



**HAL**  
open science

# Photomarquage d'affinité couplé à la spectrométrie de masse pour l'identification de protéines interagissant avec des modifications épigénétiques de l'ADN

Frédéric Thiebaut

► **To cite this version:**

Frédéric Thiebaut. Photomarquage d'affinité couplé à la spectrométrie de masse pour l'identification de protéines interagissant avec des modifications épigénétiques de l'ADN. Chimie organique. Université Pierre et Marie Curie - Paris VI, 2017. Français. NNT : 2017PA066490 . tel-01906598

**HAL Id: tel-01906598**

**<https://theses.hal.science/tel-01906598>**

Submitted on 27 Oct 2018

**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 Pierre et Marie Curie

ED 406 Chimie moléculaire

*Laboratoire des Biomolécules – UMR7203 – UPMC/ENS/CNRS*

*Structure et instabilité des génomes – UMR 7196 – MNHN/CNRS/INSERM*

## **Photomarquage d'affinité couplé à la spectrométrie de masse pour l'identification de protéines interagissant avec des modifications épigénétiques de l'ADN.**

Par Frédéric Thiebaut

Thèse de doctorat de Chimie

Dirigée par Dominique Guianvarc'h et Carole Saintomé

Présentée et soutenue publiquement le 25 octobre 2017

Devant un jury composé de :

Christine Enjalbal, Professeur, Rapporteur

Marie Lopez, Chargée de recherche, Rapporteur

Marcel Hollenstein, Chargé de recherche, Examinateur

Christine Ménager, Professeur, Examinatrice

Dominique Guianvarc'h, Professeur, Directrice de thèse

Carole Saintomé, Maître de conférences, Directrice de thèse

Emmanuelle Sachon, Maître de conférences, Encadrante de thèse

## Table des matières

Liste des Abréviations .....	8
Remerciements .....	10
Introduction .....	12
1 Modifications épigénétiques de l'ADN.....	13
Epigénétique : définition .....	15
Méthylation de l'ADN .....	15
1.1 1.2.1 Méthylation de novo par les DNMT3 .....	17
1.2 1.2.2 Maintien de la méthylation par la DNMT1 .....	17
1.2.2.1 Une première proposition de mécanisme de maintenance .....	17
1.2.2.2 Maintien de la méthylation de l'ADN : nouveau modèle. ....	18
1.2.2.3 UHRF1 : un acteur épigénétique majeur.....	19
1.3 Régulation de la méthylation de l'ADN.....	20
1.3.1 La déméthylation passive de l'ADN .....	21
1.3.2 La déméthylation active de l'ADN : les premières hypothèses. ....	22
1.3.2.1 La déméthylation enzymatique .....	22
1.4 1.3.2.2 Les mécanismes de réparation de l'ADN.....	22
Les nouveaux acteurs de l'épigénétique et leurs rôles dans la régulation de la méthylation .....	26
1.4.1 L'oxydation des 5mC comme étape cruciale de la déméthylation.....	26
1.4.1.1 Les enzymes TET et l'oxydation de la 5mC .....	27
1.4.1.2 La famille enzymatique TET.....	27
1.4.1.3 Réaction catalysée par les enzymes TET .....	28
1.4.1.4 Substrats des enzymes TET : nouvelles bases épigénétiques .....	29
1.4.2 Le rôle des bases oxydées dans la déméthylation de l'ADN. ....	29
1.4.2.1 La dilution passive des bases oxydées (AM-PD).....	29

1.4.2.2	Une élimination enzymatique directe des bases oxydées.....	30
1.4.2.3	Réparation des bases oxydées par le BER (AM-AR) .....	32
	Bilan des mécanismes généraux de la méthylation de l'ADN .....	34
2	Méthodes de caractérisation de protéines affines pour les nouvelles bases épigénétiques dérivées de la 5mC .....	36
1.5	EMSA (Electrophoretic mobility shift assay) ou gel retard .....	36
	2.1.1 Principe général de l'EMSA .....	36
2.1	2.1.2 Application de l'EMSA.....	37
	L'analyse protéomique pour la caractérisation de protéines partenaires .....	40
2.2	2.2.1 Le « pull-down » ou la purification d'affinité : principe.....	40
	2.2.2 Le « pull-down » ou la purification d'affinité : applications .....	41
	2.2.2.1 Stratégie « Label-free » .....	42
	2.2.2.2 Stratégie avec marquage.....	44
2.3	Bilan des études de protéomiques permettant l'identification des protéines partenaires des dérivées oxydés de la 5mC.....	47
3	Méthodes de pontage entre deux partenaires d'interaction. ....	49
3.1		
3.2	Le Formaldéhyde : un agent de pontage chimique.....	49
	Les agents de pontage photo-chimiques.....	51
3.2.1	L'ADN comme agent de photomarquage .....	51
3.2.2	Les agents de photomarquage classiques .....	52
	3.2.2.1 Les azotures d'aryle .....	53
	3.2.2.2 Les diazirines.....	55
	3.2.2.3 La benzophénone.....	57
3.3	3.2.3 Les agents de photomarquage analogues de nucléobases .....	59
	3.2.3.1 Les dérivés halogénés.....	59
	3.2.3.2 Les dérivés de nucléobases thiolés.....	61
	Incorporation des agents photoactivables dans des sondes oligonucléotidiques. 65	
3.3.1	La synthèse enzymatique .....	66

3.3.2	La synthèse chimique .....	67
3.3.2.1	Stratégie de fonctionnalisation pré-synthétique .....	68
3.3.2.2	Stratégie de fonctionnalisation post-synthétique.....	68
	Le PAL (photoaffinity labeling) couplé à la spectrométrie de masse. ....	70
	Projet de thèse .....	74
	Partie résultats .....	76
3.4	4 Conception et évaluation de différentes sondes oligonucléotidiques photoactivables et mise au point des conditions de photomarquage .....	77
	Synthèse de sondes ODN photoactivables .....	77
4.1	4.1.1 Conception des sondes .....	77
	4.1.2 Synthèse des séquences oligonucléotidiques modifiées.....	78
	4.1.2.1 La synthèse d'oligonucléotides sur support solide.....	79
	4.1.2.2 Incorporation des agents photoactivables benzophénone et diazirine.....	81
	4.1.3 Récapitulatif des oligonucléotides utilisés .....	83
4.2	Comparaison et évaluation des sondes ODN photoactivables simple brin .....	85
	4.2.1 La hRPA comme protéine modèle pour les sondes ADN simple brin.....	85
	4.2.1.1 Evaluation de l'affinité de hRPA pour les différents ODN par EMSA ...	86
	4.2.1.2 Évaluation de l'efficacité de photomarquage des différents ODN avec hRPA	87
	4.2.1.3 Identification de hRPA dans les complexes covalents formés par spectrométrie de masse.....	90
	4.2.2 Évaluation de l'efficacité de photomarquage des différentes sondes ODN dans un mélange protéique .....	94
	4.2.2.1 Identification de hRPA dans les complexes covalents formés.....	96
	4.2.3 Photomarquage avec la sonde ODN-4 <sup>S</sup> T dans un extrait cellulaire eucaryote	99
	4.2.3.1 Identification de hRPA dans les complexes covalents formés.....	101
	4.2.3.2 Détermination du caractère sélectif de nos sondes modifiées simple brin	

4.2.4	Conclusion sur la comparaison et l'évaluation des sondes ODN photoactivables simple brin.....	104
	Comparaison et évaluation des sondes ODN photoactivables double brin.....	105
4.3.1	Etudes des complexes duplex ODN-DNMT3a .....	106
4.3.2	Etude des complexes duplex ODN-histone H3 .....	107
4.3	4.3.2.1 Photomarquage de l'histone H3 par les différentes sondes.....	107
	4.3.2.2 Identification par spectrométrie de masse des complexes ODN-H3.....	108
4.3.3	Photomarquage avec la sonde duplexe ODN-4 <sup>S</sup> T dans un mélange protéique complexe. ....	115
4.3.4	Conclusion sur la comparaison et l'évaluation des sondes photoactivables double brin.....	116
5	Caractérisation au niveau moléculaire de complexes ODN-protéines par spectrométrie de masse MALDI-TOF.....	118
5.1	Avant-propos .....	118
5.2	Etude des acides nucléiques par spectrométrie de masse MALDI-TOF.....	121
5.3	Mise au point des expériences MALDI pour l'analyse des oligonucléotides avec la matrice 3-HPA .....	122
	5.3.1 Optimisation des dépôts : mise au point du ratio matrice : échantillon .....	122
	5.3.2 Optimisation des dépôts : dessalage des oligonucléotides .....	123
	5.3.2.1 Le citrate de diammonium comme additif dans les dépôts MALDI .....	123
	5.3.2.2 Utilisation de billes échangeuses de cations : résine Dowex .....	124
	5.3.3 Optimisation des dépôts : mise au point d'un protocole de cristallisation de la matrice 3-HPA en présence de citrate de diammonium.....	125
5.4	5.3.4 Optimisation des paramètres d'acquisition du spectromètre de masse MALDI-TOF (Voyager-DE Pro, Applied Biosystems) .....	127
	Analyse de complexes ADN-protéines .....	130
5.4.1	Identification des complexes ODN-protéines modèles .....	130
5.4.2	Etude d'un complexe modèle : ADN-peptide (ADN-ARTK).....	132
	5.4.2.1 Structure du peptide ARTK.....	132

5.4.2.2	Test de la liaison ARTK-ADN : identification de la formation d'un complexe ARTK-ADN sur gel.....	133
5.4.3	Identification du complexe ODN-ARTK.....	134
	Caractérisation de la liaison ODN 4 <sup>ST</sup> -ARTK.....	136
5.5.1	Le modèle ODN 4 <sup>ST</sup> -ARTK .....	136
5.5.2	Les modèles ODN 4 <sup>ST</sup> -acides aminés.....	144
5.5	5.5.3 Photochimie de la 4 <sup>ST</sup> : proposition de mécanisme vis-à-vis d'un peptide ...	146
	Conclusion de l'étude méthodologique de complexes photomarqués .....	148
6	Identification de protéines de liaisons aux modifications épigénétiques de l'ADN en présence des sondes ODN. ....	150
	Essais préliminaires.....	150
6.1	6.1.1 Essai préliminaires sur un extrait protéique de cellules HeLa .....	150
	6.1.2 Essai préliminaires sur un extrait protéique nucléaire de cerveau de rat .....	152
	6.1.3 Bilan des essais préliminaires .....	153
6.2	Conception de sondes ADN « boucles » photoactivables.....	155
	6.2.1 Un système de pull-down « boucle » .....	155
	6.2.2 Un système de pull-down « boucle » photoactivable.....	157
	6.2.3 Préparation des sondes ADN.....	159
	6.2.3.1 Choix de la séquence d'ADN.....	159
6.3	6.2.3.2 Synthèse des déterminants d'affinité.....	160
	6.2.3.3 Préparation des phases chromatographiques « boucle » .....	163
	Mise en place de l'analyse protéomique des pull-down « boucle ».....	164
	6.3.1 Mise en place des conditions d'élution .....	164
	6.3.2 Les différentes phases chromatographiques utilisées.....	167
	6.3.3 Description de la méthode de quantification « label free » des protéines par nanoLC ESI-MS/MS utilisée .....	168
	6.3.4 Validation des phases chromatographiques pour l'enrichissement de protéines affines pour l'ADN.....	170

	L'analyse protéomique des pull-down « boucle » .....	171
6.4.1	Méthode employée pour l'analyse des résultats obtenus .....	171
6.4.2	Le pull-down classique pour identifier les protéines affines de la 5mC et de la 5hmC	173
6.4		
6.4.2.1	Enrichissement des protéines affines pour la 5mC .....	173
6.4.2.2	Enrichissement des protéines affines pour la 5hmC .....	175
6.4.2.3	Identification des protéines sélectives de la 5mC et de la 5hmC .....	176
6.4.3	Le pull-down couplé au photomarquage pour identifier les protéines affines de la 5mC et de la 5hmC .....	178
6.4.3.1	Enrichissement de protéines affines pour l'ADN, non spécifiques pour la 4 <sup>ST</sup>	178
6.4.3.2	L'apport du photomarquage aux résultats de l'analyse protéomique du pull-down classique .....	179
6.5	Comparaison des résultats de l'analyse protéomique des pull-down classique et avec photomarquage .....	182
6.6	Conclusion de l'analyse protéomique .....	183
	Conclusion générale et perspective .....	188
	Annexe .....	192
	Matériels et méthodes.....	195
	Bibliographie.....	209
	Table des illustrations.....	242
	Table des tableaux .....	248

## Liste des Abréviations

3-HPA	Acide 3-hydroxypicolinique
4 <sup>S</sup> T	4-thiothymidine
5caC	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxyméthylcytosine
5IU	5-iodouridine
5mC	5-méthylcytosine
AdoMet	S-Adénosyl-L-Méthionine
ADN	Acide désoxyribonucléique
AM	Modification active
AM-AR	Modification active suivi d'une réparation active de l'ADN
AM-PD	Modification active suivi d'une dilution passive
AP	Site abasique de l'ADN
AR	Réparation active de l'ADN
ARN	Acide ribonucléique
ARTK	Peptide correspondant aux 16 premiers acides aminés du domaine N-terminal de l'histone H3
BER	Réparation par excision de base
Bzp	Benzophénone
CpG	Dinucléotide CG
Dia	Diazirine
DNMT	<i>DNA methyltransferase</i> , Méthyltransférase de l'ADN

EMSA	<i>Electrophoretic mobility shift assay</i> , gel retard d'électrophorèse
H3	Histone H3
HCCA	Hydroxy- $\alpha$ -cyano-cinnamique acide
LC-MS / MS.	Chromatographie en phase liquide couplée à la spectrométrie de masse en tandem
LP-BER	<i>Long Patch BER</i> , BER long
NEIL	<i>Endonucléase VIII-like</i>
NER	<i>Nucleotide excision repair</i> , réparation par excision de nucléotides
MALDI-TOF	<i>Matrix assisted laser desorption ionization-Time of flight</i>
MBD	<i>Methyl-CpG binding domain protein</i>
ODN	Oligodésoxynucléotide
PCR	Amplification en chaîne par polymérase
PD	Dilution passive
(h)RPA	Protéine de replication A
SAM	S-Adénosyl-L-Méthionine
SDS	Dodécyl sulfate sodium
SDS-PAGE	Electrophorèse dénaturante en gel de polyacrylamide en presence de SDS
SILAC	<i>Stable isotope labelling by amino acids in cell culture</i>
SP-BER	<i>Short Patch BER</i> , BER court
TDG	<i>Thymine DNA glycosylase</i>
TET	<i>Ten Eleven Translocation</i>
T <sub>m</sub>	Température de fusion
UHRF	<i>Ubiquitin-like plant Homeodomain and RING finger domain-containing protein</i>
UV	Ultraviolet

## Remerciements

Je remercie vivement le professeur Christine Enjalbal et le docteur Marie Lopez d'avoir accepté d'être les rapporteuses de ce travail de thèse. Je remercie aussi le docteur Marcel Hollenstein ainsi que le professeur Christine Ménager d'avoir évalué mon travail en tant qu'examinateur.

Je tiens à remercier le programme doctoral « Interfaces pour le vivant » de l'université Pierre et Marie Curie sans lequel je n'aurais pu effectuer cette thèse.

J'exprime ma reconnaissance au Dr Sandrine Sagan pour m'avoir permis d'effectuer ma thèse au sein du laboratoire des biomolécules de l'université Pierre et Marie Curie. Je tiens aussi à remercier le Pr Jean-François Riou qui m'a accueilli dans le laboratoire Structure et Instabilité des Génomes du muséum national d'histoire naturel, pendant ces trois années de thèse.

Je veux exprimer ici toute ma gratitude à mes directrices de thèse, officielles et officieuses : le Docteur Emmanuelle Sachon, dont le dynamisme et l'engagement ont été pour moi un exemple, le Dr Carole Saintomé, pour son franc-parler qui a pu, à chaque fois, me faire avancer sur mon projet et enfin le Pr Dominique Guianvarc'h, dont la gentillesse et la bienveillance m'ont constamment motivé. Je ne verrai plus le ski de fond, ni le site de Gif-sur-Yvettes de la même manière!

Je remercie chaleureusement le Dr Rachid Benhida et son équipe pour m'avoir accueilli dans leur laboratoire et m'avoir enseigné la synthèse d'oligonucléotides sur support solide.

Je remercie aussi Vincent Guérineau de l'ICSN dont les conseils ont été très précieux! Je remercie aussi le Dr Damarys Loew, Bérangère Lombard et Guillaume Arras de la plateforme de spectrométrie de masse de l'Institut Curie : sans vous, la dernière partie de mon travail n'aurait pas pu être possible.

Je veux remercier le Dr Gérard Bolbach, le Dr Lucrèce Matheron de la plateforme de spectrométrie de masse et de protéomique de l'IBPS, UPMC pour nos discussions et leurs conseils. Je tiens à remercier Gilles Clodic pour sa sympathie et son soutien dans les moments où je ne voyais « rien ».

Je vous remercie tout particulièrement Guy Assou, Tania Almeida Santos et Sabine Rajkumar dont le travail a permis de faire grandement avancer ma thèse.

Merci également tous les membres du laboratoire Structure et Instabilité des Génomes, les membres de l'équipe du « débrief » pour nos discussions sobrement scientifiques, Emmanuelle pour sa gentillesse (sans toi, pas de RPA), à ma partenaire de chimie et de fête de la science Pauline, à mes pauvres partenaires de bureau qui ont su me subir plus que m'apprécier. Et à tous les autres que j'oublie de ce côté de la rue Cuvier.

Merci aux membres du laboratoire des biomolécules, tout particulièrement les membres de l'équipe 2 et encore plus particulièrement les doctorants de ce laboratoire, sans vous rien n'aurait été pareil ! Un grand merci à Mehdi, Pierre, Anaïs x2, Lucille, Sébastien, Mathilde, Aude, Emilie, Claudia, Damien, Benjamin, Simon, Didier et Leïla !

Je remercie mes parents qui m'ont soutenu pendant ses huit années d'études. Merci Maman. Merci Papa. Merci Lulu. Je tiens à remercier mes amis dont le soutien, parfois douteux, a été des plus précieux : Steve, Flavien, Fabien, Martin, Christophe, Louise, Julien et tous les autres !

Enfin, je te remercie, toi, qui m'as découvert dans mes heures les plus sombres (seulement éclairé par le rétroéclairage d'un ordinateur), qui m'as soutenu tout de même, qui m'as fait sourire et m'as fait tenir. Je remercie mes pieds de m'avoir mené d'une montagne à un mont, qu'importe la couleur de celui-ci. Merci Marie.

Je vous remercie tous, j'ai pris plaisir à partager les moments que nous avons vécu.

## **Introduction**

# 1 Modifications épigénétiques de l'ADN

La diversité du monde du vivant provient essentiellement d'une expression différenciée de l'information génétique. Cette dernière est contenue dans l'acide désoxyribonucléique (ADN). Depuis la découverte de la structure de l'ADN par Watson et Crick dans les années 1950<sup>1</sup>, cette molécule n'a cessé d'être étudiée notamment pour sa fonctionnalité en tentant de répondre à cette question : Comment un phénotype est-il exprimé à partir d'un génotype ? Le code génétique repose sur la lecture des quatre bases de l'ADN : l'adénine (A), la thymine (T), la cytosine (C) et la guanine (G). La séquence formée par l'enchaînement de ces bases est lue puis est transcrite en acide ribonucléique (ARN), avant d'être traduite en protéine. On ne dénombre pas moins d'une centaine de modifications de l'ARN que ce soit au niveau des ARN de transfert qu'au niveau des ARN messagers, qui permettent, entre autre, une expression des gènes différenciés<sup>2</sup>. Ces modifications de l'ARN sont répertoriées dans plusieurs bases de données comme RMBase (RNA Modification dataBase)<sup>3</sup>.

L'ADN n'a pas une composition aussi diversifiée que l'ARN mais A, T, C et G ne sont pas les seules bases de l'ADN présentes au sein des organismes<sup>4,5</sup>. Un nombre important de modifications a été répertorié comme on peut le voir sur la Figure 1, issue de la base de données DNAmo<sup>5</sup>.

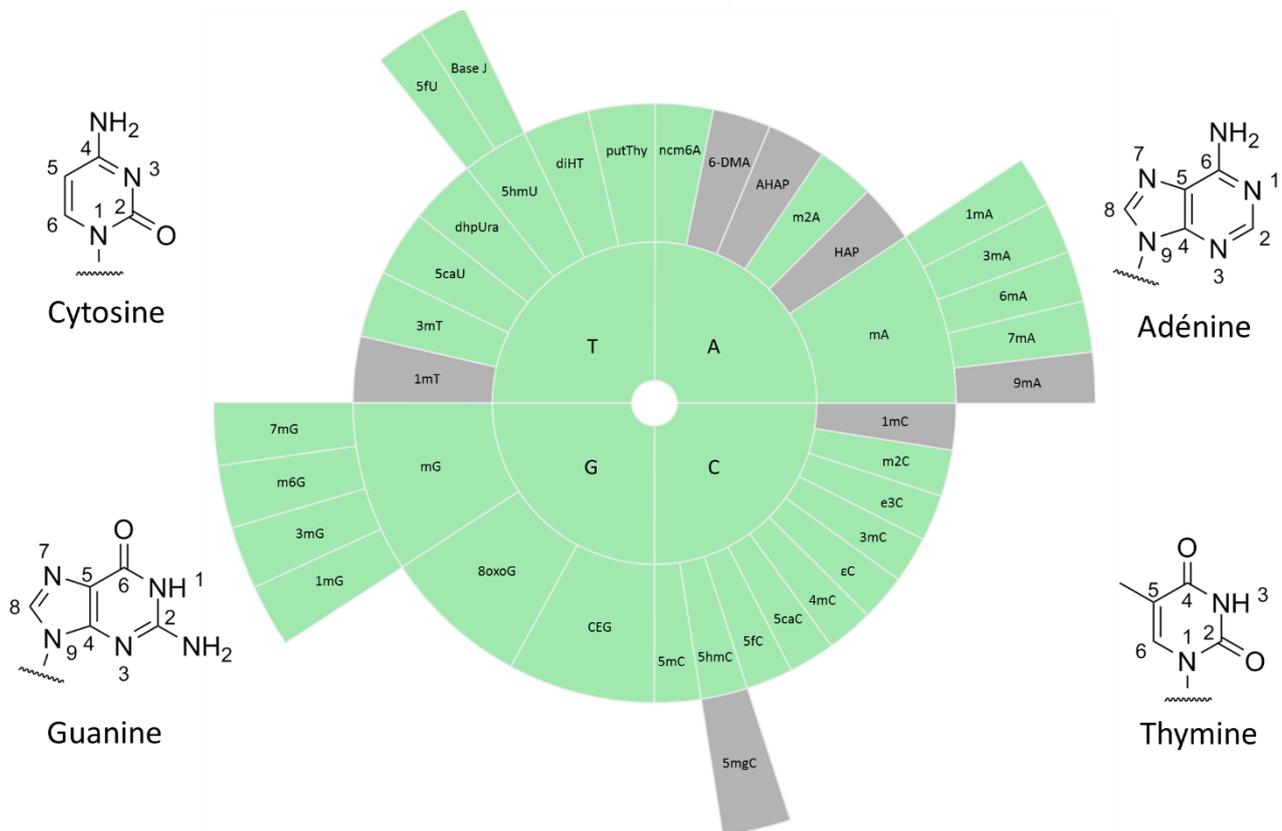


Figure 1 : Présentation des modifications covalentes de l'ADN. En gris sont représentées les bases de l'ADN provenant d'une synthèse chimique (d'après DNAmo), on ne les retrouve pas naturellement. Les modifications de l'ADN sont ici représentées sous leurs formes abrégées. Graphique tiré de la base de donnée DNAmo<sup>5</sup>

Ces bases modifiées peuvent provenir :

- De dégâts oxydatifs<sup>6</sup>, chimiques ou encore photochimiques<sup>7</sup>. On peut citer la 8-oxoguanine (8oxoG) et la 5-hydroxyméthyluracile (5hmU)<sup>6</sup> ou encore la 6-méthylguanine (m6G) et la 3-méthyladénine (3mA). Ces dégâts peuvent induire des mutations et causer la mort cellulaire<sup>8,9,10</sup>. Par exemple, la m6G induit des mésappariements lors de la réplication de l'ADN provoquant des mutations de l'ADN<sup>11,12</sup>.
  
- De modifications introduites par des enzymes au sein du biopolymère ou sur le nucléotide avant son incorporation dans l'ADN par une polymérase. Ces modifications ont un rôle biologique :
  - Elles peuvent jouer un rôle au sein des mécanismes de défense entre organismes. C'est le cas de la *N*<sup>6</sup>-méthyladénine (6mA) qui fait partie du système de restriction-modification chez les bactéries<sup>13</sup>. Ce système permet de distinguer l'ADN de la cellule hôte, de l'ADN d'un organisme étranger. Plus précisément, la présence de 6mA dans la séquence d'ADN de la cellule hôte empêche la digestion enzymatique de son génome par une enzyme de restriction, sensible à la méthylation de l'ADN<sup>14</sup>. L'ADN étranger non-méthylé est dégradé lors de son entrée dans la cellule hôte, garantissant ainsi la préservation du code génétique de l'organisme.
  - Elles peuvent jouer un rôle lors de la régulation de l'expression des gènes. C'est le cas de la 5-méthylcytosine (5mC), aussi nommée la 5<sup>ème</sup> base de l'ADN. Cette base fait partie des modifications épigénétiques de l'ADN.

