



**HAL**  
open science

# Biochimie analytique de complexes de réparation de l'ADN : élaboration d'un système analytique intégré

Vivien Berthelot

► **To cite this version:**

Vivien Berthelot. Biochimie analytique de complexes de réparation de l'ADN : élaboration d'un système analytique intégré. Chimie analytique. Université Paris Sud - Paris XI, 2014. Français. NNT : 2014PA112386 . tel-01135373

**HAL Id: tel-01135373**

**<https://theses.hal.science/tel-01135373>**

Submitted on 25 Mar 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE 470  
CHIMIE DE PARIS-SUD

Laboratoire de Chimie-physique  
LCP – UMR CNRS 8000

THÈSE DE DOCTORAT  
CHIMIE

par

**M. Vivien BERTHELOT**

*Biochimie analytique de complexes de réparation de l'ADN :  
élaboration d'un système analytique intégré*

**Date de soutenance : 12/12/2014**

**Composition du jury :**

Directeur de thèse : D<sup>r</sup> Filippo RUSCONI, HDR, CR CNRS (LCP–Université Paris-Sud)

Rapporteurs : D<sup>r</sup> Évelyne SAGE, DR CNRS émérite (Institut Curie, Orsay)

D<sup>r</sup> Régis DANIEL, DR CNRS (Université d'Évry)

Examineurs : P<sup>r</sup> Chantal HOUÉE-LEVIN (Université Paris-Sud)

D<sup>r</sup> Pascale BERTRAND, HDR, CR CEA (CEA, Fontenay-aux-Roses)

D<sup>r</sup> Thierry RABILLOUD, DR CNRS (CEA, Grenoble)



## RÉSUMÉ

Les cassures double-brin de l'ADN sont réparées dans les cellules eucaryotes selon deux voies principales : la voie de recombinaison homologue et la voie de jonction des extrémités non-homologues, toutes deux bien décrites dans la littérature. Cependant, quelques zones d'ombre persistent quant à deux aspects singuliers de leur mise en œuvre :

- \* Si ces deux mécanismes peuvent opérer dans les cellules, quels sont les déterminismes qui président au choix d'une voie de réparation plutôt que l'autre ?
- \* Dans le cas où les cassures double-brin sont induites dans l'ADN par des rayonnements ionisants — comme ceux employés en radio-thérapie anticancéreuse — comment s'opère la réparation lorsque les extrémités générées ne sont pas compatibles avec une ligation immédiate ? Connaître les protéines impliquées dans ce cas permettrait d'élaborer des adjuvants à la thérapie anticancéreuse.

Afin de contribuer à répondre à ces questionnements, nous avons voulu élaborer un système analytique intégré qui permet 1) le recrutement spécifique de complexes de réparation des cassures double-brin de l'ADN sur des phases chromatographiques élaborées au laboratoire ; 2) la résolution des complexes purifiés en 1) sur gel d'acrylamide non dénaturant et leur visualisation ; 3) la caractérisation biochimique fine des complexes séparés en 2).

La méthodologie élaborée au cours de cette thèse a concerné chacun des trois points ci-dessus : 1) nous avons conçu et mis en œuvre un système chromatographique nous permettant de distinguer les protéines recrutées spécifiquement sur des oligonucléotides duplexes d'ADN dotés d'extrémités libres (mimant des cassures double-brin) des autres protéines se liant sur la séquence interne des oligonucléotides ; 2) nous avons adapté à notre problématique une méthodologie d'électrophorèse non dénaturante permettant la résolution des complexes purifiés tout en garantissant leur intégrité au cours de la migration électrophorétique ; 3) grâce à la visualisation directe des complexes résolus dans le gel, nous avons pu déterminer par spectrométrie de masse leur composition en protéines.

L'étude biochimique des complexes purifiés a permis de montrer qu'ils étaient fonctionnels, c'est-à-dire qu'ils étaient capables de liguer deux oligonucléotides entre eux. La fouille des données de spectrométrie de masse, obtenues à partir d'un grand nombre d'expériences indépendantes, nous a permis de montrer que nous purifions bien des complexes de la physiologie de l'ADN et en particulier des complexes représentatifs de la diversité des mécanismes de réparation de l'ADN.

De manière intéressante, nous avons pu observer que certaines protéines recrutées spécifiquement sur les mimes de cassures double-brin, ne sont pourtant pas connues pour participer aux processus de réponse aux dommages de l'ADN (synthèse de nucléotides, moteur moléculaire, checkpoint, topologie de l'ADN, cytosquelette).

Le rôle des protéines évoquées ci-dessus sera prochainement caractérisé en cellule, notamment avec des stratégies de type RNAi. D'autre part, nous utiliserons les développements méthodologiques décrits ci-dessus pour étudier les mécanismes de réparation des cassures double-brin radio-induites dans l'ADN, tels qu'ils sont mis en jeu dans les cellules tumorales irradiées. Les phases chromatographiques seront pré-irradiées avec des rayonnements ionisants et seront utilisées pour purifier les protéines d'intérêt dont l'expression aura été induite auparavant, dans des cellules en culture, par irradiation  $\gamma$ .

Au cours de cette thèse, j'ai concentré mes efforts sur deux projets scientifiques.

Le premier projet a été la caractérisation des sites d'oxydation par les rayonnements ionisants sur une protéine modèle, la protéine fluorescente cyan. Ce travail, qui a requis la maîtrise de la chromatographie hautes performances de peptides et de la spectrométrie de masse haute résolution, a donné lieu à une publication <sup>a</sup>.

Le deuxième volet porte sur la réparation des cassures double-brin de l'ADN, qui fait l'objet de ce mémoire. Une publication est en cours de rédaction, qui reprendra les résultats décrits dans ce mémoire.

---

<sup>a</sup>. V. Berthelot, V. Steinmetz, L. A. Alvarez, C. Houée-Levin, F. Merola, F. Rusconi\*, et M. Erard\*. An analytical workflow for the molecular dissection of irreversibly modified fluorescent proteins (2013) *Anal Bioanal Chem* (27) : 8789–8798. \*Corresponding authors.

## REMERCIEMENTS

Je remercie le P<sup>r</sup> Mehran MOSTAFAVI pour m'avoir permis d'effectuer ma thèse au sein de son laboratoire.

Je voudrais remercier la D<sup>r</sup> Carine GIOVANNANGELI et le P<sup>r</sup> Jean-François RIOU qui m'ont accueilli dans leur laboratoire, au Muséum, pendant ces trois années de thèse.

Je veux exprimer ici toute ma gratitude à mon directeur de thèse, le D<sup>r</sup> Filippo RUSCONI pour m'avoir guidé dans ce travail par ses conseils et communiqué son enthousiasme.

Je remercie vivement mon encadrante au Muséum, le D<sup>r</sup> Danièle PRASEUTH, pour ses encouragements, ses conseils ainsi que son aide indispensable pour nos expériences « à quatre mains ».

Je remercie chaleureusement la D<sup>r</sup> Marie ERARD pour avoir accepté ma contribution à son travail sur les protéines fluorescentes.

Je veux dire ici mes remerciements aux membres du jury — la P<sup>r</sup> Chantal HOUÉE-LEVIN, les D<sup>rs</sup> Évelyne SAGE, Pascale BERTRAND, Régis DANIEL et Thierry RABILLOU — pour avoir accepté de juger mon travail de thèse et de participer à ce jury.

Je veux remercier le D<sup>r</sup> François GUILLONNEAU, et les membres de la plateforme de spectrométrie de masse de l'université Paris Descartes, ainsi que M. Olivier LANGELLA et les membres de la plateforme de protéomique de l'université Paris Sud pour les analyses par spectrométrie de masse des bandes obtenues sur gel natif.

Merci également à Gildas MOUTA-CARDOSO pour m'avoir formé à la technique du Blue Native PAGE, à Loïc PERROUAULT et Gwénola DORÉ pour l'aide qu'ils m'ont apportée. Je veux aussi remercier MM. Frédéric DA COSTA et Bernard TARROU, ainsi que M. Gilles GARCIA pour la conception et la réalisation du bloc de teflon utilisé pour le photoclivage. Je remercie M. Julien COCHENNEC pour les discussions intéressantes sur l'analyse des données de spectrométrie de masse. Je remercie aussi les D<sup>r</sup> Anne DE CIAN et Emmanuelle DELAGOUTTE pour leurs conseils et l'intérêt qu'elles ont porté à mes travaux.

Je remercie le D<sup>r</sup> Érika Brunet pour nos discussions scientifiques.

Je remercie vivement tous les membres du Laboratoire de chimie physique (LCP, université Paris Sud) et du Laboratoire de Structure et instabilité des génomes (Muséum) pour m'avoir soutenu dans les moments les plus difficiles, ainsi que pour tous les moments de convivialité partagés. La liste des noms des personnes auxquelles je fais référence serait trop longue pour être contenue sur cette page, je voudrais cependant mentionner particulièrement Karine, Armêl et Marine qui m'ont plus particulièrement épaulé et Élisabeth, Mathilde, Astrid, Mélody et Florian qui ont partagé mon bureau.

Je remercie la Présidence de l'Université Paris-Sud pour le contrat doctoral « Attractivité-Présidence », sans lequel je n'aurais pu effectuer cette thèse.

Enfin, je dédie cette thèse à mon épouse Aurélie, qui m'a soutenu tout au long de ces trois années, à ma fille Alice, qui a été si patiente avec moi et à notre futur enfant, qui aura, à sa manière, partagé avec nous les derniers moments de cette thèse.



# Table des matières

Remerciements	5
Abréviations — Glossaire	xiii
<b>1 Introduction</b>	<b>1</b>
1.1 Les origines des dommages de l'ADN . . . . .	4
1.1.1 Mécanismes impliquant des coupures de l'ADN . . . . .	5
1.1.1.1 La méiose . . . . .	5
1.1.1.2 V(D)J et recombinaison de conversion de classe . . . . .	5
1.1.2 Facteurs endogènes toxiques pour l'ADN . . . . .	6
1.1.2.1 Les produits secondaires du métabolisme . . . . .	6
1.1.2.2 Les dysfonctionnements enzymatiques . . . . .	7
1.1.3 Facteurs exogènes toxiques pour l'ADN . . . . .	7
1.2 Réparation de l'ADN . . . . .	8
1.2.1 Le BER (« <i>base-excision repair</i> ») . . . . .	8
1.2.2 Le NER (« <i>nucleotide excision repair</i> ») . . . . .	11
1.2.3 Le MMR (« <i>mismatch repair</i> ») . . . . .	12
1.2.4 La HR (« <i>homologous recombination</i> ») . . . . .	13
1.2.5 NHEJ « <i>non homologous end joining</i> » . . . . .	15
1.2.6 NHEJ « alternatif » . . . . .	16
1.2.7 Conclusions. . . . .	17
1.3 Choix de la voie de réparation . . . . .	18
1.3.1 Régulation des différentes voies de réparation . . . . .	18
1.3.1.1 Le niveau de différenciation cellulaire . . . . .	18
1.3.1.2 La phase du cycle cellulaire . . . . .	18
1.3.1.3 La nature chimique des extrémités . . . . .	19
1.3.1.4 Le contexte chromatinien de la lésion . . . . .	19
1.3.2 Rôle des trois ligases humaines dans les mécanismes de jonction . . . . .	20
1.3.3 Signatures moléculaires des voies de réparation . . . . .	20
1.3.4 Choix de la voie de réparation . . . . .	23
1.3.4.1 Équilibre entre HR et NHEJ . . . . .	23

## BIOCHIMIE ANALYTIQUE DES COMPLEXES DE RÉPARATION DE L'ADN

1.3.4.2	Équilibre entre les voies du NHEJ canonique et alternatif . . . . .	23
1.3.4.3	Hiérarchie des mécanismes mis en jeu . . . . .	24
1.3.5	Quels enjeux du choix de la voie de réparation? . . . . .	24
1.4	La réponse intégrée aux dommages de l'ADN . . . . .	26
1.4.1	Détection des dommages et transduction du signal . . . . .	26
1.4.1.1	Détection des CDB dans le NHEJ canonique . . . . .	27
1.4.1.2	Détection des CDB dans le NHEJ alternatif . . . . .	28
1.4.1.3	Détection des CDB dans la HR . . . . .	28
1.4.2	Remodelage de la chromatine . . . . .	28
1.4.3	Cycle cellulaire et apoptose . . . . .	29
1.5	Réparation et thérapie anticancéreuse . . . . .	30
1.5.1	Irradiation des cellules tumorales . . . . .	30
1.5.2	Exploitation des défauts des cellules tumorales . . . . .	31
1.6	Étude des interactions protéine-ADN . . . . .	34
1.6.1	Méthodes d'étude <i>in vivo</i> . . . . .	35
1.6.1.1	ChIP « <i>chromatin immunoprecipitation</i> » . . . . .	35
1.6.1.2	Double et triple hybride . . . . .	35
1.6.2	In silico . . . . .	36
1.6.3	Méthodes d'étude <i>in vitro</i> . . . . .	36
1.6.3.1	EMSA (« <i>electrophoretic mobility shift assay</i> ») . . . . .	36
1.6.3.2	Empreinte aux endonucléases . . . . .	38
1.6.3.3	Puces à ADN . . . . .	38
1.6.3.4	Southwestern blot . . . . .	38
1.6.3.5	Purification par affinité . . . . .	39
1.6.4	Conclusion . . . . .	41
<b>2</b>	<b>Matériels et méthodes</b> . . . . .	<b>43</b>
2.1	Tampons pour les phases chromatographiques . . . . .	45
2.2	Préparation des phases chromatographiques . . . . .	45
2.2.1	Préparation du matériel nucléaire . . . . .	46
2.2.1.1	Synthèse des oligonucléotides par PCR . . . . .	46
2.2.1.2	Purification des produits de PCR . . . . .	47
2.2.2	Constitution des phases chromatographiques . . . . .	48
2.3	Vérification de la phase chromatographique contrôle . . . . .	49
2.4	Tampons pour la purification des protéines . . . . .	49
2.5	Purification de complexes nucléoprotéiques . . . . .	50
2.5.1	Étape de purification . . . . .	50
2.5.2	Le photoclivage UV . . . . .	50
2.6	Séparations électrophorétiques . . . . .	52
2.6.1	Tampons pour les électrophorèses et solutions pour les colorations . . . . .	52

## TABLE DES MATIÈRES

2.6.2	Électrophorèse en conditions natives :	
	migration et coloration des gels . . . . .	53
2.6.2.1	La migration électrophorétique en BN PAGE	53
2.6.2.2	La coloration analytique des gels . . . . .	54
2.6.2.2.1	La coloration à l'argent . . . . .	54
2.6.2.2.2	La coloration au bleu de Coomassie colloïdal . . . . .	55
2.6.3	Électrophorèse en conditions dénaturantes :	
	migration et coloration des gels . . . . .	55
2.6.3.1	La migration électrophorétique dénaturante	55
2.6.3.2	La coloration au SYPRO . . . . .	56
2.7	Expérience d'empreinte aux nucléases . . . . .	56
2.8	Séquençage de l'ADN des complexes . . . . .	56
2.9	Activité de ligation des complexes purifiés . . . . .	57
2.9.1	Ligation par les extraits nucléaires bruts . . . . .	57
2.9.2	Ligation avec des complexes purifiés . . . . .	57
2.9.2.1	Ligation avec des complexes purifiés attachés aux billes . . . . .	58
2.9.2.2	Ligation en l'absence de supports chromatographiques . . . . .	58
2.10	Spectrométrie de masse . . . . .	58
<b>3</b>	<b>Résultats</b>	<b>61</b>
3.1	Un système chromatographique intégré . . . . .	67
3.1.1	Production des déterminants d'affinité par PCR . . . . .	69
3.1.2	Purification des produits de PCR . . . . .	70
3.1.3	Constitution des phases chromatographiques . . . . .	73
3.1.4	Contrôle de la qualité des phases produites . . . . .	73
3.1.5	Élimination du support chromatographique . . . . .	76
3.2	Purification de protéines de réparation . . . . .	78
3.2.1	Électrophorèse en conditions natives . . . . .	78
3.2.1.1	Les bandes contiennent des complexes multiprotéiques . . . . .	81
3.2.2	Points techniques critiques . . . . .	82
3.2.3	Étude qualitative des produits de purification . . . . .	83
3.2.3.1	Les complexes visualisés comportent-ils de l'ADN ? . . . . .	84
3.2.3.2	Quelle est la topologie des complexes nucléoprotéiques ? . . . . .	86
3.2.3.3	Les complexes sont-ils spécifiques des extrémités de l'ADN ? . . . . .	88
3.2.4	Étude fonctionnelle des produits de purification . . . . .	89
3.2.4.1	Ligation par les complexes encore sur les billes	90

## BIOCHIMIE ANALYTIQUE DES COMPLEXES DE RÉPARATION DE L'ADN

3.2.4.2	Ligation par les complexes détachés des billes	92
3.2.4.3	Réflexions sur les expériences de ligation . . .	94
3.3	Analyses par spectrométrie de masse . . . . .	95
3.3.1	Répartition fonctionnelle (en nombre) . . . . .	99
3.3.2	Répartition fonctionnelle (en quantité) . . . . .	99
3.3.3	Comparaison des bandes obtenues sur phase ✱ . . .	101
3.3.3.1	Protéines exclusives d'une bande donnée . . .	102
3.3.3.1.1	La méthode de calcul en moyenne « totale » . . . . .	102
3.3.3.1.2	La méthode de calcul en moyenne « restreinte » . . . . .	104
3.3.3.1.3	Les protéines désignées comme ex- clusives de la bande à 720 kDa . . . . .	107
3.3.3.1.4	Les protéines désignées comme ex- clusives de la bande à 480 kDa . . . . .	109
3.3.3.1.5	L'analyse des protéines exclusives de la bande à 240 kDa . . . . .	109
3.3.3.1.6	Les protéines exclusives de la bande « aspécifique » . . . . .	110
3.3.3.2	Analyse de la répartition des protéines de réparation dans les quatre bandes . . . . .	112
3.3.3.2.1	Interactions protéine-protéine. . . . .	113
3.3.3.2.2	Spécialisation fonctionnelle des com- plexes. . . . .	114
3.3.4	Comparaison des bandes obtenues sur phase ✱/✧ . . .	115
3.3.4.1	Protéines spécifiques de la phase ✱, bande à 720 kDa . . . . .	121
3.3.4.2	Protéines spécifiques de la phase ✱, bande à 480 kDa . . . . .	122
3.3.4.3	Protéines spécifiques de la phase ✱, bande à 240 kDa . . . . .	124
3.3.4.4	Protéines spécifiques de la phase ✱, bande aspécifique . . . . .	125
3.3.5	Conclusions sur la fouille des données . . . . .	126
3.3.5.1	Répartition des protéines dans les bandes à différents poids moléculaires (phase ✱) . . . . .	126
3.3.5.1.1	Caractère significatif des assemblages non covalents tels qu'ils sont résolus sur gel. . . . .	126
3.3.5.1.2	Recherche de protéines « exclusives » d'une bande donnée. . . . .	126

## TABLE DES MATIÈRES

3.3.5.1.3	Abondance relative des protéines de réparation dans les différentes bandes des pistes ✱. . . . .	127
3.3.5.2	Répartition des protéines dans les produits de purification sur phases ✱ et ✧ (par poids moléculaire) . . . . .	128
3.3.5.3	Tableau récapitulatif de toutes les données de masse . . . . .	129
<b>4</b>	<b>Conclusions &amp; discussion — Perspectives</b>	<b>135</b>
4.1	Conclusions & discussion . . . . .	137
4.1.1	Coexistence de trois voies de réparation de l'ADN . . . . .	138
4.1.2	Activité des complexes des différentes bandes . . . . .	139
4.1.3	Recherche de nouvelles protéines de réparation . . . . .	140
4.1.3.1	Présentation de protéines « candidats » . . . . .	140
4.1.3.2	Stratégies d'étude des protéines candidats . . . . .	142
4.1.4	Comment envisager l'identification de nouvelles protéines candidats? . . . . .	144
4.1.5	Certaines protéines sont absentes, pourquoi? . . . . .	146
4.2	Perspectives . . . . .	146
4.2.1	Les dommages à chimie complexe . . . . .	147
4.2.2	Les sites à dommages multiples . . . . .	149
<b>5</b>	<b>Annexes</b>	<b>151</b>
	Fiches relatives aux protéines d'intérêt . . . . .	153
5.0.3	Protéines révélées par la fouille des données . . . . .	153
5.0.4	Protéines commentées dans le chapitre « Résultats » . . . . .	191
	Fiches relatives aux protéines commentées . . . . .	191
	Manuel de l'utilisateur du logiciel <i>dsbRepair</i> (anglais) . . . . .	195
5.1	Introduction . . . . .	197
5.2	Getting the software and installing it . . . . .	197
5.3	Overview of the biochemical experiments . . . . .	198
5.4	General overview of the graphical user interface . . . . .	204
5.4.1	The main window . . . . .	204
5.4.2	The results window . . . . .	207
5.4.3	The history window . . . . .	207
5.4.4	The console window . . . . .	207
5.4.5	Common features . . . . .	207
5.5	Various calculations using sample lists . . . . .	211
5.5.1	The three general procedures available . . . . .	211
5.5.1.1	List exclusive proteins . . . . .	212
5.5.1.2	Ratio of calculated values . . . . .	215

## BIOCHIMIE ANALYTIQUE DES COMPLEXES DE RÉPARATION DE L'ADN

5.5.1.2.1	Example with a <i>Sum</i> -based computation . . . . .	216
5.5.1.2.2	Example with <i>Mean total</i> -based computation . . . . .	216
5.5.1.2.3	Example with <i>Mean restricted</i> -based computation . . . . .	217
5.5.1.3	Calculation on ratio values . . . . .	217
5.5.2	Specific handling of the <i>sample units</i> . . . . .	218
5.5.3	Global options . . . . .	218
5.6	User-crafted custom SQL queries . . . . .	218
5.6.1	Crash course in SQL <i>SELECT</i> queries . . . . .	219
5.7	Appendix . . . . .	226
	GNU General Public License Text . . . . .	226

# Table des figures

<b>Introduction</b>	<b>1</b>
1.1 Phases du cycle cellulaire . . . . .	4
1.2 Différentes voies de réparation en fonction des agents créant les dommages . . . . .	9
1.3 Réparation des dommages à l'ADN par la voie de l'excision de base (BER) . . . . .	10
1.4 Différentes voies du NER . . . . .	12
1.5 Voies de réparation des cassures double-brin de l'ADN . . . . .	14
1.6 Spécificité des trois DNA-ligases humaines . . . . .	21
1.7 Utilisation des microhomologies pour la réparation des cassures double-brin . . . . .	22
1.8 Translocations induites par apparition de cassures double-brin . . . . .	25
1.9 Modifications post traductionnelles et interactions impliquées dans le choix de la voie de réparation des CDB . . . . .	27
1.10 Structures d'oligonucléotides duplexes d'ADN à dommages multiples . . . . .	31
1.11 Principe du gel retard en deux dimensions . . . . .	37
1.12 Principe de l'empreinte aux nucléases . . . . .	39
<b>Matériels et méthodes</b>	<b>43</b>
2.1 Dispositif de photoclivage . . . . .	51
<b>Résultats</b>	<b>61</b>
3.1 Stratégie analytique mise en œuvre . . . . .	66
3.2 Structure des phases chromatographiques employées . . . . .	67
3.3 Groupement photoclivable . . . . .	68
3.4 Utilité d'un lien photoclivable entre l'oligonucléotide et les billes . . . . .	69
3.5 Production des oligonucléotides fonctionnalisant les billes . . . . .	71
3.6 Élimination des amorces de PCR résiduelles par chromatographie . . . . .	72

## BIOCHIMIE ANALYTIQUE DES COMPLEXES DE RÉPARATION DE L'ADN

3.7	Contrôle de la qualité de la phase dénuée d'extrémités libres de l'ADN . . . . .	74
3.8	Fixation aspécifique des oligonucléotides sur les billes . . . . .	75
3.9	Double fixation des oligonucléotides de la phase contrôle . . . . .	77
3.10	Structure du bleu de Coomassie G-250 . . . . .	79
3.11	Séparation électrophorétique des produits de purification sur les deux phases chromatographiques . . . . .	80
3.12	Électrophorèse de deuxième dimension en conditions dénaturantes . . . . .	83
3.13	Suivi des oligonucléotides dans les complexes purifiés . . . . .	85
3.14	Empreinte aux nucléases et séquençage de l'ADN protégé . . . . .	87
3.15	Ligation par les complexes nucléoprotéiques encore sur les billes . . . . .	91
3.16	Ligation d'oligonucléotides attachés aux billes . . . . .	92
3.17	Ligation d'oligonucléotides détachés des billes . . . . .	93
3.18	Gel non dénaturant obtenu pour une expérience type de purification sur les phases $\star$ et $\diamond$ . . . . .	98
3.19	Répartition fonctionnelle (en nombre) des cent protéines les plus abondantes . . . . .	100
3.20	Répartition fonctionnelle (en quantité) des cent protéines les plus abondantes (sans PARP1) . . . . .	100
3.21	Répartition fonctionnelle (en quantité) des cent protéines les plus abondantes (sans PARP1 ni RLA1) . . . . .	101
3.22	Description des méthodes de calcul utilisées pour déterminer la proportion moyenne d'une protéine dans une bande donnée par rapport aux autres . . . . .	103
3.23	Répartition d'un panel de protéines impliquées dans la réparation de l'ADN au sein de différentes bandes . . . . .	113
3.24	Description de la méthode de calcul utilisée pour déterminer le facteur d'enrichissement d'une protéine dans une des deux phases chromatographiques . . . . .	116
	<b>Manuel de l'utilisateur du logiciel <i>dsbRepair</i> (anglais)</b>	<b>195</b>
5.1	Structure of the two different phases . . . . .	199
5.2	Typical blue native PAGE obtained throughout this study . . . . .	201
5.3	Relational database structure . . . . .	203
5.4	Program's main window . . . . .	205
5.5	Program's main window (Sample lists table view) . . . . .	206
5.6	Results window . . . . .	208
5.7	History window . . . . .	209
5.8	Console window . . . . .	210
5.9	Table view contents for the List exclusive proteins procedure . . . . .	212
5.10	Program's main window . . . . .	218

TABLE DES FIGURES

5.11 Program's main window . . . . . 219  
5.12 SQL queries window . . . . . 220

# BIOCHIMIE ANALYTIQUE DES COMPLEXES DE RÉPARATION DE L'ADN

# Liste des tableaux

<b>Matériels et méthodes</b>	<b>43</b>
2.1 Tampons utilisés pour la constitution des phases chromatographiques . . . . .	45
2.3 Programme de PCR . . . . .	47
2.2 Composition du milieu réactionnel de PCR . . . . .	47
2.4 Tampons pour la purification des protéines . . . . .	49
2.5 Tampons pour les électrophorèses et les colorations . . . . .	52
<b>Résultats</b>	<b>61</b>
3.1 Protéines exclusives par poids moléculaire . . . . .	105
3.2 Protéines exclusives de la bande « aspécifique » . . . . .	111
3.3 Protéines spécifiques de la phase * . . . . .	117
3.4 Protéines spécifiquement recrutées sur la phase * (tous poids moléculaires confondus) . . . . .	129
3.5 Résumé général des protéines d'intérêt . . . . .	130

# BIOCHIMIE ANALYTIQUE DES COMPLEXES DE RÉPARATION DE L'ADN

## Abréviations — Glossaire

- \* BN PAGE « *blue native PAGE* »  
électrophorèse native au bleu de Coomassie ;
- \* CDB cassures double-brin (« *double-strand breaks* ; *DSB* ») ;
- \* CES chromatographie d'exclusion stérique, qui permet de séparer de manière non destructrice des molécules (ou assemblages) selon leur taille. En anglais : « *size exclusion chromatography* ; *SEC* » ;
- \* CRISPR « *clustered regularly interspaced short palindromic repeats* »  
motifs génomiques retrouvés dans les bactéries et les archae ;
- \* CSB cassures simple-brin (« *single-strand breaks* ; *SSB* ») ;
- \* CSR (« *class-switch recombination* »)  
recombinaison de conversion de classe ; RCC ;
- \* CDS (« *clustered damage sites* »)  
sites de dommages groupés ;
- \* DTT dithiothréitol (agent réducteur) ;
- \* DSB voir CDB ;
- \* EDTA acide éthylène diamine tétraacétique (agent chélateur des métaux divalents) ;
- \* GPL « *General Public License* »  
licence publique générale (informatique) ;
- \* GUI « *graphical user interface* »  
interface graphique utilisateur (informatique) ;
- \* HEPES acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique (tampon) ;
- \* HPLC « *high performance liquid chromatography* »  
chromatographie liquide haute performance (séparation et purification) ;
- \* HTML « *hypertext markup language* » langage de balises hypertexte (format standard de l'internet) ;
- \* HR « *homologous recombination* »  
recombinaison homologue ;

- \* ICAT « *isotope-coded affinity tag* »  
étiquette d'affinité codée par des isotopes (quantification d'analytes) ;
- \* IEF isoélectrofocalisation (séparation et purification) ;
- \* IgM immunoglobuline M ;
- \* IPG « *immobilized pH gradient* »  
gradient de pH immobilisé (séparation et purification) ;
- \* LC « *liquid chromatography* »  
chromatographie liquide (séparation et purification) ;
- \* LC-MS « *liquid chromatography-mass spectrometry* »  
chromatographie liquide couplée à la spectrométrie de masse (séparation et analyse) ;
- \* MMEJ « *microhomology end-joining* »  
réparation par jonction d'extrémités à microhomologie ;
- \* MMR « *mismatch repair* »  
réparation des mésappariements
- \* MS « *mass spectrometry* »  
spectrométrie de masse
- \* NER « *nucleotide excision repair* »  
réparation par excision de nucléotide ;
- \* NP-40 Nonidet P-40 (détergent non ionique et non dénaturant) ;
- \* PAGE « *polyacrylamide gel electrophoresis* »  
électrophorèse en gel de polyacrylamide ;
- \* PCR « *polymerase chain reaction* »  
réaction de polymérisation en chaîne ;
- \* PMF « *peptide mass fingerprint* »  
empreinte peptidique massique ;
- \* PMSF Fluorure de phénylméthylsulfonyl (inhibiteur de protéases à sérine) ;
- \* RCC voir CSR ;
- \* RP « *reversed-phase* »  
phase inversée (séparation et purification) ;
- \* RPS résonance plasmonique de surface  
en anglais, (« *surface plasmon resonance* ; *SPR* ») ;
- \* SDS « *sodium dodecylsulfate* »  
dodécylsulfate de sodium (détergent puissant) ;
- \* SM spectrométrie de masse ;
- \* SM-MI spectrométrie de masse couplée à la mobilité ionique ;

- \* SM/SM spectrométrie de masse en tandem (fragmentation d'un pré-curseur peptidique pour en déterminer la séquence) ;
- \* TAP-TAG « *tandem affinity-purification tag* »  
étiquette de purification par affinité en tandem ;
- \* TFA « *trifluoroacetic acid* »  
acide trifluoroacétique ;
- \* V(D)J recombinaison somatique (processus de fabrication des séquences variables des anticorps) ;
- \* XML « *extended markup language* »  
langage de balises étendu (héritier du HTML) ;



# CHAPITRE 1

## INTRODUCTION



L'ADN (acide désoxyribonucléique) est la molécule qui contient l'information génétique, présente chez tous les êtres vivants. Cette information génétique permet le développement des cellules et la transmission de caractères héréditaires aux générations suivantes. Dans les cellules eucaryotes, l'ADN est situé dans un compartiment appelé noyau, où se déroulent la transcription (utilisation de l'information contenue dans les gènes pour produire des protéines), la réplication (duplication des molécules d'ADN afin de permettre la transmission de l'information génétique dans chacune des deux cellules filles lors de la division cellulaire), mais également la réparation (restauration de l'intégrité des molécules d'ADN suite à un dommage). Dans le noyau, l'ADN est présent sous la forme d'une hélice [Watson et Crick, 1953] constituée de deux brins associés par un réseau de liaisons hydrogène. Chaque brin est composé d'une succession de quatre nucléotides différents : A (adénine), T (thymine), G (guanine), C (cytosine), dont l'enchaînement détermine ce qui est appelé la « séquence » de la molécule d'ADN. Les séquences des deux brins sont complémentaires (un A sur l'un des brins correspond à un T sur l'autre brin, un C à un G, un T à un A et un G à un C).

L'ADN comme molécule, dont le rôle est de fournir aux cellules filles le patrimoine génétique de la cellule mère, voit son intégrité structurale mise à mal un grand nombre de fois par cellule et par jour. Or, si la stabilité chimique et structurale de l'ADN est mise à mal, la stabilité génétique du lignage cellulaire l'est également. C'est pour cela que l'évolution a mis en place des mécanismes de réparation de l'ADN. Comme les dommages à l'ADN peuvent être d'une grande variété, les mécanismes de réparation se sont adaptés à cette variété.

Dans notre laboratoire, nous nous intéressons à la réparation des cassures double-brin de l'ADN. Comme décrit plus bas, la réparation de ces cassures constitue un enjeu important du point de vue biomédical, car ces cassures sont induites par les rayonnements ionisants (ou des radiomimétiques) que l'on emploie dans la thérapie anticancéreuse. La réparation des cassures double-brin radioinduites dans l'ADN est problématique pour de multiples raisons et nous avons voulu, dans ce travail, initier un projet dans lequel nous avons commencé par élaborer une méthodologie nouvelle de purification des complexes de réparation de l'ADN.

Dans cette introduction, nous présenterons l'intérêt de l'étude des protéines de réparation de l'ADN, nous aborderons brièvement les différentes voies qui permettent cette réparation (pour des revues détaillées sur ce sujet, voir Ciccia et Elledge [2010]; Thompson [2012]), ainsi que les méthodologies disponibles pour étudier les interactions protéines-ADN en général (pour des revues détaillées, voir Dey *et al.* [2012]; Simicevic et Deplancke [2010]; Tacheny *et al.* [2013]).

Pour le lecteur n'étant pas familier du jargon des biologistes cellulaires,

nous prenons la liberté de représenter ici très schématiquement le cycle cellulaire (Figure 1.1). En effet, tant dans cette introduction que dans les chapitres qui suivent, les différentes phases du cycle cellulaire seront évoquées maintes et maintes fois. Nous verrons, en effet, combien la phase du cycle cellulaire dans laquelle se trouve une cellule conditionne le traitement des dommages intervenant sur son ADN.

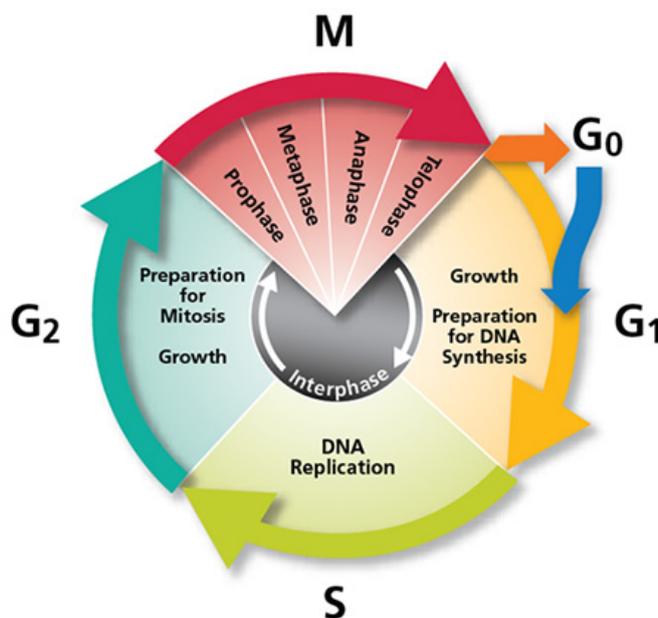


FIGURE 1.1 – **Phases du cycle cellulaire.** Pour les préoccupations qui sont les nôtres, le concept fondamental est le suivant : durant quelles phases du cycle cellulaire une cellule dispose-t-elle de deux copies de son génome ? Une cellule possède deux copies complètes de son génome au terme de la réplication (en fin de phase S), pendant la phase G<sub>2</sub> et pour une partie de la phase M. En revanche, durant la phase G<sub>1</sub> et pour une bonne partie des phases S et M, la cellule ne dispose que d'une copie complète de son génome. Tiré de <http://www.bdbiosciences.com>.

## 1.1 LES ORIGINES DES DOMMAGES DE L'ADN

Alors que l'information contenue dans la séquence nucléotidique est primordiale pour les cellules, de nombreux facteurs peuvent altérer cette séquence (substitution d'un nucléotide, insertion ou délétion d'une partie de la séquence). Parmi ces facteurs, les dommages à l'ADN sont une source importante d'instabilité pour le génome.

On estime à plus de  $10^5$  le nombre de dommages subis par l'ADN dans chacune de nos cellules chaque jour. Les causes de ces dommages sont extrêmement variées : certains sont mis en place physiologiquement, d'autres sont une conséquence indésirable du fonctionnement cellulaire, d'autres encore proviennent de facteurs environnementaux. Dans les sections suivantes, nous présenterons quelques exemples reflétant la diversité des contextes dans lesquels l'ADN peut être endommagé (pour une revue, voir [Cadet et Wagner \[2013\]](#)).

### 1.1.1 MÉCANISMES IMPLIQUANT DES COUPURES DE L'ADN

Certains mécanismes cellulaires impliquent des remaniements de la séquence nucléotidique. Ces mécanismes, qui interviennent physiologiquement dans la vie cellulaire, emploient donc des enzymes capables de générer des coupures dans l'ADN, nécessaires à ces remaniements.

#### 1.1.1.1 LA MÉIOSE

Dans les cellules sexuelles, la méiose permet de créer, à partir d'une cellule mère, quatre cellules filles dont les patrimoines génétiques diffèrent. Cette diversité résulte de deux phénomènes, dont l'un, appelé *enjambement*, implique que des coupures aient lieu sur deux molécules d'ADN afin de permettre d'échanger certaines de leurs parties. Ces coupures sont générées dans les cellules germinales mâles par des enzymes spécifiques, telles que la protéine de recombinaison méiotique Spo11 [[Cao et al., 1990](#)]. Lors de la méiose, l'ADN est donc physiologiquement « endommagé » afin de permettre une diversification du patrimoine génétique.

« *Crossing over* »

#### 1.1.1.2 V(D)J ET RECOMBINAISON DE CONVERSION DE CLASSE

Dans le système immunitaire, les anticorps produits par les cellules B sont composés de deux parties : l'une d'entre elle est commune à tous (partie constante) et l'autre, responsable de la reconnaissance spécifique d'un antigène, varie d'une molécule à l'autre (partie variable). Les anticorps sont constitués d'une chaîne lourde et de deux chaînes légères, chacune étant composée d'une partie constante et d'une partie variable.

Les gènes qui codent les parties variables de ces chaînes comportent des répertoires de séquences variables, V, de diversité, D (spécifiques de la chaîne lourde) et de jonction, J. La recombinaison V(D)J permet de produire des millions d'anticorps différents à partir d'une même séquence génomique en générant des coupures par des enzymes spécialisées, Rag1 et Rag2, de la famille des transposases. Les extrémités des fragments V/D/J sont ensuite liguées de manière à combiner ces différents fragments entre eux, ce qui constitue la source de la diversité générée (pour une revue, lire [Jones](#)

et Gellert [2004]). Les récepteurs à l'antigène exprimés par les cellules T impliquent le même processus de diversification que les anticorps.

« *Class-switch recombination* »; CSR

Dans les cellules B, le mécanisme de recombinaison de conversion de classe permet de placer une région variable, produite par V(D)J sur différentes parties constantes qui déterminent l'isotype. Ce mécanisme a lieu, par exemple, lorsqu'un anticorps membranaire (IgM) reconnaît son antigène et qu'il est alors nécessaire de produire des anticorps libres (IgE, par exemple) comportant la même séquence variable afin de pouvoir reconnaître et éliminer spécifiquement l'antigène détecté. Le processus permettant la constitution d'une nouvelle classe d'anticorps dotée de la même spécificité de reconnaissance que ceux qui sont présents à la surface des cellules du système immunitaire est mis en œuvre par le biais de l'excision d'une partie de la séquence génomique (pour une revue, consulter [Geha et al. \[2003\]](#)). La production de coupures dans les séquences « signal » est suivie de la ligation des séquences appropriées.

Les cellules peuvent donc endommager leur propre ADN de manière transitoire et organisée (mais non totalement contrôlée). Ce procédé permet notamment d'introduire de la diversité génétique (et donc moléculaire) qui est ensuite mise à profit par les cellules.

### 1.1.2 FACTEURS ENDOGÈNES TOXIQUES POUR L'ADN

Dans les cellules, l'ADN est soumis à de nombreuses attaques par des agents endogènes. Parmi les lésions ainsi provoquées, nous avons distingué celles liées à la présence de composés réactifs dans les cellules, capables d'interagir avec les groupements chimiques des molécules d'ADN et celles provoquées par des dysfonctionnements de protéines qui interviennent dans la physiologie de l'ADN.

#### 1.1.2.1 LES PRODUITS SECONDAIRES DU MÉTABOLISME

Certains produits secondaires du métabolisme possèdent une réactivité suffisante pour induire des modifications chimiques sur les molécules d'ADN. L'exemple le plus frappant concerne les espèces réactives de l'oxygène qui sont produites, par exemple, tout au long de la chaîne respiratoire, dans les mitochondries [[Simsek et Jasin, 2011](#)]. Les espèces radicalaires, en particulier, sont extrêmement réactives et peuvent modifier la nature chimique des bases de l'ADN. Les lésions provoquées par ces composés ne sont pas individuellement graves, mais elles le deviennent lorsqu'elles sont présentes en quantité importante.

### 1.1.2.2 LES DYSFONCTIONNEMENTS ENZYMATIQUES

De nombreux enzymes sont impliqués dans la physiologie de l'ADN. Ces protéines, qui sont nécessaires à la vie des cellules, peuvent toutefois également provoquer l'apparition de lésions dans l'ADN lorsqu'elles ne fonctionnent pas correctement.

Les polymérase à ADN (DNA-polymérase) sont les enzymes responsables de la réplication de l'ADN (mécanisme nécessaire afin de transmettre l'information génétique d'une cellule mère à deux cellules filles). La synthèse de l'ADN est dite semi-conservative, car elle utilise la complémentarité des bases d'ADN afin de produire un nouveau brin d'ADN à partir d'un ancien (chaque nouveau duplexe d'ADN comporte donc le brin néosynthétisé hybridé à l'un des anciens). Bien que ces enzymes soient fidèles (moins d'une erreur tous les  $10^5$  nucléotides) et, de plus, dotés d'une activité de correction (3'-5' exonucléase), il peut arriver que des mésappariements soient observés après la réplication, ce qui perturbe la structure de la double-hélice (apparition de « *bulges* »). Certaines de ces DNA-polymérase sont utilisées uniquement après un stress génomique lors d'un processus particulier de réparation appelé *synthèse translésionnelle*. Or, ces polymérase sont bien moins fiables que leurs contreparties spécialisées dans la réplication et peuvent ainsi introduire une proportion non négligeable de mésappariements [Saugar *et al.*, 2014].

En anglais, « *translesion synthesis* »

Dans les noyaux, la double-hélice d'ADN subit un enroulement qui permet sa compaction (les molécules d'ADN ont une taille importante, alors que l'espace disponible dans le noyau est limité). Cet enroulement est en partie contrôlé par des enzymes, les topoisomérase, dont le mécanisme peut faire intervenir un complexe covalent entre un brin d'ADN clivé et la protéine [mécanisme de TOP1 (p 187), par exemple]. Ces intermédiaires covalents, toxiques pour l'ADN dans certaines conditions, peuvent être stabilisés par des inhibiteurs de topoisomérase.

### 1.1.3 FACTEURS EXOGÈNES TOXIQUES POUR L'ADN

En plus des différents facteurs endogènes pouvant endommager l'ADN, l'environnement constitue également une source d'exposition à des composés pouvant induire des lésions. Parmi ces derniers, on peut notamment citer la fumée de cigarette, la lumière ultraviolette (lumière solaire), les radiations ionisantes, ou encore des composés chimiques comme le plomb contenu dans certaines peintures. Ces composés sont donc nocifs pour les cellules et, à plus grande échelle, pour l'organisme tout entier. De manière intéressante, nous verrons également par la suite qu'il peut être utile d'induire ces dommages dans nos cellules afin d'en exploiter la dangerosité (voir § 1.5.1, page 30).

**Conclusion.** Dans les cellules, l'ADN est continuellement soumis à des agressions, aussi bien endogènes qu'exogènes. Les dommages provoqués entravent le bon fonctionnement du métabolisme de l'ADN et donc celui de la cellule. Afin de permettre aux cellules de survivre et de continuer à se diviser, l'évolution a sélectionné des mécanismes destinés à restaurer l'intégrité de l'ADN.

La diversité des dommages à l'ADN reflète la variété des facteurs qui peuvent les causer. La détection, la caractérisation et la quantification de ces dommages induits par l'un des agents mentionnés ci-dessus, par exemple, fait l'objet d'une discipline à part entière (pour une revue, voir [Cadet et Wagner \[2013\]](#)). Les cassures de la chaîne phosphodiester, sur l'un des brins (CSB, pour cassure simple-brin) ou les deux (CDB, pour cassure double-brin) sont des types de dommages particuliers qui nécessitent l'intervention de ligases à ADN. Ces cassures peuvent également constituer des intermédiaires pour la réparation d'autres dommages, plus variés et plus complexes.

## 1.2 RÉPARATION DE L'ADN

Différents mécanismes de réparation sont mis en œuvre par des acteurs cellulaires répondant spécifiquement à une catégorie particulière de dommages (Figure 1.2, page ci-contre). Ces mécanismes, ou « voies de réparation », se matérialisent par le recrutement dynamique et ordonné de protéines au niveau du dommage, qui forment des complexes permettant de réparer ce dommage selon un processus en plusieurs étapes. Certaines cassures, dites de « chimie complexe », peuvent cependant mettre en difficulté les systèmes de réparation de l'ADN en raison de la structure particulière de l'ADN au niveau des extrémités générées (seules les extrémités 5'phosphate et 3'OH, qui sont dites « canoniques », sont reconnues par les ligases).

D'excellentes revues discutent des différents mécanismes connus pour réparer les dommages à l'ADN [[Ciccia et Elledge, 2010](#); [Iyama et Wilson, 2013](#)].

### 1.2.1 LE BER (« *base-excision repair* »)

Certaines lésions de l'ADN interviennent au niveau de la base azotée modifiée et ne nécessitent donc que le remplacement de cette base, sans introduire de discontinuité dans le squelette phospho-carboné de la molécule. La réparation de lésions consistant en l'apparition, dans la molécule d'ADN, d'un site abasique, d'une base oxydée, ou plus généralement d'une base altérée peut être effectuée par le mécanisme de réparation par excision de base (BER, Figure 1.3, page 10).

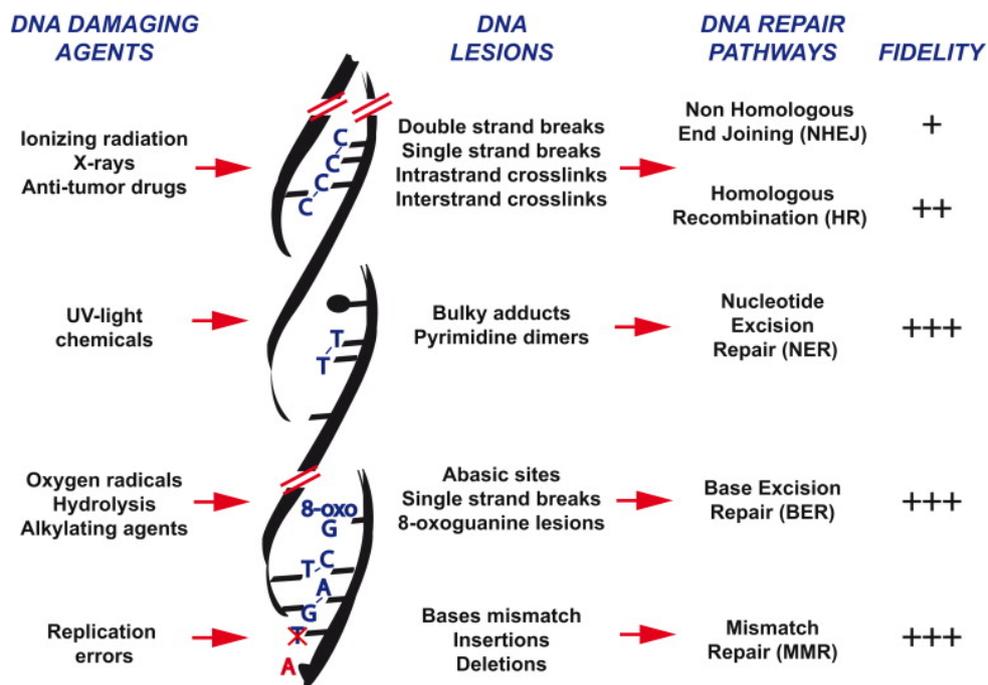


FIGURE 1.2 – Différentes voies de réparation en fonction des agents créant les dommages. Tiré de [Blanpain et al. \[2011\]](#).

Ce mécanisme, qui permet une réparation rapide des dommages, se déroule en cinq étapes :

1. la base altérée est excisée par une glycosylase, ce qui crée un site abasique ;
2. une endonucléase coupe le brin d'ADN au niveau du site abasique ;
3. les extrémités générées sont prises en charge, créant une structure reconnue par les DNA-polymérase et les ligases ;
4. le nucléotide complémentaire est inséré (processus de « *short patch* ») ;
5. la ligation par la DNA-ligase 3 restaure la continuité du brin d'ADN.

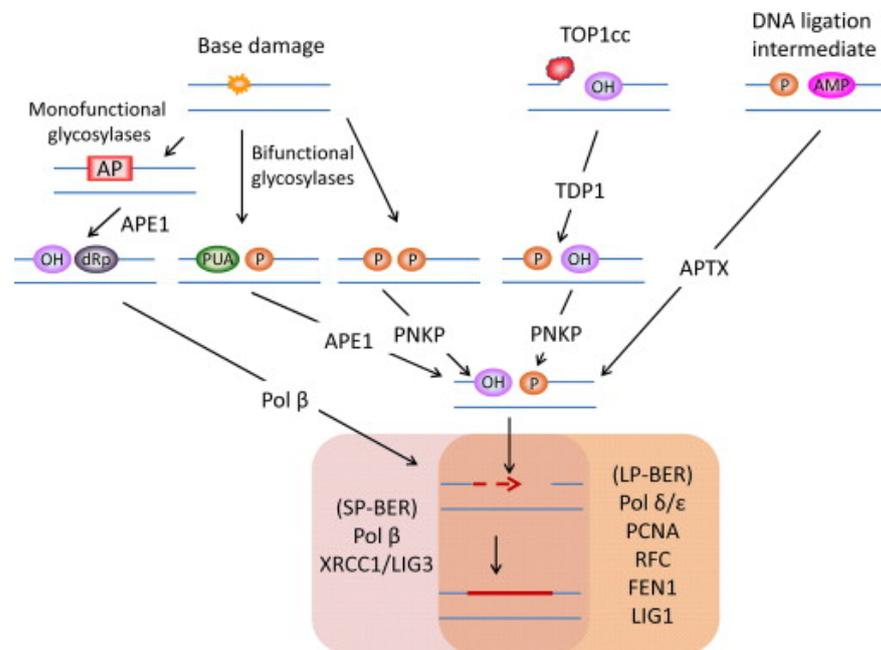


FIGURE 1.3 – Réparation des dommages à l'ADN par la voie de l'excision de base (BER). La voie BER de réparation comporte deux chemins, LP-BER (« *long patch* ») et SP-BER (« *short patch* »). Tiré de [Iyama et Wilson \[2013\]](#).

Les cellules possèdent plusieurs glycosylases à ADN (il en existe 11 différentes), qui sont capables d'exciser la base endommagée. L'élimination de la base produit un site abasique, une structure reconnue par l'endonucléase APE1 (« *apurinic endonucléase 1* »), qui effectue une coupure sur le brin endommagé en 5' de la lésion, ce qui constitue la seconde étape de ce mécanisme. Le site ainsi dévoilé est reconnu et pris en charge par la DNA-polymérase  $\beta$ , qui retire le nucléoside et le remplace par un nouveau nucléotide. Enfin, la continuité du brin d'ADN est restaurée par le complexe composé de la

DNLI3 (p 161) et de XRCC1 (p 189). Ce premier mécanisme constitue le processus « court » (« *short patch* ») de la réparation par excision de base par lequel la plus grande partie des dommages aux bases des nucléotides est réparée.

Il existe un processus « long » du BER (« *long patch BER* »), qui implique un remaniement plus important, avec synthèse d'un fragment d'ADN plus long, accompagné d'un déplacement de brin, qui fait appel d'autres enzymes spécifiques (DNA-polymérase  $\delta$  et  $\epsilon$ , assistées par le facteur de processivité PCNA, et également par *Fen1* et la DNA-ligase 1).

« *Proliferating cell nuclear antigen* »

« *Flap endonuclease 1* »

### 1.2.2 LE NER (« *nucleotide excision repair* »)

Le mécanisme du NER est déclenché par la détection d'une distorsion de la double-hélice d'ADN, due à la présence d'un dommage. Contrairement au BER, ce mécanisme met en œuvre l'élimination du fragment d'ADN endommagé. Deux chemins sont possibles, selon que le dommage est présent dans une région transcrite (« *transcription coupled NER* ») ou non (« *global genome NER* »). Ces deux voies ne diffèrent que par les protéines qui sont impliquées dans la reconnaissance des lésions (Figure 1.4, page suivante).

Abréviations en TC-NER et GG-NER

Dans le GG-NER, les protéines RPA, XPA et XPC reconnaissent les dommages. La protéine de réplication RPA (trimère de RFA1, RFA2 (p 178) et RFA3) ne reconnaît pas directement le dommage, mais se fixe sur l'ADN simple-brin qui en résulte, ce qui induit l'ouverture de la double hélice par l'activité hélicase de TFIIH. Dans le cas du TC-NER, la RNA-polymérase II, bloquée par le dommage au cours de la transcription, recrute directement les protéines de réparation.

Dans les deux cas, les nucléases XPG et XPF-ERCC1 réalisent ensuite une incision dans le brin qui provoque l'apparition d'extrémités 3' et 5'. La formation du complexe d'incision PIC3, constitué des protéines XPA, RPA, TFIIH, XPG et XPF-ERCC1, est dépendante de l'ATP. Les incisions réalisées de part et d'autre du dommage sont situées respectivement 6 nucléotides en 5' du dommage, pour la première et 20 nucléotides en 3', pour la seconde. Le complexe d'excision génère donc un oligonucléotide simple-brin d'une vingtaine d'oligonucléotides comportant le dommage, qui sera ensuite éliminé. Les protéines du complexe d'excision se détachent ensuite, à l'exception de RPA, puis la synthèse d'une nouvelle portion de brin est réalisée par les protéines RFC (protéine de réplication), RPA et par les DNA-polymérase  $\delta$  ou  $\epsilon$ . La restauration de la continuité du brin d'ADN est effectuée par l'action de la DNA-ligase 1 [Reardon et Sancar, 2005].

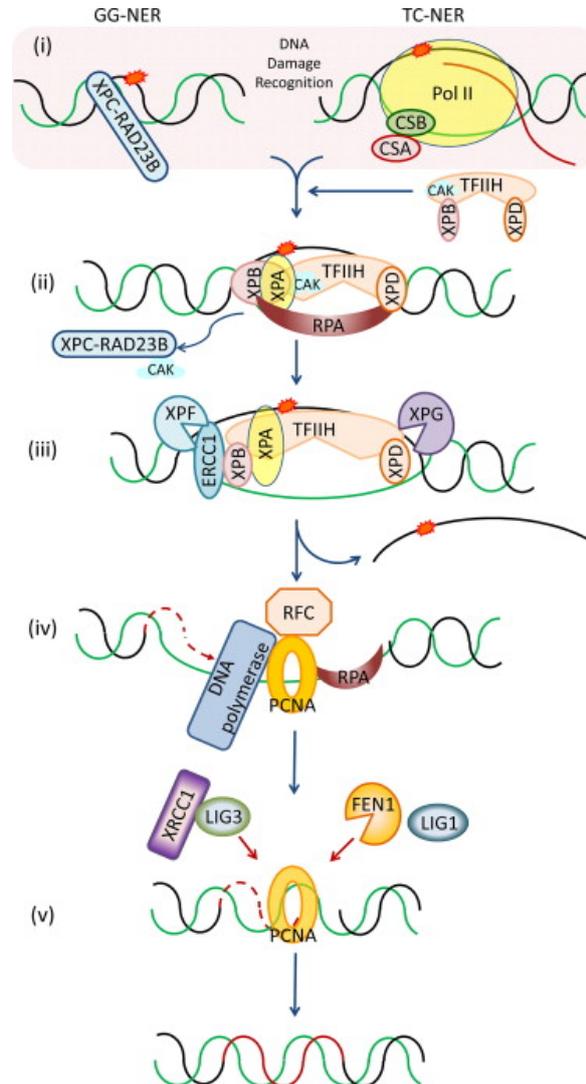


FIGURE 1.4 – **Différentes voies du NER.** La voie de réparation NER comporte deux chemins : le GG-NER (« *global genome* ») et le TC-NER (« *transcription coupled* »). Tiré de [Iyama et Wilson \[2013\]](#).

### 1.2.3 LE MMR (« *mismatch repair* »)

La réparation des mésappariements est initiée par la reconnaissance d'un défaut dans la structure de la double hélice d'ADN par le complexe MSH2–MSH6 (homologue de la protéine bactérienne MutS $\beta$ ) ou par la reconnaissance d'une boucle formée suite à la délétion ou à l'ajout de nucléotides sur l'un des deux brins par le complexe MSH2–MSH3 (homologue de la protéine bactérienne MutS $\alpha$ ).

Cette reconnaissance provoque le recrutement du complexe homologue de la protéine bactérienne MutL, formée par les protéines MLH1 et PMS1. Le complexe homologue de MutL a ensuite lui-même pour fonction de recruter les protéines qui vont permettre de prendre en charge le dommage : des nucléases (telles que la nucléase EXO1) pour exciser le fragment d'ADN endommagé, des DNA-polymérases (telles que la polymérase  $\beta$ ) pour synthétiser la nouvelle portion du brin remplaçant le fragment excisé et la DNA-ligase 1 afin de permettre de rétablir les liaisons phosphodiester entre les nucléotides de part et d'autre des coupures.

Il est intéressant de noter que le MMR est un mécanisme impliquant ses propres senseurs des dommages, mais qu'il fait intervenir des machineries moléculaires également impliquées dans les mécanismes d'excision de base et de nucléotide décrits ci-dessus.

#### 1.2.4 LA HR (« *homologous recombination* »)

Les cassures double brin sont parmi les cassures les plus dangereuses et ont la particularité de n'être pas *immédiatement* réparables par emploi d'une quelconque homologie, contrairement à nombre de mécanismes décrits plus haut. Afin de réaliser une réparation fidèle des dommages en utilisant la complémentarité des paires de bases, il est nécessaire, lorsque cela est possible, d'exploiter l'information contenue dans la chromatide sœur au moment de la réplication du chromosome. La voie de réparation par recombinaison homologue (HR) utilise la chromatide sœur pour réparer des cassures double-brin en phases S et G2 du cycle cellulaire (Figure 1.5, page suivante).

La recombinaison homologue est initiée par la génération d'un ADN simple-brin 3' sortant à partir du site de la cassure double-brin. Cette étape, dite de *résection* est initiée par le complexe MRN (constitué des protéines Mre11, Rad50 et Nbs1) ainsi que l'endonucléase CtIP (cette dernière protéine est spécifique de la phase S) et pourrait être en partie contrôlée par BRCA1 [Germani *et al.*, 2003]. La résection est ensuite amplifiée par l'activité exonucléase de 5' vers 3' de Exo1 afin d'augmenter la taille de la région simple-brin. RPA recouvre le fragment simple-brin ainsi exposé et le protège de l'action des nucléases, puis elle en est chassée par la protéine Rad51 qui initie de manière irréversible la réparation par recombinaison homologue et provoque l'invasion de brin [Krejci *et al.*, 2012]. Ce brin d'ADN, recouvert par Rad51 est alors placé en vis-à-vis de sa séquence homologue située sur la chromatide sœur du chromosome. L'action des protéines BRCA2 et Rad52 est nécessaire pour déplacer RPA et permettre la recherche de la région homologue et la formation d'une boucle permettant au brin d'ADN recouvert par Rad51 de s'y fixer. Les extrémités 3' des coupures sont alors allongées par des DNA-polymérases en se servant des deux brins de la chromatide sœur comme matrice. Les jonctions de Holliday formées par le croisement des

Terme emprunté à la chirurgie

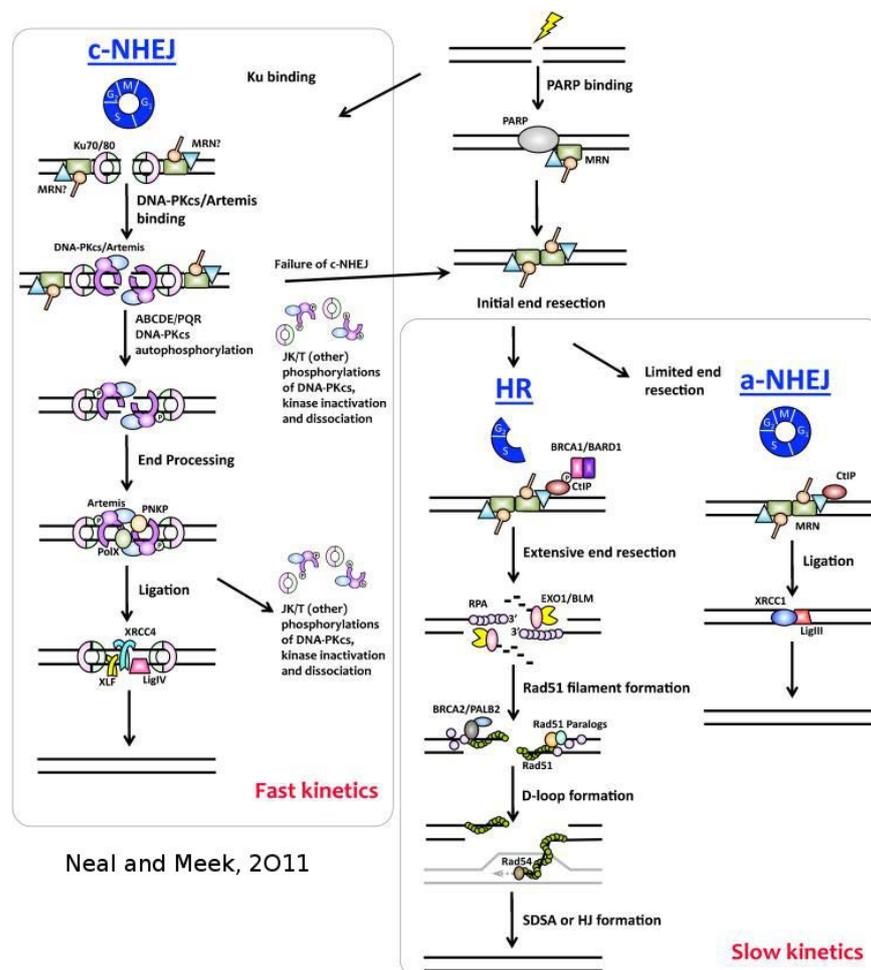


FIGURE 1.5 – Voies de réparation des cassures double-brin de l'ADN. Les cassures double-brin de l'ADN sont prises en charge par trois voies de réparation, la voie du NHEJ canonique (c-NHEJ), celle du NHEJ alternatif (a-NHEJ) et enfin celle de la recombinaison homologue (HR). Le contexte cellulaire et la nature plus spécifique du dommage peuvent conduire la cellule à choisir une voie plutôt qu'une autre. PNKP et Artemis sont des enzymes de rectification (« processing ») des extrémités de chimie complexe. Tiré de Neal et Meek [2011].

brins sont résolues par des nucléases spécifiques, appelées résolvas, ainsi que par des hélicases (appartenant à la famille RecQ). Enfin, la DNA-ligase 1 restaure la continuité des deux brins d'ADN.

### 1.2.5 NHEJ « *non homologous end joining* »

Il n'est pas toujours possible d'utiliser une séquence complémentaire de la séquence endommagée pour réparer fidèlement les cassures double-brin de l'ADN. C'est notamment le cas pendant les phases du cycle cellulaire durant lesquelles les chromosomes ne sont pas encore dupliqués (phases M, G1 et début de la phase S). Afin de réparer les cassures double-brin hors phase S/G2, les cellules font intervenir un processus de jonction des extrémités, et cela indépendamment de la séquence (NHEJ, Figure 1.5, page ci-contre) Lieber [2010]; Mahaney *et al.* [2009]. Le mécanisme NHEJ est potentiellement actif à chaque étape du cycle cellulaire et il constitue donc la voie de réparation par défaut des cassures double-brin dans l'ADN, et la seule voie possible pour les cellules qui ne se divisent pas [Iyama et Wilson, 2013]. Ce mécanisme implique toujours une étape ultime : la ligation des extrémités libérées par la cassure double-brin. Comme aucune homologie de séquence ne vient fiabiliser le processus d'aboutement, le NHEJ peut occasionner des erreurs par ligation d'extrémités appartenant, par exemple, à deux chromosomes différents. La fiabilité de ce mécanisme, longtemps remise en doute, semble cependant acceptable lorsque la quantité de dommages n'est pas trop importante [Betermier *et al.*, 2014].

La voie NHEJ de réparation des cassures double-brin est importante car les cassures double-brin sont les plus dangereuses pour les cellules : elles provoquent une rupture de la continuité chromosomique, des translocations et des délétions qui induisent à terme l'apparition de cancers. Afin de permettre la réparation des lésions, le cycle cellulaire est interrompu dès que la cellule en détecte un nombre suffisant, ce qui permet d'exclure du site du dommage toute machinerie qui ne soit pas utile à la réparation, comme celles de la réplication ou de la transcription, par exemple. Lorsque la cellule détecte que le nombre de dommages est tellement important qu'elle ne sera pas à même de tout réparer avant la reprise du cycle cellulaire, elle peut choisir la voie de l'apoptose, la mort cellulaire programmée. La signalisation moléculaire, qui consiste à informer la cellule dans son ensemble dès qu'une lésion de l'ADN a été perçue, est donc un processus complexe qui rapproche des protéines de différentes voies en des complexes non covalents transitoires. On peut citer ici les protéines de signalisation à proprement parler : d'abord les protéine-kinases, ensuite les protéines de « *checkpoint* », mais aussi les protéines qui régulent le niveau de compaction de la chromatine, par exemple. Un certain nombre des protéines citées ci-dessus ont été identifiées par spectrométrie de masse au cours de nos expériences et font l'objet de fiches individuelles descriptives en annexe (voir annexe « [Fiches relatives aux protéines d'intérêt](#) », page 153).

L'hétérodimère KU (constitué des sous-unités Ku70 et de Ku80, connues également sous les noms de XRCC6 et de XRCC5) possède une affinité très

importante pour les extrémités de l'ADN (avec un  $K_d$  de l'ordre de  $10^{-10}$  M) autour desquelles il forme un anneau [Walker *et al.*, 2001]. Ce complexe, abondant dans les noyaux (il a été estimé à  $10^5$  le nombre de molécules de ce dimère dans une cellule), est le premier de la voie du NHEJ à se fixer au niveau des extrémités de la cassure double-brin. KU est le domaine régulateur de la sous-unité catalytique de la protéine kinase ADN-dépendante (DNA-PKcs) (voir PRKDC, page 177). Lorsque le complexe ternaire est formé, l'enzyme est sous sa forme holoenzyme. PRKDC appartient à la même famille que les protéines ATM et ATR (phosphatidylinositol 3-kinases), elles-mêmes impliquées dans la signalisation des dommages de l'ADN (voir paragraphe 1.4, page 26) [Falck *et al.*, 2005]. Elle est capable de se fixer seule sur les extrémités libres de l'ADN mais son interaction avec KU augmente son affinité pour l'ADN [Uematsu *et al.*, 2007].

La fixation de PRKDC sur le dimère KU provoque le glissement ATP-dépendant, sur une dizaine de paires de bases, de ce dernier le long de la séquence nucléotidique, ce qui permet de libérer l'extrémité d'ADN qu'il protégeait. Ces extrémités peuvent alors être prises en charge par des enzymes de *traitement des extrémités*. On peut distinguer deux types de protéines impliquées dans le traitement des extrémités : les nucléases, qui permettent de générer une nouvelle extrémité, canonique, et les enzymes qui permettent la conversion des groupements chimiques, telles que PNKP par exemple, qui permet de régénérer une extrémité 5' phosphate à partir d'une extrémité 5' OH et/ou de générer une extrémité 3' OH à partir d'une extrémité 3' phosphate. Les régions simple-brin qui peuvent être mises en évidence, s'il y a un début de résection lors du NHEJ, sont prises en charge par les DNA-polymérase  $\lambda$  et  $\mu$ , après appariement — simple ou partiel — avec le brin complémentaire. Lorsque la nature des extrémités est compatible avec une ligation, PRKDC effectue la *synapsis* des deux extrémités à liguer, et un complexe de ligation est recruté pour réparer la cassure. Ce complexe est constitué de la DNA-ligase 4 [DNLI4 (p 161)], de XRCC4 (p 189) et de XLF (p 193) (pour cette dernière, voir Ahnesorg *et al.* [2006]; Buck *et al.* [2006]). Le rôle de XLF dans ce complexe serait d'augmenter la processivité de DNLI4, alors que le rôle de XRCC4 est de stabiliser la ligase 4 sur le complexe déjà assemblé au niveau des extrémités à liguer.

« End processing »

Terme grec signifiant *conjonction*, de *syn*, « ensemble » et de *haptain* « fixer »

### 1.2.6 NHEJ « ALTERNATIF »

Le processus de réparation selon la voie NHEJ décrite ci-dessus a longtemps été considéré comme l'unique voie possible pour la réparation des cassures double-brin qui ne soit pas la recombinaison homologe. Une voie, dite « alternative », a cependant été caractérisée plus récemment (Figure 1.5, page 14). En effet, des études ont permis d'observer, dans des cellules déficientes en PRKDC, que la réparation des cassures double-brin pouvait

toujours avoir lieu en dehors des phases G2 et S du cycle cellulaire, mais de façon très ralentie [DiBiase *et al.*, 2000]. Des études plus approfondies ont confirmé que la réparation observée dans ces conditions était indépendante des protéines de la recombinaison homologue et ont permis de mettre en évidence l'existence de cette voie alternative au NHEJ « canonique », qui aurait lieu de manière indépendante de KU et de la DNA-ligase 4 (pour une revue voir Rass *et al.* [2012]).

Les protéines de cette voie de réparation sont les protéines PARP1 (p 175), XRCC1 (p 189) et la DNA-ligase 3 (DNLI3, p 161), qui sont par ailleurs impliquées dans la réparation des cassures simple-brin de l'ADN par la voie du BER (voir plus haut). Il est également possible que l'activité de ligation soit portée par la DNA-ligase 1. Ce mécanisme de réparation peut également faire intervenir des enzymes impliqués dans le traitement des extrémités (Figure 1.5, page 14), lorsque la structure de ces dernières n'est pas compatible avec une ligation directe (Figure 1.7, page 22). La voie de réparation alternative du NHEJ est également parfois appelée MMEJ (« *micro-homology mediated end-joining* ») en raison de l'utilisation d'une homologie inférieure à une vingtaine de paires de bases afin de réaliser la synapsis des deux extrémités à liguer. Cette homologie est révélée par l'activité exonucléase du complexe MRN, en association avec CtIP et BRCA1 (qui sont des protéines de la recombinaison homologue ; voir plus haut) [Rass *et al.*, 2012; Zhang *et Jasin*, 2011].

Les deux voies du NHEJ coexistent dans les cellules eucaryotes. Les protéines impliquées dans la reconnaissance des dommages, pour chacune de ces deux voies (KU pour la voie canonique et PARP1 pour la voie alternative) sont en compétition.

### 1.2.7 CONCLUSIONS.

L'apparition de cassures double-brin de l'ADN est un événement qui nécessite une prise en charge rapide et efficace des dommages. Nous avons décrit trois voies spécialisées dans leur réparation. Cette spécialisation est fondée sur des mécanismes de réparation différents qui peuvent, au besoin, tirer parti du contexte cellulaire (présence de chromatides sœur dans le noyau, pour la recombinaison homologue, par exemple). Comme ces voies peuvent éventuellement être opérantes durant plusieurs phases du cycle, des mécanismes de régulation existent, qui favorisent l'une d'entre elles pour permettre la réparation de la cassure de la façon la plus satisfaisante possible en termes de rapidité et de fidélité. Collectivement, ces concepts sont traités dans la littérature comme relevant de la problématique du *choix de la voie de réparation* des cassures double-brin de l'ADN.

## 1.3 DÉTERMINISMES DU CHOIX DE LA VOIE DE RÉPARATION DES CASSURES DOUBLE-BRIN DE L'ADN

### 1.3.1 RÉGULATION DES DIFFÉRENTES VOIES DE RÉPARATION

Nous pouvons observer que, contrairement aux mécanismes qui permettent de réparer les lésions intervenant sur l'un des deux brins de l'ADN (BER, NER et MMR), les voies de réparation des cassures double-brin de l'ADN (HR et NHEJ) reconnaissent un même type de dommages. La sélection de l'un ou l'autre de ces mécanismes ne dépend donc pas uniquement de la présence d'une extrémité d'ADN. Dans les paragraphes suivants nous proposons un bref récapitulatif des principaux facteurs pouvant influencer le choix d'une voie de réparation particulière.

#### 1.3.1.1 LE NIVEAU DE DIFFÉRENCIATION CELLULAIRE

Les cellules somatiques et les cellules germinales, les cellules souches et les cellules tumorales peuvent mettre en jeu différents mécanismes de réparation. Les différences qui existent entre ces différents types de cellules peuvent d'ailleurs être exploitées avec profit pour sensibiliser préférentiellement les cellules tumorales par rapport aux cellules saines (voir § 1.5.1, page 30). Chacun des mécanismes mentionnés fait intervenir des enzymes spécifiques de sorte que la « boîte à outils » pour la réparation n'est pas la même selon le type cellulaire. Les cellules du système immunitaire, par exemple, expriment des enzymes supplémentaires qui leur permettent, comme décrit plus haut, d'introduire des cassures double-brin dont la finalité leur est bien spécifique. C'est le cas, en particulier des terminal transférases, détaillées à la § 1.3.3, page 20 et à la Figure 1.7, page 22, panneau C.

#### 1.3.1.2 LA PHASE DU CYCLE CELLULAIRE

Le cycle cellulaire joue un rôle absolument prépondérant sur la manière de réparer l'ADN. Si l'on prend le cas particulier de résection des extrémités de l'ADN libérées par la cassure, par exemple, les différentes phases du cycle exercent un contrôle sur les interactions protéine–protéine et protéine–ADN nécessaires à cette résection [Ferretti *et al.*, 2013; Huertas, 2010; Mladenov *et al.*, 2013]. BRCA1 favorise la résection de 5' en 3', conduisant à la production d'extrémités 3'-débordantes. Mais cette résection ne peut avoir lieu que si BRCA1 interagit avec la nucléase CtIP, sous le contrôle de kinases liées au cycle cellulaire (CDK2). Tout ceci dans le cadre de la recombinaison homologue, c'est-à-dire quand la cellule est dans une phase du cycle cellulaire pour laquelle la réplication a produit la chromatide sœur du chromosome sur lequel se trouve le dommage.

### 1.3.1.3 LA NATURE CHIMIQUE DES EXTRÉMITÉS

Toutes les cassures double-brin de l'ADN ne produisent pas nécessairement des extrémités immédiatement ligables (5'P et 3'OH, dites canoniques). Les extrémités peuvent être, en effet, de chimie complexe, comportant des groupements chimiques qui ne peuvent facilement être pris en charge par les systèmes de réparation. Plus les extrémités libérées par une cassure sont proches des extrémités canoniques, plus la réparation par NHEJ sera favorisée. Au contraire, un degré de complexité croissant augmentera la probabilité d'une réparation par recombinaison homologue.

