawa, M., Ashida, M., Matsunaga, T.M., Azuma, T., Uchiyama, S. & Fukui, K. (2012). RBMX: A Regulator For Maintenance And Centromeric Protection Of Sister Chromatid Cohesion. *Cell Reports*, 1(4), Pp.299–308. Doi:<https://doi.org/10.1016/J.Celrep.2012.02.005>.
- Matunis, M.J. & Rodriguez, M.S. (2016). Concepts And Methodologies To Study Protein Sumoylation: An Overview. *Methods In Molecular Biology (Clifton, N.J.)*, [Online] 1475, Pp.3–22. Doi:[https://doi.org/10.1007/978-1-4939-6358-4\\_1](https://doi.org/10.1007/978-1-4939-6358-4_1).

- Matunis, M.J., Coutavas, E. & Blobel, G. (1996). A Novel Ubiquitin-Like Modification Modulates The Partitioning Of The Ran-Gtpase-Activating Protein Rangap1 Between The Cytosol And The Nuclear Pore Complex. *Journal Of Cell Biology*, 135(6), Pp.1457–1470. Doi:Https://Doi.Org/10.1083/Jcb.135.6.1457.
- Mayo, L.D. (2012). Directing P53 To Induce Autophagy. *Cell Cycle*, 11(18), Pp.3354–3354. Doi:Https://Doi.Org/10.4161/Cc.21849.
- Mcaleenan, A., Cordon-Preciado, V., Clemente-Blanco, A., Liu, I-Chun., Sen, N., Leonard, J., Jarmuz, A. & Aragón, L. (2012). Sumoylation Of The A-Kleisin Subunit Of Cohesin Is Required For DNA Damage-Induced Cohesion. *Current Biology*, 22(17), Pp.1564–1575. Doi:Https://Doi.Org/10.1016/J.Cub.2012.06.045.
- Mccance KL, Huether SE. In: Pathophysiology: The Biologic Basis For Disease In Adults And Children. 7th Edition. Brashers VL, Rote NS, Editors. Elsevier; 2014.
- Melchior, F., Schergaut, M. & Pichler, A. (2003). SUMO: Ligases, Isopeptidases And Nuclear Pores. *Trends In Biochemical Sciences*, 28(11), Pp.612–618. Doi:Https://Doi.Org/10.1016/J.Tibs.2003.09.002.
- Meluh, P.B. & Koshland, D. (1995). Evidence That The MIF2 Gene Of Saccharomyces Cerevisiae Encodes A Centromere Protein With Homology To The Mammalian Centromere Protein CENP-C. *Molecular Biology Of The Cell*, 6(7), Pp.793–807. Doi:Https://Doi.Org/10.1091/Mbc.6.7.793.
- Minty, A., Dumont, X., Kaghad, M. & Caput, D. (2000). Covalent Modification Of P73 $\alpha$  By Sumo-1: Two-Hybrid Screening With P73 Identifies Novel Sumo-1-Interacting Proteins And A Sumo-1 Interaction Motif \*. *Journal Of Biological Chemistry*, [Online] 275(46), Pp.36316–36323. Doi:Https://Doi.Org/10.1074/Jbc.M004293200.
- Miura, K., Rus, A., Sharkhuu, A., Yokoi, S., Karthikeyan, A.S., Raghothama, K.G., Baek, D., Koo, Y.D., Jin, J.B., Bressan, R.A., Yun, D.-J. & Hasegawa, P.M. (2005). The Arabidopsis SUMO E3 Ligase SIZ1 Controls Phosphate Deficiency Responses. *Proceedings Of The National Academy Of Sciences*, 102(21), Pp.7760–7765. Doi:Https://Doi.Org/10.1073/Pnas.0500778102.
- Miyauchi, Y., Yogosawa, S., Honda, R., Nishida, T. & Yasuda, H. (2002). Sumoylation Of Mdm2 By Protein Inhibitor Of Activated STAT (PIAS) And Ranbp2 Enzymes. *Journal Of Biological Chemistry*, 277(51), Pp.50131–50136. Doi:Https://Doi.Org/10.1074/Jbc.M208319200.
- Mohideen, F., Capili, A.D., Bilimoria, P.M., Yamada, T., Bonni, A. & Lima, C.D. (2009). A Molecular Basis For Phosphorylation-Dependent SUMO Conjugation By The E2 UBC9. *Nature Structural & Molecular Biology*, 16(9), Pp.945–952. Doi:Https://Doi.Org/10.1038/Nsmb.1648.