Certaines de ces modifications des bases de l'ADN sont bien connues et leur rôle biologique bien caractérisé, d'autres identifiées récemment sont actuellement sujettes à de nombreuses études afin de mieux comprendre leur fonction.

## Epigénétique : définition

1.1 La modification de l'ADN la plus étudiée est la 5-méthylcytosine (5mC) qui est impliquée dans la régulation épigénétique. Le concept de l'épigénétique provient d'une interrogation simple : comment expliquer une si grande variation phénotypique entre cellules d'un même individu, d'une même espèce, si ceux-ci possèdent la même information génétique ? Conrad Waddington fut le premier à tenter de répondre à cette question en introduisant le terme « épigénétique » dans les années 40. Il définit celui-ci comme étant la branche de la biologie qui étudie l'interaction entre les gènes et leurs produits, c'est-à-dire les processus permettant d'accéder à un phénotype donné (« the branch of biology which studies the causal interaction between genes and their products, which bring the phenotype into being »<sup>15</sup>). Cependant, la compréhension des mécanismes impliqués dans l'expression des génomes n'a cessé d'évoluer depuis lors, amenant l'émergence d'une définition consensus de l'épigénétique comme étant l'ensemble des changements des phénotypes héréditaires ne provenant pas d'une modification directe de la séquence d'ADN (« stably heritable phenotypes resulting from changes in chromosome without changes in gene sequence »<sup>16</sup>).

La régulation de l'expression d'un gène provient de trois processus épigénétiques : les ARN non codants<sup>17</sup>, les modifications des histones (protéines qui régulent l'état de compaction de la chromatine, son accessibilité) et la méthylation de l'ADN.

1.2

### Méthylation de l'ADN

La méthylation de la cytosine en position C-5 est une modification épigénétique importante et très conservée de l'ADN chez différents organismes<sup>18,19</sup> (plantes, champignons, animaux). En effet, la présence des 5mC dans les régions promotrices modifie la transcription des gènes associés. Le groupement méthyle empêche la liaison de certains facteurs de transcription sur l'ADN et permet aussi de recruter des protéines spécifiques des 5mC. Ces protéines induisent une compaction de l'ADN, réduisant ainsi la transcription. La 5mC est ainsi considérée comme une modification épigénétique répressive. Elle intervient dans différents processus de régulation de l'expression des gènes -par exemple la différenciation cellulaire, l'inactivation d'un des deux chromosomes X chez les femelles des mammifères ou encore la stabilité chromosomique<sup>20</sup>. On retrouve aussi des variations dans les profils de méthylation de l'ADN dans certaines pathologies parmi lesquelles certains cancers<sup>21,22,23</sup>. Chez les mammifères, la méthylation des cytosines se fait principalement sur un motif formé de deux nucléotides,

cytosine et guanine, appelé CpG. Les motifs CpG ne sont pas uniformément répartis dans le génome humain. Ces dinucléotides sont regroupés en îlots CpG concentrés dans les régions promotrices des gènes, les transposons ou les séquences répétées<sup>24</sup>. La présence de groupements méthyles permet aussi de recruter des partenaires protéiques impliqués dans le remodelage de la chromatine *via* la modification des histones.

La méthylation des cytosines est catalysée par une famille d'enzymes : les ADN méthyltransférases (DNA methyltransferase, DNMT). On en dénombre trois ayant une activité catalytique : DNMT1, DNMT3A et DNMT3B et une ayant une activité régulatrice DNMT3L. Les DNMT fonctionnent à l'aide d'un cofacteur appelé SAM ou AdoMet (pour *S*-Adénosyl-L-méthionine), qui leur fournit un groupement méthyle. Le mécanisme de ces enzymes est bien caractérisé<sup>25</sup> et est représenté Figure 2<sup>25</sup>.

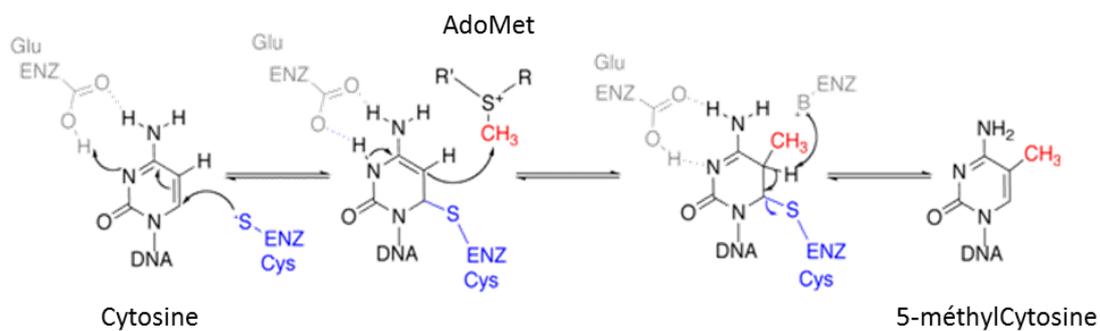


Figure 2: Mécanisme enzymatique de la méthylation de l'ADN par les DNMT (ENZ=enzyme).

Dans un premier temps, la cytosine bascule de la double hélice d'ADN vers le site catalytique de l'enzyme (base-flipping en anglais), ce qui permet à une cystéine catalytique de réaliser une attaque nucléophile sur la position 6 du cycle pyrimidique. La position 5 est alors activée (forme énamine), ce qui va lui permettre d'attaquer le groupement méthyle du SAM (en rouge sur la Figure 2). Au cours d'une dernière étape, l'élimination du proton en 5 entraîne la rupture de la liaison entre l'enzyme et la cytosine pour former la cytosine méthylée<sup>25</sup>.

Chez les eucaryotes, les DNMT peuvent être classées en deux groupes basés sur leur fonction : les DNMT3 assurent la méthylation *de novo* tandis que la DNMT1 assure le maintien de la méthylation.

### 1.2.1 Méthylation de novo par les DNMT3

La DNMT3A et la DNMT3B présentent une affinité spécifique pour les sites CpG non-méthylés. C'est pourquoi on les appelle enzymes de méthylation *de novo*. Leur rôle est schématiquement présenté sur la partie gauche de la Figure 3. Ces enzymes ont aussi la capacité de méthyler de l'ADN hémiméthylé, ainsi que de pouvoir méthyler des cytosines qui ne font pas partie d'îlots CpG<sup>26</sup>. Si la protéine DNMT3B possède une grande similarité de séquence et de fonction avec la DNMT3A, ces deux enzymes ont cependant un rôle biologique différent. La DNMT3B est responsable de la méthylation des régions centromériques. L'activité de la DNMT3A est nécessaire pendant l'embryogenèse et le développement des cellules germinales. La protéine DNMT3L ne possède aucune activité catalytique, elle a pour rôle de stimuler la réaction de méthylation *de novo* effectuée par les DNMT3 A et B<sup>27</sup> en favorisant l'interaction de ces enzymes avec l'ADN et la SAM<sup>28</sup>, améliorant ainsi leur efficacité.

### 1.2.2 Maintien de la méthylation par la DNMT1

#### 1.2.2.1 Une première proposition de mécanisme de maintenance

L'enzyme DNMT1 assure le maintien de la méthylation de l'ADN. Au cours des divisions cellulaires, l'ADN (méthylé sur chaque brin) est répliqué ce qui entraîne la formation d'ADN hémiméthylé, c'est-à-dire la présence de CpG méthylés sur seulement un des brins de la double hélice nouvellement formée. La DNMT1 reconnaît de façon préférentielle les CpGs hémiméthylés, lui permettant ainsi de restaurer le profil de méthylation de l'ADN (Figure 3).

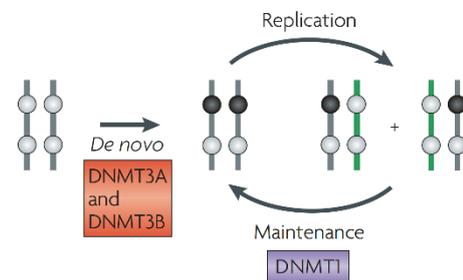


Figure 3: Schéma de la méthylation de novo et du maintien de la méthylation par les DNMT1, 3A et 3B d'après Jones et Liang<sup>29</sup>. Les boules grises représentent des motifs CpG non méthylés, les noires représentent des motifs CpG méthylés.

Ce modèle d'acquisition et de maintien de la méthylation des îlots CpG, couramment utilisé au début des années 2000, se base sur les hypothèses émises par Riggs<sup>29</sup> et Holliday et Pugh<sup>30</sup>. Celles-ci se résument assez simplement : les DNMT3 construisent le profil de méthylation et celui-ci est maintenu au cours des divisions cellulaires par l'enzyme DNMT1 seule. Cependant, ce modèle présente des lacunes. En effet, il ne propose aucun mécanisme correctif : comment la cellule peut être sûre de l'exacte fidélité du maintien de certains motifs de méthylation ? Il ne propose pas d'explication quant à la perte graduelle de positions méthylées observée au cours

de la division des cellules souches embryonnaires qui ne possèdent pas les enzymes DNMT3<sup>31,32</sup>. Il n'explique pas non plus l'absence de méthylation sur certains îlots CpG<sup>33</sup>.

### 1.2.2.2 Maintien de la méthylation de l'ADN : nouveau modèle.

Un nouveau modèle expliquant le maintien de la méthylation de l'ADN a été formulé en 2009 par Jones et Liang<sup>33</sup>.

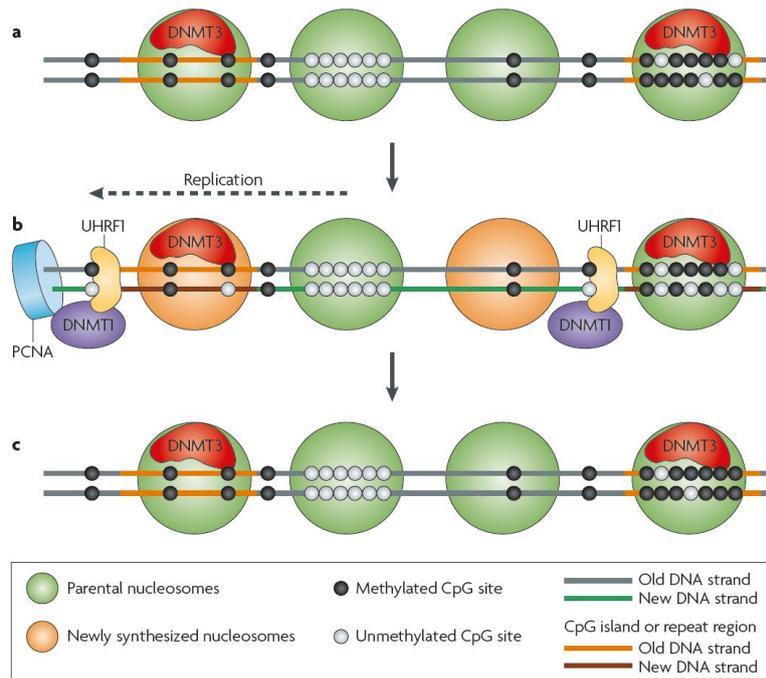


Figure 4 : Schéma représentant le nouveau modèle de maintien de la méthylation de l'ADN formulé en 2009 par Jones et Liang<sup>29</sup>. a) Représentation de l'ADN après la méthylation de novo de l'ADN, les DNMT3 étant toujours fixées sur l'ADN au sein du nucléosome. b) Mécanisme de maintien de la méthylation comprenant PCNA, UHRF1 et DNMT1 sur de l'ADN hémiméthylé obtenu après réplication de l'ADN. c) Étape correctrice de méthylation des CpG oubliés lors de l'étape de maintien de la méthylation.

Ce modèle (Figure 4) prend en compte plusieurs découvertes/paramètres nouveaux qui seront décrit ultérieurement. La première étape (Figure 4a) correspond à une étape de post-méthylation *de novo* de l'ADN par les enzymes DNMT3, incluses au sein d'un nucléosome qui contient des îlots CpG méthylés<sup>34</sup>. Au cours de la réplication de l'ADN (Figure 4b, partie gauche), DNMT1 est recrutée par deux enzymes, PCNA (Proliferating Cell Nuclear Antigen) et UHRF1 (Ubiquitin-like protein Homeodomain and RING finger domain-containing protein 1). PCNA possède des domaines d'interaction avec DNMT1 lui permettant de recruter cette dernière au niveau des fourches de réplication<sup>35</sup> – lieu de localisation de PCNA. UHRF1 (décrite dans le paragraphe suivant) qui présente une affinité pour l'ADN hémiméthylé et donc néosynthétisé, permet le recrutement de DNMT1 au sein du complexe de réplication. Le maintien de la méthylation est donc assuré directement au cours de la réplication. La dernière étape de ce

modèle est une étape correctrice (Figure 4b, partie droite et Figure 4c). Les différents sites CpG oubliés au cours du processus de réplication, vont être méthylés soit *de novo* par les DNMT3, s'ils sont contenus au sein de nucléosomes, soit par le complexe UHRF1-DNMT1 s'ils se trouvent hors des îlots CpG, UHRF1 assurant la reconnaissance de l'ADN hémiméthylé à corriger.

Deux nouveaux acteurs épigénétiques semblent se dégager de ce modèle : PCNA et UHRF1. La capacité de PCNA à recruter DNMT1 au niveau des fourches de réplication attire toujours l'attention des épigénéticiens bien que l'absence de PCNA n'induit pas de changement sur la méthylation de l'ADN<sup>36</sup>. Cela n'est pas le cas d'UHRF1 qui joue un rôle essentiel dans le maintien de la méthylation, puisque l'absence de cette protéine diminue d'environ 80 % le taux de méthylation<sup>37</sup>. UHRF1 constitue ainsi un acteur épigénétique majeur.

### 1.2.2.3 UHRF1 : un acteur épigénétique majeur

Le rôle crucial d'UHRF1 dans le maintien de la méthylation a été décrit dès 2007 par Sharif<sup>37</sup> et Bostick<sup>38</sup>. Il a aussi été montré *in vitro* qu'UHRF1 pouvait augmenter l'activité de la DNMT1 par un facteur 5<sup>39</sup>. Cette protéine est composée de 5 domaines structuraux (Figure 5) : un domaine *N*-terminal UBL (UBiquitin-Like), un domaine TTD (Tandem Tudor Domain), un domaine PHD (Plant HomeoDomain, PHD), un domaine SRA (SET and RING Associated) et enfin un domaine *C*-terminal RING (Really Interesting New Gene)<sup>40</sup>.

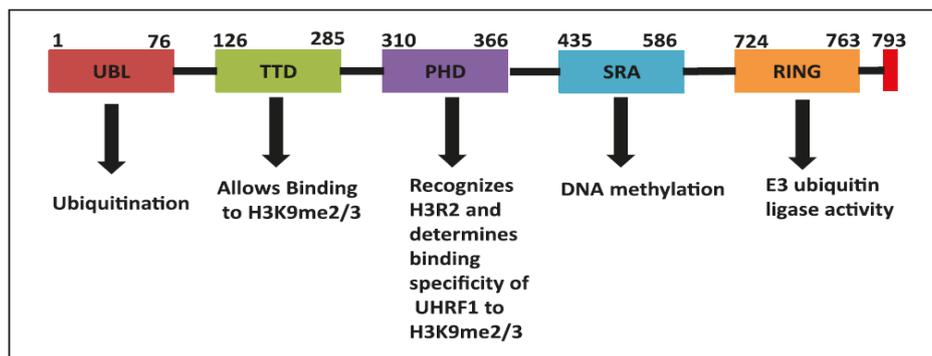


Figure 5: Les différents domaines de la protéine UHRF1 et leurs activités. D'après Sharif et al. <sup>37</sup>

Chacun de ces domaines possède une activité spécifique. Le domaine TTD permet à la protéine de se lier à la 9<sup>ème</sup> lysine de l'histone H3 quand celle-ci est di ou tri-méthylée (H3K9me2/3)<sup>41</sup>. Le domaine PHD permet de déterminer la spécificité de la liaison d'UHRF1 aux histones grâce à la reconnaissance d'une arginine non modifiée de l'histone H3 (H3R2). L'action du TTD et du PHD est coopérative<sup>42</sup> et permet de lire efficacement la modification H3K9me2/3 ce qui

conduit au réarrangement de la protéine UHRF1. Le domaine SRA a la particularité de reconnaître de façon préférentielle l'ADN hémiméthylé. La reconnaissance des cytosines méthylées s'effectue par le basculement de la base modifiée hors de la double hélice vers la poche d'interaction. Le rôle de ce basculement n'est pas encore bien compris<sup>43</sup>. Ces trois domaines assurent la fixation d'UHRF1 sur l'ADN hémiméthylé. Le dernier domaine d'intérêt est le domaine RING, qui dirige l'activité ubiquitine ligase de l'enzyme. Le recrutement de DNMT1 par UHRF1 peut s'effectuer comme schématisé sur la Figure 6.

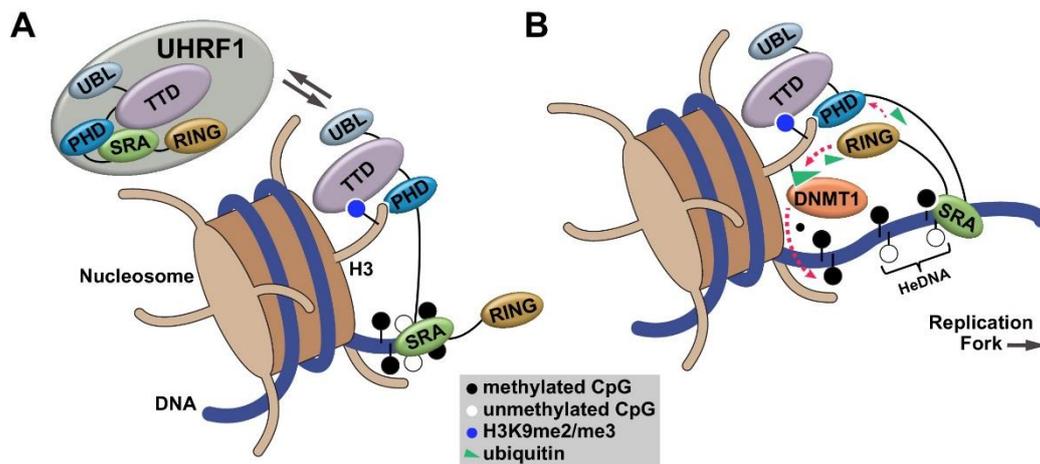


Figure 6 : Modèle présentant le recrutement de DNMT1 par l'enzyme UHRF1 d'après Qin, W. et al<sup>47</sup>.

Dans un premier temps (Figure 6A), UHRF1 va venir se fixer aux niveaux d'un nucléosome possédant la modification H3K9me2/3 à l'aide des domaines TTD et PHD, ainsi qu'à de l'ADN méthylé ou non, *via* le domaine SRA. Le domaine RING sera alors hors de portée du nucléosome et ne pourra pas diriger une réaction d'ubiquitinylation. Lorsque le domaine SRA va rencontrer de l'ADN hémiméthylé (à l'extérieur du nucléosome), la structure d'UHRF1 va être modifiée<sup>44,45</sup> (Figure 6B), le domaine RING va alors être en contact avec le nucléosome. Une réaction d'ubiquitinylation va avoir lieu sur l'histone H3. Cette modification va participer au recrutement de DNMT1<sup>46,47</sup> qui possède un domaine d'interaction avec l'ubiquitine. Le

1.3 maintien de la méthylation est assuré.

### Régulation de la méthylation de l'ADN

La méthylation de l'ADN est une modification relativement stable par rapport aux autres modifications épigénétiques. Cependant, la déméthylation de l'ADN s'observe dans différents contextes biologiques. En effet, au cours de la spécialisation des cellules, le transcriptome change rapidement, cela implique la nécessité d'avoir un épigénome dynamique et flexible. On

observe, par exemple, qu'entre 4-8 heures après la fertilisation le taux de 5-méthylcytosine (5mC) chute rapidement chez la souris<sup>48,49</sup>. D'autres études ont montré une perte rapide de la méthylation de l'ADN lors de la différenciation des cellules germinales<sup>50</sup> ou encore suite à des réponses à des stimuli environnementaux<sup>51</sup>. Cette déméthylation peut avoir lieu de façon passive ou de façon active. Les mécanismes amenant à la perte de la méthylation de l'ADN sont encore sujets à vives discussions. Ainsi, seules les hypothèses mécanistiques les plus pertinentes seront décrites dans ce manuscrit.

### 1.3.1 La déméthylation passive de l'ADN

La méthylation de l'ADN était initialement considérée comme irréversible. En effet, la nature de la liaison carbone-carbone rendait une rupture directe de cette liaison trop coûteuse en énergie. Il a donc été proposé et accepté que la déméthylation s'effectuait à travers une dilution ou une synthèse *de novo* de l'ADN<sup>52</sup> comme présenté Figure 7.

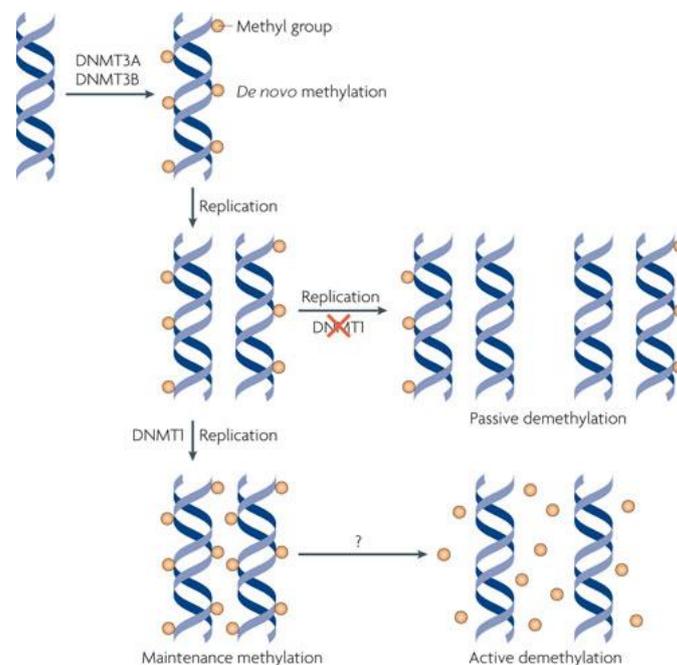


Figure 7 : Schéma représentant les différentes voies de déméthylation de l'ADN d'après Wu et Zhang<sup>49</sup>

Au cours de cette voie (Figure 7, partie haute), la méthylation de l'ADN va progressivement être diluée à travers plusieurs phases de réplication de l'ADN. L'hypothèse sous-jacente de ce mécanisme repose sur une inhibition du système UHRF1-DNMT1 qui assure la maintenance de la méthylation de l'ADN comme décrit précédemment. Cette absence d'activité n'est pas

encore bien comprise mais certaines découvertes récentes, (discutées dans la paragraphe 1.4.1, p.26) ont permis de lever le voile sur la compréhension de ce phénomène.

La déméthylation passive de l'ADN est un processus lent qui s'étend sur plusieurs générations. Cette lenteur d'exécution ne corrèle pas avec les pertes rapides du taux de 5mC observées chez les zygotes des souris<sup>48</sup>. Cela laisse supposer l'existence d'un ou plusieurs mécanismes de déméthylation actifs, indépendants de la réplication de l'ADN (Figure 7, partie basse).

### *1.3.2 La déméthylation active de l'ADN : les premières hypothèses.*

#### *1.3.2.1 La déméthylation enzymatique*

La voie la plus simple et la plus directe de déméthylation repose sur une élimination enzymatique du groupement méthyle des 5mC. Cette réaction est, comme évoquée précédemment, thermodynamiquement défavorable car elle nécessite la rupture d'une liaison C-C. Une étude a proposé qu'une telle réaction puisse être catalysée par une enzyme, la MBD2 (methyl-CpG binding domain protein 2)<sup>53</sup>, permettant ainsi d'obtenir une cytosine non modifiée et du méthanol. Malheureusement, aucun autre laboratoire n'a réussi à reproduire ce résultat dans des conditions physiologiques<sup>54</sup> et il a été rapporté que l'absence de MBD2 n'impactait pas la perte de la méthylation chez les ovocytes fertilisés<sup>55</sup>. A ce jour, aucune preuve convaincante ne supporte ce mécanisme.

#### *1.3.2.2 Les mécanismes de réparation de l'ADN*

D'autres mécanismes de déméthylation reposent sur des systèmes biologiques déjà existants : les deux principaux systèmes de réparation de l'ADN : la réparation par excision de nucléotides (nucleotide excision repair, NER) et la réparation par excision de base (base excision repair, BER).