Si l'on s'intéresse au cas d'extrémités immédiatement ligables, deux cas peuvent se présenter :

- \* Les extrémités sont telles qu'une hybridation est possible car les deux brins ont des bouts débordants compatibles et de même taille. Dans ce cas — le plus favorable qui soit — l'étape ultime de réparation par le NHEJ ne nécessitera, en théorie, que le complexe de ligation composé de XRCC4, de la DNA-ligase 4 et éventuellement de XLF (p 193) ;
- \* Les extrémités sont à bouts francs. Ce cas est aussi assez facile et peut être pris en charge par le NHEJ, quoique avec une cinétique plus lente que dans le cas ci-dessus. En effet, la synapsis mise en œuvre par PRKDC est moins stable, puisqu'il n'y a pas d'hybridation possible ;
- \* Les extrémités sont non compatibles, car avec des bouts débordants complémentaires mais pas de même taille ou pas complémentaires du tout. Ce cas particulièrement complexe ralentit encore la ligation car d'autres activités enzymatiques de « rectification » doivent intervenir avant l'étape ultime de ligation (voir § 1.2.5, page 15).

Il faut noter que, si la réparation (en particulier dans le dernier cas ci-dessus) est trop lente, la résection des extrémités peut être amorcée (voir discussion sur les sites multiples de dommages dans la § 1.5.1, page 30). Dans le cas où la résection est entamée, alors elle est suivie de la recherche d'homologies de séquence, puis par l'appariement des séquences simple-brin complémentaires issues de la résection, situées à proximité de la cassure. Noter enfin que, si la ligation des extrémités protégées par KU [XRCC5 et XRCC6 (p 190)] n'est pas réalisée rapidement par le NHEJ en phase G1, alors la phase G2 ouvre d'autres possibilités de réparation, par la HR ou le NHEJ alternatif.

### 1.3.1.4 LE CONTEXTE CHROMATINIEN DE LA LÉSION

La position de la lésion dans l'hétérochromatine (condensée : inactive) ou l'euchromatine (décondensée : active) conditionne la cinétique de la

réparation. En effet, après induction de cassures double-brin dans l'ADN des cellules, une cinétique biphasique est observée, avec la réparation de plus de 80 % des lésions radioinduites lors d'une première phase de cinétique rapide. Ces lésions rapidement réparées sont typiquement, soit de chimie des extrémités simple, soit situées dans l'euchromatine. Lors d'une deuxième phase — de cinétique lente — ce sont les extrémités de chimie complexes et/ou celles situées dans l'hétérochromatine qui sont réparées [Shibata *et al.*, 2011].

### 1.3.2 RÔLE DES TROIS LIGASES HUMAINES DANS LES MÉCANISMES DE JONCTION

Les trois ligases humaines, DNA-ligase 1, 3 et 4 (voir DNLI3 et DNLI4, page 161), sont toutes capables de liguer des cassures simple-brin et double-brin de l'ADN, mais ont des souplesses fonctionnelles différentes : DNLI4, spécifiquement dédiée au NHEJ, apparaît ainsi comme la plus flexible des ligases, en particulier pour réparer des extrémités non-directement ligables et nécessitant l'intervention d'une DNA-polymérase. DNLI3 intervient dans la ligation d'extrémités non-homologues par le NHEJ alternatif. Les ligases 1 et 3 sont également impliquées dans d'autres mécanismes que la réparation des cassures double-brin : DNLI1 est régulée selon le cycle cellulaire par phosphorylation et peut liguer les fragments d'Okazaki mis en jeu lors de la réplication, mécanisme auquel elle est plus particulièrement dédiée [Levin *et al.*, 2000]. DNLI1 est également la ligase préférentiellement utilisée lors de la réparation par recombinaison homologue. DNLI1 et DNLI3 sont toutes deux impliquées dans la réparation des cassures simple-brin (voies NER et « *long patch* » BER, pour la DNA-ligase 1 [Levin *et al.*, 2000] ; « *short patch* » BER et SSBR, pour DNLI3 [Petermann *et al.*, 2006]). DNLI3 intervient dans le processus de la méiose, étant très fortement exprimée sous forme d'une isoforme spécifique dans les cellules germinales mâles. Une autre de ses isoformes est également active dans les mitochondries. Ces trois ligases peuvent se compenser dans une certaine mesure mais ne sont, dans certains cas, pas interchangeable. Par ailleurs, les rôles spécifiques de la DNA-ligase 1 et de la DNA-ligase 3 dans les différentes voies de réparation ne sont toujours pas complètement éclaircis.

### 1.3.3 SIGNATURES MOLÉCULAIRES DES VOIES DE RÉPARATION

Il est actuellement possible d'utiliser des modèles cellulaires dans lesquels on peut induire des coupures ciblées dans le génome grâce à la mise en place d'outils appropriés [Boubakour-Azzouz *et al.*, 2012; Ghezraoui *et al.*, 2014; Guirouilh-Barbat *et al.*, 2007] pour analyser la manière dont elles sont réparées (par séquençage des jonctions au voisinage du site initial

	Nick*		DSB Ligation*		Activity
	sensing	ligation	for compatible ends	for incompatible ends	Repair pathways
<p><b>Ligase I</b> - 919 aa</p>	∅	✓	✓	∅	- Long patch BER - NER - HR - (Replication)
<p><b>Ligase IIIα (nuclear)</b> - 922 aa</p>	✓ <sup>1</sup>	✓	✓ <sup>1</sup>	∅	-SSBR -Short patch BER - NER
<p><b>Ligase IV</b> - 911 aa</p>	/	✓ <sup>2</sup>	✓ <sup>2</sup>	✓ <sup>2</sup>	- NHEJ

FIGURE 1.6 – **Spécificité des trois DNA-ligases humaines.** Partie gauche : les trois ligases humaines ont en commun leur domaine de liaison à l'ADN, représenté en bleu, et un domaine catalytique, représenté en vert, dans lequel la lysine responsable de l'activité apparaît en rouge. Les domaines N- et C-terminaux qui sont spécifiques d'une ligase donnée sont également présentés : PIP box indique un motif de fixation pour PCNA, ZnF indique la présence d'un domaine à « doigt de zinc » et BRCT indique la présence du domaine C-terminal de la protéine de BRCA1. Les principaux partenaires d'interaction sont représentés en jaune. Partie droite : ce tableau résume les caractéristiques des trois DNA-ligases humaines et présente (dans la colonne de droite) les divers mécanismes dans lesquels elles sont impliquées. Tiré de Erika Brunet, HDR 2014.

de coupure). Les conséquences de la réparation, soit directement visibles au niveau des cellules, soit détectables par le séquençage au voisinage de la coupure, renseignent sur le(s) mécanisme(s) de réparation impliqué(s). Plusieurs cas sont possibles :

- \* La réparation restitue une séquence intacte : cette réparation a pu indistinctement être prise en charge par la HR ou le NHEJ canonique ;
- \* La réparation occasionne des délétions/additions de nucléotides : ceci constitue une signature du NHEJ. Les délétions peuvent impliquer des homologies de séquence au voisinage du dommage (SDSA : « *synthesis-dependent strand annealing* » ; [Thompson, 2012]). Plusieurs mécanismes peuvent être distingués selon la taille de la délétion, la taille de la microhomologie ou le positionnement de cette dernière par rapport au site de coupure. On parle de MMEJ lorsque la taille de la microhomologie est inférieure à 20 paires de bases (20 pb), seuil au-delà duquel on parlera de recombinaison homologue. Cette dernière implique, en

effet, des régions d'homologies beaucoup plus longues, de l'ordre de plusieurs centaines de pb. Les ajouts de séquence, au contraire des délétions, peuvent être dus à des enzymes tels que la terminal transférase (TdT), qui ajoute des nucléotides aléatoires en 3' d'une séquence [Thompson, 2012].

Le décryptage des mécanismes associés à un profil de réparation nécessite de nombreuses expériences de mutagenèse ou d'extinction de l'expression de protéines intervenant dans la réparation des cassures double-brin.

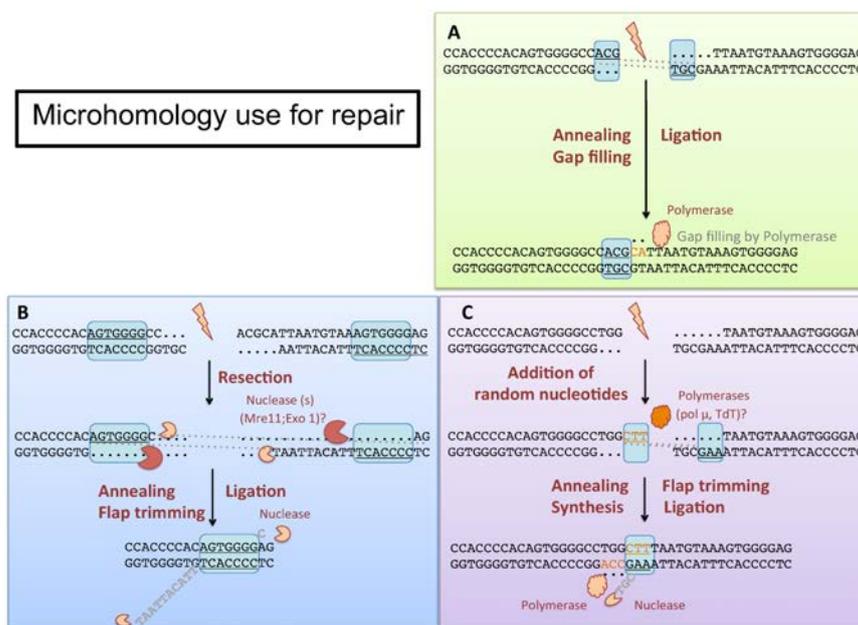


FIGURE 1.7 – **Utilisation des microhomologies pour la réparation des cassures double-brin.** (A) Si des microhomologies sont déjà disponibles sur les brins à rabouter, une hybridation peut se mettre en place immédiatement pour stabiliser les extrémités. Une DNA-polymérase peut combler les séquences non appariées (trous). La ligation peut alors avoir lieu. (B) Quand les microhomologies sont masquées dans une double-hélice, il faut les démasquer par résection du brin complémentaire (action de Exo1 ou Mre11). (C) Quand les séquences de microhomologie sont absentes, des DNA-polymérases particulières (terminal transférase ou polymérase  $\mu$ ) peuvent ajouter en 3' des nucléotides aléatoires supplémentaires qui peuvent créer une nouvelle homologie. À ce stade, l'hybridation peut se faire, on a élimination de l'ADN non hybridé par la Fen1 et la ligation peut se produire. Cette partie de la figure ne correspond pas à une réparation par la voie alternative du NHEJ. Tiré de Erika Brunet, HDR 2014.

### 1.3.4 CHOIX DE LA VOIE DE RÉPARATION

#### 1.3.4.1 ÉQUILIBRE ENTRE HR ET NHEJ

L'équilibre entre HR et NHEJ dépendrait essentiellement de deux protéines : BRCA1, pour la recombinaison homologue et 53BP1 pour la jonction des extrémités non-homologues [Bunting *et al.*, 2010; Cao *et al.*, 2009; Chapman *et al.*, 2012; Thompson, 2012]. Ces protéines ne semblent pas engagées de façon inambiguë dans l'une ou l'autre de ces voies, puisque BRCA1 peut également promouvoir le NHEJ [Thompson, 2012], notamment par interaction avec XRCC5 (Ku80) en phase G1. Le complexe MDC1 (p 172)- $\gamma$ H2AX est un intermédiaire commun entre les voies HR et NHEJ et peut aussi bien stimuler l'une ou l'autre de ces voies, en interagissant avec RAD51 pour promouvoir la première, et probablement avec PRKDC (p 177) pour promouvoir la seconde, dont on sait, par ailleurs, qu'elle inhibe la recombinaison homologue en fonction des groupes de sites sur lesquels elle est phosphorylée [Neal *et al.*, 2011].

#### 1.3.4.2 ÉQUILIBRE ENTRE LES VOIES DU NHEJ CANONIQUE ET ALTERNATIF

En phase G1, la seule voie de réparation disponible est le NHEJ. Si la lésion n'est pas réparée en G1, elle sera sollicitée par plusieurs mécanismes opérant en parallèle en phase G2/M où (potentiellement) les trois mécanismes de réparation des CDB, décrits ci-dessus, sont opérationnels. Le choix de la voie pourrait ainsi résulter d'une compétition entre quelques protéines clef, telles que la nucléase Mre11 du complexe MRN et KU (et/ou PRKDC), qui protégeraient, par ailleurs, les extrémités de l'ADN des nucléases [Shao *et al.*, 2012; Sun *et al.*, 2012]. La mise en évidence de mutants de sites de phosphorylation rendant PRKDC incapable de se dissocier des extrémités, et inhibant la résection de ces dernières, tend à montrer qu'une hiérarchie opère en donnant la priorité aux complexes du NHEJ avant l'interaction avec les complexes impliqués dans la résection des extrémités. Si les extrémités sont directement ligables, elles seront réparées rapidement par le NHEJ ; si la complexité des extrémités ralentit le processus (besoin d'étapes supplémentaires de rectification par des enzymes spécifiques, la fenêtre d'opportunité pour la résection des extrémités s'agrandit, ouvrant la voie aux mécanismes faisant intervenir des appariements de séquences alt-NHEJ (résection sur de courtes séquences) ou la recombinaison homologue (résection sur de longues séquences qui rend le processus irréversible). Ce scénario est compatible avec plusieurs études mettant en jeu différents types de dommages [Shibata *et al.*, 2011]. MDC1 interagit avec  $\gamma$ H2AX et RAD51 et influence la stabilité de RAD51 (interaction contrôlée par phosphorylation). 53BP1 promeut la voie du NHEJ en stimulant directement la ligation par DNLI4 d'une part, et en inhibant la HR d'autre part. Dans des cellules déficientes en BRCA1, 53BP1

MRN est responsable de l'initiation de l'étape de résection avant la HR

inhibe également la résection des extrémités. Cette conclusion résulte de l'observation que l'extinction de l'expression de BRCA1 n'est plus létale lorsque l'expression de 53BP1 est également éteinte, car le frein de la recombinaison homologue est levé [Cao *et al.*, 2009; Tomimatsu *et al.*, 2012]. 53BP1 bloque aussi l'étape d'activation de CtIP par BRCA1, due à son ubiquitinylation [Yu *et al.*, 2006].

#### 1.3.4.3 HIÉRARCHIE DES MÉCANISMES MIS EN JEU

En général, l'utilisation de la voie alternative du NHEJ se révèle lorsque le NHEJ canonique est défectueux, cette hiérarchie des voies pouvant tout simplement s'expliquer par des facteurs cinétiques [Schipler et Iliakis, 2013]. Toutefois, les deux mécanismes peuvent fonctionner en parallèle. C'est le cas, par exemple, dans les cellules immunitaires où la voie alternative du NHEJ peut aussi être impliquée dans les recombinaisons V(D)J et de conversion de classe [Simsek et Jasin, 2010; Weinstock *et al.*, 2007].

#### 1.3.5 QUELS ENJEUX DU CHOIX DE LA VOIE DE RÉPARATION ?

La régulation des différentes voies de réparation fait l'objet d'un contrôle très strict dont l'efficacité conditionne la stabilité génétique d'une cellule tout au long des divisions successives. Le choix de la voie de réparation, qui résulte de cette régulation, conditionne la stabilité génétique et le destin cellulaire, car tous les mécanismes de réparation n'ont pas le même degré de fiabilité. Autant, et nous l'avons vu plus haut, la cellule veut pouvoir réparer avec la plus grande fiabilité, autant, parfois, elle souhaite introduire efficacement des « erreurs » au voisinage d'une coupure (V(D)J).

La recombinaison homologue et le NHEJ canonique sont plus fiables que la réparation par MMEJ ou par le NHEJ alternatif. La nature intrinsèque des processus de réparation fondés sur une homologie partielle ou pas d'homologie du tout implique des risques de mutagenèse. Il faut noter que la HR et le alt-NHEJ nécessitent la résection des extrémités et dépendent donc tous deux du cycle cellulaire, avec une activité maximale en phase S/G2. Signalons que, dans le cas d'extrémités de chimie complexe, il a été montré que la réparation de tels dommages a impliqué l'étape de résection par la protéine CtIP qui persiste en phase G1, pour *in fine* une prise en charge du dommage par le NHEJ alternatif [Yajima *et al.*, 2013].

Une compréhension approfondie des (dé)-régulations intervenant dans les cellules tumorales pourrait déboucher sur des cibles potentielles pour augmenter spécifiquement l'apoptose dans les cellules cancéreuses. On pourrait imaginer, en effet, tirer partie de ces (dé)-régulations pour cibler des enzymes clef dont le comportement est spécifique des cellules tumorales, ce

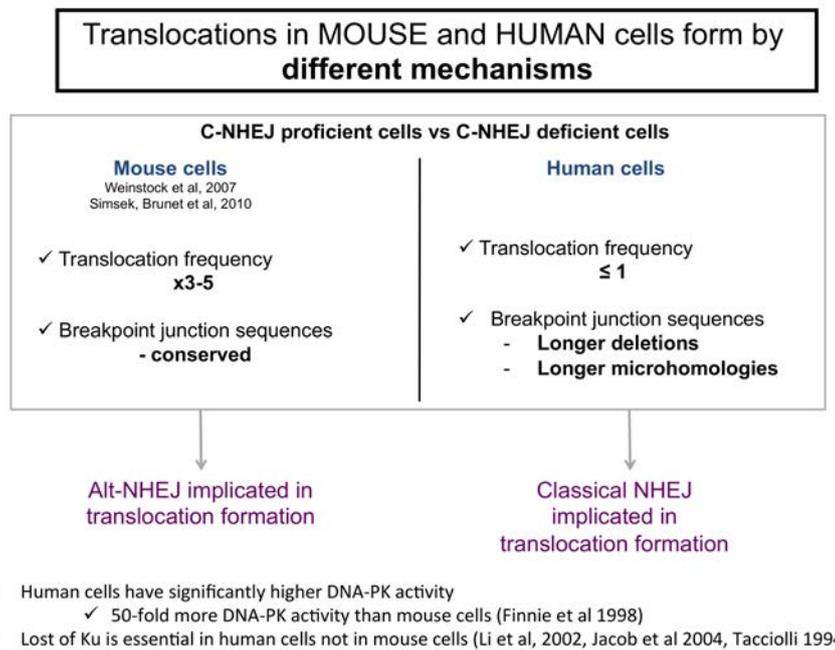


FIGURE 1.8 – **Translocations induites par apparition de cassures double-brin.** La réparation de cassures double-brin de l'ADN peut provoquer si elle est mal effectuée des translocations, qui joignent des portions de chromosomes différents. De manière intéressante, les mécanismes par lesquels ces translocations se produisent diffèrent chez l'homme et chez la souris. En effet, chez la souris il s'agit du NHEJ alternatif, alors que chez l'homme le mécanisme est principalement le NHEJ canonique [Finnie *et al.*, 1995; Li *et al.*, 2002; Taccioli *et al.*, 1994]. Tiré de Erika Brunet, HDR 2014.

qui permettrait d'avoir des traitements adjuvants aux autres protocoles thérapeutiques.

La jonction d'extrémités non-homologues peut, par sa nature, aboutir à associer deux fragments de chromosomes différents s'ils ont subi des dommages du type CDB. Ces événements, appelés translocations peuvent être responsables de l'oncogénèse. Alors que le NHEJ canonique n'est pas impliqué dans ce mécanisme chez la souris, où il jouerait un rôle de protecteur de l'intégrité du génome, il y serait, en revanche, impliqué chez l'homme [Ghezraoui *et al.*, 2014; Piganeau *et al.*, 2013] (voir Figure 1.8). Ceci a été démontré par des mesures de fréquences de translocation, induites artificiellement dans des cellules murines ou humaines dans lesquelles l'expression de protéines a été éteinte soit de manière stable (DNLI3 ou DNLI4, page 161) soit de manière transitoire (DNA-ligase 1). Les protéines dont l'expression aura été éteinte sont complétées ou non par des équivalents mutés dans certains

de leurs domaines. Les résultats montrent que, chez la souris l'inactivation du NHEJ canonique démasque la voie alternative du NHEJ et conduit à une augmentation des translocations. Dans ce cas, cela serait la DNA-ligase 3 qui serait impliquée dans le mécanisme de translocation et non pas la DNA-ligase 1 (Figure 1.8, page précédente).

## 1.4 LA RÉPONSE INTÉGRÉE AUX DOMMAGES DE L'ADN

La réponse que la cellule oppose aux dommages de l'ADN ne se limite pas à la mise en œuvre de voies de réparation mais procède bien d'un processus intégré qui voit la physiologie de la cellule temporairement redéfinie afin d'optimiser la réparation des lésions détectées. Si l'on songe à toutes les étapes « physiologiques » qui précèdent la réparation à proprement parler — ou à tout le moins qui en sont concomitantes —, on peut citer les protéines qui détectent les dommages (protéines « senseur », comme KU), celles qui transduisent l'information (ATM et ATR, par exemple et les protéines de signalisation du « *checkpoint* », par exemple Chk1 et 2), les acteurs qui interviennent au voisinage de la lésion, comme les protéines de remodelage de la chromatine et *enfin* les protéines de réparation.

L'action concertée de l'ensemble de ces acteurs est typiquement nommée « *réponse aux dommages de l'ADN* ». La portée des activités des différents acteurs ci-dessus est telle que la cellule met en place une reprogrammation du transcriptome et du protéome. On observe une inhibition rapide des activités globales de transcription et de traduction, à l'exception de certains gènes favorisant la réparation. Ces reprogrammations doivent permettre de moduler des activités cellulaires spécifiques. Ainsi, à court terme, la cellule doit pouvoir induire des signaux sous forme de modifications post-traductionnelles comme la phosphorylation, l'acétylation, la méthylation, la parylation, par exemple (voir Figure 1.9, page ci-contre). À moyen terme, la cellule doit pouvoir induire l'expression de gènes codant pour les protéines de réparation, permettant ainsi de recruter de nouveaux complexes au niveau des cassures. Les étapes précoces de cette réponse cellulaire à la présence de cassures double-brin impliquent une réorganisation spatiale rapide des protéines, avec des regroupements sur les dommages, que l'on appelle *foyers*, et qui sont souvent exploités par les expérimentateurs pour localiser et compter dans des cellules fixées les événements de réparation.

### 1.4.1 DÉTECTION DES DOMMAGES ET TRANSDUCTION DU SIGNAL

La première étape de la réparation (au sens large) des cassures double-brin (CDB) de l'ADN consiste en leur détection et en l'activation d'une voie de signalisation aboutissant au recrutement des protéines de réparation au niveau des extrémités de l'ADN. Chacune des trois voies de réparation possède des protéines qui détectent les cassures double-brin de l'ADN et

En anglais, « *DNA damage response* »; DDR

En anglais, « *foci* »

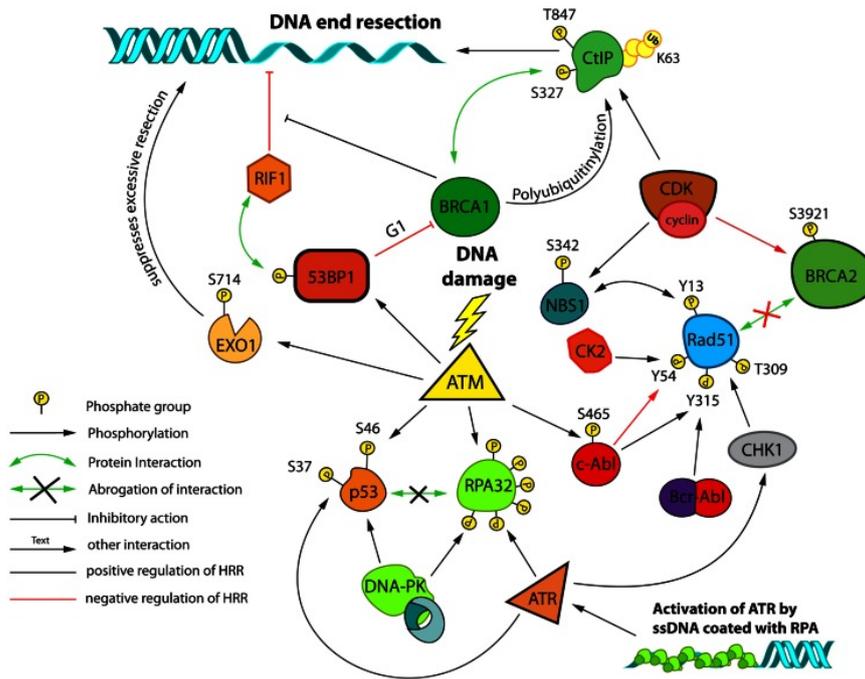


FIGURE 1.9 – Modifications post traductionnelles et interactions impliquées dans le choix de la voie de réparation des CDB. Un réseau de protéines contrôle la recombinaison homologue et est régulé par le cycle cellulaire par le biais de modifications post traductionnelles. Elles reposent en premier lieu sur les phosphorylations cycline/CDK dépendantes. Ces régulations en réponse aux dommages reposent sur les activités des kinases ATM, ATR et DNA-PKcs (PRKDC, p 177) et le contrôle de la résection des brins y joue un rôle central. Les fragments d'ADN simple-brin produits par la résection et recouverts de RPA constituent une plateforme pour l'activation de ATR. Tiré de [Mladenov et al. \[2013\]](#).

enclenchent une voie de signalisation aboutissant à la mise en place de cette voie.

#### 1.4.1.1 DÉTECTION DES CDB DANS LE NHEJ CANONIQUE

Dans la réparation des CDB par la voie canonique du NHEJ, l'hétérodimère KU reconnaît spécifiquement les extrémités de l'ADN double-brin et s'y fixe. Ce recrutement est extrêmement rapide (de l'ordre de quelques secondes ; voir [Mahaney et al. \[2009\]](#)) et est suivi de celui de PRKDC (p 177). La fixation de PRKDC sur l'ADN provoque son activation par autophosphorylation. Cette fixation ne nécessite pas la présence de l'hétérodimère KU, mais ce dernier permet de la stabiliser et augmente son activité. KU recrute

également le complexe de ligation en interagissant avec XRCC4, recrutement qui ne dépend pas de la présence de PRKDC [Mari *et al.*, 2006].

Bien que PRKDC soit capable de phosphoryler de nombreuses protéines *in vitro*, seul un nombre plus restreint de substrats, parmi lesquels elle-même et Artemis, ont pu être identifiés *in vivo*, ce qui souligne encore plus le fait que KU porte la responsabilité principale de la mise en œuvre du NHEJ canonique [Meek *et al.*, 2007].

#### 1.4.1.2 DÉTECTION DES CDB DANS LE NHEJ ALTERNATIF

Dans la voie alternative du NHEJ, les CDB de l'ADN sont reconnues par PARP1, qui entre en compétition avec KU au niveau des extrémités de l'ADN [Wang *et al.*, 2006]. PARP1 phosphoryle les protéines à proximité de la cassure, ce qui provoque le recrutement de celles impliquées dans la résection des extrémités (MRN, CtIP et BRCA1; voir Rass *et al.* [2009]; Zhang et Jasin [2011]), résection qui peut cependant être inhibée par 53BP1 [Bothmer *et al.*, 2010] et contrôlée par des facteurs annexes (RIF1, par exemple).

#### 1.4.1.3 DÉTECTION DES CDB DANS LA HR

La voie de la recombinaison homologue est celle qui nécessite la mise en œuvre des mécanismes moléculaires les plus complexes. Les cassures double-brin sont reconnues par le complexe MRN qui recrute ensuite ATM et TIP60. L'activité acétyl-transférase de TIP60, associée aux extrémités, permet d'induire l'activation par autophosphorylation de ATM, qui déclenche ensuite *via* son activité kinase la réponse aux cassures double-brin de l'ADN [Bhatti *et al.*, 2011], caractérisée en particulier par la phosphorylation des variants d'histones H2AX en  $\gamma$ H2AX [Fillingham *et al.*, 2006; van Attikum et Gasser, 2009].  $\gamma$ H2AX sert ensuite de plateforme pour le recrutement des différentes protéines de réparation par recombinaison homologue, et en particulier MDC1 (p 172) [Meyer *et al.*, 2013; Stucki et Jackson, 2006], dont les foyers sur l'ADN constituent un marqueur caractéristique de la présence de cassures double-brin. En raison du mécanisme intrinsèque de la recombinaison homologue (voir § 1.2.4, page 13), cette reconnaissance des dommages n'est pas suffisante pour initier la réparation, mais elle doit être accompagnée d'une décondensation de la chromatine dans le voisinage de la cassure.

### 1.4.2 REMODELAGE DE LA CHROMATINE

Dans les cellules, l'enroulement de l'ADN pour former la chromatine peut constituer un obstacle important à la réparation d'une cassure située dans une région condensée du génome (en particulier, comme nous venons de l'indiquer, pour la réparation par HR). Le remodelage de la structure

chromatinienne autour d'une lésion est donc l'une des réponses essentielles qui permettent de préparer la réparation (voir [Gospodinov et Herceg \[2013\]](#)). Ce remodelage a lieu *via* des modifications post-traductionnelles des histones, protéines responsables de la formation des nucléosomes qui compactent l'ADN sous forme de chromatine. La modification la plus caractéristique de la présence de cassures double-brin de l'ADN est la phosphorylation de la protéine H2AX [[van Attikum et Gasser, 2009](#)]. Cette phosphorylation provoque elle-même la parylation par PARP1 de la chromatine entourant les dommages et le recrutement de la protéine MDC1, qui sert de point d'ancrage pour les facteurs de remodelage de la chromatine. MDC1, phosphorylée par ATM recrute ensuite les ubiquitine ligases RNF8 et RNF168, dont l'activité d'ubiquitylation des histones provoque la décondensation de la chromatine. RNF8 peut s'associer avec CHD4 (p 155) afin de former le complexe NurRD qui permet la décondensation de structures chromatiniennes plus complexes [[Luijsterburg et al., 2012](#)]. Dans le cas des dommages présents dans l'hétérochromatine, des mécanismes de décompaction nécessitant des protéines spécifiques (KAP1) peuvent être nécessaires afin de permettre la réparation des cassures double-brin [[Goodarzi et Jeggo, 2012](#)].

### 1.4.3 CYCLE CELLULAIRE ET APOPTOSE

Il existe trois points de contrôle principaux (« *checkpoints* ») pour la progression du cycle cellulaire. Le premier, situé en phase G1 contrôle l'entrée des cellules en phase S, en fonction des conditions environnementales (qui permettent ou non la division cellulaire).

Le second est situé en phase G2. Il permet de vérifier l'intégrité de l'ADN avant l'entrée des cellules en mitose, et donc de s'assurer que les deux cellules filles recevront bien la même quantité d'ADN. En cas de présence de foyers  $\gamma$ H2AX, les protéines Chk1 et Chk2, en interaction avec les kinases cycline-dépendantes, bloquent cette progression et empêchent le passage en phase G2/M afin de permettre aux cellules de réparer leur ADN [[Smith et al., 2010](#)]. Si ce blocage est prolongé, les dommages sont considérés comme trop importants pour pouvoir être réparés et la voie apoptotique dépendante de p53/TP53 est déclenchée par ATM [[Shiloh et Ziv, 2013](#)], afin de préserver la stabilité génétique de l'organisme [[Surova et Zhivotovsky, 2013](#)].

Enfin, le troisième point de contrôle est celui relatif à la bonne marche du fuseau achromatique. Ce point de contrôle est essentiel à la bonne répartition du matériel génétique dans les deux cellules filles, répartition qui ne peut bien se faire que si les systèmes de rail (microtubules du fuseau) et de locomotives (protéines moteurs, comme la dynamine) fonctionnent parfaitement.

## 1.5 RÉPARATION ET THÉRAPIE ANTICANCÉREUSE

### 1.5.1 IRRADIATION DES CELLULES TUMORALES

Les rayonnements ionisants employés en radiothérapie sont destinés à induire une régression tumorale, comme effet des très nombreux dommages induits dans l'ADN. Ces rayonnements ont des caractéristiques physiques qui les classent en deux catégories : les rayonnements de particules déposant leur énergie très rapidement dès leur contact avec la matière molle (tissus) et les rayonnements dont les particules déposent leur énergie de manière moins rapide. Les particules ionisantes qui déposent leur énergie sur une trajectoire courte sont dites, en anglais, « *high linear energy transfer* » (*high LET*) ; les particules qui déposent leur énergie sur une trajectoire plus longue sont dites « *low linear energy transfer* » (*low LET*).

Plusieurs auteurs ont revu les études portant sur les dommages complexes générés par les rayonnements ionisants dans les cellules tumorales [Sage et Harrison, 2011; Schipler et Iliakis, 2013]. Les dommages induits dans l'ADN par les rayonnements à particules *high LET* sont d'une chimie hautement complexe et ce degré de complexité concourt à l'efficacité du traitement par induction, dans les cellules irradiées, d'une forte apoptose. Ces dommages peuvent être classés en fonction de leur complexité et l'on y trouve les sites de dommages groupés « *clustered damage sites, CDS* ». Ces dommages particuliers consistent en un ensemble de lésions (CSB, CDB, dommages aux bases ou aux nucléotides) qui se forment dans une courte région de l'ADN (un tour d'hélice) sur l'un ou les deux brins (voir quelques exemples à la Figure 1.10). La proximité de plusieurs dommages sur une courte séquence d'ADN peut en empêcher la réparation par les systèmes dédiés pour cause d'encombrement stérique. La prise en charge de ces lésions par les cellules d'eucaryotes peut produire des CDB après les actions combinées d'endonucléases et de DNA-polymérases engagées dans les systèmes de réplication ou de réparation (BER), ce qui affecte la survie cellulaire.

Des modèles d'étude de combinaisons de lésions ont été réalisés afin de mieux comprendre leurs conséquences biologiques, en particulier résultant du mode de réparation mis en œuvre par les cellules. Ces travaux montrent que les lésions proches sur le même brin (lésions en tandem) sont peu réparées et peuvent se révéler létales. Dans le cas de sites abasiques en vis-à-vis sur les deux brins, il y a tendance à la formation de CDB. Or, ces CDB ne sont correctement réparées qui si aucun dommage ne se trouve à proximité. Dans le cas contraire, la réparation défailante de ces CDB conduit à l'instabilité génétique bien connue [Malyarchuk *et al.*, 2013; Sage et Harrison, 2011].

Dans les sites de dommages multiples, les CDB sont prises en charge par la voie du NHEJ, induite par la phosphorylation de PRKDC (p 177). La flexibilité structurale de cet enzyme lui permet en effet de s'accomoder à

une grande variété de structures, ce qui est un facteur déterminant dans ce contexte. L'intervention de la protéine Artemis peut être nécessaire afin de rendre ces extrémités aptes pour la ligation [Malyarchuk *et al.*, 2013]. Si, cependant, la complexité des dommages multiples est trop importante pour permettre rapidement la réparation par le NHEJ, ces derniers peuvent être éliminés par résection des extrémités puis par le NHEJ alternatif ou par la voie HR.

En conclusion, ces études montrent que l'efficacité de ces cassures en thérapie anticancéreuse pourrait encore être augmentée en inhibant les protéines impliquées dans les étapes tardives des différentes voies de réparation [Sage et Harrison, 2011].

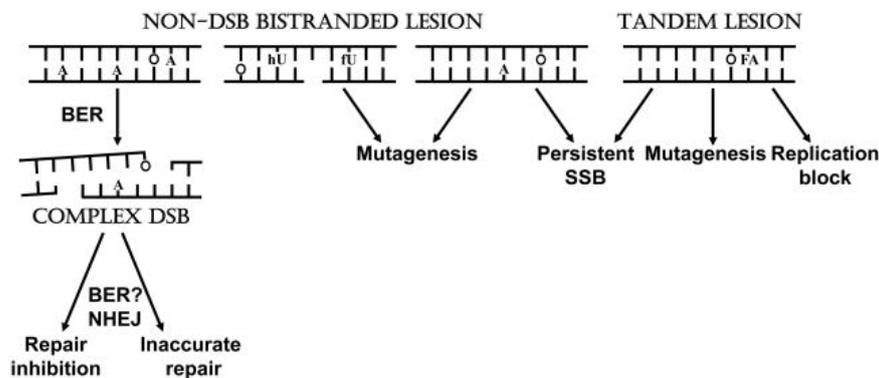


FIGURE 1.10 – Structures d'oligonucléotides duplexes d'ADN à dommages multiples. Les différentes structures représentées ont été utilisées pour étudier la réparation des dommages multiples. Ces structures sont représentatives des dommages multiples qui peuvent être induits par des rayonnements ionisants « *high LET* ». A, O, fU, hU, FA symbolisent respectivement sur la figure un site abasique, une 8-oxoG, un 5-formyluracil, un 5-hydroxyuracil et une formylamine. Voir texte. Tiré de Sage et Harrison [2011].

### 1.5.2 EXPLOITATION DES DÉFAUTS DES CELLULES TUMORALES

Dans les cellules tumorales, les mécanismes qui régulent les voies de réparation de l'ADN peuvent être altérés en raison, par exemple, de mutations dans certains gènes ou de réarrangements chromosomiques. Ces événements peuvent rendre les cellules incapables d'utiliser une voie de réparation donnée, et aggravent donc leur instabilité génétique, avec l'émergence de cellules qui ont un avantage sélectif pour la prolifération. Les défauts, qu'ils soient observés dans les voies de réparation de l'ADN, dans l'induction de l'apoptose ou encore dans la régulation de la prolifération cellulaire, sont caractéristiques

des cellules cancéreuses. En contre-partie, les cellules tumorales sont souvent appauvries en mécanismes de réparation, ce qui les rends plus vulnérables, et peut être exploité pour concevoir de nouvelles stratégies thérapeutiques.

Le cas du cancer du sein illustre bien cette corrélation entre la diminution de l'efficacité de la réparation de l'ADN et le risque de développer un cancer. Parmi les 19 gènes dont les mutations sont impliquées dans le développement de ce cancer selon COSMIC (« *Catalog of somatic mutations in cancer* », <http://cancer.sanger.ac.uk/cancergenome/projects/census>), 11 interviennent dans la réponse aux dommages de l'ADN. De manière intéressante, cette diminution de la capacité des cellules cancéreuses à réparer leur ADN peut être mise à profit pour provoquer leur apoptose. En effet les cellules dont l'ADN est fortement endommagé, et qui ne peuvent pas le réparer, induisent la voie apoptotique médiée par p53/TP53 [Shiloh et Ziv, 2013]. De manière générale les cellules tumorales sont sensibles à l'inhibition des mécanismes de réparation de l'ADN, qui augmente le temps nécessaire aux cellules pour restaurer l'intégrité de leur génôme (en particulier en phase G2 du cycle cellulaire, un point de contrôle empêche l'entrée des cellules en mitose). L'utilisation de cette stratégie est d'autant plus efficace que les cellules cancéreuses prolifèrent plus rapidement que les autres cellules, ce qui se traduit par un temps de division plus court. Il est encore plus intéressant d'utiliser un agent permettant d'induire un type particulier de dommages de l'ADN afin de traiter des cancers associés à des défauts des voies de réparation spécifiques de ces dommages. Cette stratégie permet d'induire plus spécifiquement l'apoptose des cellules tumorales. Ainsi, les sels de platine, comme le cisplatine, peuvent être utilisés pour induire des pontages intra- et inter-brins, dommages normalement réparés par le NER et la voie HR, pour traiter les patientes atteintes de cancer des ovaires. Cette stratégie est efficace car les cellules tumorales retrouvées dans ce type de cancer présentent souvent des défauts de la réparation par recombinaison homologue (le plus souvent dus à des mutations dans le gène codant BRCA1 ou BRCA2). L'utilisation de ce composé est également efficace chez les personnes atteintes de cancer du poumon et dont les cellules tumorales présentent des mutations dans le gène codant pour ERCC1, une protéine du NER.

Le concept de « létalité synthétique » a été formulé au milieu du siècle dernier sur la base d'expériences menées sur des populations de drosophiles [Dobzhansky, 1946; Lucchesi, 1968]. Selon ce concept, une protéine peut devenir essentielle à la survie d'un organisme lorsqu'il développe des mutations de gènes impliqués dans des mécanismes pourtant différents de ceux dans lesquels la protéine devenue essentielle intervient. Il a été établi par la suite que la létalité synthétique révélait l'inactivation conjointe de deux voies qui étaient capables de se suppléer l'une l'autre. Il a fallu attendre quelques années pour voir apparaître des applications de ce principe à la

réparation de l'ADN dans le cadre de la thérapie anti-cancéreuse [Hartwell *et al.*, 1997; Jr, 2009]. Des inhibiteurs ont été développés pour le traitement spécifique de certains cancers faisant intervenir des défauts dans une voie de réparation de l'ADN. Les inhibiteurs des protéines PARP, impliquées dans les voies de réparation du BER et du NHEJ alternatif, se sont avérés particulièrement efficaces pour induire l'apoptose dans les cellules tumorales présentant des défauts dans la voie de réparation par recombinaison homologue, et en particulier des mutations dans les gènes des protéines BRCA. Cette efficacité repose sur le fait que le BER est un mécanisme très utilisé par les cellules pour réparer les cassures simple-brin de l'ADN (qui sont des lésions peu nocives pour les cellules, mais qui apparaissent avec une fréquence élevée). L'inhibition de PARP1, une protéine essentielle pour la mise en place de ce mécanisme, provoque donc l'accumulation de ces lésions dans le génome, qui empêcheront la progression de la fourche de réplication en phase S, mais aussi potentiellement la formation de cassures double-brin de l'ADN si elles sont situées à proximité les unes des autres. La persistance prolongée de dommages de l'ADN jusques à la phase G2 enclenche alors la voie apoptotique, médiée par la protéine p53. Cette entrée en apoptose est évitée dans les cellules saines par la réparation des dommages, qui est réalisée à ce moment du cycle cellulaire par recombinaison homologue (la présence d'une chromatide soeur permet en effet d'utiliser ce mécanisme, qui présente l'avantage de réparer plus fidèlement l'ADN). En revanche, dans les cellules tumorales, les défauts de cette voie de réparation constituent donc un véritable verrou pour la restauration de l'intégrité génétique et permettent d'induire spécifiquement l'apoptose.

Actuellement, plusieurs de ces inhibiteurs sont en phase clinique. Le couple PARP/BRCA est celui qui a permis d'obtenir le plus de résultats, mais certaines équipes travaillent également sur l'utilisation d'inhibiteurs de protéines impliquées dans la réparation des bases oxydées pour traiter des cancers présentant des défauts dans la réparation des mésappariements.

Le principe de la létalité synthétique montre, de manière générale, que les voies de réparation de l'ADN interagissent étroitement les unes avec les autres et que, bien que chacune d'entre elle permette la réparation d'un type particulier de dommage, elles peuvent également se suppléer en cas d'inactivation de l'une d'entre elles. La redondance entre ces voies est moindre dans les cellules tumorales, qui sont appauvries en systèmes de réparation et donc plus sensibles aux dommages de l'ADN. Cette observation, et les résultats prometteurs des études qui en découlent, montrent l'importance de l'étude des voies de réparation de l'ADN. En particulier, mieux comprendre la manière dont les cellules déterminent quelle voie de réparation doit être utilisée permettrait de souligner les voies qui sont complémentaires et de trouver de nouveaux couples de protéines dont l'inactivation pourrait être mise à profit en thérapie anti-cancéreuse [Mladenov *et al.*, 2013].

Ce cas de figure est illustré par les cassures double-brin de l'ADN générées par des rayonnements ionisants. La structure de ces extrémités est en effet incompatible avec l'activité catalytique des enzymes permettant habituellement de réparer les cassures double-brin de l'ADN (DNA-polymérase et DNA-ligase 4). La réparation de ces cassures nécessite donc l'intervention d'enzymes capables de restaurer la structure canonique aux extrémités libres de l'ADN. Bien que les rayonnements ionisants puissent induire la formation d'une grande variété de structures au niveau des extrémités des cassures double-brin générées, seuls quelques enzymes sont connus à ce jour. Des stratégies nouvelles doivent être imaginées pour rechercher de nouveaux partenaires moléculaires des voies de réparation déjà connues. Dans les sections suivantes, nous décrirons les différentes méthodes qui ont été utilisées pour l'étude des interactions protéine-ADN. Il s'agit de montrer ce qui a été fait, mais aussi d'introduire la méthodologie de purification par affinité de protéines reconnaissant spécifiquement les extrémités libres de l'ADN qui est développée au laboratoire depuis quelques années.

## 1.6 ÉTUDE DES INTERACTIONS PROTÉINE-ADN

Dans les complexes nucléoprotéiques, les acides nucléiques servent souvent de plateforme au recrutement de protéines. Les protéines peuvent s'assembler spécifiquement sur une séquence nucléotidique donnée (séquence de promoteur, par exemple) ou, alors, reconnaître une structure particulière formée par l'acide nucléique (extrémités libres de l'ADN, renflements, par exemple). Si l'on s'intéresse à la spécificité de telle ou telle protéine interagissant avec l'ADN, le motif reconnu, qu'il soit de séquence ou éminemment structural, constitue l'objet recherché.

En fonction des situations évoquées ci-dessus, un certain nombre de méthodes ont été élaborées afin d'étudier les interactions protéine-ADN. Ces méthodes peuvent être distinguées selon que l'expérimentateur se trouve dans l'une ou l'autre de ces situations :

- \* La séquence de l'ADN ciblée par la/les protéine/s recherchée/s est connue (facteur de transcription reconnaissant une séquence régulatrice donnée, par exemple) ;
- \* La protéine est connue et l'on recherche la séquence sur laquelle elle se lie spécifiquement ;
- \* L'ADN a une structure particulière (G-quadruplexe, renflement ou cassure, par exemple) et l'on recherche toute protéine qui la reconnaît ;
- \* On dispose d'une protéine dont on veut tester la capacité à se lier à

une structure ou une séquence donnée.

Les méthodes que nous décrivons ci-dessous peuvent être mises en œuvre *in vitro* ou *in vivo*.

### 1.6.1 MÉTHODES D'ÉTUDE *in vivo*

#### 1.6.1.1 CHIP « *chromatin immunoprecipitation* »

L'immunoprécipitation de la chromatine est une technique en cellule qui permet de définir les sites de fixation sur l'ADN d'une protéine connue dont on dispose d'un anticorps. Tout d'abord, les interactions entre la protéine d'intérêt et l'ADN sont figées par l'usage d'un agent pontant. L'ADN génomique est alors fragmenté aux ultrasons et les fragments qui ont lié la protéine d'intérêt sont immunoprécipités grâce à un anticorps dirigé contre elle. On peut alors scinder le lien covalent entre la protéine et l'ADN afin de séquencer ce dernier. Une variante de cette technique est réalisée en conditions natives. Dans ce cas la fragmentation de la chromatine est effectuée par des nucléases en conditions douces, ce qui malheureusement diminue la précision avec laquelle la séquence reconnue par la protéine d'intérêt est déterminée [Fullwood et Ruan, 2009].

#### 1.6.1.2 DOUBLE ET TRIPLE HYBRIDE

Il est également possible d'identifier des ARN reconnaissant une protéine spécifique en utilisant la technique de triple hybride. Cette technique demande la constitution préalable d'une banque de plasmide qui permettent de produire un pool d'ARN comportant chacun une séquence variable (faisant l'objet de l'étude d'interaction) et une séquence constante (mise à profit pour la récupération). Des levures sont ensuite transformées avec l'un de ces plasmides, et les ARN sont recrutés en amont d'un gène rapporteur *via* une protéine reconnaissant la partie constante de cet ARN et fusionnée à un domaine de liaison à l'ADN. Les protéines dont l'interaction avec l'ARN cible est testée, fusionnées avec un domaine activateur spécifique du promoteur du gène rapporteur sont exprimées dans les cellules. Si ces protéines sont capables d'interagir avec la partie variable de l'ARN, l'interaction induit l'expression du gène rapporteur qui permet ainsi de la détecter.

Les techniques décrites ci-dessus permettent donc d'identifier *in vivo* à partir d'une protéine donnée (présente dans un échantillon purifié) les séquences nucléiques reconnues par cette protéine. Ces données peuvent être exploitées pour déterminer par exemple les promoteurs qui sont régulés par un facteur de transcription, ce qui pourrait permettre d'identifier les gènes responsables de l'apparition d'un cancer sachant que les personnes qui en sont atteintes présentent un dysfonctionnement pour un ou plusieurs facteurs de

transcription. Cette technique a notamment permis de déterminer l'affinité de la protéine Daz1 pour la séquence 5'UTR de l'ARNm de la protéine Cdc25C, qui est impliquée dans la transition en phase M du cycle cellulaire [Venable *et al.*, 2001].

### 1.6.2 IN SILICO

Lorsque la séquence nucléotidique est connue de l'expérimentateur et qu'une protéine d'intérêt est déjà définie, il est possible de tirer parti de bases de données qui listent les sites de liaison de protéines en fonction de leur rôle (par exemple TRANSFAC et JASPAR regroupent les données relatives aux facteurs de transcription connus). A partir de ces données, il est possible de rechercher, en utilisant des algorithmes, des sites de reconnaissance putatifs sur la séquence nucléotidique d'intérêt. L'intérêt de ces outils (parmi lesquels nous pouvons citer MATCH, Kel *et al.* [2003] et MatInspector, Cartharius *et al.* [2005], qui sont destinés à la prédiction de sites de fixation pour des facteurs de transcription) est qu'ils sont rapides à prendre en main, qu'ils ne génèrent pas de coûts importants et qu'ils donnent rapidement des résultats, mais ils sont limités par plusieurs paramètres importants :

- \* Le manque de sensibilité et de spécificité de ces méthodes engendre des erreurs qu'il n'est pas facile de repérer en se fondant exclusivement sur des tests bioinformatiques ;
- \* La présence d'une séquence consensus sur un acide nucléique n'est pas le seul paramètre selon lequel une protéine s'y fixe ou non (en particulier nous avons discuté précédemment de l'importance des modifications post-traductionnelles ainsi que des interactions protéines - protéines) ;
- \* Ces analyses peuvent ne pas prendre en compte la structuration de l'acide nucléique qui rend la séquence consensus accessible ou non aux protéines qui la reconnaissent et les conclusions de ces études ne peuvent donc pas facilement être étendues au contexte cellulaire ;
- \* Ils sont limités à la recherche de protéines reconnaissant spécifiquement une séquence et ne peuvent donc pas être utilisés pour identifier les structures reconnues par des protéines.

### 1.6.3 MÉTHODES D'ÉTUDE *in vitro*

#### 1.6.3.1 EMSA (« *electrophoretic mobility shift assay* »)

Cette méthode, appelée en français « gel retard », permet de détecter des interactions entre un oligonucléotide de séquence connue (et parfois radiomarqué) et une ou plusieurs protéine(s) [Hellman et Fried, 2007; Tsai *et al.*, 2012]. L'expérimentateur définit la séquence de l'oligonucléotide qu'il met en présence de l'échantillon à tester. Le mélange est ensuite soumis

<http://www.gene-regulation.com/pub/databases.html>

<http://jaspar.genereg.net/>

à une électrophorèse non dénaturante et la migration de l'oligonucléotide renseigne sur le fait qu'il ait fixé ou pas une ou plusieurs protéines. Sa migration sera, en effet, retardée sur gel par rapport à celle de l'oligonucléotide nu. Si l'on soupçonne que plusieurs protéines d'un mélange se sont liées à l'oligonucléotide, il est possible de réaliser une seconde dimension afin d'appréhender les diverses espèces moléculaires en présence (voir Figure 1.11) et on peut également ensuite effectuer une incubation supplémentaire en présence d'un anticorps dirigé contre une protéine qui se serait liée afin de voir s'il y a un retard *supplémentaire* [Vetchinova *et al.*, 2006]. L'interaction ADN-protéine est alors validée en mutant la séquence de l'oligonucléotide, ce qui doit abolir le retard sur gel.

« Supershift mobility assay »

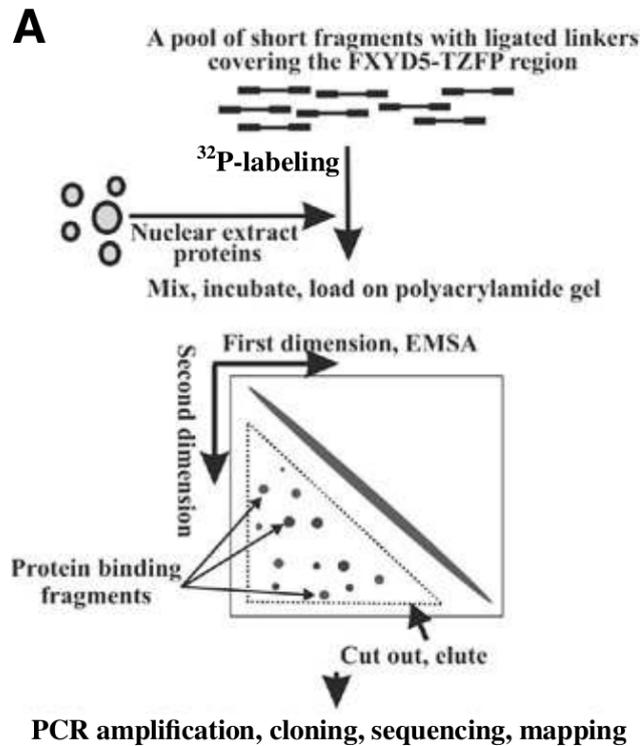


FIGURE 1.11 – Principe du gel retard en deux dimensions. L'EMSA en deux dimensions permet de visualiser la diversité des complexes formés lors de l'incubation d'un ou plusieurs oligonucléotide(s) d'intérêt avec un extrait protéique. Les protéines peuvent ensuite être identifiées par spectrométrie de masse, et les oligonucléotides par PCR puis séquençage (il est alors nécessaire de modifier les oligonucléotides avant de réaliser l'EMSA pour placer des régions sur lesquelles des amorces génériques viendront s'hybrider). D'après Vetchinova *et al.* [2006].

### 1.6.3.2 EMPREINTE AUX ENDONUCLÉASES

Les oligonucléotides sont d'abord incubés avec un échantillon protéique et ensuite soumis à une digestion nucléolytique par des nucléases non spécifiques. Les éléments de séquence de l'oligonucléotide sur lesquels des protéines se seraient fixées sont alors protégés et ne peuvent être digérés par les nucléases. Après endonucléolyse, les fragments d'ADN sont analysés sur gel et leur dimension renseigne sur les régions qui étaient protégées. Il est évidemment possible de séquencer les fragments. Historiquement l'enzyme utilisée est la DNase I [Galas et Schmitz, 1978], mais il est également possible d'utiliser d'autres enzymes, comme la benzonase, ou encore des agents chimiques. Comparée à l'EMSA, l'empreinte aux nucléases présente l'intérêt de pouvoir être fournir un renseignement sur l'élément de séquence reconnu par la protéine, et ainsi que sur l'affinité avec laquelle cette dernière est fixée [Bailly *et al.*, 2005; Fletcher et Fox, 1996] (voir Figure 1.12, page ci-contre).

### 1.6.3.3 PUCES À ADN

Les puces à ADN rassemblent sur un support de dimensions réduites et à des emplacement clairement définis (une sorte de grille) plusieurs milliers d'oligonucléotides de séquences différentes connues. Ces puces sont utilisées avec des mélanges plus ou moins complexes de protéines connues. Après incubation, l'on peut connaître quelles séquences oligonucléotidiques auront été reconnues par des protéines. Cette technique correspond à une forme de parallélisation massive, en solution, de l'EMSA [Hu *et al.*, 2009].

### 1.6.3.4 SOUTHWESTERN BLOT

Le southwestern blot est une technique qui permet de déterminer la masse moléculaire apparente d'une protéine reconnaissant un oligonucléotide donné sans avoir à utiliser d'anticorps. Cette technique est réalisée en deux étapes : dans un premier temps un échantillon complexe est résolu en conditions dénaturantes sur un gel SDS-PAGE, puis les protéines ainsi séparées sont transférées sur une membrane dans des conditions qui permettent leur renaturation. Ces membranes sont ensuite incubées avec l'oligonucléotide d'intérêt, qui est radiomarqué. La position de cet oligonucléotide sur la membrane est ensuite révélée par radiographie, et comparée à celles des bandes de gel SDS-PAGE, ce qui fournit des indices sur l'identité des protéines qui reconnaissent l'oligonucléotide d'intérêt [Siu *et al.*, 2008]. Les protéines peuvent également être identifiées de manière inabiguë par spectrométrie de masse si la seconde étape (transfert des protéines sur une membrane) est remplacée par une seconde dimension d'électrophorèse [Jiang *et al.*, 2009]. Cette technique est intéressante, car elle permet d'identifier de manière exhaustive les protéines reconnaissant directement un oligonucléotide donné,

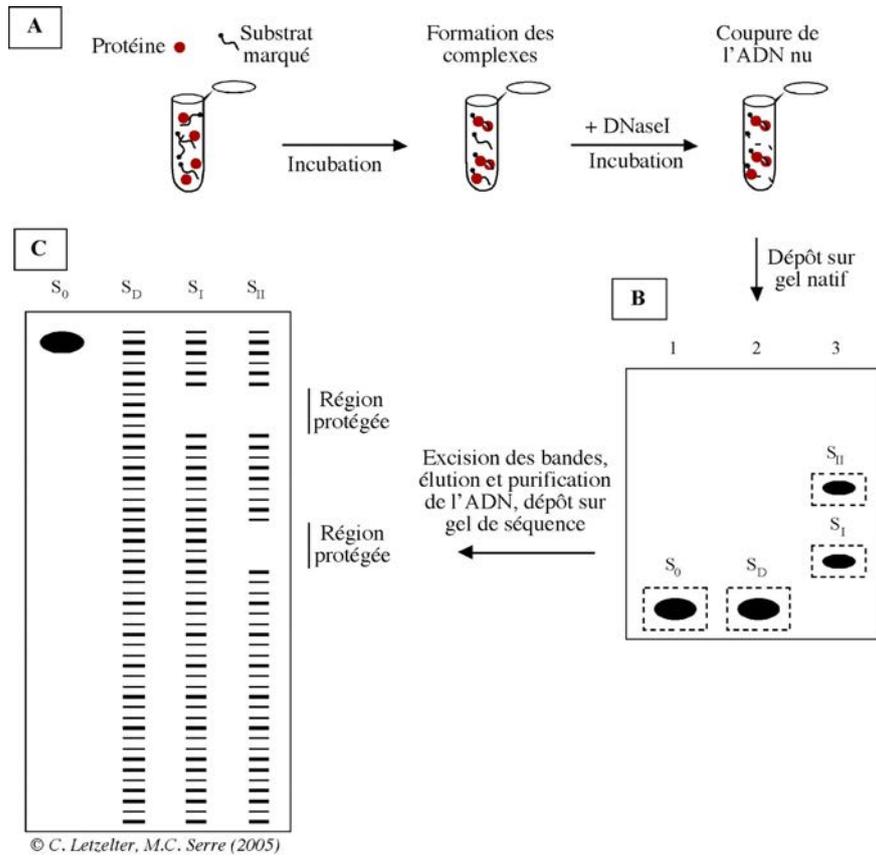


FIGURE 1.12 – **Principe de l’empreinte aux nucléases.** L’empreinte aux nucléases permet d’identifier dans un oligonucléotide les portions de séquence qui sont reconnues par une ou plusieurs protéines. Dans l’expérience schématisée ci-dessus, la bande  $S_0$  correspond à un oligonucléotide témoin, qui n’a jamais vu de protéines. La bande  $S_D$  correspond à un oligonucléotide nu, digéré par des nucléases et les bandes  $S_1$  et  $S_2$  correspondent à des complexes ADN-protéines soumis à une digestion nucléotidique. Les régions sur lesquelles des protéines se sont fixées correspondent aux parties du gel de séquence pour lesquelles il n’y a pas de signal de radioactivité.

mais elle ne permet pas en revanche l’étude de complexes multiprotéiques en raison des conditions dénaturantes utilisées lors de la première étape (gel SDS-PAGE).

### 1.6.3.5 PURIFICATION PAR AFFINITÉ

Les techniques décrites ci-dessus permettent de tester *in vitro* des interactions entre des protéines et des séquences dont certains éléments au moins sont connus. Il faut remarquer que toutes les méthodes fondées sur l’emploi

d'anticorps imposent un *a priori* quant à l'identité des protéines étudiées. La chromatographie d'affinité est fondée sur l'idée que l'un des partenaires d'une interaction est connu et que les autres (inconnus) s'y lieront. Ainsi, l'un des partenaires est fixé sur un support chromatographique — devenant ainsi le « *déterminant d'affinité* ». Un mélange comportant des molécules interagissant avec ce déterminant d'affinité est alors chargé sur le support. Si la nature de l'interaction entre déterminant d'affinité et interactant(s) est très spécifique, la purification peut avoir un haut rendement en une seule étape. Dans d'autres cas, le déterminant d'affinité peut avoir une spécificité moins marquée et le nombre de molécules s'y liant peut être plus important.

Pour constituer une phase chromatographique d'affinité, on commence par coupler l'un des partenaires de l'interaction connue sur le support. Ce couplage peut être effectué chimiquement, par formation de liaisons covalentes, mais aussi non covalentes. Dans ce dernier cas, un couple particulièrement facile à mettre en œuvre est le couple streptavidine–biotine, dont le  $K_d$  est de  $10^{-14}$  M. Typiquement, les billes (support chromatographique) sont recouvertes de la protéine streptavidine. Le déterminant d'affinité est lié covalamment à la biotine. L'ensemble entre alors fortement en interaction, ce qui fait que les billes sont dites « fonctionnalisées » avec le déterminant d'affinité. Nous avons choisi d'utiliser cette stratégie pour notre étude, en raison de sa facilité d'utilisation.

La purification par affinité se prête ordinairement à des étapes analytiques ultérieures, comme l'identification par spectrométrie de masse des protéines ayant préalablement été purifiées sur une phase d'affinité [Snijder et Heck, 2014].

La purification de protéines interagissant avec des acides nucléiques par chromatographie d'affinité est une stratégie bien connue, qui se réalise en plusieurs étapes :

- \* Le déterminant d'affinité (acide nucléique) recrute une ou plusieurs protéines d'intérêt d'un mélange complexe ;
- \* Les autres protéines du mélange, qui pourraient interagir faiblement avec la phase chromatographique, sont éliminées par différents lavages (faisant intervenir dans la majorité des cas une augmentation de la force ionique) ;
- \* La ou les protéines d'intérêt sont alors purifiées et peuvent ensuite être récupérées en les détachant de l'acide nucléique sur lequel elles se sont fixées (élution).

L'élution des analytes retenus sur une phase d'affinité peut être faite avec différents facteurs physico-chimiques, comme la force ionique, les détergeants, le pH, la température. Comme nous le détaillons plus bas, dans le chapitre des résultats, ces différentes méthodes ont des inconvénients

importants, dont nous pouvons citer celui qui, à nos yeux, est le plus important : le fait que lorsque l'on élue les protéines d'intérêt liées au déterminant d'affinité, immanquablement d'autres protéines liées sur le support chromatographique seront également détachées, qui viendront contaminer le produit de purification. Des études ont montré qu'il était possible d'augmenter spectaculairement le rendement de purification si l'élution est remplacée par le détachement pur et simple du produit de purification (déterminant d'affinité compris) du support chromatographique [Hégarat *et al.*, 2007]. Pour des revues, voir Hégarat *et al.* [2008]; Tacheny *et al.* [2013].

#### 1.6.4 CONCLUSION

Lorsqu'une interaction semble détectée par l'une ou l'autre des différentes techniques décrites ci-dessus, de nombreuses techniques biophysiques permettent de vérifier la validité des résultats obtenus. Il s'agit en général d'employer une technique orthogonale à la technique initialement employée. On peut distinguer, parmi elles, les méthodes d'étude *in vitro* (l'électrophorèse en conditions natives couplée à un double marquage, le FRET, la RMN, le dichroïsme circulaire, la microcalorimétrie, la SPR ou encore la microscopie à force atomique, par exemple) et *in vivo* (FRET *in situ*). Pour une revue, voir Dey *et al.* [2012].

De nombreuses techniques ont été développées afin d'étudier le rôle des facteurs de transcription [Jolma et Taipale, 2011; Tacheny *et al.*, 2013], dont l'importance est évidente afin de comprendre les mécanismes qui régulent l'expression des gènes. Ces techniques tirent parti des avancées qui ont été faites dans le domaine du séquençage des acides nucléiques, ainsi que du nombre important de bases de données permettant d'identifier une séquence par alignement de séquence. Il semble désormais important de développer d'autres méthodologies, adaptées à l'identification et à la caractérisation de protéines dont la particularité n'est pas de reconnaître une séquence donnée, mais une structure, de l'ADN par exemple, liée à une fonction métabolique. Dans le cadre de la réparation de l'ADN, nous nous sommes consacrés dans ces travaux de thèse au développement d'un ensemble de techniques permettant la purification et l'identification de complexes protéiques reconnaissant des extrémités libres de l'ADN. Nous avons choisi d'utiliser la chromatographie d'affinité, qui permet d'étudier des protéines sans *a priori* tant sur leur nombre que sur leur identité.



# CHAPITRE 2

## MATÉRIELS ET MÉTHODES



## 2.1 TAMPONS POUR LES PHASES CHROMATOGRAPHIQUES

Tableau 2.1: **Tampons utilisés pour la constitution des phases chromatographiques**

Composant	Quantité (1 L)	Conc. finale
<b>tampon A</b>		
Bis-TRIS HCL pH 7	41.85 g	20 mM
EDTA 500 mM (stock)	4 mL	2 mM
NaCl	11.69 g	200 mM
<b>tampon B</b>		
Bis-TRIS HCL pH 7	41.85 g	20 mM
EDTA 500 mM (stock)	4 mL	2 mM
NaCl	58.44 g	1 mM
<b>Tampon tampon TE</b>		
TRIS HCL pH 7	20.93 g	10 mM
EDTA 500 mM (stock)	2 mL	1 mM
<b>Tampon TEN 100<sup>1</sup></b>		
TRIS base HCL pH 7.5	1.211 g	10 mM
EDTA 500 mM (stock)	2 mL	1 mM
NaCl	5.844 g	100 mM
<b>Tampon TEN 1000</b>		
TRIS base HCl pH 7.5	1.211 g	10 mM
EDTA 500 mM (stock)	2 mL	1 mM
NaCl	58.44 g	1 M
<b>Tampon BN 500</b>		
Bis-TRIS	10.462 g	50mM
EDTA 500 mM (stock)	2 mL	1 mM
Glycérol	100 g	108.6 mM
Acide aminocaproïque	65.585 g	500 mM

## 2.2 PRÉPARATION DES PHASES CHROMATOGRAPHIQUES

L'élaboration, au laboratoire, des phases chromatographiques employées dans le cadre des études décrites dans ce mémoire s'appuie sur le couple

1. Ce tampon avait été utilisé pour l'incubation des billes avec les oligonucléotides, mais il a été remplacé par le tampon TEN 1000.

biotine–streptavidine. Nous ajoutons à ce couple un élément supplémentaire : un lien photoclivable dans l’ultraviolet. Ainsi, il est recommandé de travailler dans la pénombre dès que le matériel servant à la constitution des phases comporte ce lien photoclivable.

La création de phases chromatographiques fonctionnalisées avec des oligonucléotides passe par l’obtention de ces oligonucléotides duplexes d’ADN. S’il est possible de les acheter pré-synthétisés (ce qui a été fait au laboratoire par le passé), nous avons souhaité « conserver la main » sur le processus de production. Ainsi, avons-nous opté pour une synthèse par PCR.

### 2.2.1 PRÉPARATION DU MATÉRIEL NUCLÉIQUE

#### 2.2.1.1 SYNTHÈSE DES OLIGONUCLÉOTIDES PAR PCR

Les amorces utilisées pour la PCR sont achetées chez Eurogentech. La polymérase à ADN de T4 et le tampon qui lui est associé, ainsi que les désoxyribonucléotides tri-phosphate (dNTP) sont achetés chez New England Biolabs (NEBs). Le plasmide pUC19 servant de matrice est un produit d’Invitrogen. Le thermocycleur utilisé pour la réaction de PCR est une machine Eppendorf (modèle MasterCycler). Le marqueur de taille utilisé pour le contrôle des produits de PCR sur gel d’agarose est le marqueur « 50 bp ladder » de chez NEBs. La préparation des milieux réactionnels est effectuée sous hotte.

Comme détaillé plus loin, nous préparons deux types de phases chromatographiques, que nous décrivons dans le texte avec ces vignettes : \* et ✧. En fonction de la phase à préparer, la synthèse de l’oligonucléotide se fera avec des amorces bien définies (voir § 3.1, page 67, intitulée « *Un système chromatographique intégré* »). Les réactions de PCR sont effectuées par séries de 100 tubes, contenant chacun 50  $\mu\text{L}$  de milieu réactionnel. La composition du milieu de réaction est présentée dans le Tableau 2.2, page ci-contre. Pour des raisons pratiques, le milieu est préparé pour 100 tubes de PCR. Il faut alors préparer les deux mélanges 1 et 2 (pour 100 tubes, multiplier toutes les valeurs du tableau par 100). Une fois les mélanges préparés isolément, ils sont *homogénéisés* par inversion douce du tube une dizaine de fois et ils sont distribués dans 100 tubes de PCR (48.75  $\mu\text{L}$  du mélange 1 et 1.25  $\mu\text{L}$  du mélange 2, ce qui reconstitue le volume de 50  $\mu\text{L}$ ). Noter, que l’on pratique toujours deux contrôles, dans lesquels la matrice est absente, pour l’un, et l’enzyme est absent, pour l’autre.

Attention, le glycérol a tendance à sédimenter (avec l’enzyme). Ne pas faire mousser

température	durée
1 T 94°C	00 :04 :00
2 T 94°C	00 :00 :30
3 T 53°C	00 :01 :00
4 T 70°C	00 :00 :30
GOTO 2	REP 28
5 T 70°C	00 :10 :00
HOLD 15°C	

Tableau 2.3 – **Programme de PCR** Les étapes [2 ... 4] sont répétées 28 fois.Tableau 2.2: **Composition du milieu réactionnel de PCR**

Composant	Quantité (1 L)	Conc. finale
Mélange 1		
Amorce sens (20 $\mu$ M stock)	1.25 $\mu$ L	0.5 $\mu$ M
Amorce anti-sens (20 $\mu$ M stock)	1.25 $\mu$ L	0.5 $\mu$ M
dNTP (25 mM stock)	0.5 $\mu$ L	0.3 mM
Tampon Taq polymérase (10x)	5 $\mu$ L	1x
$\Delta\Delta$ H <sub>2</sub> O	40.5 $\mu$ L	—
Mélange 2		
Matrice pUC19 (2 ng/ $\mu$ L stock)	1 $\mu$ L	0.04 ng/ $\mu$ L
Taq DNA-polymérase (5 u/ mL stock)	0.25 $\mu$ L	5 mu/ mL

Les tubes de PCR sont ensuite placés dans l'appareil pour la réaction de PCR à proprement parler. L'enchaînement des cycles effectué est décrit dans le Tableau 2.3. La taille des produits de PCR obtenus est contrôlée sur gel à 2% d'agarose (dépôt sur gel de 2  $\mu$ L de plusieurs tubes pris au hasard, et de 2  $\mu$ L de chaque tube contrôle). Une fois que l'on a montré que le produit de PCR obtenu correspondait bien à celui attendu, les deux tubes contrôle ne sont plus utiles et sont éliminés. Le contenu de tous les tubes est alors assemblé dans un tube de 15 mL et conservé à  $-20^{\circ}\text{C}$  en attendant d'être purifié par chromatographie échangeuse d'anions (voir ci-dessous et § 3.1.1, page 69).

### 2.2.1.2 PURIFICATION DES PRODUITS DE PCR

L'excédent d'amorces non utilisées lors des PCR doit être éliminé. Pour cela, nous avons recours à une séparation par chromatographie sur colonne

échangeuse d'anions. La chaîne HPLC est un Äkta Purifier (GE Healthcare) ; la colonne est une colonne échangeuse d'anions Mini Q (GE Healthcare). Deux tampons sont utilisés pour cette séparation :

- \* Un tampon de charge (tampon A, page 45) ;
- \* Un tampon d'élution (tampon B, page 45).

Les acides nucléiques chargés négativement interagissent avec la colonne chargée positivement. L'élution des composés fixés sur la colonne se fait par augmentation de la force ionique du tampon (gradient croissant en tampon B). L'équivalent de 100 tubes de PCR est chargé en sept fois sur la colonne. Chaque purification est précédée et suivie d'un gradient à « blanc » afin de vérifier l'état de la colonne. L'éluat correspondant aux produits de PCR est stocké en tampon TRIS-EDTA à  $-20^{\circ}\text{C}$ .

### 2.2.2 CONSTITUTION DES PHASES CHROMATOGRAPHIQUES

Les supports chromatographiques sont des billes magnétiques recouvertes de streptavidine (« *streptavidin magnetic particles* », Roche). L'incubation des billes et des déterminants d'affinité est réalisée sur roue. Les phases chromatographiques sont constituées dans des tubes Eppendorf d'une contenance de 2 mL à fond *arrondi*. Lors des changements de tampon, les billes magnétiques sont sédimentées sur un aimant pendant 1 min. Lorsque le nouveau tampon a été ajouté, il est nécessaire d'homogénéiser la solution en réalisant de petits mouvements circulaires jusqu'à ce que le culot de billes soit entièrement dissous.

Les billes magnétiques utilisées pour la constitution des phases chromatographiques sont stockées dans un tampon de composition inconnue (10 mg/ mL). Il est donc nécessaire de les laver avant de les utiliser. Inverser un certain nombre de fois le flacon dans lequel elles sont contenues afin d'homogénéiser la suspension (ces billes sont lourdes et sédimentent vite). 35  $\mu\text{L}$  de billes sont nécessaires à la constitution de chaque phase chromatographique. Ces billes sont lavées deux fois avec 200  $\mu\text{L}$  de tampon BN 500, puis équilibrées deux fois dans 200  $\mu\text{L}$  de tampon TEN 1000 (voir page 45). Enfin, elles sont resuspendues dans 160  $\mu\text{L}$  de ce même tampon. Les oligonucléotides duplexes d'ADN biotinylés sont alors ajoutés (15 pmol pour le produit de PCR monobiotinylé et 7.5 pmol pour le produit de PCR bibiotinylé) et les tubes Eppendorf sont fixés sur une roue pour rotation sur la nuit à  $4^{\circ}\text{C}$  ( $\approx 20$  rpm).

Détail important : lorsque l'on met les tubes en rotation sur roue, si le fond du tube est conique, le volume de liquide ne coule pas lorsque le tube est inversé car il est « piégé » par capillarité dans le fond conique

### 2.3 VÉRIFICATION DE LA PHASE CHROMATOGRAPHIQUE CONTRÔLE

La double fixation des oligonucléotides biotinylés sur les billes est vérifiée par incubation pendant 2 h à 25°C de la phase chromatographique  $\diamond$  avec l'enzyme de restriction Sma I. Pour digérer les 7.5 pmol d'ADN de la phase chromatographique  $\diamond$ , on utilise 20 u d'enzyme.

### 2.4 TAMPONS POUR LA PURIFICATION DES PROTÉINES

Tableau 2.4: **Tampons pour la purification des protéines**

Composant	Quantité (1 L)	Conc. finale
Tampon TEN 1000		
voir § 2.1, page 45		
Tampon Ku 1x		
HEPES-KOH, pH 7.8	9.532 g	40mM
MgCl <sub>2</sub>	0.476 g	5 mM
KCl	4.473 g	60 mM
EDTA 500 mM (stock)	800 $\mu$ L	0.4 mM
Glycérol	34 g	37 mM
NP-40	0.1 g	0.01 %
DTT	0.771 g	0.5 mM
Tampon BN 0		
Bis-TRIS	10.462 g	50mM
EDTA 500 mM (stock)	2 mL	1 mM
Glycérol	100 g	108.6 mM
Tampon BN 500 : BN 0 +		
Acide 6-amino-caproïque	65.585 g	500 mM
Tampon pour protéines		
HEPES-KOH, pH 7.8	4.766 g	20 mM
KCl	7.455 g	100 mM
EDTA 500 mM (stock)	400 $\mu$ L	0.2 mM
Glycérol	200 g	217 mM
PMSF	87 mg	0.5 mM
DTT	77 mg	0.5 mM

## 2.5 PURIFICATION DE COMPLEXES NUCLÉOPROTÉIQUES

Les complexes nucléoprotéiques sont purifiés en deux étapes : la purification à proprement parler de protéines sur un déterminant d'affinité et ensuite le détachement des produits de purification des billes.