- Mossessova, E. & Lima, C.D. (2000). Ulp1-SUMO Crystal Structure And Genetic Analysis Reveal Conserved Interactions And A Regulatory Element Essential For Cell Growth In Yeast. *Mol Cell.*, 5(5), Pp.865–876. Doi:Https://Doi.Org/10.1016/S1097-2765(00)80326-3.

- Naidu, S.R., Lakhter, A.J. & Androphy, E.J. (2012). Piasy-Mediated Tip60 Sumoylation Regulates P53-Induced Autophagy. *Cell Cycle*, 11(14), Pp.2717–2728. Doi:<https://doi.org/10.4161/Cc.21091>.
- Naik, M.T., Kang, M., Ho, C.-C., Liao, P.-H., Hsieh, Y.-L., Naik, N.M., Wang, S.-H., Chang, I., Shih, H.-M. & Huang, T.-H. (2017). Molecular Mechanism Of K65 Acetylation-Induced Attenuation Of Ubc9 And The NDSM Interaction. *Scientific Reports*, 7(1). Doi:<https://doi.org/10.1038/S41598-017-17465-0>.
- Namanja, A.T., Li, Y., Su, Y., Wong, S., Lu, J., Colson, L., Wu, C., Shawn S.-C. Li & Chen, Y. (2012). Insights Into High Affinity Small Ubiquitin-Like Modifier (SUMO) Recognition By SUMO-Interacting Motifs (Sims) Revealed By A Combination Of NMR And Peptide Array Analysis. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, 287(5), Pp.3231–3240. Doi:<https://doi.org/10.1074/Jbc.M111.293118>.
- Nestler EJ & Greengard P., Protein Phosphorylation Is Of Fundamental Importance In Biological Regulation. Basic Neurochemistry: Molecular, Cellular And Medical Aspects 6th Edition. Published Online 1999. <https://www.ncbi.nlm.nih.gov/books/NBK28063/>
- Nico Pietack, Becher, D., Schmidl, S.R., Saier, M.H., Hecker, M., Commichau, F.M. & Jörg Stülke (2010). In Vitro Phosphorylation Of Key Metabolic Enzymes From *Bacillus Subtilis*: Prkc Phosphorylates Enzymes From Different Branches Of Basic Metabolism. *Microbial Physiology*, 18(3), Pp.129–140. Doi:<https://doi.org/10.1159/000308512>.
- Nishida, T. & Yasuda, H. (2002). PIAS1 And Pias $\alpha$  Function As SUMO-E3 Ligases Toward Androgen Receptor And Repress Androgen Receptor-Dependent Transcription. *Journal Of Biological Chemistry*, 277(44), Pp.41311–41317. Doi:<https://doi.org/10.1074/Jbc.M206741200>.
- Novatchkova, M., Bachmair, A., Eisenhaber, B. & Eisenhaber, F. (2005). Proteins With Two SUMO-Like Domains In Chromatin-Associated Complexes: The Reni (Rad60-Esc2-NIP45) Family. *BMC Bioinformatics*, 6(1), P.22. Doi:<https://doi.org/10.1186/1471-2105-6-22>.
- Novatchkova, M., Tomanov, K., Hofmann, K., Stuible, H.-P. & Bachmair, A. (2012). Update On Sumoylation: Defining Core Components Of The Plant SUMO Conjugation System By Phylogenetic Comparison. *New Phytologist*, 195(1), Pp.23–31. Doi:<https://doi.org/10.1111/J.1469-8137.2012.04135.X>.
- Oh, Y. & Chung, K.C. (2013). UHRF2, A Ubiquitin E3 Ligase, Acts As A Small Ubiquitin-Like Modifier E3 Ligase For Zinc Finger Protein 131. *Journal Of Biological Chemistry*, 288(13), Pp.9102–9111. Doi:<https://doi.org/10.1074/Jbc.M112.438234>.
- Ohshima, T. & Shimotohno, K. (2003). Transforming Growth Factor-B-Mediated Signaling Via The P38 MAP Kinase Pathway Activates Smad-Dependent Transcription Through SUMO-1 Modification Of Smad4. *Journal Of Biological Chemistry*, 278(51), Pp.50833–50842. Doi:<https://doi.org/10.1074/Jbc.M307533200>.
- Okura, T., Gong, L., Kamitani, T., Wada, T., Okura, I., Wei, C.F., Chang, H.M. & Yeh, E.T. (1996). Protection Against Fas/APO-1- And Tumor Necrosis Factor-Mediated Cell

- Death By A Novel Protein, Sentrin. *Journal Of Immunology (Baltimore, Md.: 1950)*, [Online] 157(10), Pp.4277–4281. Available At: <https://pubmed.ncbi.nlm.nih.gov/8906799/>.