### 1.3.2.2.1 BER : la réparation par excision de base

Le BER est utilisé pour corriger les bases endommagées ou les mésappariements entre bases de l'ADN –par exemple un mésappariement G:T. On recense deux mécanismes distincts au sein du BER (détailés dans <sup>56</sup>, et représenté Figure 8) :

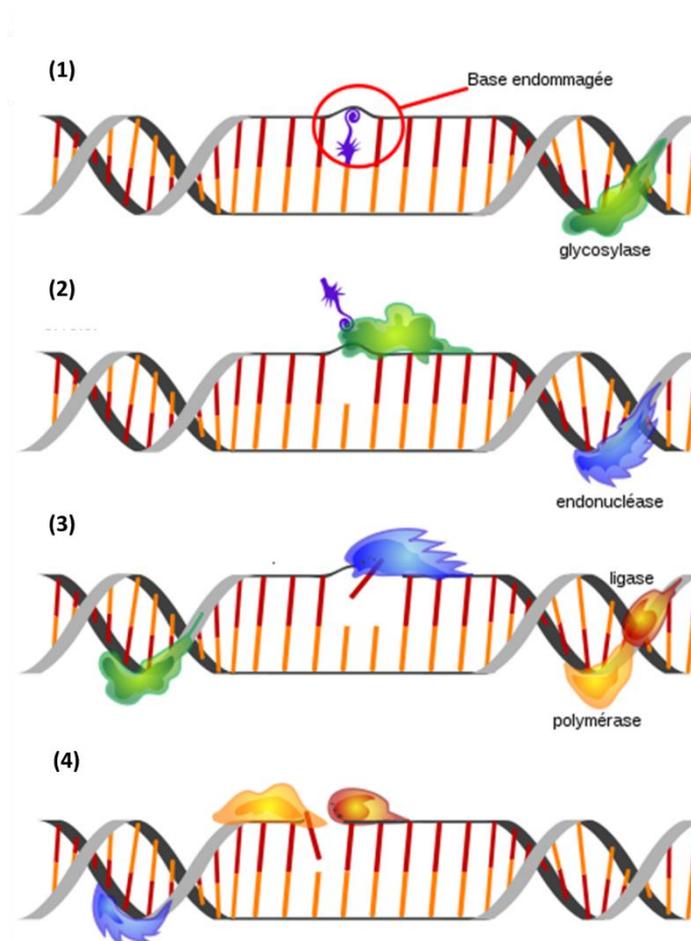


Figure 8 : Schéma des étapes du BER court (d'après une figure de Mariana RUIZ<sup>53</sup>)

- Le BER court (Short Patch BER, SP-BER) (Figure 8) commence par la reconnaissance de la base endommagée par une ADN glycosylase (1). Cette base est ensuite sortie du double brin par basculement, effectué par la même ADN glycosylase spécifique au dommage (2). La base modifiée est alors ôtée par hydrolyse de la liaison *N*-glycosidique entre la base et le désoxyribose. Cette réaction produit un site abasique (AP). Ce site est identifié par une endonucléase spécifique (AP endonucléase) qui va couper l'ADN de part et d'autre de l'AP (3). Une ADN polymérase reconnaît ensuite le site de coupure (4), et ajoute un nucléotide sur le site de l'AP, et une ligase vient réaliser la ligation du nucléotide incorporé (4).

- Le BER long (Long Patch BER, LP-BER) est une variante du SP-BER, où l'ADN polymérase va former un nouveau brin d'ADN sur quelques nucléotides. Le brin déplacé par cet ADN néo-synthétisé sera ensuite coupé avant l'intervention d'une ligase.

L'implication du BER dans la déméthylation de l'ADN a été supposée dès 1995 par Jost *et al.* qui ont découvert et purifié une glycosylase capable d'exciser une base 5mC d'un ADN hémiméthylé<sup>57</sup>. Cette protéine a, plus tard, été identifiée comme étant une thymine DNA glycosylase (TDG)<sup>58</sup>. Cette enzyme, capable de réparer les mésappariements T:G, possède un site actif capable d'accepter tous les dérivés oxydés de la cytosine et est capable d'établir des interactions avec eux<sup>59</sup>. Une autre enzyme a aussi été présentée comme ayant les mêmes caractéristiques que TDG, il s'agit de MBD4 (Methyl-CpG-Binding domain protein 4)<sup>60</sup>. Cependant, l'activité glycosylase de ces deux enzymes vis-à-vis de la base 5mC est très faible (de 30 à 40 fois inférieure) comparée à leurs activités pour des mésappariements avec une thymine ou un uracile. Cette faible activité ainsi que d'autres résultats ont mis en doute l'implication de ces deux glycosylases comme voie majoritaire de déméthylation de l'ADN.

#### 1.3.2.2.2 NER : la réparation par excision de nucléotides

La réparation par excision de nucléotides permet de réparer les lésions étendues sur l'ADN. Ces lésions proviennent généralement de dommages induits par une exposition à une irradiation ou à des produits chimiques. Le procédé général de cette voie de réparation est le suivant : une fois la zone endommagée reconnue, des enzymes vont venir exciser l'ADN de part et d'autre de la lésion. La région d'ADN simple brin ainsi libérée va être reconstruite par des ADN polymérases et des ligases, qui vont conduire à la réparation de cette région par synthèse d'un nouveau brin complémentaire<sup>61</sup>.

Mise en avant la première fois en 2007<sup>62</sup>, la famille des protéines Gadd45 (growth arrest and DNA damage inducible protein 45) est capable de stimuler la déméthylation active en interagissant avec XPG, une endonucléase du NER. Bien que d'autres études ont confirmé cette activité, elle a aussi été remise en cause<sup>63</sup>. Actuellement, la balance penche pour une activité de Gadd45a sur la perte de la méthylation de l'ADN avec la participation du NER. En effet,

Gadd45a jouerait le rôle d'un adaptateur entre les sites de déméthylation et les protéines de réparation de l'ADN du NER<sup>64,65</sup> mais aussi du BER<sup>66</sup>.

#### 1.3.2.2.3 Désamination et implication du BER

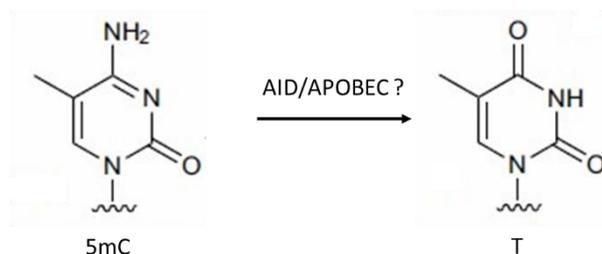


Figure 9 : Réaction de désamination de la 5mC en T par l'AID/APOBEC

La désamination en position 4 de la 5mC génère une thymine (T) provoquant l'apparition d'un mésappariement T:G qui peut être réparé par une multitude d'ADN glycosylases (TDG par exemple) et par le BER. Plusieurs enzymes capables de désaminer des cytosines ont été proposées. Les plus importantes désaminases sont AID (Activation-Induced Deaminase) et APOBEC (Apolipoprotein B pre-mRNA editing enzyme, catalic polypeptide)<sup>67</sup> (Figure 9). L'implication de ces protéines dans la déméthylation est discutée. En effet, il a été montré in vitro que l'AID/APOBEC pouvait désaminer 5mC en T<sup>68</sup>. L'implication de ce couple enzymatique dans la perte de la méthylation a été confirmée à plusieurs reprises<sup>69,70</sup>, mais il a aussi été mis en doute<sup>71,72</sup>. Dans ces dernières études, il est montré que l'AID/APOBEC agit de façon préférentielle sur l'ADN simple brin, et que leur activité est plus faible (d'un facteur 10 à 20) pour la 5mC que pour la cytosine. Ces deux résultats limitent donc l'implication de ce système pour une déméthylation efficace de l'ADN double brin. A l'heure actuelle, le rôle de l'AID/APOBEC est toujours discuté.

## Les nouveaux acteurs de l'épigénétique et leurs rôles dans la régulation de la méthylation

### 1.4.1 L'oxydation des 5mC comme étape cruciale de la déméthylation

1.4 Plusieurs équipes de chercheurs ont émis l'hypothèse que la déméthylation de l'ADN se ferait *via* une oxydation de la 5mC. Cette base oxydée, ne serait plus reconnue par la machinerie de maintien de la méthylation assurant ainsi une déméthylation passive viable. Elle pourrait, aussi, être mieux reconnue par des glycosylases, augmentant ainsi l'apport du BER dans le mécanisme de déméthylation.

Cette hypothèse se base sur des études de systèmes de déméthylation déjà établis comme la déméthylation des histones. Celle-ci est réalisée successivement par une oxydation du groupement méthyle et une perte de formaldéhyde<sup>73</sup>. Un mécanisme similaire est probable pour l'ADN puisqu'une protéine AlkB (de la famille des dioxygénases à Fe (II), dépendantes du 2-oxyglutarate (2-OG)) a été montrée comme capable de déméthyliser de la 3-méthylcytosine (3mC) et de la 1-méthyladénine (1mA) chez les bactéries<sup>74</sup> (Figure 10).

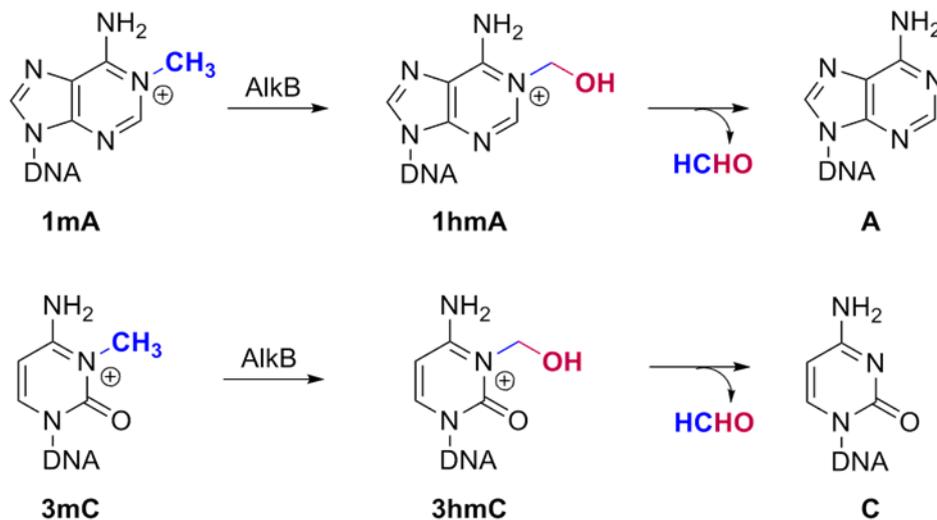


Figure 10 : Réaction de déméthylation de la 1mA et de la 3mC par AlkB

AlkB assure l'oxydation des groupements méthyle. Dans ce cas de figure, l'hydroxylation fragilise la liaison C-N et le relargage de formaldéhyde se fait spontanément. C'est une différence notable avec l'oxydation des 5mC, puisque dans ce cas, la liaison C5-CH<sub>2</sub>OH est toujours stable et il faut imaginer l'intervention d'autres phénomènes pour aboutir à un résultat similaire.

L'oxydation du groupement méthyle des thymines – structure la plus proche des cytosines méthylées- est bien connue dans plusieurs organismes. Chez les trypanosomes, certaines

thymines sont oxydées en 5-hydroxyméthyluridines (5hmU), auxquelles un glucose est, par la suite, transféré pour former la base J ( $\beta$ -D-glucosyl-hydroxyméthyluracile)<sup>75</sup>. Les enzymes J-binding protein 1 et 2 (JBP1, 2) ont été identifiées comme étant des dioxygénases responsables de la formation de la 5hmU<sup>76</sup>. Il s'agit de protéine de la famille des dioxygénases à Fe(II) dépendante du 2-oxoglutarate. La recherche d'homologues des JBP chez les mammifères à partir de leur domaine catalytique supposé a permis au groupe de Rao d'identifier les enzymes Ten Eleven Translocation (TET)<sup>77</sup>. Cette découverte a eu un très grand impact dans le monde de l'épigénétique.

#### 1.4.1.1 Les enzymes TET et l'oxydation de la 5mC

En 2009, deux publications parues simultanément dans *Science* ont mis en avant la présence de la 5-hydroxyméthylcytosine (5hmC) dans l'ADN des mammifères, et l'implication des enzymes TET dans sa formation. La 5hmC avait déjà été découverte il y a plus de 40 ans<sup>78</sup>, et était depuis lors considérée comme un dommage oxydatif. Ces deux études ont contredit ce postulat. Dans la première étude, Kriaucionis et Heintz détectent et identifient la 5hmC dans des cellules de Purkinje et dans des cellules de cerveau de souris<sup>79</sup>. Pour ce qui est de la seconde publication, Tahiliani *et al.* identifient la 5hmC dans des cellules souches embryonnaires de souris mais surtout, montrent que la protéine TET1 est capable d'oxyder la 5mC en 5hmC<sup>80</sup>. Cette modification n'est pas due à un quelconque dommage oxydatif, mais est bien le résultat d'une action enzymatique contrôlée.

#### 1.4.1.2 La famille enzymatique TET

Chez les mammifères, la famille des enzymes TET comprend trois membres : TET1, TET2 et TET3. Ce sont toutes des dioxygénases à Fe (II) dépendantes du 2-oxoglutarate, présentant la même activité catalytique<sup>81</sup>. Ces protéines se composent principalement de deux domaines :

- Un domaine CXXC *N*-terminal correspondant à un domaine à doigts de zinc. Ce domaine possède une ou plusieurs séquence(s) caractéristique(s) de type CGXCXXC(X)<sub>n</sub>C, où C est une cystéine, G une glycine et X un acide aminé quelconque. Ces régions CXXC sont caractéristiques des protéines associées à la réorganisation de la chromatine comme les MDB ou encore DNMT1.
- Un domaine catalytique constitué d'une zone riche en cystéines et d'un domaine *C*-terminal de type DSBH (double-stranded beta helix). C'est au niveau de cette dernière région qu'est

situé le cœur catalytique de type dioxygénase à Fe (II) dépendante du 2-oxoglutarate. Il comprend un motif HX...H, où X est soit un acide aspartique, soit un acide glutamique, tous deux responsables de la complexation de l'atome de fer impliqué dans le mécanisme catalytique.

### 1.4.1.3 Réaction catalysée par les enzymes TET

La réaction d'oxydation de la 5mC en 5hmC n'est pas encore définitivement établie. Un mécanisme probable a toutefois été émis. Il se base sur le cycle catalytique des dioxygénases à Fe (II) dépendantes du 2-oxoglutarate (Figure 11).

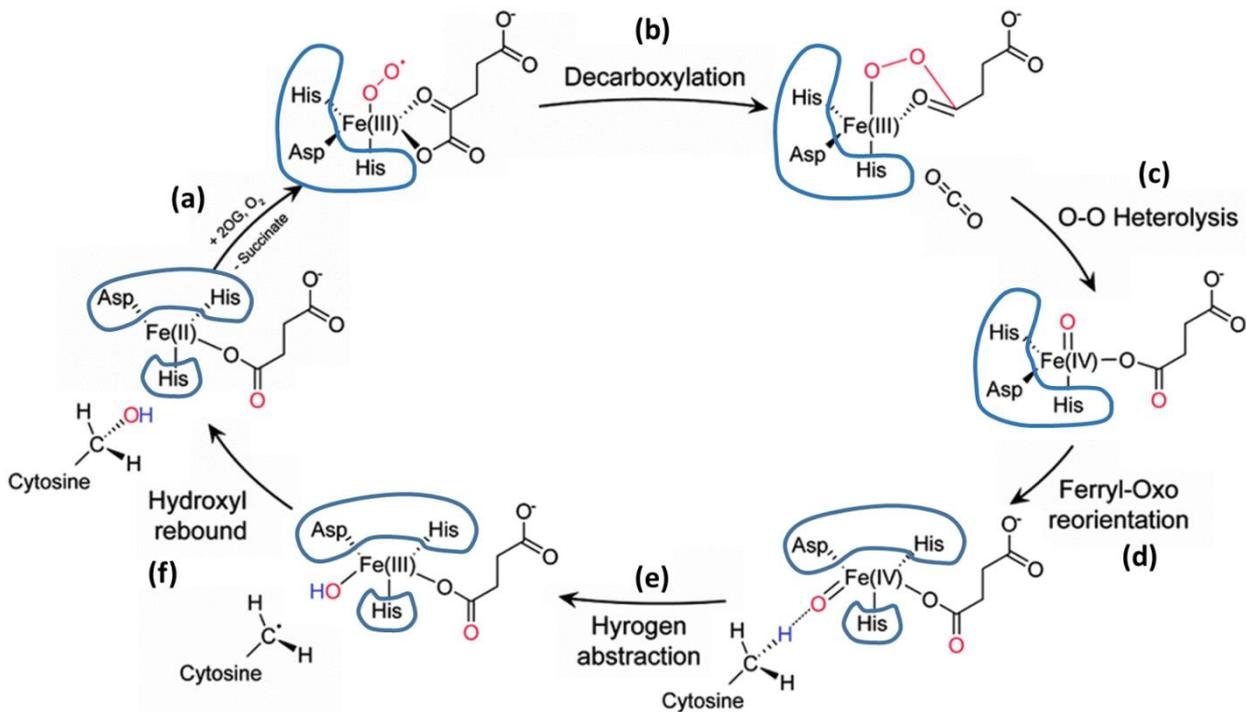


Figure 11 : Cycle catalytique caractéristique des dioxygénases à Fe (II) dépendantes du 2-oxoglutarate (2OG) adapté de Mantri et al.<sup>82</sup> pour l'oxydation de la 5mC en 5hmC. Les zones entourées en bleu correspondent aux acides aminés du site catalytique des enzymes TET.

Au cours de cette réaction, la coordination de l'atome de fer par le dioxygène et le 2-oxoglutarate permet d'oxyder le Fe (II) en Fe (III) (a). Après une étape de décarboxylation (b) suivie d'une hétérolyse (c), le Fe (III) est oxydé en Fe (IV) (d). C'est sous ce degré d'oxydation que l'atome de fer va réaliser un transfert d'hydrogène du groupement méthyle de la 5mC (e) après une réorientation du complexe catalytique. Une dernière étape radicalaire (f) permettra l'obtention de la 5hmC avant régénération du cycle catalytique avec la coordination d'une nouvelle molécule de 2-oxoglutarate, d'une nouvelle molécule de dioxygène et de l'élimination du succinate.

#### 1.4.1.4 Substrats des enzymes TET : nouvelles bases épigénétiques

Deux années après la découverte du rôle des enzymes TET, Ito *et al.* ont démontré qu'elles sont aussi capables de catalyser *in vitro* l'oxydation successive de la 5-méthylcytosine (5mC) en 5-hydroxyméthylcytosine (5hmC) puis en 5-formylcytosine (5fC) et ensuite en 5-carboxycytosine (5caC)<sup>82</sup> (Figure 12). Ces bases oxydées ne correspondent pas à un biais réactionnel puisqu'elles ont été identifiées *in vivo* dans l'ADN génomique de vertébrés<sup>82,83</sup>. Ces résultats ont permis d'établir la présence de trois nouvelles bases épigénétiques ainsi que leur voie de formation.

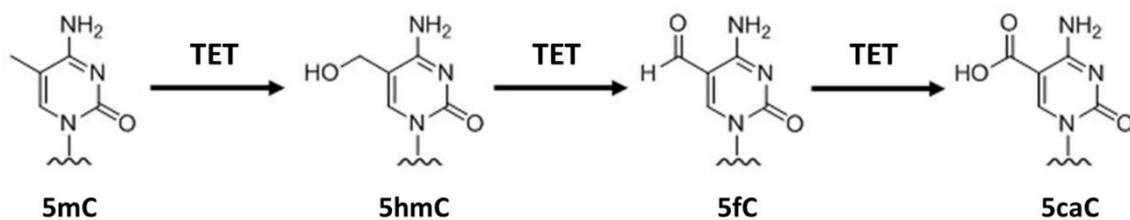


Figure 12: Oxydations successives de la 5mC par les enzymes TET.

#### 1.4.2 Le rôle des bases oxydées dans la déméthylation de l'ADN.

Depuis la détection de ces formes oxydées de la 5mC, le rôle de celles-ci est débattu. L'une des hypothèses porte sur l'implication de ces bases oxydées dans la perte de la méthylation de l'ADN. Trois voies potentielles de déméthylation sont envisageables après oxydation de la 5mC : une dilution passive des bases oxydées à l'aide de la réplication, une élimination enzymatique directe et un système se basant sur les mécanismes de réparation de l'ADN. Ces voies permettent de compléter les mécanismes déjà décrits précédemment en répondant à certaines lacunes de ces systèmes.

##### 1.4.2.1 La dilution passive des bases oxydées (AM-PD).

La dilution passive des bases oxydées se déroule selon le même procédé précédemment évoqué (paragraphe 1.3.1, p.21). L'oxydation de la 5mC permet de faciliter la dilution de la méthylation en répondant à l'hypothèse de ce mécanisme, c'est-à-dire, l'implication d'une inhibition potentielle du système UHRF1-DNMT1 qui assure la maintenance de la méthylation de l'ADN. En effet, bien qu'il existe des résultats contradictoires quant à savoir si UHRF1 peut reconnaître efficacement l'ADN hémi-hydroxy-méthylé<sup>84,85</sup>, d'autres études ont établi que DNMT1 est significativement moins efficace vis-à-vis d'un substrat hémi-hydroxy-méthylé. Son activité

chute de plus de 90 % par rapport à un substrat hémiméthylé<sup>86</sup>. Ce résultat a aussi été rapporté dans le cas des bases 5fC et 5caC<sup>87</sup>, où l'activité de DNMT1 est réduite.

Si l'on résume, les protéines TET vont initier un mécanisme de déméthylation en deux étapes. La première étape correspond à une modification active (AM) des 5mC à travers une ou des oxydations, ces bases oxydées ne seront plus reconnues par les DNMT1. La deuxième étape, quant à elle, correspond à la dilution passive (PD) de l'ADN par réplication. La grande différence de ce modèle par rapport à celui décrit précédemment est la capacité de l'AM-PD à être efficace en présence d'une machinerie de maintien de la méthylation fonctionnelle.

Il est à noter que les enzymes DNMT3 sont capables de reconnaître les sites héli-hydroxyméthylés aussi bien que les sites héli-méthylés<sup>88</sup>. Elles sont alors capables de conserver leurs rôles, c'est-à-dire maintenir la méthylation, bien que l'implication de ces protéines seules ne suffise pas<sup>89</sup>, d'autres études sont nécessaires pour mieux appréhender les rôles de l'oxydation sur le mécanisme de maintenance de la méthylation.

#### 1.4.2.2 Une élimination enzymatique directe des bases oxydées

L'élimination directe d'un groupement méthyle demande le franchissement d'une haute barrière énergétique. En comparaison l'élimination d'un groupement méthyle oxydé paraît plus réalisable. On compte deux mécanismes possibles, représentés Figure 13, pour cette voie de déméthylation :

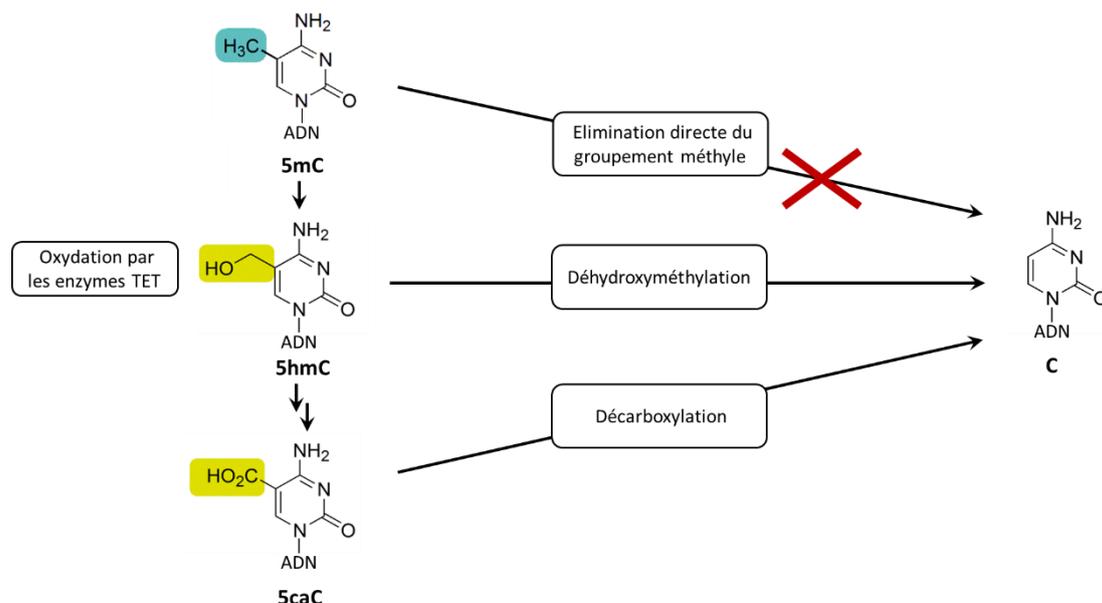


Figure 13 : Représentation des trois mécanismes possibles d'élimination directe de la méthylation de l'ADN.

- Le premier est soutenu par la capacité des DNMT à réaliser la déshydroxyméthylation des 5hmC (Figure 14). En effet, des études *in vitro* suggèrent que les DNMT3 peuvent catalyser la déshydroxyméthylation des 5hmC via un intermédiaire covalent entre la cystéine de la DNMT et le C6 de la cytosine pour régénérer des cytosines dans des conditions oxydantes et surtout en absence d'AdoMet (substrat des DNMT)<sup>90,91</sup>. La molécule AdoMet étant présente dans des taux relativement élevés dans tous les types cellulaires, cette voie est peu probable, elle est en tout cas à nuancer.

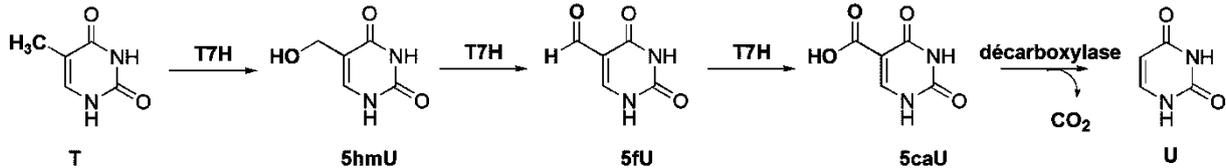


Figure 14 : Réaction de déshydroxyméthylation des 5hmC par les enzymes DNMT.