### 2.5.1 ÉTAPE DE PURIFICATION

La purification par chromatographie d'affinité de protéines spécifiques des cassures double-brin de l'ADN est menée à partir d'extraits nucléaires commerciaux (cellules HeLa, CilBiotech). Le tube d'extrait nucléaire est clarifié par centrifugation en chambre froide puis aliquoté par 300  $\mu\text{L}$  dans des tubes Eppendorf qui sont conservés à  $-80^\circ\text{C}$ . L'appareil « thermomixer » utilisé pour les incubations est de la marque Eppendorf.

Avant utilisation, les extraits nucléaires, décongelés dans un bain-marie à  $37^\circ\text{C}$ , sont centrifugés 20 min à 14000 rpm en chambre froide. Le surnageant est séparé du culot (débris membranaires et protéines agrégées). Il est ensuite dilué avec 120  $\mu\text{L}$  de tampon pour protéines frais (Tableau 2.4) pendant 30 min à  $30^\circ\text{C}$  et sous agitation à 300 rpm dans un appareil « thermomixer ». La concentration des extraits étant de 10 mg/mL, le volume d'extraits à employer est de 80  $\mu\text{L}$  par purification.

Les phases chromatographiques lavées sont équilibrées deux fois avec 200  $\mu\text{L}$  de tampon Ku puis reprises dans 80  $\mu\text{L}$  de ce même tampon. Les extraits nucléaires sont alors ajoutés et incubés avec les phases chromatographiques pendant 10 min à  $30^\circ\text{C}$  et sous agitation à 300 rpm. Le surnageant d'incubation est séparé des billes par sédimentation à l'aimant, puis ces dernières sont lavées quatre fois avec 200  $\mu\text{L}$  de tampon Ku afin de détacher les protéines n'interagissant pas assez fortement avec les déterminants d'affinité.

### 2.5.2 LE PHOTOCUVAGE UV

Cette étape, qui permet de détacher les complexes protéines-ADN des billes, est opérée par irradiation des phases avec un rayonnement ultraviolet ( $\lambda_{>300\text{ nm}}$ ) produit par une lampe à xénon-mercure UXM-200HO (Lot-Oriel, Palaiseau, France) dont la puissance, à  $\lambda_{365\text{ nm}}$ , est de 90 mW/cm<sup>2</sup> [Hégarat *et al.*, 2007].

Teflon Le dispositif d'irradiation est présenté à la Figure 2.1, page ci-contre. Afin d'éviter une montée en température trop importante pendant l'irradiation, le dispositif en *polytétrafluoroéthylène* est placé dans un récipient en plastique rempli de glace salée (l'ajout de sel permet de faire descendre la température vers  $-10^\circ\text{C}$ ). Le diamètre des puits est de 3.2 mm, ce qui fait que la surface du fond du puits permet aux billes de s'arranger en une monocouche. Le récipient en plastique est disposé sur un agitateur de type « Vortex », qui

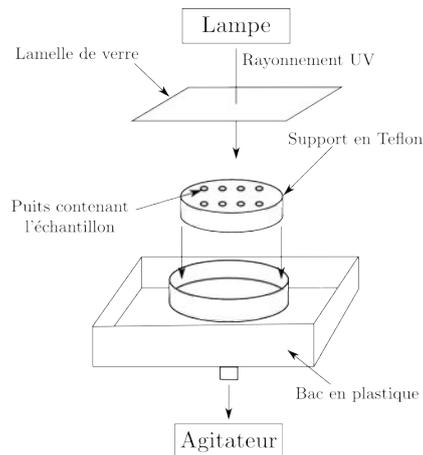


FIGURE 2.1 – **Dispositif de photoclivage** Les phases chromatographiques sont placées dans des puits creusés dans une galette épaisse de polytétrafluoroéthylène. Cette galette est enchâssée dans un bac en plastique rempli de glace salée, lui-même posé sur un agitateur de type « Vortex », qui assure une agitation continue. Le rayonnement UV est émis par une lampe à xénon-mercure située au-dessus de la galette. Une lamelle en verre est placée au dessus des puits de manière à filtrer les rayonnements trop énergétiques et à prévenir d'éventuels pontages protéine–protéine ou protéine–ADN.

permet de fournir une agitation modérée pendant l'irradiation. La lame en verre (typique lame de microscope) filtre les ultraviolets et ne laisse passer que les longueurs d'ondes les plus grandes (donc moins énergétiques). L'irradiation dure 20 min, puis les échantillons sont récupérés et les billes sont éliminées par sédimentation magnétique. Les surnageants contenant les produits de purification sont transférés dans de nouveaux tubes Eppendorf (pour la suite des expériences, les tubes de 1.5 mL sont plus pratiques).

Au cours de certaines expériences, 10 min d'irradiation ont suffi

## 2.6 SÉPARATIONS ÉLECTROPHORÉTIQUES

Plusieurs types de séparations électrophorétiques ont été utilisées tout au long de notre travail. La plus largement employée est l'électrophorèse en conditions non dénaturantes (BN PAGE) et l'autre est l'électrophorèse en conditions dénaturantes (SDS PAGE).

### 2.6.1 TAMPONS POUR LES ÉLECTROPHORÈSES ET SOLUTIONS POUR LES COLORATIONS

Tableau 2.5: **Tampons pour les électrophorèses et les colorations**

Composant	Quantité (1 L)	Conc. finale
<b>Tampon BN 500</b>		
voir § 2.1, page 45		
<b>Tampon cathode 10x</b>		
Bis TRIS HCl pH 7	31.386 g	150 mM
Tricine	89.586 g	500 mM
Bleu de Coomassie G-250 (foncé)	2 g	2.4 mM
Bleu de Coomassie G-250 (clair)	0.2 g	0.24 mM
<b>Tampon anode 10x</b>		
Bis TRIS HCl pH 7	104.62 g	500 mM
<b>Tampon de dénaturation</b>		
TRIS-HCl pH 6.8	1.514 g	12.5 mM
SDS	40 g	138.7 mM
Glycérol	200 g	2.172 M
Bleu de bromophénol	0.2 g	0.3 mM
DTT	10 g	65 mM
<b>Tampon de migration 10x</b>		
TRIS-base	37.785 g	250 mM
Glycine	144.13 g	1.92 M
SDS	10 g	34.7 mM
<b>Coloration à l'argent</b>		
Composant	Quantité (100 mL)	Conc. finale
<b>Solution de sensibilisation</b>		
Thiosulfate de sodium·(H <sub>2</sub> O) <sub>5</sub>	20 mg	806 mM
<b>Solution de coloration</b>		
(continue)		

Tableau 2.5: Tampons pour les électrophorèses et les colorations (suite)

Composant	Quantité (1 L)	Conc. finale
AgNO <sub>3</sub>	204 mg	12 mM
Solution de développement		
K <sub>2</sub> CO <sub>3</sub>	3 g	217 mM
Formaline	25 μL	6.78 μM
Thiosulfate de sodium 10 %	12.5 μL	50.4 μM
Solution d'arrêt du développement		
TRIS base	4g	330 mM
Acide acétique	5 mL	87.3 mM
<b>Coloration au bleu de Coomassie colloïdal</b>		
Composant	Quantité (2 L)	Conc. finale
Solution de coloration		
CBB G-250	0.4 g	0.020%
Sulfate d'aluminium-(14-18)-hydrate	100 g	5%
Éthanol absolu	200 mL	10%
Acide orthophosphorique	47 mL	2%
Eau bidistillée	qsp 2 L	
Solution de décoloration		
Éthanol absolu	200 mL	10%
Acide orthophosphorique	47 mL	2%
Eau bidsitillée	qsp 2 L	
(continue)		

## 2.6.2 ÉLECTROPHORÈSE EN CONDITIONS NATIVES : MIGRATION ET COLORATION DES GELS

### 2.6.2.1 LA MIGRATION ÉLECTROPHORÉTIQUE EN BN PAGE

Cette électrophorèse est menée dans une cuve particulière (XCell MiniLock de chez LifeTechnologies). Les gels utilisés sont des gels à gradient d'acrylamide (4–16 %, NativePAGE ; LifeTechnologies). Le marqueur de taille (NativeMarker ; LifeTechnologies) a été utilisé pour estimer le poids moléculaire apparent des complexes protéiques. Les deux tampons cathode (Tableau 2.5, page ci-contre) ont été soniqués pendant 15 min avant leur utilisation afin de détruire les agrégats de bleu de Coomassie qui ont tendance à se former.

Le gel est placé dans la cuve à électrophorèse et les puits sont rincés avec du tampon cathode à forte concentration en bleu de Coomassie. Cette étape permet d'éliminer les résidus d'acrylamide qui peuvent être présents

au fond des puits. Le marqueur de taille et les échantillons sont ensuite chargés sur le gel et la migration débute à 100 V pendant 1 h avec le tampon cathode concentré en bleu de Coomassie. Ce tampon est ensuite remplacé par le tampon cathode moins concentré en bleu de Coomassie et la tension est augmentée à 200 V pour la fin de la migration. La migration s'arrête lorsque le front de migration atteint l'extrémité basse du gel. Le gel est alors démoulé, décoloré pour être ensuite recoloré, soit au bleu de Coomassie colloïdal, soit à l'argent.

### 2.6.2.2 LA COLORATION ANALYTIQUE DES GELS

La coloration à visée analytique se fait après décoloration du bleu de Coomassie de migration. En effet, au terme de cette dernière, le gel est entièrement bleu et doit être décoloré afin qu'une coloration analytique puisse être faite avec succès. Un premier bain dans le fixateur 1 (30 % d'éthanol et 10 % d'acide acétique) est effectué la nuit sous agitation afin de décolorer autant que possible le gel. Deux bains de 10 min dans le fixateur 2 (20 % d'éthanol) sont ensuite effectués afin de retirer l'acide acétique du gel. Enfin, deux bains de 10 min dans l'eau bidistillée permettent de réhydrater le gel. Ce dernier est alors prêt à être coloré.

Deux colorations analytiques peuvent être effectuées en fonction de l'utilisation qui doit être faite du gel : la coloration à l'argent est ultrasensible, alors que la coloration au bleu de Coomassie colloïdal l'est un peu moins, mais, en contrepartie, est compatible avec les analyses par spectrométrie de masse.

**2.6.2.2.1 La coloration à l'argent** est effectuée selon le protocole publié par [Chevallet \*et al.\*](#), en 2006. Ce protocole est fondé sur l'utilisation de quatre solutions permettant successivement de sensibiliser les protéines à l'argent, ce qui augmentera leur propension à lier les ions  $\text{Ag}^+$  en solution, de colorer les protéines en réduisant l'ion argent en métal  $\text{Ag}^0$ , ce qui permet de l'observer et enfin de stopper la réduction afin d'avoir le meilleur contraste possible (voir Tableau 2.5, page 52).

La coloration d'un gel à l'argent est réalisée en quatre étapes, précédées d'une décoloration/fixation sur la nuit. Après chaque étape, le gel est lavé avec de l'eau bidistillée. Le volume de solution utilisé, à la fois pour les lavages et les incubations, est à chaque fois de 100 mL. Les incubations sont réalisées sous légère agitation, même lorsqu'elles sont de courte durée.

Dans un premier temps, le gel devant être coloré est placé dans 100 mL de solution de sensibilisation pendant 1 min. Après lavage, il est incubé pendant 1 h dans la solution de coloration. Après un lavage particulièrement soigné, le gel est ensuite mis au contact de la solution de développement. La durée de cette étape est laissée à l'appréciation de l'expérimentateur ; lorsque

la coloration a atteint l'intensité voulue, la solution de développement est remplacée par la solution d'arrêt du développement (fixateur). Une fois la coloration du gel stoppée, elle est stable et le gel peut donc être conservé durablement.

**2.6.2.2.2 La coloration au bleu de Coomassie colloïdal** a été mise en œuvre selon le protocole décrit par [Dyballa et Metzger](#), en 2012. Bien que ce protocole ait été développé afin de permettre la coloration d'un gel de polyacrylamide directement après la migration électrophorétique (sans avoir à réaliser d'étape de fixation), nous avons également eu recours à l'utilisation des fixateurs 1 (30 % d'éthanol et 10 % d'acide acétique) et fixateur 2 (20 % d'éthanol) préalablement à la coloration, afin de permettre la décoloration du gel.

Le gel décoloré et réhydraté est placé dans la solution de coloration. Les bandes correspondant à la présence de protéines sont visibles dès 1 h d'incubation et l'intensité maximale est atteinte après 6 h. Le contraste peut être amélioré en incubant le gel dans la solution de décoloration. La durée de cette étape est typiquement de 15 min à température ambiante, mais il est recommandé de contrôler régulièrement l'état de coloration du gel.

### 2.6.3 ÉLECTROPHORÈSE EN CONDITIONS DÉNATURANTES : MIGRATION ET COLORATION DES GELS

Après une électrophorèse en conditions natives, il peut être d'intérêt de vérifier si telle ou telle bande dans un gel natif comporte ou non des assemblages multiprotéiques. Pour cela, il faut effectuer une deuxième électrophorèse avec la piste du gel natif.

#### 2.6.3.1 LA MIGRATION ÉLECTROPHORÉTIQUE DÉNATURANTE

Ce type de migration électrophorétique est effectué dans une cuve MiniVE (Hofer) avec un gel SDS-PAGE préparé par nos soins. Le marqueur de taille « *Perfect protein marker* » (Millipore) a été utilisé pour déterminer le poids moléculaire apparent des protéines.

Une piste de gel natif est découpée et incubée dans 15 mL de tampon de dénaturation (Tableau 2.5, page 52) pendant 10 min sous légère agitation. Le tube est chauffé au four à micro-ondes pendant 10 secondes afin de faciliter la dénaturation des complexes protéiques dans le gel. La piste de gel natif est laissée à refroidir dans le tube pendant 20 min, à température ambiante et sous légère agitation. La bande de gel correspondant à la piste de gel natif est déposée horizontalement sur un gel SDS-PAGE (Tableau 2.5) coulé sans puits, de telle sorte que le bas du gel natif soit situé à gauche (voir § 3.2.1.1, page 81 et Figure 3.2.1.1). La migration électrophorétique

est effectuée pendant 1 h à 100 V puis à 200 V jusqu'à ce que le front de migration atteigne le bas du gel.

### 2.6.3.2 LA COLORATION AU SYPRO

Cette coloration est effectuée avec le réactif SYPRO Ruby (Life Technologies), qui marque les protéines avec un réactif fluorescent détectable ensuite au moyen de l'appareil Typhoon (GE Healthcare). Le gel est incubé pendant 45 min dans le tampon de fixation (ci-dessous). Ce tampon est ensuite remplacé par le SYPRO Ruby, dans lequel le gel est incubé sur la nuit, sous agitation et dans le noir. Le colorant est alors retiré et le gel est de nouveau incubé dans le tampon de fixation, pour 1 h. Le gel est ensuite numérisé avec un laser qui permet de révéler les protéines. Composition du tampon de fixation pour la coloration SYPRO :

Éthanol 10 %	50 mL
Acide acétique 7.5 %	37.5 mL
Eau bidistillée	qsp 500 mL

## 2.7 EXPÉRIENCE D'EMPREINTE AUX NUCLÉASES

Les oligonucléotides devant permettre la constitution des phases sont marqués avec de l'ATP  $\gamma$   $^{32}$ [P] (Perkin-Elmer) par action de la polynucléotide kinase (PNK, New England Biolabs) pendant 30 min à 37°C.

Les produits de purification obtenus avec les oligonucléotides radiomarqués sont digérés, après détachement des billes, à 37°C pendant 1 h sous légère agitation, en présence de 1.5 u de DNase I (Invitrogen) et de 1 u de benzonase (Merck). Les produits de cette digestion sont ensuite soumis à une électrophorèse BN PAGE ordinaire. Pour effectuer une autoradiographie, le gel (qui peut avoir été coloré au bleu de Coomassie colloïdal) est couché sur une feuille de papier absorbant 3MM (Whatman). Il est ensuite recouvert d'un film en plastique étanche (Saran wrap) et séché à 80°C plaqué sous un tapis en gomme (sécheur à gels sous vide). Il est ensuite exposé contre une cassette pour 1 h et celle-ci est révélée par lecture avec l'appareil Typhoon. L'autoradiographie permet alors d'établir la position de l'ADN radiomarqué dans les pistes du gel.

## 2.8 SÉQUENÇAGE DE L'ADN DES COMPLEXES

Après avoir résolu les complexes nucléoprotéiques après digestion par les nucléases, les bandes révélées par le bleu de Coomassie colloïdal (éventuellement correspondant à un signal d'autoradiographie) sont découpées et réduites en purée. L'ADN qui s'y trouve contenu est extrait par des lavages

multiplés à l'eau pure. L'eau d'extraction est alors séparée des microbilles de polyacrylamide (très petites) par centrifugation dans une colonne Spin-X (« *centrifuge tube filters* ») (Costar). L'ADN en solution dans l'eau d'extraction est dénaturé à 90°C pendant 5 min, puis soumis à une migration électrophorétique en gel de séquence (gel urée-acrylamide 12% à chaud). Un oligonucléotide radiomarqué de 186 pb, séquencé selon la méthode de Maxam et Gilbert, est également déposé sur le gel afin de déterminer la taille des fragments d'oligonucléotides provenant de la digestion des produits de purification par les nucléases.

## 2.9 ACTIVITÉ DE LIGATION DES COMPLEXES PURIFIÉS

Les expériences ont été réalisées en utilisant des déterminants d'affinité produits par digestion des oligonucléotides biotinylés avec la nucléase Pst I (New England Biolabs). Cet enzyme, qui ne coupe qu'à un endroit les oligonucléotides, génère deux fragments de 78 pb et de 108 pb à extrémités 5' cohésives de 4 pb.

L'étape de ligation pour les différentes expériences décrites ci-dessous a été réalisée pendant 2 h à 37°C sous agitation douce. Au terme de l'étape de ligation, les complexes nucléoprotéiques sont déprotéinisés par action de la protéinase K à 55°C pendant 1 h en tampon dénaturant (1 mg/mL de protéinase K, 0.5% de SDS, 10 mM EDTA dans du TRIS-HCl 20mM pH 7.5. Une extraction au au phénol/chloroforme/alcool isoamylique permet d'éliminer les résidus protéiques et les acides nucléiques sont précipités dans 10 volumes d'éthanol absolu en présence de 0.3 M d'acétate de sodium et de traces de glycogène, avant d'être analysés sur un gel à 2% d'agarose imprégné de bromure d'éthidium.

Mélange appelé SEVAG par les biologistes moléculaires

### 2.9.1 LIGATION PAR LES EXTRAITS NUCLÉAIRES BRUTS

Les extraits nucléaires totaux (320 µg) provenant de cellules HeLa en culture ont été utilisés pour liguer 15 pmol d'oligonucléotides biotinylés digérés par Pst I pendant 2 h à 30°C dans un tampon de ligation complet (50 mM de TRIS-HCl pH 7.5, 12 mM de HEPES pH 7.8, 10 mM de MgCl<sub>2</sub>, 1 mM d'ATP, 10 mM de DTT, 12 mM de KCl, 0.024 mM d'EDTA, 2.6% de PEG 8000, 5% de glycérol, 0.06 mM de PMSF). Cette expérience a été effectuée dans un volume de 60 µL.

### 2.9.2 LIGATION AVEC DES COMPLEXES PURIFIÉS

Les complexes purifiés ont été obtenus en utilisant 320 µg d'extraits nucléaires sur 350 µg de billes magnétiques fonctionnalisées avec 7.5 pmol

d'oligonucléotides biotinylés digérés par Pst I. Une même quantité d'oligonucléotides digérés a été ajoutée lors de la ligation dans un volume final de 35  $\mu\text{L}$  de tampon BN 0, supplémenté avec un mélange de ligation comme décrit ci-dessus. Deux cas se présentent : selon que la ligation est menée sur des complexes nucléoprotéiques purifiés encore attachés aux billes ou préalablement détachés.

### 2.9.2.1 LIGATION AVEC DES COMPLEXES PURIFIÉS ATTACHÉS AUX BILLES

La mise en place de l'expérience est comme décrit ci-dessus. La ligation (ajout de l'oligonucléotide exogène et du tampon ligase) a lieu alors que les complexes sont encore liés aux billes. Après réaction, les complexes nucléoprotéiques sont lavés une fois avec le tampon stringent BN 0, puis les phases sont irradiées dans 35  $\mu\text{L}$ , dans ce même tampon et pour 20 min. Les complexes libérés dans le surnageant sont soumis à une protéolyse (protéinase K) et les oligonucléotides sont extraits au SEVAG. La taille des oligonucléotides est déterminée sur gel d'agarose. Il faut noter que le surnageant de lavage, obtenu juste après la ligation, a subi le même traitement et a également été analysé sur gel d'agarose.

### 2.9.2.2 LIGATION EN L'ABSENCE DE SUPPORTS CHROMATOGRAPHIQUES

Cette réaction de ligation est comparable à la précédente à ceci près que les complexes nucléoprotéiques sont d'abord détachés des billes. Après élimination de ces dernières, le matériel purifié est additionné d'oligonucléotide exogène et de tampon de ligation.

## 2.10 SPECTROMÉTRIE DE MASSE

Les bandes de gel non dénaturant ont été excisées du gel et soumises à un traitement protéolytique, dans le gel, à la trypsine selon un protocole standard. Les analyses par spectrométrie de masse ont été effectuées après séparation des peptides sur une chaîne HPLC Ultimate 3000 Rapid Separation (Dionex). Les peptides élués ont été analysés dans un spectromètre de masse LTQ-ORBITRAP « Velos » (Thermo Fisher Scientific).

Les peptides issus de la protéolyse ont été acidifiés avant chargement sur une pré-colonne de phase inversée  $\text{C}_{18}$  (taille des billes : 3  $\mu\text{m}$ , taille des pores : 100 Å, diamètre interne de la colonne : 75  $\mu\text{m}$ , longueur de la colonne : 2 cm) avec le tampon de chargement (98 % eau, 2 % acétonitrile et 0.1 % d'acide trifluoroacétique). Le débit était de 5  $\mu\text{L}/\text{min}$ . Les peptides ont été séparés sur une colonne analytique de phase inversée  $\text{C}_{18}$  (taille des billes : 2  $\mu\text{m}$ , taille des pores : 100 Å, diamètre interne de la colonne : 75  $\mu\text{m}$ , longueur de la colonne : 15 cm). Le gradient chromatographique a été

developpé comme suit : de 99 % de tampon A (5 % d'acétonitrile dans 0.1 % d'acide formique dans l'eau) à 40 % de tampon B (80 % d'acétonitrile dans 0.085 % d'acide formique dans l'eau) en 45 min.

Le spectromètre de masse couplant un piège à ions quadrupolaire linéaire à une cellule ORBITRAP faisait l'acquisition des données de masse tout au long de l'élution chromatographique et fonctionnait en mode dépendance des données ((« *data-dependent acquisition* ») selon le schéma suivant : des spectres de masse étaient enregistrés sur toute la gamme de  $m/z$  et étaient systématiquement suivis de l'acquisition de spectres SM/SM (fragmentation induite par des collisions, « *CID* ») des dix ions les plus intenses. La latence d'exclusion au cours de laquelle un ion de même rapport  $m/z$  ne pouvait être fragmenté était de 24 s. Seuls des précurseurs de charges 2, 3, ou 4 étaient susceptibles d'être fragmentés.

Les données brutes (fichiers *raw*) enregistrées par le spectromètre de masse ont été converties au format *XML* avec le logiciel *msconvert* de la suite logicielle *Proteowizard* tournant sous *MS-Windows* [Chambers *et al.*, 2012]. Le format de fichier en sortie est *mzXML*, avec le mode de compatibilité TPP activé et la précision de codage des données numériques réglé à 64 bits. Toutes les étapes successives de traitement des données ont été menées sur un système *Debian GNU/Linux* ([debian.org](http://debian.org)) en faisant usage des logiciels libres suivants : *X!Tandem*, comme moteur d'identification de protéines, qui fait usage des données de fragmentation en tandem des peptides analysés [Bowden *et al.*, 2009] en conjonction avec la base de données de protéines SwissProt ([www.expasy.ch](http://www.expasy.ch)) restreinte aux protéines humaines ; le programme *X!TandemPipeline*, qui sert d'interface utilisateur avec *X!Tandem* et qui fournit des fonctionnalités utiles, tant pour filtrer que pour grouper les données d'identification fournies par *X!Tandem* ([pappso.inra.fr/bioinfo/xtandempipeline/](http://pappso.inra.fr/bioinfo/xtandempipeline/)) ; la base de données relationnelle *Sqlite3* ([www.sqlite.org](http://www.sqlite.org)) et des logiciels développés au laboratoire. En effet, deux logiciels ont été développés spécifiquement pour les besoins de notre projet, tous deux en langage C++. Le premier logiciel a pour tâche de lire le fichier d'identification de protéines tel qu'il est produit par *X!TandemPipeline* et d'injecter les données dans une base de données *Sqlite3*. Le second logiciel — *dsbRepair* — a été développé pour permettre la fouille des données bio-structurales qui peuplent la base de données. Le manuel pour le deuxième logiciel est disponible en annexe [voir § 5.0.4, page 195, intitulée *Manuel de l'utilisateur du logiciel dsbRepair (anglais)*].



# CHAPITRE 3

## RÉSULTATS



**Position du travail.** Nous avons vu dans l'introduction que les cellules subissent continuellement des dommages au niveau de leur ADN. Ces dommages peuvent être causés par de nombreux facteurs exogènes, comme les contaminants génotoxiques, les rayonnements ionisants et certaines molécules dites « poisons » d'activités enzymatiques (comme ceux des topoisomérases par exemple) [Hoshi *et al.*, 2014]. D'autres facteurs, endogènes ceux-là, peuvent également produire des lésions sur l'ADN, comme celles dues à des enzymes dysfonctionnels ou à des espèces réactives de l'oxygène. La très grande diversité des dommages que peut subir l'ADN a déterminé la mise en place, au cours de l'évolution, de voies de réparation spécifiques. Ces voies sont matérialisées dans les cellules par des « boîtes à outils », en quelque sorte, qui sont spécifiquement adaptées à tel ou tel type de dommage (cassures double-brin, méthylations de base, pontages inter-brins, dimères de thymine, par exemple). Les cassures double-brin de l'ADN sont les lésions les plus dangereuses pour les cellules, car elles peuvent conduire — si elles ne sont pas réparées à temps — à des aberrations chromosomiques pouvant se traduire, à terme, par l'apparition du cancer.

De manière intéressante, le caractère dangereux des cassures double-brin de l'ADN peut être exploité à des fins thérapeutiques, comme nous le discuterons plus loin. En effet, en thérapie anticancéreuse, il est souvent nécessaire de provoquer ces cassures dans les tissus tumoraux (on peut aussi utiliser des molécules qui les induisent, comme la bléomycine) afin d'induire l'apoptose et donc la régression tumorale.

Dans notre travail de thèse, nous nous sommes intéressés plus particulièrement à la réparation des cassures double-brin, en particulier du point de vue de la composition en protéines des « boîtes à outils » évoquées plus haut, ainsi que du point de vue de la dynamique d'assemblage des complexes de réparation.

La méthode la plus couramment employée pour l'étude biochimique des complexes de réparation est sans conteste la purification de ces complexes par chromatographie d'affinité suivie de l'identification des protéines purifiées. D'une manière très générale, ce type de chromatographie est mis en œuvre avec une phase chromatographique constituée de deux entités :

- \* Un support chromatographique, typiquement des billes de sépharose, de polyacrylamide ou de tout autre type de polymère, parfois recouvert d'une couche de matériau ferro-magnétique ;
- \* Un déterminant d'affinité choisi spécifiquement en fonction des protéines que l'on souhaite purifier. Ce déterminant d'affinité est souvent fixé au support par des interactions covalentes. Cependant, depuis la généralisation de l'utilisation du couple biotine–streptavidine, nombre de phases chromatographiques sont constituées par couplage du dé-

terminant d'affinité, préalablement lié à un résidu de biotine, avec le support chromatographique recouvert de streptavidine [Chodosh et Buratowski, 2001; Srisawat et Engelke, 2002].

Dans le cas d'une purification par affinité de protéines de réparation des cassures double-brin de l'ADN, le déterminant d'affinité sera typiquement un oligonucléotide duplexe d'ADN présentant une extrémité libre destinée à mimer une cassure double-brin et sur laquelle les protéines d'intérêt devront se fixer. Après le lavage de la phase, l'élution des protéines, qui peut mettre en œuvre plusieurs types de facteurs physico-chimiques, comme la température, la force ionique, le pH ou encore des détergents, par exemple, est censée détacher des oligonucléotides uniquement les protéines fixées spécifiquement sur leurs extrémités. Ce type d'expérience, qui a produit un grand nombre de résultats majeurs [Qi *et al.*, 2008; Samuel *et al.*, 2005], souffre, de notre point de vue, de quelques défauts :

- \* La phase chromatographique, constituée — par définition — de deux entités, le support chromatographique et le déterminant d'affinité, se comporte comme une entité unique tout au long de la purification. Les billes, dont la surface d'interaction avec le milieu est plus grande que celle des extrémités libres de l'ADN — de plusieurs ordres de grandeur — lieront de manière non spécifique une quantité considérable de protéines. Au moment de l'élution, et cela quel que soit l'agent d'élution utilisé et quel que soit le protocole de lavage, bon nombre des ces protéines seront relâchées, qui contamineront l'éluat ;
- \* L'oligonucléotide duplexe d'ADN, qui forme le déterminant d'affinité reconnu spécifiquement par les protéines d'intérêt, en ce qu'il présente une extrémité libre mimant une cassure double-brin, est pourtant reconnu sur sa séquence « interne » par des protéines présentant une affinité générique pour la double hélice d'ADN. Certaines de ces protéines « contaminantes » seront inmanquablement éluées avec les protéines d'intérêt.

Les deux inconvénients ci-dessus provoquent une diminution drastique du facteur d'enrichissement des protéines d'intérêt, car ils sont tous deux à l'origine de la contamination du produit de purification final. Dans le cas des protéines liées non spécifiquement aux billes, l'importance de la contamination du produit de purification est si grande qu'elle le complexifie au point qu'une analyse par gel de polyacrylamide (1D ou 2D) ne pourra donner de signal utile (présence d'une *traînée* pour les gels 1D et de trop nombreuses taches pour les gels 2D ; Hégarat *et al.* [2007]). Pour contourner cette difficulté, les auteurs ont élaboré des stratégies analytiques fondées sur l'emploi d'anticorps (« *western blot* ») pour analyser le contenu des

En anglais, « *smear* »

fractions chromatographiques et afin d'établir la présence ou l'absence de telle ou telle protéine dans le mélange complexe obtenu [Samuel *et al.*, 2005]. L'inconvénient majeur de cette stratégie analytique réside dans le fait que de disposer d'un anticorps suppose la connaissance de la protéine recherchée. Or, nous voulions, précisément, pouvoir identifier des protéines non encore connues pour être impliquées dans la réparation des cassures double-brin de l'ADN et l'emploi d'anticorps implique un *a priori* dans la démarche de recherche d'une protéine donnée.

**Objectifs du travail.** Au cours de ce travail de thèse, nous souhaitions élaborer une stratégie analytique qui permît de limiter l'impact des deux modes de contamination du produit de purification décrits ci-dessus. Cette stratégie analytique est décrite à la Figure 3.1, page suivante. La voie A de cette figure illustre la stratégie destinée à contrer la contamination massive du produit de purification par les protéines se liant de manière non spécifique aux billes. Il s'agit — après la purification — de se départir des billes grâce à un système de lien photoclivable particulièrement adapté aux stratégies de purification fondées sur des interactions avec des acides nucléiques [Martinez *et al.*, 2002; Olejnik *et al.*, 1996]. La voie B illustre la stratégie destinée à pallier la difficulté due à la liaison, à l'oligonucléotide de la phase chromatographique, de protéines présentant une affinité générique pour la double-hélice d'ADN. Il s'agit d'employer un système chromatographique « contrôle », avec l'emploi d'une phase qui n'exposerait au milieu environnant qu'un oligonucléotide duplexe d'ADN *sans* extrémité libre disponible.

La mise en œuvre d'un système chromatographique intégré, qui constitue le noyau de notre stratégie analytique, devait nous permettre de purifier avec un facteur d'enrichissement inégalé des complexes de réparation de l'ADN. L'utilisation ultérieure d'une technique de séparation électrophorétique en conditions natives devait nous permettre, pour la première fois, de visualiser individuellement différents complexes de réparation purifiés. Enfin, les protéines appartenant aux différents complexes de réparation ainsi purifiés devaient être identifiées par spectrométrie de masse, après digestion dans le gel. Une quantification de ces protéines a été effectuée sur la base d'une grandeur dite « emPAI », décrite plus loin.

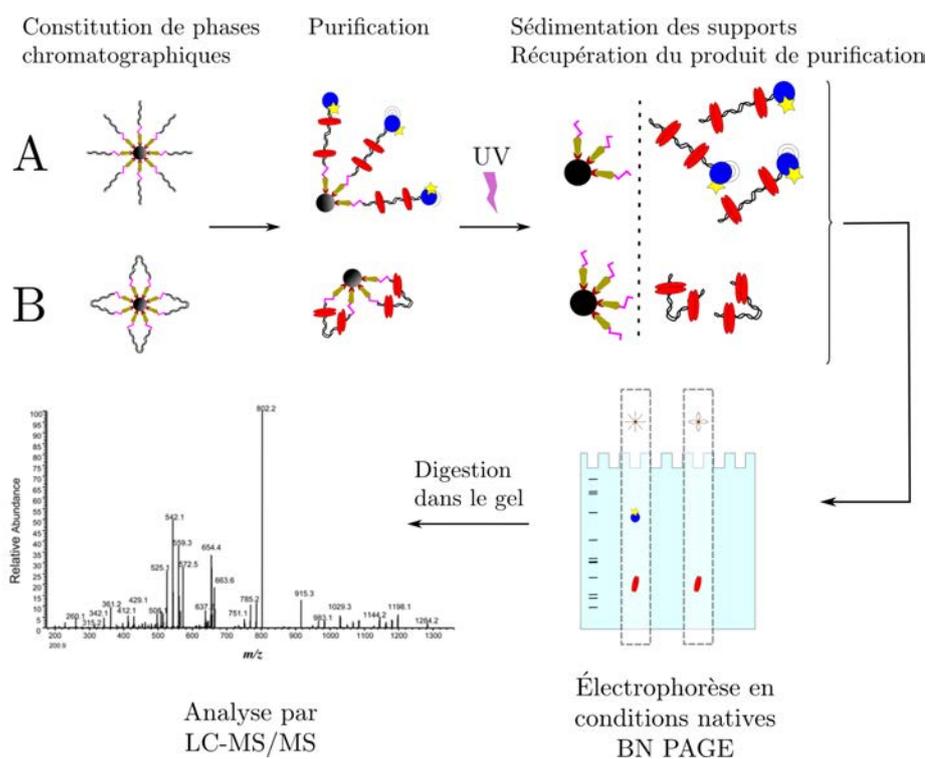


FIGURE 3.1 – **Stratégie analytique mise en œuvre.** L'utilisation de phases chromatographiques présentant des extrémités libres de l'ADN (voie A) ou n'en présentant pas (voie B) permet la purification et l'identification de protéines spécifiques des cassures double-brin de l'ADN en cinq étapes.

### 3.1 UN SYSTÈME CHROMATOGRAPHIQUE INTÉGRÉ

Le système chromatographique sur lequel repose notre méthodologie consiste en l'utilisation de deux phases chromatographiques distinctes, dont les structures sont présentées dans la Figure 3.2. Ces phases sont constituées, d'une part, de billes magnétiques recouvertes de streptavidine, servant de *support chromatographique* et, d'autre part, d'oligonucléotides duplexes d'ADN, d'une longueur de 186 paires de bases, qui constituent les *déterminants d'affinité* de nos phases chromatographiques. Olejnik *et al.* ont décrit en 1996 la synthèse d'un groupement chimique sensible aux rayonnements ultraviolets et qui se scinde sous l'action de ces derniers (Figure 3.3, page suivante). La rupture de ce lien « photoclivable » génère en 5' de la séquence nucléotidique libérée le groupement 5' phosphate canonique et permet de détacher du support chromatographique le produit de purification [déterminant d'affinité + protéines d'intérêt].

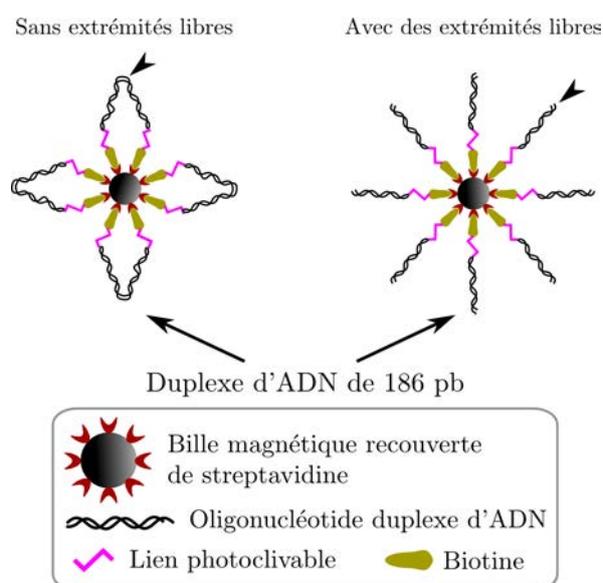


FIGURE 3.2 – **Structure des phases chromatographiques employées.** La phase contrôle (phase de gauche) ne présente pas d'extrémité d'ADN libre. La phase de droite est dotée d'extrémités libres de l'ADN. Un lien photoclivable, placé entre la biotine et la séquence oligonucléotidique, permet de récupérer les produits de purification.

L'utilisation du lien photoclivable décrit ci-dessus a été inspirée par les résultats obtenus par notre équipe lors de précédents travaux, portant sur la purification de la protéine TetR par chromatographie d'affinité. La protéine TetR est impliquée dans la régulation de la transcription du gène de résistance à la tétracycline [Kisker *et al.*, 1995], qui code pour un antibiotique

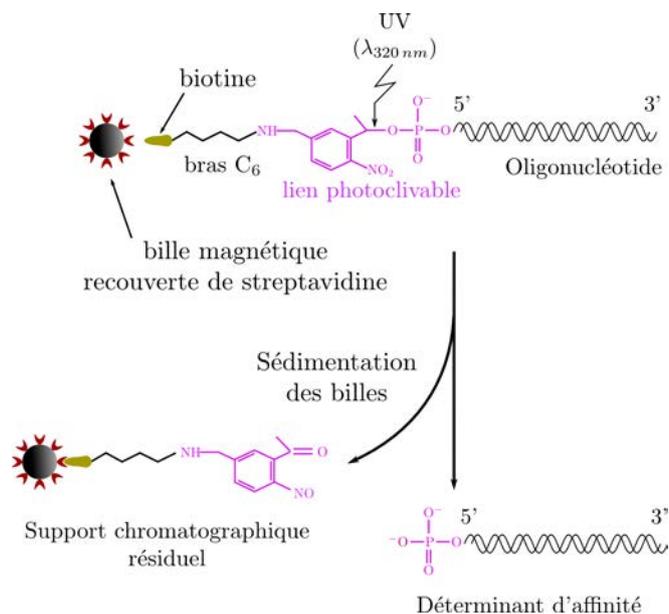


FIGURE 3.3 – **Groupe photoclivable.** Le groupement chimique illustré en violet est sensible aux ultraviolets (UV). Sous l'action d'un rayonnement UV, ce groupement se scinde pour libérer l'oligonucléotide avec son extrémité 5'-phosphate canonique [Olejnik *et al.*, 1996].

inhibant la synthèse des protéines bactériennes *via* des interactions avec la sous-unité 30S des ribosomes. La protéine TetR reconnaît le promoteur de son propre gène, qu'elle régule positivement en présence de tétracycline, y compris lorsque cet antibiotique est présent en faible quantité. La phase chromatographique utilisée pour la purification de cette protéine était constituée de billes magnétiques recouvertes de streptavidine et fonctionnalisées avec des oligonucléotides biotinylés comportant la séquence nucléotidique TetO promotrice reconnue par TetR. Lors de ces travaux, le problème de la contamination du produit de purification par les protéines liées non spécifiquement au support chromatographique avait été soulevé. Ce problème avait été résolu par l'insertion du groupement photoclivable décrit ci-dessus, entre le déterminant d'affinité et la biotine. La Figure 3.4, page ci-contre montre la différence d'enrichissement que l'on obtient lorsque l'on pratique l'analyse des produits de purification selon deux modes différents :

- \* Au terme de la purification, l'élution des protéines retenues sur l'oligonucléotide comportant la séquence promotrice TetO est effectuée par l'emploi de détergent (gel 2D de gauche) ;
- \* Au terme de la purification, la phase chromatographique est irradiée

avec des UV ( $\lambda_{>300 \text{ nm}}$ ) de manière à détacher des billes l'oligonucléotide et les protéines qui s'y étaient liées.

On constate que le fait de détacher le complexe protéines–oligonucléotide des billes permet de s'affranchir d'une grande quantité de protéines contaminantes liées non spécifiquement aux billes. Ainsi, le gel 2D de droite ne comporte que deux protéines, l'actine (protéine très abondante dont il est parfois difficile de se départir) et la protéine TetR. Ce résultat montre que l'utilisation du lien photoclivable publié par [Olejnik \*et al.\*](#) permet de limiter fortement la contamination du produit de purification.

Dans l'optique de notre projet, dont l'objectif était de réussir à identifier des protéines interagissant spécifiquement avec les extrémités libres de l'ADN, ce type de système était absolument souhaitable. Nous avons donc choisi de l'utiliser pour la constitution de nos phases, afin de pouvoir libérer le produit de purification du support chromatographique sans avoir à utiliser d'agents d'éluion, dont nous avons discuté plus haut des effets indésirables sur la pureté finale de l'échantillon (voir page 64).

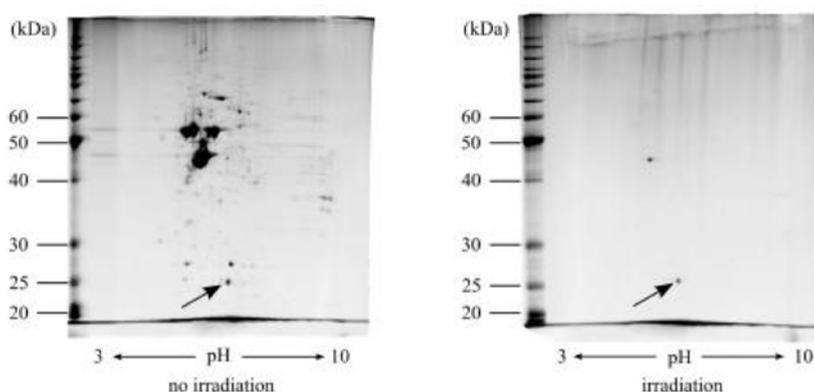


FIGURE 3.4 – **Utilité d'un lien photoclivable entre l'oligonucléotide et les billes.** Gel 2D de gauche : la grande quantité de protéines liées non spécifiquement au support chromatographique est relâchée au moment de l'éluion des protéines d'intérêt liées à l'oligonucléotide. Gel 2D de droite : les complexes protéines–oligonucléotide sont détachés des billes avant l'analyse par électrophorèse 2D. On constate qu'il y a uniquement deux taches sur le gel : l'actine et la protéine d'intérêt (flèche). Tiré de [Hégarat \*et al.\*](#) [2007].

### 3.1.1 PRODUCTION DES DÉTERMINANTS D'AFFINITÉ PAR PCR

Afin de constituer nos phases chromatographiques, nous avons réalisé au laboratoire la synthèse des oligonucléotides duplexes d'ADN servant de

déterminants d'affinité. Nous avons, pour cela, utilisé la technique d'amplification en chaîne par polymérase qui permet la production d'une grande quantité d'ADN. Les amorces sens (représentée en bleu sur la Figure 3.5, page ci-contre) et antisens (représentée en vert sur la même figure) s'hybrident de part et d'autre de la cassette de clonage multiple (panneau A). Cette cassette de multiclonage contient des sites uniques de coupure pour plusieurs enzymes de restriction, parmi lesquels on peut mentionner Sma I (Figure 3.5, panneau B, schéma 3) et Pst I, dont l'utilisation sera décrite ultérieurement. Au cours de l'amplification, le fragment d'ADN produit a une taille de 186 pb. Selon que les amorces portent ou non une biotine, les amplicons seront — ou non — capables de se lier aux billes recouvertes de streptavidine. Si les deux amorces sont munies d'une biotine, le produit de PCR sera bibiotinylé et se liera aux billes par ses deux extrémités (Figure 3.5, panneau B, schéma 1). Si, en revanche, seule l'amorce sens est munie d'une biotine, le produit de PCR sera monobiotinylé et ne pourra se lier aux billes que par une extrémité, laissant libre l'autre extrémité (schéma 2). Ainsi, il est possible de préparer les deux phases décrites à la Figure 3.2, page 67.

### 3.1.2 PURIFICATION DES PRODUITS DE PCR

Au terme de la PCR, il est possible que des amorces non utilisées soient encore présentes dans le milieu réactionnel. Leur présence, lors de la fixation des produits de PCR sur les billes recouvertes de streptavidine, serait problématique car les amorces biotinylées feraient d'excellents compétiteurs pour nos oligonucléotides duplexes d'ADN. Par ailleurs, bien que ces amorces soient courtes (20 paires de bases environ), elles pourraient également recruter des protéines spécifiques de l'ADN simple-brin, ce qui n'est en aucun cas souhaitable. Pour ces raisons, les amorces résiduelles dans le produit de PCR ont été éliminées par séparation chromatographique sur une phase échangeuse d'anions (phase dont la résine est chargée positivement et qui est donc capable de retenir des composés chargés négativement, comme les acides nucléiques).

Un chromatogramme typique de purification du produit de PCR (pour une centaine de PCR regroupées) est illustré à la Figure 3.6, page 72. Dans un premier temps, l'ensemble des acides nucléiques du milieu de PCR se fixent à la résine échangeuse d'anions « *chargement de l'échantillon sur la colonne* ». Nous avons ensuite développé un « *gradient croissant en NaCl* » par paliers. Les amorces, de petite taille, sont éluées très tôt dans le gradient alors que le produit de PCR est élué pour une concentration en NaCl d'environ 650 mM. Le plasmide matrice (pUC19) est élué sous forme d'un épaulement du pic contenant les produits de PCR, ce qui n'est pas un inconvénient majeur, car il n'a pas de biotine et ne sera donc pas retenu sur les billes (*vide infra*). Le rapport des aires  $\lambda_{260 \text{ nm}}/\lambda_{280 \text{ nm}}$  du pic correspondant au produit de

En anglais, « *polymerase chain reaction* », PCR.

En anglais, « *multicloning site* », MCS.

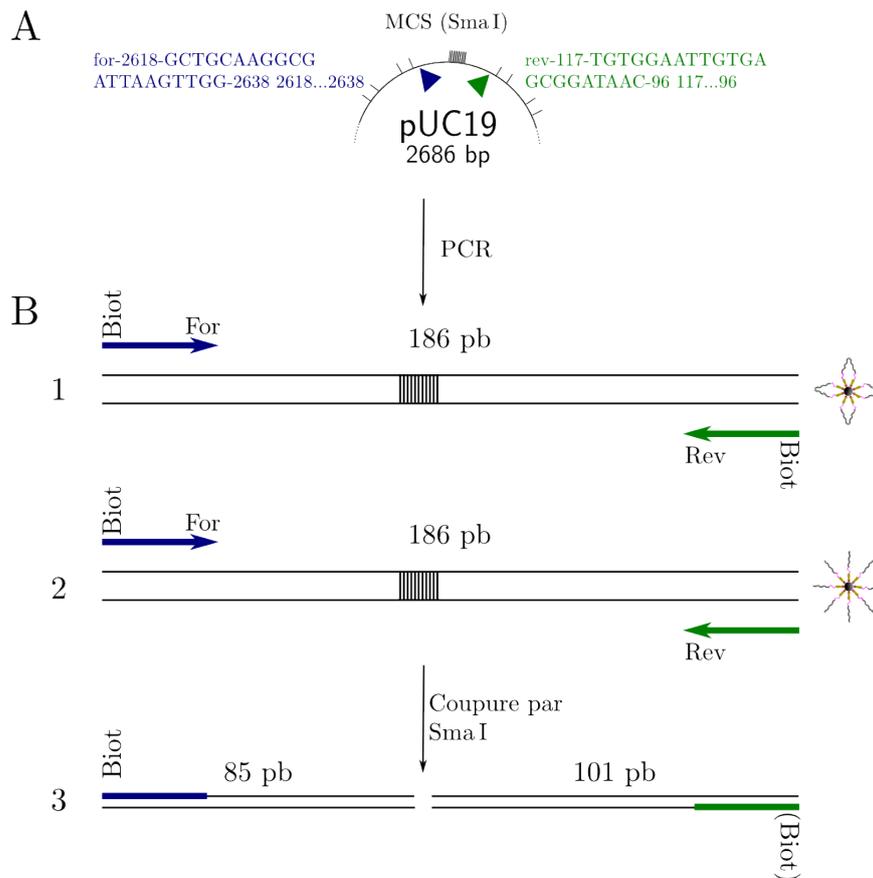


FIGURE 3.5 – **Production des oligonucléotides fonctionnalisant les billes.** Panneau A : les amorces de PCR s’hybrident de part et d’autre de la cassette de multiclonage du plasmide pUC19. Panneau B : différentes combinaisons d’amorces sens (« For ») et antisens (« Rev ») biotinylées ou non en vue de produire des oligonucléotides bi- ou monobiotinylés (voies 1 et 2, respectivement). Le site de clonage multiple comporte un site de coupure pour Sma I, produisant deux fragments de 85 pb et de 101 pb (schéma 3).

PCR est de l’ordre de 2. Ce rapport nous renseigne sur le fait que l’ADN est très fortement enrichi ( $\lambda_{260 \text{ nm}}$ ) et que la composante protéique ( $\lambda_{280 \text{ nm}}$ ) est faible.

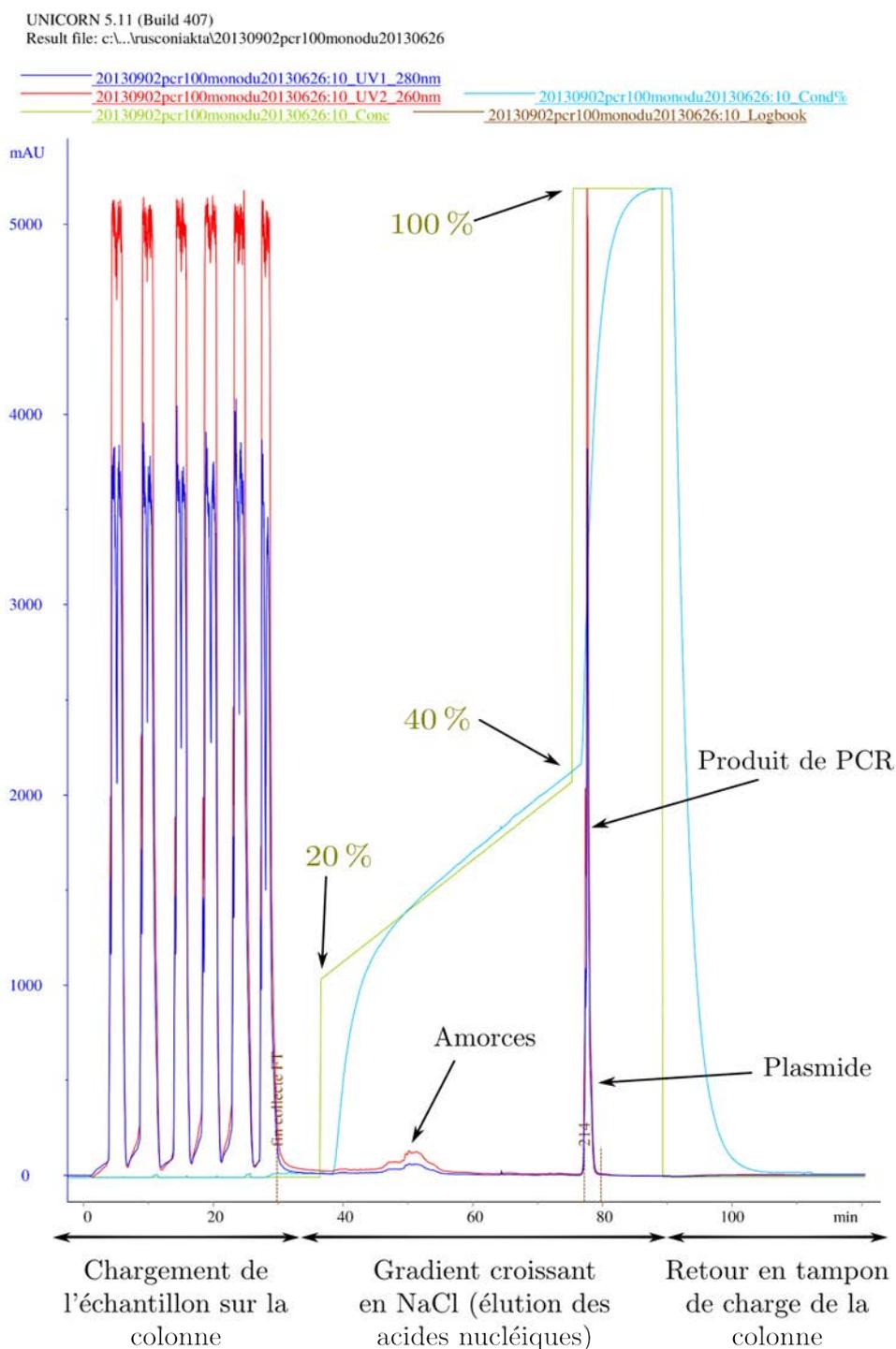


FIGURE 3.6 – **Élimination des amorces de PCR résiduelles par chromatographie.** Les amorces résiduelles sont éliminées par séparation chromatographique sur colonne échangeuse d'anions. Traces : en rouge, DO  $\lambda_{260}$  nm ; en bleu,  $\lambda_{280}$  nm ; en vert, consigne sur le % B ; cyan, conductimétrie. L'élué du plasmide se traduit par un léger épaulement à la droite du pic correspondant à l'élué du produit de PCR. Tampon A : TRIS-HCl 20 mM, EDTA 2 mM, NaCl 200 mM, pH 7 ; tampon B : comme pour le tampon A mais avec 1 M NaCl.

### 3.1.3 CONSTITUTION DES PHASES CHROMATOGRAPHIQUES

Les phases chromatographiques sont constituées par incubation, dans du tampon TEN 100 (page 45), des billes magnétiques recouvertes de streptavidine et d'oligonucléotides duplexes d'ADN tels que produits par PCR et purifiés par chromatographie (sections précédentes). Cette incubation débute la veille de l'expérience et dure une nuit à 4°C sous rotation. Après cette incubation, les phases chromatographiques sont lavées deux fois avec du tampon TEN 1000 afin d'éliminer tous les oligonucléotides qui se seraient liés aux billes de manière non spécifique, c'est-à-dire non *via* la biotine. Les phases chromatographiques sont alors prêtes pour la purification.

### 3.1.4 CONTRÔLE DE LA QUALITÉ DES PHASES PRODUITES

L'obtention de deux phases chromatographiques distinctes, l'une dotée d'extrémités libres d'ADN double-brin et l'autre en étant dépourvue, est cruciale pour la suite de notre démarche analytique. Nous avons testé la qualité de nos phases sur la base de deux critères :

- \* la fixation correcte des oligonucléotides sur les billes magnétiques lors de la constitution des phases  $\leftarrow$  et  $\rightarrow$ ;
- \* l'absence d'extrémité libre de l'ADN sur la phase chromatographique dite « contrôle » ( $\leftarrow$ ).

La Figure 3.7, page suivante illustre l'expérience de contrôle de la qualité des phases que nous produisons. Cette expérience est menée après l'incubation sur la nuit des billes avec l'oligonucléotide choisi (mono ou bibiotinylé). Les billes sont sédimentées et le surnageant S1, qui contient tous les nucléotides non liés aux billes au terme de cette incubation, est prélevé. Le culot de billes (C1) est alors lavé avec du tampon TEN 1000 (page 45) et le surnageant de lavage est *additionné* au surnageant S1. Le culot de billes est alors soumis à un traitement par l'enzyme Sma I qui coupe les oligonucléotides (qu'ils soient mono- ou bibiotinylés) approximativement en leur milieu (Figure 3.5, page 71, panneau B, schéma 3). Ainsi, si, dans le cas d'une phase bibiotinylée, certains oligonucléotides devaient être fixés aux billes par une seule biotine, la moitié de l'oligonucléotide non fixée devrait être récupérée dans le surnageant S2. Les billes sont sédimentées dans le culot C2 et sont lavées (les milieux de lavage sont conservés et additionnés au surnageant S2). Après lavage, les billes sont reprises dans du tampon BN 500 et soumises à une irradiation UV pour en détacher les oligonucléotides qui s'étaient fixés *via* une interaction biotine-streptavidine. Ces oligonucléotides sont alors récupérés sous forme de surnageant S3. L'ensemble des surnageants ci-dessus (S1 à S3) sont analysés sur gel d'agarose et les résultats sont montrés à la Figure 3.8, page 75.

Ceci est d'importance pour l'interprétation des résultats (*vide infra*)

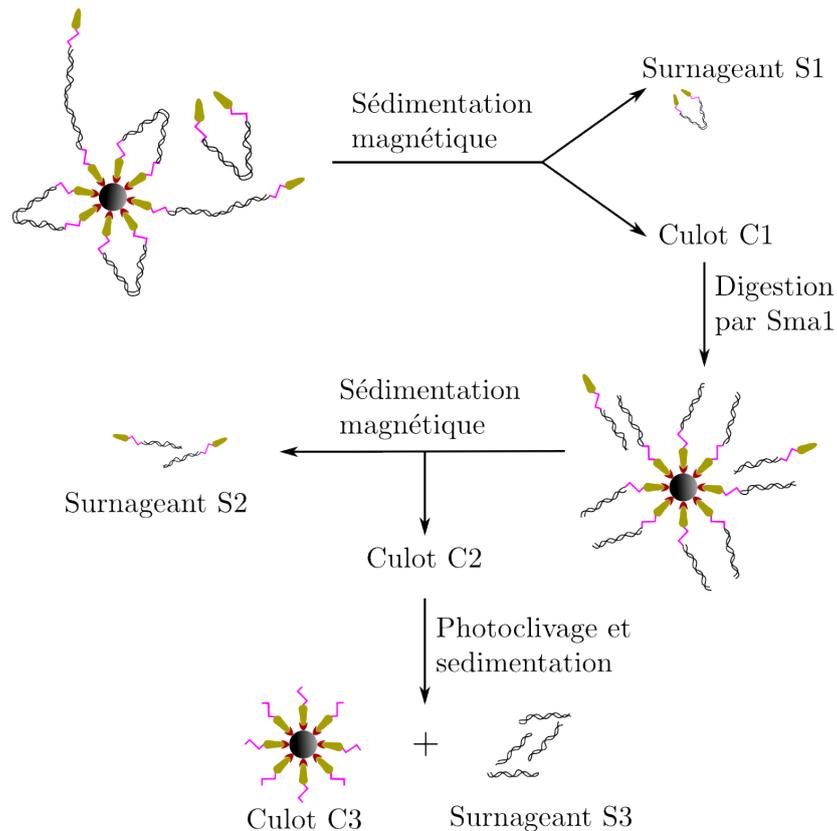


FIGURE 3.7 – **Contrôle de la qualité de la phase dénuée d’extrémités libres de l’ADN.** Après constitution des phases, les oligonucléotides biotinylés non fixés (S1) sont séparés des phases par sédimentation à l’aimant des billes (C1). Les oligonucléotides fixés sur les billes sont alors digérés par l’enzyme de restriction Sma I. Les moitiés d’oligonucléotides potentiellement libérées (S2) sont alors séparées des billes par sédimentation de ces dernières (C2). Enfin les fragments d’oligonucléotides fixés spécifiquement aux billes par leur extrémité biotinylée sont libérés par irradiation UV (S3) et séparés des supports chromatographiques par sédimentation des billes (C3).

Dans un premier temps, il s’agissait de vérifier que l’incubation des oligonucléotides biotinylés avec les billes recouvertes de streptavidine permettait d’en lier, à tout le moins, une forte proportion. Ainsi, nous attendions-nous à n’observer, dans le surnageant S1, que des traces d’oligonucléotides non fixés. Pourtant, le surnageant S1 (piste S1 de la Figure 3.8) contenait une quantité *massive* d’oligonucléotides, qui ne s’étaient donc pas fixés aux billes ou alors, qui interagissaient avec celles-ci de manière labile. En effet, ces oligonucléotides pouvaient, soit avoir été récupérés dans le surnageant d’incubation sur la nuit, soit avoir été décrochés des billes lors du lavage

de celles-ci avec le tampon TEN 1000 (comme nous l'avions fait remarquer plus haut). Ce résultat montre, quoi qu'il en soit, qu'une grande majorité des oligonucléotides ne s'étaient pas fixés aux billes par une interaction biotine–streptavidine — dont le  $K_d$  est très bas ( $\approx 10^{-14}M$ ) — qui n'est pas dissociée par une concentration de 1 M de NaCl. Pour ce qui est du surnageant S2, l'on voit sur le gel qu'il contient un peu d'oligonucléotide digéré (la bande a migré à 100 pb, ce qui est cohérent avec ce que l'on attend pour le surnageant S2). Enfin, le surnageant S3 contient bien les oligonucléotides libérés par photoclivage. Ces oligonucléotides ont les tailles attendues, ce qui prouve que la digestion par Sma I a été efficace.

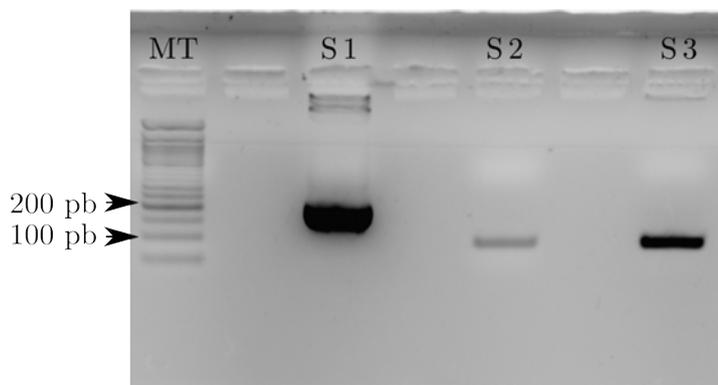


FIGURE 3.8 – **Fixation aspécifique des oligonucléotides sur les billes.** L'ADN observable sur la piste S1 montre que, lors de la constitution de la phase contrôle, les oligonucléotides ne se sont pas fixés aux billes *via* une interaction spécifique biotine–streptavidine.

Deux hypothèses peuvent être formulées pour expliquer le bien trop faible rendement de fixation des oligonucléotides aux billes :

1. la quantité d'extrémités d'ADN biotinylées mises au contact des billes dépasse de beaucoup la capacité de fixation de celles-ci, provoquant la perte de tous les oligonucléotides n'ayant pu avoir accès à un site de fixation ;
2. le tampon d'incubation utilisé (TEN 100, page 45) ne permet pas la fixation spécifique des oligonucléotides *via* une interaction biotine–streptavidine. Les oligonucléotides pourraient donc interagir faiblement avec les billes mais en seraient détachés lors des lavages effectués avec un tampon plus stringant (TEN 1000). En effet, l'on peut imaginer qu'un oligonucléotide qui se soit « couché » sur quelques molécules de streptavidine, sans pour autant se lier *via* sa biotine, ne soit pas conduit à se détacher de la bille, lorsque l'incubation est menée dans

un tampon faiblement salin. Si l'oligonucléotide mal lié n'a pas de raison de se détacher des billes, il n'aura aucune opportunité pour s'y attacher de nouveau, mais cette fois par le biais de la biotine, ce qui rendrait cette interaction définitive.

Pour tester la première hypothèse, nous avons diminué la quantité d'ADN utilisée pour constituer les phases, tout en augmentant la quantité de billes utilisées, ce qui nous a permis d'observer une amélioration du rendement de fixation des oligonucléotides sur les billes. Pour cette nouvelle tentative, nous avons analysé séparément le surnageant d'incubation des oligonucléotides avec les billes et les surnageants des deux lavages effectués au terme de celle-ci. Nous avons mis en évidence que l'ADN détecté sur la piste S1 de la Figure 3.8 provenait en grande partie de ces lavages, ce qui ajoutait du crédit à notre hypothèse numéro 2 ci-dessus. Nous avons supposé que les conditions peu stringentes dans lesquelles l'incubation était effectuée favorisaient des interactions non spécifiques entre les oligonucléotides et la streptavidine, qui entraient en compétition avec l'interaction biotine-streptavidine souhaitée. La solution à ce problème passait par l'établissement de conditions d'incubation des oligonucléotides avec les billes telles que les premiers ne pouvaient plus interagir avec les dernières par des interactions faibles, mais uniquement par la biais du couple biotine-streptavidine. Or, comment labiliser des interactions non spécifiques sans avoir recours à des détergents ou des solvants organiques ? En augmentant la force ionique. Nous avons donc décidé de faire l'incubation des oligonucléotides avec les billes dans un tampon à 1 M de NaCl (tampon TEN 1000), dont nous savions qu'il permettait de détacher les oligonucléotides liés de manière non spécifique aux billes. Les nouveaux résultats sont montrés à la Figure 3.9, page ci-contre.

Les résultats étaient, cette fois, totalement convaincants, avec une fixation complète des oligonucléotides sur les billes, au terme de l'incubation sur la nuit. En effet, la piste S1 montre uniquement la présence du plasmide pUC19, matrice de la PCR, dont nous avons vu qu'il copurifiait avec les produits de PCR ; voir § 3.1.1 et Figure 3.6, page 72). De manière intéressante, rien n'est visible dans la piste S2, ce qui démontre que les oligonucléotides biotinylés se fixent bien aux billes par les deux résidus biotine pour constituer une phase  $\diamond$ . Ces résultats encourageants devaient être confirmés dans la mesure où il nous fallait vérifier qu'il était bien possible de détacher les oligonucléotides des billes par irradiation des phases chromatographiques avec des rayons ultraviolets.

### 3.1.5 ÉLIMINATION DU SUPPORT CHROMATOGRAPHIQUE

Nous évoquons plus haut le problème analytique lié à la présence du support chromatographique au moment de l'élution des protéines d'intérêt (page

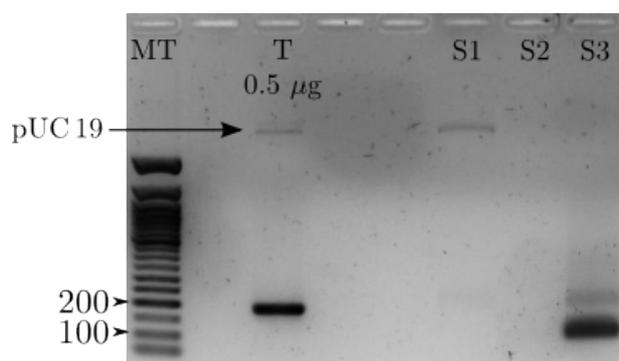


FIGURE 3.9 – **Double fixation des oligonucléotides de la phase contrôle.** L'absence d'ADN dans la piste S1 montre que les oligonucléotides biotinylés sont fixés au moins par une extrémité. La piste S2, correspondant au surnageant de digestion par l'enzyme de restriction Sma I, est également vierge ce qui montre que les oligonucléotides biotinylés se sont fixés aux billes recouvertes de streptavidine par leurs deux extrémités. La taille de l'ADN provenant du surnageant d'irradiation UV des phases chromatographiques, observé dans la piste S3, montre que la digestion par Sma I n'a pas été empêchée par la structure en « hanse » de l'ADN doublement lié aux billes.

64). Nous avons également présenté le dispositif expérimental envisagé pour détacher le produit de purification des supports chromatographiques sans avoir à utiliser aucun type d'agent d'éluion (Figure 3.3, page 68). Les phases, dont nous avons commencé la préparation, comme décrit ci-dessus, ont été soumises à une irradiation UV ( $\lambda_{>300 \text{ nm}}$ ) en vue de détacher les oligonucléotides des billes. Après sédimentation des billes à l'aimant, le surnageant S3 a été récupéré et a été analysé (piste S3 de la Figure 3.9). On peut remarquer que la bande de la piste S3 est épaisse. En réalité, cela est tout à fait attendu, car les oligonucléotides contenus dans le surnageant S3 doivent avoir deux tailles très proches (85 pb et 101 pb, Figure 3.5, page 71) puisqu'ils sont le produit de la digestion par Sma I de l'oligonucléotide de 186 pb. La digestion par Sma I de cet oligonucléotide doublement lié aux billes est quasiment quantitative, puisqu'il reste très peu d'oligonucléotide migrant à  $\approx 200$  pb (186 pb, précisément). Cette observation est très positive, car elle suggère que l'oligonucléotide, même doublement lié aux billes, présente suffisamment de flexibilité pour être digéré par un enzyme de restriction. Enfin, nous avons établi que la quantité d'ADN présente dans le surnageant S3 correspondait à 70% de la quantité d'ADN initialement utilisée pour constituer les phases chromatographiques, ce qui est un rendement tout à fait appréciable si l'on songe aux étapes qu'il a fallu pour constituer ces phases et en observer les

caractéristiques.

Au total, après un important travail d'optimisation des conditions de fabrication de nos phases, nous sommes parvenus à préparer de manière amplement satisfaisante les deux phases  $\langle \diamond \rangle$  (phase « contrôle ») et  $\star$  (censée mimer les cassures double-brin de l'ADN).

## 3.2 PURIFICATION DE PROTÉINES DE RÉPARATION DE L'ADN

Après avoir montré, dans la section précédente, qu'il nous avait été possible de mettre en place les deux développements méthodologiques envisagés pour permettre une purification par affinité efficace de protéines spécifiques des cassures double-brin de l'ADN, nous avons effectué une purification à partir d'extraits nucléaires de cellules *HeLa* en culture.

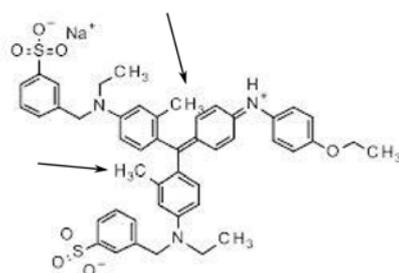
Cette purification est effectuée par incubation des phases chromatographiques, constituées puis lavées (§ 3.1.3, page 73), avec 800  $\mu\text{g}$  d'extraits nucléaires commerciaux provenant de cellules *HeLa*. Après une incubation de 10 min à 30°C, la suspension [billes + protéines] est lavée quatre fois avec du tampon Ku (page 49) pour éliminer toutes les protéines fixées de manière non spécifique. La suspension est ensuite équilibrée dans du tampon BN 0 avant d'être irradiée pendant 20 min à  $\lambda_{>300 \text{ nm}}$  en tampon BN 500. Au terme de l'irradiation UV, les produits de purification sont récupérés dans le surnageant, pour des analyses ultérieures.

### 3.2.1 ÉLECTROPHORÈSE EN CONDITIONS NATIVES

Comme le but de notre projet était de caractériser de nouvelles protéines de réparation de l'ADN et que celle-ci n'opère — dans une vaste majorité des cas — que par assemblage de complexes supramoléculaires au niveau du dommage, il était impératif que la méthode par laquelle nous visualiserions les complexes fût de nature à conserver les interactions non covalentes entre partenaires des assemblages. Les produits de purification obtenus plus haut ont donc été résolus sur un gel de polyacrylamide de type « *Blue-Native PAGE* » [Schagger *et al.*, 1994, abrégé en BN PAGE]. Ce type d'électrophorèse utilise les propriétés du bleu de Coomassie G-250, qui se lie aux protéines au niveau de leurs amino-acides basiques.

Le bleu de Coomassie G-250, un colorant des protéines bien connu (Figure 3.10), se distingue du bleu de Coomassie R-250 par la présence de deux groupements méthyle supplémentaires qui lui donnent une teinte verte (alors que le bleu R-250 a une teinte rouge, lorsqu'ils sont tous deux très concentrés). La fixation de molécules de bleu de Coomassie sur les protéines leur confère des charges négatives. Les protéines ainsi chargées, placées

Du nom de la personne  
qui donna, à son insu, les  
cellules cancéreuses  
utilisées pour la création  
de la toute première lignée  
de cellules tumorales  
humaines en culture :  
Henrietta Lacks



Bleu de Coomassie G-250

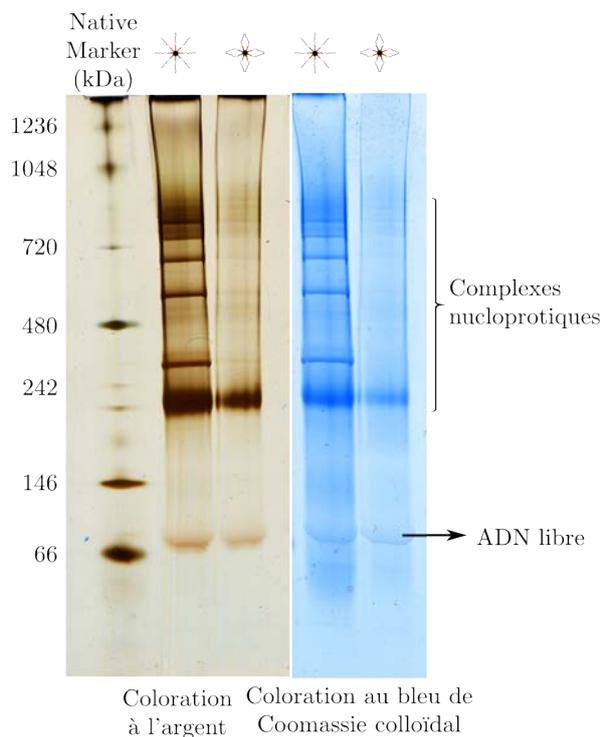
FIGURE 3.10 – **Structure du bleu de Coomassie G-250.** Le bleu de Coomassie G-250 se fixe aux acides aminés basiques des protéines. Les flèches pointent vers les deux groupements méthyle qui sont absents de la molécule R-250. Tiré de <http://structuresearch.merck-chemicals.com>.

dans un champ électrique sont soumises à une force qui leur impose une migration vers l'anode. Il faut noter que l'électrophorèse BN-PAGE est *non dénaturante*, car le bleu de Coomassie a des propriétés physico-chimiques telles qu'il n'est pas de nature à dissocier les protéines en interaction, à la différence du dodécylsulfate de sodium (SDS), qui est employé pour dénaturer les protéines avant une migration électrophorétique dénaturante (SDS PAGE).

Le profil d'électrophorèse obtenu avec nos produits de purification, tel que l'on peut l'observer à la Figure 3.11, page suivante, a été reproduit lors de plus de vingt expériences indépendantes. Dans cette figure, les pistes ✨ et ✨ ont été dupliquées afin de comparer les colorations obtenues avec deux protocoles différents. La coloration au bleu de Coomassie colloïdal [Dyballa et Metzger, 2012] tient ses promesses, avec une qualité de rendu comparable à celle obtenue en coloration à l'argent [Chevallet *et al.*, 2006]. C'est donc la coloration au bleu de Coomassie colloïdal qui a été retenue pour la coloration systématique des gels. Il faut noter qu'une telle qualité de séparation (avec, non seulement la séparation du matériel biologique en quelques bandes, mais également avec une résolution aussi fine) est rarement obtenue pour ce type d'électrophorèse *native* BN PAGE (*vide infra*). Si l'on regarde la piste ✨, dans laquelle sont résolus les produits de purification obtenus avec la phase dotée d'extrémités libres de l'ADN, l'on constate la présence de six bandes dont la migration correspond approximativement aux masses moléculaires apparentes suivantes : 720 kDa, 700 kDa, 500 kDa, 300 kDa, 240 kDa et enfin 66 kDa. Le gel de polyacrylamide employé ici est donc capable de permettre la résolution d'édifices moléculaires dont la masse moléculaire apparente est comprise entre le MDa et quelques dizaines de kDa.