- Ouyang, J., Yujiang Geno Shi, Valin, A., Xuan, Y. & Gill, G. (2009). Direct Binding Of Corest1 To SUMO-2/3 Contributes To Gene-Specific Repression By The LSD1/Corest1/HDAC Complex. *Mol Cell*, 34(2), Pp.145–154. Doi:<https://doi.org/10.1016/j.molcel.2009.03.013>.
- Pan, X., Li, H., Zhang, P., Jin, B., Man, J., Tian, L., Su, G., Zhao, J., Li, W., Liu, H., Gong, W., Zhou, T. & Zhang, X. (2006). Ubc9 Interacts With SOX4 And Represses Its Transcriptional Activity. *Biochemical And Biophysical Research Communications*, [Online] 344(3), Pp.727–734. Doi:<https://doi.org/10.1016/j.bbrc.2006.03.194>.
- Park, H.C., Kim, H., Koo, S.C., Park, H.J., Cheong, M.S., Hong, H., Baek, D., Chung, W.S., Kim, D.H., Bressan, R.A., Lee, S.Y., Bohnert, H.J. & Yun, D.-J. (2010). Functional Characterization Of The Siz/Pias-Type Sumo E3 Ligases, Ossiz1 And Ossiz2 In Rice. *Plant, Cell & Environment*, 33(11), Pp.1923–1934. Doi:<https://doi.org/10.1111/j.1365-3040.2010.02195.x>.
- Park, H.J., Kim, W.-Y., Park, H.C., Lee, S.Y., Bohnert, H.J. & Yun, D.-J. (2011). SUMO And Sumoylation In Plants. *Molecules And Cells*, 32(4), Pp.305–316. Doi:<https://doi.org/10.1007/s10059-011-0122-7>.
- Pascal, J.M. (2018). The Comings And Goings Of PARP-1 In Response To DNA Damage. *DNA Repair*, (71), Pp.177–182. Doi:<https://doi.org/10.1016/j.dnarep.2018.08.022>.
- Peterson, B.G., Glaser, M.L., Rapoport, T.A. & Baldrige, R.D. (2019). Cycles Of Autoubiquitination And Deubiquitination Regulate The ERAD Ubiquitin Ligase Hrd1. *Elife*, 8. Doi:<https://doi.org/10.7554/elifesciences.50903>.
- Pichler, A., Fatouros, C., Lee, H. & Eisenhardt, N. (2017). SUMO Conjugation – A Mechanistic View. *Biomolecular Concepts*, 8(1), Pp.13–36. Doi:<https://doi.org/10.1515/bmc-2016-0030>.
- Pichler, A., Gast, A., Seeler, J.S., Dejean, A. & Melchior, F. (2002). The Nucleoporin Ranbp2 Has SUMO1 E3 Ligase Activity. *Cell*, 108(1), Pp.109–120. Doi:[https://doi.org/10.1016/S0092-8674\(01\)00633-X](https://doi.org/10.1016/S0092-8674(01)00633-X).
- Pond, K. W., Christelle De Renty, Yagle, M.K. & Ellis, N.A. (2019). Rescue Of Collapsed Replication Forks Is Dependent On NSMCE2 To Prevent Mitotic DNA Damage. *PLOS Genetics*, [Online] 15(2), Pp.E1007942–E1007942. Doi:<https://doi.org/10.1371/journal.pgen.1007942>.
- Potenza, N. (2006). Hybridase Activity Of Human Ribonuclease-1 Revealed By A Real-Time Fluorometric Assay. *Nucleic Acids Research*, 34(10), Pp.2906–2913. Doi:<https://doi.org/10.1093/nar/gkl368>.
- Potts, P.R. & Yu, H. (2005). Human MMS21/NSE2 Is A SUMO Ligase Required For DNA Repair. *Molecular And Cellular Biology*, 25(16), Pp.7021–7032. Doi:<https://doi.org/10.1128/mcb.25.16.7021-7032.2005>.

- Prudden, J., Perry, J.J.P., Arvai, A.S., Tainer, J.A. & Boddy, M.N. (2009). Molecular Mimicry Of SUMO Promotes DNA Repair. *Nature Structural & Molecular Biology*, [Online] 16(5), Pp.509–516. Doi:<https://doi.org/10.1038/Nsmb.1582>.
- Psakhye, I. & Jentsch, S. (2012). Protein Group Modification And Synergy In The SUMO Pathway As Exemplified In DNA Repair. *Cell*, 151(4), Pp.807–820. Doi:<https://doi.org/10.1016/J.Cell.2012.10.021>.