- Le deuxième mécanisme se base sur une décarboxylation des 5caC. Dans la voie de recyclage des pyrimidines, la conversion des thymines (T) en uraciles (U) est réalisée par deux enzymes. Au cours de ce processus, T est premièrement transformée par la thymine hydroxylase (THase), une dioxygénase à Fe (II) dépendante du 2-oxoglutarate, en 5-hydroxyméthyluracile (5hmU), puis en 5-formyluracile (5fU) et enfin en 5-carboxyluracile (5caU) au cours de trois oxydations successives (Figure 15). La conversion est finalisée par l'action de l'iso-orotate décarboxylase qui assure la décarboxylation de 5caU pour former U<sup>92</sup>. La similarité chimique entre la conversion de T en U et l'oxydation de la 5mC suggère que les enzymes TET fonctionnent comme la THase, et qu'il existerait une décarboxylase qui jouerait un rôle similaire à celui joué par l'iso-orotate décarboxylase<sup>55</sup>. Une seule étude à ce jour rapporte que la décarboxylation des 5caC peut avoir lieu au sein de lysat de cellules souches embryonnaires de souris<sup>93</sup>. Une décarboxylase capable de réaliser cette réaction reste cependant à être identifiée.

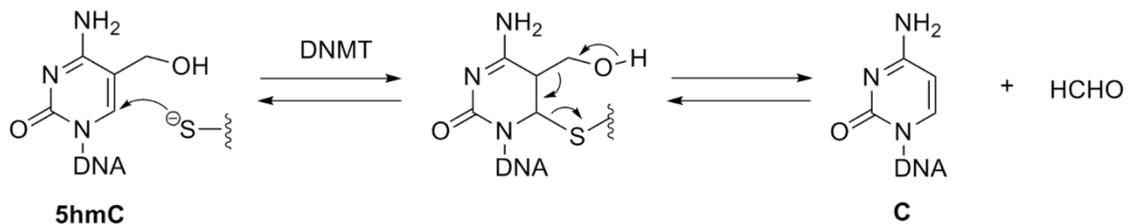


Figure 15 : Voie de recyclage de la thymine impliquant des oxydations successives du méthyle de la thymine par la thymine-7-hydroxylase (T7H), Fe(II)/2-oxoglutarate dépendante, et une décarboxylation.

### 1.4.2.3 Réparation des bases oxydées par le BER (AM-AR)

Ce mécanisme, appelé AM-AR, se caractérise par une modification active (AM) des 5mC par les TET suivie d'une réparation active (AR) de ces bases oxydées par le BER.

#### - TDG

Le principal facteur limitant de cette voie de déméthylation (déjà décrite dans le paragraphe 1.3.2.2.1, p.23) était l'activité faible des enzymes TDG et MBD4 (glycosylases) envers la 5mC. L'activité de MBD4 pour les nouvelles bases épigénétiques reste faible, ce qui n'est pas le cas pour TDG. En effet, il a été montré que contrairement à 5hmC, TDG présente une forte activité d'excision *in vitro* envers les bases 5fC et 5caC<sup>83,94</sup>. Ce résultat a été confirmé sur des extraits de cellules HeLa où seuls les oligonucléotides contenant de la 5fC ou de la 5caC entraînaient une réparation par le BER dépendant de TDG<sup>95</sup>. D'autres travaux montrent qu'une déplétion de TDG chez des cellules souches embryonnaires murines entraîne une augmentation de cinq à dix fois du taux de 5fC et 5caC, indiquant une élimination efficace de ces produits d'oxydation de la 5mC dans des conditions physiologiques<sup>96,97</sup>. On retrouve aussi ce mécanisme chez les embryons de *Xenopus laevis*<sup>95</sup>, il semble donc que cette voie de déméthylation soit répandue et tout particulièrement chez les vertébrés.

Il a été postulé que la réparation des 5fC et 5caC s'effectue par le BER court (SP-BER)<sup>95</sup>. Weber *et al.* ont par ailleurs réussi à reconstituer *in vitro* un système de déméthylation alliant TET1, TDG et les protéines du SP-BER<sup>98</sup> validant ce mécanisme comme une voie de déméthylation probable.

#### - TDG/NEIL

Une étude de protéomique ciblée sur l'identification des protéines partenaires de 5fC et 5caC a permis d'identifier, en dehors de TDG, une deuxième famille de glycosylases, les NEIL<sup>99</sup>. Cela suggère que ces enzymes pourraient elles aussi jouer un rôle dans l'AM-AR.

La famille des glycosylases NEIL (endonucléase VIII-like) compte trois membres. NEIL1 et 2 sont bifonctionnelles. Elles sont à la fois des glycosylases et des AP lyases<sup>100</sup>, c'est-à-dire qu'elles vont pouvoir rompre la liaison *N*-glycosidique entre le désoxyribose et la base oxydée/endommagée, et cliver le brin d'ADN en position 3' du site abasique nouvellement formé. NEIL3 est monofonctionnelle et a principalement le rôle de glycosylase<sup>101</sup>. Ces trois enzymes jouent un rôle dans la réparation des dommages oxydatifs de l'ADN, à l'instar d'autres glycosylases comme OGG1 (8-oxoguanine DNA glycosylase). Actuellement seules NEIL 1 et

2 feraient partie du processus de déméthylation de l'ADN dépendant du BER. Deux hypothèses ont été émises sur leurs rôles dans cette voie de réparation :

- NEIL1 et 2 seraient les glycosylases de secours de TDG. En effet, au cours d'expériences de « knockdown » *in vitro*, il a été montré que la transcription d'un gène n'était possible qu'après déméthylation de l'ADN. Cette déméthylation est réalisée à l'aide de TDG. En absence de cette dernière, elle est partiellement réalisée par les enzymes NEIL<sup>102</sup>. Cependant, NEIL1 et NEIL2 purifiées, ne prennent pas en charge les résidus 5fC et 5caC<sup>95,103</sup>, rendant cette hypothèse peu probable.
- Les enzymes NEIL permettent d'améliorer l'efficacité de TDG. Cette hypothèse se base sur l'observation du rôle de NEIL1 au cours de la réparation de la base 8-oxoguanine (8oxoG). Lors de l'excision de cette base endommagée, NEIL1 améliore le « turnover » enzymatique et donc, l'activité d'OGG1 en déplaçant OGG1 du site abasique<sup>104</sup>. Schomacher *et al.* ont constaté que NEIL1 et 2 possèdent la même activité envers TDG<sup>95</sup> (Figure 16). Une fois au niveau des sites abasiques, les NEIL sont capables de réaliser la rupture du brin d'ADN *via* leur activité lyase. Cela donne ainsi une direction à cette réaction de déméthylation. Le modèle coopératif de TDG et de NEIL est supporté par différentes expériences, où il a été montré qu'une réduction de l'expression de TDG s'accompagnait d'une surexpression de NEIL2, ou bien qu'une absence de NEIL impliquait une augmentation du taux de 5fC et de 5caC<sup>95</sup>.

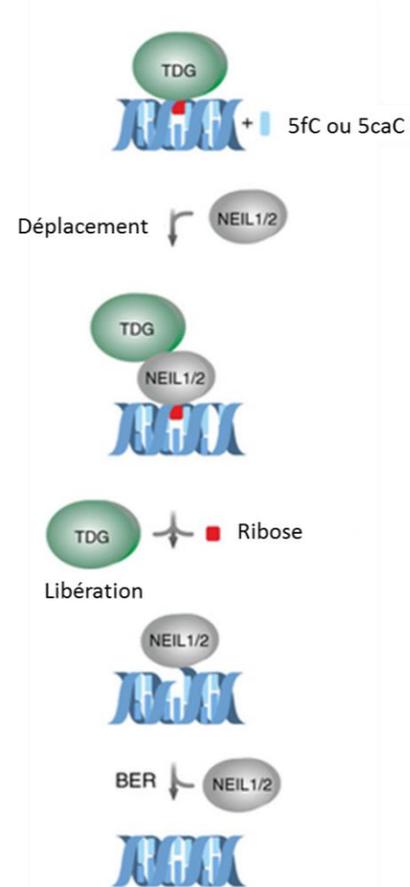


Figure 16: Rôle de NEIL1/2 d'après Schomacher *et al.*<sup>95</sup>

En résumé, cette voie de déméthylation implique trois étapes successives. Dans un premier temps les 5mC sont oxydées jusqu'à l'obtention de 5fC et 5caC par les enzymes TET. Les glycosylases TDG vont ensuite venir reconnaître et exciser ces bases oxydées. Les enzymes

NEIL vont, quant à elles, venir déloger les TDG des sites abasiques, leur permettant un meilleur « turnover » et aussi rompre le brin d'ADN. Cette cassure sera ensuite réparée par le BER. Il est à noter que la présence de TDG ou NEIL au niveau d'un site abasique sur un brin d'ADN empêche de façon stérique l'excision des bases 5fC et 5caC situées sur le brin opposé. Cela induit un mécanisme séquentiel non simultané qui diminue le risque de cassure double brin de l'ADN<sup>98</sup>.

## Bilan des mécanismes généraux de la méthylation de l'ADN

1.5

La méthylation de l'ADN est un processus biologique important et bien régulé que ce soit par des mécanismes de maintenance que par différents mécanismes de déméthylation, qui sont résumés sur la Figure 17.

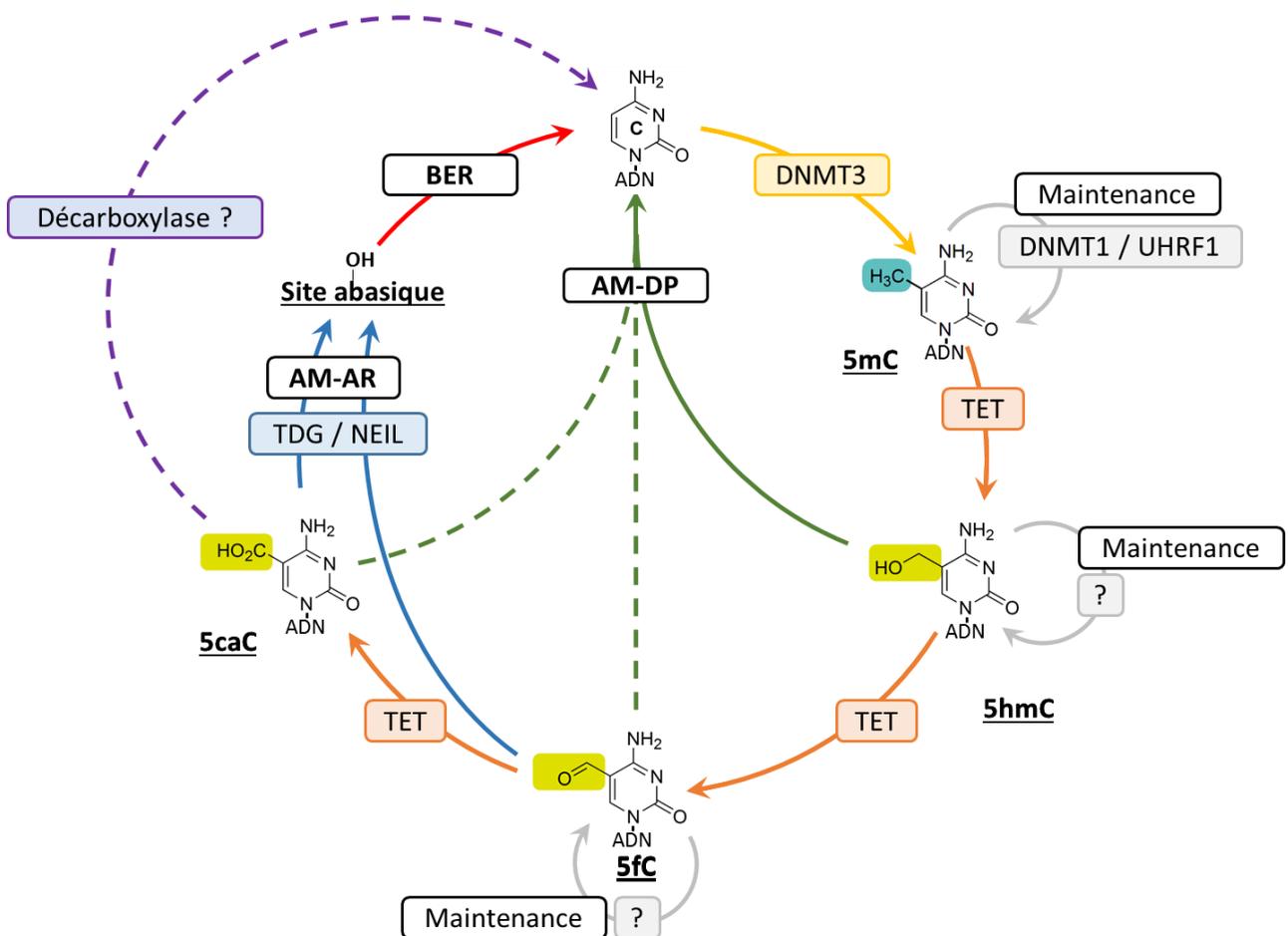


Figure 17 : Schéma bilan des voies de régulation de la méthylation de l'ADN. AM-PD correspond à la dilution passive de la méthylation dépendante de la réplication. AM-AR représente la voie de déméthylation se basant sur le BER. Les traits pleins correspondent aux mécanismes acceptés par la communauté scientifique, ceux en pointillés sont toujours discutés.

L'hypothèse de ces mécanismes de déméthylation se base essentiellement sur la découverte des bases oxydées 5hmC, 5fC et 5caC, qui interviennent en tant qu'intermédiaires dans les processus précédemment évoqués. Certaines de ces bases semblent aussi jouer un rôle épigénétique bien défini. En effet, la 5hmC est présente dans des quantités importantes dans le cerveau ainsi que dans d'autres tissus –par exemple la 5hmC représente 40 % des 5mC dans les cellules de Purkinje<sup>105</sup>. Ces résultats sont en contradiction avec le rôle d'intermédiaire, qu'on lui a attribué, puisque dans ce cas-là, la 5hmC aurait un temps de vie très court et donc n'aurait pas le temps de s'accumuler. La 5hmC est donc une modification épigénétique stable<sup>106</sup>. Il est d'ailleurs accepté que celle-ci soit impliquée dans la régulation des gènes pendant le développement, la carcinogénèse et dans la transcription active de certains gènes<sup>81,107,108</sup>, bien que les mécanismes biochimiques dont dépendent ces résultats ne sont pas encore élucidés. Plus récemment, il a été suggéré que la 5fC aurait une fonction propre au sein de l'ADN. En effet, Bachman *et al.* ont, à l'instar des résultats trouvés pour la 5hmC, montré que la 5fC était une modification stable<sup>109</sup>. Ce résultat a ensuite été confirmé par Su *et al.* ; qui considèrent la 5fC comme une base semi-permanente. Cette considération provient d'un de leur résultat où seulement 50 % des 5fC étaient excisés par TDG<sup>110</sup>.

La stabilité de la 5hmC, tout comme celle de la 5fC, amène à penser que ces deux bases possèdent une fonction qui leur est propre au sein des mécanismes de la régulation épigénétique. Cela suggère la présence de processus de maintien de ces bases, de façon analogue à la 5mC. Il est d'ailleurs curieux de noter que le deuxième membre de la famille UHRF (UHRF2) présente une spécificité pour la 5hmC<sup>99</sup>. On ne connaît pas encore son rôle exact mais il semblerait qu'elle recrute les enzymes TET pour continuer l'oxydation de la 5mC. UHRF2 pourrait assurer un rôle analogue à celui d'UHRF1 envers la 5hmC. A ce jour, aucun mécanisme de maintien de la 5hmC ou de la 5fC n'est décrit (Figure 17). L'identification de protéines partenaires de ces modifications de l'ADN pourrait permettre une meilleure compréhension de leur fonction et des différents processus qui entrent en jeu lors de leur régulation.

## 2 Méthodes de caractérisation de protéines affines pour les nouvelles bases épigénétiques dérivées de la 5mC

Les interactions entre l'ADN et les protéines jouent un rôle majeur dans la plupart des processus biologiques liés à l'expression des gènes. L'étude de ces interactions est importante pour appréhender les mécanismes liés à la régulation de cette expression, et l'identification des partenaires protéiques des modifications épigénétiques de l'ADN, que ce soit les protéines dites « *writers* » (qui les introduisent), « *readers* » (qui reconnaissent ces modifications) ou *erasers* » (qui les éliminent) est un domaine de recherche en plein essor.

Une grande variété de méthodes ont été développées pour explorer les interactions ADN-protéines<sup>111,112</sup>. La plupart de ces techniques se basent sur la caractérisation des protéines partenaires soit à l'aide d'anticorps spécifiques, soit à l'aide d'une analyse protéomique en spectrométrie de masse. Le domaine de la protéomique - définie comme l'étude de l'ensemble des protéines d'un échantillon biologique dans des conditions données - s'est développé de façon spectaculaire dans le début des années 2000 et jusqu'à aujourd'hui<sup>113</sup>. Ces méthodes ont d'ailleurs permis l'identification de protéines partenaires des modifications épigénétiques.

La découverte des dérivés oxydés de la 5mC<sup>80,79,82</sup> a conduit de nombreux groupes de recherche à se poser la même question : Quelles sont les protéines partenaires de ces nouvelles bases ?

Pour répondre à cette question, deux techniques ont été principalement utilisées : le gel retard (EMSA) et l'analyse protéomique en spectrométrie de masse.

2.1

### EMSA (Electrophoretic mobility shift assay) ou gel retard

#### 2.1.1 Principe général de l'EMSA

La méthode de gel retard (ou EMSA) est devenue un protocole standard pour étudier les interactions ADN-protéines<sup>114</sup>. Introduite par Fried et Crother<sup>115</sup> ainsi que Garner et Revsin<sup>116</sup>, le principe de cette technique est simple. Il se base sur le fait suivant : les complexes ADN-protéines migrent moins efficacement que l'ADN libre lors d'une électrophorèse sur gel non dénaturant, que le gel soit de polyacrylamide ou d'agarose<sup>117,118</sup>. L'EMSA est utilisé pour étudier les protéines se liant à des sondes oligonucléotides d'ADN connues comportant

classiquement un marquage radioactif ou fluorescent, et peut être utilisé pour évaluer le degré d'affinité ou de spécificité de l'interaction étudiée. Cette technique présente aussi la capacité de différencier les complexes de différentes stœchiométries. Par contre, elle nécessite la connaissance d'un des partenaires d'interactions. Cette méthode de caractérisation nécessite un *a priori* sur la protéine partenaire.

Un des inconvénients de cette technique est la possibilité que la bande retardée soit en réalité le résultat de la complexation de plusieurs protéines sur l'oligonucléotide rapporteur. Ces protéines peuvent soit correspondre à des protéines se liant directement à l'ADN soit des protéines recrutées par ces dernières. Toutefois, il est possible de séparer ces protéines en réalisant une électrophorèse de seconde dimension (EMSA-2D), isolant ainsi les partenaires de l'ADN. Les protéines peuvent ensuite être identifiées à l'aide d'anticorps spécifiques vis-à-vis d'une protéine ou être récupérées pour être analysées par spectrométrie de masse.

Le couplage de l'EMSA à la spectrométrie de masse présente plusieurs avantages. En effet, l'EMSA permet de simplifier le système d'analyse en ciblant les protéines d'intérêt. La simplicité relative de l'échantillon obtenu permet alors de réaliser son analyse en spectrométrie de masse. Par contre, la sensibilité du système dépend grandement de la quantité de protéines observable au sein de la bande d'intérêt du gel. De manière générale, l'utilisation d'un gel d'électrophorèse lors de la préparation d'un échantillon diminue la sensibilité de système d'analyse<sup>119</sup>. Cette technique a permis au groupe de Lawrence J. Abraham d'identifier le facteur de transcription YY1 comme étant capable de se lier à des séquences répétées de l'ADN de type CCAT<sup>120</sup>. Elle a aussi permis au groupe de Kenneth J. Macdonald de caractériser AtrA<sup>121</sup>, un facteur de transcription, comme se liant sur la région promotrice du gène actII-ORF4 responsable de la biosynthèse de l'actinorhodine, un antibiotique chez *Streptomyces coelicolor*<sup>122</sup>.

### 2.1.2 Application de l'EMSA

L'identification des protéines partenaires des dérivés oxydés de la 5mC par l'EMSA part d'un *a priori* : les protéines testées sont déjà connues et on suppose qu'elles vont être spécifiques vis-à-vis d'une base particulière.

Pour illustrer cette technique, on peut citer les travaux de Jin *et al.*<sup>123</sup>, où l'affinité des protéines de la famille des MBD (Methyl CpG Binding Domain), protéines spécifiques de la 5mC<sup>124</sup>, a été testée vis-à-vis d'un oligonucléotide contenant des 5mC et un oligonucléotide contenant des 5hmC (Figure 18).

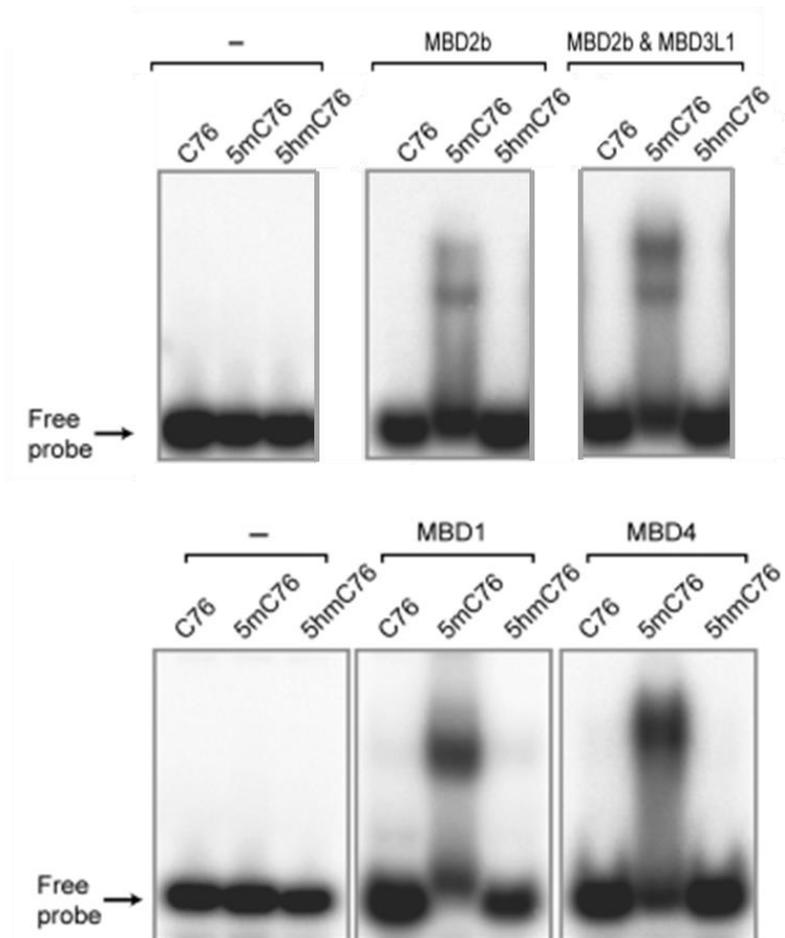


Figure 18 : Affinité des protéines MBD par rapport aux oligomères contenant de la 5mC et de la 5hmC adapté de Jin *et al.*<sup>124</sup>.  
 (A) Liaison de MBD2b, du complexe MBD2b / MBD3L1 (complexe MIRA), de MBD1 et MBD4 (100 ng de chaque protéine) avec des oligonucléotides 76-mères contenant des cytosines modifiées symétriquement, 5mC ou 5hmC dans des sites CpG. Les gels se situant à gauche correspondent à des expériences contrôles sans protéines.

Les protéines MBD2b, MBD2b / MBD3L1 (complexe MIRA), MBD1 et MBD4 (100 ng de chaque protéine) ont été incubées avec des oligonucléotides 76-mères contenant des cytosines modifiées symétriquement, 5mC (5mC76) ou 5hmC (5hmC76) dans des sites CpG ainsi qu'avec un oligonucléotide contrôle (C76), avant d'être déposées sur un gel retard. L'apparition de bandes retardées est visible uniquement en présence de protéines (Figure 18). De plus, ces bandes retardées s'observent seulement sur les pistes correspondant aux oligonucléotides contenant des 5mC (Figure 18, pistes 5mC76) quel que soit la protéine utilisée. Jin *et al.* en ont déduit que les protéines de la famille des MBD ne présentaient pas d'affinité pour les oligonucléotides présentant des 5hmC.

L'EMSA a aussi permis de mettre en évidence l'affinité de la glycosylase TDG vis-à-vis de la 5fC et de la 5caC<sup>125</sup>. En effet, Zhang *et al.* ont comparé l'affinité d'une version inactive de la TDG (hTDG<sup>cat(N140A)</sup>) en l'incubant avec des oligonucléotides 23-mères dont une cytosine était modifiée soit par de la 5mC soit par une forme oxydée de la 5mC (Figure 19).