Lorsque nous présentâmes ces gels à nos collaborateurs massistes, l'un de nos interlocuteurs nous corrigea, indiquant qu'il s'agissait sans doute de gels SDS-PAGE !

D'un point de vue biochimique, il ressort immédiatement de ce gel que



**FIGURE 3.11 – Séparation électrophorétique des produits de purification sur les deux phases chromatographiques.** Les pistes d'intérêt ont été dupliquées afin de tester deux colorations, au bleu de Coomassie colloïdal et à l'argent. \* : phase chromatographie dotée d'extrémités libres mimant des cassures de l'ADN; ✧ : phase chromatographie contrôle dénuée d'extrémités libres de l'ADN. Gel de polyacrylamide à gradient 4–16 % d'acrylamide.

les produits de purification obtenus sur les deux phases sont significativement différents, avec une diversité moléculaire plus grande dans le cas de la phase à extrémités libres de l'ADN. En effet, seule une bande paraît être commune aux deux échantillons (bande 5 à 240 kDa ; pour la bande à 66 kDa, voir plus bas). Ainsi, le résultat illustré à la Figure 3.11 constituait pour notre projet une étape clef pour deux raisons principales : la première est que nous démontrions que les deux phases étaient bel et bien capables de recruter des espèces moléculaires différentes ; la seconde est que la présence de plusieurs bandes, dont certaines de haut poids moléculaire apparent, permettait d'envisager la caractérisation fine de *complexes* de compositions différentes. En effet, la protéine de réparation de plus haute masse moléculaire est la sous-unité catalytique de la protéine kinase ADN-dépendante (460 kDa). La présence de bandes à  $\approx 720$  kDa (et plus encore, quoique moins bien résolues) semble indiquer que ces bandes sont bien le fait d'*assemblages*

protéiques n'ayant pas subi de dénaturation au cours de la migration.

Il convient de remarquer la bande de poids moléculaire apparent le plus faible (66 kDa), qui est observée dans les deux pistes \* et <math>\langle \diamond \rangle</math>. Cette bande correspond à l'oligonucléotide duplexe d'ADN de 186 pb employé pour préparer nos phases chromatographiques. Cet ADN a sans doute migré dans un état « libre », c'est-à-dire sans avoir lié de protéines, à une position correspondant à une masse apparente de 66 kDa, alors que sa masse calculée est de l'ordre de 120 kDa. Cette migration *apparemment* rapide peut être due au fait qu'il est lourdement chargé négativement, ce qui l'accélère considérablement sous l'effet du champ électrique. D'autre part, les marqueurs de taille sont des protéines globulaires et leur comportement électrophorétique n'est évidemment pas comparable à celui d'un ADN parfaitement flexible.

Si l'on s'attarde sur la bande à 240 kDa, qui se trouve être commune aux deux purifications, sur les phases \* et <math>\langle \diamond \rangle</math>, elle est probablement due à l'interaction de protéines n'ayant pas de spécificité pour les extrémités libres de l'ADN. C'est pour cette raison que cette bande a été nommée « bande aspécifique », par opposition à toutes les autres bandes que l'on n'observe que dans la piste \*. Le fait que nous puissions observer cette bande constitue une justification de notre système à deux phases chromatographiques, puisque nous pouvions d'emblée concentrer notre attention sur les bandes spécifiques des cassures double-brin de l'ADN (phase \*).

Remarquons que la bande 2 n'a pas été observée sur l'ensemble des gels obtenus et n'a donc pas été étudiée par la suite, de sorte que la bande 1 a été appelée « bande à 720 kDa ».

### 3.2.1.1 LES BANDES CONTIENNENT DES COMPLEXES MULTIPROTÉIQUES

Comme nous l'évoquions plus haut, les produits de purification obtenus sur les phases \* migrent à des masses moléculaires apparentes relativement importantes, ce qui suggérerait qu'il s'agissait là d'assemblages non covalents de protéines. Afin de vérifier cette hypothèse, nous avons découpé une piste \* de BN PAGE et en avons soumis le contenu à une deuxième électrophorèse *dénaturante* (SDS PAGE). La seconde migration électrophorétique se faisant dans une direction orthogonale à celle de la première, les protéines d'une même bande de première dimension devaient être résolues en de nouvelles bandes au cours de la deuxième dimension, si cette première bande contenait un ou plusieurs complexes non covalents.

Afin d'obtenir la meilleure dénaturation possible des protéines contenues dans le gel de première dimension, la bande de gel BN PAGE découpée est incubée pendant 10 min dans un tampon de dénaturation<sup>1</sup> avant de la chauffer quelques secondes dans un four à micro-ondes.

1. Composition : 12.5 mM de TRIS-HCl pH 7, 4 % de SDS, 65 mM de DTT, 0.3 mM de bleu de bromophénol et 2,1 M de glycérol.

Attention, le glycérol est très vite chauffé par les micro-ondes

Le profil électrophorétique de deuxième dimension, à la Figure 3.12, page ci-contre, révèle que certaines bandes de première dimension sont bel et bien résolues en plusieurs taches de deuxième dimension. Cette observation nous permet d'affirmer que les bandes observées sur les gels BN-PAGE correspondent bien à des *complexes* protéiques d'interactions non covalentes. Notons la présence de certaines protéines dans plusieurs complexes séparés en première dimension, comme c'est le cas des protéines XRCC5 (p 190) et XRCC6 (70 kDa et 80 kDa, respectivement ; flèches sur la figure), ce qui se traduit par la visualisation d'un chapelet de taches continues dans la dimension horizontale du gel SDS-PAGE.

D'un point de vue technique, faisons remarquer que la migration des protéines en deuxième dimension laisse apparaître un motif en hyperbole (tracé jaune). Ce motif est dû à la migration, au cours de la deuxième électrophorèse, de protéines qui existaient *sous une forme monomérique* dans la piste de première dimension. En effet, si une protéine se trouve sous forme monomérique au cours de la première dimension, sa migration au cours de la seconde ne devrait pas être modifiée (la masse apparente de la protéine n'a pas de raison de varier de la première à la deuxième migration). Ainsi, la protéine devrait migrer sur une diagonale orientée, sur le gel, de haut à droite en bas à gauche. Mais comme le deuxième gel de polyacrylamide est un gel de propriétés physico-chimiques différentes du premier, la migration de deuxième dimension n'est pas l'exacte réplique de celle de première dimension. Ainsi, l'on obtient en deuxième dimension une « compression » des protéines vers les hautes masses, ce qui produit non pas une diagonale, mais une courbe hyperbolique.

### 3.2.2 POINTS TECHNIQUES CRITIQUES

Si le profil de migration visualisé a été obtenu de manière reproductible, la qualité du gel peut varier rapidement en fonction de la manière dont l'électrophorèse a été effectuée.

Qualité évaluée tant par la netteté des bandes, qui doivent être bien résolues, que par le fait que les pistes sont bien rectilignes et dénuées d'artefact de migration

**Deux tampons de cathode sont utilisés en séquence.** Lors de la migration électrophorétique en gel BN PAGE, le tampon du compartiment cathodique est trop opaque pour que l'expérimentateur puisse suivre l'avancée du front de migration. Cette opacité provient de la forte concentration en bleu de Coomassie dans ce tampon. Il est donc nécessaire de préparer un second tampon, moins concentré en bleu de Coomassie, et de remplacer le premier avec celui-ci lorsque les différentes protéines présentes dans l'échantillon ont bien pénétré dans le gel (après 1 h de migration lente, par exemple). Ainsi, le front de migration devient visible sous la forme d'une ligne noire sur fond bleu clair. Il est alors possible de séparer plus longtemps les complexes sans risquer de perdre de matériel.

Ligne correspondant au bleu de Coomassie contenu dans le premier tampon et qui ne s'est pas fixé sur des protéines

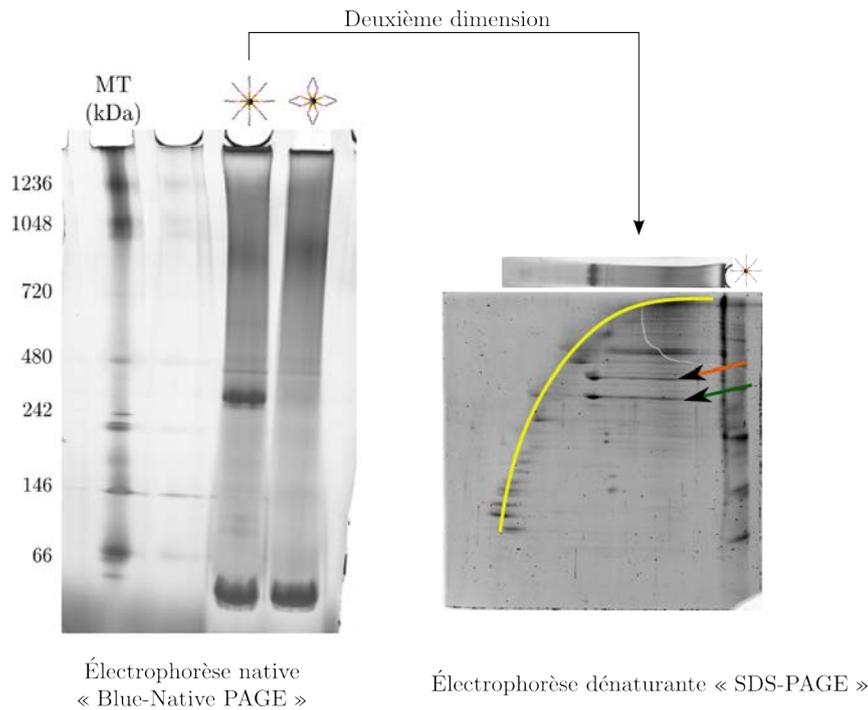


FIGURE 3.12 – **Électrophorèse de deuxième dimension en conditions dénaturantes.** La piste \* provenant d'un gel BN-PAGE (première dimension, à gauche ; conditions natives), est placée sur un gel SDS-PAGE (deuxième dimension, à droite ; conditions dénaturantes). Cette seconde dimension permet de résoudre en leurs différents constituants des complexes non covalents séparés en première dimension.

**Préparation des tampons d'électrophorèse.** Le bleu de Coomassie utilisé dans les tampons de cathode a tendance à s'agréger lorsqu'il est en solution, ce qui produit des granules qui s'accumulent dans les dents du gel de polyacrylamide. Une forte sonication pendant  $\frac{1}{4}$  h, juste avant la migration électrophorétique permet de réduire la dimension de ces agrégats et donc d'améliorer considérablement la qualité visuelle du gel.

### 3.2.3 ÉTUDE QUALITATIVE DES PRODUITS DE PURIFICATION

Si nous avons pu observer des complexes protéiques par électrophorèse, c'est que ces complexes se sont nécessairement assemblés sur nos phases chromatographiques et qu'ils ont bien été détachés des billes, puisque celles-ci sont éliminées après la purification. Or, ce qui est détaché des billes est, *stricto sensu*, l'oligonucléotide, qui — précisément — est lié aux billes par le biais du lien photoclivable. Ainsi, nous avons tout lieu de penser que le matériel purifié était constitué de protéines *et* d'ADN.

Les sections suivantes décrivent les expériences qui nous ont permis d'établir de quelle nature étaient ces produits de purification et comment ils intervenaient sur les phases chromatographiques, en particulier sur la phase ✱.

### 3.2.3.1 LES COMPLEXES VISUALISÉS COMPORTENT-ILS DE L'ADN ?

La manière que nous avons eue de tester la présence d'ADN au niveau des complexes purifiés sur la phase ✱ est illustrée à la Figure 3.13, page ci-contre, voie de gauche. Il s'agit de radiomarquer l'oligonucléotide avant même de l'incuber avec les billes, ce qui produit une phase ✱. La réaction enzymatique recherchée est mise en œuvre par la polynucléotide kinase (PNK), qui phosphoryle l'extrémité 5'-OH d'un oligonucléotide en utilisant le groupement phosphate en  $\gamma$  de l'ATP. Si l'atome de phosphore de ce groupement phosphate est l'isotope  $^{32}\text{P}$ , alors la radioactivité sera transférée sur l'oligonucléotide au cours de la réaction enzymatique. Cette réaction de phosphorylation ne peut être mise en œuvre que sur l'oligonucléotide monobiotinylé, car il dispose d'une extrémité 5'OH libre (contrairement à l'oligonucléotide bibiotinylé, qui a ses deux extrémités 5' engagées dans le lien photoclivable). Ainsi, au terme de la réaction de la PNK, la radioactivité sera portée, par l'oligonucléotide, à son extrémité libre — c'est-à-dire l'extrémité *distale* par rapport aux billes.

Une fois la phase chromatographique ✱ ainsi constituée, les étapes de purification qui suivent sont en tous points identiques à ce qui a été décrit plus haut. Au terme de la migration électrophorétique des produits de purification, le gel BN PAGE est coloré comme d'ordinaire et est ensuite soumis à autoradiographie. Le gel coloré au bleu de Coomassie colloïdal et l'autoradiographie sont présentés à la Figure 3.13. On constate d'emblée que les complexes présents dans le gel BN PAGE sont radioactifs, puisqu'ils impressionnent le film à leur emplacement précis. Une exception notable est à remarquer : la bande que nous qualifions d'« aspécifique » ne comporte pas d'ADN radioactif, puisqu'elle n'est pas « répliquée » sur l'autoradiographie.

Noter que la topologie de la radioactivité sur l'oligonucléotide est essentielle pour l'interprétation des résultats (*vide infra*)

Dans la figure, le rapport  $\frac{\text{signal}}{\text{bruit}}$  dans ce gel est sub-optimal, mais le profil électrophorétique est bien le même que celui observé en routine

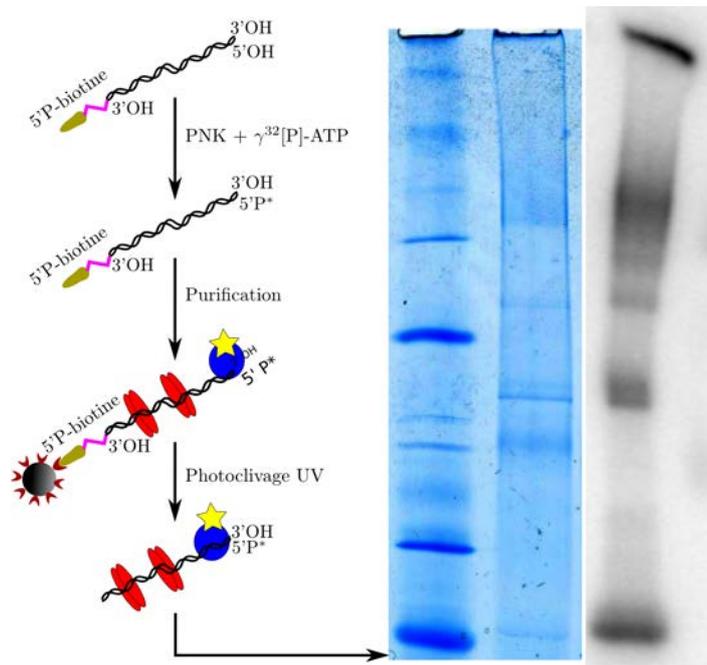


FIGURE 3.13 – **Suivi des oligonucléotides dans les complexes purifiés.** Marquage des oligonucléotides constituant les phases monobiotinylées dans le but de visualiser par autoradiographie ces oligonucléotides sur le gel BN PAGE. À chaque complexe protéique spécifique de la phase monobiotinylée, visible par coloration au bleu colloïdal de Coomassie (gel à gauche), correspond une bande sur l'autoradiographie adjacente (autoradiographie à droite).

Ces résultats nous permettent de tirer plusieurs conclusions :

- \* Les assemblages protéiques formés sur les oligonucléotides de la phase perdurent au cours de la migration électrophorétique sous forme de complexes « *nucléoprotéiques* » puisqu'ils comportent encore l'ADN sur lequel ils se sont formés ;
- \* Les oligonucléotides qui migrent associés aux assemblages protéiques sont radioactifs : ils sont *nécessairement* intacts. En effet, s'ils avaient dû subir une quelconque digestion, la radioactivité aurait été perdue ou leur migration aurait été altérée, ce qui se serait vu en comparant les bandes détectées en autoradiographie à celles visibles au bleu de Coomassie ;
- \* Le fait que les oligonucléotides portent encore la radioactivité prouve que leur extrémité est parfaitement protégée, vraisemblablement par

des protéines la reconnaissant spécifiquement ;

- \* La bande aspécifique ne contient pas d'ADN radioactif : soit elle ne comporte pas du tout d'ADN, soit l'extrémité radioactive a été digérée en cours de purification. Ainsi, en écho à ce qui a été dit au point précédent, cela pourrait signifier que les protéines de ce complexe ne se sont pas liées à cet ADN au niveau de son extrémité distale. Une autre hypothèse pourrait être qu'il s'agit là de complexes nucléoprotéiques qui se sont formés sur la phase chromatographique mais dont la stabilité est faible.

### 3.2.3.2 QUELLE EST LA TOPOLOGIE DES COMPLEXES NUCLÉOPROTÉIQUES ?

Le fait qu'il soit encore possible, au terme de la migration électrophorétique, de détecter de la radioactivité au niveau des bandes de plus hauts poids moléculaires apparents était extrêmement positif et nous a conduit à tenter de localiser plus précisément les assemblages protéiques au niveau de la séquence des oligonucléotides. L'approche que nous avons tentée est de type « empreinte aux nucléases », bien connue des biologistes moléculaires. Le principe de l'expérience est décrit à la Figure 3.14, page ci-contre, panneau A.

Dans notre cas, il s'est agi de reprendre l'expérience de radiomarquage de l'ADN avant la constitution de la phase ✱, d'effectuer la purification des complexes comme d'ordinaire et enfin de photoclivrer la suspension de billes afin de ne conserver dans le surnageant que les complexes nucléoprotéiques. C'est à ce moment qu'intervient une étape supplémentaire, par rapport à l'expérience de radiomarquage décrite plus haut : les complexes nucléoprotéiques purifiés en solution sont soumis à un traitement endonucléolytique par une combinaison de deux endonucléases, benzonase et DNase I. Ces nucléases sont censées provoquer des coupures non spécifiques de la séquence nucléotidique à chaque emplacement non « occupé » par les assemblages protéiques recrutés au cours de la purification. Dans une expérience d'empreinte aux nucléases ordinaire, l'ADN ainsi morcelé est analysé sur gel (d'agarose ou de polyacrylamide) pour tenter de déduire de la taille des fragments obtenus les sites de coupure par les endonucléases. Or, ces sites de coupure sont le reflet direct de l'inoccupation de la séquence par les protéines, ce qui constitue, de manière indirecte, l'empreinte d'occupation de l'oligonucléotide par les protéines. Dans notre cas, juste après le traitement aux endonucléases, les produits de purification digérés ont été soumis à une électrophorèse BN PAGE ordinaire en vue d'effectuer une autoradiographie, comme décrit plus haut.

Le profil de migration électrophorétique obtenu après digestion aux endonucléases est comparé à celui obtenu pour des produits de purification



intacts (Figure 3.14, panneau B). L'on peut constater un déplacement des trois complexes de haut poids moléculaire apparent vers les basses masses (migration accélérée). Ainsi, la bande à 720 kDa devient plus discrète et se déplace un peu en-dessous de cette taille, alors que les bandes à 480 kDa et à 240 kDa disparaissent toutes deux. Enfin, une nouvelle bande apparaît en-dessous de la bande aspécifique.

Le fait que l'on observe un différentiel de migration avant et après traitement aux endonucléases des complexes purifiés, confirme que l'ADN est une composante intrinsèque de ces complexes. D'autre part, le fait que la radioactivité soit encore présente dans les complexes nucléoprotéiques après digestion par les nucléases montre que l'extrémité de ces complexes est protégée, sans quoi l'on aurait perdu la radioactivité. Il y a là l'idée que les complexes purifiés sur la phase  $\star$  pourraient effectivement être des complexes spécifiques des extrémités libres de l'ADN.

Il faut noter, malgré ces résultats encourageants, que notre expérience ne permet pas de déduire la filiation des bandes apparaissant après digestion aux endonucléases par rapport aux bandes originelles (avant digestion). S'il est possible de penser que la fine bande présente en-dessous de 720 kDa après traitement aux nucléases provient de la bande plus large au dessus de 720 kDa en absence de nucléases, il est impossible de dire de quel(s) complexe(s) provient la bande apparaissant en-dessous de la bande aspécifique. Remarquons également que la migration électrophorétique de la bande aspécifique n'est pas altérée par le traitement aux nucléases, ce qui renforce l'idée que le matériel contenu dans celle-ci n'est pas de nature nucléoprotéique.

### 3.2.3.3 LES COMPLEXES SONT-ILS SPÉCIFIQUES DES EXTRÉMITÉS DE L'ADN ?

Pour établir de manière irréfragable la position des assemblages protéiques sur l'oligonucléotide de la phase  $\star$ , nous avons découpé les bandes de polyacrylamide (numérotées sur la Figure 3.14, panneau B) et les avons réduites en purée afin d'en extraire l'ADN. L'ADN contenu dans les différents complexes nucléoprotéiques a ensuite été soumis à une migration électrophorétique haute résolution sur un gel de séquence (Figure 3.14, panneau C). Pour pouvoir déduire la séquence de ces oligonucléotides sur la base de leur migration électrophorétique, l'oligonucléotide originel ayant servi pour la constitution de la phase  $\star$  a été séquencé selon la méthode de Maxam et Gilbert (pistes G+A latérales). Les oligonucléotides extraits des bandes 1, 2, 3, 4 du gel BN PAGE (panneau B) ont été déposés dans les puits de même numéro du gel de séquence (panneau C). Leur migration et le motif d'échelle qu'ils donnent permettent d'en déterminer la séquence par comparaison avec les pistes G+A. La piste C du gel de séquence comporte l'ADN contrôle non séquencé et servait à localiser la position de migration des deux oligonucléotides simple-brin de 186 bases. La piste 4 comporte

l'ADN tel qu'extrait de la bande 4 du gel BN PAGE (ADN libre). Cet ADN paraît intact, puisqu'il migre à la même position que celle à laquelle migre l'ADN contrôle. L'ADN des pistes 1, 2 et 3 a une taille d'une trentaine de nucléotides ou moins. Les résultats de cette expérience permettent de tirer plusieurs conclusions :

- \* Les complexes nucléoprotéiques purifiés sur la phase ✨ sont des complexes d'extrémité libre de l'ADN (une trentaine de nucléotides peut être considéré comme l'extrémité d'un oligonucléotide de 186 pb) ;
- \* Le fait de conserver des complexes supramoléculaires de plusieurs centaines de kDa sur une séquence aussi courte que trente paires de bases montre qu'il s'agit là majoritairement de complexes d'interactions protéine-protéine qui se sont formés sur l'ADN à un stade précoce de leur assemblage.

Il est intéressant de noter que les trois séquences d'ADN obtenues pour les bandes numérotées 1, 2 et 3 (panneau B et pistes du panneau C), sont toutes de tailles grossièrement équivalentes. Cependant, l'on peut noter que les deux pistes 1 et 2 du gel de séquence, correspondant aux complexes de poids moléculaires réduits, contiennent des ADN plus petits que ceux de la piste 3, qui correspond aux complexes de poids moléculaires plus importants.

### 3.2.4 ÉTUDE FONCTIONNELLE DES PRODUITS DE PURIFICATION

Les complexes que nous purifions sur notre phase ✨ s'assemblent à l'extrémité libre de l'oligonucléotide. Cette extrémité était censée, depuis la conception du projet, mimer une cassure double-brin de l'ADN. Nous avons voulu tester la capacité des produits de purification à effectuer la ligation d'oligonucléotides duplexes d'ADN.

Le principe de l'expérience est d'effectuer une purification ordinaire sur des billes fonctionnalisées avec un oligonucléotide duplexe d'ADN monobiotinylé. Afin de faire un test de ligation avec des oligonucléotides à bouts cohésifs, nous avons utilisé les fragments d'ADN provenant d'une coupure par Pst I de l'oligonucléotide *bibiotinylé* de 186 pb. La coupure de l'oligonucléotide par Pst I produit deux fragments, de 78 pb et 108 pb, avec des bouts cohésifs de 4 bases en 5'. Ces deux fragments sont biotinylés et peuvent donc servir pour la fonctionnalisation des billes recouvertes de streptavidine. Au terme de la purification des complexes de réparation sur cette phase chromatographique, il faut apporter un oligonucléotide duplexe d'ADN exogène, qui est additionné au milieu avec du tampon de ligation (l'ATP est essentiel à la ligation). L'oligonucléotide exogène aura été obtenu comme ci-dessus, par restriction enzymatique avec Pst I, à ceci près que l'ADN biotinylé de

L'oligonucléotide employé pour constituer la phase contrôle ✨

186 pb aura été préalablement irradié aux UV afin d'en détacher la biotine (et par là-même régénérer l'extrémité 5'-phosphate canonique).

Les complexes nucléoprotéiques formés spécifiquement sur la phase  $\star$ , dont nous voulons tester l'activité de ligation, peuvent exister sous deux formes :

- \* tels qu'ils sont assemblés sur l'oligonucléotide duplexe d'ADN avant photoclivage UV, c'est-à-dire que les complexes nucléoprotéiques sont encore attachés aux billes ;
- \* libres en solution, c'est-à-dire que la phase aura été irradiée aux UV pour les détacher des billes, ces dernières ayant été éliminées du milieu.

Les sections qui suivent décrivent les résultats de ligation que nous avons obtenus dans chacun des cas ci-dessus. Pour des raisons de cohérence biochimique et analytique, l'expérience de ligation a été faite *en une seule fois* selon les deux possibilités expérimentales ci-dessus et l'ensemble des résultats est montré sur un seul gel d'agarose, Figure 3.15, page ci-contre. Une expérience de ligation contrôle s'imposait d'emblée : la vérification de la capacité des extraits nucléaires de cellules HeLa en tant que tels à faire la ligation. En effet, il aurait été problématique que ces extraits n'eussent l'activité de ligation, ce qui nous était arrivé par le passé. Ce contrôle positif est donc effectué systématiquement et le résultat est visible à la piste 4 de la Figure 3.15, avec un taux de ligation de 19%, ce qui est tout à fait correct au vu des données disponibles dans la littérature pour ce genre d'expérience.

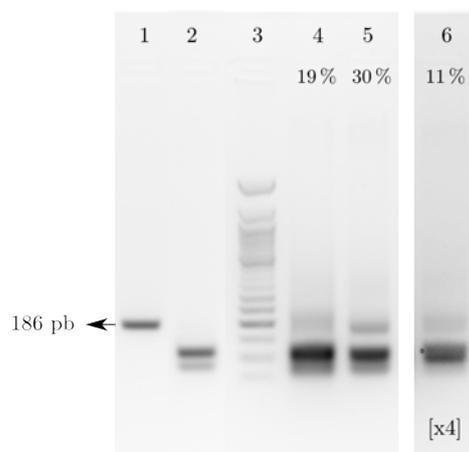
Plusieurs expériences ont permis d'obtenir des résultats concordants

#### 3.2.4.1 LIGATION PAR LES COMPLEXES ENCORE SUR LES BILLES

Le principe de l'expérience pour laquelle les complexes nucléoprotéiques sont encore attachés aux billes au moment où l'on additionne dans le milieu réactionnel du tampon de ligation et de l'oligonucléotide duplexe d'ADN exogène est décrit à la Figure 3.16, page 92.

L'apport du tampon de ligation et de l'oligonucléotide duplexe d'ADN exogène est effectué au terme de la purification des complexes nucléoprotéiques, *avant* de les dissocier des billes. Ce n'est qu'après incubation du milieu réactionnel pendant 1 heure à 37°C que les complexes nucléoprotéiques sont détachés des billes par photoclivage UV et les protéines digérées par la protéinase K à 55°C pendant 1 heure. Les résidus protéiques sont alors éliminés du milieu par extraction au phénol/chloroforme/alcool isoamylique ( $\Phi/\chi$ , dans la figure). L'ADN ainsi purifié est alors analysé sur gel d'agarose afin de quantifier les événements de ligation.

Le rendement de ligation obtenu lorsque les complexes nucléoprotéiques sont encore attachés aux billes est de 11% (Figure 3.15, page ci-contre, piste 6). Il est donc inférieur à ce qui est obtenu pour le contrôle de ligation



**FIGURE 3.15 – Ligation par les complexes nucléoprotéiques encore sur les billes.** La piste 1 montre l'ADN biotinylé de 186 pb utilisé comme substrat de départ. Cet ADN est digéré par l'enzyme Pst I, qui produit deux fragments de 78 pb et 108 pb (piste 2, la coupure est totale). Piste 4 : contrôle positif de ligation avec les extraits nucléaires totaux de cellules HeLa. Piste 5 : milieu de ligation après réaction des complexes nucléoprotéiques détachés des billes. Piste 6 : milieu de ligation après réaction des complexes nucléoprotéiques encore attachés aux billes. La piste 3 contient les marqueurs de taille « 50 bp ladder » (la quatrième bande à partir du bas, plus intense que les autres, est à 200 pb) Gel à 2 % d'agarose. Toutes les pistes appartiennent au même gel et ont été chargées au même moment. De manière reproductible, le cas de la ligation effectuée lorsque les complexes sont encore attachés aux billes (piste 6) donne une extraction de l'ADN moins efficace que dans les autres cas. Il a donc fallu augmenter le contraste d'un facteur 4 ([x4]), mais le rendement de ligation n'a pas changé au cours du traitement de l'image. Ceci a été vérifié avec deux logiciels de quantification différents. Le rendement de ligation a été calculé en comparant le rapport d'intensité entre la bande à 186 paires de base et les autres bandes de la piste.

tel que décrit plus haut (19 %, piste 4). La valeur de 11 % de rendement de ligation est cependant comparable aux valeurs obtenues pour des conditions expérimentales analogues [Audebert *et al.*, 2004].

Afin de montrer que la ligation détectée était bien le fait des complexes nucléoprotéiques, une expérience contrôle de ligation a été faite en effectuant une purification factice, au cours de laquelle les billes étaient « nues », c'est-à-dire dénuées d'ADN. Le protocole de ligation ne changeait en rien par rapport à ce qui vient d'être décrit. Les résultats de cette expérience nous ont étonnés, car une ligation non négligeable était observée. Une telle expérience contrôle est rare dans la littérature. Nous avons voulu la faire pour deux raisons : 1) car nous le pouvions, simplement grâce à notre système chromatographique à lien photoclivable ; 2) car nous avons une conscience

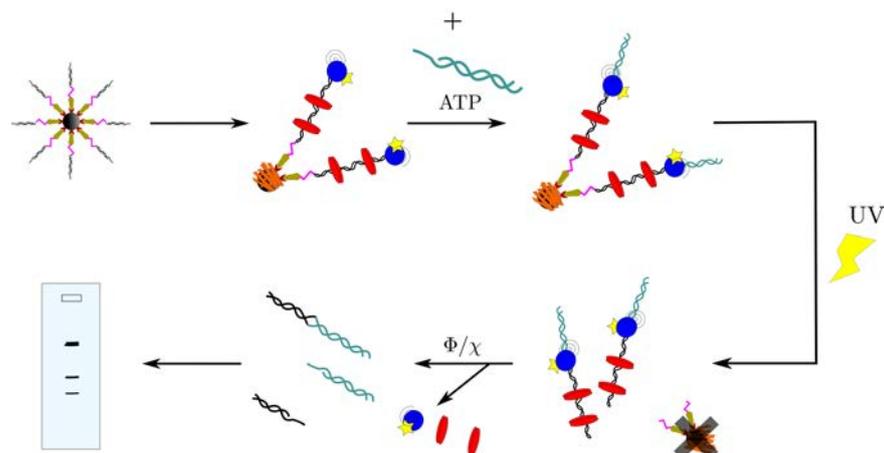


FIGURE 3.16 – **Ligation d’oligonucléotides attachés aux billes.** Les complexes nucléoprotéiques purifiés sont encore attachés aux billes au moment où l’on additionne dans le milieu réactionnel le tampon de ligation et l’oligonucléotide duplexe d’ADN exogène. Après incubation du milieu réactionnel, les complexes sont détachés des billes par irradiation UV, déprotéinisés, et l’ADN est extrait au phénol-chloroforme-alcool isoamylique ( $\Phi/\chi$ ) pour être analysé sur gel d’agarose. Si un événement de ligation a lieu, l’ADN résultant aura une taille approximativement double de celle de l’ADN non ligué.

aiguë de ce que les billes étaient fortement contaminées par des protéines qui s’y liaient de manière non spécifique [Hégarat *et al.*, 2007].

Ces résultats justifiaient pleinement d’effectuer une expérience de ligation par des complexes nucléoprotéiques préalablement détachés des billes.

#### 3.2.4.2 LIGATION PAR LES COMPLEXES DÉTACHÉS DES BILLES

Le principe de l’expérience pour laquelle la ligation est testée sur des complexes nucléoprotéiques préalablement détachés des billes est décrit à la Figure 3.17, page ci-contre.

Le rendement de ligation obtenu lorsque les complexes nucléoprotéiques sont détachés des billes est de 30 % (Figure 3.15, page précédente, piste 5). C’est un rendement très élevé, que nous avons obtenu de manière reproductible. Il est donc largement supérieur à celui obtenu dans le cas où les complexes sont encore attachés aux billes. Cela pourrait s’expliquer par le fait que les complexes encore attachés aux billes ont moins de degrés de liberté pour effectuer leur travail. En effet, malgré une agitation douce, censée maintenir les billes en suspension (et donc les complexes nucléoprotéiques qui s’y trouvent attachés), nous avons observé que les billes avaient

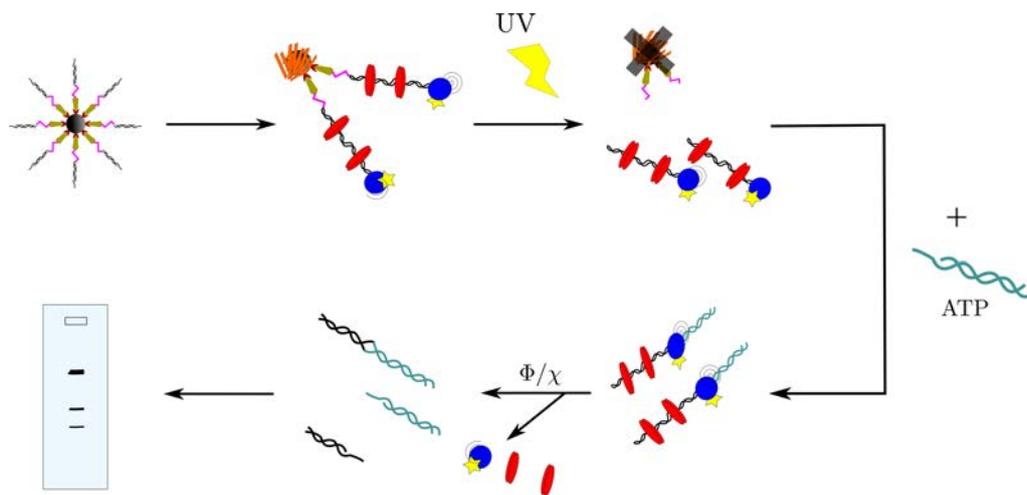


FIGURE 3.17 – **Ligation d’oligonucléotides détachés des billes.** Les complexes nucléoprotéiques purifiés sont d’abord détachés des billes par photoclivage UV et se retrouvent dans le surnageant (les billes sont éliminées). Ce n’est qu’alors que l’on additionne dans le milieu réactionnel le tampon de ligation et l’oligonucléotide duplexe d’ADN exogène. Après incubation du milieu réactionnel, les complexes sont déprotéinisés et l’ADN est extrait au phénol-chloroforme-alcool isoamylique ( $\Phi/\chi$ ) pour être analysé sur gel d’agarose. Si un événement de ligation a lieu, l’ADN résultant aura une taille approximativement double de celle de l’ADN non ligué.

malgré tout tendance à sédimenter. Inversement, dans le cas où les complexes nucléoprotéiques sont détachés des billes, ils évoluent librement en solution et sont donc mieux à même de rencontrer l'oligonucléotide duplexe d'ADN exogène additionné au milieu.

Une autre observation importante est que le rendement de ligation de 30 % est également très supérieur à celui observé pour les extraits nucléaires bruts (totaux). Ceci nous paraît très significatif, car cela rend compte, selon nous, de la purification hautement spécifique — sur notre phase  $*$  — de complexes de réparation de l'ADN tout à fait fonctionnels. En effet, au cours de la purification, seule une fraction très faible des protéines de l'extrait protéique nucléaire reste liée aux oligonucléotides de la phase  $*$ . À titre d'indication, l'expérience de ligation contrôle, menée avec les extraits nucléaires bruts, impliquait l'emploi de 320  $\mu\text{g}$  de protéines (piste 4, Figure 3.15, page 91) alors que l'on estime à 5–25  $\mu\text{g}$  la quantité de protéines retenues collectivement sur les oligonucléotides de la phase  $*$  (piste 5 de la même figure).

### 3.2.4.3 RÉFLEXIONS SUR LES EXPÉRIENCES DE LIGATION

Les résultats de ligation de deux oligonucléotides duplexes d'ADN à bouts collants de 4 bases décrits ci-dessus montrent que les extraits nucléaires comportent une ou plusieurs activités de ligation. Le fait que nous ayons pu effectuer une ligation très efficace avec les complexes nucléoprotéiques préalablement détachés des billes montre également que les assemblages purifiés sont indéniablement dotés d'une activité de ligation.

Il faut noter que l'emploi d'oligonucléotides à bouts cohésifs s'est révélé important pour pouvoir obtenir des rendements de ligation suffisants, afin de détecter les événements de ligation sur gel d'agarose imprégné de bromure d'éthidium. Ces expériences de ligation ont ainsi pu être menées sans avoir recours au radiomarquage, qui est typiquement employé dans les expériences décrites dans la littérature. Par ailleurs, le fait de n'avoir pas eu à recourir à la radioactivité pour suivre les événements de ligation nous a permis de rendre compte de tels événements même quand l'extrémité libre des ADN à liguer aurait été modifiée au cours de la purification, typiquement par des exonucléases. En effet, si les bouts cohésifs se retrouvaient raccourcis, cela ne poserait pas de problème de détection des produits de ligation sur gel d'agarose coloré au bromure d'éthidium. En revanche, si les oligonucléotides avaient été radiomarqués à leur extrémité distale, un seul événement d'exonucléation leur aurait fait perdre la radioactivité, ce qui aurait fait perdre la trace d'un éventuel événement de ligation ultérieur.

Les résultats de ligation décrits plus haut étaient extrêmement encourageants, mais nous avons voulu vérifier s'ils étaient effectivement le fait d'un processus de ligation canonique. Comme la ligation est rédhitoirement dépendante de la présence d'ATP, nous avons effectué une expérience de

ligation après avoir consommé tout l'ATP de l'extrait nucléaire. Aucun événement de ligation n'avait alors pu être observé.

Afin de vérifier que l'activité de ligation observée au cours de nos expériences provenait bien des protéines recrutées sur notre phase chromatographique, nous avons également réalisé une expérience contrôle, par l'emploi d'extraits nucléaires préalablement dénaturés par incubation de quelques min à 80°C. Aucune ligation ne put être observée dans ces conditions, ce qui suggérait que la ligation était bien le fait d'enzymes contenus dans les extraits nucléaires et qui avaient été purifiés sur notre phase chromatographique.

Les rendements de ligation observés sont tellement importants (rendement de 30 %, dans le cas de la ligation effectuée par les complexes nucléoprotéiques préalablement détachés des billes) qu'il est fort probable que l'ADN impliqué dans ces événements de ligation ait été l'objet de modifications chimiques en cours de purification, comme par l'action d'endo- ou d'exo-nucléases (observation que nous avons faites au cours d'autres expériences). Ceci est intéressant, car cela montre que la ligation peut s'opérer, dans notre système biochimique reconstitué, même sans bouts cohésifs complets (de 4 bases). On se rapprocherait alors de conditions de diversité des extrémités rencontrées dans la cellule, avec des complexes de réparation dynamiques et fortement actifs.

Les tests fonctionnels de ligation — menés sur des complexes nucléoprotéiques encore attachés aux billes — ont un défaut majeur : les protéines contaminantes présentes sur les billes peuvent interférer avec l'activité de ligation que l'on souhaite doser et qui est censée refléter uniquement la capacité fonctionnelle des complexes d'extrémité de l'ADN. Notre système chromatographique à lien photoclivable a permis de révéler ce biais, dont nous prétendons qu'il peut avoir des effets biochimiques parasites majeurs et conduire à des artefacts problématiques.

### 3.3 ANALYSES PAR SPECTROMÉTRIE DE MASSE

Les méthodologies analytiques décrites dans les sections précédentes nous ont permis de recruter sur des phases chromatographiques d'affinité des complexes protéine(s)-ADN (dits « nucléoprotéiques »). Ces complexes ont été séparés en gel de polyacrylamide non dénaturant et ont pu être visualisés directement par coloration au bleu de Coomassie colloïdal. Trois des bandes de migration électrophorétique comportent de l'ADN et ne sont présentes dans le gel que pour des échantillons purifiés sur une phase mimant les cassures double-brin (phase \*). Une dernière bande est présente quelle que soit la phase (↔ ou \*) et paraît ne pas contenir d'ADN (bande dite « aspécifique »).

Pour chaque expérience de purification, deux échantillons ont été systé-

L'on consomme tout l'ATP d'un échantillon en l'incubant avec du glucose et de l'hexokinase. Celle-ci phosphoryle le glucose en glucose-6-phosphate, consommant ainsi tout l'ATP présent dans le milieu.

matiquement déposés sur gel non dénaturant : l'échantillon purifié sur une phase ✱ et l'échantillon purifié sur la phase contrôle ✧). Une expérience type est montrée à la Figure 3.18, page 98. On constate qu'il y a bien quatre paires de bandes, les quatre bandes clairement visibles dans la piste ✱ et leur contrepartie (même si non visible) dans la piste ✧. Noter que le jeu de données sur lequel nous avons effectué les analyses porte sur six expériences indépendantes (n=6).

L'ensemble des observations qui ont pu être faites dans les sections précédentes suggère fortement que les complexes nucléoprotéiques séparés dans les différentes bandes représentent différents complexes de réparation de l'ADN. Grâce à de la chimie des protéines standard (digestion par la trypsine des protéines contenues dans les différentes bandes de gel de polyacrylamide préalablement excisées) et à des analyses par HPLC(C<sub>18</sub>)–SM/SM, nous avons identifié et quantifié les protéines constitutives des complexes nucléoprotéiques visualisés sur gel (voir *Matériels et méthodes*, page 43). Il s'agit là d'une stratégie analytique bien connue en protéomique, dite « *bottom-up* », car elle consiste à déconstruire un tout pour le reconstruire de manière indirecte (on digère toutes les protéines d'un mélange et l'on se sert des peptides pour toutes les identifier). Dans l'ensemble des analyses qui suivent, les protéines auront été identifiées sur la base d'au minimum 3 peptides séquencés par SM en tandem correspondant à la séquence de la protéine identifiée (« *peptide matches* »).

Comme nous avons travaillé exclusivement sur des extraits nucléaires de cellules humaines, toutes les recherches ont été effectuées dans la base de données de protéines SwissProt<sup>2</sup>. Compte-tenu du système analytique décrit plus haut, plusieurs manières s'offrent à nous pour la fouille des données de spectrométrie de masse :

1. Comparaison des protéines contenues dans différentes bandes appartenant à un même produit de purification sur phase dotée d'extrémités libres de l'ADN (phase ✱). Ainsi, l'on comparera, par exemple, les protéines identifiées dans la bande à 480 kDa à celles identifiées dans les autres bandes de la même piste de gel. On voit ici que l'on s'intéresse uniquement aux produits de purification sur phase ✱. On peut concevoir ces analyses comme des comparaisons « intra-pistes » de bandes ;
2. Comparaison des produits de purification obtenus sur les deux types

---

2. Il s'agit d'une base de données de très grande qualité car les curateurs vérifient les données en étudiant la bibliographie. Cette base est ainsi bien distincte d'autres bases dont les données sont obtenues de manière quasi-automatique à partir d'annotations de séquences génomiques. Cette base de données protéiques est à l'adresse suivante : <http://www.expasy.ch>.

de phases. Ici, chaque comparaison consiste à regarder la composition en protéines d'une bande (à un poids moléculaire donné) dans une piste  $\diamond$  et à la comparer à la composition d'une bande (de même poids moléculaire) dans une piste  $\ast$ . Par exemple, l'on comparera les protéines identifiées dans la bande à 720 kDa après purification sur phase contrôle  $\diamond$  à celles de la même bande mais pour une purification sur la phase  $\ast$ . On peut concevoir ces analyses comme des comparaisons « inter-pistes » de bandes.

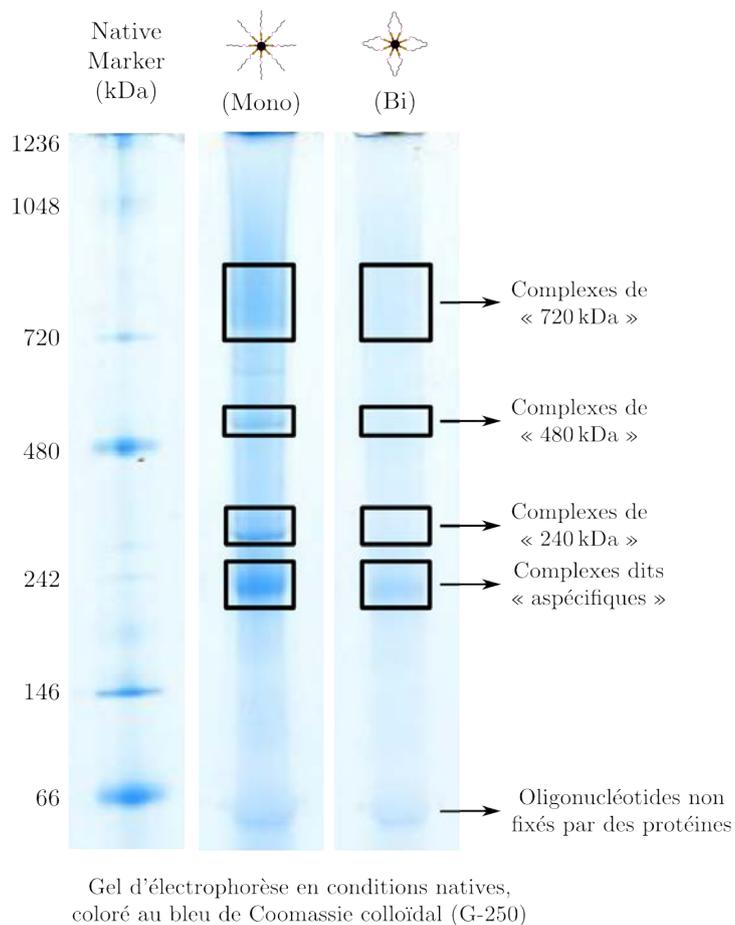
Nous avons ainsi pu obtenir le contenu protéique de chacune des bandes analysées, ce qui nous a conduit à identifier plus de 300 protéines. Nous nous sommes alors posés les questions suivantes :

- \* Dans quelle mesure les protéines purifiées sont-elles spécifiques des acides nucléiques ? Il s'agit d'une métrique d'intérêt, car elle rendrait compte de l'efficacité de la stratégie qui consiste à détacher les billes des produits de purification ;
- \* Dans quelle mesure les protéines purifiées sur nos phases chromatographiques sont-elles connues pour intervenir dans la réparation de l'ADN ? Ici aussi, il s'agirait d'une métrique intéressante, car elle permettrait de conclure sur la validité de l'approche de chromatographie d'affinité pour ce type de projet, en corrélation avec le point ci-dessus ;
- \* La composition en protéines des complexes séparés dans différentes bandes permet-elle de caractériser des complexes de natures différentes (voies de réparation, par exemple). En quoi la présence simultanée, dans un échantillon donné, de différents complexes nucléoprotéiques peut-elle renseigner sur le mode d'assemblage des protéines sur l'ADN (dynamique d'assemblage, caractère séquentiel ordonné de l'assemblage, par exemple) ?
- \* Nos résultats mettent-ils en relief la présence de protéines dont le rôle dans la réparation de l'ADN n'est pas décrit mais qui, pourtant, sont présentes uniquement dans les produits de purification obtenus sur phase  $\ast$  ?

Un échantillon, ici, est une piste de gel BN PAGE, soit le produit d'une purification

N'ayant pas voulu nous satisfaire, pour la fouille des données, d'un critère purement binaire (présence ou absence d'une protéine donnée dans un échantillon donné), nous avons fondé toutes nos analyses sur une grandeur numérique, l'emPAI, qui quantifie l'abondance d'une protéine donnée dans un échantillon. La quantification de protéines sur la base de l'emPAI ne requiert pas d'étape préliminaire de chimie des protéines (marquage, soit des protéines, soit des peptides issus de celles-ci) et est donc dite « *label-free* ».

*In extenso :*  
« *exponentially-modified Protein Abundance Index* »



**FIGURE 3.18 – Gel non dénaturant obtenu pour une expérience type de purification sur les phases  $\star$  et  $\star$ .** Les bandes correspondant aux complexes protéiques spécifiques des cassures double-brin de l'ADN (piste  $\star$ ) ont été excisées. Les bandes en vis-à-vis ont également été systématiquement excisées de la piste  $\star$ .

L'emPAI est le produit d'une découverte fortuite de la part de [Ishihama et al.](#), en 2005, qui appliquaient dans leurs travaux le PAI tel qu'il avait été défini par [Rappsilber et al.](#), en 2002. Ces auteurs avaient remarqué que le calcul suivant :

$$\text{emPAI} = 10^{\text{PAI}} - 1$$

« Protein Abundance  
Index »

reflétait plus directement la concentration d'une protéine donnée dans un échantillon. Dans cette équation, PAI quantifie l'abondance relative des protéines présentes dans un échantillon et est défini comme suit :

$$\text{PAI} = \frac{N_{\text{obs}}}{N_{\text{tot}}}$$

avec

- \*  $N_{\text{obs}}$  : nombre de peptides détectés correspondant à la séquence d'une protéine donnée ;
- \*  $N_{\text{tot}}$  : nombre théorique de peptides attendus sur la base de la séquence protéique et d'autres paramètres gouvernant l'analyse par spectrométrie de masse, comme la gamme de  $m/z$  employée pour les spectres de SM ;

Nous avons ainsi pu réaliser une analyse fondée à la fois sur l'identification de protéines au sein de complexes de réparation des cassures double-brin de l'ADN, mais aussi sur la variation de leur quantité au sein des différents complexes et en fonction de la phase chromatographique utilisée (avec ou sans extrémités libres de l'ADN). Nous avons également développé des outils d'analyse permettant d'appliquer notre stratégie de fouille de données au plus grand nombre possible de situations.

### 3.3.1 RÉPARTITION FONCTIONNELLE (EN NOMBRE)

La Figure 3.19, page suivante montre, sur les cent protéines les plus abondantes, que la très grande majorité des protéines identifiées au cours des six expériences utilisées pour la fouille de données appartiennent à des groupes fonctionnels étroitement liés à la physiologie des acides nucléiques, avec une forte prévalence de la physiologie de l'ADN. Les protéines impliquées dans la régulation de la topologie de l'ADN (protéines histones et topoisomérases) sont les plus nombreuses (29 %) devant les protéines de réparation de l'ADN (protéines impliquées dans la réparation des cassures simple-brin et double-brin de l'ADN — 20 %).

### 3.3.2 RÉPARTITION FONCTIONNELLE (EN QUANTITÉ)

Lorsque l'on s'intéresse, non pas au nombre de protéines classées par catégorie, comme ci-dessus, mais à la somme des abondances (mesurées par l'emPAI) des protéines purifiées, l'on constate d'emblée que la très forte prépondérance de la protéine PARP1 (p 175) masque la contribution des autres protéines purifiées (l'on obtient, en effet, 99 % des protéines classées dans la catégorie « Réparation de l'ADN », ce qui est manifestement un résultat artificiel. Si PARP1 est ôtée de la liste des protéines sur lesquelles le calcul est effectué, l'on obtient la répartition fonctionnelle décrite à la Figure 3.20, page suivante. Les protéines impliquées dans la réparation de l'ADN sont représentées à hauteur du tiers de la quantité de toutes les

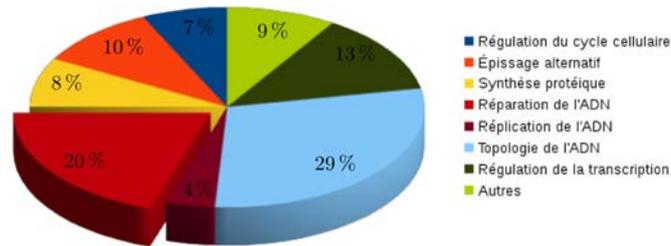


FIGURE 3.19 – Répartition fonctionnelle (en nombre) des cent protéines les plus abondantes. La majorité des cent protéines les plus abondantes identifiées dans ce travail se répartissent en huit catégories fonctionnelles. La physiologie de l'ADN est majoritairement représentée (67%).

protéines recrutées sur nos phases. Noter que la présence de la protéine RLA1 (p 180) augmente d'une manière très significative la représentativité des protéines de synthèse protéique. Il s'agit là d'un cas artefactuel comparable à celui de PARP1, évoqué plus haut.

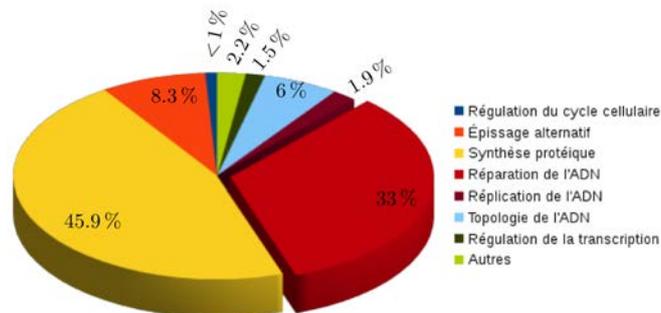


FIGURE 3.20 – Répartition fonctionnelle (en quantité) des cent protéines les plus abondantes (sans PARP1). Même lorsque l'on ignore la protéine PARP1, la quantité des protéines de réparation de l'ADN purifiées représente un tiers de la quantité totale de protéines dans les complexes analysés. L'importante proportion de protéines impliquées dans la synthèse des protéines est due à l'abondance de la protéine RLA1.

Si l'on ôte la protéine RLA1 de la liste des protéines, la répartition quantitative est alors modifiée comme montré à la Figure 3.21, page 101. On constate alors que, logiquement la proportion quantitative des protéines de la synthèse protéique est diminuée, à la faveur d'une augmentation des proportions pour toutes les autres catégories. La proportion quantitative des protéines de réparation est alors de 58 %.

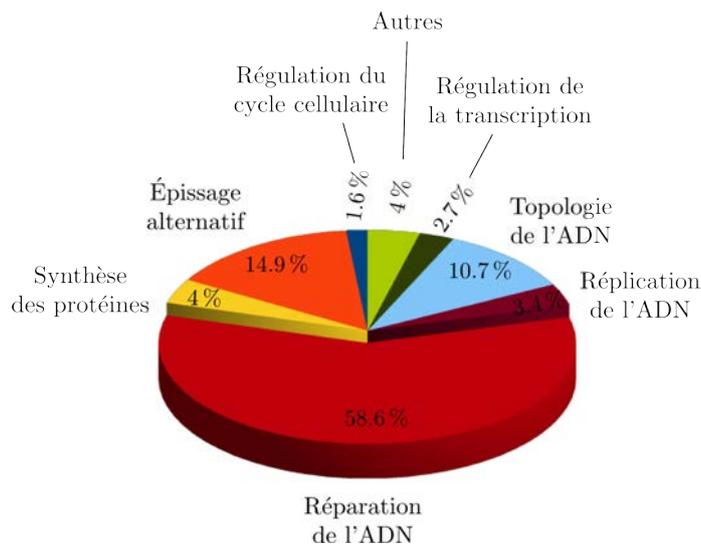


FIGURE 3.21 – **Répartition fonctionnelle (en quantité) des cent protéines les plus abondantes (sans PARP1 ni RLA1).** Lorsque l'on ignore les protéines PARP1 et RLA1, la quantité des protéines de réparation de l'ADN purifiées représente plus de la moitié de la quantité totale de protéines dans les complexes analysés. L'importante proportion de protéines impliquées dans la synthèse des protéines était donc bien à l'abondance de la protéine RLA1 (voir Figure 3.20, page ci-contre).

### 3.3.3 COMPARAISON DES BANDES OBTENUES SUR PHASE ✱

Nous avons vu plus haut que les complexes que nous purifions sur la phase ✱ sont résolus en quatre bandes, dont trois sont observées uniquement pour la phase ✱ (contrairement à la bande « aspécifique », qui est présente dans les deux cas). Nous nous référerons, dans ce qui suit, aux différentes bandes par leur poids moléculaire apparent : 720, 480 et 240 kDa.

Comme évoqué plus haut, une manière de comparer les complexes qui se sont assemblés sur la phase ✱ — et qui sont résolus dans les trois bandes — est de comparer la composition en protéines de ces bandes l'une par rapport à l'autre (raisonnement 1, page 96). Ces comparaisons s'appuient sur l'abondance des protéines dans les différentes bandes de gel BN PAGE.

L'abondance d'une protéine donnée dans un échantillon donné est quantifiée par sa valeur d'emPAI, décrite plus haut.

Une première approche consiste simplement à énumérer les protéines que l'on ne trouve que dans une bande et pas dans les deux autres ; ces protéines sont dites « exclusives » de la bande dans laquelle elles sont trouvées.

Une deuxième approche consiste à véritablement ne s'intéresser qu'aux protéines qui sont au moins communes à deux bandes et, dans ce cas, à comparer leur abondance — emPAI — dans les différentes bandes.

Le volume de calculs décrits dans les sections suivantes auraient été ingérables manuellement (même avec l'aide d'un tableur). Nous avons donc développé un outil *ad hoc* qui nous permit d'effectuer la majorité des calculs nécessaires à la fouille de nos données. Il s'agit du logiciel *dsbRepair* dont la documentation (en anglais) est fournie en annexe (voir § 5.0.4, page 195). Cet outil est conçu pour s'interfacer avec la base de données relationnelles dans laquelle un autre logiciel développé dans le cadre de nos études a permis l'injection des données d'identification de protéines. Les requêtes à la base de données sont effectuées en fonction des *desiderata* de l'utilisateur et les calculs sont menés sur les données extraites de la base.

### 3.3.3.1 PROTÉINES EXCLUSIVES D'UNE BANDE DONNÉE

Pour pouvoir déterminer quelles sont les protéines exclusives d'une bande donnée, il est nécessaire d'établir — pour toutes les protéines présentes dans les trois bandes — quelle est leur proportion dans chacune des bandes. Si une protéine est présente dans une ou deux bandes seulement (absente, donc, d'au moins une bande), sa valeur d'emPAI dans cette bande est nulle. Comme notre jeu de données comporte plusieurs échantillons ( $n=6$ ), nous pouvons faire une moyenne des proportions (pour une protéine donnée) dans chacune des bandes. Il y a deux manières de calculer la moyenne des proportions, qui sont illustrées à la Figure 3.22 (moyennes « totale » ou « restreinte »). Ces deux méthodes de calcul sont détaillées ci-dessous.

**3.3.3.1.1 La méthode de calcul en moyenne « totale »** est illustrée à la Figure 3.22, voie 1. Ce calcul est mené isolément pour *chaque* protéine trouvée dans au moins l'une des trois bandes d'une piste de gel BN PAGE. Dans cette analyse de données, il faut concevoir « bande » comme « poids moléculaire apparent ». Comme notre jeu de données comporte 6 échantillons ( $n=6$  pistes de gel), la moyenne « totale » doit être le ratio d'une somme sur 6. Si une protéine donnée est totalement absente de l'une des 6 pistes de gel BN PAGE, son pourcentage d'emPAI vaut 0. Cette méthode de calcul est donc « stringente », car un pourcentage d'emPAI nul ne contribue en rien à la somme des pourcentages, qui est pourtant bien divisée par le nombre total d'échantillons ( $n=6$ ). Dans l'exemple de la figure, une protéine donnée

*Protéine*, ici, vaut « identification par spectrométrie de masse » dans un échantillon donné, c'est-à-dire dans une bande de gel BN PAGE

Ici, « échantillon » a l'acception statistique et non pas analytique

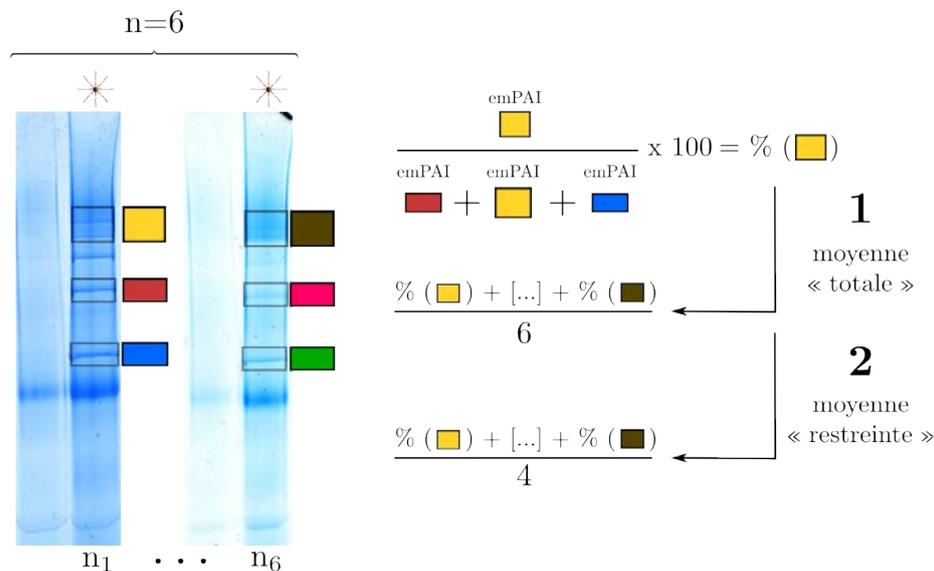


FIGURE 3.22 – Description des méthodes de calcul utilisées pour déterminer la proportion moyenne d’une protéine dans une bande donnée par rapport aux autres. Deux méthodes de calcul sont proposées. La voie 1, dite de moyenne « totale » est plus sélective et produit une liste de protéines limitée. La voie 2 est moins stringente (voir texte).

est présente dans les trois bandes du premier gel  $n_1$  (jaune, rouge, bleu). Le calcul procède alors comme ci-dessous :

1. La valeur  $\text{emPAI}_{\text{jaune}}$  est conservée comme valeur « de référence » (elle se trouvera au numérateur de la fraction) ;
2. La somme des  $\text{emPAI}_{\text{jaune+rouge+bleu}}$  est calculée pour déterminer la quantité cumulée de cette protéine dans les trois bandes d’une piste de gel BN PAGE ;
3. Le rapport  $\frac{\text{emPAI}_{\text{jaune}}}{\text{emPAI}_{\text{jaune+rouge+bleu}}}$  est calculé et multiplié par 100, ce qui donne le pourcentage d’abondance de la protéine dans la bande de référence ( $\%_{\text{jaune}}$ ) ;
4. Ce calcul du pourcentage pour la piste  $n_1$  est reproduit depuis l’étape 1 pour le gel  $n_2$  et ainsi de suite pour  $n_3 \dots n_6$ .
5. La somme des six pourcentages obtenus lors des six itérations [étape 1 → étape 4] est calculée et divisée par le nombre de pistes  $n$  (soit,  $n=6$ ). L’on obtient alors la moyenne des pourcentages d’abondance de la protéine pour un poids moléculaire apparent donné ;

6. L'ensemble des étapes ci-dessus est alors reproduit pour la deuxième bande de la première piste (rouge en lieu et place de la jaune) et enfin pour la troisième (bleue).

Ce pourcentage est calculé à l'étape 5, page précédente de l'algorithme dont les étapes sont énumérées

Au terme des calculs décrits ci-dessus, toute protéine pour laquelle le pourcentage d'abondance moyen, pour un poids moléculaire apparent donné, est supérieur ou égal à 90 % est considérée comme *exclusive* de ce poids moléculaire. Ce calcul est effectué pour toutes les protéines identifiées dans les bandes analysées.

Avec cette méthode de calcul, le nombre de protéines exclusives que l'on obtient est relativement limité. Il nous a paru que cette méthode de calcul était trop « stringente », eu égard au fait que l'absence d'une protéine d'une piste BN PAGE était un événement de nature technique et qu'il pouvait ne pas refléter une dimension biologique. Il nous paraît rationnel, en effet, de considérer qu'une protéine, en particulier si elle est présente en faible quantité dans les cellules, puisse ne pas être identifiée à chaque fois par spectrométrie de masse. Cela est d'autant plus vrai que nous avons utilisé des paramètres *restrictifs* pour notre étude. Nous avons décidé que les protéines pouvaient être considérées comme purifiées de manière reproductible dès lors qu'elles étaient présentes dans la moitié ou plus des échantillons analysés (3 échantillons au moins sur 6). Plutôt que de diminuer le pourcentage à partir duquel nous considérons qu'une protéine est exclusive d'un poids moléculaire (une bande), nous avons modifié la méthode de calcul de ce pourcentage afin de pallier les aléas expérimentaux. Dans cette variante, le pourcentage d'abondance moyen de la protéine à un poids moléculaire donné est calculé uniquement sur la base des bandes dans lesquelles la protéine a été identifiée.

Nous avons restreint, en effet, les identifications protéiques à celles résultant du séquençage par SM de trois peptides concordants (« *matches* »)

**3.3.3.1.2 La méthode de calcul en moyenne « restreinte »** est illustrée à la Figure 3.22, voie 2. Ici, si une protéine n'est pas présente dans une bande de référence donnée (étape 1, page précédente de l'énumération ci-dessus) toute la piste de BN PAGE la contenant est éliminée du jeu de données. Le calcul est alors en tous points comparable à ce qui vient d'être décrit, si ce n'est que la moyenne des pourcentages (étape 5 de l'énumération ci-dessus) est calculée en divisant la somme des pourcentages par le nombre de bandes de référence dans lesquelles la protéine a été identifiée. L'on pourrait formuler l'effet logique de cette nouvelle méthode de calcul ainsi : « *cette nouvelle moyenne nous permet d'identifier comme exclusives d'une bande donnée des protéines qui y sont retrouvées exclusivement mais pas systématiquement* ». Ainsi, cette méthode de calcul produit des listes de protéines plus longues et, d'un point de vue biologique, plus intéressantes, car elle permet de « lisser » la variabilité technique de nos expériences. Les protéines que l'on a pu définir comme exclusives de l'une des trois bandes

à 720, 480 ou 240 kDa sont répertoriées dans le Tableau 3.1, ci-dessous.

Tableau 3.1: **Protéines exclusives par poids moléculaire**

Pour chaque protéine, le pourcentage moyen d'abondance dans la bande spécifiée est fourni sous les deux versions (moyennes totale et restreinte). N/n : nombre de fois où la protéine a été prise en compte/nombre d'expériences (n=6).

Protéine	% moy. tot	% moy. rest	N/n
Bande à 720 kDa			
SEPT9 (p 182)	100 %	100 %	6/6
DIDO1 (p 160)	100 %	100 %	6/6
TOP2A (p 187)	100 %	100 %	6/6
TCOF (p 186)	100 %	100 %	6/6
CHD4 (p 155)	100 %	100 %	6/6
TOP2B (p 187)	100 %	100 %	6/6
CHD7 (p 155)	100 %	100 %	6/6
PRKDC (p 177)	100 %	100 %	6/6
MDC1 (p 172)	98.8 %	98.8 %	6/6
SMCA5 (p 184)	93.6 %	93.6 %	6/6
BAZ1A (p 154)	91 %	91 %	6/6
AFF4 (p 153)	80 %	100 %	5/6
SEPT7 (p 182)	80 %	100 %	5/6
EHMT1 (p 162)	80 %	100 %	5/6
SEPT2 (p 182)	80 %	100 %	5/6
PRP8 (p 177)	80 %	100 %	5/6
EMSA1 (p 162)	66 %	100 %	4/6
CUL1 (p 158)	50 %	100 %	3/6
DHE4 (p 160)	50 %	100 %	3/6
RAD50 (p 177)	50 %	100 %	3/6
ADNP (p 153)	50 %	100 %	3/6
LDB1 (p 172)	50 %	100 %	3/6
SEP11 (p 182)	50 %	100 %	3/6
EHMT2 (p 162)	50 %	100 %	3/6
Bande à 480 kDa			
SF3A1 (p 182)	100 %	100 %	6/6
RD23B (p 178)	100 %	100 %	6/6

(continue)

Tableau 3.1: **Protéines exclusives par poids moléculaire (suite)**

Protéine	% moy. tot	% moy. rest	N/n
SF3A2 (p 183)	80 %	100 %	5/6
SF3A2 (p 183)	80 %	100 %	5/6
XPC (p 189)	80 %	100 %	5/6
CHD2 (p 155)	66 %	100 %	4/6
RPA1 (p 180)	66 %	100 %	4/6
PCF11 (p 175)	66 %	100 %	4/6
CLIP2 (p 156)	50 %	100 %	3/6
TBL3 (p 186)	50 %	100 %	3/6
WDR36 (p 189)	50 %	100 %	3/6
RFC4 (p 179)	50 %	100 %	3/6
RFC3 (p 179)	50 %	100 %	3/6
RPAB1 (p 181)	50 %	100 %	3/6
RFC1 (p 179)	48 %	95.3 %	3/6
RPA2 (p 181)	63 %	95.2 %	4/6
H13 (p 165)	63 %	95.2 %	4/6
RFX5 (p 179)	62 %	93.8 %	4/6
CAF1A (p 154)	61 %	92.5 %	4/6
RPAC1 (p 181)	50 %	91.8 %	3/6
MORC2 (p 172)	75 %	90.1 %	5/6

## Bande à 240 kDa

CTCF (p 158)	100 %	100 %	6/6
PDS5B (p 175)	98 %	98 %	6/6
ABCF1 (p 153)	93.1 %	93.1 %	6/6
CLAP1 (p 156)	80 %	100 %	5/6
SRRT (p 185)	80 %	100 %	5/6
NOP2 (p 174)	80 %	100 %	5/6
KHDR1 (p 171)	80 %	100 %	5/6
KIFC1 (p 164)	80 %	100 %	5/6
SMU1 (p 184)	80 %	100 %	5/6
RBM26 (p 178)	66 %	100 %	4/6
SNUT1 (p 184)	66 %	100 %	4/6
CDK7 (p 155)	66 %	100 %	4/6
RED (p 178)	66 %	100 %	4/6
TTF2 (p 188)	66 %	100 %	4/6
SAFB1 (p 181)	50 %	100 %	3/6
CARF (p 154)	50 %	100 %	3/6

(continue)

Tableau 3.1: **Protéines exclusives par poids moléculaire (suite)**

Protéine	% moy. tot	% moy. rest	N/n
DHX9 (p 160)	50 %	100 %	3/6
TAOK1 (p 186)	50 %	100 %	3/6
CC037 (p 155)	50 %	100 %	3/6
ANLN (p 153)	48 %	95.8 %	3/6
DNMT1 (p 161)	80 %	95.5 %	5/6
ERCC2 (p 163)	62 %	93.4 %	4/6
TOP1 (p 187)	74 %	92.3 %	5/6
NEUA (p 173)	73 %	91.7 %	5/6

**3.3.3.1.3 Les protéines désignées comme exclusives de la bande à 720 kDa** révèlent que les complexes auxquelles elles participent sont fonctionnellement impliqués dans la régulation de la structure chromatinienne. Citons les protéines CHD4 et CHD7 (p 155), qui interviennent dans l'acétylation des histones [Tong *et al.*, 1998]. Les protéines SMCA5 (p 184) et BAZ1A (p 154), sont décrites pour interagir l'une avec l'autre dans un complexe qui est doté d'une activité hélicase (SMCA5) et qui permet de réguler la position des nucléosomes [Collins *et al.*, 2002]. SMCA5 est impliquée indirectement dans un grand nombre de voies de régulation du métabolisme de l'ADN. Ainsi, est-elle impliquée dans le maintien de la structure pendant la réplication et, par phosphorylation de l'histone H2AX (p 192) en  $\gamma$ H2AX, réparation de l'ADN [Pessina et Lowndes, 2014]. Enfin, elle peut participer à la régulation de la transcription. D'autres protéines, comme les topoisomérases TOP2A et TOP2B (p 187) permettent de contrôler la topologie de l'ADN et sont chargées de le relâcher quand il est nécessaire de séparer les brins de la double-hélice. Il est possible que ces deux molécules reconnaissent l'extrémité libre sur notre phase comme un produit intermédiaire de réaction, car le cycle catalytique de ces enzymes comporte une étape de d'endonucléolyse dans la double-hélice d'ADN. Les protéines PRKDC (p 177)<sup>3</sup> et MDC1 (p 172) sont bien connues pour être impliquées dans la réparation de l'ADN. PRKDC est un enzyme essentiel de la voie de la réparation des cassures double-brin de l'ADN par jonction des extrémités non homologues [Tang *et al.*, 2014]. Elle y intervient dès que le dimère Ku70/Ku80 (protéines senseurs des dommages)

3. Également connue sous le nom de DNA-PKcs, pour sous-unité catalytique de la protéine kinase dépendante de l'ADN. PRKDC s'associe aux deux sous-unités régulatrices Ku70 : XRCC5 (p 190) et Ku80 : XRCC6 (p 190) (qui forment le dimère KU). L'ensemble hétérotrimérique est l'holoenzyme DNA-PK.

s'est lié aux extrémités libérées par la cassure (ce qui forme l'holoenzyme doté d'activité catalytique). MDC1 joue également un rôle de senseur des dommages de l'ADN. Elle est décrite pour interagir avec les protéines ATM (p 191), CHK1, CHK2 (p 192) et P53 (p 192), toutes impliquées dans l'arrêt du cycle cellulaire dès que la cellule détecte des dommages à l'ADN (« *checkpoint* ») et dans la voie de l'apoptose [Bencokova *et al.*, 2009; Nakanishi *et al.*, 2007; Xu et Stern, 2003]. La protéine DIDO1 (p 160) serait également impliquée dans cette voie de régulation, ce qui permet de la rattacher aux mécanismes de réponse aux dommages de l'ADN [Garcia-Domingo *et al.*, 2003]. Une telle association est cependant moins directe pour deux dernières protéines, SEPT9 (p 182) et TCOF (p 186), qui sont impliquées dans des processus de formation de microfilaments cytosquelettiques et (pour TCOF) dans le transport nucléocytoplasmique.

L'utilisation de la moyenne restreinte nous permet d'isoler de nouvelles protéines exclusives. Parmi celles-ci, remarquons encore les protéines impliquées dans la formation du cytosquelette [septines SEPT2, SEPT7 et SEP11 (p 182)]. L'on voit aussi apparaître des protéines impliquées dans d'autres mécanismes cellulaires, comme la régulation de la transcription [ADNP (p 153), LDB1 (p 172), DHE4 (p 160)]. D'autres protéines sont impliquées dans le contrôle de la progression du cycle cellulaire, une voie étroitement liée avec la prise en charge des dommages de l'ADN. C'est le cas de la protéine CUL1 (p 158), qui intervient dans l'ubiquitinylation de protéines permettant la progression du cycle cellulaire [Lovly *et al.*, 2008], mais aussi des protéines EHMT1 et EHMT2 (p 162), qui sont impliquées dans la méthylation des histones<sup>4</sup>. Enfin, cette nouvelle méthode de calcul permet également de faire ressortir la présence exclusive de RAD50 (p 177), une protéine majeure des étapes précoces de la réparation de l'ADN [Carney *et al.*, 1998].