- Rabut, G. & Peter, M. (2008). Function And Regulation Of Protein Neddylolation. *EMBO Reports*, 9(10), Pp.969–976. Doi:<https://doi.org/10.1038/Embor.2008.183>.
- Ramazi, S., Allahverdi, A. & Zahiri, J. (2020) Evaluation Of Post-Translational Modifications In Histone Proteins: A Review On Histone Modification Defects In Developmental And Neurological Disorders. *J. Biosci.*, 45, 135.
- Reindle, A., Belichenko, I., Bylebyl, G.R., Chen, X.L., Gandhi, N. & Johnson, E.S. (2006). Multiple Domains In Siz SUMO Ligases Contribute To Substrate Selectivity. *Journal Of Cell Science*, 119(22), Pp.4749–4757. Doi:<https://doi.org/10.1242/Jcs.03243>.
- Ribet, D. & Cossart, P. (2010). Post-translational modifications in host cells during bacterial infection. *FEBS Letters*, 584(13), pp.2748–2758. doi:<https://doi.org/10.1016/j.febslet.2010.05.012>.
- Rodriguez, M.S., Dargemont, C. & Hay, R.T. (2001). SUMO-1 Conjugation in Vivo Requires Both a Consensus Modification Motif and Nuclear Targeting. *Journal of Biological Chemistry*, 276(16), pp.12654–12659. doi:<https://doi.org/10.1074/jbc.m009476200>.
- Roscic, A., Möller, A., Calzado, M.A., Renner, F., Wimmer, V.C., Gresko, E., Lüdi, K.S. & Schmitz, M. Lienhard (2006). Phosphorylation-Dependent Control Of Pc2 SUMO E3 Ligase Activity By Its Substrate Protein HIPK2. *Molecular Cell*, 24(1), Pp.77–89. Doi:<https://doi.org/10.1016/J.Molcel.2006.08.004>.
- Ryšlavá, H., Doubnerová, V., Kavan, D., & Vaněk, O. (2013). Effect Of Posttranslational Modifications On Enzyme Function And Assembly. *Journal Of Proteomics*, 92, 80–109. <https://doi.org/10.1016/J.Jprot.2013.03.025>
- Rytz, T.C., Miller, M., Mcloughlin, F., Augustine, R.L., Marshall, R.P., Juan, Y., Charng, Y., Scalf, M., Smith, L.M. & Vierstra, R.D. (2018). Sumoylome Profiling Reveals A Diverse Array Of Nuclear Targets Modified By The SUMO Ligase SIZ1 During Heat Stress. *The Plant Cell*, 30(5), Pp.1077–1099. Doi:<https://doi.org/10.1105/Tpc.17.00993>.
- S. Goffinont, Coste, F., Prieu-Serandon, P., Mance, L., Virginie Gaudon, Garnier, N., Castaing, B. & Marcin Suskiewicz (2023). Structural Insights Into The Regulation Of The Human E2~SUMO Conjugate Through Analysis Of Its Stable Mimetic. *Journal Of Biological Chemistry*, 299(7), Pp.104870–104870. Doi:<https://doi.org/10.1016/J.Jbc.2023.104870>.
- Sachdev, S. (2001). Piasy, A Nuclear Matrix-Associated SUMO E3 Ligase, Represses LEF1 Activity By Sequestration Into Nuclear Bodies. *Genes & Development*, 15(23), Pp.3088–3103. Doi:<https://doi.org/10.1101/Gad.944801>.

- Salas-Lloret, D., Jansen, N.S., Easa Nagamalleswari, Van, Ekaterina Gracheva, De, H., Marie, A., Peter, Pichler, A., Goedhart, J., Alfred C.O. Vertegaal & Román González-Prieto (2023). SUMO-Activated Target Traps (Satts) Enable The Identification Of A Comprehensive E3-Specific SUMO Proteome. *Science Advances*, 9(31). Doi:<https://doi.org/10.1126/Sciadv.Adh2073>.
- Sanyal, S., Mondal, P., Sen, S., Sengupta (Bandyopadhyay), S. & Das, C. (2020). SUMO E3 Ligase CBX4 Regulates Htert-Mediated Transcription Of *CDH1* And Promotes Breast Cancer Cell Migration And Invasion. *Biochemical Journal*, 477(19), Pp.3803–3818. Doi:<https://doi.org/10.1042/Bcj20200359>.