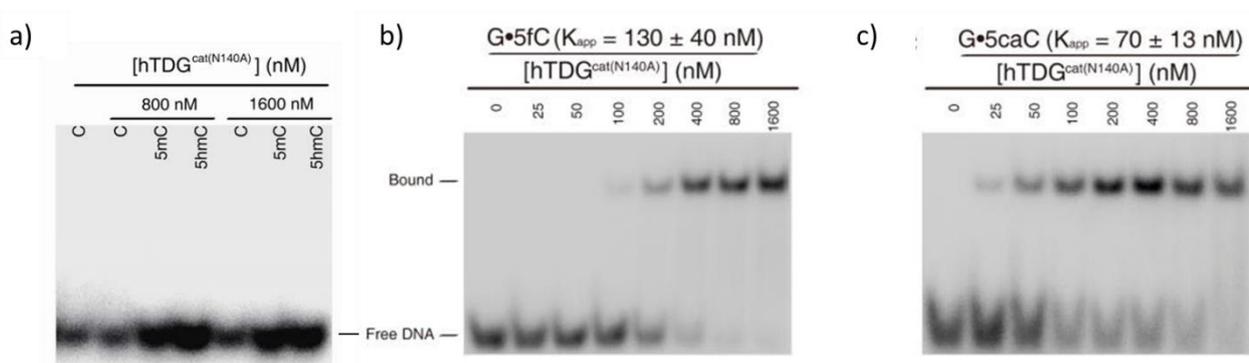


Figure 19 : Affinité de TDG vis-à-vis d'oligonucléotides 23-mères contenant une cytosine modifiée. (a) EMSA d'une version inactive de TDG (hTDG<sup>cat(N140A)</sup>) à deux concentrations différentes, et des oligonucléotides contenant soit une cytosine, soit une 5mC, soit une 5hmC (4 nM) ; (b) EMSA de la hTDG<sup>cat(N140A)</sup> avec un oligonucléotide contenant une 5fC ; (c) EMSA de la hTDG<sup>cat(N140A)</sup> avec un oligonucléotide contenant une 5caC. Figure adaptée des travaux de Zhang *et al.* <sup>126</sup>.

Leurs travaux ont montrés que TDG ne présentait pas d'affinité envers les oligonucléotides contenant de la cytosine, de la 5mC ou de la 5hmC puisqu'aucune bande retardée correspondant à un complexe TDG-ADN n'est observable lors de la réalisation de l'EMSA avec ces partenaires d'études (Figure 19a). Par contre, TDG se lie bien aux oligonucléotides constitués d'une 5fC ou d'une 5caC comme on peut l'observer sur la Figure 19, b et c. Ces deux gels retards montrent que hTDG<sup>cat(N140A)</sup> se lie de façon préférentielle à l'ADN contenant de la 5caC par rapport à celui contenant de la 5fC. En effet, la constante de dissociation apparente obtenue avec la 5caC est inférieure à celle obtenue avec la 5fC (70 contre 130 nM, Figure 19, b et c).

La contrainte de l'*a priori* sur les protéines utilisées lors de l'EMSA limite l'application de cette technique dans l'identification de nouvelles protéines partenaires, que celles-ci soient connues ou non. Actuellement, l'EMSA est utilisé pour caractériser le rôle des protéines en interaction avec l'ADN. Par exemple, elle a été utilisée pour montrer le rôle du domaine SRA au sein de la protéine UHRF1<sup>41</sup>. En effet, seule la protéine UHRF1 possédant ce domaine non muté, est capable d'interagir avec de l'ADN. L'EMSA a aussi permis d'élucider le rôle stimulateur de la

DNMT3L sur l'activité des protéines DNMT3A et DNMT3B<sup>27</sup>, par des expériences de co-incubation.

### **L'analyse protéomique pour la caractérisation de protéines partenaires**

L'identification des protéines partenaires des dérivés oxydés de la 5-méthylcytosine a été principalement reportée lors de deux études protéomiques<sup>126,99</sup>. Ces études se basent sur une 2.2 stratégie de préparation de l'échantillon dite « pull-down » couplée à une analyse en spectrométrie de masse.

#### *2.2.1 Le « pull-down » ou la purification d'affinité : principe*

Le « pull-down » ou purification d'affinité se base sur l'interaction entre deux molécules ayant de l'affinité l'une envers l'autre. Il s'agit d'une méthode assimilable à une technique chromatographique, qu'on appelle aussi « promotor immobilisation ». Elle nécessite la connaissance d'un des partenaires d'interaction, dans notre cas soit une séquence spécifique d'ADN soit d'une protéine partenaire. Le partenaire connu est fixé sur un support –il devient ainsi le « déterminant d'affinité » (Figure 20). Un mélange comportant des molécules d'intérêt est alors chargé sur le support greffé. Le partenaire d'interaction se fixe sur le déterminant d'affinité. Le support est ensuite lavé pour éliminer toutes les interactions non spécifiques. La dernière étape de ce système consiste en un relargage soit du déterminant d'affinité, soit du partenaire d'affinité à l'aide de conditions d'élution drastiques. Plus l'affinité entre partenaires est forte, plus l'enrichissement sera important.

Il existe aussi une variante à ce système de capture de protéine, appelée «promotor trapping »: cette fois l'ADN (déterminant d'affinité) est directement incubé avec un extrait protéique, avant d'être récupéré sur un support solide. Les protéines affines sont ensuite lavées et éluées de la même façon que pour le pull-down.

Pour constituer une phase chromatographique d'affinité, l'un des partenaires doit être couplé sur un support. Ce couplage peut être de type covalent : pont disulfure, liaison amide, liaison de coordination à un métal ou bien de type non covalent. L'un des couplages non covalents les plus classiques repose sur l'interaction entre un groupement de type biotine et un analogue de l'avidine. L'interaction entre ces deux partenaires correspond à l'interaction non covalente la plus forte connue avec un  $K_d$  de  $10^{-14}$  M<sup>127</sup> pour le couple biotine-streptavidine. Typiquement, des billes magnétiques recouvertes de streptavidine sont utilisées comme support et fonctionnalisées par le déterminant d'affinité muni d'une biotine (Figure 20).

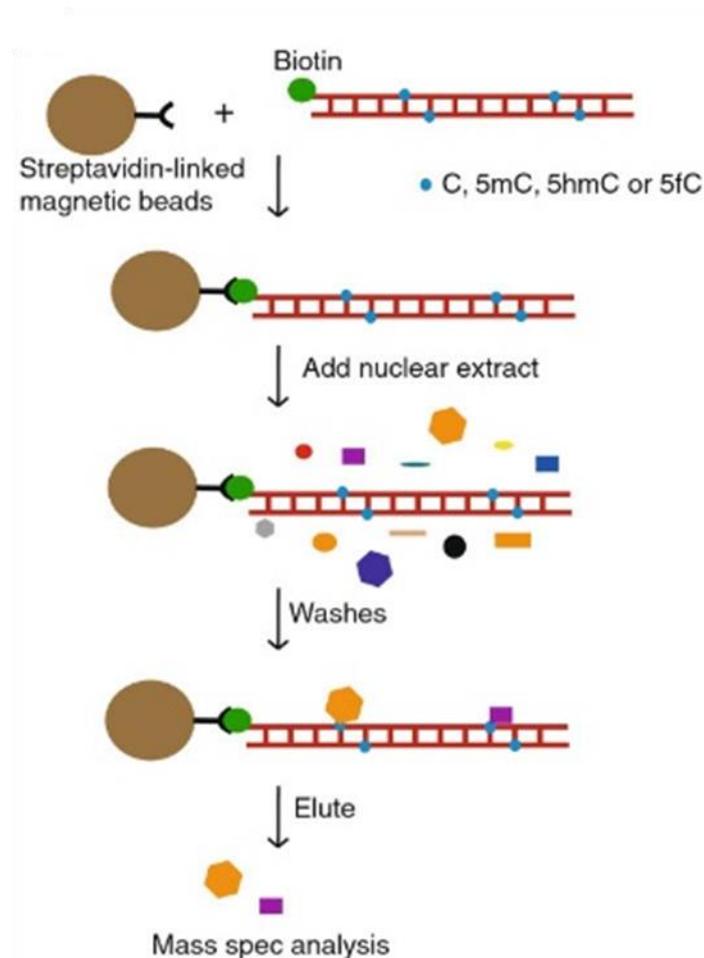


Figure 20 ; Schéma représentant une purification d'affinité de type « DNA pulldown » d'après Iurlaro, M. et al<sup>4</sup>. L'oligonucléotide, possédant ou non des modifications épigénétiques, est fixé à l'aide de son groupement biotine sur des billes magnétiques comportant des streptavidines. Le déterminant d'affinité obtenu est incubé avec un extrait nucléaire, avant d'être lavé. Les protéines affines envers l'oligonucléotide sont ensuite éluées avant d'être analysées par spectrométrie de masse.

### 2.2.2 Le « pull-down » ou la purification d'affinité : applications

La purification d'affinité basée sur l'ADN est un concept simple mais son analyse est difficile à mettre en place. En effet, les protéines spécifiques d'une modification, d'un locus ou d'une

séquence particulière de l'ADN sont faiblement abondantes par rapport aux protéines non spécifiques se liant à l'ADN- par exemple l'ensemble des protéines impliqué dans la réplication de l'ADN. L'analyse des échantillons nécessite donc l'utilisation de contrôles dans le but d'éliminer au maximum les interactions non spécifiques que ce soit pendant le pull-down même, ou en cours d'analyse par comparaison des échantillons. Pour ce faire, on utilise classiquement une analyse protéomique quantitative. C'est-à-dire, qu'en plus d'être identifiées simplement, les protéines sont quantifiées par comparaison de leur enrichissement respectif sur une séquence d'ADN contenant une modification par rapport à une séquence d'ADN contrôle. Typiquement, plus une protéine est présente sur une séquence modifiée par rapport à la séquence contrôle, plus cette protéine est spécifique de la modification introduite. On parle aussi de protéomique comparative. Cette technique d'analyse présente l'avantage d'exclure en partie les interactions non spécifiques puisqu'elles vont être présentes dans des proportions équivalentes au sein de la séquence modifiée et de la séquence contrôle. Théoriquement, seules les protéines spécifiques pour l'une ou l'autre des séquences sont identifiées dans ce genre d'analyse.

Il existe de plus en plus de méthodes de protéomique quantitatives ou comparatives. Elles peuvent être classées en deux groupes : les techniques sans marquage ou « label-free » et les techniques avec marquage<sup>128,129</sup>.

#### 2.2.2.1 Stratégie « Label-free »

La stratégie sans marquage, comme son nom l'indique, ne nécessite pas de modification préalable ou post-interaction des différents partenaires biologiques d'étude. Elle est donc facile à mettre en place. Actuellement, on peut distinguer deux stratégies de quantification « label-free » largement utilisées mais fondamentalement différentes :

- Une quantification basée sur l'intensité du signal des ions correspondant aux peptides caractéristiques d'une protéine particulière en spectrométrie de masse<sup>130</sup>. Dans cette approche, les chromatogrammes ioniques de chaque peptide sont extraits d'une analyse LC-MS/MS. Cela permet d'accéder à leur concentration relative par mesure de l'aire de leur pic sur l'échelle de temps chromatographique. Ces concentrations relatives sont ensuite corrélées à l'intensité des pics obtenus en spectrométrie de masse pour ces mêmes peptides. Une fois étalonné, le suivi de l'intensité des ions peptidiques d'une même protéine permet de déduire la quantité relative de celle-ci. On peut notamment

citer les méthodes de TIC (total ions chromatograms) ou XIC (extracted ions chromatograms).

- Une quantification par comptage spectral<sup>131</sup> (« *spectral counting* ») où l'on compte et compare le nombre de spectres de fragments identifiant les peptides d'une protéine donnée. Cette approche est basée sur l'observation empirique selon laquelle : plus une protéine est présente dans un échantillon, plus les spectres MS en tandem recueillis sont nombreux pour des peptides de cette protéine. Par conséquent, la quantification relative peut être obtenue en comparant le nombre de certains spectres caractéristiques d'une protéine, entre un ensemble d'expériences.

C'est cette dernière méthode de quantification comparative qui a permis de réaliser un des deux criblages des protéines partenaires de la 5-hydroxyméthylcytosine et de la 5-formylcytosine par spectrométrie de masse<sup>126</sup>. Iurlaro *et al.* ont appliqué une stratégie de purification d'affinité basée sur l'ADN pour identifier les différents profils de protéines interagissant avec les dérivés oxydés de la cytosine. En premier lieu, ils ont travaillé avec deux séquences promoteurs de gènes : Pax6 (280 paires de bases) et Fgf15 (248 paires de bases), qui correspondent respectivement à une région sans îlots CpG et à une région avec îlots CpG, contenant des cytosines, des 5mC, des 5hmC et des 5fC. Ces séquences biotinylées ont été ensuite indépendamment incubées avec des billes magnétiques fonctionnalisées par de la streptavidine pour former des déterminants d'affinité (Figure 20). S'en est suivi une incubation avec des extraits protéiques de cellules souches embryonnaires de souris. Ils ont ensuite vérifié par western blot que leur système de criblage était capable d'enrichir la protéine UHRF1. L'analyse par spectrométrie de masse des différents échantillons de protéines enrichies leur a permis d'identifier des protéines spécifiques pour la 5mC telles que des *methyl-CpG binding proteins* comme MBD4, mais aussi d'autres protéines comme TET1 ou OGT1. Peu de protéines semblaient spécifiques pour la 5hmC d'après les résultats de leur étude, on peut citer RBM14, PRP8 et PNKP. Il est à noter qu'en fonction de l'absence ou de la présence d'îlots CpG, certains profils protéiques étaient modifiés. Par contre, ils ont identifié 21 protéines qui ne se liaient qu'à la 5fC. La majorité de ces protéines présentent une fonction dans la régulation de la transcription et de l'organisation de la chromatine, comme par exemple la famille des facteurs de transcription Fox (dont Foxk2) ou encore le complexe Nurd. Pour démontrer le rôle de ces lecteurs de 5fC, ils ont réalisés le knockdown ces protéines identifiées. Aucune modification des taux de 5fC et 5caC n'a été constaté sauf dans le cas de la protéine TDG. Ces résultats

permettent de valider le rôle de TDG en tant que protéine impliquée dans la déméthylation de l'ADN, tout en écartant les autres protéines identifiées de cette voie biologique. En effet, ces protéines reconnaîtraient la 5fC comme une modification épigénétique. Aucune fonction propre de la 5fC n'est établie à ce jour.

#### 2.2.2.2 Stratégie avec marquage

Il existe de nombreuses techniques de protéomique quantitative avec marquage. La majorité de ces techniques s'appuie aujourd'hui sur l'utilisation d'isotopes stables pour une analyse différentielle en spectrométrie de masse. Pour comparer deux échantillons, les protéines de l'un des échantillons sont marquées avec des isotopes lourds (protéines « lourdes ») alors que les protéines du deuxième échantillon sont marquées avec des isotopes légers (protéines « légères ») (Figure 21). L'analyse en spectrométrie de masse permet de repérer des paires de peptides lourds et légers dont l'abondance relative peut être calculée (Figure 21). Le marquage isotopique dans les échantillons à étudier peut se faire de différentes manières : par marquage métabolique<sup>132,133</sup>, par marquage chimique (ICAT ou *isotope coded affinity tags*, iTRAQ ou *isobaric tag for relative and absolute quantification*)<sup>134,135</sup>, par marquage enzymatique<sup>136</sup> ou par marquage post-enzymatique<sup>137</sup>.

La technique de marquage métabolique la plus couramment utilisée est appelée SILAC (*stable isotope labelling by amino acids in cell culture*). Développée en 2002 par le groupe Mathias Mann<sup>133</sup>, la méthode SILAC repose sur la culture de deux populations différentes de cellules identiques (Figure 21). L'une des populations de cellules est cultivée dans un milieu de croissance contenant des acides aminés non modifiés (milieu « light »). En revanche, la deuxième population est cultivée dans un milieu de croissance contenant des acides aminés marqués avec des isotopes lourds stables (non radioactifs, dit « heavy »). Par exemple, le milieu peut contenir de l'arginine marquée avec six atomes de carbone 13 (<sup>13</sup>C) au lieu du carbone 12 (<sup>12</sup>C). Lorsque les cellules poussent dans ce milieu, elles incorporent l'arginine lourde dans toutes leurs protéines. Par la suite, tous les peptides contenant une seule arginine sont 6 Da plus lourds que leurs homologues non marqués. Les protéines des deux populations de cellules sont ensuite combinées et analysées ensemble par spectrométrie de masse. Des paires de peptides chimiquement identiques de différentes compositions isotopiques peuvent être différenciées dans un spectromètre de masse en raison de leur différence de masse. Le rapport des intensités

des pics dans le spectre de masse pour ces paires de peptides reflète le rapport d'abondance pour les deux protéines.

Au cours de la seconde étude portant sur l'identification des protéines partenaires des cytosines modifiées, Spruijt *et al.* ont utilisé des séquences d'ADN de 23 nucléotides comportant de la 5mC, de la 5hmC, de la 5fC et de la 5caC<sup>99</sup> et ils les ont incubées lors d'un pull-down avec des extraits protéiques de cellules souches embryonnaires préparés selon la méthode SILAC (Figure

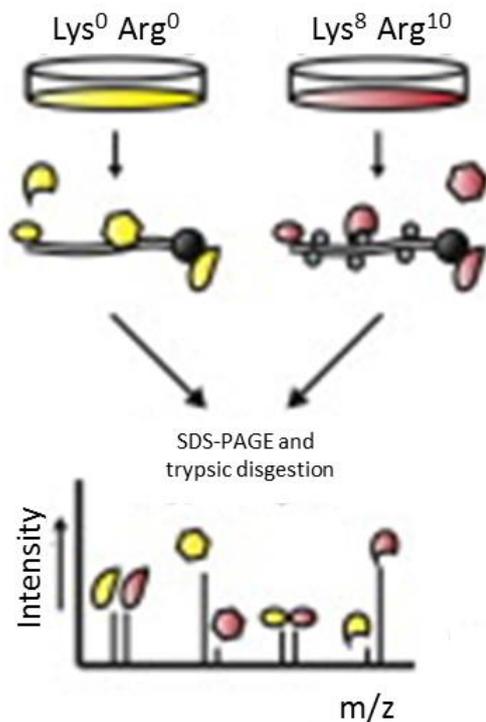


Figure 21 : Schéma représentatif de la méthode SILAC couplée à du pull-down utilisée par Spruijt *et al.*<sup>100</sup>. La culture cellulaire des cellules souches embryonnaires est réalisée dans deux milieux différents : un « léger » et un « lourd » composé de lysines et d'arginines comportant respectivement 8 et 10 <sup>13</sup>C. Les extraits cellulaires « légers » (jaune) et « lourds » (rouge) sont incubés respectivement avec un ADN non modifié et un ADN modifié au cours d'un pull-down, avant que les protéines enrichies ne soient mélangées avant d'être séparées à l'aide d'un gel d'électrophorèse et digérées à la trypsine pour être analysés par spectrométrie de masse. Le rapport d'intensité des pics correspondant à une même protéine entre sa version « légère » et sa version « lourde » permet de déterminer l'affinité de cette protéine pour la modification de l'ADN.

21). Ils ont aussi travaillé sur des extraits de cellules souches neurales de souris et des extraits de cerveaux de souris adultes mais la technique SILAC n'a pu être utilisée dans ces cas-là, car ces cellules ne peuvent pas être cultivées. L'analyse protéomique de ces deux dernières expériences a été réalisée selon une stratégie label-free, déjà décrite précédemment au cours des travaux d'Iurlaro *et al.*<sup>126</sup>

Après le pull-down d'ADN sur billes magnétiques, les protéines enrichies ont été séparées par un gel d'électrophorèse dénaturant avant d'être digérées à la trypsine pour être analysées en spectrométrie de masse (Figure 21, partie basse). Les résultats obtenus pour les trois types cellulaires sont les suivants :

- Au sein des cellules souches embryonnaires de souris, les protéines MeCP2 et UHRF1 ont été confirmées comme des lecteurs des 5mC et des 5hmC. Il a aussi été montré que Thy28, une protéine liée à l'apoptose, pouvait lire spécifiquement ces modifications.

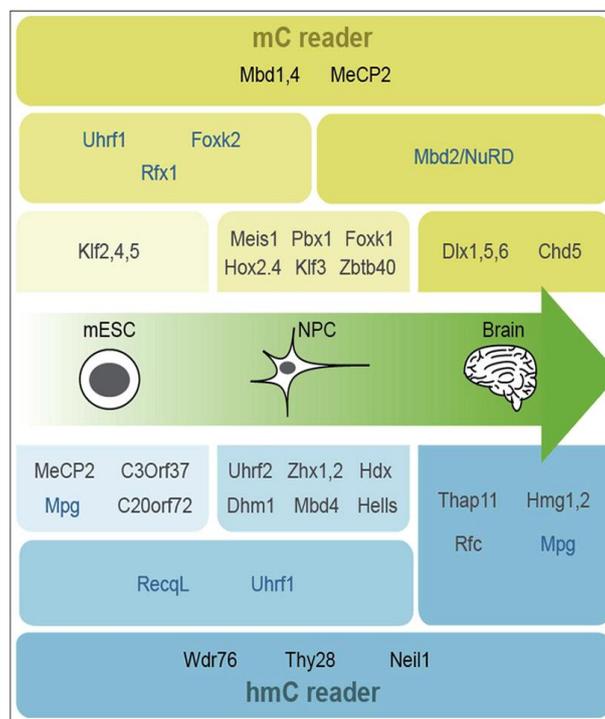
Par contre, les protéines de la famille des methyl-binding domains proteins MBD2 et MBD3 semblent préférentiellement affines pour les cytosines modifiées non-oxydées. Ce résultat a d'ailleurs été confirmé avec les protéines purifiées. Ils ont aussi montré que deux glycosylases : MPG et NEIL1 ainsi qu'une hélicase Recql semblent spécifiques de la 5hmC. Ces protéines peuvent être impliquées dans une voie de déméthylation active de l'ADN pour convertir la 5hmC en C *via* le BER. Pour ce qui concerne les modifications 5fC et 5caC, elles recrutent principalement des protéines impliquées dans la réparation des dégâts oxydatifs, avec peu de recouvrement entre la 5fC et 5caC. Ce phénomène suggère des rôles différents pour ces deux modifications. La glycosylase TDG se lie préférentiellement avec la 5fC et la 5caC, plutôt qu'avec la 5hmC, confirmant ainsi son rôle dans la déméthylation de l'ADN. La 5fC, au contraire de la 5caC, enrichit spécifiquement des protéines impliquées dans la réparation de l'ADN, comme p53, un facteur de transcription régulant la mitose et l'apoptose des cellules et dont les mutations se retrouvent dans un grand nombre de cancers. Les lecteurs de la 5caC comprennent, quant à eux, des protéines appartenant à des complexes modeleurs de la chromatine mais aussi, de façon surprenante la DNMT1. Globalement, cette étude montre que la 5hmC recrute moins de protéines que la C, la 5mC, la 5fC et la 5caC. Cette observation permet d'émettre deux hypothèses : la 5hmC serait un intermédiaire dans la déméthylation de cytosines alors que la 5fC et la 5caC joueraient un rôle de modifications épigénétiques.

- Dans les cellules souches neurales de souris (NPC), l'oxydation des 5mC entraîne une perte d'affinité avec plusieurs facteurs de transcription. La majorité des protéines en interaction avec la 5hmC est impliquée dans des processus de réparation de l'ADN, on peut citer NEIL1, NEIL3 et Recql. De façon plus intéressante, ce travail a permis d'identifier la protéine UHRF2 comme un lecteur spécifique de la 5hmC. Cette activité, ainsi que la capacité de cette protéine à se lier à la 5fC et la 5caC ont été testées et vérifiées avec la protéine recombinante. UHRF2 se lie préférentiellement à 5hmC puis à 5caC et reconnaît peu la 5fC et la 5mC. Ils ont aussi montré que la co-expression de la protéine UHRF2 avec le domaine catalytique de TET1 dans des cellules HEK293T, entraînait l'augmentation des niveaux de 5hmC, 5fC et 5caC. Cette observation permet de proposer UHRF2 comme une protéine facilitant le recrutement de TET1 ou en tout cas, l'oxydation successive des 5mC. Cette hypothèse se base sur la capacité de UHRF2

à faire basculer la 5hmC en dehors de la double hélice d'ADN<sup>138</sup>, dans un mécanisme similaire à celui de UHRF1 avec la 5mC.

- Les expériences sur les extraits protéiques de cerveaux de souris adultes ont permis d'identifier un plus grand nombre de protéines spécifiques pour la 5hmC que pour la 5mC, notamment Wdr76, Thy28 et Thap11. La caractérisation d'un plus grand nombre de protéines affines pour la 5hmC au sein de ces extraits est certainement due au fait que le cerveau adulte est l'organe contenant les plus hauts taux de 5hmC<sup>139</sup>.

Un schéma récapitulatif de toutes les protéines identifiées par Spruijt *et al.*<sup>99</sup> est présenté Figure 22. Cette étude correspond actuellement au criblage le plus abouti des protéines partenaires des formes oxydées de la 5-méthylcytosine.