On peut donc conclure que les protéines qui sont retrouvées dans la bande à 720 kDa sont majoritairement documentées comme intervenant plus ou moins directement dans les voies de réponse aux dommages de l'ADN. Les protéines dont nous avons discuté les fonctions ci-dessus constituent un échantillon révélateur de la complexité de ces réponses. Il est intéressant de souligner que la protéine PRKDC est la plus abondante de toutes les protéines décrites, ce qui permet de penser que les complexes situés dans cette bande sont majoritairement orientés vers la voie de réparation canonique du NHEJ.

---

4. Cette observation est intéressante et devrait faire l'objet d'une recherche plus approfondie. En effet, selon les sites de méthylation, cette dernière peut conduire à une activation ou à une inactivation de la séquence d'ADN. Or, au cours de la réparation, on peut imaginer que la chromatine doit être décondensée à l'endroit du dommage. Il serait donc intéressant de vérifier de quelle méthylation il s'agit dans le cas d'espèce.

Ces septines forment une plateforme d'interaction (« *scaffold* ») au niveau du fuseau achromatique permettant de maintenir CENPE (p 191) localisée au kinétochores et donc permettant la migration des chromosomes

**3.3.3.1.4 Les protéines désignées comme exclusives de la bande à 480 kDa** ne sont qu'au nombre de deux (moyenne totale) : SF3A1 (p 182) et RD23B (p 178). SF3A1 (p 182) est impliquée dans l'épissage. Aucune relation n'a été établie entre cette protéine et la réponse aux dommages de l'ADN, mais on sait qu'elle interagit avec SAF-A [HNRPU (p 169)], protéine impliquée dans l'épissage et dont on sait qu'elle était un substrat de PRKDC (p 177) [Britton *et al.*, 2009]. RD23B (p 178), quant à elle, est directement liée à la réparation de l'ADN, mais pas par la voie de réparation des cassures double-brin, car elle intervient dans la réparation par le mécanisme du NER, en tant que partenaire du complexe XPC (p 189) [Sugasawa *et al.*, 2009].

Il va de soi qu'une bande ne peut être caractérisée sur la base de deux identités de protéines ; cela démontre bien l'utilité du calcul de l'abondance par la voie de la moyenne restreinte. Cette méthode permet, en effet, de faire émerger deux nouvelles catégories de protéines. Bon nombre des protéines exclusives de la bande à 480 kDa sont impliquées dans le métabolisme de l'ARN, avec des fonctions variées [TBL3 (p 186), WDR36 (p 189), PCF11 (p 175), RPA1 (p 180), RPA2, RPAB1 et RPAC1 (p 181)]. Les autres protéines sont reliées à la réparation des lésions de l'ADN. Les protéines XPC (p 189), RFC4 et RFC3 (p 179) interviennent directement dans la réparation des dommages par la voie du NER [Overmeer *et al.*, 2010; Sugasawa *et al.*, 2009] ; les protéines CHD2 (p 155) et CAF1A (p 154) interviennent dans le remodelage de la chromatine en réponse aux dommages de l'ADN [Kadyrova *et al.*, 2011; Rajagopalan *et al.*, 2012].

On peut donc observer que les protéines exclusives de la bande à 480 kDa sont très liées à la réparation des cassures simple-brin de l'ADN par la voie du NER. On trouve également bon nombre de protéines spécifiques des ARN. Nous ne sommes pas certains de devoir en être fortement étonnés, étant donné que beaucoup de protéines interagissant avec l'ADN ont été reconnues récemment comme se liant à des ARN [Hudson et Ortlund, 2014].

**3.3.3.1.5 L'analyse des protéines exclusives de la bande à 240 kDa** fait ressortir trois protéines, dont deux interviennent dans la régulation de la transcription [CTCF (p 158) et ABCF1 (p 153)] et la troisième (PDS5B, p 175) est impliquée dans le contrôle de la progression du cycle cellulaire [Shintomi et Hirano, 2009]. Il est intéressant de noter que la protéine CTCF se fixe dans les cellules humaines à des endroits qui correspondent à des « *hot spot* » pour les cassures double-brin de l'ADN [Tchurikov *et al.*, 2014]. Ces « *hot spots* » sont des loci au niveau desquels les cassures double-brin ont lieu plus souvent que dans le reste du génome. Il est donc probable que la fonction de régulateur épigénétique de CTCF soit en lien avec ces structures.

L'utilisation de la moyenne restreinte fait ressortir un grand nombre de protéines, dont les fonctions dans les cellules sont diverses. Les protéines de

« *Spliceosome* » la machinerie d'épissage RED (p 178) et SMU1 (p 184) ont un rôle encore mal compris. Les protéines CLAP1 (p 156) et KIFC1 (p 164) sont associées aux microtubules et les protéines TTF2 (p 188) et DNMT1 (p 161) sont impliquées dans la régulation de la transcription. On a également pu identifier plusieurs protéines impliquées dans le contrôle de la progression du cycle cellulaire, comme NOP2 (p 174), KHDR1 (p 171) ou encore CDK7 (p 155) [Barlat *et al.*, 1997; Schneider *et al.*, 1998; Valdez *et al.*, 1992]. Enfin, deux protéines sont impliquées dans la réparation des dommages à l'ADN : ERCC2 (p 163) et TAOK1 (p 186), la première impliquée dans la réparation par NER (en tant que sous unité de l'hélicase XPD), et la seconde intervenant dans la voie de régulation de l'apoptose régie par p38 en réponse aux dommages de l'ADN [Gdowicz-Klosok *et al.*, 2013; Raman *et al.*, 2007].

Indirectement, pour DNMT1, *via* son interaction avec HDAC2 [Rountree *et al.*, 2000]

Les protéines identifiées exclusivement dans la bande à 240 kDa sont donc en grande partie des protéines intervenant dans des voies de signalisation et dans la régulation du cycle cellulaire. On observe, à l'inverse, une diminution des protéines impliquées dans la réparation de l'ADN comparée aux résultats obtenus pour les bandes à 720 et 480 kDa. On peut donc être tenté d'ordonner les bandes de la plus spécialisée pour la réparation de l'ADN à la moins spécialisée et cela en fonction de leur poids moléculaire apparent décroissant. Selon ce raisonnement, les complexes présents dans la bande à 240 kDa seraient les premiers formés et auraient pour rôle, *in vivo*, de signaler la présence de dommages.

### 3.3.3.1.6 Les protéines exclusives de la bande « aspécifique ».

Nous pouvons observer, d'après les résultats décrits ci-dessus, que les bandes migrant à des poids moléculaires apparents plus petits semblent contenir une proportion de protéines exclusives connexes avec la réparation de l'ADN plus limitée que les bandes de plus haut poids moléculaire. Cela nous a conduit à nous demander ce qu'il pouvait en être pour une bande migrant à un poids moléculaire apparent encore plus faible : la bande « aspécifique ». Rappelons les éléments dont nous disposons, quant à cette bande, qui la placent à part :

- \* elle est observée quelle que soit la phase sur laquelle est faite la purification (phase  $\diamond$  ou  $\star$ , voir § 3.2.1, page 78) ;
- \* quand elle est visualisée au bleu de Coomassie colloïdal, elle ne comigre avec aucun signal d'autoradiographie (voir § 3.2.3.1, page 84) ;
- \* après digestion aux nucléases des produits de purification, cette bande ne voit pas sa migration électrophorétique altérée (§ 3.2.3.2, page 86).

DNase I et benzonase

Les résultats de l'analyse des protéines exclusives de la bande aspécifique sont présentés dans le Tableau 3.2, page ci-contre.

Tableau 3.2: **Protéines exclusives de la bande « aspécifique »**

Pour chaque protéine, le pourcentage moyen d'abondance dans la bande « aspécifique » est fourni selon les deux moyennes, totale et restreinte. N/n : nombre de fois où la protéine a été prise en compte/nombre d'expériences (n=6).

Protéine	% moy. tot	% moy. rest	N/n
<b>Bande « aspécifique »</b>			
FOXK1 (p 164)	100 %	100 %	6/6
C1TC (p 154)	100 %	100 %	6/6
COR1C (p 156)	100 %	100 %	6/6
SMAL1 (p 183)	100 %	100 %	6/6
RLA2 (p 180)	100 %	100 %	6/6
RLA1 (p 180)	100 %	100 %	6/6
FL2D (p 163)	100 %	100 %	6/6
PARP1 (p 175)	99.7 %	99.7 %	6/6
RLA0 (p 180)	98.3 %	98.3 %	6/6
SK2L2 (p 183)	94.1 %	94.1 %	6/6
DDX23 (p 159)	93.5 %	93.5 %	6/6
NFIA (p 173)	93 %	93 %	6/6
SFPQ (p 183)	91 %	91 %	6/6
DHX9 (p 160)	90 %	90 %	6/6
CSK21 (p 157)	100 %	83.3 %	5/6
G3P (p 164)	100 %	83.3 %	5/6
IF2G (p 170)	100 %	83.3 %	5/6
IF2A (p 169)	100 %	83.3 %	5/6
T2FA (p 185)	100 %	83.3 %	5/6
CSK23 (p 157)	100 %	66.7 %	4/6
DPOG1 (p 161)	100 %	66.7 %	4/6
IF2GL (p 170)	100 %	66.7 %	4/6
CREB1 (p 157)	100 %	50 %	3/6
ATF1 (p 154)	100 %	50 %	3/6
DDX47 (p 160)	100 %	50 %	3/6
HP1B3 (p 169)	100 %	50 %	3/6
VIME (p 188)	100 %	50 %	3/6
ERCC3 (p 163)	100 %	50 %	3/6
NFIB (p 174)	90.5 %	75.4 %	5/6

Une première observation — immédiate — est qu’une grande partie des protéines de ce tableau ne sont pas impliquées dans la réparation de l’ADN. Ceci paraît, à première vue, confirmer notre sentiment, décrit précédemment, selon lequel seuls les complexes de haut poids moléculaire étaient fonctionnellement reliés à la réparation de l’ADN.

Quelques protéines paraissent absolument « hors thème » : les protéines ribosomales RLA2, RLA0 et RLA1 (p 180), que nous avons évoquées pour indiquer à quel point elles étaient abondantes (RLA1 ; voir § 3.3.2, page 99). Certaines protéines impliquées dans la régulation de la transcription sont exclusives de cette bande, comme NFIA (p 173) et NFIB (p 174) ; d’autres interviennent dans la traduction [IF2A et IF2G (p 170)]. Ainsi, la tentation est grande de confirmer que la bande aspécifique est bien telle. Cependant, quelques protéines ont retenu notre attention et paraissent justifier que l’on ne se hâte pas pour tirer des conclusions.

Les protéines PARP1 (p 175), DPOG1 (p 161) et ERCC3 (p 163) sont des protéines impliquées dans la réparation des lésions de l’ADN par la voie du BER [Orta *et al.*, 2014; Parlanti *et al.*, 2004; van Vuuren *et al.*, 1994]. La première est, de plus, une protéine centrale de la détection et de la signalisation des dommages, qui intervient aussi dans la réparation des cassures double-brin de l’ADN par la voie alternative du NHEJ [Ahel *et al.*, 2008; Soni *et al.*, 2014].

La protéine SMAL1 (p 183) intervient dans la réponse aux dommages de l’ADN, *via* sa liaison à la protéine RPA, au niveau des fourches de réplication, qu’elle bloque en présence de dommages [Bansbach *et al.*, 2009]. La protéine SFPQ (p 183), décrite initialement comme intervenant dans l’épissage, se lie à la protéine NONO (p 174) pour favoriser la réparation des cassures double-brin par la voie canonique du NHEJ, et cela en se fixant sur l’hétérodimère KU à proximité des extrémités libres de l’ADN [Bladen *et al.*, 2005]. La présence de ces cinq protéines de la physiologie de la réparation de l’ADN, remet en question les hypothèses posées d’après les résultats des précédentes expériences.

RPA se lie à l’ADN  
simple-brin

KU : XRCC5/XRCC6,  
page 190

### 3.3.3.2 ANALYSE DE LA RÉPARTITION DES PROTÉINES DE RÉPARATION DANS LES QUATRE BANDES

Dans l’ensemble des protéines que nous avons caractérisées comme exclusives dans les deux tableaux ci-dessus, un bon nombre de celles-ci sont connues pour être étroitement liées à la physiologie de la réparation de l’ADN. Nous avons alors voulu explorer comment ces protéines se répartissaient dans les quatre bandes et si ces dernières matérialisaient des voies de réparation distinctes. En effet, la répartition des différentes protéines dans les différentes bandes, pourrait révéler l’existence d’interactions protéine–protéine dans une bande donnée.

Dans l’histogramme représenté à la Figure 3.23, page ci-contre, les pro-

téines ont été triées (de gauche à droite) en fonction de leur profil de répartition dans les différentes bandes : d'abord dans la bande à 720 kDa, puis dans la bande à 480 kDa et enfin dans la bande à 240 kDa. Ces données ont été obtenues en calculant le pourcentage d'abondance selon la méthode de la moyenne « totale » (voir Figure 3.22, page 103, voie 1).

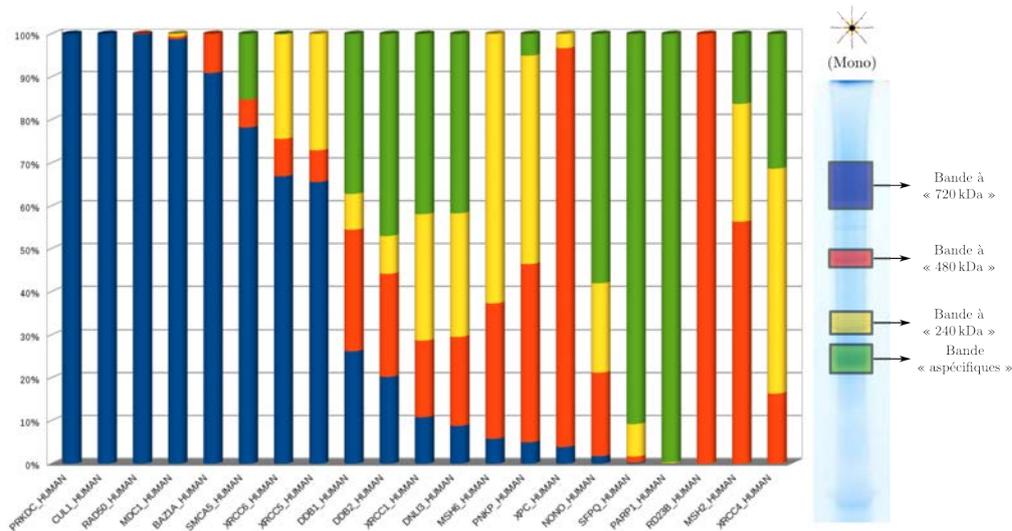


FIGURE 3.23 – Répartition d'un panel de protéines impliquées dans la réparation de l'ADN au sein de différentes bandes. Les pourcentages de répartition des protéines ont été calculés pour les bandes à 720 kDa (bleu), 480 kDa (orange), 240 kDa (jaune) et, enfin, pour la bande « aspécifique » (vert). Calcul par moyenne totale.

**3.3.3.2.1 Interactions protéine–protéine.** Comme remarque préliminaire, l'on peut indiquer que cette présentation des données a l'avantage de regrouper les protéines selon qu'elles sont prépondérantes dans telle ou telle bande. Ainsi, l'on obtient un aperçu des interactions physiques possibles entre protéines. Si l'on regarde l'histogramme de gauche à droite, l'on peut observer les protéines qui sont majoritairement présentes dans la bande à 720 kDa. De manière tout à fait intéressante, ces protéines appartiennent principalement aux voies de réparation des cassures double-brin. Les protéines XRCC6, XRCC5 et PRKDC, sous-unités de la protéine kinase dépendante de l'ADN (voie NHEJ) sont bien représentées. D'autres protéines, décrites pour interagir ensemble sont, pour la majorité, proches les unes des autres sous forme de barres contiguës, comme XRCC5 et XRCC6, BAZ1A et SMCA5, XRCC1 et DNLI3 ou encore SFPQ et NONO. Cette observation est très intéressante, car elle permet d'entrevoir que l'analyse exhaustive

des profils de migration des protéines identifiées pourrait permettre de faire ressortir des protéines interagissant entre elles mais dont l'interaction ne serait pourtant pas encore documentée.

**3.3.3.2.2 Spécialisation fonctionnelle des complexes.** Il est possible de faire une nouvelle lecture des données de l'histogramme, si l'on tire parti des fonctions connues des protéines décrites. On peut alors mettre en évidence une spécialisation fonctionnelle des complexes séparés dans les différentes bandes.

Ainsi, pouvons-nous observer que la bande de haut poids moléculaire (720 kDa) contient des complexes majoritairement orientés vers le NHEJ canonique (en bleu dans l'histogramme), alors que les autres complexes de plus petites masses ont d'autres orientations (NER, MMR). Pourtant les autres complexes « spécifiques » de plus petite masse comportent aussi le dimère XRCC5/XRCC6 (p 190) (couleurs jaune et orange des barres XRCC5 :XRCC6). Si l'on analyse le patron de coloration des barres de l'histogramme, qui est le reflet de la répartition des protéines correspondantes dans les différentes bandes, l'on peut voir que certaines ont des proportions de couleurs absolument comparables. Par exemple, les protéines XRCC1 (p 189) et DNLI3 (p 161) sont présentes dans les quatre bandes, à chaque fois exactement avec les mêmes proportions. On décèle ainsi, en lisant l'histogramme de gauche à droite, une spécialisation fonctionnelle croissante vers le BER et le NHEJ alternatif, par exemple, pour les bandes verte, jaune et rouge, alors que la composante bleue (bande à 720 kDa) diminue fortement.

Si l'on s'intéresse à nouveau aux deux barres XRCC5 et XRCC6, l'on peut constater que la répartition de ces deux protéines dans la bande rouge (480 kDa) est nettement plus faible que pour les bandes bleue et jaune (respectivement 720 kDa et 240 kDa). On aurait pu s'attendre, vu le processus de migration électrophorétique, à ce que la répartition de KU, dans les différentes bandes, soit progressivement plus limitée à chaque poids moléculaire apparent plus faible. En effet, il était parfaitement imaginable que les assemblages de haut poids moléculaire se dissocient progressivement au cours de la migration. Or, le fait que l'on ait une proportion de KU plus importante à 240 kDa qu'à 480 kDa montre que les complexes que nous séparons sur gel sont effectivement le reflet fidèle de mécanismes moléculaires d'assemblage définis. Pour renforcer encore ce concept, citons l'exemple des protéines DDB1, DDB2 (p 158) et XPC (p 189) et RD23B (p 178).

Les protéines DDB1 et DDB2 sont impliquées dans la reconnaissance des dommages dus aux UV, qui sont réparés par le mécanisme NER. De manière intéressante, ces protéines sont préférentiellement recrutées dans la bande à 480 kDa par rapport à la bande à 240 kDa, alors que les protéines XPC

et RD23B, qui sont impliquées dans la réparation à proprement parler de ces dommages, sont véritablement exclusives de cette bande à 480 kDa. On assiste donc bien à une transition fonctionnelle au sein de la bande à 480 kDa qui fait céder du terrain au NHEJ canonique à la faveur du NER. De manière analogue, PARP1 (p 175), qui est impliquée dans le NHEJ alternatif, est exclusivement présente dans la bande « aspécifique », la seule qui ne contient pas KU. Les autres protéines de cette voie de réparation [DNLI3 (p 161) et XRCC1 (p 189)] sont également de moins en moins abondantes dans les complexes séparés dans les bandes de plus haut poids moléculaire apparent.

Enfin, il peut être intéressant de rechercher les protéines qui sont systématiquement présentes dans l'ensemble des bandes obtenues par purification sur la phase ✱. En effet, l'étude des protéines qui ne sont pas des contaminants — qui sont donc absentes de la phase ✧ — serait particulièrement intéressante, car elle permettrait de mettre en évidence des mécanismes communs pour l'assemblage des différents complexes, certainement essentiels à la réparation de l'ADN. Les protéines ayant ces caractéristiques ont été identifiées et sont présentées dans le Tableau 3.4, page 129.

### 3.3.4 COMPARAISON DES BANDES OBTENUES SUR PHASE ✱/✧

Dans le processus d'analyse que nous avons décrit dans les sections précédentes, ce qui nous avait motivé était le fait de déterminer en quoi chaque bande d'une piste ✱ donnée de BN PAGE se singularisait par rapport aux autres. Nous avons montré qu'effectivement les différentes bandes étaient significativement différentes les unes des autres, car nous avons pu faire ressortir que bon nombre de protéines qui composaient une bande donnée étaient absentes des autres. Cette méthodologie d'analyse a, cependant, quelques défauts :

Bandes à 720, 480, 240 kDa et aspécifique

- \* Une limitation technique est liée au fait que la phase ✱ est constituée par un déterminant d'affinité qui comporte — en plus de son extrémité libre — la séquence « interne » de l'oligonucléotide. Or, les analyses décrites plus haut ne permettent pas de différencier les protéines recrutées sur l'une ou l'autre de ces deux parties ;
- \* Comme le fondement des méthodes d'analyse décrites jusqu'ici consistait à rechercher les protéines exclusives d'une bande donnée par rapport aux autres, nous avons éliminé toute protéine qui était présente dans plus d'une bande. Or, ce critère n'est pas le bon si l'on recherche les protéines qui sont spécifiques des extrémités libres (celles que l'on ne doit retrouver que sur la phase ✱) par différence d'avec celles qui sont recrutées sur la séquence interne et donc également sur la phase contrôle ✧.

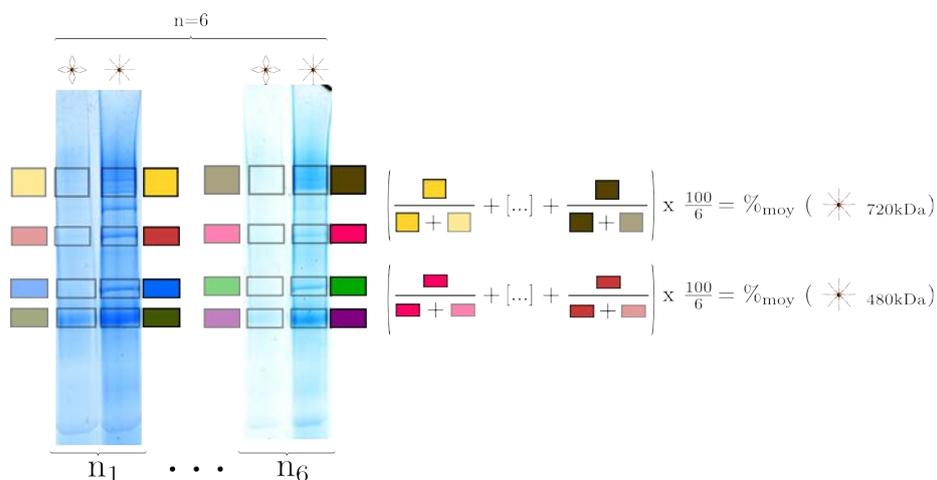


FIGURE 3.24 – Description de la méthode de calcul utilisée pour déterminer le facteur d'enrichissement d'une protéine dans une des deux phases chromatographiques. Les pourcentages de répartition moyens des protéines dans la phase présentant des extrémités libres de l'ADN ont été calculés pour chaque complexe. Un pourcentage total a ensuite pu être calculé.

Sur les phases  $\star$  et  $\star$

Comme, pour nous, une expérience complète est constituée de *deux* séparations chromatographiques, suivies d'une migration électrophorétique des produits de purifications obtenus, nous disposons d'un jeu de données d'identifications des protéines contenues dans l'ensemble des paires de bandes (voir Figure 3.18, page 98).

La méthode d'analyse des données proposée dans cette section vise donc à rechercher les protéines qui sont *spécifiques* de la phase  $\star$  et qui pourraient donc être impliquées dans les voies de réparation des cassures double-brin de l'ADN. Pour cela, nous effectuons le calcul décrit à la Figure 3.24, qui est effectué pour chaque protéine.

1. Pour une paire de bandes  $n_1$  — à un poids moléculaire donné — le ratio des valeurs d'emPAI  $\frac{\ast}{\ast+\diamond}$  est calculé et stocké. Cette étape est répétée pour toutes les paires restantes  $n_2 \dots n_6$  ;
2. La moyenne des ratios pour les paires  $n_1 \dots n_6$  obtenus à l'étape 1) est calculée ( $\%_{\text{moy}}$ ) ;
3. L'ensemble des étapes ci-dessus est alors reproduit en séquence pour les autres bandes d'un autre poids moléculaire.

Au terme des calculs décrits ci-dessus, chaque protéine se voit attribuer une *valeur numérique* caractérisant sa propension à se trouver davantage ou non dans les bandes obtenues avec la phase  $\ast$ . Le Tableau 3.3 présente les données obtenues avec cette nouvelle méthode de comparaison.

La moyenne des ratios d'abondance dans la bande  $\ast$  par rapport à la bande  $\diamond$

Tableau 3.3: **Protéines spécifiques de la phase  $\ast$**

Pour chaque protéine, le pourcentage d'abondance moyen dans une bande  $\ast$  par rapport à la bande  $\diamond$  correspondante est calculé. Les protéines présentées sont celles qui ont un pourcentage d'abondance moyen dans la bande  $\ast$  supérieur à 90 %. N/n : nombre de fois où la protéine a été prise en compte/nombre d'expériences (n=6).

Protéine	Ratio moyen	N/n
Bande à 720 kDa		
SEPT7 (p 182)	100 %	6/6
CKAP5 (p 156)	100 %	6/6
DYN2 (p 161)	100 %	6/6
MSH6 (p 173)	100 %	4/6
PRC1 (p 176)	100 %	4/6
CUL1 (p 158)	100 %	3/6
NFIB (p 174)	100 %	3/6
ACINU (p 153)	100 %	3/6
KIF2C (p 171)	100 %	3/6
PP1A (p 176)	100 %	3/6
SEPT2 (p 182)	97 %	5/6
H4 (p 168)	96 %	6/6
PP1RA (p 176)	95 %	6/6
TERF2 (p 186)	95 %	5/6
XRCC5 (p 190)	94 %	6/6

(continue)

Tableau 3.3: Protéines spécifiques de la phase ✱ (suite)

Protéine	Ratio moyen	N/n
XRCC6 (p 190)	94 %	6/6
LDB1 (p 172)	93 %	5/6
SEP11 (p 182)	93 %	3/6
PAR14 (p 174)	93 %	5/6
PRP8 (p 177)	93 %	5/6
DDB1 (p 158)	92 %	6/6
PRKDC (p 177)	91 %	6/6
EHMT2 (p 162)	91 %	3/6
EMSA1 (p 162)	91 %	5/6
RAD50 (p 177)	90 %	3/6
Bande à 480 kDa		
DKC1 (p 160)	100 %	3/6
PWP2 (p 177)	100 %	3/6
TBL3 (p 186)	100 %	3/6
RU17 (p 181)	100 %	3/6
WIZ (p 189)	100 %	3/6
CPSF5 (p 157)	100 %	3/6
H15 (p 165)	100 %	3/6
CHD2 (p 155)	100 %	4/6
RFC4 (p 179)	100 %	4/6
RFC3 (p 179)	100 %	4/6
CKAP5 (p 156)	100 %	4/6
U520 (p 188)	100 %	4/6
PCF11 (p 175)	100 %	4/6
MSH6 (p 173)	100 %	5/6
DYN2 (p 161)	100 %	6/6
MSH2 (p 172)	96 %	5/6
PAR14 (p 174)	96 %	4/6
XRCC5 (p 190)	95 %	6/6
XRCC6 (p 190)	94 %	6/6
ABCF1 (p 153)	93 %	4/6
WDR36 (p 189)	93 %	3/6
DDB2 (p 158)	93 %	5/6
PP1RA (p 176)	92 %	4/6
SFPQ (p 183)	91 %	3/6
DDX5 (p 159)	90 %	5/6

(continue)

Tableau 3.3: Protéines spécifiques de la phase \* (suite)

Protéine	Ratio moyen	N/n
DDB1 (p 158)	90 %	6/6
Bande à 240 kDa		
KIFC1 (p 164)	100 %	6/6
MNT (p 172)	100 %	5/6
CC037 (p 155)	100 %	5/6
RBM26 (p 178)	100 %	4/6
ILF3 (p 170)	100 %	4/6
DYN2 (p 161)	100 %	4/6
H2B1H (p 166)	100 %	3/6
H2B1D (p 166)	100 %	3/6
H2B1C (p 165)	100 %	3/6
NFIB (p 174)	100 %	3/6
CARF (p 154)	100 %	3/6
ILF2 (p 170)	100 %	3/6
H2BFS (p 168)	100 %	3/6
RL13 (p 180)	100 %	3/6
RLA0 (p 180)	100 %	3/6
HDAC2 (p 168)	100 %	3/6
WIZ (p 189)	100 %	3/6
DHX9 (p 160)	100 %	3/6
EXOSX (p 163)	100 %	3/6
POTE1 (p 176)	100 %	3/6
CPSF7 (p 157)	100 %	3/6
H2B2F (p 168)	100 %	3/6
H2B1N (p 167)	100 %	3/6
H2B1M (p 167)	100 %	3/6
H2B1L (p 166)	100 %	3/6
SK2L2 (p 183)	100 %	3/6
H2B1K (p 166)	100 %	3/6
RBM15 (p 178)	100 %	3/6
TF2H4 (p 187)	100 %	3/6
XRCC6 (p 190)	97 %	6/6
DDX5 (p 159)	97 %	6/6
NOP2 (p 174)	97 %	5/6
TR150 (p 188)	96 %	4/6
ANLN (p 153)	95 %	3/6
PRC1 (p 176)	95 %	5/6

(continue)

Tableau 3.3: **Protéines spécifiques de la phase \*** (suite)

Protéine	Ratio moyen	N/n
XRCC5 (p 190)	95 %	6/6
PP1RA (p 176)	94 %	5/6
CLAP1 (p 156)	93 %	5/6
IMDH1 (p 171)	92 %	3/6
MSH6 (p 173)	91 %	6/6
MSH2 (p 172)	90 %	4/6
SNUT1 (p 184)	90 %	4/6
Bande « aspécifique »		
FOSL2 (p 164)	100 %	3/6
MSH2 (p 172)	100 %	3/6
CSK23 (p 157)	100 %	4/6
DDX3X (p 159)	100 %	4/6
H2B1H (p 166)	100 %	4/6
H2B1D (p 166)	100 %	4/6
H2B1C (p 165)	100 %	4/6
H2B1B (p 165)	100 %	4/6
ILF3 (p 170)	100 %	4/6
H2BFS (p 168)	100 %	4/6
T2FB (p 185)	100 %	4/6
H2B2F (p 168)	100 %	4/6
H2B2E (p 167)	100 %	4/6
H2B1O (p 167)	100 %	4/6
H2B1N (p 167)	100 %	4/6
H2B1M (p 167)	100 %	4/6
H2B1L (p 166)	100 %	4/6
H2B1K (p 166)	100 %	4/6
H2B1J (p 166)	100 %	4/6
H2B3B (p 168)	100 %	4/6
CSK21 (p 157)	100 %	5/6
ILF2 (p 170)	100 %	5/6
G3P (p 164)	100 %	5/6
DPOG1 (p 161)	100 %	5/6
DDX47 (p 160)	100 %	5/6
SMCA5 (p 184)	100 %	5/6
DYN2 (p 161)	100 %	6/6
ATF1 (p 154)	98 %	4/6
FL2D (p 163)	94 %	6/6
(continue)		

Tableau 3.3: Protéines spécifiques de la phase ✱ (suite)

Protéine	Ratio moyen	N/n
RBM15 (p 178)	92 %	6/6
DDB2 (p 158)	91 %	6/6
DHX9 (p 160)	91 %	6/6
DDB1 (p 158)	90 %	6/6

### 3.3.4.1 PROTÉINES SPÉCIFIQUES DE LA PHASE ✱, BANDE À 720 kDa

La comparaison des protéines recrutées sur les phases ✱ et ✧, pour la bande à 720 kDa met en évidence des protéines bien connues pour intervenir dans la réparation de l'ADN. C'est le cas, par exemple, des protéines PRKDC (p 177), XRCC5 et XRCC6 (p 190), qui interviennent dans la voie de réparation des cassures double-brin de l'ADN par la voie du NHEJ canonique. L'analyse des résultats obtenus pour la bande à 720 kDa montre, d'une part un recoupement avec ceux obtenus en utilisant l'ancienne approche [la protéine RAD50 (p 177), par exemple, est identifiée dans les deux cas] et, d'autre part que de nouvelles protéines d'intérêt peuvent ressortir, comme les protéines MSH6 (p 173), DDB1 (p 158), PP1A, et PP1RA (p 176) qui ne ressortaient pas dans l'étude précédente. La protéine DDB1 (p 158) agit en tant qu'inducteur de la réparation des dommages induits par les rayonnements ultraviolets, qu'elle contribue à détecter [Sugasawa *et al.*, 2005]. Les protéines PP1A et PP1RA (p 176) sont impliquées dans la signalisation des dommages à l'ADN et sont activées par la protéine ATM [Mi *et al.*, 2007].

Plus généralement, parmi les 24 protéines apparaissant comme spécifiques dans la bande à 720 kDa, seules 9 d'entre elles étaient décrites comme étant exclusives. Ces nombres montrent l'intérêt de notre nouvelle stratégie d'analyse et paraissent attester qu'une partie importante des protéines impliquées dans la réparation des cassures de l'ADN recrutées sur la phase ✱ est commune à au moins deux bandes.

Il est intéressant de remarquer que la protéine TERF2 (p 186) est présente spécifiquement sur la phase ✱, car, bien qu'elle ne soit pas impliquée dans la réparation des cassures double-brin de l'ADN, elle en reconnaît les extrémités libres. TERF2 est recrutée au niveau des télomères, où elle joue un rôle important dans la prévention des dommages liés au stress répliatif, et cela en interaction avec TOP2A (p 187) [Ye *et al.*, 2010].

Il est également important de remarquer que bon nombre des protéines qui ressortent avec ce nouveau calcul sont impliquées dans des mécanismes de transport : les protéines DYN2 (p 161), CKAP5 (p 156), KIF2C (p 171)

et PRC1 (p 176) sont des protéines associées aux microtubules [Cassimeris et Morabito, 2004; Diatloff-Zito *et al.*, 1995; Zhu et Jiang, 2005]. La présence des septines 2, 7 et 11, préalablement présentées comme exclusives de la bande à 720 kDa, renforce l'impression selon laquelle une « mécanique » moléculaire encore à identifier formerait un lien entre les dommages à l'ADN et le système cytosquelettique en général, peut-être en relation avec la nécessité d'induire le « *checkpoint* » mitotique.

Il semble que EMSA1 possède un domaine de liaison à l'ADN  
<http://www.ebi.ac.uk/interpro/entry/IPR000949>

Il serait intéressant d'étudier la protéine EMSA1 (p 162), vu que sa fonction n'est pas encore documentée et que nous ne disposons que de très peu d'informations à son sujet. La protéine PRP8 (p 177) est également un candidat intéressant pour la réparation de l'ADN, car elle est impliquée dans l'épissage. Or, nous avons vu que beaucoup de protéines connues initialement pour être impliquée dans cette fonction cellulaire ont ensuite été caractérisées comme participant de la réparation de l'ADN [SFPQ (p 183), par exemple]. Les autres protéines identifiées sont impliquées dans la régulation de la transcription [NFIB (p 174) et LDB1 (p 172)], dans la structure de la double-hélice d'ADN [H4 (p 168), EHMT2 (p 162), ainsi que ACINU (p 153), dans des conditions particulières Sahara *et al.* [1999]], dans des voies de signalisation [PAR14 (p 174)] et dans la régulation de la progression du cycle cellulaire (CUL1 (p 158) et PRC1 (p 176)).

Il ressort donc de ces analyses que la comparaison des quantités de protéine entre la phase  $\star$  et la phase  $\diamond$  montre bien que les complexes protéiques assemblés sur la première de ces deux phases sont fonctionnellement plus orientés vers la reconnaissance et la réparation des dommages de l'ADN. Il sera intéressant de comprendre pour quelles raisons de si nombreuses protéines associées aux fonctions cytosquelettiques (microfilaments, microtubules et moteurs moléculaires) sont également aussi fortement spécifiques de la phase  $\star$  pour la bande à 720 kDa.

#### 3.3.4.2 PROTÉINES SPÉCIFIQUES DE LA PHASE $\star$ , BANDE À 480 kDa

La comparaison des protéines recrutées sur les phases  $\star$  et  $\diamond$ , pour la bande à 480 kDa met en évidence 20 protéines qui n'avaient pas été discutées lors de la caractérisation de ce complexe par détermination des protéines exclusives. Six protéines spécifiques de la phase  $\star$ , au contraire, sont également exclusives de cette bande (CHD2, RFC4, RFC3, PCF11, WDR36 et DDX5) et 8 protéines ont déjà été identifiées comme étant spécifiques de la phase  $\star$  pour la bande à 720 kDa (CKAP5, MSH6, DYN2, PAR14, XRCC5, XRCC6, PP1RA, SFPQ et DDB1).

Les protéines impliquées dans la réponse aux dommages de l'ADN sont également très nombreuses au sein de cette liste : les protéines DDB1 et DDB2 (p 158) (voie de réparation par excision de nucléotide, NER) sont connues pour permettre la détection des dommages à l'ADN induits par les

rayonnements ultraviolets. Ces protéines, en interaction avec RBX1, permettraient d'augmenter l'accessibilité de l'ADN aux protéines de réparation en provoquant l'ubiquitinylation des histones par l'E3 ligase CUL4A/CUL4B. Les protéines RFC3 et RFC4 (p 179), qui sont également impliquées dans la voie de réparation NER, apparaissent elles aussi comme exclusives de la phase  $\star$ . Les protéines MSH2 et MSH6 (p 173), qui interagissent entre elles pour former MutS $\alpha$ , sont impliquées dans la reconnaissance des mésappariements devant être pris en charge par le mécanisme du MMR, en particulier après la réplication [Iaccarino *et al.*, 1998]. Les protéines XRCC5 et XRCC6 (p 190) sont les protéines de réparation caractéristiques du NHEJ canonique et interagissent avec SFPQ (p 183) au niveau des extrémités de l'ADN. Le rôle de la protéine PP1RA (p 176) dans la signalisation des dommages de l'ADN (en interaction avec la protéine ATM) a déjà été discuté ci-dessus; l'hélicase DDX5 (p 159), qui est également une protéine exclusive de la bande à 480 kDa, possède elle aussi un rôle de signalisation, spécifiquement dans l'induction de l'apoptose en réponse aux dommages de l'ADN [Rossow et Janknecht, 2003]. La protéine CHD2 (p 155) joue un rôle dans l'accessibilité des régions d'ADN endommagées aux complexes de réparation, ce qui pourrait également être le rôle de la protéine WIZ (p 189) qui intervient dans la méthylation des histones [Ueda *et al.*, 2006].

La comparaison des protéines purifiées sur la phase  $\star$  par rapport à celles obtenues sur phase  $\diamond$  produit, dans le cas de la bande à 480 kDa, une série de protéines impliquées dans le métabolisme de l'ARN. Ce résultat est frappant et les protéines ainsi isolées sont regroupées en deux catégories :

- \* les protéines de ce groupe se confondent en partie avec la liste des protéines impliquées dans la réparation de l'ADN. On y retrouve notamment les protéines SFPQ (p 183) et DDX5 (p 159), mais aussi d'autres protéines comme U520 (p 188) qui est présente dans la snRNP U5 au même titre que PRP8 (p 177) (spécifique de la phase  $\star$  dans la bande à 720 kDa). La protéine RU17 (p 181) fait également partie du spliceosome et participe à la formation de la snRNP U1. Les protéines PCF11 (p 175) et CPSF5 (p 157) font partie de complexes de clivage des pré-ARNm lors de l'épissage et contrôlent également la polyadénylation de l'ARNm [Kim *et al.*, 2010].
- \* les protéines impliquées dans la voie de biogenèse des ribosomes, telles que les protéines DKC1 (p 160), PWP2 (p 177) et ABCF1 (p 153) [Paytubi *et al.*, 2009] semblent également spécifiques de la phase  $\star$ . Cette observation paraît étonnante, mais ce résultat est renforcé par la présence exclusive de la protéine TBL3 (p 186), qui interagit également avec les ARN et qui est, de plus, impliquée chez la levure dans la régulation du cycle cellulaire *via* le contrôle de la quantité d'ARNr

[Dasil et Bustelo, 2004; Dragon *et al.*, 2002].

La présence d'un grand nombre de protéines intervenant dans la physiologie des ARN et, en particulier, pour l'épissage des ARNm et pour la voie de la biogenèse des ribosomes, semble donc être un trait particulièrement spécifique de la phase  $\star$  par rapport à la phase  $\diamond$  pour la bande à 480 kDa. Il est cependant intéressant de noter que l'on retrouve de manière également spécifique dans la bande  $\star$  de ce poids moléculaire des protéines connexes avec les fonctions cytosquelettiques de la cellule qui étaient déjà ressorties lors des calculs pour la bande à 720 kDa [les protéines DYN2 (p 161), CKAP5 (p 156)] et un variant de l'histone H1, la protéine H15 (p 165). Ceci est également le cas pour l'une des isoformes de PARP, la protéine PAR14 (p 174), qui est impliquée dans des voies de régulation grâce à son activité de parylation [Cho *et al.*, 2011].

### 3.3.4.3 PROTÉINES SPÉCIFIQUES DE LA PHASE $\star$ , BANDE À 240 kDa

La comparaison des protéines recrutées sur les phases  $\star$  et  $\diamond$ , pour la bande à 480 kDa met en évidence un nombre encore plus important de protéines n'ayant pas été décrites comme exclusives de cette bande. L'apparition d'un grand nombre d'histones (9 variants différents de l'histone H2B), ainsi que la présence de protéines directement impliquées dans les mécanismes de remodelage de la chromatine [comme HDAC2 (p 168)] soulève la question de la spécificité observée. Il en va de même pour la protéine POTE1 (p 176), qui fait partie du complexe de protection des télomères et reconnaît donc spécifiquement les extrémités libres de l'ADN en ce qu'elles miment les télomères. Il est probable que cette protéine ait été recrutée par TERF2 (p 186) dont on sait qu'elle est exclusive de la bande à 240 kDa. L'on retrouve également bon nombre de protéines impliquées dans la réparation de l'ADN, comme l'hétérodimère KU, MSH2 et MSH6. Nous pouvons cependant remarquer que ces protéines ont déjà été identifiées comme étant spécifiques de la phase  $\star$  lors de l'analyse des bandes de plus haut poids moléculaire, ce qui laisserait supposer que les complexes de plus bas poids moléculaires sont des complexes moins spécialisés. Il est intéressant cependant de noter la spécificité de la protéine CARF (p 154), impliquée dans les étapes précoces de la réponse aux dommages de l'ADN et exclusive de la bande à 240 kDa [Cheung *et al.*, 2014; Reinhardt *et al.*, 2010].

De la même manière que lors de l'analyse des données relatives aux bandes à 480 kDa, de nombreuses protéines de la physiologie de l'ARN sont identifiées comme étant spécifiques de la phase  $\star$  à 240 kDa. De manière rassurante, on fait également ressortir des protéines décrites pour interagir les unes avec les autres, comme le couple CPSF5 et CPSF7 (p 157), qui intervient dans le pré-clivage des ARNm. Le moteur moléculaire DYN2 (p 161) et

la protéine de structure PRC1 (p 176), toutes deux fonctionnant sur les microtubules, font également partie des protéines soulignées par l'analyse. Or, ces deux protéines étaient déjà ressorties des études des bandes à 720 et 480 kDa, auxquelles s'ajoutent trois autres protéines également impliquées dans la physiologie du cytosquelette [CLAP1 (p 156), KIFC1 (p 164) et ANLN (p 153)], et qui sont, de plus, exclusives de la bande à 240 kDa.

Le nombre de protéines qui ressortent grâce à cette dernière étude, ainsi que la variété croissante des voies physiologiques dans lesquelles elles interviennent (épissage, biogenèse des ribosomes, réparation de l'ADN, régulation de la transcription, organisation du cytosquelette, contrôle du cycle cellulaire, organisation de la chromatine et voies métaboliques) fragilisent l'idée que nous nous faisons d'une interaction absolument spécifique avec les extrémités libres de l'ADN de la phase ✱. Nous nous sommes alors tournés vers l'analyse des données obtenues avec les bandes « aspécifiques », dans l'espoir de ne pas y trouver de protéine de réparation de l'ADN, ce qui en ferait une nouvelle bande contrôle. En effet, les protéines qui s'y trouveraient seraient alors susceptibles d'être éliminées des jeux de protéines identifiées plus haut.

#### 3.3.4.4 PROTÉINES SPÉCIFIQUES DE LA PHASE ✱, BANDE ASPÉCIFIQUE

La comparaison des protéines recrutées sur les phases ✱ et ✧, pour la bande aspécifique fait de nouveau ressortir un nombre important d'histones (14 variants de l'histone H2B) ainsi que des protéines qui interviennent dans le remodelage de la chromatine, comme la protéine SMCA5 (p 184). Cette analyse permet également d'observer que la protéine DYN2 (p 161) est toujours spécifique de la phase ✱.

Il est intéressant de noter, en revanche, l'absence des protéines XRCC5 (p 190) et XRCC6 (p 190), ce qui permet d'envisager d'utiliser le contenu de cette bande pour éliminer toutes les protéines qui n'interviennent pas dans la voie de réparation du NHEJ canonique. Il serait néanmoins dommage de ne pas considérer certaines protéines qui semblent dignes d'intérêt, qui sont ressorties lors de cette dernière étude. Notons les kinases CSK21 (p 157) et CSK23 (p 157), par exemple, qui sont impliquées dans la régulation du cycle cellulaire, en particulier car elles participent à la mise en place du « *checkpoint* » relatif au fuseau achromatique. Malgré l'absence des deux sous-unités de la protéine KU, nous avons la protéine G3P (p 164) (*GAPDH*), qui est décrite pour interagir avec PRKDC (p 177) [Kornberg *et al.*, 2010]. Enfin, ILF3 (p 170) pourrait permettre la liaison stable de PRKDC sur l'ADN.

Glycéraldéhyde-3-phosphate déshydrogénase

### 3.3.5 CONCLUSIONS SUR LA FOUILLE DES DONNÉES

Au cours des analyses effectuées jusqu'ici, nous avons tenté de quantifier deux caractéristiques principales pour chaque protéine identifiée dans nos échantillons :

- \* la protéine est-elle préférentiellement dans une bande à un poids moléculaire donné pour un échantillon \* donné ?
- \* la protéine est-elle préférentiellement recrutée sur la phase \* par rapport à la phase \* ?

#### 3.3.5.1 RÉPARTITION DES PROTÉINES DANS LES BANDES À DIFFÉRENTS POIDS MOLÉCULAIRES (PHASE \*)

Les approches choisies pour cette étape de la fouille des données font usage uniquement des identifications de protéines purifiées sur la phase qui mime les cassures double-brin de l'ADN (\*).

**3.3.5.1.1 Caractère significatif des assemblages non covalents tels qu'ils sont résolus sur gel.** Nous avons pu observer plus haut, par électrophorèse en deux dimensions, que les produits de purification résolus sur BN PAGE étaient séparés ultérieurement sur gel dénaturant en leurs différents constituants. Ce résultat nous avait permis de conclure que ces produits de purification étaient des complexes multiprotéiques. Cette conclusion a été confirmée grâce aux analyses par spectrométrie de masse du contenu des différentes bandes de BN PAGE.

On constate, en effet, que bon nombre de protéines se retrouvent dans des bandes migrant à des poids moléculaires apparents largement supérieurs à leur propre masse moléculaire. Par exemple, le cas de la protéine PRKDC (p 177), présente dans la bande à 720 kDa est particulièrement intéressant. Ayant une taille de 470 kDa, cette protéine pourrait être identifiée dans la bande à 480 kDa — si elle était séparée de ses partenaires d'interaction lors de la migration électrophorétique — cependant elle n'y est jamais identifiée. Ce résultat montre que nos conditions expérimentales conservent bien les interactions non covalentes les plus significatives entre les protéines.

**3.3.5.1.2 Recherche de protéines « exclusives » d'une bande donnée.** C'est la première approche que nous avons choisie dans le but d'établir dans quelle mesure les complexes contenus dans telle ou telle bande étaient fonctionnellement différents. En effet, certaines protéines sont bien connues pour constituer des signatures de chemins physiologiques particuliers (réparation de l'ADN, régulation du cycle cellulaire, cytosquelette et moteurs moléculaires, par exemple).

Les résultats de la mise en œuvre de cette approche sont résumés ci-dessous, dans la colonne « Exclusivité » du Tableau 3.5, page 130. Si l'on regarde la composition en protéines spécifiques de la réparation de l'ADN dans les trois grandes bandes, à 240 kDa, 480 kDa et 720 kDa, l'on peut constater que les protéines identifiées n'y sont pas uniformément réparties, avec une forte prédominance pour les complexes de haute masse, pour lesquels la voie NHEJ canonique semble bien représentée. Les complexes migrant à 480 kDa se distinguent de ceux à 720 kDa par l'absence totale de PRKDC (p 177). En revanche, à 480 kDa, certaines protéines de la voie de réparation NER sont présentes de manière exclusive, ce qui donne à penser qu'il y a une spécialisation fonctionnelle des complexes contenus dans cette bande. Pour ce qui est de la bande à 240 kDa, la majorité des protéines exclusives paraît impliquée dans la régulation de la progression du cycle cellulaire (cyclines, « *checkpoint* »). Il faut noter que le dimère KU est systématiquement présent dans ces trois bandes (Figure 3.23, page 113), malgré le fait que seule la bande à 720 kDa paraisse représentative de la voie NHEJ. On perçoit alors, que la recherche de protéines exclusives de telle ou telle bande constitue une approche un peu limitée, car elle ne permet pas d'embrasser l'ensemble des voies fonctionnelles représentées dans les différentes bandes. Ainsi, avons-nous opté pour une stratégie plus large, qui consiste à mesurer l'abondance relative (la répartition) des protéines de réparation de l'ADN dans l'ensemble des bandes des pistes ✱.

**3.3.5.1.3 Abondance relative des protéines de réparation dans les différentes bandes des pistes ✱.** Cette approche analytique s'est traduite par l'établissement de l'histogramme à la Figure 3.23, page 113, qu'il est possible d'analyser selon deux modes produisant des conclusions complémentaires :

- \* Nous avons pu montrer que des protéines connues pour interagir physiquement entre elles avaient des profils de répartition très proches (sinon identiques) dans les différentes bandes ;
- \* Nous avons pu montrer la spécialisation fonctionnelle des complexes séparés dans les quatre bandes résolues sur gel natif. La bande migrant à 720 kDa est indéniablement orientée vers la réparation des cassures double-brin par la voie NHEJ canonique, alors que les bandes de plus bas poids moléculaires sont orientées vers les voies NER, BER ou NHEJ alternatif.

Ce qui se traduit par le fait que les trois couleurs de chacune des barres d'histogramme correspondantes ont la même dimension

### 3.3.5.2 RÉPARTITION DES PROTÉINES DANS LES PRODUITS DE PURIFICATION SUR PHASES \* ET † (PAR POIDS MOLÉCULAIRE)

L'approche choisie pour cette étape de la fouille des données a consisté à tirer partie du fait que chaque expérience que nous menions était constituée de deux expériences : l'une faite sur la phase \* et l'autre faite sur la phase contrôle (†). Nous avons voulu quantifier la propension des protéines à se lier à une phase plutôt qu'à l'autre, ce qui aurait dû nous aider à faire ressortir des protéines spécifiques des extrémités libres de l'ADN.

La fouille des données a permis de montrer que les protéines de réparation des cassures double-brin de l'ADN avaient une préférence absolue pour la phase mimant ces cassures (phase \*), puisqu'elles étaient systématiquement absentes de l'autre phase. Par ailleurs, deux classes de protéines de réparation se singularisent : les protéines de la voie NER (dans la bande à 480 kDa) et les protéines de la voie BER (dans la bande aspécifique), dont nous montrons qu'elles n'ont pas de préférence particulière pour la phase \*. Si l'on s'intéresse, maintenant, aux protéines « senseur » des dommages à l'ADN, nous montrons qu'elles sont préférentiellement recrutées sur la phase \*, à l'instar de protéines impliquées dans le métabolisme des nucléotides [IMDH1, IMDH2 (p 171) et G3P (p 164)] ou dans la régulation du cycle cellulaire (cyclines, différents « checkpoints »). Enfin, mentionnons les protéines que l'on pourrait qualifier « de structure », comme les histones, que l'on ne retrouve que sur la phase à extrémités libres, précisément car elles requièrent des degrés de liberté au niveau de l'ADN pour pouvoir s'y lier.

Il est toute une classe de protéines pour lesquelles l'interprétation de leur préférence pour l'une ou l'autre des deux phases chromatographiques pose un problème : la classe des protéines apparaissant n'avoir pas de spécificité particulière. Certaines de ces protéines pourraient avoir été purifiées en raison d'interactions avec d'autres protéines de la physiologie générale de l'ADN ; d'autres protéines pourraient avoir des interactions avec des partenaires des complexes de réparation d'intérêt. Dans les deux cas, les fonctions connues de ces protéines ne paraissent pas correspondre à des tâches de réparation de l'ADN. Pourtant, il est possible, intellectuellement, d'établir des connexions entre ces protéines. L'on songe, par exemple, aux protéines de signalisation/régulation de la progression du cycle cellulaire ou encore aux protéines du cytosquelette.

Si l'on considère l'ensemble des bandes aux différents poids moléculaires, quelques protéines d'intérêt majeur ressortent des analyses, qui sont présentées dans le Tableau 3.4.

La présence inattendue de DYN2 (p 161) en tête de tableau montre bien que cette protéine était presque systématiquement (22 fois sur 24) et exclusivement (100 %) recrutée sur la phase \*. D'autres protéines se trouvaient dans une situation analogue mais, pour celles-ci, ce résultat était nettement plus attendu.

Protéine	Ratio moyen $\ast/\ast$	Nombre d'identifications
DYN2 (p 161)	100 %	22/24
XRCC5 (p 190)	95 %	18/24
DDX5 (p 159)	90 %	20/24
XRCC6 (p 190)	90 %	21/24
SFPQ (p 183)	88 %	23/24
DDB1 (p 158)	85 %	22/24
TERF2 (p 186)	85 %	18/24
DDB2 (p 158)	84 %	20/24

Tableau 3.4 – **Protéines spécifiquement recrutées sur la phase  $\ast$  (tous poids moléculaires confondus)**. Les protéines décrites ont un pourcentage moyen d'abondance pour la phase  $\ast \geq 84$  %.

### 3.3.5.3 TABLEAU RÉCAPITULATIF DE TOUTES LES DONNÉES DE MASSE

Le Tableau 3.5, page suivante présente une vue synoptique de l'ensemble des données d'intérêt obtenues par l'analyse du contenu des bandes de gels natifs par spectrométrie de masse. La colonne « Exclusivité » comporte le nom des protéines qui n'ont été identifiées que dans telle ou telle bande d'un poids moléculaire donné et sur phase chromatographique  $\ast$ . La colonne « Spécificité » comporte le nom des protéines qui sont spécifiquement recrutées sur la phase  $\ast$ . La colonne « MM » précise la masse moléculaire des protéines se trouvant dans la première colonne. On peut remarquer d'emblée que si l'on fait la somme des masses de toutes les protéines identifiées dans une bande donnée, cette masse dépasse de beaucoup le poids moléculaire apparent de la bande. Il s'agit là de la manifestation expérimentale dans nos conditions analytiques de ce qui est reconnu comme une des grandes difficultés dans le domaine de l'assemblage dynamique et prétendument ordonné des complexes de réparation. Si l'on prend, par exemple, la masse de toutes les protéines connues pour interagir physiquement avec les protéines Ku70/Ku80, leur somme serait de plusieurs mégaDaltons et l'on peine à croire que l'ensemble de ces protéines puisse être recruté par KU à un instant donné [Grundy *et al.*, 2014]. Si, dans ce contexte, l'on cherchait dans nos données un cas de figure illustrant cette difficulté, l'on citerait l'exemple de la bande à 720 kDa, dans laquelle PRKDC (p 177) est identifiée de manière exclusive et dans laquelle la protéine SFPQ (p 183) est spécifiquement absente.

PRKDC n'est jamais présente dans les autres bandes

**Tableau 3.5: Résumé général des protéines d'intérêt**

Ce tableau regroupe les informations utiles pour toutes les protéines d'intérêt qui ont été extraites de la masse des données par les différentes méthodes d'analyse décrites.

Exclusivité	MM (kDa)	Spécificité *
<b>Bande à 720 kDa</b>		
PRKDC (p 177)	470kDa	PRKDC
SEPT7 (p 182)	50kDa	SEPT7
SEPT2 (p 182)	41kDa	SEPT2
PRP8 (p 177)	273kDa	PRP8
AFF4 (p 153)	127kDa	AFF4
EMSA1 (p 162)	115kDa	EMSA1
SEP11 (p 182)	49kDa	SEP11
RAD50 (p 177)	154kDa	RAD50
CUL1 (p 158)	90kDa	CUL1
EHMT2 (p 162)	132kDa	EHMT2
SEPT9 (p 182)	65kDa	H4 (p 168)
DIDO1 (p 160)	244kDa	DYN2 (p 161)
TOP2A (p 187)	174kDa	CKAP5 (p 156)
TCOF (p 186)	152kDa	TERF2 (p 186)
CHD4 (p 155)	218kDa	MSH6 (p 173)
TOP2B (p 187)	183kDa	DDB2 (p 158)
CHD7 (p 155)	336kDa	PP1RA (p 176)
EHMT1 (p 162)	141kDa	XRCC5 (p 190)
DHE4 (p 160)	61kDa	XRCC6 (p 190)
ADNP (p 153)	123kDa	DDB1 (p 158)
LDB1 (p 172)	47kDa	SSBP (p 185)
MDC1 (p 172)	227kDa	IMDH1 (p 171)
SMCA5 (p 184)	122kDa	
BAZ1A (p 154)	179kDa	
<b>Bande à 480 kDa</b>		
CHD2 (p 155)	211kDa	CHD2
PCF11 (p 175)	173kDa	PCF11
TBL3 (p 186)	89kDa	TBL3
WDR36 (p 189)	105kDa	WDR36

(continue)

Tableau 3.5: **Résumé général des protéines d'intérêt (suite)**

Exclusivité	MM (kDa)	Spécificité *
RFC4 (p 179)	40kDa	RFC4
RFC3 (p 179)	41kDa	RFC3
CAF1A (p 154)	107kDa	CAF1A
SF3A1 (p 182)	89kDa	IMDH1 (p 171)
RD23B (p 178)	43kDa	DYN2 (p 161)
SF3A2 (p 183)	49kDa	MSH6 (p 173)
XPC (p 189)	106kDa	CKAP5 (p 156)
RPA1 (p 180)	195kDa	DDX5 (p 159)
CLIP2 (p 156)	116kDa	NONO (p 174)
RPAB1 (p 181)	25kDa	U520 (p 188)
RFC1 (p 179)	128kDa	TOX4 (p 187)
RPA2 (p 181)	128kDa	ABCF1 (p 153)
H13 (p 165)	22kDa	XRCC5 (p 190)
RFX5 (p 179)	65kDa	MSH2 (p 172)
RPAC1 (p 181)	39kDa	PAR14 (p 174)
MORC2 (p 172)	118kDa	XRCC6 (p 190) SFPQ (p 183)
Bande à 240 kDa		
NOP2 (p 174)	89kDa	NOP2
KIFC1 (p 164)	123kDa	KIFC1
CLAP1 (p 156)	169kDa	CLAP1
SMU1 (p 184)	58kDa	SMU1
CDK7 (p 155)	39kDa	CDK7
SNUT1 (p 184)	90kDa	SNUT1
RED (p 178)	66kDa	RED
DHX9 (p 160)	141kDa	DHX9
CARF (p 154)	61kDa	CARF
CC037 (p 155)	41kDa	CC037
SAFB1 (p 181)	103kDa	SAFB1
ANLN (p 153)	124kDa	ANLN
CTCF (p 158)	83kDa	DDX5 (p 159)
SRRT (p 185)	101kDa	RBM26 (p 178)
KHDR1 (p 171)	48kDa	TR150 (p 188)
RBM26 (p 178)	114kDa	DYN2 (p 161)

(continue)

Tableau 3.5: **Résumé général des protéines d'intérêt (suite)**

Exclusivité	MM (kDa)	Spécificité *
TTF2 (p 188)	130kDa	IMDH1 (p 171)
TAOK1 (p 186)	116kDa	XRCC6 (p 190)
PDS5B (p 175)	165kDa	PRC1 (p 176)
DNMT1 (p 161)	183kDa	XRCC5 (p 190)
ERCC2 (p 163)	87kDa	MSH6 (p 173)
ABCF1 (p 153)	96kDa	MSH2 (p 172)
TOP1 (p 187)	91kDa	
NEUA (p 173)	48kDa	
Bande aspécifique		
SMAL1 (p 183)	106kDa	SMAL1
FOXK1 (p 164)	75kDa	FOXK1
FL2D (p 163)	44kDa	FL2D
T2FA (p 185)	58kDa	T2FA
G3P (p 164)	36kDa	G3P
DPOG1 (p 161)	140kDa	DPOG1
CSK23 (p 157)	45kDa	CSK23
HP1B3 (p 169)	61kDa	HP1B3
DDX47 (p 160)	51kDa	DDX47
CREB1 (p 157)	37kDa	CREB1
ATF1 (p 154)	29kDa	ATF1
NFIB (p 174)	47kDa	NFIB
DHX9 (p 160)	140kDa	DHX9
RLA2 (p 180)	11kDa	DDX5 (p 159)
RLA1 (p 180)	12kDa	DYN2 (p 161)
COR1C (p 156)	53kDa	CSK21 (p 157)
C1TC (p 154)	102kDa	RBM15 (p 178)
IF2G (p 170)	51kDa	DDX3X (p 159)
IF2A (p 169)	36kDa	ILF3 (p 170)
CSK21 (p 157)	45kDa	SMCA5 (p 184)
IF2GL (p 170)	51kDa	PNKP (p 175)
VIME (p 188)	54kDa	ILF2 (p 170)
ERCC3 (p 163)	89kDa	IN35 (p 171)
PARP1 (p 175)	113kDa	MSH2 (p 172)
RLA0 (p 180)	34kDa	DDB2 (p 158)
SK2L2 (p 183)	118kDa	SP16H (p 184)

(continue)

Tableau 3.5: **Résumé général des protéines d'intérêt (suite)**

Exclusivité	MM (kDa)	Spécificité *
DDX23 (p 159)	96kDa	SRRT (p 185)
NFIA (p 173)	56kDa	SUH (p 185)
SFPQ (p 183)	76kDa	RFA2 (p 178)



# CHAPITRE 4

## CONCLUSIONS & DISCUSSION PERSPECTIVES



L'un des objectifs majeurs de cette étude consistait en l'élaboration d'une méthodologie qui permît de purifier et de séparer des complexes de réparation de l'ADN. Dans la section suivante, nous discuterons les conclusions que nous pouvons faire sur les résultats obtenus dans le cadre de nos travaux de thèse. Une perspective à moyen terme de ce travail consiste à appliquer cette méthode en vue d'apporter quelques réponses aux questions difficiles relatives à la réparation des cassures radioinduites dans l'ADN ([Perspectives](#), § 4.2, page 146).

## 4.1 CONCLUSIONS & DISCUSSION

Au cours des travaux de thèse décrits plus haut, nous avons obtenu des résultats qui, bien souvent, nous conduisaient à nous interroger, comme décrit dans l'énumération suivante :

1. Les édifices moléculaires se formant sur notre phase ✱ comportent des complexes représentant indubitablement différentes voies de réparation de l'ADN. De manière intéressante, ces voies de réparation sont représentées par des complexes qui sont résolus dans des bandes de poids moléculaires assez différents. Comment faire, alors, pour en étudier la formation ?
2. Les complexes — tout au moins certains — purifiés sur la phase ✱ sont capables de liguer des oligonucléotides, ce qui montre qu'ils sont natifs et fonctionnels. Nous ne savons cependant pas quels complexes portent effectivement l'activité ;
3. L'étude de la composition du matériel biologique séparé par électrophorèse native BN PAGE révèle une grande diversité moléculaire. Un grand nombre des protéines identifiées dans les différentes bandes de gel d'électrophorèse ne sont pas connues pour être impliquées dans la réparation et, pour bon nombre d'entre elles, ne sont pas connues pour appartenir à une quelconque voie de la physiologie de l'ADN. Un autre élément concernant la diversité moléculaire est celui relatif à l'observation qu'une seule bande de gel natif contient certainement plusieurs complexes différents ;
4. Un certain nombre de protéines non connues pour être impliquées dans la réparation de l'ADN sont pourtant *hautement* spécifiques de la phase ✱, qui mime les cassures double-brin. Comment en poursuivre l'étude pour établir leur degré d'implication dans les processus de réparation que nous cherchons à caractériser ?

## 4.1.1 COEXISTENCE DE TROIS VOIES DE RÉPARATION DE L'ADN

Nous avons pu observer, en analysant les données de spectrométrie de masse, que les protéines impliquées dans la réparation des dommages de l'ADN ne représentent qu'une minorité de l'ensemble des protéines identifiées dans chaque bande. Cependant, la quantification des protéines, grâce au paramètre emPAI, révèle que les protéines impliquées dans la réparation des cassures double-brin de l'ADN sont — et de loin — les plus abondantes. Si l'on étudie la répartition des protéines de réparation de l'ADN au sein des différentes bandes de gel BN PAGE, l'on constate que la bande de plus haut poids moléculaire (720 kDa) est, d'une certaine façon, celle qui contient les complexes les plus « aboutis ». Quand nous disons *aboutis*, il ne s'agit pas nécessairement du fait qu'ils soient les complexes les plus gros, mais plutôt qu'ils soient les complexes fonctionnellement les plus cohérents vu qu'ils ont été purifiés sur une phase chromatographique mimant des cassures double-brin : ce sont des complexes de réparation de ces cassures par la voie NHEJ. Les données, présentées dans la Figure 3.23, page 113, montrent bien que les différentes bandes semblent peuplées de complexes de réparation selon des voies différentes (toutes n'étant, par ailleurs, pas directement liées à la réparation des cassures double-brin de l'ADN). De manière intéressante, nous avons mis en évidence une forme de spécialisation fonctionnelle des complexes en fonction de la bande dans laquelle ils sont identifiés. Nous pouvons ainsi discuter de la mise en place de *trois voies* de réparation dans les quatre bandes visualisées, dont deux sont en relation avec la réparation des cassures double-brin (voies du NHEJ). Notons l'absence relative de protéines de réparation des cassures double-brin par recombinaison homologue, qui limite la portée de notre étude. Cependant, ce résultat peut s'expliquer, au vu de nos conditions expérimentales : nos oligonucléotides sont de petite taille et il y a absence d'un contexte chromosomique adapté.

Les voies NHEJ  
canonique, NHEJ  
alternatif/BER et NER

Le fait que nous purifions sur notre phase \* un tel nombre de complexes différents pose la question de la dynamique de l'assemblage de ces complexes sur les oligonucléotides. Notre perception est que, lorsque l'on ajoute l'extrait nucléaire à la phase chromatographique, un processus de tâtonnement se déroule qui consiste à « tester » le substrat sur lequel les protéines de réparation s'assemblent. Au début, l'on peut imaginer que le nombre de protéines est faible et qu'il y a compétition pour le substrat : les protéines les moins affines pour les extrémités libres de l'ADN se détachent à la faveur de la venue de protéines spécifiques. Ainsi, par tâtonnement moléculaire, nous imaginons que les complexes finissent pas atteindre l'état « abouti » que nous évoquons plus haut. Mais nous ne disposons, pour cette spéculation, que de données statiques. Pour tester cette hypothèse, il serait intéressant d'introduire une dimension temporelle dans nos analyses. Il serait donc nécessaire d'effectuer des purifications « résolues en temps »

afin de faire des études cinétiques sur la formation des différents complexes. Nous y avons songé, bien entendu, depuis du temps déjà, mais la contrainte majeure que nous avons est liée à l'étape de photoclivage. En effet, si nous voulons « piéger » des complexes de réparation en cours de formation à différents temps, il faut pouvoir photoclivier très rapidement les complexes nucléoprotéiques pour les détacher des billes. Or, cette phase est longue. Nous songeons à l'emploi de lasers dans l'UV dont la fluence est telle qu'il pourrait être envisageable de détacher suffisamment vite les complexes pour les faire migrer dans le gel. Noter qu'une autre stratégie pourrait fournir les renseignements de cinétique recherchés *vide infra*.

#### 4.1.2 ACTIVITÉ DES COMPLEXES DES DIFFÉRENTES BANDES

Les études de ligation, décrites § 3.2.4, page 89, ont montré que les complexes que nous purifions sur la phase \* sont capables de liguer deux oligonucléotides entre eux. Le fait que nous puissions recruter sur cette phase chromatographique des complexes fonctionnels est un résultat important, qui montre que les protéines purifiées permettent la prise en charge d'une cassure double-brin de l'ADN. Nous ne pouvons toutefois pas établir la nature des complexes qui sont fonctionnels. En effet, nous résolvons les produits de purification en quatre bandes principales, mais nous n'avons pas testé la ligation pour chaque bande isolément. Si nous pouvions le faire, nous nous attendrions sans doute à ce que les complexes de la bande à 720 kDa soient capables de liguer. Peut-être ceux de la bande prétendument aspécifique feraient aussi de la ligation, grâce aux protéines de la voie alternative du NHEJ [PARP1 (p 175), XRCC1 (p 189) et DNLI3 (p 161)].

Un certain nombre d'études ont impliqué la détection d'activités enzymatiques après leur migration en gel de polyacrylamide [Mailloux *et al.*, 2008; Shimizu *et al.*, 1998]. D'autres études, encore plus ambitieuses, ont consisté à faire copolymériser dans le gel une activité enzymatique donnée (une protéine-kinase) afin de rechercher, après migration d'un extrait protéique, les protéines substrat de cette protéine-kinase. La contrainte était liée à la renaturation des protéines (de l'extrait, d'une part et copolymérisée, d'autre part) après la migration [Li *et al.*, 2007]. Nous n'aurions pas autant de contraintes : notre gel est non dénaturant et nous pourrions simplement découper les bandes à tester, les équilibrer dans le tampon ligase, auquel nous aurions préalablement ajouté l'oligonucléotide radiomarqué.

Une autre manière de faire, serait de résoudre les complexes nucléoprotéiques de manière non destructrice, c'est-à-dire sans les faire migrer dans un gel, même non dénaturant. Par exemple, nous pourrions effectuer une séparation par chromatographie d'exclusion stérique. Même si les performances de ce genre de chromatographie sont moyennes, nous aurions sans doute une résolution des complexes à 720 kDa et à 250 kDa, peut-être même

de ceux à 480 kDa. Chaque fraction serait alors testée pour ses propriétés fonctionnelles de ligation. Cette expérience est tout à fait *envisageable*, car nous savons que nos complexes sont stables plus d'une heure à température ambiante dans du tampon d'acétate d'ammonium 50 mM.

De manière anecdotique, nous avons déjà rassemblé les résines pour la mettre en œuvre

### 4.1.3 RECHERCHE DE NOUVELLES PROTÉINES DE RÉPARATION

L'analyse des données de spectrométrie de masse nous a permis d'identifier, dans le produit de purification de chacune des phases  $\diamond$  et  $\star$ , un nombre important de protéines n'étant pas nécessairement impliquées dans la réparation des dommages à l'ADN. Le but de ce travail consistait à identifier de *nouvelles* protéines impliquées dans des processus de réparation, nous avons pour chacune d'entre elles essayé d'établir un lien entre sa répartition dans les différentes bandes et les mécanismes de réparation qui y semblaient initiés.

Cette implication dans les processus de réparation peut être directe ou indirecte, comme dans le cas de la régulation du cycle cellulaire, par exemple

#### 4.1.3.1 PRÉSENTATION DE PROTÉINES « CANDIDATS »

Ce travail a permis de faire ressortir certaines protéines, dont le « comportement » sur nos phases chromatographiques était similaire à celui de « protéines de référence ». Cette similarité a été établie sur la base de deux critères :

Une protéine de référence est une protéine dont le rôle dans la réparation de l'ADN est bien caractérisé

- \* La protéine n'est recrutée que sur la phase chromatographique  $\star$ , révélant une affinité spécifique pour les extrémités libres de l'ADN ;
- \* la protéine se répartit dans les quatre bandes de différents poids moléculaires de la même manière que le fait une protéine de référence, ce qui permet d'imaginer une interaction — directe ou indirecte — entre ces deux protéines.

Les protéines que nous avons révélées au cours de la fouille des données de spectrométrie de masse, qui ne sont pas connues pour être impliquées dans la réparation de l'ADN, constituent d'excellents candidats pour des études fonctionnelles ultérieures. Ces différents candidats sont détaillés ci-dessous, regroupés selon leur(s) fonction(s) connue(s).

Notion large, regroupant les microfilaments et les microtubules

**Les protéines du cytosquelette et les protéines moteurs.** Ces protéines sont intéressantes, car elles sont liées à la dynamique intracellulaire, qui fait l'objet d'un contrôle très rigoureux par la cellule, en particulier avec des « points de contrôle » (« *checkpoints* »). Trois types de protéines figurent dans cette catégorie, les protéines de structure, les protéines régulatrices et enfin les moteurs moléculaires.

Pour les protéines de structure, mentionnons plusieurs septines (SEPT7, p 182), qui ont été purifiées sur la phase ✱ et résolues exclusivement dans la bande à 720 kDa. La fonction de ces protéines à activité GTPasique est encore assez mal connue chez les mammifères. Chez la levure, certaines septines pourraient être impliquées dans la réponse aux dommages à l'ADN, sous forme d'une coordination des fonctions cytosquelettiques (actine) et microtubulaires pendant la division cellulaire, avec la mise en place d'un « *checkpoint* ». Une étude par [Kremer et al.](#), publiée en 2007, a montré que les septines régulent bien, chez les mammifères également, l'organisation du cytosquelette à actine et la mise en place de l'arrêt du cycle cellulaire à la détection de dommages à l'ADN. Ces études ne montrent en aucun cas d'interaction directe des septines avec l'ADN ou même avec des protéines de réparation.

Le moteur moléculaire dynamine (DYN2, p 161) est une protéine présentant des caractéristiques intéressantes, dans nos expériences : elle est identifiée dans toutes les bandes de la phase ✱ et cela d'une manière parfaitement reproductible ; qui plus est, elle n'apparaît jamais dans les bandes de la phase ✧. Dans les cellules, cette protéine est essentielle pour l'organisation du fuseau achromatique, la migration des chromosomes le long des microtubules et la cytokinèse [[Chua et al., 2009](#); [Samora et al., 2011](#)]. À noter que cette protéine interagit également avec une autre protéine recrutée sur nos phases : CLASP1 (p 156), qui stabilise les microtubules. Une spéculation immédiate serait que le recrutement massif de CLASP1 sur des dommages l'empêcherait de stabiliser les microtubules du fuseau, les rendant ainsi moins stables, ce qui permettrait d'enclencher un signal cellulaire de type « *checkpoint* ». D'autre part, le rôle que la dynamine pourrait avoir dans la transition S/G2 n'a pas encore été relié à la réponse aux dommages de l'ADN. Il serait donc intéressant de comprendre pourquoi cette protéine qui, selon la littérature, n'établit pas d'interaction directe avec l'ADN, est aussi spécifiquement et systématiquement purifiée sur la phase ✱. La protéine KIFC1 (p 164), qui se lie aux extrémités positives des microtubules lors de la formation du fuseau, n'est identifiée que dans la bande à 240 kDa et est elle aussi spécifiquement recrutée sur la phase ✱. Elle constitue donc également une protéine d'intérêt.