- Schellenberg, M.J., Jenna Ariel Lieberman, Andrés Herrero-Ruiz, Butler, L.R., Williams, J., Muñoz-Cabello, A.M., Mueller, G.A., London, R.E., Cortés-Ledesma, F. & R. Scott Williams (2017). ZATT (ZNF451)–Mediated Resolution Of Topoisomerase 2 DNA-Protein Cross-Links. *Science*, [Online] 357(6358), Pp.1412–1416. Doi:<https://doi.org/10.1126/Science.Aam6468>.
- Schramm, F.D., Schroeder, K. & Jonas, K. (2020). Protein Aggregation In Bacteria. *FEMS Microbiology Reviews*, [Online] 44(1), Pp.54–72. Doi:<https://doi.org/10.1093/Femsre/Fuz026>.
- Shaheen, M. & Broxmeyer, H.E. (2018). Chapter 16 - Cytokine/Receptor Families And Signal Transduction. In: R. Hoffman, E.J. Benz, L.E. Silberstein, H.E. Heslop, J.I. Weitz, J. Anastasi, M.E. Salama And S.A. Abutalib, Eds., *Sciencedirect*. [Online] Elsevier, Pp.163–175. Available At: <https://www.sciencedirect.com/science/article/abs/pii/B9780323357623000160> [Accessed 17 Jan. 2024].
- Shen, Z., Pardington-Purtymun, P.E., Comeaux, J.C., Moyzis, R.K. & Chen, D.J. (1996). UBL1, A Human Ubiquitin-Like Protein Associating With Human RAD51/RAD52 Proteins. *Genomics*, 36(2), Pp.271–279. Doi:<https://doi.org/10.1006/Geno.1996.0462>.
- Shi, X., Du, Y., Li, S. & Wu, H. (2022). The Role Of SUMO E3 Ligases In Signaling Pathway Of Cancer Cells. *International Journal Of Molecular Sciences*, 23(7), P.3639. Doi:<https://doi.org/10.3390/Ijms23073639>.
- Shi, Y., Zhang, Y., Lin, S., Wang, C., Zhou, J., Peng, D. & Xue, Y. (2020). Dbpsp 2.0, An Updated Database Of Protein Phosphorylation Sites In Prokaryotes. *Scientific Data*, 7(1). Doi:<https://doi.org/10.1038/S41597-020-0506-7>.
- Shi, Y., Zhang, Y., Lin, S., Wang, C., Zhou, J., Peng, D. & Xue, Y. (2020). Dbpsp 2.0, An Updated Database Of Protein Phosphorylation Sites In Prokaryotes. *Scientific Data*, 7(1). Doi:<https://doi.org/10.1038/S41597-020-0506-7>.
- Smit, J.J. & Sixma, T.K. (2014). RBR E3-Ligases At Work. *EMBO Reports*, 15(2), Pp.142–154. Doi:<https://doi.org/10.1002/Embr.201338166>.
- Song, J., Zhang, Z., Hu, W. & Chen, Y. (2005). Small Ubiquitin-Like Modifier (SUMO) Recognition Of A SUMO Binding Motif. *Journal Of Biological Chemistry*, 280(48), Pp.40122–40129. Doi:<https://doi.org/10.1074/Jbc.M507059200>.

- Soria-Bretones, I., Cepeda-García, C., Checa-Rodríguez, C., Heyer, V., Reina-San-Martin, B., Soutoglou, E. & Huertas, P. (2017). DNA End Resection Requires Constitutive Sumoylation Of CtIP By CBX4. *Nature Communications*, 8(1).  
Doi:<https://doi.org/10.1038/S41467-017-00183-6>.
- Strumillo, M., & Beltrao, P. (2015). Towards The Computational Design Of Protein Post-Translational Regulation. *Bioorganic & Medicinal Chemistry*, 23(12), 2877–2882.  
<https://doi.org/10.1016/J.Bmc.2015.04.056>
- Strunnikov, A.V., Aravind, L. & Koonin, E.V. (2001). *Saccharomyces Cerevisiae* SMT4 Encodes An Evolutionarily Conserved Protease With A Role In Chromosome Condensation Regulation. *Genetics*, 158(1), Pp.95–107.  
Doi:<https://doi.org/10.1093/Genetics/158.1.95>.
- Su, V. & Lau, A.F. (2009). Ubiquitin-Like And Ubiquitin-Associated Domain Proteins: Significance In Proteasomal Degradation. *Cellular And Molecular Life Sciences*, 66(17), Pp.2819–2833. Doi:<https://doi.org/10.1007/S00018-009-0048-9>.