2.3

Figure 22 : Vue d'ensemble de toutes les protéines identifiées comme étant spécifiques de la 5mC ou de la 5hmC au sein de trois milieux cellulaires différents d'après Spruijt *et al.*<sup>100</sup>

### Bilan des études de protéomiques permettant l'identification des protéines partenaires des dérivés oxydés de la 5mC

La comparaison entre les deux études protéomiques permettant l'identification des protéines partenaires des dérivés oxydés de la 5-méthylcytosine peut être faite. En effet, Spruijt *et al.* et Iurlaro *et al.* ont utilisé des extraits protéiques de cellules souches embryonnaires de souris lors de leur pull-down respectif. Malgré des séquences d'ADN différentes et une stratégie de quantification en spectrométrie de masse différente (SILAC et comptage spectral), ils ont trouvé

des résultats similaires. Pour illustrer cette similitude, on peut citer l'identification de la protéine UHRF1 comme étant capable de reconnaître des oligonucléotides modifiés par de la 5mC et de la 5hmC ou encore le fait que la protéine MBD3 ne soit pas plus spécifique pour la 5hmC que pour la 5mC, comme Yildirim *et al.* l'avaient reporté. Cependant, il existe des résultats différents comme le cas du facteur de transcription Foxk2. D'un côté, Spruijt *et al.* ont montré que Foxk2 était un partenaire protéique des 5mC et qu'il ne semblait pas spécifique des formes oxydées de cette base. D'un autre côté, Iurlaro *et al.* ont identifié Foxk2 comme une protéine spécifique de la 5fC. Cette différence de résultats peut s'expliquer par le fait qu'Iurlaro *et al.* ont réussi à identifier Foxk2 uniquement dans leur pull-down avec de l'ADN contenant de la 5fC. Bien que cela puisse traduire une affinité de Foxk2 pour la 5fC, cela peut aussi correspondre à un enrichissement globalement faible de cette protéine. En effet, si la protéine est faiblement exprimée et si elle interagit de façon transitoire sur l'ADN ou si sa constante d'affinité pour l'ADN est faible, la méthode de pull-down ne permet pas d'enrichir efficacement cette protéine qui sera typiquement éliminée lors des étapes de lavages. Or, ces trois caractéristiques peuvent s'appliquer aux protéines partenaires des formes oxydées de la 5mC qui sont faiblement présentes dans le génome, rendant ainsi l'identification de leurs partenaires difficiles.

Une approche permettant de capturer des protéines partenaires de modification de l'ADN de façon covalente serait intéressante :

- Elle permettrait de piéger les protéines ayant une faible constante d'affinité pour l'ADN ainsi que les protéines interagissant de façon dynamique. Celles-ci se trouveront alors fixées sur l'ADN.
- Elle permettrait aussi des conditions de lavages plus drastiques, éliminant ainsi un plus grand nombre de protéines non-spécifiques.

Les protéines affines seraient enrichies plus efficacement, ce qui rendrait leur identification par spectrométrie de masse plus facile. Dans le chapitre suivant, nous présenterons des méthodes permettant d'établir un lien covalent entre deux partenaires d'interaction.

### 3 Méthodes de pontage entre deux partenaires d'interaction.

Les méthodes d'études des interactions ADN-protéines par pontage sont fondées sur l'établissement d'un lien covalent entre les deux partenaires d'interaction. Ce lien peut avoir plusieurs origines qui nécessitent l'intervention d'un agent de pontage. Ce lien peut être de nature chimique ou photochimique.

Les agents pontants chimiques sont utilisés pour fixer des interactions en formant un lien covalent entre les deux partenaires en interaction. Ils sont généralement constitués de deux fonctions réactives séparées par un bras espaceur plus ou moins long permettant de ponter covalamment des groupements chimiques plus ou moins éloignés dans l'espace. Certains agents pontants n'ont pas de bras espaceur et sont appelés « zéro-length », c'est le cas du formaldéhyde.

#### Le Formaldéhyde : un agent de pontage chimique

3.1

Le formaldéhyde est utilisé depuis des décennies pour sonder la structure et la fonction de macromolécules ainsi que pour piéger des partenaires d'interaction<sup>140</sup>. Cette molécule électrophile est susceptible de réagir avec les groupements nucléophiles (-NH<sub>2</sub>, -OH, SH) d'un grand nombre de macromolécules biologiques. La réaction chimique qui a lieu avec le formaldéhyde se déroule en plusieurs étapes avec un ou plusieurs partenaires, différents ou non<sup>141</sup> (Figure 23). Lors de la première étape, un groupement nucléophile forme un lien covalent avec le formaldéhyde. On obtient un adduit de type hydroxyméthyle qui va être converti en une base de Schiff. Dans un second temps, la base de Schiff va réagir avec un autre groupement nucléophile, générant ainsi un pont covalent (crosslink) entre les différents intervenants de ces étapes (Figure 23, étape 2). Cette réaction présente l'avantage de réaliser le pontage de deux macromolécules différentes -comme les protéines ou l'ADN- tout en étant réversible.

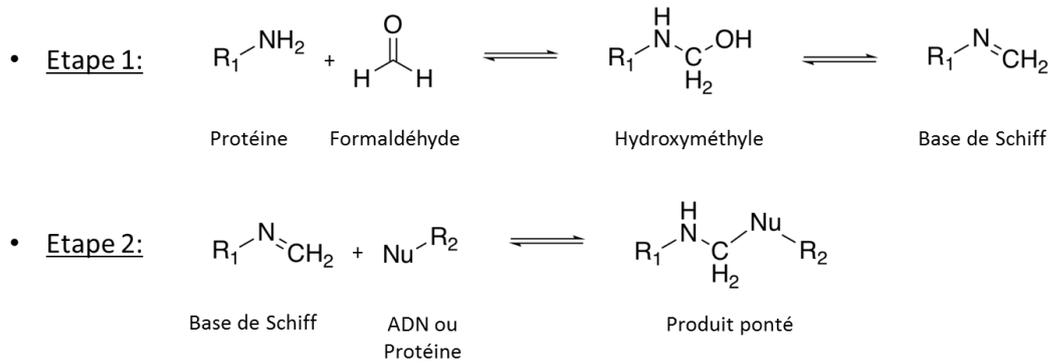


Figure 23 : Réaction chimique ayant lieu durant le pontage de biomolécules par le formaldéhyde.

Dans le cas des protéines, le formaldéhyde réagit de façon préférentielle avec les amines libres *N*-terminales, et les chaînes latérales de la cystéine (fonction thiol), de l'histidine (fonction imidazole), de la lysine (fonction amine), du tryptophane (indole) et de l'arginine (guanidine)<sup>142,143</sup>. Il a été montré que les lysines accessibles, c'est-à-dire exposées, étaient plus réactives que les autres résidus<sup>144</sup>. Cela se passe sans modification majeure de la structure tertiaire de la protéine qui reste active<sup>144</sup>. La capacité du formaldéhyde à ponter des acides aminés à des bases de l'ADN a été très étudiée *in vitro*. En comparant les produits de réactions contenant de la lysine, de la cystéine de l'histidine ou du tryptophane avec chacune des quatre bases de l'ADN, le plus haut taux de pontage est obtenu avec de la lysine et une désoxyguanosine<sup>145</sup>. Ce résultat a aussi été confirmé en utilisant des peptides courts et des trinuécléotides<sup>145</sup>.

Actuellement, le pontage au formaldéhyde est principalement utilisé pour l'immuno-précipitation de la chromatine (ChIP) grâce à la réversibilité de la réaction engendrée. Le principe consiste à réaliser *in cellulo* un pontage entre l'ADN et ses protéines partenaires. Les cellules sont alors lysées et l'ADN libre est dégradé. Les complexes covalents ADN-protéines sont isolés à l'aide d'anticorps anti-protéines partenaires (immuno-précipitation), avant de dissocier les partenaires pour pouvoir analyser la séquence d'ADN restante. Cette technique a permis de décrire les interactions physiologiques entre des séquences d'ADN, des facteurs de transcription et la chromatine modifiée ou non<sup>146,147</sup>. Le couplage de cette technique à la spectrométrie de masse est appelé ChIP-MS<sup>148</sup> et permet d'identifier des protéines partenaires de modification de l'ADN. Cependant, elle nécessite l'utilisation d'anticorps et donc d'un *a priori* sur les protéines d'interaction.

L'utilisation d'agents pontant chimiques comme le formaldéhyde présente quelques inconvénients. En effet, ces agents sont souvent sélectifs d'une fonction chimique, leur application est donc limitée à des protéines contenant cette fonction. De plus, seules les interactions stables entre partenaires, c'est-à-dire non transitoires, sont piégées par ces agents car la cinétique de réaction des agents pontant chimiques est plus longue que le temps de présence des protéines transitoires sur l'ADN. On peut encore ajouter que le pontage par des agents chimiques ne permet pas un contrôle temporel de la liaison covalente puisque ceux-ci réagissent dès la mise en présence de fonctions avec lesquelles ils vont pouvoir réagir.

Pour surmonter ces inconvénients, on peut choisir d'utiliser des agents de pontage photoactivables. L'utilisation de la lumière pour stimuler la formation de liens covalents est intéressante car elle peut fournir un contrôle temporel de la capture des partenaires d'interactions. Sous l'effet d'une irradiation, la plupart des agents photoactivables forment des intermédiaires hautement réactifs (par exemple, des carbènes, des nitrènes et des radicaux), qui réagissent de manière non spécifique, c'est-à-dire sur n'importe quel type d'atome.

### 3.2 Les agents de pontage photo-chimiques

Le principe des agents de pontage photo-chimiques est le suivant : une fois l'ADN en contact avec une protéine affine, l'échantillon est soumis à une phase d'irradiation. Au cours de cette phase, l'interaction va être fixée de façon covalente. Le groupement photo-réactif est généralement greffé sur l'un des partenaires de l'interaction, le plus souvent sur l'ADN. Il existe plusieurs types d'agents de photomarquage, le paragraphe suivant donne une description des composés les plus couramment utilisés.

#### 3.2.1 *L'ADN comme agent de photomarquage*

L'un des agents de photomarquage ADN-protéines les plus classiques est l'ADN lui-même. Le photomarquage direct de protéines avec l'ADN a été réalisé dès 1962 par irradiation UV à 254 nm<sup>149</sup>. Le site le plus réactif au cours d'une irradiation UV est la double liaison entre les carbones 5 et 6 des pyrimidines<sup>150</sup>. Celle-ci forme un di-radical qui va pouvoir réagir avec les chaînes latérales des acides aminés proches. Ce pontage représente une preuve directe d'une proximité entre une protéine donnée et une séquence d'ADN et n'implique aucune modification chimique des partenaires.

Cette technique de photomarquage a été utilisée dans de nombreuses applications. On peut citer l'étude des interactions entre l'ADN et les histones<sup>151,152</sup>, des études portant sur l'ARN polymérase et son mode de liaison sur l'ARN<sup>153,154</sup>. Elle a aussi été couplée à une analyse en spectrométrie de masse dès 1993 pour identifier le site de liaison du GTP au niveau de la  $\beta$ -tubuline lors de la formation des microtubules<sup>155</sup>. Elle présente l'avantage d'être non-invasive et peut être réalisée directement sur des cellules, sans modification de celles-ci, traduisant ainsi des conditions réelles. Malheureusement, les rendements de photomarquage sont assez faibles avec cette technique et l'irradiation dans l'UV provoque des dommages par exemple sur l'ADN<sup>156</sup>. Des groupements photo-activables ont donc été développés pour pallier ces problèmes. On peut les classer en deux groupes correspondant aux agents de photomarquage dit « classiques » et ceux analogues de nucléobases.

### 3.2.2 Les agents de photomarquage classiques

Les agents de photomarquage doivent répondre à certains critères pour pouvoir être utilisés dans un contexte biologique :

- La structure de la biomolécule étudiée ne doit pas être fortement perturbée par le groupement photo-réactif ce qui diminuerait l'affinité de la sonde pour sa cible.
- Ils doivent être chimiquement inertes et stables dans des conditions biologiques
- Ils doivent réagir uniquement à des longueurs d'ondes qui n'induiront pas de dégradation d'autres molécules biologiques, typiquement à des longueurs d'ondes supérieures à 300 nm.
- Le temps de réaction doit être court, si possible inférieur au temps de dissociation du complexe à étudier. La réaction doit pouvoir former des liens covalents avec n'importe quel résidu ou groupement chimique.

Il n'existe pas d'agent de photomarquage idéal capable de remplir tous ces critères. Malgré tout, trois fonctions principales se dégagent grâce à leurs formes particulièrement réactives après irradiation : les azotures d'aryle<sup>157</sup>, les diazirines<sup>158</sup> et la benzophénone<sup>159</sup> (Figure 24).

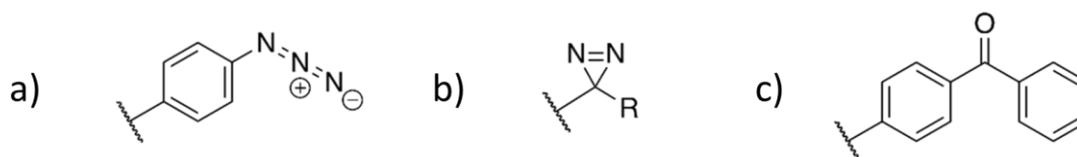


Figure 24: Groupements photoactivables les plus couramment utilisés: (a) azoture d'aryle; (b) diazirine; (c) benzophénone

### 3.2.2.1 Les azotures d'aryle

L'utilisation des azotures d'aryle a été reportée dès 1969 lors de l'identification de sites d'interactions d'un anticorps avec un haptène photoactivable<sup>157</sup>. Ils sont, depuis lors, très utilisés du fait des nombreux avantages dont ils disposent. Le principal est leur taille réduite qui leur permet d'être facilement introduits au sein de composés biologiques naturels aromatiques comme la phénylalanine<sup>160</sup> ou l'adénine<sup>161,162</sup>, sans pour autant altérer significativement la structure de ces composés.

La réactivité des azotures d'aryle est intéressante et se base sur la réactivité des nitrenes. En effet, suite à la phase d'irradiation, le groupement azoture (**1**) va perdre une molécule de diazote entraînant la formation d'un nitrene singulet **2** (Figure 25). Celui-ci va pouvoir réagir selon trois mécanismes réactionnels différents.

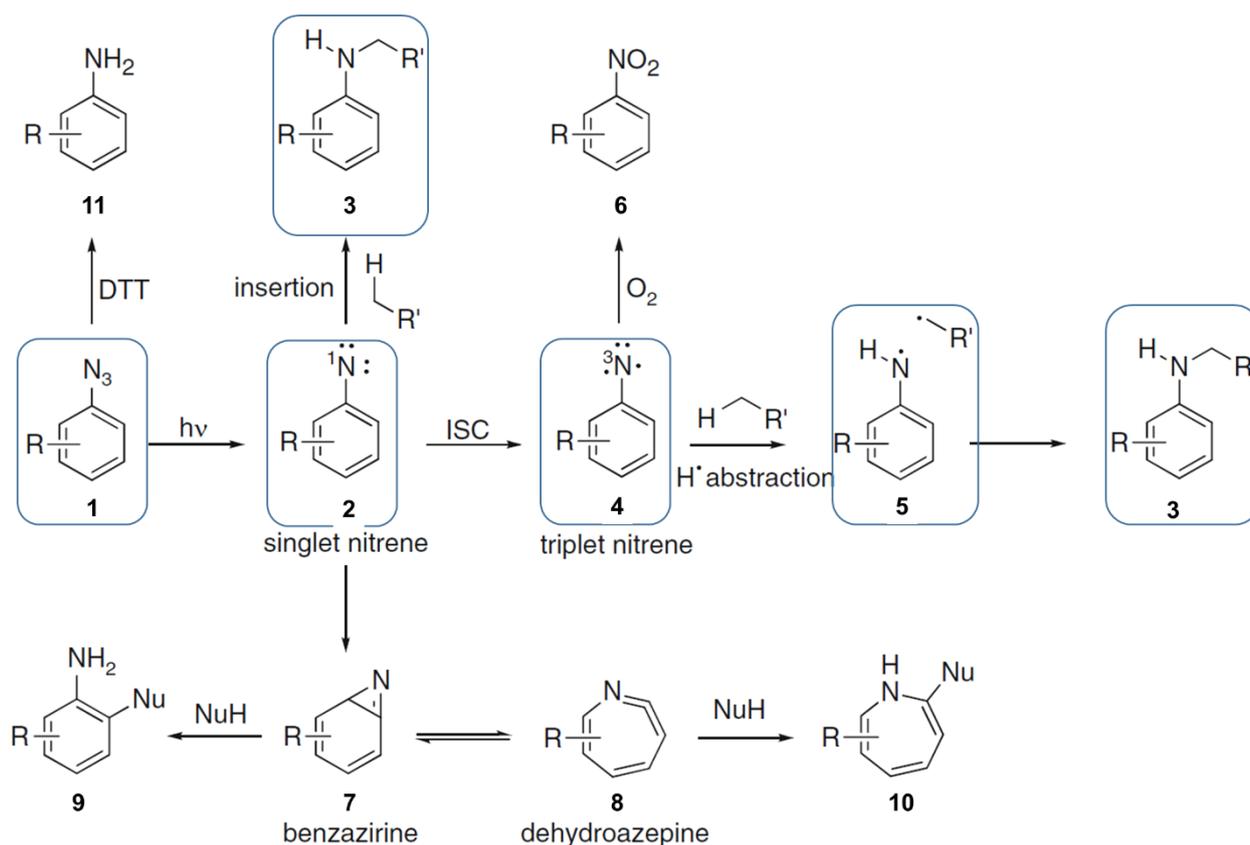


Figure 25 : Principaux mécanismes des réactions des azotures d'aryles activés. Le chemin réactionnel encadré en bleu représente la réaction de pontage désirée. Les autres correspondent à des réactions secondaires. D'après Geurink et al.<sup>163</sup>

Premièrement, le nitrene singulet **1** va pouvoir réaliser une insertion RH au sein de liaisons C-H ou X-H – X étant un hétéroatome- présente dans la molécule d'intérêt. La molécule pontée **3** est alors obtenue. Une seconde voie réactionnelle permet la formation de la molécule **3**. Le nitrene singulet va rapidement être converti en nitrene triplet **4** par un mécanisme de croisement

intersystème (Intersystem Crossing, ISC). Il va ensuite pouvoir réaliser une abstraction d'hydrogène à partir d'une liaison R-H formant un radical R· qui va se coupler sur l'atome d'azote (**5**) conduisant à la formation de **3**. L'avantage de ces deux mécanismes est la formation d'un lien covalent non spécifique, c'est-à-dire qu'il peut s'établir sur n'importe quel groupement fonctionnel présent dans les biomolécules. Toutefois, lorsque l'établissement du pont covalent passe par le nitrene triplet, celui-ci peut réagir de façon parasitaire avec l'oxygène<sup>163</sup>, formant ainsi le dérivé nitro **6** et réduisant le rendement de photomarquage.

Pour ce qui est de la troisième réaction possible du nitrene singulet, il s'agit d'un réarrangement intramoléculaire rapide qui aboutit à la formation d'une forme benzazirine (**7**). Cette forme est en équilibre avec une forme déhydroazépine (**8**) par expansion/réduction de cycle<sup>164</sup>. Les molécules **7** et **8** sont des molécules électrophiles particulièrement réactives envers des nucléophiles. On peut alors obtenir un pontage uniquement au niveau de groupements nucléophiles des biomolécules (**9** et **10**), les chaînes latérales de certains acides aminés par exemple. Pour éviter cette réactivité, on prône l'utilisation d'azotures d'aryles fluorés qui vont stabiliser le nitrene singulet, rendant ainsi la transformation en benzazirine plus lente, favorisant la réaction d'insertion NuH.

Mise à part la réaction parasite des nitrenes triplet avec l'oxygène<sup>163</sup>, les azotures d'aryles peuvent être réduits en amine **11** par des dithiols comme le DTT (Dithiothréitol), et de façon moins significative par les thiols<sup>165</sup>. Cette réaction secondaire peut s'avérer problématique puisque les thiols, comme le glutathion, sont présents dans tous les milieux biologiques et les dithiols sont couramment utilisés comme agents réducteurs. Tout cela limite l'utilisation des azotures d'aryles dans des conditions expérimentales classiques. Une autre limite de ces agents de photomarquage est leur longueur d'onde d'excitation qui est inférieure à 300 nm, pouvant ainsi endommager les protéines et les acides nucléiques présents en solution. Il existe certains dérivés/analogues d'azotures d'aryles où le cycle aromatique est substitué par des groupements nitro, imino ou encore acyl, qui ont des longueurs d'onde maximales légèrement augmentées, remédiant ainsi à cette limite.

Les azotures d'aryles, bien qu'ayant des rendements de photomarquage assez faibles causés par les nombreuses réactions secondaires possibles après irradiation, et nécessitant l'utilisation de longueurs d'onde d'activation difficilement compatibles avec des conditions biologiques, restent couramment utilisés. Cette popularité est basée sur la relative facilité de leur synthèse et de leur incorporation.

### 3.2.2.2 Les diazirines

Les diazirines ont été développées peu de temps après les azotures d'aryles. La première description de diazirines en tant qu'agent de photomarquage date de 1973 avec la 3-aryle-3H-diazirine<sup>158</sup> **12** (Figure 26). Depuis lors, et après plusieurs phases d'optimisation les trifluorométhylaryldiazirines<sup>166</sup> **13** se sont révélées être les plus efficaces et sont couramment utilisées.

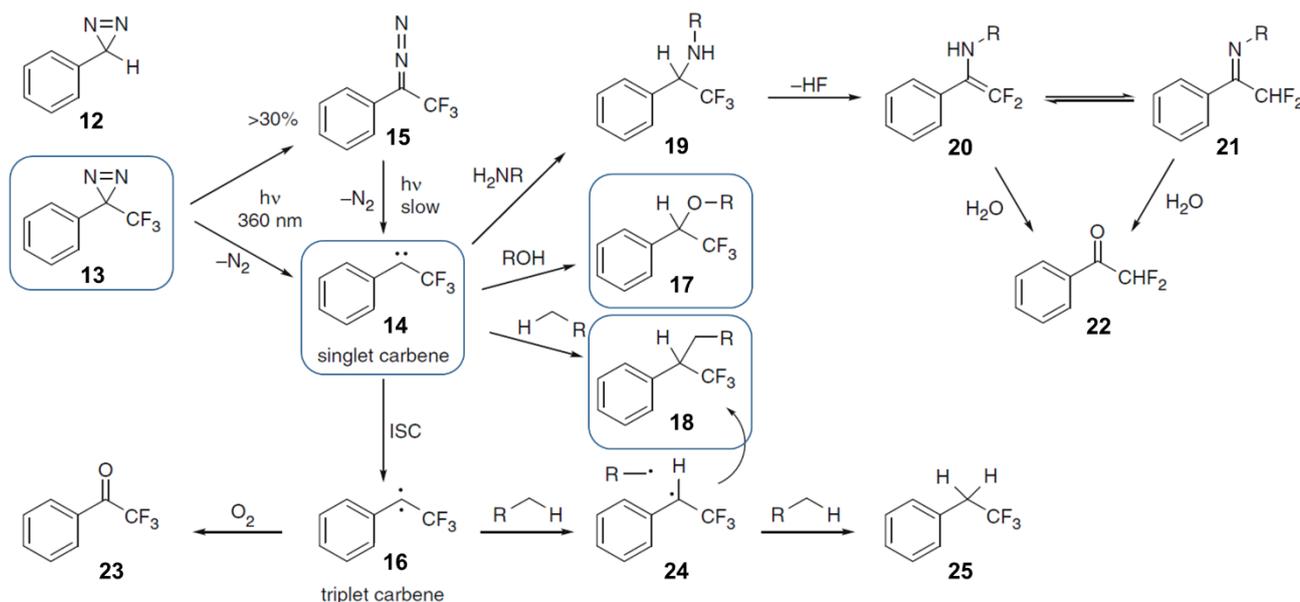


Figure 26 : Mécanismes réactionnels de l'excitation de la trifluorométhylaryldiazirine. Le chemin réactionnel encadré en bleu représente la réaction de pontage désirée. Les autres correspondent à des réactions secondaires. D'après Geurink et al.<sup>163</sup>

La longueur d'onde d'activation des groupements diazirines est comprise entre 320 et 380 nm, ce qui est bien au-dessus de la limite physiologique des 300 nm. Ces groupements ne risquent donc pas d'induire des dommages significatifs lors de leur activation. Une fois activée par irradiation, la diazirine **13** perd une molécule de diazote formant un carbène singulet **14** (Figure 26). Toutefois, environ 30 % de la diazirine initiale est convertie en une forme diazoisomère linéaire **15**. Ce composé peut lui aussi former le carbène singulet sous l'influence d'un rayonnement, cela dit aux longueurs d'onde typiquement utilisées pour les diazirines, ce processus est relativement lent. L'espèce diazoisomère, ayant un temps de demi-vie assez long, va pouvoir être hydrolysé, diminuant le nombre d'entités qui pourront réagir avec les biomolécules ciblées. Ce problème est en partie inexistant pour les trifluorométhylaryldiazirines **13**, car les fortes propriétés attractrices d'électrons du groupement trifluorométhyle stabilisent le diazoisomère<sup>166</sup>, ce qui le rend moins photolabile. Le carbène singulet **14** est une espèce ayant un temps de demi-vie très court : environ 1 ns, qui va pouvoir être transformé en carbène triplet **16** par conversion intersystème. Un carbène singulet peut

réagir en tant qu'électrophile, nucléophile ou bien les deux en fonction de la nature de ses substituants, pour ce qui est de l'état triplet celui-ci va se comporter comme un di-radical.

Le carbène singulet **14** va réaliser très rapidement des réactions d'insertion sans réelle discrimination entre différents sites R-H. Cependant, les insertions au sein de groupement hydroxyles **17** donnent de meilleurs rendements que les insertions au sein de liaison C-H<sup>167</sup> **18**. Cette préférence réactionnelle est problématique puisqu'elle entraîne l'inactivation partielle des diazirines en milieu aqueux. Toutefois, cette réaction permet le pontage des acides aminés serine et tyrosine qui possèdent un groupement hydroxyle. L'insertion au niveau d'une amine primaire **19** ou secondaire peut mener à une réaction secondaire non désirée en milieu aqueux (Figure 26). La molécule **19** va rapidement perdre une molécule de HF pour former l'énamine **20** en équilibre avec sa forme imine **21**. Ces formes vont ensuite être hydrolysées pour donner la cétone correspondante **22** ce qui va conduire à la rupture du lien avec le substrat photo-ponté.