**Les protéines du métabolisme cellulaire.** Ces protéines sont également intéressantes, en particulier le cas de la protéine IMDH1 (p 171), qui est impliquée dans la synthèse des nucléotides et qui est présente dans les trois bandes de plus haut poids moléculaire après purification sur phase ✱. Nous avons également été intéressés par la spécificité de la protéine G3P (p 164), un enzyme de la glycolyse qui n'est identifié que dans la bande de plus bas poids moléculaire. Cette protéine est connue pour interagir avec

G3P est aussi connue sous le nom de GAPDH

APE1, un enzyme de la réparation des sites abasiques de l'ADN [Azam *et al.*, 2008]. Au cours de l'apparition de dommages à l'ADN, cette protéine est connue pour se transloquer au noyau. Enfin, elle pourrait être impliquée dans l'organisation du cytosquelette et dans le NHEJ [Kornberg *et al.*, 2010; Tisdale, 2002].

### Les protéines impliquées dans la régulation du cycle cellulaire.

Citons, dans cette catégorie, la protéine DDX5 (p 159), une hélicase à ARN qui est impliquée dans la réponse aux dommages de l'ADN en participant à la voie apoptotique médiée par p53 [Bates *et al.*, 2005]. DDX5 est spécifiquement identifiée dans les quatre bandes de la phase  $\ast$ . La protéine CARF (p 154) interagit avec la cycline kinase-dépendante 2 et induit l'activation de la voie pro-apoptotique p53. La protéine CSK21 (p 157) est une kinase impliquée dans la régulation de cette même voie et permet également l'arrêt du cycle cellulaire en phase G2 [Keller *et al.*, 2001]. Enfin, une protéine connue pour intervenir dans la voie de signalisation médiée par une cytokine, mais dont la fonction n'est pas encore connue, est également recrutée de manière spécifique sur la phase  $\ast$  (IN35, p 171).

#### 4.1.3.2 STRATÉGIES D'ÉTUDE DES PROTÉINES CANDIDATS

Plusieurs types d'expériences peuvent être envisagés pour déceler, parmi l'ensemble des protéines citées ci-dessus, celles qui présentent le plus de potentiel :

**Immunofluorescence contre la protéine avant/après dommage à l'ADN.** Il s'agit ici de suivre la localisation subcellulaire de la protéine dans des cellules en culture grâce à un anticorps dirigé contre elle. L'on fera typiquement des expériences d'immunofluorescence dans les cellules avant et après induction de dommages à l'ADN (par des agents radiomimétiques ou irradiation  $\gamma$ ). Deux situations peuvent se présenter : soit la protéine est cytoplasmique et, après induction de dommages à l'ADN, on la voit se relocaliser au noyau ; soit la protéine est nucléaire et elle se relocalise en foyers (« *foci* ») au niveau des chromosomes. Dans les deux cas, l'éventuelle relocalisation de la protéine étudiée après induction de dommages à l'ADN constitue un indice important de son implication — plus ou moins directe, certes — dans la réparation de l'ADN.

**Interférence à ARN et ciblage d'une coupure dans le génome.** Il s'agit d'une méthode d'extinction de l'expression d'une protéine donnée fondée sur une des grandes découvertes en biologie de ces quinze dernières années [Fire *et al.*, 1998]. Pour mettre en œuvre ce type d'expérience, il faut disposer de la séquence de l'ARNm de la protéine dont l'expression doit être

éteinte. On transfecte alors les cellules en culture avec des ARN double-brin de 21–23 pb dont l’un des brins est complémentaire de l’ARNm à cibler. Ces petits ARN interférents sont pris en charge par la protéine Argonaute, qui les sépare en deux brins. Le brin complémentaire de l’ARNm est mis en vis-à-vis de l’ARNm ciblé et ce dernier est dégradé. Le laboratoire de biophysique du Muséum possède une grande compétence dans ce domaine et, également, dans la réparation de l’ADN, telle qu’elle peut être étudiée dans des cellules en culture. En particulier, nos collaborateurs maîtrisent l’emploi de nucléases hyper-spécifiques (CRISPR) qui permettent de cibler sans difficulté des séquences génomiques précises [Barrangou, 2013; Barrangou *et al.*, 2007; Ghezraoui *et al.*, 2014], ce qui leur permet de bien caractériser les événements de réparation qui pourraient avoir lieu dans les cellules. Pour tester l’effet de l’extinction de l’expression d’une protéine donnée sur la réparation de l’ADN, les cellules en culture sont transfectées avec des siRNA. En parallèle, les cellules sont induites pour activer la coupure spécifique à un site précis de leur génome. Après une durée de  $\approx 72$  h (et toutes les 24 h ensuite), une réaction de PCR sur la séquence ciblée par la nucléase permet d’établir si la réparation s’est faite, d’une part et, si oui, comment elle s’est faite. En effet, selon la séquence de la région réparée (la jonction), l’on peut établir dans quelle mesure la voie de réparation employée est le NHEJ, le NHEJ alternatif ou la recombinaison homologue. L’observation de délétions au niveau de cette jonction est en effet associée à l’utilisation de microhomologies lors de la réparation, ce qui est l’une des caractéristiques de la voie alternative du NHEJ [Fattah *et al.*, 2010; Smith *et al.*, 2003]. Il faut noter que, si la protéine étudiée par extinction à l’ARNi a un rôle dans la réparation, cette dernière peut intervenir plus lentement que pour des cellules contrôle ou peut ne pas intervenir du tout, avec des effets bien visibles au niveau de la culture cellulaire dans son ensemble (temps de doublement, apoptose).

siRNA, pour « *small interfering RNAs* »

**La résonance plasmonique de surface (RPS).** Cette technique dont le principe physique est assez complexe, permet de détecter avec une grande sensibilité des interactions moléculaires intervenant à la surface d’un objet métallique (souvent une plaque d’or) [Liedberg *et al.*, 1983]. Nous imaginons une approche combinant la purification de complexes de réparation sur notre phase chromatographique, puis par chromatographie d’exclusion stérique, en premier lieu. Les complexes ainsi simplifiés pourraient être dissociés en conditions douces (hausse de la température, par exemple) pour les déposer ensuite sur la plaque d’or préalablement fonctionnalisée avec des oligonucléotides [Wang *et al.*, 2014; Zagorodko *et al.*, 2014]. L’expérience permettrait ainsi de suivre en temps réel le *ré*-assemblage du complexe de réparation sur la plaque et d’en étudier la cinétique de formation. Il serait alors absolument intéressant de refaire ces expériences en présence

de substances à rôle régulateur ou de cofacteur (ATP,  $Mg^{2+}$ , par exemple). Il serait très intéressant également de suivre une étape de ligation et de constater que le complexe de réparation se détache de l'ADN substrat, par exemple. L'ajout d'anticorps dirigés contre telle ou telle protéine permettrait de tester sa contribution dans la formation de l'assemblage de réparation, soit car ils viendraient se fixer sur le complexe déjà assemblé (ce qui pourrait donner des renseignements topologiques), soit car ils inhiberaient la formation de l'assemblage.

#### 4.1.4 COMMENT ENVISAGER L'IDENTIFICATION DE NOUVELLES PROTÉINES CANDIDATS ?

Le jeu de données que nous avons dans nos mains à l'issue de ce projet de thèse est très important et nous sommes conscients de ce que la fouille des données n'est pas achevée. Un corollaire à cette affirmation est l'idée, déjà mentionnée maintes fois, que nous avons dans chaque bande de gel non dénaturant un certain nombre de complexes différents, dont les protéines sont identifiées (identifications qui font partie de ces nombreuses données, précisément). Deux problèmes se posent :

1. **Comment améliorer la fouille des données ?** Nous entrevoyons bien que le développement de nouveaux outils d'analyse devraient permettre de tirer plus d'informations des données dont nous disposons déjà ;
2. **Comment améliorer la séparation des complexes ?** Il nous faudrait trouver le moyen de résoudre les complexes qui se trouvent dans une même bande afin de permettre l'élaboration d'un modèle plus précis de la manière dont les complexes s'assemblent sur la phase  $*$ .

Si l'on s'attarde sur le point numéro 1, nous pensons que de nouvelles fonctionnalités ajoutées au logiciel de fouille de données devraient permettre de « tracer » de manière quantitative le comportement de deux protéines (ou plus) au cours de différentes purifications. Or, si l'on pouvait montrer que certaines protéines voient leur quantité varier de la même manière dans telle série d'expériences par rapport à telle autre série d'expériences, l'on aurait là un indice fort d'une interaction entre ces protéines. Cette approche de biochimie des interactions non covalentes *in silico* pourrait nous orienter vers des protéines passées inaperçues jusqu'à présent, sans avoir à envisager une séparation ultérieure des complexes. L'on peut également s'interroger sur l'opportunité de refaire un nombre limité d'expériences, telles que celles décrites dans ce mémoire, mais en faisant usage de la quantification par « étiquetage » des peptides tryptiques [Ross *et al.*, 2004; Unwin, 2010]. L'obtention de données quantitatives réellement précises pourrait, à terme, mieux nous renseigner sur la « coprécipitation » de protéines nouvelles avec

des protéines bien connues de différentes voies de réparation de l'ADN et, par là-même, pourrait nous aider à comprendre comment, d'un point de vue moléculaire, telle ou telle voie de réparation prend le pas sur les autres (voir plus haut la discussion concernant les différentes voies de réparation représentées dans les différentes bandes de gel natif).

Le point numéro 2, [page ci-contre](#) ne se satisfait pas d'expériences *in silico*. Il s'agit ici d'envisager différentes méthodes de nature soit à séparer factuellement les différents complexes de masses apparentes approximativement égales soit à les détecter sans les confondre.

**L'immunoprécipitation** est sans doute une méthode de choix, qui utilise un anticorps dirigé contre une protéine donnée pour « pêcher » cette protéine dans un mélange. Si l'anticorps est couplé à des billes, l'expérience permet de retenir sur lesdites billes la protéine d'intérêt et toutes les protéines qui interagissent avec elle (les interactants). Cette stratégie pourrait permettre, par exemple, de déterminer quels sont les interactants de la dynamine dans les produits de purification obtenus sur phase  $\ast$ . On voit bien ici que l'on introduit une nouvelle dimension séparative à notre système de purification. L'identification par spectrométrie de masse des protéines contenues dans le matériel immunoprécipité fournirait des éléments beaucoup plus fiables pour l'étude des interactions protéines-protéines mises en jeu [[Li et al., 2012](#)].

**La spectrométrie de masse à mobilité ionique SM-MI** pourrait apporter une dimension supplémentaire à nos analyses. Il faudrait d'abord simplifier le mélange de complexes, probablement par chromatographie d'exclusion stérique ou par immunoprécipitation (ci-dessus). Chaque fraction obtenue serait alors analysée par spectrométrie de masse à mobilité ionique, qui couple une séparation des espèces moléculaires sur la base du rapport masse-sur-charge ( $m/z$ ) à une séparation des analytes sur la base de leur encombrement stérique<sup>1</sup>. La dimension stérique, tout à fait orthogonale à la dimension purement massique, permet d'envisager l'étude de mélanges complexes et, qui plus est, de procéder à des analyses des composants des complexes, car il est possible de « peler » ces complexes comme des oignons en les soumettant à des collisions et en regardant quels partenaires moléculaires en sont détachés (sur la base de la perte d'une masse caractéristique de telle ou telle protéine connue).

Un instrument de type SM-MI est installé cette année dans notre laboratoire

---

1. En fait, leur résistance à un vol dans une cellule du spectromètre où règne une pression quasi-atmosphérique. On quantifie ainsi la « prise au vent » de différents analytes (assemblages, en l'occurrence) qui ne seraient pas résolus simplement sur la base de leur rapport  $m/z$ .

### 4.1.5 CERTAINES PROTÉINES SONT ABSENTES, POURQUOI ?

Parallèlement à l'identification de protéines *a priori* étrangères aux mécanismes de réparation de l'ADN, nous avons aussi remarqué, à l'inverse, l'absence de protéines majeures de la réparation des cassures double-brin de l'ADN [par exemple XLF (p 193), aussi connue sous le nom de NHEJ1 ou de Cernunnos] qui n'ont jamais été identifiées malgré la présence de protéines avec lesquelles elles interagissent physiquement [dans notre exemple XRCC4 (p 189) et DNLI4 (p 161)]. Des observations de ce type étaient attendues au vu des poids moléculaires apparents des bandes de gel natif. En effet, quelle masse aurait un complexe formé de toutes les protéines connues pour interagir avec le dimère KU, par exemple ? Ce type de discussion est de plus en plus souvent retrouvé dans la littérature (revue récente : Grundy *et al.* [2014]). La notion même d'assemblage dynamique — et ordonné, peut-être — est centrale dans la réparation de l'ADN. Toutes les protéines de la voie NHEJ ne sont pas nécessairement en interaction entre elles et, *in fine*, avec les extrémités à religuer, à un instant donné. Malgré cet argumentaire de nature fonctionnel, nous ne pouvons pas ne pas nous interroger sur le fait que, peut-être, certaines protéines manquent par des causes éminemment expérimentales, comme le fait que la migration électrophorétique non dénaturante ne le serait pas absolument, par exemple, avec une dissociation des sous-unités les plus labiles des complexes initialement formés pour pouvoir les recruter dans nos conditions expérimentales. Il serait intéressant de voir si l'utilisation d'un tel agent renforcerait l'intensité des bandes de haut poids moléculaire dans le gel natif, ou permettrait l'apparition de nouvelles bandes de plus haut poids moléculaire encore, correspondant à des complexes plus importants en taille.

Question évoquée à la  
§ 4.1.1, page 138

Il faut noter, pour rebondir sur la question de la cinétique de formation des complexes, que l'on peut envisager l'emploi d'agents pontants à des moments plus ou moins longs après la mise en contact de l'extrait nucléaire avec la phase chromatographique. Nous figerions alors des complexes à différents états « d'aboutissement », ce qui permettrait de mieux interpréter la présence de différentes voies de réparation dans les différentes bandes de gel natif. Si ces différents états peuvent alors être corrélés avec la présence de telle ou telle protéine, nous pourrions envisager des hypothèses structurales quant au déterminisme du choix de la voie de réparation employée pour un dommage donné, ce qui était l'une des perspectives à moyen/long terme de notre étude.

## 4.2 PERSPECTIVES

L'objectif principal pour la suite de ce projet est la recherche de protéines participant à la réparation des dommages radioinduits dans l'ADN.

Dans les cellules, la réparation de l'ADN est mise en œuvre par différents

systèmes moléculaires qui sont sélectionnés par la cellule en fonction de la nature du dommage. Ces différentes *voies de réparation* de l'ADN se matérialisent par l'assemblage sur le dommage, selon un processus dynamique ordonné, de complexes protéiques qui effectuent une réparation multiétapes. Certaines cassures double-brin de l'ADN — *induites par les rayonnements ionisants* — peuvent mettre en échec la réparation de l'ADN de deux manières :

- \* Les extrémités de l'ADN qui sont libérées au cours de la cassure peuvent avoir des structures chimiques complexes non compatibles avec l'action des ligases à ADN ou des polymérases à ADN ;
- \* Des dommages peuvent être générés sur plusieurs sites d'une courte région de l'ADN, qui sont dits dommages « en cluster » ou sites à dommages multiples. Ces dommages interviennent principalement lorsque l'énergie des particules ionisantes est déposée sur une *courte distance* (notion de transfert linéique d'énergie, TLE ; en anglais « *linear energy transfer* » ; LET).

L'énergie déposée sur une faible distance de pénétration est alors grande

#### 4.2.1 LES DOMMAGES À CHIMIE COMPLEXE

Certains enzymes sont capables de « *rectifier* » les extrémités de chimie complexe telles qu'elles sont produites lors des cassures radioinduites dans l'ADN. Ces enzymes doivent intervenir afin de les rendre à nouveau réparables. Peu de ces enzymes sont connus à ce jour, alors qu'une grande diversité chimique caractérise les extrémités de l'ADN produites par les rayonnements ionisants.

Si l'on prend le cas de la réparation des cassures double-brin radioinduites de l'ADN, celle-ci ne peut se produire que si la rectification des extrémités de chimie complexe est opérée par des enzymes spécialisés. Les extrémités 3'-phosphoglycolate, par exemple, empêchent la religation, qui ne pourra se faire que si l'enzyme TDP1 [Pouliot *et al.*, 1999] transforme le 3'-phosphoglycolate en 3'-phosphate [Zhou *et al.*, 2005]. De manière intéressante, si l'on inactive cet enzyme, la radiosensibilisation aux rayonnements ionisants est seulement légèrement augmentée, ce qui pourrait indiquer une redondance fonctionnelle. La protéine Artemis est une protéine de rectification possédant des activités exonucléase et endonucléase employées pour rectifier les extrémités de chimie complexe de l'ADN [Drouet *et al.*, 2006]. Elle peut se substituer à TDP1 et pourrait constituer ainsi la base moléculaire de la redondance fonctionnelle évoquée ci-dessus [Moshous *et al.*, 2001; Povirk *et al.*, 2007]. Son implication dans des phénomènes physiologiques comme le V(D)J anihile, cependant, l'idée d'en faire une cible pour un inhibiteur que l'on utiliserait comme adjuvant au traitement radiothérapeutique antitumoral.

Historiquement, la découverte de nouveaux processus de réparation et des protéines impliquées dans ceux-ci, s'est faite par des études génétiques et biochimiques de cellules issues de patients ayant consulté en hôpital pour des symptômes associés à l'exposition aux rayons solaires ou pour lesquels le système immunitaire était dysfonctionnel [Fulop et Phillips, 1990; Moshous *et al.*, 2000]. Des cultures cellulaires — typiquement de fibroblastes ou de cellules du sang — sont mises en œuvre à partir de prélèvements tissulaires chez les sujets malades et chez des sujets sains (cultures contrôle). Grâce à ces cultures, des expériences de génétique moléculaire et de biochimie peuvent être menées afin de comparer les cellules du malade avec celles de sujets sains [Buck *et al.*, 2006]. Une fois le gène caractérisé, les études biochimiques du mécanisme d'intervention du produit du gène dans la réparation proprement dite peuvent commencer.

Des extraits protéiques nucléaires sont préparés à partir de ces cultures cellulaires, pour être chromatographiés sur des phases munies d'oligonucléotides duplexes d'ADN mimant à leur extrémité une cassure double-brin de l'ADN. L'objectif de ces expériences est de comparer — selon les cultures cellulaires desquelles les protéines ont été extraites — le type de complexe de réparation (et sa composition en sous-unités) qui peut s'assembler sur les oligonucléotides ADN de la phase chromatographique. Le fait — largement documenté dans ce mémoire — de pouvoir se départir des billes après purification permet de se débarrasser des protéines contaminantes liées aux billes et d'obtenir des facteurs d'enrichissement inégalés qui permet donc d'envisager, comme nous le montrons, une étude moléculaire fine des protéines interagissant spécifiquement avec nos phases.

Notre projet vise donc la caractérisation biochimique et fonctionnelle de nouveaux enzymes de rectification se liant spécifiquement aux extrémités de chimie complexe libérées par les cassures radioinduites dans l'ADN. Il faut donc constituer une phase chromatographique fonctionnalisée avec des oligonucléotides ADN dotés d'extrémités de chimie complexe. Pour ce faire, on irradie les oligonucléotides biotinylés utilisés pour la préparation de la phase  $\diamond$ , ce qui produira des cassures double-brin. Les ADN monobiotinylés ainsi obtenus seront incubés avec les billes, ce qui constituera une phase  $\star$  à ADN irradié. Des extraits protéiques nucléaires de cellules traitées (par des rayons  $\gamma$  ou des radiomimétiques) seront chromatographiés sur cette phase particulière afin d'en purifier des complexes de réparation spécifiques des cassures radioinduites dans l'ADN.

La formation de cassures double-brin de l'ADN fait partie intégrante du processus physiologique conduisant à la diversité des anticorps et donc à la performance du système immunitaire

Les éventuelles protéines découvertes pour participer activement à la réparation des cassures à chimie complexe pourraient devenir des cibles pharmacologiques afin de les inhiber, ce qui pourrait permettre de radiosensibiliser les cellules tumorales et donc de diminuer les doses déposées avec un même effet de régression tumorale.

#### 4.2.2 LES SITES À DOMMAGES MULTIPLES

La notion de dommages multiples sur une courte région de l'ADN (localement, donc) date de la fin des années 80 [Ward, 1988; Ward *et al.*, 1987] (pour des revues, voir [Gros *et al.*, 2002; Sage et Harrison, 2011; Schipler et Iliakis, 2013]). Ces dommages interviennent quand les rayonnements ionisants peuvent déposer une grande énergie sur une distance de pénétration courte (on dit que ces rayonnements sont « *high LET* »). Bien que l'existence de ces sites multiples de dommages ait été prédite par la modélisation, cette notion est une réponse conceptuelle aux problèmes liés à l'observation que tous les rayonnements ionisants ne produisent pas des dommages de même gravité, en termes de réparation par les cellules. Les dommages dus aux rayonnements ionisants sont souvent liés à des modifications chimiques des bases (ou des sucres) et produisent ce que nous avons décrit plus haut comme des dommages à chimie complexe. Le fait que ces dommages soient regroupés en « *clusters* » pose des problèmes majeurs à la cellule : comment réparer dans de brefs délais des dommages regroupés dans une séquence d'ADN courte (un tour d'hélice) ?

Des expériences élégantes ont permis de montrer que la réparation des dommages multiples est difficile pour les cellules, en partie car les enzymes de réparation introduisent des coupures simple-brin à l'endroit des sites endommagés (absence de base, en particulier). L'intermédiaire réactionnel de réparation de la cassure simple-brin introduite par le cycle catalytique de réparation peut bloquer la réparation de l'autre dommage, cependant que la réplication, elle, continue, ce qui introduit une mutation. Bien d'autres cas existent et sont de degrés de complexité différents.

D'autres expériences ambitieuses, chez la levure, ont tenté de répondre aux différents problèmes évoqués ci-dessus. Il s'agit de faire la synthèse chimique de courts ADN duplexes dans lesquels des dommages bien définis ont été introduits sur les deux brins à des distances courtes, ce qui est censé mimer des sites de dommages multiples. Ces ADN de synthèse ont été insérés dans un vecteur, lui-même transfecté dans les cellules. Le devenir du plasmide est alors étudié et des conclusions ont été tirées quant au caractère réparable de certaines combinaisons de lésions multiples [Kozmin *et al.*, 2009].

Nous ne pouvons pas envisager d'expériences de ce type, avec des cellules en culture, mais nous pouvons imaginer irradier avec des rayonnements à haut TLE de l'ADN biotinylé et nous en servir pour fonctionnaliser

des billes, exactement comme ce que nous souhaitons faire pour étudier les protéines de rectification. Les purifications seraient alors mises en œuvre de la même manière que celle qui a été décrite dans ce travail. Nous ne pouvons pas prédire ce que nous obtiendrons, mais il n'est pas exclu que certaines bandes soient différentes de celles que l'on obtient lorsque les ADN biotinylés sont irradiés avec des rayonnements à faible TLE.

# CHAPITRE 5

## ANNEXES



## FICHES RELATIVES AUX PROTÉINES D'INTÉRÊT

### 5.0.3 PROTÉINES RÉVÉLÉES PAR LA FOUILLE DES DONNÉES

#### ABCF1

---

**Descriptif** « *ATP-binding cassette sub-family F member 1* »

Impliquée dans l'initiation de la traduction des ARNm [Paytubi *et al.*, 2009]. L'isoforme 1 de cette protéine est également impliqué dans la biogenèse des ribosomes.

**Uniprot** <http://www.uniprot.org/uniprot/Q8NE71>

#### ACINU

---

**Descriptif** « *apoptotic chromatin condensation apoptosis inducer in the nucleus* »

Membre auxiliaire du complexe de jonction des exons participant à l'épissage des ARNm. ACINU est la protéine qui confère à ce complexe la propriété de pouvoir se fixer sur les ARN [Schwerk *et al.*, 2003].

**Uniprot** <http://www.uniprot.org/uniprot/Q9UKV3>

#### ADNP

---

**Descriptif** « *activity-dependant neuroprotector homeobox protein* »

Facteur de transcription potentiel.

**Uniprot** <http://www.uniprot.org/uniprot/Q9H2P0>

#### AFF4

---

**Descriptif** « *AF4/FMR2 family member 4* »

Fait partie du super complexe d'élongation qui permet d'augmenter l'activité catalytique de l'ARN polymérase II.

**Uniprot** <http://www.uniprot.org/uniprot/Q9UHB7>

#### ANLN

---

**Descriptif** « *ctin-binding protein anillin* »

Essentielle pour la cytokinèse, en particulier parce qu'elle permet de réguler la contractilité de la myosine pendant cette phase [Zhao *et Fang*, 2005].

**Uniprot** <http://www.uniprot.org/uniprot/Q9NQW6>

### ATF1

---

**Descriptif** « *cyclic AMP-dependent transcription factor ATF-1* »

Ce facteur de transcription se lie également au CRE des promoteurs de gènes viraux. Il joue également un rôle dans l'induction de la prolifération et de la transformation cellulaire [Zheng *et al.*, 2008].

**Uniprot** <http://www.uniprot.org/uniprot/P18846>

### BAZ1A

---

**Descriptif** « *bromodomain adjacent to zinc finger domain protein 1A* »

Interagit avec SMCA5 (p 184), en particulier lors de la réplication de l'hétérochromatine péri-centromérique. Ces deux protéines font de plus partie du complexe WHICH, qui médie la phosphorylation des histones  $\gamma$ H<sub>2</sub>AX et qui reconfigure les structures anormales de la chromatine (SMCA5 est le domaine ATPasique de ce complexe et BAZ1A en est la sous-unité accessoire et non catalytique).

**Uniprot** <http://www.uniprot.org/uniprot/Q9NRL2>

### C1TC

---

**Descriptif** « *C-1-tetrahydrofolate synthase* »

Impliquée dans le métabolisme du THF (tétrahydrofolate), qui permet son interconversion. Certaines des activités de cette enzyme nécessitent l'utilisation d'un équivalent d'ATP ou de NADP<sup>+</sup>.

**Uniprot** <http://www.uniprot.org/uniprot/P11586>

### CAF1A

---

**Descriptif** « *chromatin assembly factor 1 subunit A* »

Fait partie du complexe CAF-1 dont la fonction est de reformer la chromatine lors de la réplication et la réparation de l'ADN.

**Uniprot** <http://www.uniprot.org/uniprot/Q13111>

### CARF

---

**Descriptif** « *CDKN2A-interacting protein* »

Impliquée dans l'activation des voies de p53/TP53 (provoquant notamment l'apoptose).

**Uniprot** <http://www.uniprot.org/uniprot/Q9NXV6>

### CC037

---

**Descriptif** « *embryonic stem cell-specific 5-hydroxymethylcytosine-binding protein* »

Se lie spécifiquement à l'ADN présentant une 5-hydroxyméthylcytosine dans les cellules souches. Elle pourrait avoir une activité peptidase (activité prédite par similarité). Également connue sous le nom HMCES.

**Uniprot** <http://www.uniprot.org/uniprot/Q96FZ2>

### CDK7

---

**Descriptif** « *cyclin-dependant kinase 7* »

Kinase impliquée dans le contrôle du cycle cellulaire. Cette protéine est la sous-unité catalytique du complexe d'activation des kinases dépendantes de la cycline. Ce complexe est indirectement responsable de l'arrêt du cycle cellulaire provoqué lors de la réponse aux dommages de l'ADN.

**Uniprot** <http://www.uniprot.org/uniprot/P50613>

### CHD2

---

**Descriptif** « *chromodomain-helicase-DNA-binding protein 2* »

Serait impliquée dans la réponse aux dommages de l'ADN. Elle interagit avec la protéine histone H3.3 et permet ainsi le remodelage de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/O14647>

### CHD4

---

**Descriptif** « *chromodomain-helicase-DNA-binding protein 4* »

Protéine impliquée dans le remodelage de la chromatine *via* la désacétylation des protéines histones [Tong *et al.*, 1998].

**Uniprot** <http://www.uniprot.org/uniprot/Q14839>

### CHD7

---

**Descriptif** « *chromodomain-helicase-DNA-binding protein 7* »

Protéine impliquée dans le remodelage de la chromatine *via* la désacétylation des protéines histones. Sa déficience provoque l'apparition du syndrome CHARGE [Schulz *et al.*, 2014].

**Uniprot** <http://www.uniprot.org/uniprot/Q9P2D1>

### CKAP5

---

**Descriptif** « *cytoskeleton-associated protein 5* »

Protéine cytosquelettique qui se lie aux extrémités positives des microtubules. Elle joue un rôle important dans l'organisation des pôles du fuseau [Gergely *et al.*, 2003].

**Uniprot** <http://www.uniprot.org/uniprot/Q14008>

### CLAP1

---

**Descriptif** « *CLIP-associating protein 1* »

Joue un rôle dans la stabilisation des microtubules, en particulier dans leur (dé-)polymérisation en se fixant à leur extrémité positive. Cette protéine est également impliquée dans la stabilisation des kinétochores et intervient donc dans la division cellulaire.

**Uniprot** <http://www.uniprot.org/uniprot/Q7Z460>

### CLASP1

---

**Descriptif** Voir CLAP1 (p 156)

### CLIP2

---

**Descriptif** « *CAP-gly domain-containing linker protein 2* »

Protéine méconnue qui lierait les microtubules aux corps dendritiques lamellaires (des organelles membranaires des neurones lorsqu'ils sont liés par des jonctions dendrites-dendrites).

**Uniprot** <http://www.uniprot.org/uniprot/Q9UDT6>

### COR1C

---

**Descriptif** « *coronin-1C* »

Serait impliquée dans la motilité et la transduction du signal. Elle pourrait jouer un rôle dans la motilité en intervenant dans l'organisation du cytosquelette d'actine.

**Uniprot** <http://www.uniprot.org/uniprot/Q9ULV4>

### CPSF5

---

**Descriptif** « *cleavage and polyadenylation specificity factor subunit 5* »  
Protéine impliquée dans le complexe de pré-clivage des ARNm, qui intervient lors de l'épissage. Ce facteur interagit avec CPSF7 (décrite ci dessus).

**Uniprot** <http://www.uniprot.org/uniprot/043809>

### CPSF7

---

**Descriptif** « *cleavage and polyadenylation specificity factor subunit 7* »  
Protéine impliquée dans le complexe de pré-clivage des ARNm, qui intervient lors de l'épissage.

**Uniprot** <http://www.uniprot.org/uniprot/Q8N684>

### CREB1

---

**Descriptif** « *cyclic AMP-responsive element-binding protein 1* »  
Facteur de transcription dépendant de la phosphorylation. Il stimule la transcription des gènes viraux en se fixant au CRE (« *cAMP response element* »).

**Uniprot** <http://www.uniprot.org/uniprot/P16220>

### CSK21

---

**Descriptif** « *casein kinase II subunit alpha* »  
Constituant d'une serine/threonine kinase. Lors de la mitose, cette protéine fait partie du point de contrôle de l'assemblage du fuseau dépendant de p53/TP53 qui est responsable de l'arrêt du cycle en G2 en réponse à l'endommagement du fuseau *via* son interaction avec CDK1 [Keller *et al.*, 2001]. Cette protéine est également impliquée dans la régulation négative de l'apoptose en raison de son activité de phosphorylation des caspases 9 et 2 et de NOL3 un régulateur apoptotique. Cette phosphorylation inhibe l'activation de la cascade de phosphorylation médiée par la caspase 8 [Sayed *et al.*, 2001].

**Uniprot** <http://www.uniprot.org/uniprot/P68400>

### CSK23

---

**Descriptif** « *casein kinase II subunit alpha 3* »

Probable sous-unité catalytique d'une sérine/thréonin kinase ayant une grande variété de substrats. Elle promeut la prolifération cellulaire ainsi que la tumorigénèse en régulant négativement l'expression du suppresseur de tumeur PML.

**Uniprot** <http://www.uniprot.org/uniprot/Q8NEV1>

### CTCF

---

**Descriptif** « *transcriptional repressor CTCF* »

Interagit de préférence avec l'ADN non méthylé. Elle n'est pas décrite pour intervenir dans la réparation de l'ADN, mais joue un rôle important dans le remodelage de la chromatine en favorisant la formation de boucles de chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/P49711>

### CUL1

---

**Descriptif** « *cullin-1* »

CUL1 est une protéine centrale des complexes E3 ubiquitine-ligase, qui sont impliquées dans le contrôle de la progression du cycle cellulaire, de la transduction du signal et de la transcription [Lovly *et al.*, 2008].

**Uniprot** <http://www.uniprot.org/uniprot/Q13616>

### DDB1

---

**Descriptif** « *DNA damage-binding protein 1* »

DDB1 interagit avec la protéine DDB2 (p 158) et reconnaît les dommages de l'ADN induits par rayonnement ultraviolet. Ce complexe permet ensuite le recrutement des protéines impliquées dans la voie de réparation par excision de nucléotide (NER). En association avec CUL4, est aussi impliquée dans l'élimination des protéines histones par ubiquitinylation, ce qui permet aux protéines de réparation d'accéder aux régions endommagées de l'ADN [Hu *et al.*, 2004; Wang *et al.*, 2006].

**Uniprot** <http://www.uniprot.org/uniprot/Q16531>

### DDB2

---

**Descriptif** « *DNA damage-binding protein 2* »

DDB2 interagit avec DDB1 (décrite ci-dessus) pour initier la réparation de l'ADN par la voie du NER, mais également au sein d'un complexe ubiquitine

ligase dont l'un des rôles est de permettre l'accessibilité aux dommages de l'ADN par élimination des protéines histones [Wakasugi *et al.*, 2002].

**Uniprot** <http://www.uniprot.org/uniprot/Q92466>

### DDX3X

---

**Descriptif** « *ATP-dependent RNA helicase DDX3X* »

DDX3X est une hélicase ATP-dépendante multifonctionnelle : elle peut *in vitro* dérouler partiellement la double-hélice d'ADN et *in vivo* elle est impliquée dans la transcription, la maturation des ARNm ainsi que dans leur export et leur traduction. Elle s'associe au complexe de jonction des exons lors de l'épissage des ARNm, avec les ribosomes lors de la traduction. Elle jouerait un rôle anti apoptotique en interagissant avec GKS3A/B et BIRC2 [Sun *et al.*, 2008]. Elle est également utilisée par le virus du HIV pour se répliquer dans les cellules humaines.

**Uniprot** <http://www.uniprot.org/uniprot/000571>

### DDX5

---

**Descriptif** « *probable ATP-dependent RNA helicase DDX5* »

Impliquée dans la régulation de l'épissage des pré ARNm ; Se lie à l'exon 10 du pré ARNm de la protéine tau. DHX5 est également impliquée dans la régulation de la transcription du récepteur de l'oestrogène et de l'androgène. Peut également réprimer la transcription en interagissant avec les histones désacétylases [Wilson *et al.*, 2004]. DDX5 est également impliquée dans la réponse aux dommages de l'ADN en participant à la voie apoptotique médiée par p53/TP53 [Bates *et al.*, 2005].

**Uniprot** <http://www.uniprot.org/uniprot/P17844>

### DDX23

---

**Descriptif** « *probable ATP-dependent RNA helicase DDX23* »

Protéine impliquée dans l'épissage des pré ARNm. Sous sa forme phosphorylée elle permet également la formation du complexe B du spliceosome [Mathew *et al.*, 2008].

**Uniprot** <http://www.uniprot.org/uniprot/Q9BUQ8>

**DDX47**

---

**Descriptif** « *probable ATP-dependent RNA helicase DDX47* »

Hélicase impliquée dans l'apoptose [Lee *et al.*, 2005] qui pourrait également jouer un rôle dans le processing des ARNr et l'épissage des ARNm.

**Uniprot** <http://www.uniprot.org/uniprot/Q9H0S4>

**DHE4**

---

**Descriptif** « *glutamate deshydrogenase 2, mitochondrial* »

Impliqué dans le métabolisme du glutamate (elle est en particulier importante pour sa régénération).

**Uniprot** <http://www.uniprot.org/uniprot/P49448>

**DHX9**

---

**Descriptif** « *ATP-dependant RNA helicase A* »

Possède une activité hélicase dont le rôle est d'activer la transcription des ARNm du collagène de type I.

**Uniprot** <http://www.uniprot.org/uniprot/Q08211>

**DIDO1**

---

**Descriptif** « *death-inducer obliterator 1* »

Jouerait un rôle de facteur de transcription pro-apoptotique [Liu *et al.*, 2014].

**Uniprot** <http://www.uniprot.org/uniprot/Q9BTC0>

**DKC1**

---

**Descriptif** « *H/ACA ribonucleoprotein complex subunit 4* »

Protéine de la physiologie de l'ARN.

Protéine essentielle pour la biogenèse des ribosomes et la maintenance des télomères. Elle pourrait également faire partie du complexe chargé de la pseudouridylation des ARNr.

**Uniprot** <http://www.uniprot.org/uniprot/060832>

### DNLI3

---

**Descriptif** « *DNA ligase 3* »

Protéine impliquée dans la réparation de l'ADN, en particulier dans les voies de réparation de l'ADN par excision de base (BER) et dans la voie alternative du NHEJ, et qui interagit avec la protéine XRCC1 (p 189) [Simsek et Jasin, 2011].

**Uniprot** <http://www.uniprot.org/uniprot/P49916>

### DNLI4

---

**Descriptif** « *DNA ligase 4* »

Permet la jonction des extrémités des oligodéoxynucléotides double-brin de manière ATP-dépendante. DNLI4 est impliquée dans la réparation des cassures double-brin de l'ADN dans la voie canonique du NHEJ, en interaction avec les protéines XRCC4 (p 189) et XLF (p 193) qui sont nécessaires à son activité. Le recrutement de ce complexe de ligation au niveau des extrémités de l'ADN est réalisé par PRKDC (p 177) [Chen *et al.*, 2000].

**Uniprot** <http://www.uniprot.org/uniprot/P49917>

### DNMT1

---

**Descriptif** « *DNA (cytosine-5)-methyltransferase 1* »

Responsable de la méthylation des îlots CpG de l'ADN, et s'associe avec l'ADN en phase G2/M afin de maintenir cette méthylation. La liaison directe de DNMT1 avec l'histone désacétylase 2 (HDAC2 (p 168)) permet la répression de la transcription.

**Uniprot** <http://www.uniprot.org/uniprot/P26358>

### DPOG1

---

**Descriptif** « *DNA polymerase subunit gamma-1* »

Polymérase impliquée dans la réplication de l'ADN mitochondrial.

**Uniprot** <http://www.uniprot.org/uniprot/P54098>

### DYN2

---

**Descriptif** « *dynamine-2* »

Protéine associée aux microtubules et productrice de force qui se lie au GTP et pourrait être de plus impliquée dans la transition des cellules en phase G2/M [Diatloff-Zito *et al.*, 1995].

**Uniprot** <http://www.uniprot.org/uniprot/P50570>

L'étude de cette protéine serait extrêmement intéressante car il s'agit de la protéine la plus spécifique de la phase ✱, identifiées à tous les poids moléculaires. Il semblerait donc qu'elle ne reconnaisse que les extrémités libre de l'ADN, ce qui en ferait une bonne protéine de réparation, mais son rôle dans ce domaine n'est pas documenté. L'étude de cette protéine nous permettrait également d'argumenter en faveur ou non de l'utilisation des données de masse fournies par l'analyse de la bande « aspécifique » pour éliminer de manière systématique des protéines dont la spécificité pour la phase ✱ serait artéfactuelle.

### EHMT1

---

**Descriptif** « *histone-lysine N-methyltransferase 1* »

Aussi connue sous le nom de Glp, méthyle spécifiquement la lysine 9 des histones H3, mais peut également méthyler l'ADN et d'autres protéines, telles que la protéine suppresseur de tumeur p53 [Huang *et al.*, 2010]. EHMT1 pourrait être impliquée dans la régulation de la progression du cycle cellulaire pendant la phase G0.

**Uniprot** <http://www.uniprot.org/uniprot/Q9H9B1>

### EHMT2

---

**Descriptif** « *histone-lysine N-methyltransferase 2* »

Aussi connue sous le nom de G9A, méthyle spécifiquement la lysine 9 des histones H3, mais peut également méthyler l'ADN et d'autres protéines, telles que la protéine suppresseur de tumeur p53 [Huang *et al.*, 2010].

**Uniprot** <http://www.uniprot.org/uniprot/Q96KQ7>

### EMSA1

---

**Descriptif** « *ELM2 and SANT domain-containing protein 1* »

Mal connue : on lui suppose une capacité de fixation à l'ADN et un rôle de régulateur de la transcription, mais cela n'a pas été confirmé expérimentalement.

**Uniprot** <http://www.uniprot.org/uniprot/Q6PJG2>

## ERCC2

---

**Descriptif** « *TFIIH basal transcription factor complex helicase XPD subunit* »

Hélicase ATP-dépendante qui fait partie du complexe de transcription TFIIH. ERCC2 intervient dans la réparation de l'ADN par la voie du NER, dans la ségrégation des chromosomes (au sein du complexe MMXD qui interagit avec le fuseau mitotique) et dans la transcription (permet le recrutement du complexe d'activation des kinases cycline-dépendantes sur le complexe de transcription).

**Uniprot** <http://www.uniprot.org/uniprot/P18074>

## ERCC3

---

**Descriptif** « *TFIIH basal transcription factor complex helicase XPB subunit* »

Protéine hélicase faisant partie du complexe de transcription. Cette protéine est impliquée dans la voie de réparation de l'ADN par excision de nucléotide et permet de combler les manques générés lors de l'excision [van Vuuren *et al.*, 1994].

**Uniprot** <http://www.uniprot.org/uniprot/P19447>

## EXOSX

---

**Descriptif** « *exosome component 10* »

Sous-unité catalytique putative du complexe exosomal des ARN, dont le rôle est primordial dans la physiologie des divers ARN. Il serait impliqué dans la maturation des ARNr, ARNsn et ARNsno, mais aussi dans l'élimination des ARNm et en particulier ceux des protéines histones [Mullen *et Marzluff*, 2008]

**Uniprot** <http://www.uniprot.org/uniprot/Q01780>

## FL2D

---

**Descriptif** « *pre-mRNA splicing regulator WTAP* »

Intervient dans la régulation de l'épissage. En se fixant au pré ARNm codant pour CCNA2, FL2D régulerait la transition du cycle cellulaire en phase G2/M [Horiuchi *et al.*, 2006].

**Uniprot** <http://www.uniprot.org/uniprot/Q15007>

## FOSL2

---

**Descriptif** « *Fos-related antigen 2* »

Protéine impliquée dans la formation des os par sa régulation positive de la différenciation des ostéoblastes. FOSL2 est une protéine qui dimérise également avec la protéine JUN. Ce complexe favorise la prolifération cellulaire [Hein *et al.*, 2009].

**Uniprot** <http://www.uniprot.org/uniprot/P15408>

## FOXK1

---

**Descriptif** « *forkhead box protein 1* »

Régulateur transcriptionnel qui favorise la transcription du gène de la myoglobine.

**Uniprot** <http://www.uniprot.org/uniprot/P85037>

## G3P

---

**Descriptif** « *glyceraldehyde-3-phosphate dehydrogenase* »

Protéine dotée d'une activité glycéraldéhyde-3-phosphate déshydrogénase et nitrosylase, qui intervient dans la glycolyse. Intervient aussi dans des mécanismes nucléaires comme la transcription, le transport des ARN, la réplication et l'apoptose. Ces fonctions seraient dues à son activité nitrosylase, qui aurait pour substrat les protéines SIRT1, HDAC2 et PRKDC [Joo *et al.*, 2012; Kornberg *et al.*, 2010]. Elle joue également un rôle sur l'organisation du cytosquelette [Tisdale, 2002].

**Uniprot** <http://www.uniprot.org/uniprot/P04406>

## KIFC1

---

**Descriptif** « *kinesin-like protein 1* »

Se fixe aux extrémités négatives des microtubules et est requise pour la formation des fuseaux bipolaires. Elle intervient donc, entre autres, dans les mécanismes de régulation de la progression du cycle cellulaire.

**Uniprot** <http://www.uniprot.org/uniprot/Q9BW19>

## H12

---

**Descriptif** « *histone H1.2* »

Variante de l'histone H1 qui lie les nucléosomes entre eux, formant ainsi la chromatine. Les histones H1 sont nécessaires pour la condensation de la

chromatine et peuvent permettre la régulation de la transcription de certains gènes grâce à cette activité.

**Uniprot** <http://www.uniprot.org/uniprot/P16403>

### H13

---

**Descriptif** « *histone H1.3* »

Protéine centrale pour la formation de la fibre de chromatine puisqu'elle permet l'assemblage des nucléosomes. Elle agit également comme régulateur de la transcription de par sa possibilité de modifier l'accessibilité d'un gène.

**Uniprot** <http://www.uniprot.org/uniprot/P16402>

### H15

---

**Descriptif** « *histone H1.5* »

Variante de l'histone H1 qui est chargée de maintenir entre eux les nucléosomes. Protéine impliquée dans la formation de la fibre de chromatine. Permet également de contrôler la transcription *via* le remodelage de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/P16401>

### H2B1B

---

**Descriptif** « *histone H2B type 1-B* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/P33778>

### H2B1C

---

**Descriptif** « *histone H2B type 1-C/E/F/G/I* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/P62807>

### H2B1D

---

**Descriptif** « *histone H2B type 1-D* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/P58876>

### H2B1H

---

**Descriptif** « *histone H2B type 1-H* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/Q93079>

### H2B1J

---

**Descriptif** « *histone H2B type 1-J* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/P06899>

### H2B1K

---

**Descriptif** « *histone H2B type 1-K* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/O60814>

### H2B1L

---

**Descriptif** « *histone H2B type 1-L* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la

réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/Q99880>

### H2B1M

---

**Descriptif** « *histone H2B type 1-M* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/Q99879>

### H2B1N

---

**Descriptif** « *histone H2B type 1-N* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/Q99877>

### H2B1O

---

**Descriptif** « *histone H2B type 1-O* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/P23527>

### H2B2E

---

**Descriptif** « *histone H2B type 2-E* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/Q16778>

### H2B2F

---

**Descriptif** « *histone H2B type 2-F* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/Q5QNW6>

### H2B3B

---

**Descriptif** « *histone H2B type 3-B* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/Q8N257>

### H2BFS

---

**Descriptif** « *histone H2B type F-S* »

Protéine histone constituant les nucléosomes. Ces protéines peuvent faire l'objet de modifications post-traductionnelles lors de la transcription, de la réplication ou de la réparation afin de permettre de modifier le degré de compaction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/P57053>

### H4

---

**Descriptif** « *histone H4* »

Protéine constituant les nucléosomes, structures qui enroulent l'ADN et le compactent. De ce fait elle limite l'accessibilité de la double-hélice aux protéines de la physiologie de l'ADN et doit parfois être éliminée pour permettre les processus cellulaires.

**Uniprot** <http://www.uniprot.org/uniprot/P62805>

### HDAC2

---

**Descriptif** « *histone deacetylase 2* »

Protéine responsable de la désacétylation des résidus lysines des histones constituant les nucléosomes (H2A, H2B, H3 et H4). La désacétylation est

associée à la répression de la transcription. Interagit avec la protéine DNMT1 (p 161) en fin de phase S pour réprimer la transcription. Cette protéine est impliquée dans des cascades de modifications post-traductionnelles, comme par exemple la SUMOylation [Blakeslee *et al.*, 2014].

**Uniprot** <http://www.uniprot.org/uniprot/Q92769>

### HMCES

---

**Descriptif** « *embryonic stem cell-specific 5-hydroxymethylcytosine-binding protein* »

Se lie spécifiquement à l'ADN présentant une 5-hydroxyméthylcytosine dans les cellules souches. Elle pourrait avoir une activité peptidase (activité prédite par similarité). Également connue sous le nom CC037 (p 155).

**Uniprot** <http://www.uniprot.org/uniprot/Q96FZ2>

### HNRPU

---

**Descriptif** « *heterogeneous nuclear ribonucleoprotein U* »

Protéine impliquée dans l'épissage, mais qui a également une affinité pour l'ADN double- et simple-brin. Est phosphorylée par PRKDC (p 177), mais son rôle dans la réparation de l'ADN reste putatif [Britton *et al.*, 2009].

**Uniprot** <http://www.uniprot.org/uniprot/Q00839>

### HP1B3

---

**Descriptif** « *heterochromatin protein 1-binding protein 3* »

Composant de l'hétérochromatine qui pourrait avoir un rôle dans la structure et la fonction de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/Q5SSJ5>

Il serait intéressant d'étudier cette protéine, car son rôle précis dans les cellules ne semble pas documenté. Il pourrait s'agir d'une nouvelle protéine impliquée dans les réarrangements de la chromatine permettant aux protéines de réparation d'accéder aux dommages.

### IF2A

---

**Descriptif** « *eukaryotic translation initiation factor 2 subunit 1* »

Impliquée dans la synthèse des protéines. Forme un complexe ternaire avec une molécule de GTP et l'ARNt initiateur, qui se lie ensuite à la sous-unité ribosomale 40S.

**Uniprot** <http://www.uniprot.org/uniprot/P05198>

### IF2G

---

**Descriptif** « *eukaryotic translation initiation factor 2 subunit 3* »

Impliquée dans la synthèse des protéines. Forme un complexe ternaire avec une molécule de GTP et l'ARNt initiateur, qui se lie ensuite à la sous-unité ribosomale 40S.

**Uniprot** <http://www.uniprot.org/uniprot/P41091>

### IF2GL

---

**Descriptif** « *putative eukaryotic translation initiation factor 2 subunit 3-like protein* »

Possible initiateur de la traduction. Son rôle serait semblable à celui de la protéine IF2G (p 170).

**Uniprot** <http://www.uniprot.org/uniprot/Q2VIR3>

### ILF2

---

**Descriptif** « *interleukin enhancer-binding factor 2* »

Protéine qui promeut la transcription du gène de l'interleukine lors de l'activation des lymphocytes T. Permet également la formation de la PRKCD stable sur l'ADN [Shamanna *et al.*, 2011]. Interagit avec le facteur ILF3.

**Uniprot** <http://www.uniprot.org/uniprot/Q12905>

### ILF3

---

**Descriptif** « *interleukin enhancer-binding factor 3* »

Protéine qui promeut la transcription du gène de l'interleukine lors de l'activation des lymphocytes T. Permet également la formation de la PRKCD stable sur l'ADN [Shamanna *et al.*, 2011]. Interagit avec le facteur ILF2. Peut également agir comme un inhibiteur de la traduction en se fixant sur les ARNm.

**Uniprot** <http://www.uniprot.org/uniprot/Q12906>

## IMDH1

---

**Descriptif** « *inosine-5' monophosphate deshydrogenase 1* »

Impliquée dans la synthèse *de novo* des nucléotides de guanine. L'activité de cette protéine a donc un impact important sur la croissance cellulaire.

**Uniprot** <http://www.uniprot.org/uniprot/P20839>

## IMDH2

---

**Descriptif** « *Inosine-5'-monophosphate dehydrogenase 2* »

Protéine réalisant la première étape de la synthèse des nucléotides de guanine qui en est aussi l'étape limitante. La régulation de son activité permet donc de contrôler la croissance cellulaire. Elle pourrait jouer un rôle dans le métabolisme des acides nucléiques, et dans le développement des tumeurs.

**Uniprot** <http://www.uniprot.org/uniprot/P12268>

## IN35

---

**Descriptif** « *interferon-induced 35 kDa protein* »

Cette protéine semble avoir été reliée à la voie de signalisation médiée par les cytokines.

**Uniprot** <http://www.uniprot.org/uniprot/P80217>

L'étude de cette protéine dont le rôle est méconnu serait intéressant, car il permettrait l'obtention de résultats originaux.

## KHDR1

---

**Descriptif** « *KH domain-containing, RNA-binding, signal transduction-associated protein 1* »

Impliquée dans la progression du cycle cellulaire en phase G2-M. De manière intéressante, son isoforme 3 inhibe la phase S, et n'est exprimé que dans les cellules dont la croissance a été stoppée.

**Uniprot** <http://www.uniprot.org/uniprot/Q07666>

## KIF2C

---

**Descriptif** « *kinesin-like protein KIF2C* »

Responsable de la dépolymérisation des extrémités positives des microtubules dans les cellules mitotiques, et agit en interaction avec KIF18B.

**Uniprot** <http://www.uniprot.org/uniprot/Q99661>

### LDB1

---

**Descriptif** « *LIM domain binding protein 1* »

Se fixe aux domaines LIM (domaines impliqués dans les liaisons protéine-protéine) d'un grand nombre de facteurs de transcription, et pourrait ainsi être impliqué dans la régulation de la transcription de certaines protéines.

**Uniprot** <http://www.uniprot.org/uniprot/Q86U70>

### MDC1

---

**Descriptif** « *mediator of DNA damage checkpoint protein 1* »

Connue pour son activité de senseur des dommages de l'ADN. Elle interagit avec la kinase ATM et fait partie de la voie de signalisation qui induit l'arrêt du cycle cellulaire lors de lésions de l'ADN. Cette protéine servirait également de plateforme pour le recrutement de protéines de réparation de l'ADN.

**Uniprot** <http://www.uniprot.org/uniprot/Q14676>

### MNT

---

**Descriptif** « *max-binding protein MNT* »

Répresseur transcriptionnel qui est impliqué, entre autres dans la régulation négative de l'apoptose.

**Uniprot** <http://www.uniprot.org/uniprot/Q99583>

### MORC2

---

**Descriptif** « *MORC family CW-type zinc finger protein 2* »

Répresseur transcriptionnel qui régule négativement l'expression de l'anhydrase carbonique 9 *via* son interaction avec l'histone désacétylase 4.

**Uniprot** <http://www.uniprot.org/uniprot/Q9Y6X9>

### MSH2

---

**Descriptif** « *DNA mismatch repair protein Msh2* »

Protéine impliquée dans la reconnaissance des mésappariements des bases de l'ADN en interaction avec MSH6 (p 173). Elle induit ensuite la

réparation de ces derniers par la voie du MMR (« *Mismatch repair* »). Le complexe MSH2-MSH6 (homologue de MutS $\alpha$ ) pourrait également être impliqué dans la réparation des cassures double-brin par recombinaison homologue [Gradia *et al.*, 1999; Iaccarino *et al.*, 1998]. Suite à la détection des dommages, MSH2 interagit également avec MSH3 pour former le complexe MutS $\beta$ , qui intervient dans les processus de réparation.

**Uniprot** <http://www.uniprot.org/uniprot/P43246>

### MSH3

---

**Descriptif** « *DNA mismatch repair protein Msh3* »

Protéine du complexe MutS $\beta$  qui intervient dans l'initiation de la réparation des mésappariements des bases de l'ADN.

**Uniprot** <http://www.uniprot.org/uniprot/P20585>

### MSH6

---

**Descriptif** « *DNA mismatch repair protein Msh6* »

Protéine du complexe MutS $\alpha$  (documenté dans la section consacrée à MSH2) qui intervient dans la détection et la réparation des mésappariements des bases de l'ADN.

**Uniprot** <http://www.uniprot.org/uniprot/P52701>

### NEUA

---

**Descriptif** « *N-acetylaminate cytidyltransferase* »

Enzyme dont le rôle est de synthétiser du CMP-N-acylneuramate, un marqueur des chaînes oligosaccharidique à partir de N-acylneuramate et de CTP.

**Uniprot** <http://www.uniprot.org/uniprot/Q8NFW8>

### NFIA

---

**Descriptif** « *nuclear factor 1 A-type* »

Reconnaît une séquence palindrome présente dans l'ADN des virus à ADN. Elle joue un rôle d'activateur transcriptionnel, mais est également capable de favoriser la réplication.

**Uniprot** <http://www.uniprot.org/uniprot/Q12857>

**NFIB**

---

**Descriptif** « *nuclear factor 1 B-type* »

Facteur reconnaissant la séquence des promoteurs viraux. Il peut également activer la transcription et la réplication.

**Uniprot** <http://www.uniprot.org/uniprot/000712>

**NOLC1**

---

**Descriptif** « *Nucleolar and coiled-body phosphoprotein 1* »

Protéine dotée d'une activité GTPase et ATPase, impliquée dans la nucléogénèse. Elle pourrait jouer un rôle dans le maintien de la structure du centre fibrillaire et des composants fibrillaires dense du noyau.

**Uniprot** <http://www.uniprot.org/uniprot/Q14978>

**NONO**

---

**Descriptif** « *non-POU domain-containing octamer-binding protein* »

NONO est une protéine de la physiologie de l'ADN (sous sa forme simple- ou double-brin) impliquée dans plusieurs processus nucléaires comme l'épissage, la rétention des ARN défectueux en interaction avec la protéine SFPQ (décrite ci-dessous), le déroulement de la double hélice (indirectement en modulant l'activité de TOP1 (p 187)). Le complexe SFPQ / NONO intervient aussi dans la réparation des cassures double-brin de l'ADN dans la voie du NHEJ canonique en permettant la formation d'un complexe de préligation fonctionnel (*via* son interaction avec les protéines XRCC5 (p 190) et XRCC6 (p 190)) [Li *et al.*, 2014].

**Uniprot** <http://www.uniprot.org/uniprot/Q15233>

**NOP2**

---

**Descriptif** « *probable 28S rRNA cytosine-methyltransferase* »

Jouerait un rôle dans la régulation de la progression du cycle cellulaire.

**Uniprot** <http://www.uniprot.org/uniprot/P46087>

**PAR14**

---

**Descriptif** « *poly [ADP-ribose] polymerase 14* »

Possède une activité ADP-ribosyltransférase impliquée dans la régulation de la transcription dépendante de STAT6, mais aussi le mécanisme de survie cellulaire dépendant de JNK-2 [Barbarulo *et al.*, 2013].

**Uniprot** <http://www.uniprot.org/uniprot/Q460N5>

### PARP1

---

**Descriptif** « *poly [ADP-ribose] polymerase 1* »

PARP1 est une protéine impliquée dans la réparation des dommages de l'ADN. Elle intervient à la fois dans la voie de réparation par excision de base (BER) pour la réparation des lésions de l'ADN simple-brin et dans la voie alternative du NHEJ pour la réparation des cassures double-brin de l'ADN [Beck *et al.*, 2014]. Elle joue également un rôle de signalisation *via* son activité de parylation [Ahel *et al.*, 2008] et permet ainsi, entre autres, d'initier la réponse aux dommages de l'ADN médiée par p53/TP53 [Yan *et al.*, 2013].

**Uniprot** <http://www.uniprot.org/uniprot/P09874>

### PCF11

---

**Descriptif** « *pre-mRNA cleavage complex 2 protein Pcf11* »

Protéine qui fait partie du complexe chargé de cliver les pré-ARNm.

**Uniprot** <http://www.uniprot.org/uniprot/094913>

### PDS5B

---

**Descriptif** « *sister-chromatid cohesion protein PDS5 homolog B* »

Module la cohésion des chromatides soeurs pendant la mitose et, en tant que telle, régule négativement le cycle cellulaire. La cohésion des chromatides soeurs peut être cruciale pour la réparation de l'ADN, en particulier par la voie de recombinaison homologue, mais pour l'instant l'implication de la protéine PDS5B dans cette voie de réparation n'a pas été observée.

**Uniprot** <http://www.uniprot.org/uniprot/Q9NTI5>

### PNKP

---

**Descriptif** « *bifunctional polynucleotide phosphatase/kinase* »

Joue un rôle majeur dans la réparation de l'ADN. PNKP intervient à la fois dans la voie de réparation par jonction des extrémités non homologues (NHEJ) et dans la voie de réparation par excision de base (BER). Cette phosphatase kinase permet de phosphoryler une extrémité 5'OH pour générer une extrémité 5' phosphate ou de déphosphoryler une extrémité 3' phosphate pour générer une extrémité 3'OH et de permettre ainsi la ligation

d'extrémités dont la structure initiale était incompatible avec l'action d'une ligase [Jilani *et al.*, 1999].

**Uniprot** <http://www.uniprot.org/uniprot/Q96T60>

### POTE1

---

**Descriptif** « *protection of telomeres protein 1* »

Protéine impliquée dans le complexe shelterin recruté par TERF2 (p 186) au niveau des télomères. Cette protéine est impliquée dans la réplication des extrémités des chromosomes, dans la régulation de leur longueur et dans leur protection [de Lange, 2005].

**Uniprot** <http://www.uniprot.org/uniprot/Q9NUX5>

### PP1A

---

**Descriptif** « *serine/threonine-protein phosphatase PP1-alpha catalytic subunit* »

La protéine phosphatase 1 (PP1) reconnaît plus de 200 substrats et est essentielle pour la mise en œuvre de mécanismes divers tels que la division cellulaire, la synthèse de protéines, ou encore la régulation du métabolisme du glycogène. Dans le complexe PTX / PP1 elle joue un rôle dans le contrôle de la structure chromatiniennne et permet ainsi le contrôle de la progression du cycle cellulaire. Lors de la réponse aux dommages de l'ADN, elle interagit avec la protéine NEK2 [Mi *et al.*, 2007].

**Uniprot** <http://www.uniprot.org/uniprot/P62136>

### PP1RA

---

**Descriptif** « *serine/threonine-protein phosphatase 1 regulatory subunit 10* »

La protéine PP1RA est une protéine impliquée dans la formation du complexe PTW / PP1 dont la fonction est décrite ci-dessus.

**Uniprot** <http://www.uniprot.org/uniprot/Q96QC0>

### PRC1

---

**Descriptif** « *protein regulator of cytokinesis 1* »

Régulateur essentiel de la cytokinèse qui pontent les microtubules entre eux. PRC1 est requise pour le recrutement de nombreuses protéines au niveau du fuseau lors de la cytokinèse [Subramanian *et al.*, 2010].

**Uniprot** <http://www.uniprot.org/uniprot/043663>

## PRKDC

---

**Descriptif** « *DNA-dependent protein kinase, catalytic subunit* »

La sous-unité catalytique de la protéine kinase ADN-dépendante (DNA-PKcs ou ici PRKDC) est une protéine essentielle de la voie classique de la jonction des extrémités non-homologues permettant la réparation des cassures double-brin de l'ADN.

**Uniprot** <http://www.uniprot.org/uniprot/P78527>

La présence de cette protéine sur nos phase est importante car elle montre que notre système permet de purifier des complexes de haut poids moléculaire. L'implication de cette protéine dans plusieurs voies de signalisation nous permet également de supposer que l'identification par spectrométrie de masse permettra de mettre en valeur de nombreux mécanismes cellulaires mis en place sur notre déterminant d'affinité.

## PRP8

---

**Descriptif** « *pre-mRNA-processing-splicing factor 8* »

Permet le recrutement de protéines et d'ARNsn impliqués dans l'épissage alternatif, mais également le positionnement des ARNsn sur le pré-ARNm.

**Uniprot** <http://www.uniprot.org/uniprot/Q6P2Q9>

## PWP2

---

**Descriptif** « *periodic tryptophan protein 2 homolog* »

Protéine impliquée dans la biogenèse des ribosomes [Dasil et Bustelo, 2004].

**Uniprot** <http://www.uniprot.org/uniprot/Q15269>

## RAD50

---

**Descriptif** « *Radiation repair protein 50* »

Fait partie du complexe MRN (avec les protéines Mre11 et Nbs1) qui est un élément central de la réparation des cassures double-brin de l'ADN en tant que senseur des dommages de l'ADN (*via* l'activation de la kinase ATM, [Lee et Paull, 2004]), mais aussi pour la resection des extrémités avant la recombinaison (grâce à l'activité nucléasique de Mre11) et potentiellement pour réaliser la synapse entre les extrémités d'ADN libres avant la jonction des extrémités non homologues (ce qui serait le rôle de RAD50). Ce complexe, qui a une haute affinité pour les cassures double-brin de l'ADN, participe également à la maintenance des télomères.

**Uniprot** <http://www.uniprot.org/uniprot/Q92878>

**RBM15**

---

**Descriptif** « *putative RNA-binding protein 15* »**Uniprot** <http://www.uniprot.org/uniprot/Q96T37>**RBM26**

---

**Descriptif** « *RNA-binding protein 26* »

Se fixe sur les acides nucléiques et en particulier sur les ARN poly(A), mais dont le rôle n'est pas précisément connu.

**Uniprot** <http://www.uniprot.org/uniprot/Q5T8P6>**RD23B**

---

**Descriptif** « *UV excision repair protein RAD23 homolog B* »

Décrite comme faisant partie du complexe XPC, impliqué dans la réparation des dommages à l'ADN par excision de nucléotide (NER). Cette voie de réparation est notamment utilisée pour réparer les dommages induits par l'exposition à un rayonnement ultraviolet.

**Uniprot** <http://www.uniprot.org/uniprot/P54727>**RED**

---

**Descriptif** « *function not documented* »

Fonction non documentée.

**Uniprot** <http://www.uniprot.org/uniprot/Q13123>**RFA2**

---

**Descriptif** « *replication protein A 32 kDa subunit* »

Protéine du complexe de réplication RPA/RP-A, dont le rôle est de se fixer sur l'ADN simple-brin lors de la réplication ou suite à un stress afin de le stabiliser. Cette protéine joue donc un rôle particulièrement important lors de la réplication, mais aussi dans la réponse aux dommages de l'ADN. Elle contrôle en effet l'activation du point de contrôle du cycle cellulaire et des mécanismes de réparation. Elle interagit avec les protéines de la recombinaison homologue RAD51 et RAD52 [Park *et al.*, 1996] pour permettre le remodelage de la chromatine en réponse à la détection de cassures double-brin, mais elle peut également permettre le recrutement des protéines XPA et XPG du NER [Aboussekhra *et al.*, 1995; He *et al.*, 1995] et serait également impliquée dans la réparation des dommages par la voie

du BER [DeMott *et al.*, 1998]. Son rôle dans l'activation du point de contrôle du cycle cellulaire se ferait par l'activation indirecte de CHK1 [Gong et Chen, 2011]. Lors de la phase S elle permet aussi de recruter SMARCA1/HARP, un complexe qui permet de faire redémarrer les fourches de réplifications bloquées au niveau d'un dommage de l'ADN. Enfin RFA2 pourrait être aussi impliquée dans la maintenance des télomères [Grudic *et al.*, 2007].

**Uniprot** <http://www.uniprot.org/uniprot/P15927>

### RFC1

---

**Descriptif** « *replication factor C subunit 1* »

Sous-unité de la protéine accessoire RF-C impliquée dans l'élongation du brin d'ADN lors de la répllication, mais qui est aussi essentielle lors de la réparation de l'ADN par excision de nucléotide (NER) [Overmeer *et al.*, 2010] et par recombinaison homologue (HR).

**Uniprot** <http://www.uniprot.org/uniprot/P35251>

### RFC3

---

**Descriptif** « *replication factor C subunit 3* »

Sous-unité de la protéine accessoire RF-C impliquée dans l'élongation du brin d'ADN lors de la répllication, mais qui est aussi essentielle lors de la réparation de l'ADN par excision de nucléotide (NER) [Overmeer *et al.*, 2010] et par recombinaison homologue (HR).

**Uniprot** <http://www.uniprot.org/uniprot/P40938>

### RFC4

---

**Descriptif** « *replication factor C subunit 4* »

Sous-unité de la protéine accessoire RF-C impliquée dans l'élongation du brin d'ADN lors de la répllication, mais qui est aussi essentielle lors de la réparation de l'ADN par excision de nucléotide (NER) [Overmeer *et al.*, 2010] et par recombinaison homologue (HR).

**Uniprot** <http://www.uniprot.org/uniprot/P35249>

### RFX5

---

**Descriptif** « *Regulatory factor X protein 5* »

Reconnaît les « X-boxes » contenues dans les promoteurs des gènes du complexe majeur d'histocompatibilité de classe II et active leur transcription.

**Uniprot** <http://www.uniprot.org/uniprot/P48382>

### RLA0

---

**Descriptif** « *60S acidic ribosomal protein P0* »

Protéine ribosomale qui intervient dans l'étape d'élongation de la synthèse des protéines.

**Uniprot** <http://www.uniprot.org/uniprot/P05388>

### RLA1

---

**Descriptif** « *60S acidic ribosomal protein P1* »

Protéine ribosomale qui intervient dans l'étape d'élongation de la synthèse des protéines.

**Uniprot** <http://www.uniprot.org/uniprot/P05386>

### RLA2

---

**Descriptif** « *60S acidic ribosomal protein P2* »

Protéine ribosomale qui intervient dans l'étape d'élongation de la synthèse des protéines.

**Uniprot** <http://www.uniprot.org/uniprot/P05387>

### RL13

---

**Descriptif** « *60S ribosomal protein L13* »

Protéine appartenant à la sous-unité 60S des ribosomes.

**Uniprot** <http://www.uniprot.org/uniprot/P26373>

### RPA1

---

**Descriptif** « *DNA-directed RNA polymerase I subunit 1* »

Sous-unité de l'ARN polymérase I, qui catalyse la transcription de l'ADN en ARN.

**Uniprot** <http://www.uniprot.org/uniprot/095602>

## RPA2

---

**Descriptif** « *DNA-directed RNA polymerase I subunit 2* »

Sous-unité de l'ARN polymérase I, qui catalyse la transcription de l'ADN en ARN.

**Uniprot** <http://www.uniprot.org/uniprot/Q9H9Y6>

## RPAB1

---

**Descriptif** « *DNA-directed RNA polymerase I subunit* »

Sous-unité de l'ARN polymérase I, qui catalyse la transcription de l'ADN en ARN. La sous-unité RPAB1 est également présente dans d'autres RNA polymérases (ARN polymérases II et III).

**Uniprot** <http://www.uniprot.org/uniprot/P19388>

## RPAC1

---

**Descriptif** « *DNA-directed RNA polymerase I subunit* »

Sous-unité de l'ARN polymérase I, qui catalyse la transcription de l'ADN en ARN. La sous-unité RPAC1 est également présente dans l'ARN polymérase III.

**Uniprot** <http://www.uniprot.org/uniprot/015160>

## RU17

---

**Descriptif** « *U1 small nuclear ribonucleoprotein 70 kDa* »

Protéine impliquée dans l'épissage des ARNm. Elle fait partie de la sous-unité U1 du spliceosome.

**Uniprot** <http://www.uniprot.org/uniprot/P08621>

## SAFB1

---

**Descriptif** « *scaffold attachment factor B1* »

Protéine du complexe de transcription qui serait capable d'inhiber la prolifération cellulaire (ces fonctions ont été attribuées par similarité avec les protéines de type SAFA).

**Uniprot** <http://www.uniprot.org/uniprot/Q15424>

**SEPT2**

---

**Descriptif** « *septin 2* »

Possède une activité GTPasique et peut polymériser en des filaments cytosquelettiques. Elle est requise pour une progression normale à travers la mitose. Elle contribue à l'établissement de la structure moléculaire permettant le maintien de CENPE aux kinétochores et donc la migration des chromosomes. Pendant l'anaphase, elle pourrait jouer un rôle dans la ségrégation des chromosomes.

**Uniprot** <http://www.uniprot.org/uniprot/Q15019>

**SEPT7**

---

**Descriptif** « *septin 7* »

Possède une activité GTPasique et peut polymériser en des filaments cytosquelettiques. Elle est requise pour une progression normale à travers la mitose. Est aussi impliquée dans la cytokinèse et est requise pour une association normale entre CENPE et le kinétochore.

**Uniprot** <http://www.uniprot.org/uniprot/Q16181>

**SEPT9**

---

**Descriptif** « *septin-9* »

Enzyme GTPase qui pourrait avoir un rôle dans la cytokinèse et donc dans la division cellulaire. Ce rôle a été prédit en raison de la similarité de séquence avec les autres septines.

**Uniprot** <http://www.uniprot.org/uniprot/Q9UHD8>

**SEP11**

---

**Descriptif** « *septin 11* »

Possède une activité GTPasique et polymérise pour former des filaments cytosquelettiques. Pourrait être impliquée dans la cytokinèse. Interagit avec SEPT7 et SEPT9 (p 182).

**Uniprot** <http://www.uniprot.org/uniprot/Q9NVA2>

**SF3A1**

---

**Descriptif** « *splicing factor 3A subunit 1* »

Fait partie du facteur d'épissage SF3A. Même si elle n'est, à priori, pas directement impliquée dans la réparation des dommages de l'ADN, elle

interagit avec la protéine SAF-A / hn-RPU dont il a récemment été démontré qu'elle est un substrat de l'enzyme PRKDC (p 177).

**Uniprot** <http://www.uniprot.org/uniprot/Q15459>

### SF3A2

---

**Descriptif** « *splicing factor 3A, subunit 2* »

Fait partie du même complexe d'épissage que la protéine SF3A1 (p 182).

**Uniprot** <http://www.uniprot.org/uniprot/Q15428>

### SFPQ

---

**Descriptif** « *splicing factor, proline- and glutamine-rich* »

Impliquée dans un grand nombre de mécanismes cellulaires. Outre le fait qu'elle intervienne dans l'épissage (elle est requise pour la formation du spliceosome), on peut également citer son rôle dans la réparation de l'ADN, en association avec la protéine NONO (p 174) (interaction avec les protéines XRCC5 et XRCC6 pour former un complexe de pré ligation fonctionnel dans la voie du NHEJ, [Bladen et al. \[2005\]](#)).

**Uniprot** <http://www.uniprot.org/uniprot/P23246>

### SK2L2

---

**Descriptif** « *superkiller viralicidic activity 2-like 2* »

Serait impliquée dans l'épissage des pré ARNm.

**Uniprot** <http://www.uniprot.org/uniprot/P42285>

### SMAL1

---

**Descriptif** « *SWI/SNF related matrix associated actin dependant regulator of chromatin subfamily A like protein 1* »

Hélicase dont le mécanisme nécessite l'utilisation d'ATP. Elle intervient dans l'enroulement de l'ADN simple brin lié par la protéine RPA. Elle intervient également de manière générale pour réenrouler des brins dissociés. Elle jouerait également un rôle important dans les mécanismes de réponse aux dommages de l'ADN [[Ciccia et al., 2009](#)].

**Uniprot** <http://www.uniprot.org/uniprot/Q9NZC9>

### SMCA5

---

**Descriptif** « *SWI/SNF related matrix associated actin-dependant regulator of chromatin, subfamilyA, member 5* »

Interagit avec BAZ1A (p 154), en particulier lors de la réplication de l'hétérochromatine péri-centromérique. Ces deux protéines font de plus partie du complexe WHICH, qui médie la phosphorylation des histones  $\gamma$ H<sub>2</sub>AX et qui reconfigure les structures anormales de la chromatine (SMCA5 est le domaine ATPasique de ce complexe et BAZ1A en est la sous-unité accessoire et non catalytique).

**Uniprot** <http://www.uniprot.org/uniprot/060264>

### SMU1

---

**Descriptif** « *WD40 repeat-containing protein SMU1* »

Non documenté. Présente à la fois dans le cytoplasme et dans le noyau des cellules.

**Uniprot** <http://www.uniprot.org/uniprot/Q2TAY7>

### SNUT1

---

**Descriptif** « *U4/U6.U5 tri-snRNP-associated protein 1* »

Intervient dans l'épissage des ARN messagers pour laquelle un rôle a également été décrit dans l'arrêt du cycle cellulaire et dans le déclenchement de l'apoptose [Hosokawa *et al.*, 2005].

**Uniprot** <http://www.uniprot.org/uniprot/043290>

### SP16H

---

**Descriptif** « *suppressor of Ty 16 homolog* »

Protéine qui fait partie du complexe FACT, lequel est impliqué dans de nombreux mécanismes tels que l'élongation des ARNm, la réplication et la réparation de l'ADN. Ce complexe peut agir en tant que chaperonne des histones, qui déstabilise puis reforme les nucléosomes au passe de l'ARN polymérase II (agit au niveau de la dissociation d'un dimère H2A-H2B, Orphanides *et al.* [1999]; Pavri *et al.* [2006]). En réponse à l'endommagement de l'ADN, ce complexe phosphoryle également la protéine p53 qui intervient dans la régulation du cycle cellulaire et peut provoquer l'apoptose des cellules [Keller *et al.*, 2001].

**Uniprot** <http://www.uniprot.org/uniprot/Q9Y5B9>

## SRRT

---

**Descriptif** « *Serrate RNA effector molecule homolog* »

Impliquée dans le silencing de gènes grâce à son rôle de stabilisation et de transport des micro-ARN coiffés provenant du « *cap-binding complex* ». Cette protéine est également impliquée dans le passage des cellules en phase S.