- Suk, T.R., Nguyen, T.T., Fisk, Z.A., Mitkovski, M., Geertsma, H.M., Parmasad, J.-L.A., Heer, M.M., Callaghan, S.M., Benseler, F., Brose, N., Tirard, M. & Rousseaux, M.W.C. (2023). Characterizing The Differential Distribution And Targets Of Sumo1 And Sumo2 In The Mouse Brain. *Iscience*, [Online] 26(4), P.106350.  
Doi:<https://doi.org/10.1016/J.Isci.2023.106350>.
- Sun, H. & Hunter, T. (2012). Poly-Small Ubiquitin-Like Modifier (Polysumo)-Binding Proteins Identified Through A String Search. *Journal Of Biological Chemistry*, 287(50), Pp.42071–42083. Doi:<https://doi.org/10.1074/Jbc.M112.410985>.
- Suzuki, R., Shindo, H., Tase, A., Kikuchi, Y., Shimizu, M. & Yamazaki, T. (2009). Solution Structures And DNA Binding Properties Of The N-Terminal SAP Domains Of SUMO E3 Ligases From *Saccharomyces Cerevisiae* And *Oryza Sativa*. *Proteins: Structure, Function, And Bioinformatics*, 75(2), Pp.336–347.  
Doi:<https://doi.org/10.1002/Prot.22243>.
- Swatek, K.N. & Komander, D. (2016). Ubiquitin Modifications. *Cell Research*, [Online] 26(4), Pp.399–422. Doi:<https://doi.org/10.1038/Cr.2016.39>.
- Takahashi, Y., Kahyo, T., Toh-E, A., Yasuda, H. & Kikuchi, Y. (2001). Yeast Ull1/Siz1 Is A Novel SUMO1/Smt3 Ligase For Septin Components And Functions As An Adaptor Between Conjugating Enzyme And Substrates. *J Biol Chem*, 276(52), Pp.48973–48977. Doi:<https://doi.org/10.1074/Jbc.M109295200>.
- Tesena, P., Korchunjit, W., Taylor, J. & Wongtawan, T. (2017). Comparison Of Commercial RNA Extraction Kits And Qpcr Master Mixes For Studying Gene Expression In Small Biopsy Tissue Samples From The Equine Gastric Epithelium. *Journal Of Equine Science*, 28(4), Pp.135–141.  
Doi:<https://doi.org/10.1294/Jes.28.135>.
- Tirard, M., Hsiao, H. — H., Nikolov, M., Urlaub, H., Melchior, F. & Brose, N. (2012). In Vivo Localization And Identification Of Sumoylated Proteins In The Brain Of His 6 - HA-SUMO1 Knock-In Mice. *Proceedings Of The National Academy Of Sciences Of*

- The United States Of America*, 109(51), Pp.21122–21127.  
Doi:<https://doi.org/10.1073/Pnas.1215366110>.
- Tokarz, P. & Woźniak, K. (2021). SENP Proteases As Potential Targets For Cancer Therapy. *Cancers*, 13(9), P.2059. Doi:<https://doi.org/10.3390/Cancers13092059>.
- Truong, K., Lee, T., Li, B. & Chen, Y. (2012). Sumoylation Of SAE2 C Terminus Regulates SAE Nuclear Localization. *Protein Synthesis And Degradation*, 287(51), Pp.42611–42619. Doi:<https://doi.org/10.1074/Jbc.M112.420877>.
- Truong, K., Lee, T.D. & Chen, Y. (2012). Small Ubiquitin-Like Modifier (SUMO) Modification Of E1 Cys Domain Inhibits E1 Cys Domain Enzymatic Activity\*. *Journal Of Biological Chemistry*, 287(19), Pp.15154–15163. Doi:<https://doi.org/10.1074/Jbc.M112.353789>.
- Ulrich, H.D. (2014). Two-Way Communications Between Ubiquitin-Like Modifiers And DNA. *Nature Structural & Molecular Biology*, 21(4), Pp.317–324. Doi:<https://doi.org/10.1038/Nsmb.2805>.
- Varejão, N., Ibars, E., Lascorz, J., Colomina, N., Torres-Rosell, J. & Reverter, D. (2018). DNA Activates The Nse2/Mms21 SUMO E3 Ligase In The Smc5/6 Complex. *The EMBO Journal*, 37(12). Doi:<https://doi.org/10.15252/Embj.201798306>.
- Varejão, N., Lascorz, J., Codina-Fabra, J., Bellí, G., Borràs-Gas, H., Torres-Rosell, J. & Reverter, D. (2021). Structural Basis For The E3 Ligase Activity Enhancement Of Yeast Nse2 By SUMO-Interacting Motifs. *Nature Communications*, [Online] 12(1), P.7013. Doi:<https://doi.org/10.1038/S41467-021-27301-9>.