Après le changement de l'état de spin du carbène, c'est-à-dire le passage de l'état singulet à triplet, ce dernier, réagissant comme un di-radical, va pouvoir soit réagir avec du dioxygène (**23**) soit réaliser une abstraction d'hydrogène sur le substrat conduisant à l'intermédiaire **24**. Si ce dernier réagit avec le radical formé, on obtient alors une liaison covalente carbone-carbone **18**. Par contre, s'il réalise une seconde abstraction d'hydrogène, on obtient le composé réduit **25**.

L'utilisation de sonde à base de diazirine peut être donc limitée par des réactions secondaires qui vont diminuer leur rendement de photomarquage. De plus, la synthèse et l'incorporation de ce type de fonction chimique nécessite souvent des étapes de synthèse plus contraignantes par rapport aux aryles azides par exemple.

Il existe aujourd'hui de nombreux blocs commerciaux permettant l'incorporation de ces motifs chimiques dans des biomolécules. Contrairement aux azotures d'aryles, les trifluorométhylaryldiazirines modifient la structure des biomolécules auxquelles ils sont greffés. Cela dit, la stabilité de ces agents dans une variété de milieu : acide, basique, oxydant ou réducteur, ainsi que leur grande réactivité représentent deux avantages importants pour leur utilisation dans des milieux biologiques.

### 3.2.2.3 La benzophénone

Tout comme les diazirines, la première description de la benzophénone comme agent de photomarquage date de 1973<sup>159</sup>. Peu de temps après, le photomarquage de l'albumine sérique bovine (BSA) a été réalisé à l'aide de ces sondes<sup>168</sup>. Depuis lors, les benzophénones ont été largement utilisées et incorporées dans différents type de biomolécules telles que des peptides<sup>169,170</sup>, des inhibiteurs<sup>171</sup>, ou encore des acides nucléiques<sup>172</sup>.

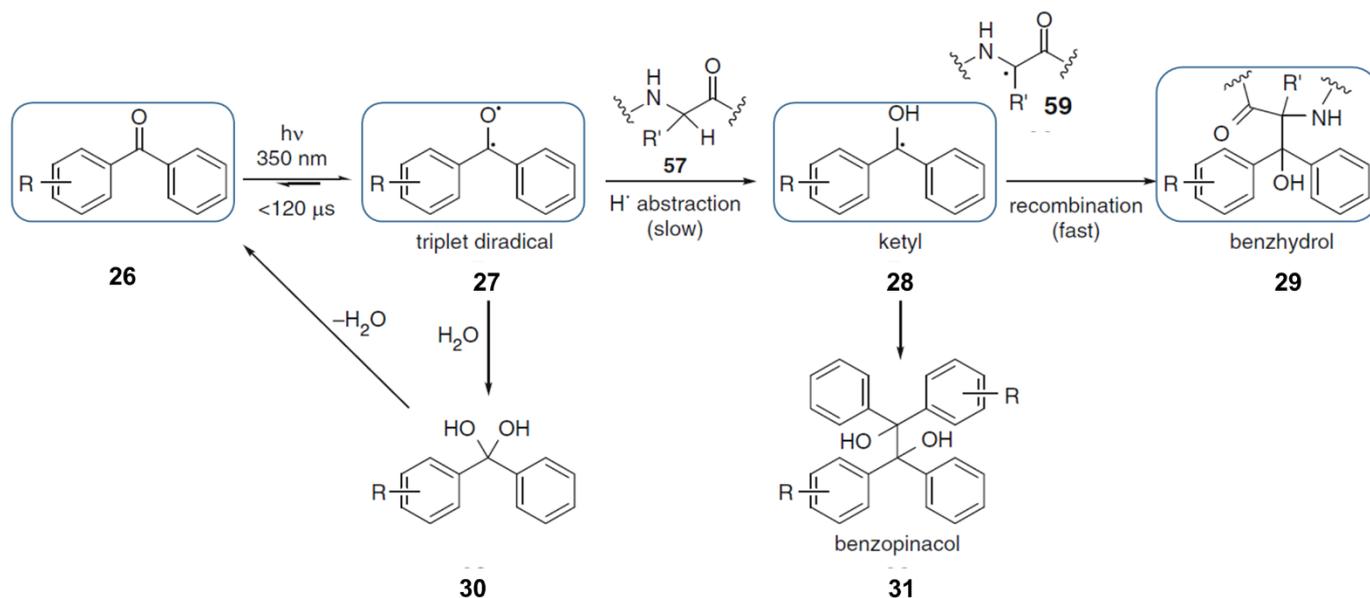


Figure 27 : Principaux mécanismes de réaction de la benzophénone irradiée. Le chemin réactionnel encadré en bleu représente la réaction de pontage désirée. Les autres correspondent à des réactions secondaires. D'après Dorman et al.<sup>173</sup>

Les benzophénones **26** sont activées à des longueurs d'ondes comprises entre 350 et 365 nm. Le mécanisme possible de l'activation et la réactivité des benzophénones après irradiation sont montrés sur la Figure 27. L'absorption d'un photon par une benzophénone, aboutit initialement à la formation rapide d'un di-radical dans un état triplet **27**. La formation de ce di-radical triplet est réversible puisque celui-ci ne peut pas exister plus de 120  $\mu$ s avant de se relaxer pour revenir à l'état fondamental **26**, s'il ne trouve pas de partenaire avec qui réagir. La première étape de la chaîne réactionnelle menant aux espèces photo-pontées correspond à une abstraction d'hydrogène (**28**). Cette étape est limitante et le rendement de la réaction est dépendant de la nature du lien X-H à proximité<sup>173</sup>. La deuxième étape correspond à la création d'une liaison C-X avec le radical carboné formé lors de la première étape (**29**).

La benzophénone est plus réactive vis-à-vis des liaisons C-H que des liaisons O-H ou N-H. Cela représente un avantage évident pour cet agent photoactivable, puisqu'il va pouvoir se lier à n'importe quel acide aminé. Un autre avantage de la benzophénone est le nombre limité de

réactions secondaires induites par le di-radical. En effet, si le di-radical réagit avec une molécule d'eau pour former l'hydrate **30**, cette espèce va rapidement se réarranger en cétone, reformant ainsi une benzophénone, qui pourra être recyclée en di-radical sous irradiation. Une homodimérisation du monoradical **31** est aussi possible mais la proximité des deux radicaux formés après l'abstraction d'hydrogène, ainsi que la rapidité de la recombinaison de ceux-ci, limite fortement cette voie réactionnelle. La benzophénone est le seul agent photoactivable capable de se régénérer au cours de l'irradiation, cela présente notamment l'avantage d'augmenter le rendement du photomarquage au cours du temps.

Cette capacité de régénération est un avantage mais peut aussi être un inconvénient. En effet, si la benzophénone n'est pas désactivée à temps ou est trop longtemps irradiée, celle-ci va pouvoir se déplacer dans le milieu réactionnel augmentant ainsi la probabilité de réaliser des pontages non spécifiques. La benzophénone présente d'autres inconvénients comme le temps d'irradiation nécessaire pour obtenir un rendement de photo-marquage acceptable qui est plus élevé que pour n'importe quelle autre fonction photoactivable<sup>174</sup>. On peut aussi citer la taille et la rigidité de la molécule qui risque de modifier la structure de la biomolécule modifiée par cet agent de photomarquage.

La benzophénone est un agent photoactivable très couramment utilisé, notamment grâce à son unique capacité de régénération ainsi qu'à son incorporation facilitée par la disponibilité de nombreux blocs commerciaux permettant un couplage sur tous types de composés.

Les diazirines aryles et les benzophénones induisent des modifications structurelles au sein des biomolécules dans lesquelles ils sont incorporés, pouvant entraîner une altération des interactions avec le partenaire d'étude. Dans le but de pallier ce problème, ces agents sont souvent couplés aux biomolécules à l'aide d'un bras espaceur. Cette stratégie présente l'avantage de minimiser l'altération de la biomolécule puisque les groupements induisant une forte gêne stérique seront éloignés. Cela signifie aussi que le groupement photoactivable sera en retrait par rapport au site d'interaction. Celui-ci pourra toujours se lier au partenaire direct d'interaction ainsi qu'à d'autres biomolécules pouvant former un complexe interactif, mais la distance imposera aussi une plus grande probabilité de former des ponts avec des partenaires non spécifiques. Pour éviter cela, on peut avoir recours à des agents photo-activables dits « zero-length » en anglais, c'est-à-dire au cœur même de l'interaction et qui perturbent le moins

possible la structure des biomolécules comme par exemple les agents de photomarquage analogues de nucléobases.

### 3.2.3 Les agents de photomarquage analogues de nucléobases

Les agents de photomarquage analogues de nucléobases sont spécifiquement utilisés dans le cadre d'études portant sur les interactions ADN (ou ARN)-protéines. Leurs réactivités se basent sur les propriétés photo-chimiques de certaines bases modifiées de l'ADN tels que des dérivés halogénés ou des dérivés thiolés. On peut aussi ajouter à cette deuxième classe des dérivés nucléosidiques de type arylazides comme la 2-azido-adénine, la 8-azido-adénine, 8-azido-guanine<sup>175</sup> ou bien encore la 5-azido-uridine<sup>176</sup>(Figure 28) qui vont donc réagir selon le principe évoqué dans le paragraphe 3.2.2.1, p.53.

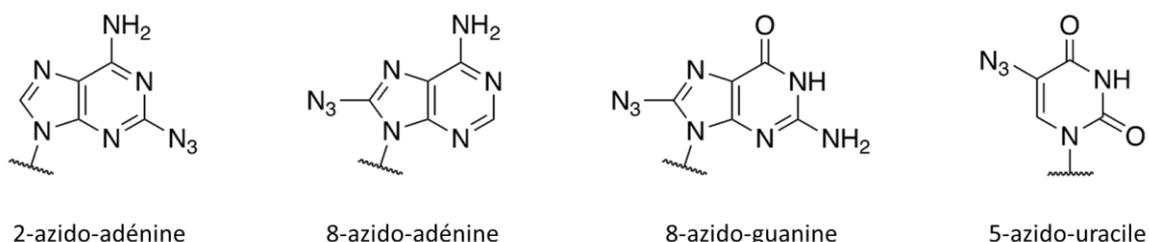


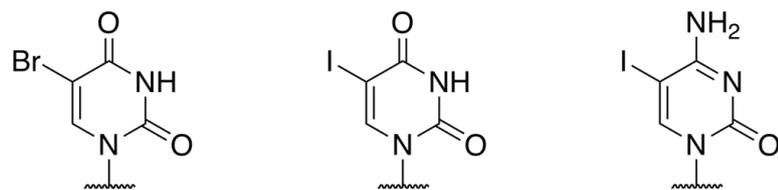
Figure 28 : Groupement photoactivables puriques et pyrimidiques de type arylazides.

#### 3.2.3.1 Les dérivés halogénés

Les premiers analogues de nucléobases utilisés pour leur activité de photomarquage ont été les dérivés bromés de l'uracile. Parmi eux, on peut citer la 5-bromouracile (5BrU) (Figure 29). Ceux-ci sont particulièrement intéressants notamment de par leur longueur d'onde d'excitation qui est proche ou supérieure à 300 nm<sup>177</sup>. De plus, le rayon de Van der Waals d'un atome de brome (1,95 Å) est similaire au rayon d'un groupement méthyle (2,0 Å), cela rend la 5BrU et la 5BrdU comparable à une thymine en taille. En conséquence, les molécules d'ADN contenant des dérivés bromés vont être reconnues avec la même affinité par les protéines spécifiques de l'ADN que l'ADN lui-même.

L'utilisation des dérivés bromés a connu un fort essor dans les années 1970 jusqu'aux années 1990. Ils ont permis l'identification de points de contact entre une séquence d'ADN de l'opéron lactose et le répresseur de cet opéron<sup>178</sup>, ils ont aussi permis la caractérisation du site de liaison de la sous-unité spécifique de l'enzyme de restriction EcoK1<sup>179</sup> ou encore l'étude de la chaîne

alpha et bêta de la protéine de liaison aux télomères et de l'ADN télomérique d'*Oxytricha nova*<sup>177</sup>.



5-Bromo-uracile (5BrU)    5-iodo-uracile (5IU)    5-iodo-cytosine (5IC)

Figure 29 : Différents dérivés halogénés pyrimidiques.

La bromouridine constitue la plus populaire des pyrimidines halogénées utilisées dans le cadre d'expériences de photomarquage. Cependant, la longueur d'onde d'activation proche des 300 nm, et son faible rendement de photomarquage n'en fait plus une sonde de choix. La substitution de l'atome de brome par un atome d'iode, conduisant à la 5-iodouracile (5IU) ou encore la 5-iodocytosine (5IC) (Figure 29), permet d'optimiser la sonde. En effet, la longueur d'onde d'activation des dérivés iodés est de 325 nm et les rendements de photomarquage peuvent aller jusqu'à 90 %<sup>180</sup>. Malgré cela, l'incorporation d'un atome d'iode peut être défavorable puisque le rayon de Van der Waals de cet atome (2,15 Å) est supérieur à celui d'un groupement méthyle. Il a d'ailleurs été prouvé que la présence de plusieurs dérivés iodés de l'uridine au sein d'un brin d'ARN pouvait causer une augmentation de la constante de dissociation de celui-ci<sup>181</sup>. Tout comme les dérivés bromés, leurs analogues iodés ont permis la caractérisation de différentes interactions : entre des ARN viraux et les protéines R17 de bactériophage<sup>180</sup>, ou encore entre la protéine RecA d'*E. Coli* et de l'ADN simple brin<sup>182</sup>.

Ces deux classes de dérivés halogénés présentent deux mécanismes photochimiques différents.

Pour ce qui est de la 5BrU, l'irradiation à 308 nm donne lieu à un état singulet de type  $\pi-\pi^*$  ou  $n-\pi^*$ . L'état singulet de type  $\pi-\pi^*$  peut induire l'homolyse de la liaison C-Br et se convertir en un état triplet par un croisement intersystème<sup>183</sup>, tout comme l'état singulet de type  $n-\pi^*$ . Ce triplet va ensuite réagir uniquement par mécanisme de transfert d'électron pour former une liaison covalente avec une molécule proche. Dans le cas de la 5IU, l'absorption d'un photon entraîne la rupture homolytique de la liaison C-I, amenant la création d'un état singulet qui va directement réagir avec les protéines présentes à proximité par un mécanisme radicalaire<sup>184</sup>. Si l'on compare l'efficacité de photomarquage entre ces deux dérivés halogénés, celle-ci est 2 à 5 fois supérieure pour la 5IU<sup>180</sup>.

La sélectivité de ces dérivés pour certains acides aminés a été établie au cours de nombreuses études (Figure 30). La tyrosine<sup>180</sup>, le tryptophane, l'histidine<sup>185</sup>, la cystéine<sup>186</sup>, la lysine<sup>187</sup> et l'alanine<sup>188</sup> vont réagir avec la 5BrU. Dans le cas de la 5IU, elle va préférentiellement réagir avec la tyrosine, la phenylalanine<sup>180</sup>, l'histidine et la méthionine<sup>182</sup>.

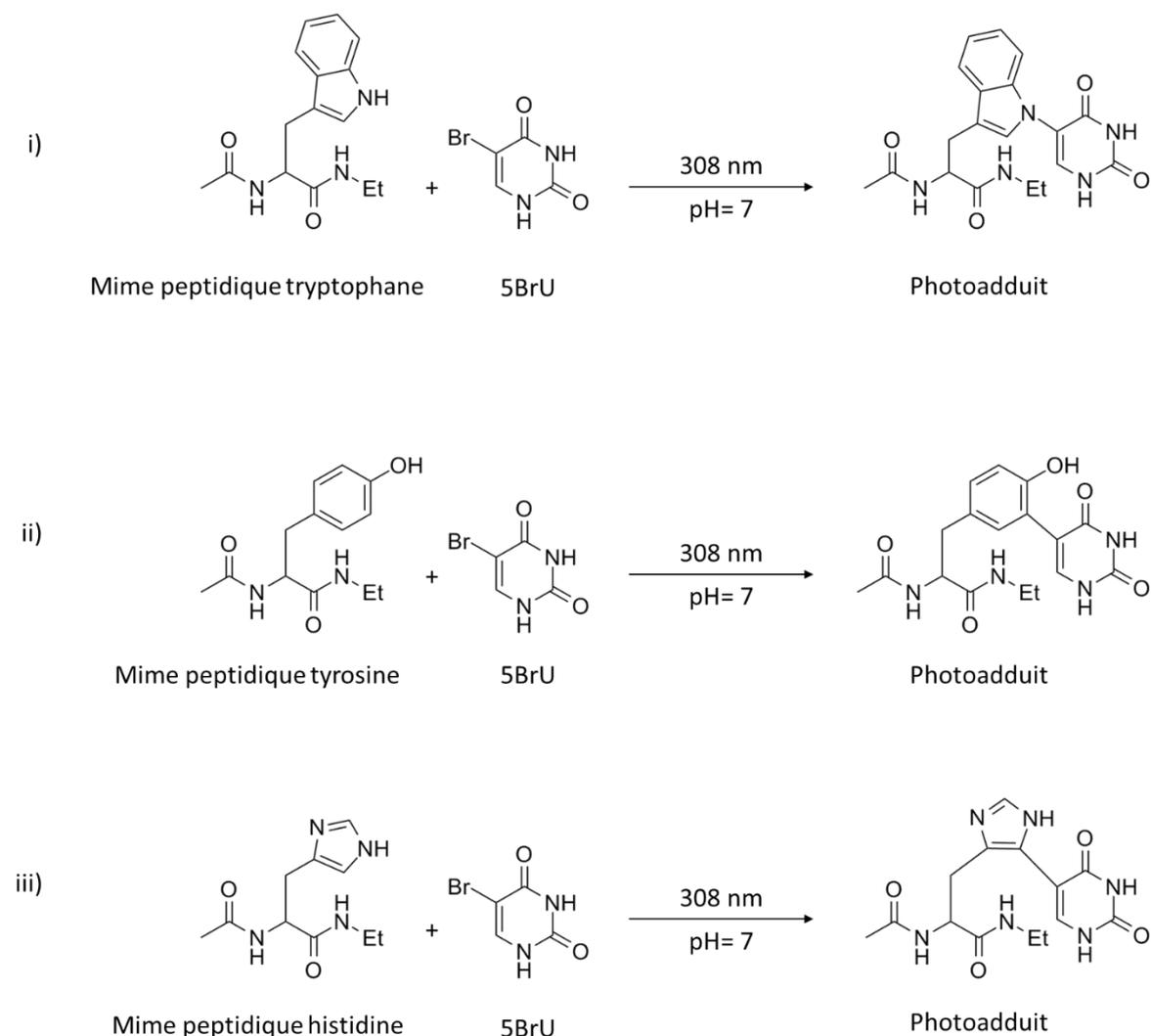


Figure 30 : Réactions photochimiques de la 5BrU avec des mimes peptidiques: i) de tryptophane, ii) de tyrosine et iii) d'histidine d'après Dietz et al<sup>185</sup>.

### 3.2.3.2 Les dérivés de nucléobases thiolés

La 4-thiouridine (4<sup>S</sup>U) est un nucléoside naturel, rare que l'on retrouve par exemple au niveau des ARN de transfert (ARNt) d'*E.Coli*<sup>189</sup> mais qui est absente chez les cellules eucaryotes. Il s'agit d'un analogue de l'uridine où l'atome d'oxygène en position 4 du cycle pyrimidique est remplacé par un atome de soufre. Cela lui apporte la capacité d'être photo-activée à une longueur d'onde supérieure à 300 nm. La présence naturelle de la 4<sup>S</sup>U dans l'ARNt a été mise

à profit pour une étude structurale de l'ARNt par photomarquage<sup>190</sup>. Elle a aussi servi d'inspiration pour la synthèse de deux autres dérivés thiolés : la 4-thiothymidine (4<sup>S</sup>T) et la 6-thio-2'-désoxyguanosine (6<sup>S</sup>G).

L'analogie structurale de ces bases leur permet de s'apparier avec les bases complémentaires de l'ADN *via* des liaisons hydrogène classiques de type Watson-Crick. Il a aussi été montré que

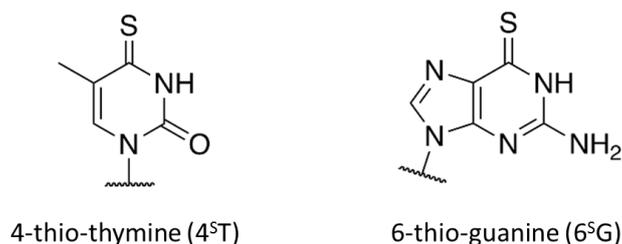


Figure 31 : Structure des bases 4<sup>S</sup>T et 6<sup>S</sup>G.

l'incorporation d'un atome de soufre à la place d'un atome d'oxygène n'entraînait qu'une variation de 0,45 Å du rayon de Van der Waals<sup>191</sup>. Cette faible variation structurale n'a qu'un effet limité sur la stabilité des duplexes d'ADN<sup>192</sup>. Cependant, ce n'est plus le cas lorsque le taux de substitution en dérivés thiolés est élevé<sup>193,194</sup>. Un autre avantage des dérivés thiolés est leur longueur d'onde d'activation supérieure à 300 nm–330 nm pour la 4<sup>S</sup>U par exemple. De plus, on peut aussi citer la bonne stabilité des oligonucléotides « thiolés » notamment ceux avec de la 4<sup>S</sup>T qui peuvent se conserver des années à l'abri de la lumière.

Un avantage non négligeable de ces bases modifiées est leur réactivité. L'agent de photomarquage 4<sup>S</sup>T ainsi que les autres analogues thiolés de nucléobases (4<sup>S</sup>U, 6<sup>S</sup>G) ont des comportements photochimiques similaires (Figure 32).

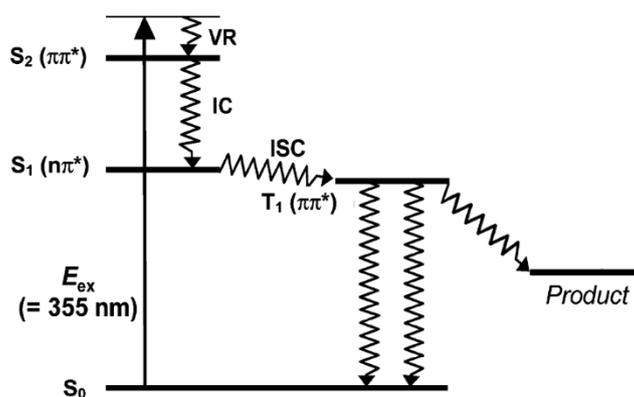


Figure 32 : Diagramme de Jablonsky de la 4<sup>S</sup>T représentant ses différents niveaux énergétiques. D'après Harada et al<sup>195</sup>.

Lors de l'irradiation de la 4<sup>S</sup>T et des autres bases thiolées, l'absorption d'un photon permet de passer de l'état fondamental S<sub>0</sub> à un état excité qui va par relaxation vibrationnelle (VR) rapidement peupler l'état excité singulet S<sub>2</sub> (π-π\*). Cet état va rapidement se relaxer par

conversion interne (IC) vers l'état excité singulet S1 ( $n-\pi^*$ ), avant que ce dernier ne se transforme quasi totalement en état excité triplet T1 ( $\pi-\pi^*$ ) par un mécanisme de conversion intersystème<sup>195</sup> (ISC). Ces deux dernières étapes sont très rapides, de l'ordre de quelques picosecondes. Le haut rendement d'interconversion -le rendement quantique d'ISC est de  $0,9 \pm 0,1$ - entre l'état singulet et l'état triplet, ainsi que le temps de vie très court de l'état singulet permettent de déduire que l'état triplet est la forme photo-réactive de ces composés thiolés. C'est donc cet état triplet qui va établir des liaisons covalentes avec les molécules à proximité. Les triplets ont tendance à fortement réagir avec le dioxygène sous forme singulet  $^1O_2$  (Figure 83). Cette réaction particulièrement bien décrite dans le cas où la  $4^S T$  est utilisée dans des traitements anticancéreux (voir revue<sup>196</sup>) correspond à une réaction secondaire et parasite dans le cas où l'on souhaite capturer une protéine, puisqu'il y aura moins de triplet disponible pour réagir avec un acide aminé à proximité.

Une fois sous forme d'espèce réactive, les dérivés thiolés vont pouvoir réagir pour former des liaisons covalentes. On ne sait que peu de chose sur la réactivité de ces espèces vis-à-vis des protéines et leurs constituants, les acides aminés. Par contre, la photochimie de ces agents photo-activables est bien décrite pour les acides nucléiques.