**Uniprot** <http://www.uniprot.org/uniprot/Q9BXP5>

## SSBP

---

**Descriptif** « *single-stranded DNA-binding protein* »

Protéine qui se fixe préférentiellement sur l'ADN simple-brin dans les mitochondries. Elle jouerait un rôle important lors de la réplication, et pourrait également contrôler l'activité de certaines hélicases.

**Uniprot** <http://www.uniprot.org/uniprot/Q04837>

## SUH

---

**Descriptif** « *suppressor of hairless protein* »

Cette protéine est impliquée dans la régulation de la transcription, et en particulier dans la communication inter-cellulaire *via* la voie de signalisation Notch [Oswald *et al.*, 2005]. Pourrait interagir avec des histones acétylases / désacétylases afin de moduler la transcription en modifiant l'état de condensation de la chromatine.

**Uniprot** <http://www.uniprot.org/uniprot/Q06330>

## T2FA

---

**Descriptif** « *general transcription factor IIF subunit 1* »

Protéine faisant partie du complexe de l'ARN polymérase II qui joue un rôle majeur dans l'activation de la transcription.

**Uniprot** <http://www.uniprot.org/uniprot/P35269>

## T2FB

---

**Descriptif** « *general transcription factor IIF subunit 2* »

Facteur général qui joue un rôle majeur dans la transcription des gènes eucaryotes par l'ARN polymérase II.

**Uniprot** <http://www.uniprot.org/uniprot/P13984>

### TAOK1

---

**Descriptif** « *serine/threonine-protein kinase TAOK1* »

Impliquée dans la réponse aux dommages de l'ADN. Elle est jouée en effet un rôle dans le point de contrôle pour la transition des cellules en phase G2/M en activant la voie de cascade de la p38 MAP kinase en réponse aux dommages de l'ADN [Raman *et al.*, 2007]. TAOK1 aurait également un rôle dans la stabilité du cytosquelette.

**Uniprot** <http://www.uniprot.org/uniprot/Q7L7X3>

### TBL3

---

**Descriptif** « *transducin beta-like protein 3* »

Interagit avec les ARN [Castello *et al.*, 2012] et serait impliquée dans la voie de signalisation de la protéine G couplée à son récepteur.

**Uniprot** <http://www.uniprot.org/uniprot/Q12788>

### TCOF

---

**Descriptif** « *treacle protein* »

Acteur du transport nucléolaire et cytoplasmique. Elle est également impliquée dans la biogenèse des ribosomes [Valdez *et al.*, 2004].

**Uniprot** <http://www.uniprot.org/uniprot/Q13428>

### TERF2

---

**Descriptif** « *telomeric repeat-binding factor 2* »

Protéine qui se lie aux répétitions 5' TTAGGG 3' d'ADN double brin présentes au niveau des télomères. Cette protéine permet aussi de recruter les complexes chargés de protéger les télomères. Elle inhibe la réparation de l'ADN médiée par homologie qui aurait des conséquences sur la longueur de ces derniers [de Lange, 2005].

**Uniprot** <http://www.uniprot.org/uniprot/Q15554>

La présence de cette protéine sur nos phase est intéressante en ce qu'elle ne reconnaît pas l'extrémité libre de l'ADN comme un mime de cassure double-brin mais comme une extrémité. Le rôle de cette protéine nous laisse malheureusement entrevoir une contamination de notre échantillon par des protéines télomériques et les protéines qui leur sont associées, contamination dont il ne sera pas possible de se départir avec notre système.

## TF2H4

---

**Descriptif** « *general transcription factor IIH subunit 4* »

Facteur de transcription impliqué dans la voie de réparation des dommages de l'ADN par NER [Kuper *et al.*, 2014] qui peut également, lorsqu'il est associé à CAK permettre la transcription des ARN par l'ARN polymérase II.

**Uniprot** <http://www.uniprot.org/uniprot/Q92759>

## TOP1

---

**Descriptif** « *DNA topoisomérase 1* »

Enzyme dont le rôle est de relâcher les superenroulements de l'ADN. Elle possède un mécanisme de « passage de brin », c'est à dire qu'elle est capable d'introduire une cassure simple-brin dans l'ADN puis de religuer le brin clivé pour relâcher l'ADN. Les complexe de clivage produits par TOP1 sont retrouvés dans le mécanisme d'apoptose en présence d'inhibiteurs de la tubuline et de la topoisomérase II [Sordet *et al.*, 2006].

**Uniprot** <http://www.uniprot.org/uniprot/P11387>

## TOP2A

---

**Descriptif** « *DNA-topoisomérases II  $\alpha$*  »

Introduit des cassures double-brin dans l'ADN afin d'en modifier la topologie. L'enzyme TOP2A a également été décrite comme ayant une activité de ligation [Bromberg *et al.*, 2004].

**Uniprot** <http://www.uniprot.org/uniprot/P11388>

## TOP2B

---

**Descriptif** « *DNA-topoisomérases II  $\beta$*  »

Introduit des cassures double-brin dans l'ADN afin d'en modifier la topologie. L'enzyme TOP2A a également été décrite comme ayant une activité de ligation [Bromberg *et al.*, 2004].

**Uniprot** <http://www.uniprot.org/uniprot/Q02880>

## TOX4

---

**Descriptif** « *thymocyte selection-associated high mobility group box protein 4* »

Protéine impliquée dans le complexe PTW/PP1 [PP1A (p 176)] qui régule la condensation de la chromatine et la progression du cycle cellulaire.

**Uniprot** <http://www.uniprot.org/uniprot/094842>

### TR150

---

**Descriptif** « *thyroin hormone receptor associated protein 3* »

Protéine impliquée dans l'épissage des pré ARNm. Cette protéine est également impliquée dans la réponse aux dommages de l'ADN, probablement au sein du complexe SNARP [Beli *et al.*, 2012].

**Uniprot** <http://www.uniprot.org/uniprot/Q9Y2W1>

Il serait particulièrement intéressant d'étudier cette protéine, car elle est décrite pour être exclue des sites de cassure, mais sa déplétion provoque toutefois une hypersensibilisation des cellules aux rayonnements ionisants. Elle serait donc importante pour la réparation des dommages, mais son rôle précis n'est pas encore documenté. Dans nos expériences cette protéine est spécifique dans la bande à 240 kDa dont elle n'est pas exclusive. La non spécificité de TR150 dans les autres bandes s'explique probablement par sa capacité à reconnaître les acides nucléiques même en dehors du contexte de la réparation de l'ADN.

### TTF2

---

**Descriptif** « *transcription termination factor 2* »

Possède une activité ATPasique dépendante de l'ADN double-brin. Elle serait impliquée dans la répression de la transcription lors de la mitose.

**Uniprot** <http://www.uniprot.org/uniprot/Q9UNY4>

### U520

---

**Descriptif** « *U5 small nuclear ribonucleoprotein 200 kDa helicase* »

Impliquée dans l'assemblage du spliceosome et joue donc un rôle essentiel dans l'épissage des pré-ARNm.

**Uniprot** <http://www.uniprot.org/uniprot/075643>

### VIME

---

**Descriptif** « *vimentin* »

Protéine formant des filaments intermédiaires de classe III, présente en particulier dans les cellules mésenchymateuse. Dans les cellules, elle est attachée au noyau, au reticulum endoplasmique ou encore aux mitochondries.

**Uniprot** <http://www.uniprot.org/uniprot/P08670>

### WDR36

---

**Descriptif** « *WD repeat-containing protein 36* »

Protéine dont l'interaction avec les ARN a été décrite par [Castello et al.](#) en 2012 et qui est impliquée dans leur processing.

**Uniprot** <http://www.uniprot.org/uniprot/Q8NI36>

### WIZ

---

**Descriptif** « *protein Wiz* »

Protéine dont le rôle pourrait être de séquestrer les protéines EHMT1 et EHMT2 (des histones méthyltransférases) afin de réprimer la transcription.

**Uniprot** <http://www.uniprot.org/uniprot/095785>

### XPC

---

**Descriptif** « *DNA repair protein complementing XP-C cells* »

Interagit avec la protéine RD23B décrite plus haut au sein du complexe XPC. Ce complexe fait partie de la voie de réparation par excision de nucléotide destiné à réparer les cassures simple-brin de l'ADN.

**Uniprot** <http://www.uniprot.org/uniprot/Q01831>

### XRCC1

---

**Descriptif** « *DNA repair protein XRCC1* »

Protéine impliquée dans la réparation de l'ADN et intervenant dans les voies de réparation du BER et du NHEJ alternatif. Elle interagit avec DNLI3 (p 161).

**Uniprot** <http://www.uniprot.org/uniprot/P18887>

### XRCC4

---

**Descriptif** « *DNA repair protein XRCC4* »

Protéine impliquée dans la réparation de l'ADN par la voie du NHEJ canonique. Forme le complexe de ligation avec DNLI4 (p 161) et XLF (p 193). Ce complexe est recruté sur les cassures par PRKDC (p 177) [[Chen et al., 2000](#)].

**Uniprot** <http://www.uniprot.org/uniprot/Q13426>

### XRCC5

---

**Descriptif** « *X-ray repair cross-complementing protein 5* »

La protéine XRCC5 (Ku80) a une activité hélicase ATP-dépendante. Elle est impliquée dans les phénomènes de translocation chromosomique. Elle interagit également avec XRCC6 (décrite ci-dessous) pour former un hétéroduplexe qui se lie à l'ADN. Elle est essentielle pour la voie de réparation des cassures double-brin du NHEJ canonique, et permet le recrutement de PRKDC (p 177) [Schulte-Uentrop *et al.*, 2008; West *et al.*, 1998]. Elle possède également une activité d'excision de nucléotide [Roberts *et al.*, 2010]. En se fixant sur l'ADN avec APEX1, elle peut aussi jouer un rôle de répresseur transcriptionnel.

**Uniprot** <http://www.uniprot.org/uniprot/P13010>

### XRCC6

---

**Descriptif** « *X-ray repair cross-complementing protein 6* »

La protéine XRCC6 possède une activité hélicase et se lie à l'ADN double-brin de préférence lorsqu'il forme des fourches. Elle interagit avec la protéine XRCC5 (décrite ci-dessus) dans la voie de réparation des cassures double-brin de l'ADN du NHEJ canonique et permet de recruter PRKDC sur les extrémités de l'ADN (pour plus d'information, consulter les références indiquées dans la description d'XRCC5).

**Uniprot** <http://www.uniprot.org/uniprot/P12956>

Nous sommes heureux d'avoir identifié les protéines XRCC5 et XRCC6 dans nos échantillons car ces protéines sont abondantes dans le noyau des cellules et sont décrites pour être des senseurs des extrémités libres de l'ADN. Ces deux protéines constituent donc un contrôle positif pour notre méthodologie. D'un point de vue analytique, leur présence est une signature de la mise en place d'un mécanisme de réparation des cassures double-brin par NHEJ canonique sur nos phases.

#### 5.0.4 PROTÉINES COMMENTÉES DANS LE CHAPITRE « RÉSULTATS »

##### ATM

---

**Descriptif** « *serine-protein kinase ATM* »

Protéine kinase qui active les voies de signalisation des cassures double-brin de l'ADN, de l'apoptose ou encore du stress génotoxique, ce qui en fait un senseur des dommages de l'ADN [Bakkenist et Kastan, 2003; Shiloh, 2014]. Son action provoque l'initiation des mécanismes de réponse aux dommages de l'ADN parmi lesquels l'arrêt du cycle cellulaire. Protéine responsable de la phosphorylation des histones  $\gamma$  H<sub>2</sub>AX qui sont des marqueurs précoces des cassures double-brin de l'ADN [Fernandez-Capetillo *et al.*, 2003]. ATM se lie également aux extrémités de l'ADN. Dans un autre contexte, ATM permet aussi la synthèse d'une chaîne lourde d'immunoglobuline en intervenant dans le mécanisme d'exclusion d'allèle [Chaumeil *et al.*, 2013].

**Uniprot**

<http://www.uniprot.org/uniprot/Q13315>

##### CENPE

---

**Descriptif** « *centromere-associated protein E* »

Joue un rôle essentiel dans la stabilité chromosomique en favorisant la stabilisation de la capture des microtubules aux kinétochores. Pendant la métaphase, elle joue un rôle dans la migration des chromosomes à la plaque équatoriale. Elle est requise pour le « *checkpoint* » mitotique et effectue sa surveillance à tous les kinétochores pour éviter les aneuploïdies dues à la perte d'un chromosome.

**Uniprot** <http://www.uniprot.org/uniprot/Q02224>

##### CHK1

---

**Descriptif** « *serine/threonine-protein kinase Chk1* »

Protéine kinase requise pour l'arrêt du cycle cellulaire médiée par checkpoint et pour l'activation des mécanismes de réparation de l'ADN endommagé ou non répliqué [Bertoli *et al.*, 2013; Walworth *et al.*, 1993]. Chk1 est impliquée dans plusieurs mécanismes permettant la sauvegarde de l'intégrité du génôme dans lesquels elle régule négativement la progression du cycle cellulaire. Interagit en particulier avec les phosphatases dont le rôle est d'induire la mitose (CDC25) dont l'inhibition impacte directement l'activité des cyclines dépendantes kinase (CDK) impliquées dans la progression du cycle cellulaire [Kramer *et al.*, 2004]. Favorise la recombinaison homologue en phosphorylant RAD51, ce qui la libère de son interaction avec BRCA2 et

lui permet d'être recrutée sur la chromatine [Sorensen *et al.*, 2005]. Phosphoryle également TP53 de manière à réguler négativement la prolifération cellulaire. Elle pourrait également être impliquée dans la réplication, dans la régulation de la transcription des gènes impliqués dans la progression du cycle cellulaire en phosphorylant les protéines histones.  
**Uniprot** <http://www.uniprot.org/uniprot/014757>

## CHK2

---

**Descriptif** « *serine/threonine-protein kinase Chk2* »

Protéine kinase impliquée, tout comme Chk1 (décrite di-dessus), dans l'arrêt du cycle cellulaire médiée par checkpoint en réponse à la présence de dommages de l'ADN [Matsuoka *et al.*, 1998]. Joue probablement un rôle similaire à celui de Chk1 dans les mécanismes permettant le maintien de l'intégrité génomique et dans la promotion de la réparation des cassures par recombinaison homologue. Protéine suppresseur de tumeur qui contrôle la transcription de gènes pro-apoptotiques en phosphorylant le facteur de transcription E2F1 [Stevens *et al.*, 2003]. Elle régule également l'apoptose en interagissant avec p53/TP53 qu'elle phosphoryle [Zhang *et al.*, 2014].

**Uniprot** <http://www.uniprot.org/uniprot/096017>

## H2AX

---

**Descriptif** « *histone H2AX* »

Variante de l'histone conventionnelle H2A qui la remplace dans certains nucléosomes. Elle est requise pour la mise en place de l'arrêt du cycle cellulaire en réponse aux rayonnements ionisants. Elle est phosphorylée sur sa queue C-terminale, pour produire la  $\gamma$ H2AX, qui est un marqueur de la présence de cassures double-brin de l'ADN [Paull *et al.*, 2000].

**Uniprot** <http://www.uniprot.org/uniprot/P16104>

## P53

---

**Descriptif** « *cellular tumor antigen p53* »

Protéine suppresseur de tumeur qui induit l'arrêt de la croissance cellulaire ou l'apoptose selon le contexte cellulaire et le type de cellule [Yonish-Rouach *et al.*, 1993]. p53 est également impliquée dans l'arrêt du cycle cellulaire en régulant la transcription des gènes qui sont impliqués dans la division cellulaire, et en inhibant l'activité de la kinase cycline-dépendante 7 (CDK7) en réponse à la détection de dommages sur l'ADN [Schneider *et al.*, 1998]. Son effet pro-apoptotique vient en partie de

son effet positif sur la transcription de longs ARN non codants qui sont eux même des répresseurs de la transcription [Huarte *et al.*, 2010].

**Uniprot** <http://www.uniprot.org/uniprot/P04637>

## XLF

---

**Descriptif** « *XRCC4 like factor* »

Protéine impliquée dans la réparation des cassures double-brin de l'ADN. Elle forme avec DNLI4 (p 161) et XRCC4 (p 189) le complexe de ligation chargé de restaurer la continuité du brin d'ADN endommagé [Ahnesorg *et al.*, 2006; Buck *et al.*, 2006]. Aussi connue sous le nom de NHEJ1.

**Uniprot** <http://www.uniprot.org/uniprot/Q9H9Q4>

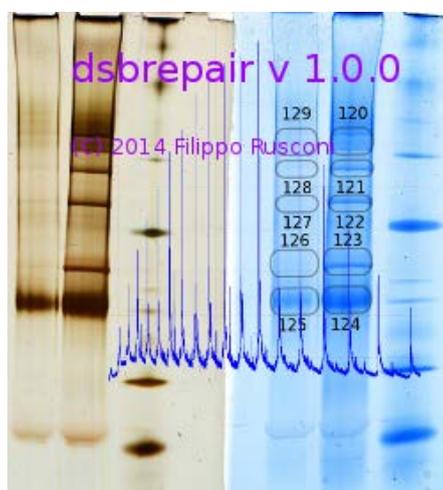


MANUEL DE L'UTILISATEUR DU LOGICIEL *dsbRepair*  
(ANGLAIS)

User Manual for the *dsbrepair*  
mass data inspection program

Implementing

an interface to the *Double-Strand Break Repair* database  
and various mass spectrometric data-mining procedures



Filippo Rusconi<sup>1</sup>

---

1. [filippo\[dot\]rusconi@dash\[dot\]psud\[dot\]fr](mailto:filippo[dot]rusconi@dash[dot]psud[dot]fr) at Laboratoire de Chimie-physique (LCP), UMR8000 CNRS — Université Paris-Sud 11 — Bâtiment 349 — F-91405 Orsay — France

## Abstract

This manual documents the *dsbrepair* program that was developed in-house to mine the mass spectrometric data obtained for a project aiming at characterizing DNA repair complexes that assembled onto specially crafted chromatographic phases. The mass spectrometric data have been accumulating for years and we needed a way to either make sense out of them or to establish ways to compare various bits of data obtained for different biochemical experiments.

*dsbrepair* provides an interface to the mass spectrometric data database in which all the mass data have been injected. *dsbrepair* also make it particularly intuitive to perform computation of various kinds on sets of sample mass data to allow the user to compare various samples both in terms of presence or absence of proteins and in terms of assessing their amount in these same samples.

## 5.1 INTRODUCTION

This manual describes the *dsbrepair* software program that was developed in-house to help mine mass spectrometric data obtained during innumerable experiments related to the purification of DNA repair complexes. This document is to be construed as a companion document to the article that describes the experiments and the obtained results.

In a first section, it will describe where to get the program for different platforms (GNU/Linux, Windows and MacOSX). Then, the installation procedure will be described. Finally, the program features are detailed.

## 5.2 GETTING THE SOFTWARE AND INSTALLING IT

All the software binary packages (pre-built binaries for different platforms) and the source code are freely available at .

Please, take some time to read the GPL-3 file shipped within the source code tarball, the binary packages and available in the appendix () or at <http://www.gnu.org/copyleft/gpl.html>. This is the license of both the software and the documentation. This software is Free Software, as defined by the Free Software Foundation<sup>2</sup>.

This program is free software : you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, version 3 of the License.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY ; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. See the GPL-3 file shipped along with the both source code for this program and the binary packages prepared therefrom.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses>.

The package for the MS-Windows system is an easy-to-install executable file. Double-click its icon to run it and follow the instructions. Please, check the boxes to both launch the program at the end of the install and show the “readme” file. The package for current GNU/Linux distributions is a binary tarball. Uncompress it in a directory.

---

2. See <http://www.fsf.org>.

### 5.3 OVERVIEW OF THE BIOCHEMICAL EXPERIMENTS

In order to be able to understand the design and the benefits apportioned by the *dsbrepair* program in our *DNA repair* scientific project, it is necessary to provide a succinct overview of the experimental setup and goals. The reader who has read the scientific article to which this document is associated, can safely skip this entire section and start reading at section 5.4, entitled [General overview of the graphical user interface](#), page 204.

**Our Scientific Project** is aimed at the characterization of novel proteins involved in the repair of DNA double-strand breaks. One of the fundamental methods used to purify DNA repair complexes is based on specific affinity chromatography using duplex DNA oligonucleotides grafted on the chromatography support, with one of their ends being free to recruit DNA repair proteins (these free ends should therefore be construed as “baits”).

**The Technical Limitations of the Above Approach** are, as described in the scientific article, of two kinds :

- \* One limitation is that the DNA oligonucleotides not only do have free ends so as to act as baits for proteins specifically recognizing these ends (because they mimic DNA double-strand breaks), but are also formed of a long duplex sequence (generally this sequence is between 40–300 bp-long). This duplex DNA sequence will be bound by proteins that have a generic affinity for double-helix DNA (mostly based on a mix of electrostatic and hydrophobic interactions). At elution time, almost of all these proteins will be eluted along with the proteins of interest. This “noise” will impair the identification of the proteins of interest, that is, the proteins that are expected to specifically bind to the free DNA ends;
- \* The other limitation of the traditional way of doing affinity chromatography to purify DNA repair complexes is related to the fact that the chromatographic support onto which the affinity determinant is grafted actually binds a huge amount of unspecific proteins, even moreso when the beads are functionalized with proteins (like streptavidin). At elution time, not only the proteins of interest—those that bound to the free DNA ends—but also the proteins bound to the beads, will be part of the purified matter, thus leading to a massive contamination of the purified product.

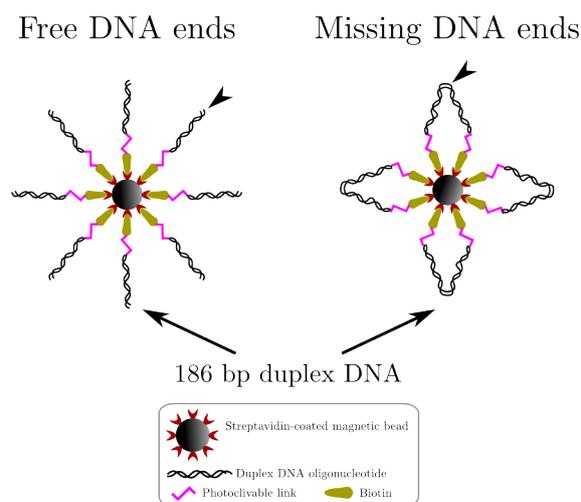


FIGURE 5.1 – **Structure of the two different phases.** The two chromatographic phases are used to perform protein purifications. The phase endowed with free DNA ends ( $\star 1$ ) is expected to allow us to purify DNA repair complexes specific of the DNA double-strand breaks.

**A New Chromatographic System** was thus devised in order to counter the two limitations above by relying on the two improvements below :

- \* Two affinity chromatography phases are produced (Figure 5.1). One phase will have the DNA oligonucleotides bearing only one biotin moiety at one end. These DNA oligonucleotides will thus bind to the streptavidin-coated beads *via* one end only, thus leaving one free end available to recruit specific DNA repair proteins (phase denoted  $\star 1$ ). The other phase will be prepared using DNA oligonucleotides that bear one biotin moiety at each one of their ends. These DNA oligonucleotides will thus bind to the streptavidin-coated beads with both of their ends, thus leaving no free end available for the DNA repair-specific proteins to bind to (phase denoted  $\diamond 1$ ). This latter chromatographic phase should be construed as a “control” phase for the purification experiments ;
- \* After the binding of the DNA-interacting proteins on the DNA oligonucleotides, be these proteins showing a generic affinity for the DNA double helix or proteins assembling specifically at the free DNA ends, we wanted to be able to get rid of the streptavidin-coated beads which, as described above, are hugely protein contaminant-loaded. In order to detach the DNA oligonucleotides (along with their interacting proteins) from the chromatographic support in the mildest conditions, we have inserted an ultraviolet photocleavable linker between the DNA oligo-

nucleotide and each biotin moiety beared by the oligonucleotide. By irradiating the chromatographic slurry in mild UV conditions, right after the purification, we could magnet-sediment the beads and recover the DNA–protein complexes left in the supernatant.

For each experiment, the HeLa human nuclear extracts used for the purifications are deposited onto our functionalized beads batch-wise. Each time, two tubes are handled : one tube containing only the chromatographic phase endowed with free DNA ends (made with mono-biotinylated DNA oligonucleotides) and another tube containing the chromatographic phase devoid of these ends (made with bi-biotinylated DNA oligonucleotides).

**The Resolution and the Visualization** of the purified complexes were performed by running the purified material in a non-denaturing electrophoresis gel. The specific kind of non-denaturing electrophoresis used throughout our project is termed “blue native polyacrylamide gel electrophoresis” (BN-PAGE), as the molecule responsible for the charge of the migrating analytes is not SDS but Coomassie blue<sup>3</sup>. Figure 5.2, page ci-contre shows the typical result obtained during the purification of proteins from a nuclear extract on our two chromatographic phases. The gel was stained with colloidal Coomassie blue. There are three major bands in the \*1 lane that are absent from the <math>\diamond</math>1 lane<sup>4</sup>. These are typically the bands of interest to this study as they are due to the presence of proteins–DNA complexes that assemble specifically of the free DNA ends.

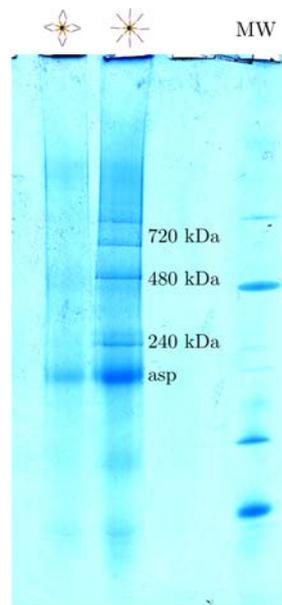
**Mass Spectrometry** was used throughout all of the project to identify the proteins contained in the bands described above. Whenever a band in a \*1 lane was excised from a gel, the same position from the <math>\diamond</math>1 lane was excised as a control. Both bands were then trypsin-treated separately and the peptidic mixtures were analysed separately by LC-MS/MS. The raw mass data were then injected into the *X!TandemPipeline* software,<sup>5</sup> which is a Java wrapper around *X!Tandem* allowing easy GUI-based inspection of the identification data. The protein identification data were then fed to the *SQLite3* database with a home-made XML-formatted data parser.

---

3. The sodium laurylsulfate molecule also known as sodium dodecylsulfate, is a very strong protein denaturing agent, while Coomassie blue is very mildly denaturing.

4. In the gel shown, there are two bands resolved around 720 kDa, but usually only one larger band is seen.

5. <http://pappso.inra.fr/bioinfo/xtandempipeline/>.



**FIGURE 5.2 – Typical blue native PAGE obtained throughout this study.** A nuclear extract was used separately on each chromatographic phase. The ✧1 vignette denotes the chromatographic phase devoid of free DNA ends, while the ✨1 vignette denotes the chromatographic phase endowed with free DNA ends onto which it is expected that DNA repair-specific complexes might assemble. The purified material was significantly different depending on the chromatographic phase used. This, of course, was the expected result. *asp* : aspecific band, since it is present in both lanes.

The **SQLite3 Database's Structure** is described in Figure 5.3. The fields of the `proHits` table are detailed below :

- \* `matches` : the number of peptides that made the protein identification possible ;
- \* `pai` : the PAI value as computed by *X!Tandem* ;
- \* `empai` : the emPAI value as computed by *X!TandemPipeline* ;
- \* `empaiPct` : the percentage of this protein's emPAI with respect to the sum of the emPAI values of all the proteins identified in this sample ;
- \* `accession_id` : key to the `accessions` table.

The fields of the `accessions` table are easily understandable : the `number` field stores the SwissProt database accession number (like `XRCC6_HUMAN`). The `desc` field stores the human-understandable description of the protein (like "X-ray repair cross-complementing protein 6").

The fields of the `experiments` table are detailed below :

- \* `expDate` : the date of the experiment ;
- \* `labBook` : the complete number of the labBook ;
- \* `sampleDesc` : the description of the sample ;
- \* `sampleNumb` : the number of the mass spectrometric experiment, which actually resolves to the number of the sample. This is the way to identify unambiguously one given mass spectrometric experiment, hence one given *mass data set* for one given HPLC-MS/MS run ;
- \* `gelScanFile` : the file name of the scanned Coomassie blue gel ;
- \* `biotinMoiety_id` : key to the `biotinMoieties` table.

The possible values for the `desc` field of the `biotinMoieties` table are easy to understand : value 1 means that the DNA oligonucleotides used to functionalize the beads have only one biotin moiety. This leads to the preparation of a chromatographic phase endowed with free DNA ends (\*1) ; value 2 means that the DNA oligonucleotides used to functionalize the beads have two biotin moieties. This leads to the preparation of a chromatographic phase devoid of free DNA ends (<math>1</math>).

The database holds all the data obtained for a large set of experiments all done in the same experimental conditions. These data are thus considered to be suitable for establishing comparisons in a number of ways, as described in detail below.

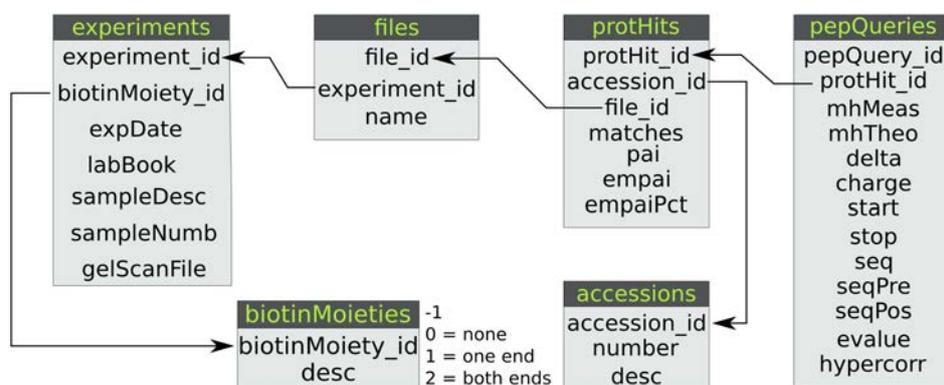


FIGURE 5.3 – **Relational database structure.** The database is structured into a number of tables. The entry point to the database is the **protHits** table, since it is the table that relates to the essential data of interest : accession numbers recorded in the **accessions** table.

**Two main ways of thoroughly comparing proteins** are possible, with our six full data sets involving the various bands observed in  $\ast 1$  lanes as well as in  $\diamond 1$  lanes :

- \* comparing the protein content of a complex from a given band (*e.g.* 720 kDa) from a  $\ast 1$  lane with the contents of the equivalent band in a  $\diamond 1$  lane. This comparison typically enables establishing which proteins are specific of the free DNA ends ( $\ast 1$ ), using the proteins identified in the band belonging to the  $\diamond 1$  lane as unspecific “control” proteins ;
- \* comparing the protein content of two complexes (that is, at different apparent molecular masses) both present in the same lane (that, is from a biochemical sample purified on a  $\ast 1$  chromatographic phase). This kind of comparison has a significant biological value, as it might demonstrate that these two bands contain complexes that are of biologically-relevant distinct protein compositions [*e.g.* the complexes might correspond to different DNA repair (sub-)pathways].

The *dsbrepair* program includes a set of flexible algorithms to perform both kinds of comparisons above, with a flexible graphical user interface that will be described in detail below.

## 5.4 GENERAL OVERVIEW OF THE GRAPHICAL USER INTERFACE

The graphical user interface is comprised of a number of windows and dialogs.

### 5.4.1 THE MAIN WINDOW

In order to set the stage, the main window of the *dsbrepair* program is described first. When *dsbrepair* is run, the main window appears like in Figure 5.4, page ci-contre.

The *SQLite3* database holding all the mass spectrometric-based protein identification data is located at the path shown in the **Database** edit widget. By default, this pathway is set at software build time and should lead to the database shipped with the distribution. It is possible at any time, once the database path has been set (and optionally modified from its default value), to execute SQL queries by clicking onto the **SQL queries** button (detailed later).

The **Sample lists** widget is a table view (like in a spreadsheet). The table view in the example is empty. It might be populated by using the buttons **Add sample list** and **Add sample**. It is possible, at any time, to remove a

5.4. GENERAL OVERVIEW OF THE GRAPHICAL USER INTERFACE 205

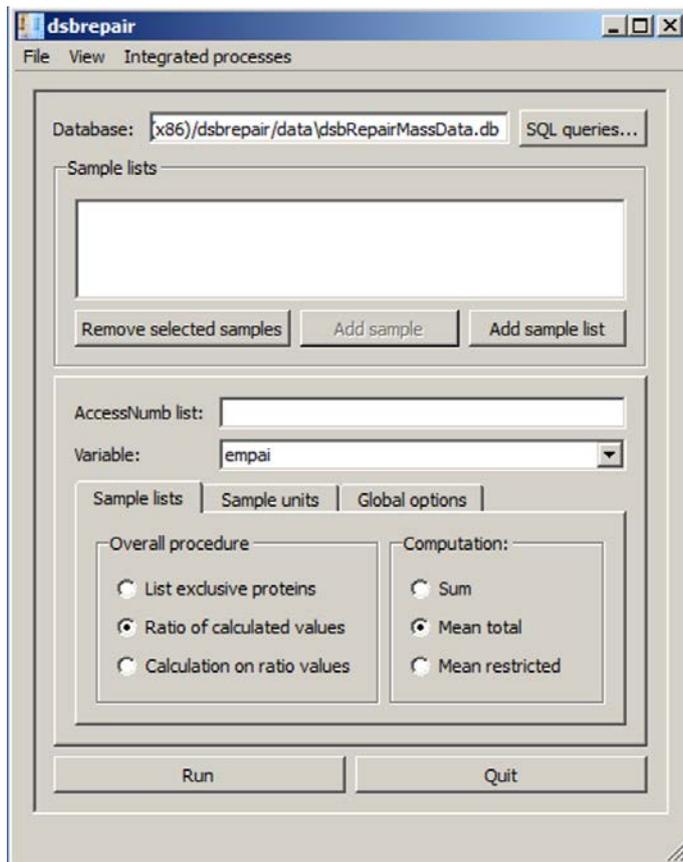


FIGURE 5.4 – **Program’s main window.** This is the aspect of the program’s main window upon executing the program.

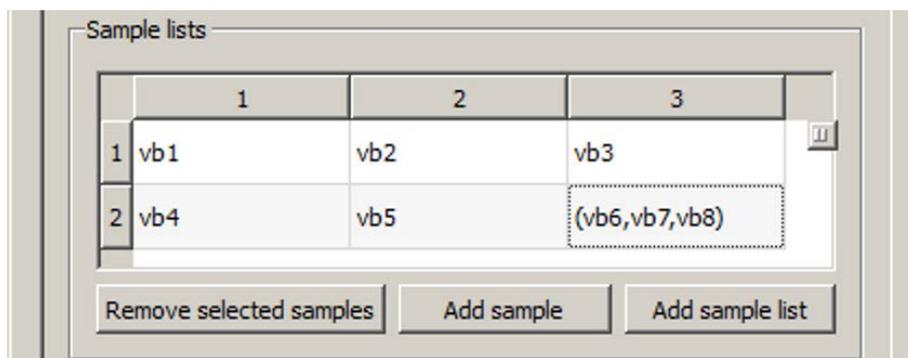


FIGURE 5.5 – Program’s main window (**Sample lists table view**). This is the aspect of the sample lists table view once two sample lists have been defined, each comprising three sample units.

sample or a set of selected samples by clicking onto the **Remove selected samples**. If the number of samples to be removed corresponds to a substantial subset of all the samples, the program alerts the user with a message box asking for confirmation (there is no undo available as of yet). To add a sample list, first click onto **Add sample list**. This will create a new row, with a single empty cell. Start entering the name of a sample (like *vb129*) and finish editing with the  $\leftarrow P$  key (the  $\leftarrow P$  vignette denotes the  $\langle \text{Enter} \rangle$  key).

Two *sample lists* (rows), each made of three *sample units* (cells) are shown in Figure 5.5. Why use the term “sample unit” and not simply “sample”? Because each cell of the table can actually hold not only a single sample (*vb1*, for example), but a set of samples [like (*vb6,vb7,vb8*)]. Note that it is only required to insert comma characters (‘,’) between the sample numbers in a sample unit, as the parentheses [‘(..., ..., ...)’] are inserted automatically by the program upon hitting the  $\leftarrow P$  key.

Once the sample lists have been defined, it might be useful to restrict the computations to a list of one or more accession numbers of interest. This list of accession numbers might be entered in the **AccessNum** list line edit widget in the form of comma-separated accession numbers (like “*XRCC5\_HUMAN,XRCC5\_HUMAN,XRCC4\_HUMAN*”). If no accession number is present in this line edit widget, then all the proteins identified in the samples are involved in the calculations.

Indeed, what do we talk about when we talk “samples”? Well, a *sample* is nothing more than a purification product that was trypsinized and analyzed by HPLC-MS/MS. In other words, as *sample* is nothing else than a HPLC-MS/MS run that identified a set of proteins. A *sample* is thus “*a set of protein accession numbers identified during an HPLC-MS/MS run*”. Indeed, the proteins identified in this sample are actually what we are looking for and what we are working on each time we talk about a “sample”. Then, in the following text, a “*sample*” is actually to be understood as a “*list of proteins identified in that sample*”.

The various calculations that can be achieved using *dsbrepair* are then specified in the various widgets located beneath this accession number list. These calculations are detailed in the next sections.

#### 5.4.2 THE RESULTS WINDOW

The results window is shown in Figure 5.6, [page suivante](#). This window is displayed by selecting the *View* → *Results* menu (or by entering  $\overline{\text{Ctrl}}+\overline{\text{R}}$ ). At each new computation, the results are appended to the other contents of the window.

#### 5.4.3 THE HISTORY WINDOW

Any time a computation could be successfully triggered and executed, the details of that computation are output to this window. This window is displayed by selecting the *View* → *History* menu (or by entering  $\overline{\text{Ctrl}}+\overline{\text{H}}$ ).

#### 5.4.4 THE CONSOLE WINDOW

Any time anything happens in the program, the details of the actions are output to this window. Upon a computation, this window will collect all the messages from the program that are not true results, but that might help understanding results. This window is displayed by selecting the *View* → *Konsole* menu (or by entering  $\overline{\text{Ctrl}}+\overline{\text{K}}$ ).

#### 5.4.5 COMMON FEATURES

All the auxiliary windows (results, history, console) have the same set of features :

- \* Traceability of the data output in various windows. A given calculation might give rise to data/information textual output to be sent to different windows. To distinguish the results of a computation from those of another computation, each computation result output is preceded

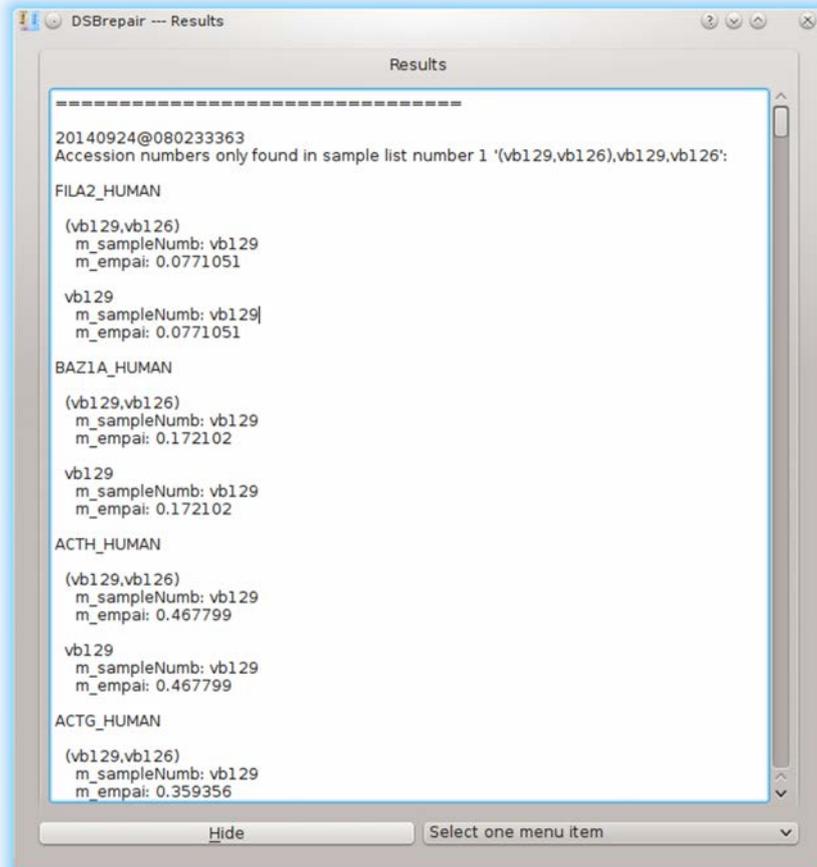


FIGURE 5.6 – **Results window.** The results window is a free-form text editor where all new results are systematically appended to the contents. The user might modify the contents of the window (that is, the window is not read-only).

#### 5.4. GENERAL OVERVIEW OF THE GRAPHICAL USER INTERFACE 209

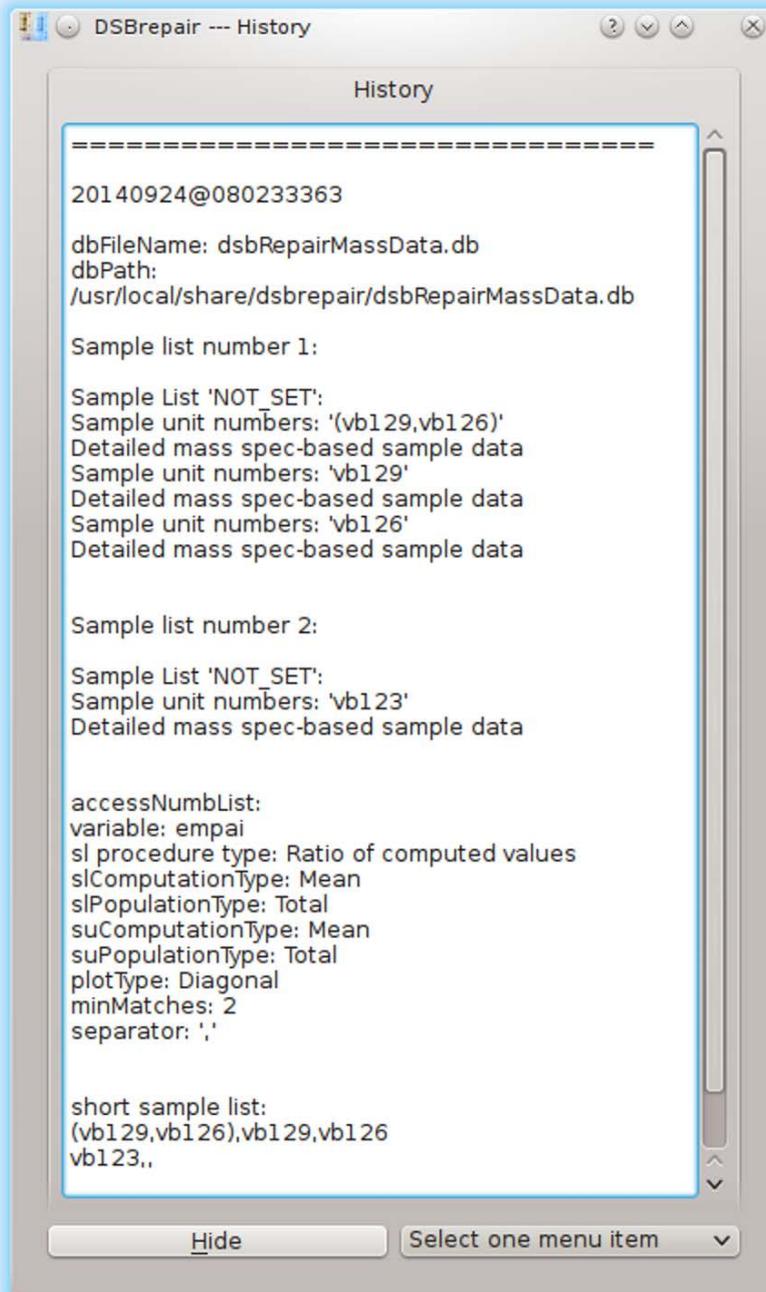


FIGURE 5.7 – **History window.** The history window is a free-form text editor where all the details about any computation started by the user are recorded. New history elements are systematically appended to the contents. The user might modify the contents of the window (that is, the window is not read-only).

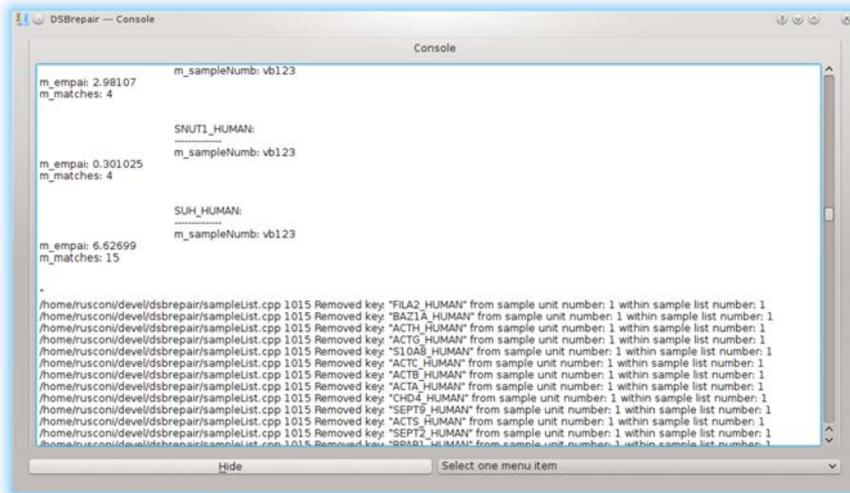


FIGURE 5.8 – **Console window.** The console window is a free-form text editor where all the details about anything happening in the program are recorded. New elements are systematically appended to the contents. The user might modify the contents of the window (that is, the window is not read-only).

by a time stamp in the form 20140924@080233363 (date@time). Thus, a given timestamp is used to identify any output from a given computation, whatever the window into which it is echoed. In this way it is easy to recollect all the data pertaining to that specific computation in all the windows ;

- \* **Hide** This button hides the window, without closing it. The contents of the windows are not lost ;
- \* **Save to file** This drop-down menu item will save the current contents of the window to a file. If no file name has been previously defined, then a file name selection dialog will allow the user to do so ;
- \* **Save as new file** This drop-down menu item will invite the user to define the name of a new file and then save the contents of the window to that new file ;
- \* **Append to file** This drop-down menu item will save the contents of the window to the current file without overwriting it, but by appending them to it. If no file name has been previously defined, the user will be invited to define the name of the new file to which to store the contents.
- \* It is possible to have *dsbrepair* store the position of the windows on

the screen estate so as to customize the working session. If a given window is in a position that suits the user's workflow, hiding it and showing it again will store the position of that window, so that the program will reinstate that position upon starting it anew.

## 5.5 VARIOUS CALCULATIONS USING SAMPLE LISTS

In *dsbrepair* there are, somehow, two kinds of sample lists, as hinted in the description of the **Sample lists** table view. Indeed, in one cell of the table view, we saw earlier that we can define a sample unit which might be "composite", that is, made of more than one sample. And then, there is the real sample list, which is made of all the cells in a row. A sample list is actually a *sample unit* list. What are the calculations that can be done on these two kinds of lists? In fact, what kind of data mining procedures are available in *dsbrepair*? This is what will be described in the sections below.

### 5.5.1 THE THREE GENERAL PROCEDURES AVAILABLE

There are three main procedures (that is, main algorithms) at work in *dsbrepair*, which are all accessed *via* the **Overall procedure** group box in the **Sample lists** tab of the main window (Figure 5.4, page 205) :

- \* **List exclusive proteins** This procedure actually does not perform any computation and is made available because of its convenience, particularly in the early steps of data mining ;
- \* **Ratio of calculated values** This procedure first performs a given calculation for each sample list<sup>6</sup>. This preliminary computation is specified in the **Computation** group box. Once the value has been calculated for each one of the two sample lists, the program then computes the ratio

$$\frac{\text{calculated value for sample list 1}}{\text{calculated value for sample list 2}}$$

- \* **Calculation on ratio values** This procedure requires that two sample lists be selected and that each have the same number of sample units<sup>7</sup>. First, a ratio calculation is performed on each pair of sample units for both sample lists, that is, ratio

$$\frac{\text{emPAI}_{\text{first sample unit in sample list 1}}}{\text{emPAI}_{\text{first sample unit in sample list 2}}}$$

---

6. Only two sample lists, that is, only two rows of the table view might be selected for analysis.

7. As for the other case above, only two rows might be selected for the analysis, and this time the sample lists must have the same number of sample units (that is, both rows must have the same number of cells).

	1	2
1	vb123	
2	(vb126,vb120)	
3	vb129	vb114

FIGURE 5.9 – **Table view contents for the List exclusive proteins procedure.** This is an example of how the List exclusive proteins procedure might be used. This procedure will list all the protein identifications that are present in a sample list (all the contents of a row are merged) and that are absent from all the other sample lists (all the contents of all the other rows are merged).

is computed, then ratio

$$\frac{\text{emPAI}_{\text{second sample unit in sample list 1}}}{\text{emPAI}_{\text{second sample unit in sample list 2}}}$$

is computed and so on for each sample unit pair in the sample lists. Finally, with the list of ratios, the computation selected in the **Computation** group box is performed.

These three algorithms are detailed in the sections below.

#### 5.5.1.1 LIST EXCLUSIVE PROTEINS

This algorithm detects all the protein identifications (accession numbers) that are only present in a given sample list and totally absent from all the other sample lists. Figure 5.9 shows the sample lists that have been defined. Upon choosing **List exclusive proteins** from the **Procedure** group box widget and then clicking onto the **Run** button, these sample lists are processed as follows :

1. All the protein identifications found in the first sample list's sample units (that is, all the cells of that row) are merged into a single protein identification list (list number 1);
2. All the protein identifications found in *all* the other sample lists' sample units (that is, all the remaining cells in the table) are merged into a single protein identification list (list number 2);
3. Each protein identification (an accession number) present in the list number 1 but absent in the list number 2 is reported;
4. Then, this same process is repeated for the next sample list *versus* all the other sample lists, and so on.

The output from this example, is provided below :

Accession numbers only found in sample list number 1 'vb123':

H2B1M\_HUMAN}

vb123

m\_sampleNumb: vb123

m\_empai: 2.98107

RBM27\_HUMAN

vb123

m\_sampleNumb: vb123

m\_empai: 0.132541

[others skipped]

Accession numbers only found in sample list number 2 '(vb126,vb120)':

MORC2\_HUMAN

(vb126,vb120)

m\_sampleNumb: vb120

m\_empai: 0.376857

SF3B1\_HUMAN

(vb126,vb120)

m\_sampleNumb: vb120

m\_empai: 0.142069

[others skipped]

Accession numbers only found in sample list number 3 'vb129,vb114':

SSRP1\_HUMAN

vb114

m\_sampleNumb: vb114

m\_empai: 0.467799

RPA49\_HUMAN

```

vb114
  m_sampleNumb: vb114
  m_empai: 0.930698

```

It might be of interest to ask some computation to be performed on the protein identifications found to be unique in a given sample list. The computation to be performed is selectable in the **Computation** group box. But then, what is the variable onto which that computation should be performed? That variable of interest might be selected within the **Variable** combo box widget (Figure 5.4, page 205). In the example, the variable is *empai*. This is how the three available computations perform :

- \* **Sum** : This operation simply computes the sum of the variable value for each accession number encountered in the samples of a sample list. For example, protein XXXX\_HUMAN is found in both the *vb129* and *vb114* samples. The sum will be calculated by adding the *empai* for that XXXX\_HUMAN accession number in the samples *vb129* and *vb114*;
- \* **Mean total** This operation simply computes the sum of the variable as above and then divides it by the number of samples in the sample list. If the example above is kept, then the sum of the *empai* for the two identifications would be divided by two, since there are two samples in the sample list ;
- \* **Mean restricted** This operation simply computes the sum of the variable as above and then divides it by the number of samples in the sample list *in which the identification was actually found*. If a protein identification is found in two samples of a sample list that comprises four samples, then the sum would be divided by two and not by four. That is, the mean is restricted to the samples that actually have the given accession number.

The output of the procedure involves listing all the proteins encountered exclusively in any given sample list, formatted for immediate insertion in a spreadsheet program, with the first line being the header describing the contents of the columns :

```

AccNumb,Mean,Count,Sample list
H2B1L_HUMAN,2.98107,1,[1]
MORC2_HUMAN,0.188429,1,[2]
RLAO_HUMAN,0.292447,2,[3]

```

Since, in this example, the computation that was asked for is **Mean** (either **total** or **restricted**), the value that is reported for each accession number is

**Mean** (second column, above). The **Count** (third column) is the number of times that a given accession number was encountered in the list. The list in which the protein was found to be unique is provided as a [number].

#### 5.5.1.2 RATIO OF CALCULATED VALUES

This algorithm works in two steps : first, values are calculated for each sample list, and then the ratio of the obtained values is computed and returned. Note that this algorithm only accepts two sample lists. Each sample list might have any number of sample units. To define a sample list, simply select as many cells as necessary in the **Sample lists** table view widget (**Ctrl** + left mouse button). For a given sample list, all the selected cells must belong to the same table view row. Do the same for the second list. The algorithm will proceed according to the following steps :

1. Search of all the proteins in the first sample list that are absent from the second sample list and do the same in the reverse order, that is, search for all the proteins that are exclusively present in the second sample list. The found proteins are stored in memory and will be listed with the output. Note that all the proteins found to be exclusively present in one sample or the other are removed from the samples. This is logical : how can a given protein be compared in two sample lists if that protein is actually only present in one of the two sample lists ? Also of interest is the fact that the presence of a protein in a sample list is to be construed “as a whole”, that is, a protein might be present in one sample unit of a given sample list and absent from another sample unit. The fact is that a protein needs to be found in at least one sample unit of a given sample list. At the end of this step, the algorithm only knows about proteins found to be common to the two sample lists ;
2. Start with the first sample list and for each accession number (that is, each protein) in that sample list, perform the calculation specified in the **Computation** group box. The options are the same as before : **Sum**, **Mean total**, **Mean restricted**. The variable on which the calculation is performed is selected, as described earlier, in the **Variable** combo box widget. This same calculation is performed for the second sample list ;
3. For each accession number, compute the ratio between the calculated value for that accession number in the sample list 1 over the calculated value in the sample list 2.

Let's take some examples to show how the flexibility of this approach can be useful. First off, let's look at one specific protein : **HNRPU\_HUMAN** (set this accession number to the **Access Numb** list line edit widget). We are interested in the *emPAI* variable. These are the values found for this protein in four

experiments (note that these values are only for the example and are not biologically significant) :

- \* Real experiment 1 ( $\ast 1$ ) : *vb120*, emPAI=0.350 ;
- \* Real experiment 2 ( $\ast 1$ ) : *vb123*, emPAI=2.675 ;
- \* Control experiment 1 ( $\diamond 1$ ) : *vb126*, emPAI=0.8233 ;
- \* Control experiment 2 ( $\diamond 1$ ) : *vb129*, emPAI=0.35 ;

Now, let's say that we define two sample lists :

- \* Sample list 1 : *vb120* | *vb123* (two distinct cells in the row) ;
- \* Sample list 2 : *vb126* | *vb129* (two distinct cells in the row).

**5.5.1.2.1 Example with a *Sum*-based computation** In this example we want to compare the HNRPU\_HUMAN raw abundance (emPAI-based) in both sample lists. To do this, we want to sum the emPAI value for each sample in list 1, then do the same for list 2 and finally compare them by computing the ratio (sum 1/sum 2). Thus, we require that the calculation to be performed on the sample lists be **Sum** (in the **Computation** group box in the **Sample lists** tab). Upon clicking **Run**, the following output is generated :

```
AccNumb,Ratio 1/2,Sum 1,Sum 2,Count 1,Count 2,
HNRPU_HUMAN,2.57738,3.02498,1.17366,2,2,
```

We can see that the emPAI values were added (the header above mentions “Sum”) for the various samples in each sample list. Each sample list is referred to by a number (1 or 2). Sum 1 is 3.02498 which matches correctly the (0.350+2.675) sum, and Sum 2 is 1.174, which matches correctly the (0.8233+0.35) sum.

Now, the ratio is computed, as indicated in the header, like this ( $\frac{3.02498}{1.174}$ ) ; the calculated ratio matches.

As a final note, the number of times that the accession number is found in a given sample list is provided in the form “Count <list number>”. In the example above, the protein was found in both samples of each sample list.

**5.5.1.2.2 Example with *Mean total*-based computation** In this example we want to compare the HNRPU\_HUMAN mean abundance (emPAI-based) in both sample lists. To do this, we want to first compute the mean of the emPAI value for each sample in list 1, then do the same for list 2 and finally compare them by computing the ratio (mean 1/mean 2). Thus, we require that the calculation to be performed on the sample lists be **Mean total** (in the **Computation** group box in the **Sample lists** tab).

Upon clicking Run, the following output is generated :

```
AccNumb,Ratio 1/2,Mean 1,Var 1,Sd 1,Mean 2,Var 2,Sd 2,Count 1,Count 2,
HNRPU_HUMAN,2.57738,1.51249,1.35065,1.16217,0.586831,0.0279701,0.167243,2,2,
```

In the header above, “Var” stands for “Variance” and “Sd” stands for “Standard deviation”. Each sample list has these data associated to it *via* the number of the sample list (“Sd<number>”, for example). Because HNRPU\_HUMAN is found in both samples of each sample list, Count 1 and Count 2 both have the 2 value. By dividing the sums obtained in the previous example by 2, we get Mean 1 and Mean 2. Finally, the ration between the means is computed.

**5.5.1.2.3 Example with Mean restricted-based computation** This example is almost identical to the previous one unless the mean computation is performed by dividing the sum of the variable values by the number of samples *in which the accession number was actually found*. This means that, if HNRPU\_HUMAN had only been found in one of the samples of list 1, then the Mean 1 would have had the value of the emPAI of that accession number in the single sample in which it was found.

### 5.5.1.3 CALCULATION ON RATIO VALUES

The aim of this algorithm is to allow performing statistics on the distribution of proteins in various complexes (typically in the same blue native gel lane). Since this kind of biochemical experiments might be done a number of times, statistics might be performed on the ratios calculated on protein abundance data “sample *vs* sample”. For example, imagine that we want to perform statistics on the abundances of proteins found in two sets of different bands :

- \* Sample list 1 : vb112 | vb120 | vb130 (these are 720 kDa bands on the \*1 phase) ;
- \* Sample list 2 : vb114 | vb123 | vb132 (these are 240 kDa bands on the \*1 phase) ;

There must be the same number of samples in each sample list, because the data processing involves computing the following *paired* ratios for each accession number :

$$r_{vb112/vb114} = \frac{\text{emPAI}_{vb112}}{\text{emPAI}_{vb114}} ; r_{vb120/vb123} = \frac{\text{emPAI}_{vb120}}{\text{emPAI}_{vb123}} ; r_{130/132} = \frac{\text{emPAI}_{vb130}}{\text{emPAI}_{vb132}}$$

Once all the ratios are known, the calculation specified in the **Computation** group box of the **Sample lists** tab is performed on these ratios. So, indeed, this kind of procedure is what it is called : “*Perform a specified calculation on a set of pre-computed ratios*”.

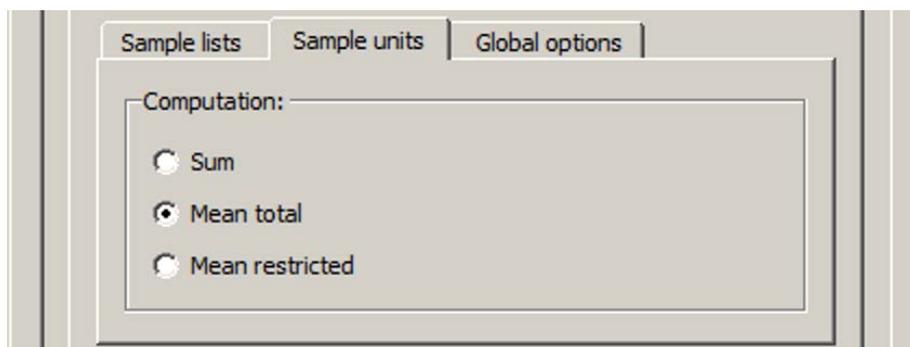


FIGURE 5.10 – **Program’s main window.** This is the aspect of the program’s main window upon executing the program.

### 5.5.2 SPECIFIC HANDLING OF THE *sample units*

In the examples for the two last procedures described above, each cell contained a single sample number (*vb112*, for example). But this does not need be the case : as explained earlier, each given cell is not a sample, but a sample unit. When a sample unit contains more than one sample, it is said to be “composite” [like in (*vb112,vb114*)]. What happens, from a statistical point of view, when a composite sample unit is found in a sample list ?

Each time a composite sample unit is found, it is first reduced to a single entity (that is, a single list of accession numbers). Then, for each accession number, a value is calculated according to what is specified by the user in the **Computation** group box of the **Sample units** tab (Figure 5.10).

### 5.5.3 GLOBAL OPTIONS

The **Global options** tab holds options that apply to any of the procedures described above. In particular, since this program interfaces seamlessly to the *SQLite3* database, there is an option that can modify the output of

## 5.6 USER-CRAFTED CUSTOM SQL QUERIES

It is often the case that the user wants to interrogate the database. Instead of opening a terminal window to perform SQL queries to the database, *dsbrepair* offers a simple graphic user interface to make SQL queries. The window depicted in Figure 5.12, page 220 can be displayed by clicking onto the **SQL queries** button in the main window of the program. While some basic structured query language is necessary to use this feature, the example that is shipped with the program (within the **History** drop-down menu widget)

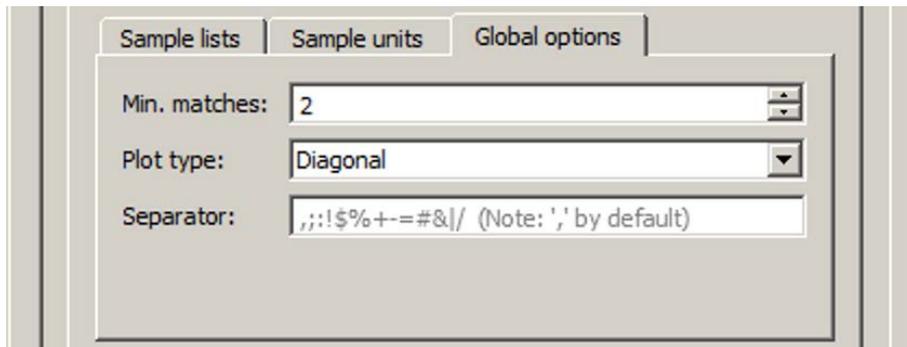


FIGURE 5.11 – **Program’s main window.** This is the aspect of the program’s main window upon executing the program.

should help anyone starting to use it. The query text must be entered in the Query text widget.

Upon successfully performing a query, the user might want to store its text in the history. To do this, simply let the query text displayed in the multiline editor widget (Query text) and click onto the Add to history button. This query will be stored so that upon closing and reexecuting the program, it will be made available in the History drop-down menu widget. Each time an item in this list is selected, the corresponding SQL query text is displayed in the multiline edit widget.

When a SQL query is successfully performed, the results are displayed in the SQL query results multiline text edit widget. The history might be erased using the Clear history button.

### 5.6.1 CRASH COURSE IN SQL *SELECT* QUERIES

The SQL queries feature in *dsbrepair* is only focused on the *SELECT* subset of SQL commands. This *SELECT* command subset is about using a database for querying it in powerful ways. All the SQL *SELECT* commands shown in the examples below are formatted for maximum syntax readability. Any line return might be replaced by a space.

The *dsbrepair* database is a relational database, that is, it is composed of more than one table. Indeed, the structure of the database, shown in Figure 5.3, page 203 is multi-table. The tables serve the purpose to store the data in a conceptually-affine manner. For example, in our case, all the file names of all the recorded experiments are stored in a `files` table. Likewise, all the details about an experiment are recorded in the `experiments` table. When one wants to make a relation between a given record in a table and a set of data recorded in another table, a *relation key* is required. That key name has always the `<name>_id` naming pattern. For example, the `files` table

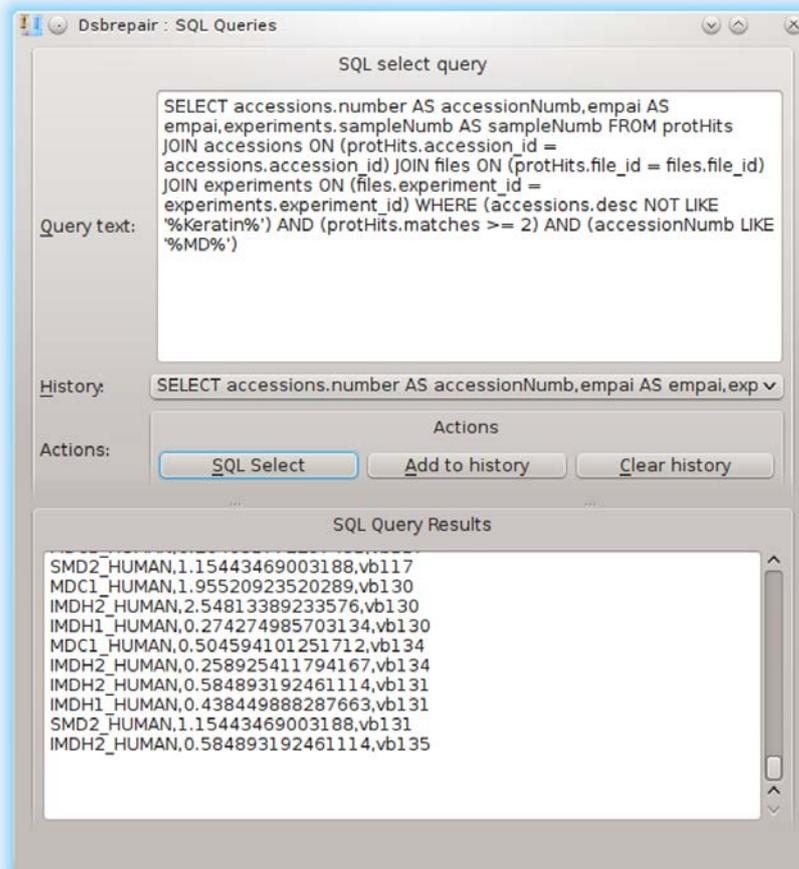


FIGURE 5.12 – **SQL queries window.** This is the aspect of the program's SQL queries window.

has the `experiment_id` key that makes the relation to the `experiments` table. In turn, the `experiments` table has the `biotinMoiety_id` key that makes the relation to the `biotinMoieties` table. To actually make a relation during a SQL query, the involved tables need to be JOINed together. This happens with a JOIN statement :

```
protHits JOIN accessions ON (protHits.accession_id = accessions.accession_id)
```

which states that when “recollecting” all the data in the `SELECT` query, the contents of the two tables must be mixed together only if the specified keys match. This JOIN statement needs to be made each time the contents of more than one table is requested.

Let’s try selecting all the data from the `protHits` table (`*` means all the fields in the table) :

```
SELECT
*
FROM protHits
```

This command will output a very long list of records :

```
201409250105235519
protHit_id,accession_id,file_id,matches,pai,empai,empaiPct
1,1,1,262,1.34408602150538,21.0844211988575,0
2,2,1,110,1.43661971830986,26.3287468865075,0
3,3,1,96,3.25925925925926,1815.59978837533,0
4,4,1,83,1.02597402597403,9.616320618413,0
5,5,1,80,3.03846153846154,1091.60086111738,0
6,6,1,77,1.69047619047619,48.0316141230326,0
7,7,1,64,0.852941176470588,6.12756483443891,0
```

The header provides the name of the variables of which the recorded values are shown in columns. This list contains all the proteins that were identified in all the experiments recorded in the database. However, this output is not very useful. Indeed, we do not know what 1 means for `accession_id`. Let’s try to actually join the `accessions` database to get to the name of the protein that was identified in the output above :

```
SELECT
*
FROM protHits
JOIN accessions ON (protHits.accession_id = accessions.accession_id)
```

```

20140925@105750097
protHit_id,accession_id,file_id,matches,pai,empai,empaiPct,accession_id,number,
1,1,1,262,1.34408602150538,21.0844211988575,0,1,PRKDC_HUMAN,DNA-dependent prote
2,2,1,110,1.43661971830986,26.3287468865075,0,2,TOP2A_HUMAN,DNA topoisomerase 2

```

We now see that with the `accessions` table having been JOINed, we can access the data specifically stored in that table. These data show up in the listing above under column header elements `accession_id`, `number`, `desc`.

It is possible to require that the data output by a `SELECT` command be filtered and shown in a specific manner by naming the field of the table (syntax : `tablename.fieldname`) to be shown and by specifying under what column header that data bit must be displayed (`AS`) :

```

SELECT
accessions.number AS accessNumb
FROM protHits
JOIN accessions ON (protHits.accession_id = accessions.accession_id)

```

This query outputs the following :

```

20140925@110315074
accessNumb
PRKDC_HUMAN
TOP2A_HUMAN
XRCC5_HUMAN
TOP2B_HUMAN

```

As we can see, now, only the accession number values of the all the records (corresponding to protein identifications) in the database are shown. There are as many records in this output as in the above one, only we have decided to only show the accession number and to provide a new name for the column header in the output (`AS` statement).

Let's try, now, to limit the output to a logical condition. We want to only list all the protein identifications in the database that have "PRKDC\_HUMAN" as the accession number :

```

SELECT

```

```
accessions.number AS accessNumb
FROM protHits
JOIN accessions ON (protHits.accession_id = accessions.accession_id)
WHERE (accessNumb LIKE '%PRKDC_HUMAN%')
```

This query outputs the following :

```
201409250111012803
accessNumb
PRKDC_HUMAN
```

Not very useful, right? We only know that the searched protein was found a number of times, but we do not know anything more than that. Let's try to recollect some more interesting data for these records. How about trying to know on what chromatographic phase this protein was found. That data bit is stored in the `biotinMoieties` table that is, in turn, related to the `experiments` table. The point is that we need to enter in the database *via* the `protHits` table, since we want to list all the `protHits` that pertain to an identification of "PRKDC\_HUMAN". In the database structure shown in Figure 5.3, we can see that the `files` table is the relation between the `protHits` table and the `experiments` table. So we need to JOIN these tables. But we also need to specify what field of the table is of interest and how the records need to be arranged in the header :

```
SELECT
accessions.number AS accessNumb,
biotinMoieties.desc AS biotinDesc
FROM protHits
JOIN accessions ON (protHits.accession_id = accessions.accession_id)
```

```

JOIN files ON (protHits.file_id = files.file_id)
JOIN experiments ON (files.experiment_id = experiments.experiment_id)
JOIN biotinMoieties ON (experiments.biotinMoiety_id = biotinMoieties.biotinMoiet
WHERE (accessNumb LIKE '%PRKDC_HUMAN%')

```

```

20140925@111916030
accessNumb,biotinDesc
PRKDC_HUMAN,1
PRKDC_HUMAN,2
PRKDC_HUMAN,1
PRKDC_HUMAN,1
PRKDC_HUMAN,2
PRKDC_HUMAN,1
PRKDC_HUMAN,2
PRKDC_HUMAN,1
PRKDC_HUMAN,2
PRKDC_HUMAN,1
PRKDC_HUMAN,2
PRKDC_HUMAN,1
PRKDC_HUMAN,2
PRKDC_HUMAN,1
PRKDC_HUMAN,2

```

And now, we can see that “PRKDC\_HUMAN” was actually purified on both chromatographic phases. But then, we may want to know which band contained it. Let’s try adding the sampleDesc data element to the output.

```

SELECT
accessions.number
AS accessNumb, biotinMoieties.desc
AS biotinDesc, experiments.sampleDesc
AS sampleDesc
FROM protHits
JOIN accessions ON (protHits.accession_id = accessions.accession_id)
JOIN files ON (protHits.file_id = files.file_id)
JOIN experiments ON (files.experiment_id = experiments.experiment_id)
JOIN biotinMoieties ON (experiments.biotinMoiety_id = biotinMoieties.biotinMoiet
WHERE (accessNumb LIKE '%PRKDC_HUMAN%')

```

```

20140925@112159613
accessNumb,biotinDesc,sampleDesc
PRKDC_HUMAN,1,[vb120] - complexe interet 720kDa

```

```
PRKDC_HUMAN,2,[vb129] - controle complexe 720kDa
PRKDC_HUMAN,1,[vb81] - complexe 720kDa
PRKDC_HUMAN,1,[vb85] - complexe 720kDa
PRKDC_HUMAN,2,[vb96] - controle complexe 720kDa
PRKDC_HUMAN,1,[vb104] - complexe 720kDa
PRKDC_HUMAN,2,[vb111] - controle complexe 720kDa
PRKDC_HUMAN,1,[vb112] - complexe 720kDa
PRKDC_HUMAN,2,[vb119] - controle complexe 720kDa
PRKDC_HUMAN,1,[vb113] - complexe 480kDa
PRKDC_HUMAN,2,[vb118] - controle complexe 480kDa
PRKDC_HUMAN,1,[vb130] - complexe 720kDa
PRKDC_HUMAN,2,[vb134] - controle complexe 720kDa
```

Now we can see that the protein was actually purified only in complexes that are of pretty large molecular masses.

## 5.7 APPENDIX

### GNU GENERAL PUBLIC LICENSE TEXT

#### GNU GENERAL PUBLIC LICENSE

Version 3, 29 June 2007

Copyright © 2007 Free Software Foundation, Inc. <http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

#### **Preamble**

The GNU General Public License is a free, copyleft license for software and other kinds of works.

The licenses for most software and other practical works are designed to take away your freedom to share and change the works. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change all versions of a program—to make sure it remains free software for all its users. We, the Free Software Foundation, use the GNU General Public License for most of our software; it applies also to any other work released this way by its authors. You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for them if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs, and that you know you can do these things.

To protect your rights, we need to prevent others from denying you these rights or asking you to surrender the rights. Therefore, you have certain responsibilities if you distribute copies of the software, or if you modify it : responsibilities to respect the freedom of others.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must pass on to the recipients the same freedoms that you received. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

Developers that use the GNU GPL protect your rights with two steps : (1) assert copyright on the software, and (2) offer you this License giving you legal permission to copy, distribute and/or modify it.

For the developers' and authors' protection, the GPL clearly explains that there is no warranty for this free software. For both users' and authors' sake, the GPL requires that modified versions be marked as changed, so that their problems will not be attributed erroneously to authors of previous versions.

Some devices are designed to deny users access to install or run modified versions of the software inside them, although the manufacturer can do so. This is fundamentally incompatible with the aim of protecting users' freedom to change the software. The systematic pattern of such abuse occurs in the area of products for individuals to use, which is precisely where it is most unacceptable. Therefore, we have designed this version of the GPL to prohibit the practice for those products. If such problems arise substantially in other domains, we stand ready to extend this provision to those domains in future versions of the GPL, as needed to protect the freedom of users.

Finally, every program is threatened constantly by software patents. States should not allow patents to restrict development and use of software on general-purpose computers, but in those that do, we wish to avoid the special danger that patents applied to a free program could make it effectively proprietary. To prevent this, the GPL assures that patents cannot be used to render the program non-free.

The precise terms and conditions for copying, distribution and modification follow.

## TERMS AND CONDITIONS

### 0. Definitions.

“This License” refers to version 3 of the GNU General Public License.

“Copyright” also means copyright-like laws that apply to other kinds of works, such as semiconductor masks.

“The Program” refers to any copyrightable work licensed under this License. Each licensee is addressed as “you”. “Licensees” and “recipients” may be individuals or organizations.

To “modify” a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a “modified version” of the earlier work or a work “based on” the earlier work.

A “covered work” means either the unmodified Program or a work based on the Program.

To “propagate” a work means to do anything with it that, without permission, would make you directly or secondarily liable for infringement under applicable copyright law, except executing it on a computer or modifying a private copy. Propagation includes copying, distribution (with or without modification), making available to the public, and in some countries other activities as well.