- Varejão, N., Lascorz, J., Li, Y. & Reverter, D. (2019). Molecular Mechanisms In SUMO Conjugation. *Biochemical Society Transactions*, 48(1), Pp.123–135. Doi:<https://doi.org/10.1042/Bst20190357>.
- Vlastaridis, P., Kyriakidou, P., Chaliotis, A., Van De Peer, Y., Oliver, S.G. & Amoutzias, G.D. (2017). Estimating The Total Number Of Phosphoproteins And Phosphorylation Sites In Eukaryotic Proteomes. *Gigascience*, [Online] 6(2). Doi:<https://doi.org/10.1093/Gigascience/Giw015>.
- Vozandychova, V., Stojkova, P., Hercik, K., Rehulka, P. & Stulik, J. (2021). The Ubiquitination System Within Bacterial Host–Pathogen Interactions. *Microorganisms*, 9(3), P.638. Doi:<https://doi.org/10.3390/Microorganisms9030638>.
- Wang, M., Jiang, Y., & Xu, X. (2015). A Novel Method For Predicting Post-Translational Modifications On Serine And Threonine Sites By Using Site-Modification Network Profiles. *Molecular Biosystems*, 11(11), 3092–3100.  
<https://doi.org/10.1039/C5mb00384a>
- Wang, R., Huang, S., Fu, X., Huang, G., Yan, X., Yue, Z., Chen, S., Li, Y. & Xu, A. (2017). The Conserved Ancient Role Of Chordate PIAS As A Multilevel Repressor Of The NF-Kb Pathway. *Scientific Reports*, [Online] 7(1), P.17063. Doi:<https://doi.org/10.1038/S41598-017-16624-7>.
- Wang, Y.-C., Peterson, S.E. & Loring, J.F. (2014) Protein Post-Translational Modifications And Regulation Of Pluripotency In Human Stem Cells. *Cell Res.*, 24, 143.

- Wasik, U. & Filipek, A. (2014). Non-Nuclear Function Of Sumoylated Proteins. *Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research*, 1843(12), Pp.2878–2885. Doi:<https://doi.org/10.1016/j.bbamcr.2014.07.018>.
- Watson, I.R., Irwin, M.S. & Ohh, M. (2011). NEDD8 Pathways In Cancer, Sine Quibus Non. *Cancer Cell*, 19(2), Pp.168–176. Doi:<https://doi.org/10.1016/j.ccr.2011.01.002>.
- Weber, A.R., Schuermann, D. & Schär, P. (2014). Versatile Recombinant Sumoylation System For The Production Of SUMO-Modified Protein. *Plos ONE*, 9(7), P.E102157. Doi:<https://doi.org/10.1371/journal.pone.0102157>.
- Weber, J., Polo, S. & Maspero, E. (2019). HECT E3 Ligases: A Tale With Multiple Facets. *Frontiers In Physiology*, [Online] 10. Doi:<https://doi.org/10.3389/fphys.2019.00370>.
- Wei, L., Xing, P., Shi, G., Ji, Z., & Zou, Q. (2019). Fast Prediction Of Protein Methylation Sites Using A Sequence-Based Feature Selection Technique. *IEEE/ACM Transactions On Computational Biology And Bioinformatics*, 16(4), 1264–1273. <https://doi.org/10.1109/TCBB.2017.2670558>
- Werner, A., Moutty, MC., Möller, U. & Melchior, F. (2009). Performing In Vitro Sumoylation Reactions Using Recombinant Enzymes. In: Ulrich, H.D. (Eds) SUMO Protocols. METHODS IN MOLECULAR BIOLOGY™, Vol 497. Humana Press, Totowa, NJ. [https://doi.org/10.1007/978-1-59745-566-4\\_12](https://doi.org/10.1007/978-1-59745-566-4_12)
- Wilkinson, K.A. & Henley, J.M. (2010). Mechanisms, Regulation And Consequences Of Protein Sumoylation. *The Biochemical Journal*, [Online] 428(2), Pp.133–145. Doi:<https://doi.org/10.1042/BJ20100158>.
- Wu, L., Pan, T., Zhou, M., Chen, T., Wu, S.-Y., Xi Lv, Li, J., Yu, F., Guan, Y., Liu, B., Zhang, W., Deng, X., Chen, Q., Liang, A., Lin, Y., Wang, L., Tang, X., Cai, W., Li, L. & He, X. (2022). CBX4 Contributes To HIV-1 Latency By Forming Phase-Separated Nuclear Bodies And Sumoylating EZH2. *EMBO Reports*, 23(7). Doi:<https://doi.org/10.15252/Embr.202153855>.