To “convey” a work means any kind of propagation that enables other parties to make or receive copies. Mere interaction with a user through a computer network, with no transfer of a copy, is not conveying.

An interactive user interface displays “Appropriate Legal Notices” to the extent that it includes a convenient and prominently visible feature that (1) displays an appropriate copyright notice, and (2) tells the user that there is no warranty for the work (except to the extent that warranties are provided), that licensees may convey the work under this License, and how to view a copy of this License. If the interface presents a list of user commands or options, such as a menu, a prominent item in the list meets this criterion.

1. Source Code.

The “source code” for a work means the preferred form of the work for making modifications to it. “Object code” means any non-source form of a work.

A “Standard Interface” means an interface that either is an official standard defined by a recognized standards body, or, in the case of interfaces specified for a particular programming language, one that is widely used among developers working in that language.

The “System Libraries” of an executable work include anything, other than the work as a whole, that (a) is included in the normal form of packaging a Major Component, but which is not part of that Major Component, and (b) serves only to enable use of the work with that Major Component, or to implement a Standard Interface for which an implementation is available to the public in source code form. A “Major Component”, in this context, means a major essential component (kernel, window system, and so on) of the specific operating system (if any) on which the executable work runs, or a compiler used to produce the work, or an object code interpreter used to run it.

The “Corresponding Source” for a work in object code form means all the source code needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to control those activities. However, it does not include the work’s System Libraries, or general-purpose tools or generally available free programs which are used unmodified in performing those activities but which are not part of the work. For example, Corresponding Source includes interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work.

The Corresponding Source need not include anything that users can regenerate automatically from other parts of the Corresponding Source. The Corresponding Source for a work in source code form is that same work.

## 2. Basic Permissions.

All rights granted under this License are granted for the term of copyright on the Program, and are irrevocable provided the stated conditions are met. This License explicitly affirms your unlimited permission to run the unmodified Program. The output from running a covered work is covered by this License only if the output, given its content, constitutes a covered work. This License acknowledges your rights of fair use or other equivalent, as provided by copyright law.

You may make, run and propagate covered works that you do not convey, without conditions so long as your license otherwise remains in force. You may convey covered works to others for the sole purpose of having them make modifications exclusively for you, or provide you with facilities for running those works, provided that you comply with the terms of this License in conveying all material for which you do not control copyright. Those thus making or running the covered works for you must do so exclusively on your behalf, under your direction and control, on terms that prohibit them from making any copies of your copyrighted material outside their relationship with you.

Conveying under any other circumstances is permitted solely under the conditions stated below. Sublicensing is not allowed; section 10 makes it unnecessary.

## 3. Protecting Users' Legal Rights From Anti-Circumvention Law.

No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures. When you convey a covered work, you waive any legal power to forbid circumvention of technological measures to the extent such circumvention is effected by exercising rights under this License with respect to the covered work, and you disclaim any intention to limit operation or modification of the work as a means of enforcing, against the work's users, your or third parties' legal rights to forbid circumvention of technological measures.

## 4. Conveying Verbatim Copies.

You may convey verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice; keep intact all notices stating that this License and any non-permissive terms added in accord with section 7 apply to the code; keep intact all notices of the absence of any warranty; and give all recipients a copy of this License along with the Program.

You may charge any price or no price for each copy that you convey, and you may offer support or warranty protection for a fee.

5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions :

- (a) The work must carry prominent notices stating that you modified it, and giving a relevant date.
- (b) The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to “keep intact all notices”.
- (c) You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it.
- (d) If the work has interactive user interfaces, each must display Appropriate Legal Notices ; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

A compilation of a covered work with other separate and independent works, which are not by their nature extensions of the covered work, and which are not combined with it such as to form a larger program, in or on a volume of a storage or distribution medium, is called an “aggregate” if the compilation and its resulting copyright are not used to limit the access or legal rights of the compilation’s users beyond what the individual works permit. Inclusion of a covered work in an aggregate does not cause this License to apply to the other parts of the aggregate.

6. Conveying Non-Source Forms.

You may convey a covered work in object code form under the terms of sections 4 and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License, in one of these ways :

- (a) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by the Corresponding Source fixed on a durable physical medium customarily used for software interchange.

- (b) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by a written offer, valid for at least three years and valid for as long as you offer spare parts or customer support for that product model, to give anyone who possesses the object code either (1) a copy of the Corresponding Source for all the software in the product that is covered by this License, on a durable physical medium customarily used for software interchange, for a price no more than your reasonable cost of physically performing this conveying of source, or (2) access to copy the Corresponding Source from a network server at no charge.
- (c) Convey individual copies of the object code with a copy of the written offer to provide the Corresponding Source. This alternative is allowed only occasionally and noncommercially, and only if you received the object code with such an offer, in accord with subsection 6b.
- (d) Convey the object code by offering access from a designated place (gratis or for a charge), and offer equivalent access to the Corresponding Source in the same way through the same place at no further charge. You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source may be on a different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements.
- (e) Convey the object code using peer-to-peer transmission, provided you inform other peers where the object code and Corresponding Source of the work are being offered to the general public at no charge under subsection 6d.

A separable portion of the object code, whose source code is excluded from the Corresponding Source as a System Library, need not be included in conveying the object code work.

A “User Product” is either (1) a “consumer product”, which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, “normally used” refers to a typical or common use of that class of product, regardless of

the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a consumer product regardless of whether the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.

“Installation Information” for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section must be accompanied by the Installation Information. But this requirement does not apply if neither you nor any third party retains the ability to install modified object code on the User Product (for example, the work has been installed in ROM).

The requirement to provide Installation Information does not include a requirement to continue to provide support service, warranty, or updates for a work that has been modified or installed by the recipient, or for the User Product in which it has been modified or installed. Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network.

Corresponding Source conveyed, and Installation Information provided, in accord with this section must be in a format that is publicly documented (and with an implementation available to the public in source code form), and must require no special password or key for unpacking, reading or copying.

#### 7. Additional Terms.

“Additional permissions” are terms that supplement the terms of this License by making exceptions from one or more of its conditions. Additional permissions that are applicable to the entire Program shall be treated as though they were included in this License, to the extent that they are valid under applicable law. If additional permissions apply only to part of the Program, that part may be used separately under those permissions, but the entire Program remains governed by

this License without regard to the additional permissions.

When you convey a copy of a covered work, you may at your option remove any additional permissions from that copy, or from any part of it. (Additional permissions may be written to require their own removal in certain cases when you modify the work.) You may place additional permissions on material, added by you to a covered work, for which you have or can give appropriate copyright permission.

Notwithstanding any other provision of this License, for material you add to a covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms :

- (a) Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License ; or
- (b) Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it ; or
- (c) Prohibiting misrepresentation of the origin of that material, or requiring that modified versions of such material be marked in reasonable ways as different from the original version ; or
- (d) Limiting the use for publicity purposes of names of licensors or authors of the material ; or
- (e) Declining to grant rights under trademark law for use of some trade names, trademarks, or service marks ; or
- (f) Requiring indemnification of licensors and authors of that material by anyone who conveys the material (or modified versions of it) with contractual assumptions of liability to the recipient, for any liability that these contractual assumptions directly impose on those licensors and authors.

All other non-permissive additional terms are considered “further restrictions” within the meaning of section 10. If the Program as you received it, or any part of it, contains a notice stating that it is governed by this License along with a term that is a further restriction, you may remove that term. If a license document contains a further restriction but permits relicensing or conveying under this License, you may add to a covered work material governed by the terms of that license document, provided that the further restriction does not survive such relicensing or conveying.

If you add terms to a covered work in accord with this section, you must place, in the relevant source files, a statement of the additional terms that apply to those files, or a notice indicating where to find the applicable terms.

Additional terms, permissive or non-permissive, may be stated in the form of a separately written license, or stated as exceptions; the above requirements apply either way.

8. Termination.

You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License (including any patent licenses granted under the third paragraph of section 11).

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, you do not qualify to receive new licenses for the same material under section 10.

9. Acceptance Not Required for Having Copies.

You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work. These actions infringe copyright if you do not accept this License. Therefore, by modifying or propagating a covered work, you indicate your acceptance of this License to do so.

10. Automatic Licensing of Downstream Recipients.

Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

An “entity transaction” is a transaction transferring control of an organization, or substantially all assets of one, or subdividing an

organization, or merging organizations. If propagation of a covered work results from an entity transaction, each party to that transaction who receives a copy of the work also receives whatever licenses to the work the party's predecessor in interest had or could give under the previous paragraph, plus a right to possession of the Corresponding Source of the work from the predecessor in interest, if the predecessor has it or can get it with reasonable efforts.

You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License. For example, you may not impose a license fee, royalty, or other charge for exercise of rights granted under this License, and you may not initiate litigation (including a cross-claim or counterclaim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for sale, or importing the Program or any portion of it.

#### 11. Patents.

A "contributor" is a copyright holder who authorizes use under this License of the Program or a work on which the Program is based. The work thus licensed is called the contributor's "contributor version".

A contributor's "essential patent claims" are all patent claims owned or controlled by the contributor, whether already acquired or hereafter acquired, that would be infringed by some manner, permitted by this License, of making, using, or selling its contributor version, but do not include claims that would be infringed only as a consequence of further modification of the contributor version. For purposes of this definition, "control" includes the right to grant patent sublicenses in a manner consistent with the requirements of this License.

Each contributor grants you a non-exclusive, worldwide, royalty-free patent license under the contributor's essential patent claims, to make, use, sell, offer for sale, import and otherwise run, modify and propagate the contents of its contributor version.

In the following three paragraphs, a "patent license" is any express agreement or commitment, however denominated, not to enforce a patent (such as an express permission to practice a patent or covenant not to sue for patent infringement). To "grant" such a patent license to a party means to make such an agreement or commitment not to enforce a patent against the party.

If you convey a covered work, knowingly relying on a patent license, and the Corresponding Source of the work is not available for anyone to copy, free of charge and under the terms of this License, through a publicly available network server or other readily accessible means, then you must either (1) cause the Corresponding Source to be so available, or (2) arrange to deprive yourself of the benefit of the patent

license for this particular work, or (3) arrange, in a manner consistent with the requirements of this License, to extend the patent license to downstream recipients. “Knowingly relying” means you have actual knowledge that, but for the patent license, your conveying the covered work in a country, or your recipient’s use of the covered work in a country, would infringe one or more identifiable patents in that country that you have reason to believe are valid.

If, pursuant to or in connection with a single transaction or arrangement, you convey, or propagate by procuring conveyance of, a covered work, and grant a patent license to some of the parties receiving the covered work authorizing them to use, propagate, modify or convey a specific copy of the covered work, then the patent license you grant is automatically extended to all recipients of the covered work and works based on it.

A patent license is “discriminatory” if it does not include within the scope of its coverage, prohibits the exercise of, or is conditioned on the non-exercise of one or more of the rights that are specifically granted under this License. You may not convey a covered work if you are a party to an arrangement with a third party that is in the business of distributing software, under which you make payment to the third party based on the extent of your activity of conveying the work, and under which the third party grants, to any of the parties who would receive the covered work from you, a discriminatory patent license (a) in connection with copies of the covered work conveyed by you (or copies made from those copies), or (b) primarily for and in connection with specific products or compilations that contain the covered work, unless you entered into that arrangement, or that patent license was granted, prior to 28 March 2007.

Nothing in this License shall be construed as excluding or limiting any implied license or other defenses to infringement that may otherwise be available to you under applicable patent law.

#### 12. No Surrender of Others’ Freedom.

If conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot convey a covered work so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not convey it at all. For example, if you agree to terms that obligate you to collect a royalty for further conveying from those to whom you convey the Program, the only way you could satisfy both those terms and this License would be to refrain entirely from conveying the Program.

13. Use with the GNU Affero General Public License.

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU Affero General Public License into a single combined work, and to convey the resulting work. The terms of this License will continue to apply to the part which is the covered work, but the special requirements of the GNU Affero General Public License, section 13, concerning interaction through a network will apply to the combination as such.

14. Revised Versions of this License.

The Free Software Foundation may publish revised and/or new versions of the GNU General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies that a certain numbered version of the GNU General Public License “or any later version” applies to it, you have the option of following the terms and conditions either of that numbered version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of the GNU General Public License, you may choose any version ever published by the Free Software Foundation.

If the Program specifies that a proxy can decide which future versions of the GNU General Public License can be used, that proxy’s public statement of acceptance of a version permanently authorizes you to choose that version for the Program.

Later license versions may give you additional or different permissions. However, no additional obligations are imposed on any author or copyright holder as a result of your choosing to follow a later version.

15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

## 16. Limitation of Liability.

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

## 17. Interpretation of Sections 15 and 16.

If the disclaimer of warranty and limitation of liability provided above cannot be given local legal effect according to their terms, reviewing courts shall apply local law that most closely approximates an absolute waiver of all civil liability in connection with the Program, unless a warranty or assumption of liability accompanies a copy of the Program in return for a fee.

## END OF TERMS AND CONDITIONS

### How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively state the exclusion of warranty ; and each file should have at least the “copyright” line and a pointer to where the full notice is found.

```
<one line to give the program's name and a brief idea of
what it does.>
```

```
Copyright (C) <year> <name of author>
```

```
This program is free software : you can redistribute it and/or
modify it under the terms of the GNU General Public License
as published by the Free Software Foundation, either version
3 of the License, or (at your option) any later version.
```

```
This program is distributed in the hope that it will be
useful, but WITHOUT ANY WARRANTY ; without even
```

the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

Also add information on how to contact you by electronic and paper mail.

If the program does terminal interaction, make it output a short notice like this when it starts in an interactive mode :

```
<program> Copyright (C) <year> <name of author>  
This program comes with ABSOLUTELY NO WARRANTY ;  
for details type show w. This is free software, and you are  
welcome to redistribute it under certain conditions ; type  
show c for details.
```

The hypothetical commands **show w** and **show c** should show the appropriate parts of the General Public License. Of course, your program's commands might be different ; for a GUI interface, you would use an "about box".

You should also get your employer (if you work as a programmer) or school, if any, to sign a "copyright disclaimer" for the program, if necessary. For more information on this, and how to apply and follow the GNU GPL, see <http://www.gnu.org/licenses/>.

The GNU General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Lesser General Public License instead of this License. But first, please read <http://www.gnu.org/philosophy/why-not-lgpl.html>.



## BIBLIOGRAPHIE

- A. Aboussekhra, M. Biggerstaff, M. K. Shivji, J. A. Vilpo, V. Moncollin, V. N. Podust, M. Protic, U. Hubscher, J. M. Egly, et R. D. Wood. Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell*, 80(6) :859 – 868, 1995.
- I. Ahel, D. Ahel, T. Matsusaka, A. J. Clark, J. Pines, S. J. Boulton, et S. C. West. Poly(ADP-ribose)-binding zinc finger motifs in DNA repair/checkpoint proteins. *Nature*, 451(7174) :81 – 85, 2008.
- P. Ahnesorg, P. Smith, et S. P. Jackson. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell*, 124(2) :301 – 313, 2006.
- M. Audebert, B. Salles, et P. Calsou. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *J. Biol. Chem.*, 279(53) :55117 – 55126, 2004.
- S. Azam, N. Jouvét, A. Jilani, R. Vongsamphanh, X. Yang, S. Yang, et D. Ramotar. Human glyceraldehyde-3-phosphate dehydrogenase plays a direct role in reactivating oxidized forms of the DNA repair enzyme APE1. *J. Biol. Chem.*, 283(45) :30632 – 30641, 2008.
- C. Bailly, J. Kluza, C. Martin, T. Ellis, et M. J. Waring. DNase I footprinting of small molecule binding sites on DNA. *Methods Mol Biol*, 288 :319 – 342, 2005.
- C. J. Bakkenist et M. B. Kastan. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature*, 421(6922) :499 – 506, 2003.
- C. E. Bansbach, R. Betous, C. A. Lovejoy, G. G. Glick, et D. Cortez. The annealing helicase SMARCAL1 maintains genome integrity at stalled replication forks. *Genes Dev*, 23(20) :2405 – 2414, 2009.
- A. Barbarulo, V. Iansante, A. Chaidos, K. Naresh, A. Rahemtulla, G. Franzoso, A. Karadimitris, D. O. Haskard, S. Papa, et C. Bubici. Poly(ADP-ribose) polymerase family member 14 (PARP14) is a novel effector of the JNK2-dependent pro-survival signal in multiple myeloma. *Oncogene*, 32(36) :4231 – 4242, 2013.
- I. Barlat, F. Maurier, M. Duchesne, E. Guitard, B. Tocque, et F. Schweighofer. A role for Sam68 in cell cycle progression antagonized by a spliced variant within the KH domain. *J. Biol. Chem.*, 272(6) :3129 – 3132, 1997.

- R. Barrangou. CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip Rev RNA*, 4(3) :267 – 278, 2013.
- R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, et P. Horvath. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819) :1709 – 1712, 2007.
- G. J. Bates, S. M. Nicol, B. J. Wilson, A.-M. F. Jacobs, J.-C. Bourdon, J. Wardrop, D. J. Gregory, D. P. Lane, N. D. Perkins, et F. V. Fuller-Pace. The DEAD box protein p68 : A novel transcriptional coactivator of the p53 tumour suppressor. *EMBO J*, 24(3) :543 – 553, 2005.
- C. Beck, I. Robert, B. Reina-San-Martin, V. Schreiber, et F. Dantzer. Poly(ADP-ribose) polymerases in double-strand break repair : Focus on PARP1, PARP2 and PARP3. *Exp Cell Res*, 2014.
- P. Beli, N. Lukashchuk, S. A. Wagner, B. T. Weinert, J. V. Olsen, L. Baskcomb, M. Mann, S. P. Jackson, et C. Choudhary. Proteomic investigations reveal a role for RNA processing factor THRAP3 in the DNA damage response. *Mol Cell*, 46(2) :212 – 225, 2012.
- Z. Bencokova, M. R. Kaufmann, I. M. Pires, P. S. Lecane, A. J. Giaccia, et E. M. Hammond. ATM activation and signaling under hypoxic conditions. *Mol Cell Biol*, 29(2) :526 – 537, 2009.
- C. Bertoli, S. Klier, C. McGowan, C. Wittenberg, et R. A. M. de Bruin. Chk1 inhibits E2F6 repressor function in response to replication stress to maintain cell-cycle transcription. *Curr Biol*, 23(17) :1629 – 1637, 2013.
- M. Betermier, P. Bertrand, et B. S. Lopez. Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet*, 10(1) :e1004086, 2014.
- S. Bhatti, S. Kozlov, A. A. Farooqi, A. Naqi, M. Lavin, et K. K. Khanna. ATM protein kinase : The linchpin of cellular defenses to stress. *Cell Mol Life Sci*, 68(18) :2977 – 3006, 2011.
- C. L. Bladen, D. Udayakumar, Y. Takeda, et W. S. Dynan. Identification of the polypyrimidine tract binding protein-associated splicing factor.P54(nrb) complex as a candidate DNA double-strand break rejoining factor. *J. Biol. Chem.*, 280(7) :5205 – 5210, 2005.
- W. W. Blakeslee, C. L. Wysoczynski, K. S. Fritz, J. K. Nyborg, M. E. A. Churchill, et T. A. McKinsey. Class I HDAC inhibition stimulates cardiac protein SUMOylation through a post-translational mechanism. *Cell Signal*, 26(12) :2912 – 2920, 2014.

- C. Blanpain, M. Mohrin, P. A. Sotiropoulou, et E. Passegue. DNA-damage response in tissue-specific and cancer stem cells. *Cell Stem Cell*, 8(1) :16 – 29, 2011.
- A. Bothmer, D. F. Robbiani, N. Feldhahn, A. Gazumyan, A. Nussenzweig, et M. C. Nussenzweig. 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J Exp Med*, 207(4) :855 – 865, 2010.
- I. Boubakour-Azzouz, P. Bertrand, A. Claes, B. S. Lopez, et F. Rougeon. Terminal deoxynucleotidyl transferase requires KU80 and XRCC4 to promote N-addition at non-V(D)J chromosomal breaks in non-lymphoid cells. *Nucleic Acids Res.*, 40(17) :8381 – 8391, 2012.
- P. Bowden, R. Beavis, et J. Marshall. Tandem mass spectrometry of human tryptic blood peptides calculated by a statistical algorithm and captured by a relational database with exploration by a general statistical analysis system. *J Proteomics*, 73(1) :103 – 111, 2009.
- S. Britton, C. Froment, P. Frit, B. Monsarrat, B. Salles, et P. Calsou. Cell nonhomologous end joining capacity controls SAF-A phosphorylation by DNA-PK in response to DNA double-strand breaks inducers. *Cell Cycle*, 8(22) :3717 – 3722, 2009.
- K. D. Bromberg, R. Velez-Cruz, A. B. Burgin, et N. Osheroff. DNA ligation catalyzed by human topoisomerase II alpha. *Biochemistry*, 43(42) :13416 – 13423, 2004.
- D. Buck, L. Malivert, R. de Chasseval, A. Barraud, M.-C. Fondaneche, O. Sanal, A. Plebani, J.-L. Stephan, M. Hufnagel, F. le Deist, A. Fischer, A. Durandy, J.-P. de Villartay, et P. Revy. Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell*, 124(2) :287 – 299, 2006.
- S. F. Bunting, E. Callen, N. Wong, H.-T. Chen, F. Polato, A. Gunn, A. Bothmer, N. Feldhahn, O. Fernandez-Capetillo, L. Cao, X. Xu, C.-X. Deng, T. Finkel, M. Nussenzweig, J. M. Stark, et A. Nussenzweig. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell*, 141(2) :243 – 254, 2010.
- J. Cadet et J. R. Wagner. DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. *Cold Spring Harb Perspect Biol*, 5(2), 2013.

- L. Cao, E. Alani, et N. Kleckner. A pathway for generation and processing of double-strand breaks during meiotic recombination in *S. Cerevisiae*. *Cell*, 61(6) :1089 – 1101, 1990.
- L. Cao, X. Xu, S. F. Bunting, J. Liu, R.-H. Wang, L. L. Cao, J. J. Wu, T.-N. Peng, J. Chen, A. Nussenzweig, C.-X. Deng, et T. Finkel. A selective requirement for 53BP1 in the biological response to genomic instability induced by Brca1 deficiency. *Mol Cell*, 35(4) :534 – 541, 2009.
- J. P. Carney, R. S. Maser, H. Olivares, E. M. Davis, M. Le Beau, J. R. Yates III, L. Hays, W. F. Morgan, et J. H. Petrini. The hMre11/hRad50 protein complex and Nijmegen breakage syndrome : Linkage of double-strand break repair to the cellular DNA damage response. *Cell*, 93(3) :477 – 486, 1998.
- K. Cartharius, K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein, et T. Werner. MatInspector and beyond : Promoter analysis based on transcription factor binding sites. *Bioinformatics*, 21(13) :2933 – 2942, 2005.
- L. Cassimeris et J. Morabito. TOGp, the human homolog of XMAP215/Dis1, is required for centrosome integrity, spindle pole organization, and bipolar spindle assembly. *Mol Biol Cell*, 15(4) :1580 – 1590, 2004.
- A. Castello, B. Fischer, K. Eichelbaum, R. Horos, B. M. Beckmann, C. Strein, N. E. Davey, D. T. Humphreys, T. Preiss, L. M. Steinmetz, J. Krijgsveld, et M. W. Hentze. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, 149(6) :1393 – 1406, 2012.
- M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T. A. Baker, M.-Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S. L. Seymour, L. M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E. W. Deutsch, R. L. Moritz, J. E. Katz, D. B. Agus, M. MacCoss, D. L. Tabb, et P. Mallick. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.*, 30(10) :918 – 920, 2012.
- J. R. Chapman, M. R. G. Taylor, et S. J. Boulton. Playing the end game : DNA double-strand break repair pathway choice. *Mol Cell*, 47(4) :497 – 510, 2012.
- J. Chaumeil, M. Micsinai, P. Ntziachristos, D. B. Roth, I. Aifantis, Y. Kluger, L. Deriano, et J. A. Skok. The RAG2 C-terminus and ATM protect genome

- integrity by controlling antigen receptor gene cleavage. *Nat Commun*, 4 : 2231, 2013.
- L. Chen, K. Trujillo, P. Sung, et A. E. Tomkinson. Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the. *J. Biol. Chem.*, 275 (34) :26196 – 26205, 2000.
- C. T. Cheung, R. Singh, R. S. Kalra, S. C. Kaul, et R. Wadhwa. Collaborator of ARF (CARF) regulates proliferative fate of human cells by dose-dependent regulation of DNA damage signaling. *J. Biol. Chem.*, 289 (26) :18258 – 18269, 2014.
- M. Chevallet, S. Luche, et T. Rabilloud. Silver staining of proteins in polyacrylamide gels. *Nat Protoc*, 1(4) :1852 – 1858, 2006.
- S. H. Cho, A. K. Ahn, P. Bhargava, C.-H. Lee, C. M. Eischen, O. McGuinness, et M. Boothby. Glycolytic rate and lymphomagenesis depend on PARP14, an ADP ribosyltransferase of the B aggressive lymphoma (BAL) family. *Proc. Natl. Acad. Sci. USA*, 108(38) :15972 – 15977, 2011.
- L. A. Chodosh et S. Buratowski. Purification of DNA-binding proteins using biotin/streptavidin affinity systems. *Curr Protoc Protein Sci*, Chapter - 9 :Unit – 9 – 7, 2001.
- J. Chua, R. Rikhy, et J. Lippincott-Schwartz. Dynamin 2 orchestrates the global actomyosin cytoskeleton for epithelial maintenance and apical constriction. *Proc. Natl. Acad. Sci. USA*, 106(49) :20770 – 20775, 2009.
- A. Ciccia et S. J. Elledge. The DNA damage response : Making it safe to play with knives. *Mol Cell*, 40(2) :179 – 204, 2010.
- A. Ciccia, A. L. Bredemeyer, M. E. Sowa, M.-E. Terret, P. V. Jallepalli, J. W. Harper, et S. J. Elledge. The SIOD disorder protein SMARCA1 is an RPA-interacting protein involved in replication fork restart. *Genes Dev*, 23(20) :2415 – 2425, 2009.
- N. Collins, R. A. Poot, I. Kukimoto, C. Garcia-Jimenez, G. Dellaire, et P. D. Varga-Weisz. An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nat Genet*, 32(4) :627 – 632, 2002.
- T. de Lange. Shelterin : The protein complex that shapes and safeguards human telomeres. *Genes Dev*, 19(18) :2100 – 2110, 2005.
- M. S. DeMott, S. Zigman, et R. A. Bambara. Replication protein A stimulates long patch DNA base excision repair. *J. Biol. Chem.*, 273(42) :27492 – 27498, 1998.

- B. Dey, S. Thukral, S. Krishnan, M. Chakrobarty, S. Gupta, C. Manghani, et V. Rani. DNA-protein interactions : Methods for detection and analysis. *Mol. Cell. Biochem.*, 365(1 - 2) :279 – 299, 2012.
- C. Diatloff-Zito, A. J. Gordon, E. Duchaud, et G. Merlin. Isolation of an ubiquitously expressed cDNA encoding human dynamin II, a member of the large GTP-binding protein family. *Gene*, 163(2) :301 – 306, 1995.
- S. J. DiBiase, Z. C. Zeng, R. Chen, T. Hyslop, W. J. Curran Jr, et G. Iliakis. DNA-dependent protein kinase stimulates an independently active, nonhomologous, end-joining apparatus. *Cancer Res*, 60(5) :1245 – 1253, 2000.
- T. Dobzhansky. Genetics of Natural Populations. Xiii. Recombination and Variability in Populations of *Drosophila Pseudoobscura*. *Genetics*, 31(3) : 269 – 290, 1946.
- M. Dosil et X. R. Bustelo. Functional characterization of Pwp2, a WD family protein essential for the assembly of the 90 S pre-ribosomal particle. *J. Biol. Chem.*, 279(36) :37385 – 37397, 2004.
- F. Dragon, J. E. G. Gallagher, P. A. Compagnone-Post, B. M. Mitchell, K. A. Porwancher, K. A. Wehner, S. Wormsley, R. E. Settlege, J. Shabanowitz, Y. Osheim, A. L. Beyer, D. F. Hunt, et S. J. Baserga. A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature*, 417(6892) :967 – 970, 2002.
- J. Drouet, P. Frit, C. Delteil, J.-P. de Villartay, B. Salles, et P. Calsou. Interplay between Ku, Artemis, and the DNA-dependent protein kinase catalytic subunit at DNA ends. *J. Biol. Chem.*, 281(38) :27784 – 27793, 2006.
- N. Dyballa et S. Metzger. Fast and sensitive coomassie staining in quantitative proteomics. *Methods Mol Biol*, 893 :47 – 59, 2012.
- J. Falck, J. Coates, et S. P. Jackson. Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature*, 434(7033) : 605 – 611, 2005.
- F. Fattah, E. H. Lee, N. Weisensel, Y. Wang, N. Lichter, et E. A. Hendrickson. Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells. *PLoS Genet*, 6(2) : e1000855, 2010.
- O. Fernandez-Capetillo, A. Celeste, et A. Nussenzweig. Focusing on foci : H2AX and the recruitment of DNA-damage response factors. *Cell Cycle*, 2(5) :426 – 427, 2003.

- L. P. Ferretti, L. Lafranchi, et A. A. Sartori. Controlling DNA-end resection : A new task for CDKs. *Front Genet*, 4 :99, 2013.
- J. Fillingham, M.-C. Keogh, et N. J. Krogan. GammaH2AX and its role in DNA double-strand break repair. *Biochem. Cell Biol.*, 84(4) :568 – 577, 2006.
- N. J. Finnie, T. M. Gottlieb, T. Blunt, P. A. Jeggo, et S. P. Jackson. DNA-dependent protein kinase activity is absent in xrs-6 cells : Implications for site-specific recombination and DNA double-strand break repair. *Proc. Natl. Acad. Sci. USA*, 92(1) :320 – 324, 1995.
- A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, et C. C. Mello. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669) :806 – 811, 1998.
- M. C. Fletcher et K. R. Fox. Dissociation kinetics of actinomycin D from individual GpC sites in DNA. *Eur. J. Biochem.*, 237(1) :164 – 170, 1996.
- M. J. Fullwood et Y. Ruan. ChIP-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.*, 107(1) :30 – 39, 2009.
- G. M. Fulop et R. A. Phillips. The scid mutation in mice causes a general defect in DNA repair. *Nature*, 347 :479–482, 1990.
- D. J. Galas et A. Schmitz. DNase footprinting : A simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.*, 5(9) : 3157 – 3170, 1978.
- D. Garcia-Domingo, D. Ramirez, G. G. de Buitrago, et C. Martinez-A. Death inducer-obliterator 1 triggers apoptosis after nuclear translocation and caspase upregulation. *Mol Cell Biol*, 23(9) :3216 – 3225, 2003.
- A. Gdowicz-Klosok, M. Widel, et J. Rzeszowska-Wolny. The influence of XPD, APE1, XRCC1, and NBS1 polymorphic variants on DNA repair in cells exposed to X-rays. *Mutat Res*, 755(1) :42 – 48, 2013.
- R. S. Geha, H. H. Jabara, et S. R. Brodeur. The regulation of immunoglobulin E class-switch recombination. *Nat Rev Immunol*, 3(9) :721 – 732, 2003.
- F. Gergely, V. M. Draviam, et J. W. Raff. The ch-TOG/XMAP215 protein is essential for spindle pole organization in human somatic cells. *Genes Dev*, 17(3) :336 – 341, 2003.

- A. Germani, A. Prabel, S. Mourah, M.-P. Podgorniak, A. Di Carlo, R. Ehrlich, S. Gisselbrecht, N. Varin-Blank, F. Calvo, et H. Bruzzoni-Giovanelli. SIAH-1 interacts with CtIP and promotes its degradation by the proteasome pathway. *Oncogene*, 22(55) :8845 – 8851, 2003.
- H. Ghezraoui, M. Piganeau, B. Renouf, J.-B. Renaud, A. Sallmyr, B. Ruis, S. Oh, A. E. Tomkinson, E. A. Hendrickson, C. Giovannangeli, M. Jasin, et E. Brunet. Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Mol Cell*, 55(6) :829 – 842, 2014.
- Z. Gong et J. Chen. E3 ligase RFWD3 participates in replication checkpoint control. *J. Biol. Chem.*, 286(25) :22308 – 22313, 2011.
- A. A. Goodarzi et P. A. Jeggo. The heterochromatic barrier to DNA double strand break repair : How to get the entry visa. *Int J Mol Sci*, 13(9) : 11844 – 11860, 2012.
- A. Gospodinov et Z. Herceg. Chromatin structure in double strand break repair. *DNA Repair (Amst)*, 12(10) :800 – 810, 2013.
- S. Gradia, D. Subramanian, T. Wilson, S. Acharya, A. Makhov, J. Griffith, et R. Fishel. HMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. *Mol Cell*, 3(2) :255 – 261, 1999.
- L. Gros, M. K. Sapparbaev, et J. Laval. Enzymology of the repair of free radicals-induced DNA damage. *Oncogene*, 21(58) :8905 – 8925, 2002.
- A. Grudic, A. Jul-Larsen, S. J. Haring, M. S. Wold, P. E. Lonning, R. Bjerkvig, et S. O. Boe. Replication protein A prevents accumulation of single-stranded telomeric DNA in cells that use alternative lengthening of telomeres. *Nucleic Acids Res.*, 35(21) :7267 – 7278, 2007.
- G. J. Grundy, H. A. Moulding, K. W. Caldecott, et S. L. Rulten. One ring to bring them all—the role of Ku in mammalian non-homologous end joining. *DNA Repair (Amst)*, 17 :30 – 38, 2014.
- J. Guirouilh-Barbat, E. Rass, I. Plo, P. Bertrand, et B. S. Lopez. Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends. *Proc. Natl. Acad. Sci. USA*, 104(52) :20902 – 20907, 2007.
- L. H. Hartwell, P. Szankasi, C. J. Roberts, A. W. Murray, et S. H. Friend. Integrating genetic approaches into the discovery of anticancer drugs. *Science*, 278(5340) :1064 – 1068, 1997.
- Z. He, L. A. Henricksen, M. S. Wold, et C. J. Ingles. RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature*, 374(6522) :566 – 569, 1995.

- N. Hégarat, G. M. Cardoso, F. Rusconi, J.-C. François, et D. Praseuth. Analytical biochemistry of DNA–protein assemblies from crude cell extracts. *Nucleic Acids Res.*, 35(13) :e92, 2007.
- N. Hégarat, J.-C. François, et D. Praseuth. Modern tools for identification of nucleic acid-binding proteins. *Biochimie*, 90(9) :1265 – 1272, 2008.
- S. Hein, S. Mahner, C. Kanowski, T. Loning, F. Janicke, et K. Milde-Langosch. Expression of Jun and Fos proteins in ovarian tumors of different malignant potential and in ovarian cancer cell lines. *Oncol Rep*, 22(1) :177 – 183, 2009.
- L. M. Hellman et M. G. Fried. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc*, 2(8) :1849 – 1861, 2007.
- K. Horiuchi, M. Umetani, T. Minami, H. Okayama, S. Takada, M. Yamamoto, H. Aburatani, P. C. Reid, D. E. Housman, T. Hamakubo, et T. Kodama. Wilms' tumor 1-associating protein regulates G2/M transition through stabilization of cyclin A2 mRNA. *Proc. Natl. Acad. Sci. USA*, 103(46) :17278 – 17283, 2006.
- N. Hoshi, T. Hirano, T. Omotehara, J. Tokumoto, Y. Umemura, Y. Mantani, T. Tanida, K. Warita, Y. Tabuchi, T. Yokoyama, et H. Kitagawa. Insight into the mechanism of reproductive dysfunction caused by neonicotinoid pesticides. *Biol. Pharm. Bull.*, 37(9) :1439 – 1443, 2014.
- M. Hosokawa, R. Kadota, S. Shichijo, K. Itoh, I. Dmitriev, V. Krasnykh, D. T. Curiel, Y. Takue, H. Wakasugi, S. Takashima, et Y. Heike. Cell cycle arrest and apoptosis induced by SART-1 gene transduction. *Anticancer Res.*, 25(3B) :1983 – 1990, 2005.
- J. Hu, C. M. McCall, T. Ohta, et Y. Xiong. Targeted ubiquitination of CDT1 by the DDB1-CUL4A-ROC1 ligase in response to DNA damage. *Nat Cell Biol*, 6(10) :1003 – 1009, 2004.
- S. Hu, Z. Xie, A. Onishi, X. Yu, L. Jiang, J. Lin, H. s. Rho, C. Woodard, H. Wang, J.-S. Jeong, S. Long, X. He, H. Wade, S. Blackshaw, J. Qian, et H. Zhu. Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell*, 139(3) :610 – 622, 2009.
- J. Huang, J. Dorsey, S. Chuikov, L. Perez-Burgos, X. Zhang, T. Jenuwein, D. Reinberg, et S. L. Berger. G9a and Glp methylate lysine 373 in the tumor suppressor p53. *J. Biol. Chem.*, 285(13) :9636 – 9641, 2010.

- M. Huarte, M. Guttman, D. Feldser, M. Garber, M. J. Koziol, D. Kenzelmann-Broz, A. M. Khalil, O. Zuk, I. Amit, M. Rabani, L. D. Attardi, A. Regev, E. S. Lander, T. Jacks, et J. L. Rinn. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, 142(3) :409 – 419, 2010.
- W. H. Hudson et E. A. Ortlund. The structure, function and evolution of proteins that bind DNA and RNA. *Nat Rev Mol Cell Biol*, 15(11) :749 – 760, 2014.
- P. Huertas. DNA resection in eukaryotes : Deciding how to fix the break. *Nat Struct Mol Biol*, 17(1) :11 – 16, 2010.
- I. Iaccarino, G. Marra, F. Palombo, et J. Jiricny. HMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSalpha. *EMBO J*, 17(9) :2677 – 2686, 1998.
- Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, et M. Mann. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics*, 4(9) :1265 – 1272, 2005.
- T. Iyama et D. M. Wilson. DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst)*, 12(8) :620 – 636, 2013.
- D. Jiang, Y. Jia, Y. Zhou, et H. W. Jarrett. Two-dimensional southwestern blotting and characterization of transcription factors on-blot. *J. Proteome Res.*, 8(7) :3693 – 3701, 2009.
- A. Jilani, D. Ramotar, C. Slack, C. Ong, X. M. Yang, S. W. Scherer, et D. D. Lasko. Molecular cloning of the human gene, PNKP, encoding a polynucleotide kinase 3'-phosphatase and evidence for its role in repair of DNA strand breaks caused by oxidative damage. *J. Biol. Chem.*, 274(34) :24176 – 24186, 1999.
- A. Jolma et J. Taipale. Methods for Analysis of Transcription Factor DNA-Binding Specificity In Vitro. *Subcell Biochem*, 52 :155 – 173, 2011.
- J. M. Jones et M. Gellert. The taming of a transposon : V(D)J recombination and the immune system. *Immunol Rev*, 200 :233 – 248, 2004.
- H.-Y. Joo, S. R. Woo, Y.-N. Shen, M. Y. Yun, H.-J. Shin, E.-R. Park, S.-H. Kim, J.-E. Park, Y.-J. Ju, S. H. Hong, S.-G. Hwang, M.-H. Cho, J. Kim, et K.-H. Lee. SIRT1 interacts with and protects glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from nuclear translocation : Implications for

- cell survival after irradiation. *Biochem. Biophys. Res. Commun.*, 424(4) : 681 – 686, 2012.
- W. G. Kaelin Jr. Synthetic lethality : A framework for the development of wiser cancer therapeutics. *Genome Med*, 1(10) :99, 2009.
- L. Y. Kadyrova, E. R. Blanco, et F. A. Kadyrov. CAF-I-dependent control of degradation of the discontinuous strands during mismatch repair. *Proc. Natl. Acad. Sci. USA*, 108(7) :2753 – 2758, 2011.
- A. E. Kel, E. Gossling, I. Reuter, E. Cheremushkin, O. V. Kel-Margoulis, et E. Wingender. MATCH : A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.*, 31(13) :3576 – 3579, 2003.
- D. M. Keller, X. Zeng, Y. Wang, Q. H. Zhang, M. Kapoor, H. Shu, R. Goodman, G. Lozano, Y. Zhao, et H. Lu. A DNA damage-induced p53 serine 392 kinase complex contains CK2, hSpt16, and SSRP1. *Mol Cell*, 7(2) : 283 – 292, 2001.
- S. Kim, J. Yamamoto, Y. Chen, M. Aida, T. Wada, H. Handa, et Y. Yamaguchi. Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. *Genes Cells*, 15(9) :1003 – 1013, 2010.
- C. Kisker, W. Hinrichs, K. Tovar, W. Hillen, et W. Saenger. The complex formed between Tet repressor and tetracycline-Mg<sup>2+</sup> reveals mechanism of antibiotic resistance. *J. Mol. Biol.*, 247(2) :260 – 280, 1995.
- M. D. Kornberg, N. Sen, M. R. Hara, K. R. Juluri, J. Van K. Nguyen, A. M. Snowman, L. Law, L. D. Hester, et S. H. Snyder. GAPDH mediates nitrosylation of nuclear proteins. *Nat Cell Biol*, 12(11) :1094 – 1100, 2010.
- S. G. Kozmin, Y. Sedletska, A. Reynaud-Angelin, D. Gasparutto, et E. Sage. The formation of double-strand breaks at multiply damaged sites is driven by the kinetics of excision/incision at base damage in eukaryotic cells. *Nucleic Acids Res.*, 37(6) :1767 – 1777, 2009.
- A. Kramer, N. Mailand, C. Lukas, R. G. Syljuasen, C. J. Wilkinson, E. A. Nigg, J. Bartek, et J. Lukas. Centrosome-associated Chk1 prevents premature activation of cyclin-B-Cdk1 kinase. *Nat Cell Biol*, 6(9) :884 – 891, 2004.
- L. Krejci, V. Altmannova, M. Spirek, et X. Zhao. Homologous recombination and its regulation. *Nucleic Acids Res.*, 40(13) :5795 – 5818, 2012.

- B. E. Kremer, L. A. Adang, et I. G. Macara. Septins regulate actin organization and cell-cycle arrest through nuclear accumulation of NCK mediated by SOCS7. *Cell*, 130(5) :837 – 850, 2007.
- J. Kuper, C. Braun, A. Elias, G. Michels, F. Sauer, D. R. Schmitt, A. Poterszman, J.-M. Egly, et C. Kisker. In TFIIH, XPD helicase is exclusively devoted to DNA repair. *PLoS Biol*, 12(9) :e1001954, 2014.
- J.-H. Lee et T. T. Paull. Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. *Science*, 304(5667) :93 – 96, 2004.
- J. H. Lee, S. B. Rho, et T. Chun. GABAA receptor-associated protein (GABARAP) induces apoptosis by interacting with DEAD (Asp-Glu-Ala-Asp/His) box polypeptide 47 (DDX 47). *Biotechnol. Lett*, 27(9) :623 – 628, 2005.
- D. S. Levin, A. E. McKenna, T. A. Motycka, Y. Matsumoto, et A. E. Tomkinson. Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. *Curr Biol*, 10(15) :919 – 922, 2000.
- G. Li, C. Nelsen, et E. A. Hendrickson. Ku86 is essential in human somatic cells. *Proc. Natl. Acad. Sci. USA*, 99(2) :832 – 837, 2002.
- K. W. Li, N. Chen, P. Klemmer, F. Koopmans, R. Karupothula, et A. B. Smit. Identifying true protein complex constituents in interaction proteomics : The example of the DMXL2 protein complex. *Proteomics*, 12(15 - 16) : 2428 – 2432, 2012.
- S. Li, Z. Li, F.-J. Shu, H. Xiong, A. C. Phillips, et W. S. Dynan. Double-strand break repair deficiency in NONO knockout murine embryonic fibroblasts and compensation by spontaneous upregulation of the PSPC1 paralog. *Nucleic Acids Res.*, 42(15) :9771 – 9780, 2014.
- X. Li, B. Guan, M. K. Srivastava, A. Padmanabhan, B. S. Hampton, et C. J. Bieberich. The reverse in-gel kinase assay to profile physiological kinase substrates. *Nat Methods*, 4(11) :957 – 962, 2007.
- M. R. Lieber. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem*, 79 :181 – 211, 2010.
- B. Liedberg, C. Nylander, et I. Lunström. Surface plasmon resonance for gas detection and biosensing. *Sensors and Actuators*, 4(0) :299 – 304, 1983.

- Y. Liu, H. Kim, J. Liang, W. Lu, B. Ouyang, D. Liu, et Z. Songyang. The death-inducer obliterator 1 (Dido1) gene regulates embryonic stem cell self-renewal. *J. Biol. Chem.*, 289(8) :4778 – 4786, 2014.
- C. M. Lovly, L. Yan, C. E. Ryan, S. Takada, et H. Piwnicka-Worms. Regulation of Chk2 ubiquitination and signaling through autophosphorylation of serine 379. *Mol Cell Biol*, 28(19) :5874 – 5885, 2008.
- J. C. Lucchesi. Synthetic lethality and semi-lethality among functionally related mutants of *Drosophila melanogaster*. *Genetics*, 59(1) :37 – 44, 1968.
- M. S. Luijsterburg, K. Acs, L. Ackermann, W. W. Wiegant, S. Bekker-Jensen, D. H. Larsen, K. K. Khanna, H. van Attikum, N. Mailand, et N. P. Dantuma. A new non-catalytic role for ubiquitin ligase RNF8 in unfolding higher-order chromatin structure. *EMBO J*, 31(11) :2511 – 2527, 2012.
- B. L. Mahaney, K. Meek, et S. P. Lees-Miller. Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem. J*, 417(3) :639 – 650, 2009.
- R. J. Mailloux, R. Darwich, J. Lemire, et V. Appanna. The monitoring of nucleotide diphosphate kinase activity by blue native polyacrylamide gel electrophoresis. *Electrophoresis*, 29(7) :1484 – 1489, 2008.
- S. Malyarchuk, R. Castore, R. Shi, et L. Harrison. Artemis is required to improve the accuracy of repair of double-strand breaks with 5'-blocked termini generated from non-DSB-clustered lesions. *Mutagenesis*, 28(3) : 357 – 366, 2013.
- P.-O. Mari, B. I. Florea, S. P. Persengiev, N. S. Verkaik, H. T. Bruggenwirth, M. Modesti, G. Giglia-Mari, K. Bezstarosti, J. A. A. Demmers, T. M. Luider, A. B. Houtsmuller, et D. C. van Gent. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc. Natl. Acad. Sci. USA*, 103(49) :18597 – 18602, 2006.
- J. Martinez, A. Patkaniowska, H. Urlaub, R. Luhrmann, et T. Tuschl. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell*, 110 (5) :563 – 574, 2002.
- R. Mathew, K. Hartmuth, S. Mohlmann, H. Urlaub, R. Ficner, et R. Luhrmann. Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nat Struct Mol Biol*, 15(5) :435 – 443, 2008.

- S. Matsuoka, M. Huang, et S. J. Elledge. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science*, 282(5395) :1893 – 1897, 1998.
- K. Meek, P. Douglas, X. Cui, Q. Ding, et S. P. Lees-Miller. Trans Autophosphorylation at DNA-dependent protein kinase's two major autophosphorylation site clusters facilitates end processing but not end joining. *Mol Cell Biol*, 27(10) :3881 – 3890, 2007.
- B. Meyer, K.-O. Voss, F. Tobias, B. Jakob, M. Durante, et G. Taucher-Scholz. Clustered DNA damage induces pan-nuclear H2AX phosphorylation mediated by ATM and. *Nucleic Acids Res.*, 41(12) :6109 – 6118, 2013.
- J. Mi, C. Guo, D. L. Brautigan, et J. M. Larner. Protein phosphatase-1alpha regulates centrosome splitting through Nek2. *Cancer Res*, 67(3) :1082 – 1089, 2007.
- E. Mladenov, S. Magin, A. Soni, et G. Iliakis. DNA double-strand break repair as determinant of cellular radiosensitivity to killing and target in radiation therapy. *Front Oncol*, 3 :113, 2013.
- D. Moshous, L. Li, R. d. Chasseval, N. Philippe, N. Jabado, M. J. Cowan, A. Fischer, et J.-P. d. Villartay. A new gene involved in DNA double-strand break repair and V(D)J recombination is located on human chromosome 10p. *Human Molecular Genetics*, 9(4) :583 – 588, 2000.
- D. Moshous, I. Callebaut, R. de Chasseval, B. Corneo, M. Cavazzana-Calvo, F. Le Deist, I. Tezcan, O. Sanal, Y. Bertrand, N. Philippe, A. Fischer, et J. P. de Villartay. Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell*, 105(2) :177 – 186, 2001.
- T. E. Mullen et W. F. Marzluff. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev*, 22(1) :50 – 65, 2008.
- M. Nakanishi, T. Ozaki, H. Yamamoto, T. Hanamoto, H. Kikuchi, K. Furuya, M. Asaka, D. Delia, et A. Nakagawara. NFBD1/MDC1 associates with p53 and regulates its function at the crossroad between cell survival and death in response to DNA damage. *J. Biol. Chem.*, 282(31) :22993 – 23004, 2007.
- J. A. Neal et K. Meek. Choosing the right path : Does DNA-PK help make the decision ? *Mutat Res*, 711(1 - 2) :73 – 86, 2011.

- J. A. Neal, V. Dang, P. Douglas, M. S. Wold, S. P. Lees-Miller, et K. Meek. Inhibition of homologous recombination by DNA-dependent protein kinase requires kinase activity, is titratable, and is modulated by autophosphorylation. *Mol Cell Biol*, 31(8) :1719 – 1733, 2011.
- J. Olejnik, E. Krzymanska-Olejnik, et K. J. Rothschild. Photocleavable biotin phosphoramidite for 5'-end-labeling, affinity purification and phosphorylation of synthetic oligonucleotides. *Nucleic Acids Res.*, 24(2) :361 – 366, 1996.
- G. Orphanides, W. H. Wu, W. S. Lane, M. Hampsey, et D. Reinberg. The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature*, 400(6741) :284 – 288, 1999.
- M. L. Orta, A. Hoglund, J. M. Calderon-Montano, I. Dominguez, E. Burgos-Moron, T. Visnes, N. Pastor, C. Strom, M. Lopez-lazaro, et T. Helleday. The PARP inhibitor Olaparib disrupts base excision repair of 5-aza-2'-deoxycytidine lesions. *Nucleic Acids Res.*, 42(14) :9108 – 9120, 2014.
- F. Oswald, M. Winkler, Y. Cao, K. Astrahantseff, S. Bourteele, W. Knochel, et T. Borggrefe. RBP-Jkappa/SHARP recruits CtIP/CtBP corepressors to silence Notch target genes. *Mol Cell Biol*, 25(23) :10379 – 10390, 2005.
- R. M. Overmeer, A. M. Gourdin, A. Giglia-Mari, H. Kool, A. B. Houtsmuller, G. Siegal, M. I. Fousteri, L. H. F. Mullenders, et W. Vermeulen. Replication factor C recruits DNA polymerase delta to sites of nucleotide excision repair but is not required for PCNA recruitment. *Mol Cell Biol*, 30(20) : 4828 – 4839, 2010.
- M. S. Park, D. L. Ludwig, E. Stigger, et S. H. Lee. Physical interaction between human RAD52 and RPA is required for homologous recombination in mammalian cells. *J. Biol. Chem.*, 271(31) :18996 – 19000, 1996.
- E. Parlanti, B. Pascucci, G. Terrados, L. Blanco, et E. Dogliotti. Aphidicolin-resistant and -sensitive base excision repair in wild-type and DNA polymerase beta-defective mouse cells. *DNA Repair (Amst)*, 3(7) :703 – 710, 2004.
- T. T. Paull, E. P. Rogakou, V. Yamazaki, C. U. Kirchgessner, M. Gellert, et W. M. Bonner. A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr Biol*, 10(15) :886 – 895, 2000.
- R. Pavri, B. Zhu, G. Li, P. Trojer, S. Mandal, A. Shilatifard, et D. Reinberg. Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell*, 125(4) :703 – 717, 2006.

- S. Paytubi, X. Wang, Y. W. Lam, L. Izquierdo, M. J. Hunter, E. Jan, H. S. Hundal, et C. G. Proud. ABC50 promotes translation initiation in mammalian cells. *J. Biol. Chem.*, 284(36) :24061 – 24073, 2009.
- F. Pessina et N. F. Lowndes. The RSF1 histone-remodelling factor facilitates DNA double-strand break repair by recruiting centromeric and Fanconi Anaemia proteins. *PLoS Biol*, 12(5) :e1001856, 2014.
- E. Petermann, C. Keil, et S. L. Oei. Roles of DNA ligase III and XRCC1 in regulating the switch between short patch and long patch BER. *DNA Repair (Amst)*, 5(5) :544 – 555, 2006.
- M. Piganeau, H. Ghezraoui, A. De Cian, L. Guittat, M. Tomishima, L. Perrouault, O. Rene, G. E. Katibah, L. Zhang, M. C. Holmes, Y. Doyon, J.-P. Concordet, C. Giovannangeli, M. Jasin, et E. Brunet. Cancer translocations in human cells induced by zinc finger and TALE nucleases. *Genome Res*, 23(7) :1182 – 1193, 2013.
- J. J. Pouliot, K. C. Yao, C. A. Robertson, et H. A. Nash. Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. *Science*, 286(5439) :552 – 555, 1999.
- L. F. Povirk, T. Zhou, R. Zhou, M. J. Cowan, et S. M. Yannone. Processing of 3'-phosphoglycolate-terminated DNA double strand breaks by Artemis nuclease. *J. Biol. Chem.*, 282(6) :3547 – 3558, 2007.
- Y. Qi, J. A. Ranish, X. Zhu, A. Kronen, J. Zhang, R. Aebersold, D. W. Rose, M. G. Rosenfeld, et C. Carriere. Atbfl is required for the Pit1 gene early activation. *Proc. Natl. Acad. Sci. USA*, 105(7) :2481 – 2486, 2008.
- S. Rajagopalan, J. Nepa, et S. Venkatachalam. Chromodomain helicase DNA-binding protein 2 affects the repair of X-ray and. *Environ Mol Mutagen*, 53(1) :44 – 50, 2012.
- M. Raman, S. Earnest, K. Zhang, Y. Zhao, et M. H. Cobb. TAO kinases mediate activation of p38 in response to DNA damage. *EMBO J*, 26(8) : 2005 – 2014, 2007.
- J. Rappsilber, U. Ryder, A. I. Lamond, et M. Mann. Large-scale proteomic analysis of the human spliceosome. *Genome Res*, 12(8) :1231 – 1245, 2002.
- E. Rass, A. Grabarz, I. Plo, J. Gautier, P. Bertrand, et B. S. Lopez. Role of Mre11 in chromosomal nonhomologous end joining in mammalian cells. *Nat Struct Mol Biol*, 16(8) :819 – 824, 2009.

- E. Rass, A. Grabarz, P. Bertrand, et B.-S. Lopez. [Double strand break repair, one mechanism can hide another : Alternative non-homologous end joining]. *Cancer Radiother*, 16(1) :1 – 10, 2012.
- J. T. Reardon et A. Sancar. Nucleotide excision repair. *Prog Nucleic Acid Res Mol Biol*, 79 :183 – 235, 2005.
- H. C. Reinhardt, P. Hasskamp, I. Schmedding, S. Morandell, M. A. T. M. van Vugt, X. Wang, R. Linding, S.-E. Ong, D. Weaver, S. A. Carr, et M. B. Yaffe. DNA damage activates a spatially distinct late cytoplasmic cell-cycle checkpoint network controlled by MK2-mediated RNA stabilization. *Mol Cell*, 40(1) :34 – 49, 2010.
- S. A. Roberts, N. Strande, M. D. Burkhalter, C. Strom, J. M. Havener, P. Hasty, et D. A. Ramsden. Ku is a 5'-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature*, 464(7292) :1214 – 1217, 2010.
- P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, et D. J. Pappin. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*, 3(12) :1154 – 1169, 2004.
- K. L. Rossow et R. Janknecht. Synergism between p68 RNA helicase and the transcriptional coactivators CBP and p300. *Oncogene*, 22(1) :151 – 156, 2003.
- M. R. Rountree, K. E. Bachman, et S. B. Baylin. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat Genet*, 25(3) :269 – 277, 2000.
- E. Sage et L. Harrison. Clustered DNA lesion repair in eukaryotes : Relevance to mutagenesis and cell survival. *Mutat Res*, 711(1 - 2) :123 – 133, 2011.
- S. Sahara, M. Aoto, Y. Eguchi, N. Imamoto, Y. Yoneda, et Y. Tsujimoto. Acinus is a caspase-3-activated protein required for apoptotic chromatin condensation. *Nature*, 401(6749) :168 – 173, 1999.
- C. P. Samora, B. Mogessie, L. Conway, J. L. Ross, A. Straube, et A. D. McAinsh. MAP4 and CLASP1 operate as a safety mechanism to maintain a stable spindle position in mitosis. *Nat Cell Biol*, 13(9) :1040 – 1050, 2011.

- S. Samuel, J.-C. Twizere, et L. R. Bernstein. YB-1 represses AP1-dependent gene transactivation and interacts with an AP-1 DNA sequence. *Biochem. J*, 388(Pt - 3) :921 – 928, 2005.
- I. Saugar, M. A. Ortiz-Bazan, et J. A. Tercero. Tolerating DNA damage during eukaryotic chromosome replication. *Exp Cell Res*, 2014.
- M. Sayed, S. Pelech, C. Wong, A. Marotta, et B. Salh. Protein kinase CK2 is involved in G2 arrest and apoptosis following spindle damage in epithelial cells. *Oncogene*, 20(48) :6994 – 7005, 2001.
- H. Schagger, W. A. Cramer, et G. von Jagow. Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal. Biochem.*, 217(2) :220 – 230, 1994.
- A. Schipler et G. Iliakis. DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. *Nucleic Acids Res.*, 41(16) :7589 – 7605, 2013.
- E. Schneider, M. Montenarh, et P. Wagner. Regulation of CAK kinase activity by p53. *Oncogene*, 17(21) :2733 – 2741, 1998.
- L. Schulte-Uentrop, R. A. El-Awady, L. Schliecker, H. Willers, et J. Dahm-Daphi. Distinct roles of XRCC4 and Ku80 in non-homologous end-joining of endonuclease- and ionizing radiation-induced DNA double-strand breaks. *Nucleic Acids Res.*, 36(8) :2561 – 2569, 2008.
- Y. Schulz, P. Wehner, L. Opitz, G. Salinas-Riester, E. M. H. F. Bongers, C. M. A. van Ravenswaaij-Arts, J. Wincent, J. Schoumans, J. Kohlhase, A. Borchers, et S. Pauli. CHD7, the gene mutated in CHARGE syndrome, regulates genes involved in neural crest cell guidance. *Hum Genet*, 133(8) :997 – 1009, 2014.
- C. Schwerk, J. Prasad, K. Degenhardt, H. Erdjument-Bromage, E. White, P. Tempst, V. J. Kidd, J. L. Manley, J. M. Lahti, et D. Reinberg. ASAP, a novel protein complex involved in RNA processing and apoptosis. *Mol Cell Biol*, 23(8) :2981 – 2990, 2003.
- R. A. Shamanna, M. Hoque, A. Lewis-Antes, E. I. Azzam, D. Lagunoff, T. Pe'ery, et M. B. Mathews. The NF90/NF45 complex participates in DNA break repair via nonhomologous end joining. *Mol Cell Biol*, 31(23) : 4832 – 4843, 2011.
- Z. Shao, A. J. Davis, K. R. Fattah, S. So, J. Sun, K.-J. Lee, L. Harrison, J. Yang, et D. J. Chen. Persistently bound Ku at DNA ends attenuates

- DNA end resection and homologous recombination. *DNA Repair (Amst)*, 11(3) :310 – 316, 2012.
- A. Shibata, S. Conrad, J. Birraux, V. Geuting, O. Barton, A. Ismail, A. Kakarougkas, K. Meek, G. Taucher-Scholz, M. Lobrich, et P. A. Jeggo. Factors determining DNA double-strand break repair pathway choice in G2 phase. *EMBO J*, 30(6) :1079 – 1092, 2011.
- Y. Shiloh. ATM : Expanding roles as a chief guardian of genome stability. *Exp Cell Res*, 2014.
- Y. Shiloh et Y. Ziv. The ATM protein kinase : Regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol*, 14(4) :197 – 210, 2013.
- N. Shimizu, M. Yoshiyama, T. Omura, A. Hanatani, S. Kim, K. Takeuchi, H. Iwao, et J. Yoshikawa. Activation of mitogen-activated protein kinases and activator protein-1 in myocardial infarction in rats. *Cardiovasc Res*, 38(1) :116 – 124, 1998.
- K. Shintomi et T. Hirano. Releasing cohesin from chromosome arms in early mitosis : Opposing actions of Wapl-Pds5 and Sgo1. *Genes Dev*, 23(18) : 2224 – 2236, 2009.
- J. Simicevic et B. Deplancke. DNA-centered approaches to characterize regulatory protein-DNA interaction complexes. *Mol Biosyst*, 6(3) :462 – 468, 2010.
- D. Simsek et M. Jasin. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat Struct Mol Biol*, 17(4) :410 – 416, 2010.
- D. Simsek et M. Jasin. DNA ligase III : A spotty presence in eukaryotes, but an essential function where tested. *Cell Cycle*, 10(21) :3636 – 3644, 2011.
- F. K. Y. Siu, L. T. O. Lee, et B. K. C. Chow. Southwestern blotting in investigating transcriptional regulation. *Nat Protoc*, 3(1) :51 – 58, 2008.
- J. Smith, E. Riballo, B. Kysela, C. Baldeyron, K. Manolis, C. Masson, M. R. Lieber, D. Papadopoulo, et P. Jeggo. Impact of DNA ligase IV on the fidelity of end joining in human cells. *Nucleic Acids Res.*, 31(8) :2157 – 2167, 2003.
- J. Smith, L. M. Tho, N. Xu, et D. A. Gillespie. The ATM-Chk2 and ATR-Chk1 pathways in DNA damage signaling and cancer. *Adv Cancer Res*, 108 :73 – 112, 2010.

- J. Snijder et A. J. R. Heck. Analytical approaches for size and mass analysis of large protein assemblies. *Annu Rev Anal Chem (Palo Alto Calif)*, 7 :43 – 64, 2014.
- A. Soni, M. Siemann, M. Grabos, T. Murmann, G. E. Pantelias, et G. Iliakis. Requirement for Parp-1 and DNA ligases 1 or 3 but not of Xrcc1 in chromosomal translocation formation by backup end joining. *Nucleic Acids Res.*, 42(10) :6380 – 6392, 2014.
- O. Sordet, A. Goldman, et Y. Pommier. Topoisomerase II and tubulin inhibitors both induce the formation of apoptotic topoisomerase I cleavage complexes. *Mol Cancer Ther*, 5(12) :3139 – 3144, 2006.
- C. S. Sorensen, L. T. Hansen, J. Dziegielewski, R. G. Syljuasen, C. Lundin, J. Bartek, et T. Helleday. The cell-cycle checkpoint kinase Chk1 is required for mammalian homologous recombination repair. *Nat Cell Biol*, 7(2) : 195 – 201, 2005.
- C. Srisawat et D. R. Engelke. RNA affinity tags for purification of RNAs and ribonucleoprotein complexes. *Methods*, 26(2) :156 – 161, 2002.
- C. Stevens, L. Smith, et N. B. La Thangue. Chk2 activates E2F-1 in response to DNA damage. *Nat Cell Biol*, 5(5) :401 – 409, 2003.
- M. Stucki et S. P. Jackson. GammaH2AX and MDC1 : Anchoring the DNA-damage-response machinery to broken chromosomes. *DNA Repair (Amst)*, 5(5) :534 – 543, 2006.
- R. Subramanian, E. M. Wilson-Kubalek, C. P. Arthur, M. J. Bick, E. A. Campbell, S. A. Darst, R. A. Milligan, et T. M. Kapoor. Insights into antiparallel microtubule crosslinking by PRC1, a conserved nonmotor microtubule binding protein. *Cell*, 142(3) :433 – 443, 2010.
- K. Sugasawa, Y. Okuda, M. Saijo, R. Nishi, N. Matsuda, G. Chu, T. Mori, S. Iwai, K. Tanaka, K. Tanaka, et F. Hanaoka. UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell*, 121(3) :387 – 400, 2005.
- K. Sugasawa, J. i. Akagi, R. Nishi, S. Iwai, et F. Hanaoka. Two-step recognition of DNA damage for mammalian nucleotide excision repair : Directional binding of the XPC complex and DNA strand scanning. *Mol Cell*, 36(4) :642 – 653, 2009.
- J. Sun, K.-J. Lee, A. J. Davis, et D. J. Chen. Human Ku70/80 protein blocks exonuclease 1-mediated DNA resection in the presence of human Mre11 or Mre11/Rad50 protein complex. *J. Biol. Chem.*, 287(7) :4936 – 4945, 2012.

- M. Sun, L. Song, Y. Li, T. Zhou, et R. S. Jope. Identification of an anti-apoptotic protein complex at death receptors. *Cell Death Differ*, 15(12) : 1887 – 1900, 2008.
- O. Surova et B. Zhivotovsky. Various modes of cell death induced by DNA damage. *Oncogene*, 32(33) :3789 – 3797, 2013.
- G. E. Taccioli, T. M. Gottlieb, T. Blunt, A. Priestley, J. Demengeot, R. Mizuta, A. R. Lehmann, F. W. Alt, S. P. Jackson, et P. A. Jeggo. Ku80 : Product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. *Science*, 265(5177) :1442 – 1445, 1994.
- A. Tacheny, M. Dieu, T. Arnould, et P. Renard. Mass spectrometry-based identification of proteins interacting with nucleic acids. *J Proteomics*, 94 : 89 – 109, 2013.
- X. Tang, F. Yuan, et K. Guo. Repair of radiation damage of U2OS osteosarcoma cells is related to DNA-dependent protein kinase catalytic subunit (DNA-PKcs) activity. *Mol. Cell. Biochem.*, 390(1 - 2) :51 – 59, 2014.
- N. A. Tchurikov, D. M. Fedoseeva, D. V. Sosin, A. V. Snezhkina, N. V. Melnikova, A. V. Kudryavtseva, Y. V. Kravatsky, et O. V. Kretova. Hot spots of DNA double-strand breaks and genomic contacts of human rDNA units are involved in epigenetic regulation. *J Mol Cell Biol*, 2014.
- L. H. Thompson. Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells : The molecular choreography. *Mutat Res*, 751(2) :158 – 246, 2012.
- E. J. Tisdale. Glyceraldehyde-3-phosphate dehydrogenase is phosphorylated by protein kinase Ciota /lambda and plays a role in microtubule dynamics in the early secretory pathway. *J. Biol. Chem.*, 277(5) :3334 – 3341, 2002.
- N. Tomimatsu, B. Mukherjee, K. Deland, A. Kurimasa, E. Bolderson, K. K. Khanna, et S. Burma. Exo1 plays a major role in DNA end resection in humans and influences double-strand break repair and damage signaling decisions. *DNA Repair (Amst)*, 11(4) :441 – 448, 2012.
- J. K. Tong, C. A. Hassig, G. R. Schnitzler, R. E. Kingston, et S. L. Schreiber. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature*, 395(6705) :917 – 921, 1998.
- C. Tsai, V. Smider, B. J. Hwang, et G. Chu. Electrophoretic mobility shift assays for protein-DNA complexes involved in DNA repair. *Methods Mol Biol*, 920 :53 – 78, 2012.

- J. Ueda, M. Tachibana, T. Ikura, et Y. Shinkai. Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP. *J. Biol. Chem.*, 281(29) :20120 – 20128, 2006.
- N. Uematsu, E. Weterings, K. i. Yano, K. Morotomi-Yano, B. Jakob, G. Taucher-Scholz, P.-O. Mari, D. C. van Gent, B. P. C. Chen, et D. J. Chen. Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J. Cell Biol.*, 177(2) :219 – 229, 2007.
- R. D. Unwin. Quantification of proteins by iTRAQ. *Methods Mol Biol*, 658 : 205 – 215, 2010.
- B. C. Valdez, L. Perlaky, Y. Saijo, D. Henning, C. Zhu, R. K. Busch, W. W. Zhang, et H. Busch. A region of antisense RNA from human p120 cDNA with high homology to mouse p120 cDNA inhibits NIH 3T3 proliferation. *Cancer Res*, 52(20) :5681 – 5686, 1992.
- B. C. Valdez, D. Henning, R. B. So, J. Dixon, et M. J. Dixon. The Treacher Collins syndrome (TCOF1) gene product is involved in ribosomal DNA gene transcription by interacting with upstream binding factor. *Proc. Natl. Acad. Sci. USA*, 101(29) :10709 – 10714, 2004.
- H. van Attikum et S. M. Gasser. Crosstalk between histone modifications during the DNA damage response. *Trends Cell Biol*, 19(5) :207 – 217, 2009.
- A. J. van Vuuren, W. Vermeulen, L. Ma, G. Weeda, E. Appeldoorn, N. G. Jaspers, A. J. van der Eb, D. Bootsma, J. H. Hoeijmakers, et S. Humbert. Correction of xeroderma pigmentosum repair defect by basal transcription factor BTF2 (TFIIH). *EMBO J*, 13(7) :1645 – 1653, 1994.
- J. P. Venables, M. Ruggiu, et H. J. Cooke. The RNA-binding specificity of the mouse Dazl protein. *Nucleic Acids Res.*, 29(12) :2479 – 2483, 2001.
- A. S. Vetchinova, S. B. Akopov, I. P. Chernov, L. G. Nikolaev, et E. D. Sverdlov. Two-dimensional electrophoretic mobility shift assay : Identification and mapping of transcription factor CTCF target sequences within an FXYD5-COX7A1 region of human chromosome 19. *Anal. Biochem.*, 354(1) :85 – 93, 2006.
- M. Wakasugi, A. Kawashima, H. Morioka, S. Linn, A. Sancar, T. Mori, O. Nikaido, et T. Matsunaga. DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair. *J. Biol. Chem.*, 277(3) :1637 – 1640, 2002.

- J. R. Walker, R. A. Corpina, et J. Goldberg. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature*, 412(6847) :607 – 614, 2001.
- N. Walworth, S. Davey, et D. Beach. Fission yeast chk1 protein kinase links the rad checkpoint pathway to cdc2. *Nature*, 363(6427) :368 – 371, 1993.
- H. Wang, L. Zhai, J. Xu, H.-Y. Joo, S. Jackson, H. Erdjument-Bromage, P. Tempst, Y. Xiong, et Y. Zhang. Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Mol Cell*, 22(3) :383 – 394, 2006.
- W. Wang, L. Yin, L. Gonzalez-Malerva, S. Wang, X. Yu, S. Eaton, S. Zhang, H.-Y. Chen, J. LaBaer, et N. Tao. In situ drug-receptor binding kinetics in single cells : A quantitative label-free study of anti-tumor drug resistance. *Sci Rep*, 4 :6609, 2014.
- J. F. Ward. DNA damage produced by ionizing radiation in mammalian cells : Identities, mechanisms of formation, and reparability. *Prog Nucleic Acid Res Mol Biol*, 35 :95 – 125, 1988.
- J. F. Ward, J. W. Evans, C. L. Limoli, et P. M. Calabro-Jones. Radiation and hydrogen peroxide induced free radical damage to DNA. *Br J Cancer Suppl*, 8 :105 – 112, 1987.
- J. D. Watson et F. H. Crick. The structure of DNA. *Cold Spring Harb Symp Quant Biol*, 18 :123 – 131, 1953.
- D. M. Weinstock, E. Brunet, et M. Jasin. Formation of NHEJ-derived reciprocal chromosomal translocations does not require Ku70. *Nat Cell Biol*, 9(8) :978 – 981, 2007.
- R. B. West, M. Yaneva, et M. R. Lieber. Productive and nonproductive complexes of Ku and DNA-dependent protein kinase at DNA termini. *Mol Cell Biol*, 18(10) :5908 – 5920, 1998.
- B. J. Wilson, G. J. Bates, S. M. Nicol, D. J. Gregory, N. D. Perkins, et F. V. Fuller-Pace. The p68 and p72 DEAD box RNA helicases interact with HDAC1 and repress transcription in a promoter-specific manner. *BMC Mol Biol*, 5 :11, 2004.
- X. Xu et D. F. Stern. NFB1/MDC1 regulates ionizing radiation-induced focus formation by DNA checkpoint signaling and repair factors. *FASEB J*, 17(13) :1842 – 1848, 2003.

- H. Yajima, H. Fujisawa, N. I. Nakajima, H. Hirakawa, P. A. Jeggo, R. Okayasu, et A. Fujimori. The complexity of DNA double strand breaks is a critical factor enhancing end-resection. *DNA Repair (Amst)*, 12(11) :936 – 946, 2013.
- Q. Yan, R. Xu, L. Zhu, X. Cheng, Z. Wang, J. Manis, et M. A. Shipp. BAL1 and its partner E3 ligase, BBAP, link Poly(ADP-ribose) activation, ubiquitylation, and double-strand DNA repair independent of ATM, MDC1, and RNF8. *Mol Cell Biol*, 33(4) :845 – 857, 2013.
- J. Ye, C. Lenain, S. Bauwens, A. Rizzo, A. Saint-Leger, A. Poulet, D. Benarroch, F. Magdinier, J. Morere, S. Amiard, E. Verhoeyen, S. Britton, P. Calsou, B. Salles, A. Bizard, M. Nadal, E. Salvati, L. Sabatier, Y. Wu, A. Biroccio, A. Londono-Vallejo, M.-J. Giraud-Panis, et E. Gilson. TRF2 and apollo cooperate with topoisomerase 2alpha to protect human telomeres from replicative damage. *Cell*, 142(2) :230 – 242, 2010.
- E. Yonish-Rouach, D. Grunwald, S. Wilder, A. Kimchi, E. May, J. J. Lawrence, P. May, et M. Oren. P53-mediated cell death : Relationship to cell cycle control. *Mol Cell Biol*, 13(3) :1415 – 1423, 1993.
- X. Yu, S. Fu, M. Lai, R. Baer, et J. Chen. BRCA1 ubiquitinates its phosphorylation-dependent binding partner CtIP. *Genes Dev*, 20(13) : 1721 – 1726, 2006.
- O. Zagorodko, J. Spadavecchia, A. Yanguas, I. Larroulet, A. Pesquera, A. Zurutuza, R. Boukherroub, et S. Szunerits. Highly sensitive detection of DNA hybridization on commercialized graphene coated surface plasmon resonance interfaces. *Anal. Chem.*, 2014.
- J. Zhang, G. Gao, L. Chen, J. Li, X. Deng, Q. s. Zhao, et C. Huang. Hydrogen peroxide/ATR-Chk2 activation mediates p53 protein stabilization and anti-cancer activity of cheliensisin A in human cancer cells. *Oncotarget*, 5 (3) :841 – 852, 2014.
- Y. Zhang et M. Jasin. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nat Struct Mol Biol*, 18(1) :80 – 84, 2011.
- W.-M. Zhao et G. Fang. Anillin is a substrate of anaphase-promoting complex/cyclosome (APC/C) that controls spatial contractility of myosin during late cytokinesis. *J. Biol. Chem.*, 280(39) :33516 – 33524, 2005.
- D. Zheng, Y.-Y. Cho, A. T. Y. Lau, J. Zhang, W.-Y. Ma, A. M. Bode, et Z. Dong. Cyclin-dependent kinase 3-mediated activating transcription

- factor 1 phosphorylation enhances cell transformation. *Cancer Res*, 68 (18) :7650 – 7660, 2008.
- T. Zhou, J. W. Lee, H. Tatavarthi, J. R. Lupski, K. Valerie, et L. F. Povirk. Deficiency in 3'-phosphoglycolate processing in human cells with a hereditary mutation in tyrosyl-DNA phosphodiesterase (TDP1). *Nucleic Acids Res.*, 33(1) :289 – 297, 2005.
- C. Zhu et W. Jiang. Cell cycle-dependent translocation of PRC1 on the spindle by Kif4 is essential for midzone formation and cytokinesis. *Proc. Natl. Acad. Sci. USA*, 102(2) :343 – 348, 2005.