- Wu, S.-Y. & Cheng, C.-M. (2009). p53 sumoylation: Mechanistic insights from reconstitution studies. *Epigenetics*, 4(7), pp.445–451. doi:<https://doi.org/10.4161/epi.4.7.10030>.
- Xie, N., Zhang, L., Gao, W., Huang, C., Huber, P.E., Zhou, X., Li, C., Shen, G. & Zou, B. (2020). NAD<sup>+</sup> Metabolism: Pathophysiologic Mechanisms And Therapeutic Potential. *Signal Transduction And Targeted Therapy*, [Online] 5. Doi:<https://doi.org/10.1038/S41392-020-00311-7>.
- Xu, Y. & Chou, K.-C. (2016) Recent Progress In Predicting Posttranslational Modification Sites In Proteins. *Curr. Top. Med. Chem.*, 16, 591–603.
- Yang, M., Chia Yang Hsu, Ting, C.-Y., Liu, L.-F. & Hwang, J. (2006a). Assembly Of A Polymeric Chain Of SUMO1 On Human Topoisomerase I In Vitro. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, 281(12), Pp.8264–8274. Doi:<https://doi.org/10.1074/Jbc.M510364200>.

- Yang, S. & Sharrocks, A.D. (2010). The SUMO E3 Ligase Activity Of Pc2 Is Coordinated Through A SUMO Interaction Motif. *Molecular And Cellular Biology*, 30(9), Pp.2193–2205. Doi:<https://doi.org/10.1128/Mcb.01510-09>.
- Yang, S.K., Galanis, A., Witty, J. & Sharrocks, A.D. (2006). An Extended Consensus Motif Enhances The Specificity Of Substrate Modification By SUMO. 25(21), Pp.5083–5093. Doi:<https://doi.org/10.1038/Sj.Emboj.7601383>.
- Yasuhiro M. & Junichi S. (2014). Sumoylation. *Circulation Research*, [Online] 115(8), Pp.686–689. Doi:<https://doi.org/10.1161/Circresaha.114.304989>.
- Yau, T.-Y., Sander, W., Eidson, C. & Courey, A.J. (2021). SUMO Interacting Motifs: Structure And Function. *Cells*, 10(11), P.2825. Doi:<https://doi.org/10.3390/Cells10112825>.
- Yeh, E.T.H. (2008). Sumoylation And De-Sumoylation: Wrestling With Life's Processes. *Journal Of Biological Chemistry*, 284(13), Pp.8223–8227. Doi:<https://doi.org/10.1074/Jbc.R800050200>.
- Yong, D., Green, S.R., Ghiabi, P., Santhakumar, V. & Vedadi, M. (2023). Discovery Of Nedd4 Auto-Ubiquitination Inhibitors. *Scientific Reports*, [Online] 13(1), P.16057. Doi:<https://doi.org/10.1038/S41598-023-42997-Z>.
- Yunus, A.A. & Lima, C.D. (2009). Structure Of The Siz/PIAS SUMO E3 Ligase Siz1 And Determinants Required For SUMO Modification Of PCNA. *Molecular Cell*, 35(5), Pp.669–682. Doi:<https://doi.org/10.1016/J.Molcel.2009.07.013>.
- Zeng, W., Gu, S., Yu, Y., Feng, Y., Xiao, M. & Feng, X.-H. (2021). ZNF451 Stabilizes TWIST2 Through Sumoylation And Promotes Epithelial-Mesenchymal Transition. *American Journal Of Cancer Research*, [Online] 11(3), Pp.898–915. Available At: <https://pubmed.ncbi.nlm.nih.gov/33791162/> [Accessed 17 Jan. 2024].
- Zhao, X. (2018). SUMO-Mediated Regulation of Nuclear Functions and Signaling Processes. *Molecular Cell*, 71(3), pp.409–418. doi:<https://doi.org/10.1016/j.molcel.2018.07.027>.
- Zhu, J., Zhu, S., Guzzo, C.M., Ellis, N.A., Sung, K.S., Choi, C.Y. & Matunis, M.J. (2008). Small Ubiquitin-Related Modifier (SUMO) Binding Determines Substrate Recognition And Paralog-Selective SUMO Modification. *Journal Of Biological Chemistry*, 283(43), Pp.29405–29415. Doi:<https://doi.org/10.1074/Jbc.M803632200>.