#### 3.2.3.2.1 La photochimie sur les bases nucléiques

La plupart des études portant sur les mécanismes photochimiques ont été effectuées sur des dinucléotides, dont l'un correspond à un dérivé thiolé, typiquement la  $4^S U$  (voir revue<sup>197</sup>). L'agencement de la base modifiée – soit en position 3', soit en position 5'- va changer la réactivité de cette dernière. On observe deux types de réactivités différentes des dérivés thiolés en position 4 : un comportement électrophile ou un comportement nucléophile de type cycloaddition [2+2] :

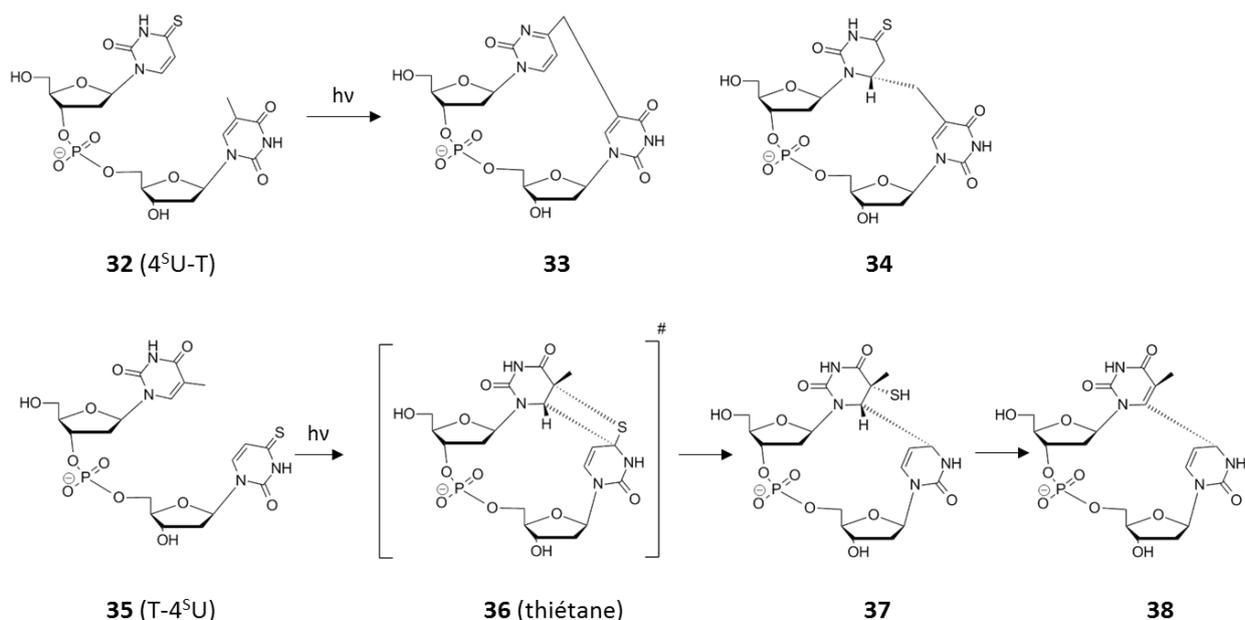


Figure 33 : Structure des photoadduits obtenus après irradiation des dinucléotides 4<sup>S</sup>U-T (haut) et T-4<sup>S</sup>U (bas). Figure adaptée de Favre et al. <sup>197</sup>

- Lorsque la 4<sup>S</sup>U est positionnée en 5' au sein d'un dinucléotide comprenant une thymine (**32**) (Figure 33, haut), elle va se comporter comme un électrophile. Dans ce cas-là, le groupe méthyle de la thymine va perdre un atome d'hydrogène lors de l'irradiation. Le radical méthylène formé va ensuite pouvoir s'additionner soit en position 4 soit en position 6 du radical dérivé de la base thiolée (Figure 33, molécules **33** et **34**). Cette réaction s'apparente à une addition 1,2 (**33**) ou 1,4 (**34**) sur une thiocétone insaturée. On peut supposer que l'ajout d'un groupement méthyle en position 5 de la 4<sup>S</sup>U (qui devient ainsi une 4<sup>S</sup>T), favorise l'addition en position 6 par une augmentation de la gêne stérique au niveau de la position 4. Il est à noter que l'addition 1,2 s'accompagne d'une perte de l'atome de soufre.
- Dans le cas inverse, c'est-à-dire la 4<sup>S</sup>U en position 3', celle-ci va être impliquée dans une cycloaddition de type [2+2] avec la double liaison C5-C6 de la thymine en 5' (**35**) (Figure 33, bas). Lors de l'approche parallèle des liaisons C4-S et C5-C6 un intermédiaire thiétane **36** va se former. Cet intermédiaire peut être en équilibre ou directement conduire à sa forme ouverte **37**. Une dernière étape d'élimination peut amener à la formation de la molécule **38**, par perte d'un groupement SH<sub>2</sub>.

### 3.2.3.2.2 La photochimie sur les acides aminés

Comme précédemment évoqué la photochimie des dérivés thiolés avec les acides aminés est peu connue. Les deux seuls exemples d'irradiation de dérivés de la  $4^S\text{U}$  vis-à-vis d'un acide aminé ont été réalisés *in vitro* par Ito *et al.* Ils ont postulé un premier mécanisme réactionnel entre une 1,3-diméthyl-4-thiouracile et une lysine en 1978<sup>198</sup> (Figure 34), où le dérivé thiolé se comporte comme un électrophile. En effet, l'amine de la chaîne latérale de la lysine s'additionne selon un mécanisme de type 1,4 sur la thio-cétone. Deux années plus tard, Ito *et al.* ont réussi à caractériser le produit de leur réaction par RMN<sup>199</sup> (Figure 34). Cette caractérisation leur a permis de montrer que l'addition nucléophile de la lysine était effectuée préférentiellement par l'amine en  $\alpha$  plutôt que l'amine de la chaîne latérale. Si ce mécanisme est effectif dans un contexte protéique, il faut envisager que cette addition nucléophile se fait via l'amine de la chaîne latérale de la lysine, l'autre étant impliquée dans une liaison peptidique.

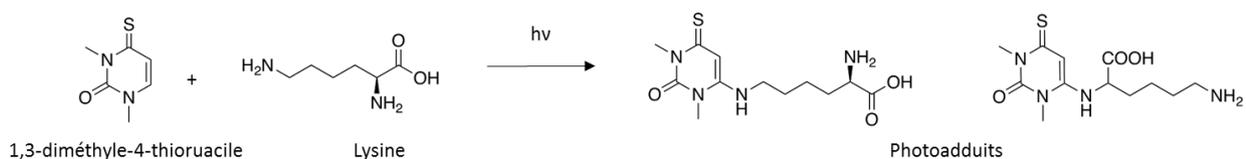


Figure 34 : Photoréaction de la 1,3-diméthyle-4-thiouracile avec la lysine d'après Ito *et al.*<sup>199</sup>. Première (A) et seconde (B) chimiosélectivité mécanistique.

Il existe donc de nombreux groupements photoactivables, certains spécifiques à l'étude des interactions mettant en jeu des acides nucléiques, tous possédant des avantages et des inconvénients et lors de la mise en place d'une étude utilisant le photomarquage, il est souvent judicieux de tester plusieurs groupements photoactivables afin de choisir celui qui convient le mieux à l'étude.

3.3

#### **Incorporation des agents photoactivables dans des sondes oligonucléotidiques.**

L'introduction d'une fonction photoactivable au sein d'une séquence d'ADN peut se faire de différentes manières. Elle peut être effectuée par voie enzymatique ou par voie chimique.

### 3.3.1 La synthèse enzymatique

L'incorporation des groupements photoactivables par voie enzymatique se fait via des dérivés triphosphates comportant ces groupements. Les enzymes impliquées dans la synthèse des acides nucléiques (polymérase) vont prendre en charge ces dérivés triphosphates photoactivables et les incorporer au sein de l'ADN ou de l'ARN. Plusieurs synthèses de dérivés triphosphates ont été répertoriés<sup>200</sup>. On peut notamment citer la synthèse du triphosphate d'uridine modifié par un groupement diazirine (U<sup>DZ</sup>TP) en position 5 de la base réalisé par C. Smith *et al.* (Figure 35)<sup>201</sup>.

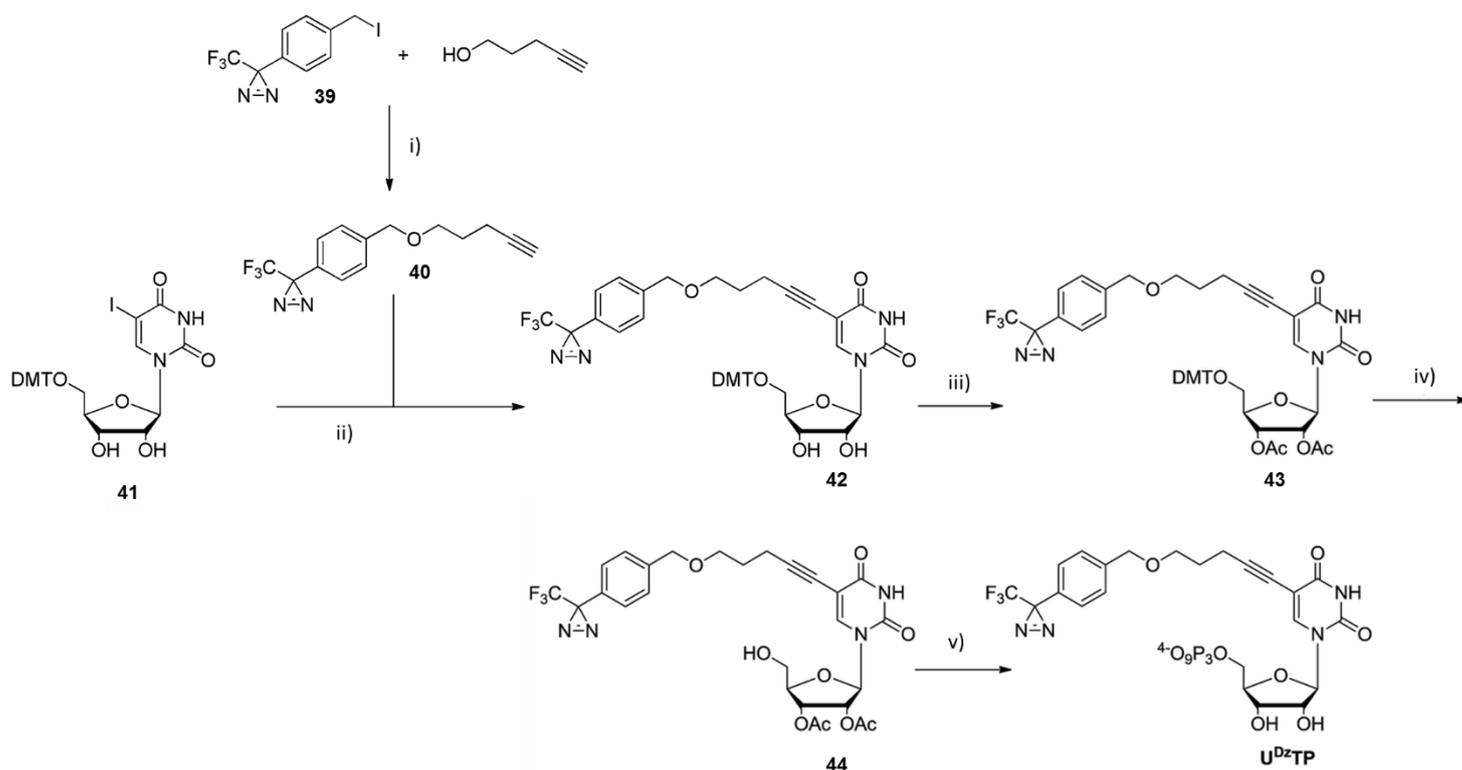


Figure 35 : Synthèse du triphosphate U<sup>DZ</sup>TP d'après Smith *et al.*<sup>202</sup>. Les réactifs et les conditions sont les suivantes : (i) 4-pentyn-1-ol, NaH, THF, 0 °C, 2 h, 68%; (ii) Pd(PPh<sub>3</sub>)<sub>2</sub>Cl<sub>2</sub>, Et<sub>3</sub>N, CuI, DMF, r.t., o.n., 84%. (iii) Ac<sub>2</sub>O, pyridine, DMAP, 0 °C to r.t., 2 h, 71%; (iv) 1% DCAA, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C, 1.5 h, 48%; (v) (1) 2-chloro-1,3,2-benzodioxaphosphorin-4-one, pyridine, dioxane; (2) (nBu<sub>3</sub>NH)<sub>2</sub>H<sub>2</sub>P<sub>2</sub>O<sub>7</sub>, nBu<sub>3</sub>N, DMF; (3) I<sub>2</sub>, pyridine, H<sub>2</sub>O; (4) NH<sub>3</sub> (aq.), H<sub>2</sub>O, 21% (sur les 4 dernières étapes)

Au cours de cette synthèse, la diazirine **39** est couplée à un bras espaceur 4-pentyn-1-ol en présence de NaH, avant la réalisation d'un couplage de Sonogashira (**40**) sur la 5-iodouridine protégée **41** qui permet de générer **42**. Cette molécule est ensuite bis-acétylée pour former le composé **43**. Après élimination du groupement protecteur DMT en 5', le composé **44** est ensuite converti en U<sup>DZ</sup> triphosphate par application de la réaction de Ludwig-Eckstein<sup>202</sup>.

Les dérivés triphosphates des analogues nucléosidiques photoactivables peuvent être incorporés de deux façons au sein d'une molécule d'ADN :

- Par réaction en chaîne par polymérase (PCR, Polymerase Chain Reaction) : il s'agit d'une méthode qui permet de dupliquer en grand nombre une séquence d'ADN. Lors de la duplication de cette séquence, appelée « *template* » ou « *amplicon* », l'ajout des nucléotides triphosphates constituant l'ADN est contrôlé, on peut donc choisir d'incorporer un analogue photoactivable à la place d'un nucléoside naturel. L'utilisation de mélanges entre nucléotides triphosphates modifiés ou naturels permet une introduction plus ou moins grande de l'analogue photoactivable dans la séquence d'ADN, mais sans pouvoir contrôler précisément la position des modifications au niveau de la séquence cible sauf si l'on atteint le taux de substitution maximal.

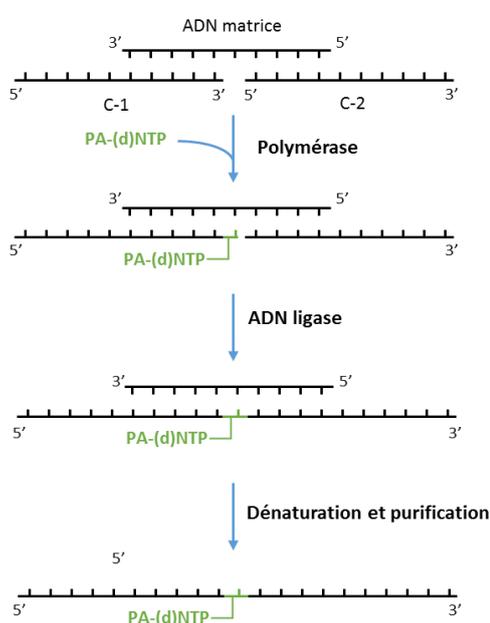


Figure 36: Principe de l'introduction d'un nucléoside photoactivable (PA-(d)NTP) à une position spécifique dans l'ADN

- Par introduction à une position spécifique (Figure 36) : l'incorporation de l'analogue photoactivable à l'extrémité 3' d'une séquence d'ADN s'effectue en présence du nucléotide triphosphate modifié, de l'ADN d'intérêt et d'une ADN polymérase. L'introduction de l'analogue photoactivable au cœur d'une séquence d'ADN est plus compliquée. Le principe de cette méthode est présenté dans la Figure 36. Deux brins d'ADN **C-1** et **C-2**, correspondant à la séquence cible délestée d'un nucléotide, sont maintenus côte à côte à l'aide d'un brin d'ADN **matrice** complémentaire à la séquence cible. Le nucléotide photoactivable **PA-(d)NTP** est incorporé par une polymérase à l'extrémité 3' de **C-1**. Une ADN ligase lie ensuite les deux brins d'ADN. Le duplexe est dénaturé et le brin contenant le nucléotide modifié est purifié.

### 3.3.2 La synthèse chimique

L'introduction de fonctions photoactivables dans des séquences d'ADN peut s'effectuer à deux moments lors de la synthèse d'un oligonucléotide : avant la synthèse du polymère au niveau des synthons phosphoramidites (stratégie de fonctionnalisation pré-synthétique) ou après la synthèse (stratégie de fonctionnalisation post-synthétique).

### 3.3.2.1 Stratégie de fonctionnalisation pré-synthétique

La modification pré-synthétique des oligonucléotides consiste à fonctionnaliser des synthons phosphoramidites avec la modification désirée. Ce synthon est ensuite incorporé à la position souhaitée au cours de la synthèse sur support solide de l'ADN cible. La synthèse des synthons phosphoramidites des bases photoactivables analogues de nucléobases comme la 5IU est bien maîtrisée<sup>203</sup> et ces synthons sont commercialement accessibles.

L'ajout d'autres groupements photoactivables comme la diazirine et la benzophénone peut se faire *via* des bras espaceurs sur la partie sucre<sup>204</sup> ou sur la partie base du nucléoside<sup>201,205</sup> (Figure 37).

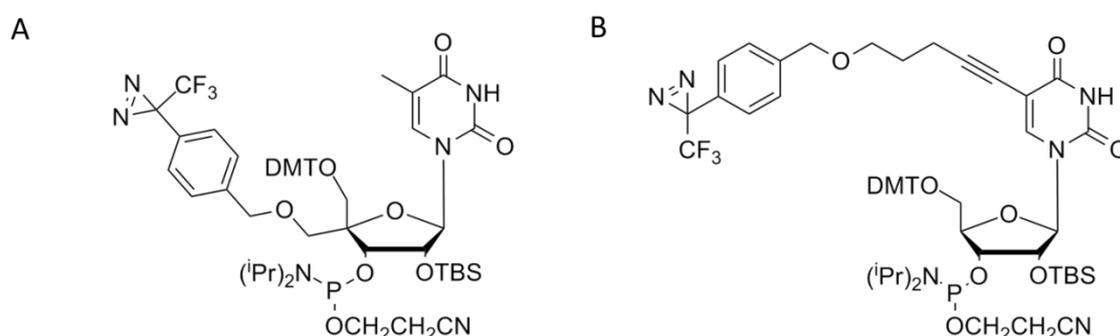


Figure 37 : Exemple de phosphoramidites fonctionnalisés par un groupement photoactivable au niveau du sucre (A) et au niveau de la base (B). D'après Leibmann et al.<sup>205</sup>, et Smith et al.<sup>202</sup>

L'accès à ce type de composés nécessite de nombreuses étapes de synthèse avant l'obtention des phosphoramidites d'intérêt. Par ailleurs, les modifications chimiques introduites doivent résister aux divers réactifs de synthèse automatique d'oligonucléotides sur support solide. Enfin, les synthons phosphoramidites hautement modifiés peuvent entraîner la chute du rendement de synthèse lorsque ceux-ci sont utilisés dans un cycle de synthèse automatique sur support solide. Pour surmonter cet obstacle, des stratégies de fonctionnalisation post-synthétiques ont été mises au point.

### 3.3.2.2 Stratégie de fonctionnalisation post-synthétique

La fonctionnalisation post-synthétique se fonde sur l'incorporation d'une étiquette chimique sur l'ADN par synthèse sur support solide ou par des procédés enzymatiques. Dans une deuxième étape, l'étiquette chimique est fonctionnalisée par le groupement d'intérêt. L'incorporation de l'étiquette à une position donnée permet une modification post-synthétique sélective. Cette stratégie d'assemblage "convergente" permet une grande diversification de la

fonctionnalisation de l'ADN. En effet, un ADN donné va pouvoir être fonctionnalisé par différents groupements chimiques.

De nombreuses méthodes de marquage covalent d'oligonucléotides ont été décrites<sup>206,207</sup>. On peut citer de façon exhaustive les méthodes de chimie « click »<sup>208,209</sup> les réactions de Huisgen-Dimroth de formation de triazole<sup>210</sup>, les réactions de Diels-Alder<sup>211</sup>, la formation d'hydrazone<sup>212</sup>, les aminations réductives<sup>213</sup>, la ligation de Staudinger<sup>214</sup>, les déplacements nucléophiles<sup>215</sup>, la ligation chimique native<sup>216</sup>, la conjugaison thiol-maléimide<sup>217</sup>, l'utilisation de phosphorothioate<sup>216</sup> ou encore la formation de liaison amide<sup>218</sup>. Lercher *et al.* ont, quant à eux, développé une méthode d'introduction de groupement photoactivable par un couplage de Suzuki-Miyaura<sup>207</sup> sur une base halogénée (la 5-iodouracile). Cette méthode est illustrée par

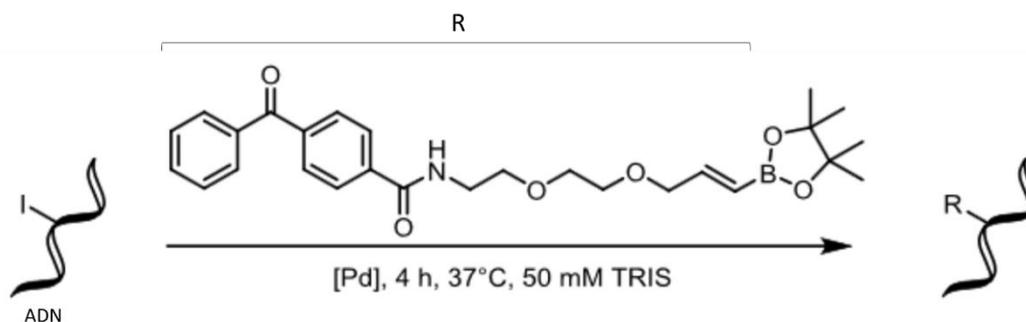


Figure 38 : Fonctionnalisation post-synthétique d'une molécule d'ADN halogénée par un groupement possédant une benzophénone selon un couplage de Suzuki-Miyaura. D'après Lercher *et al.*

l'incorporation d'un groupement benzophénone sur une molécule d'ADN (Figure 38).

Cette approche permettant d'introduire un groupement fonctionnel sur des molécules d'ADN possédant des 5IU est intéressante car les synthons phosphoramidites correspondants sont aisément accessibles.

On peut aussi citer la méthode du « nucléoside convertible ». Cette technique, développée par MacMillan et Verdine<sup>219</sup>, consiste à introduire dans la séquence un analogue nucléotidique contenant un groupement partant qui va pouvoir être substitué par un nucléophile. Ce nucléophile peut par exemple correspondre à un agent photoactivable. Ainsi, Shigdel *et al.* ont utilisé la méthode du nucléoside convertible pour introduire des diazirines aliphatiques à partir de deux analogues nucléotidiques différents : la 2-fluoro-2'-désoxyinosine et la O<sup>6</sup>-triazolyl-2'-désoxyuridine (Figure 39)<sup>220</sup>.

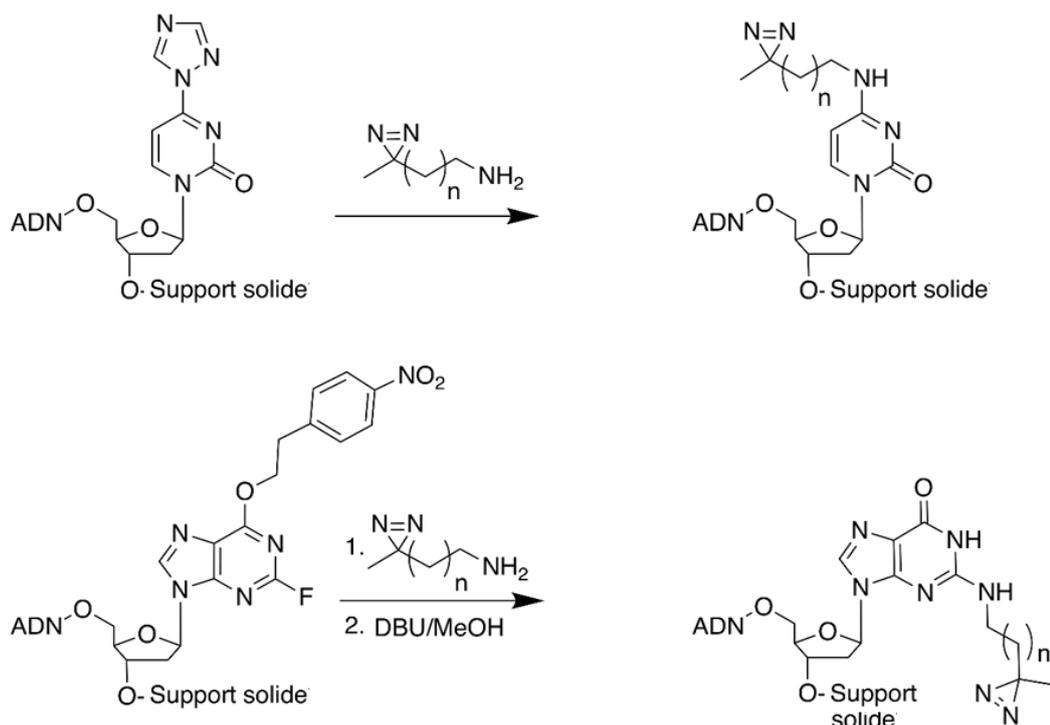


Figure 39 : Fonctionnalisation post-synthétique de la 2-fluoro-2'-désosxyinosine et de la O6-triazolyl-2'-désoxyuridine par la méthode du nucléoside convertible par des diazirines aliphatiques adapté des travaux de Shigel et al <sup>220</sup>.

Les différentes méthodes de fonctionnalisation de l'ADN permettent l'introduction d'une grande diversité de groupements chimiques au sein de la séquence d'ADN et notamment d'agents de photomarquage.

3.4

#### Le PAL (photoaffinity labeling) couplé à la spectrométrie de masse.

Les agents de photomarquage sont principalement utilisés au sein d'une technique : le photomarquage d'affinité (PhotoAffinity Labeling, PAL). Premièrement introduit en 1962 par le groupe de Westheimer<sup>221</sup>, le PAL se caractérise par la fonctionnalisation d'une molécule d'intérêt – par exemple de l'ADN- par un groupement photoactivable ; celle-ci est alors appelée sonde. Ce groupement photoactivable est stable dans des conditions biologiques, et peut être activé suite à une exposition à un rayonnement ultraviolet (UV). La sonde forme une espèce fortement réactive de façon transitoire. Cet intermédiaire réactionnel établit une liaison covalente avec les macromolécules situées à proximité de la sonde (Figure 40).

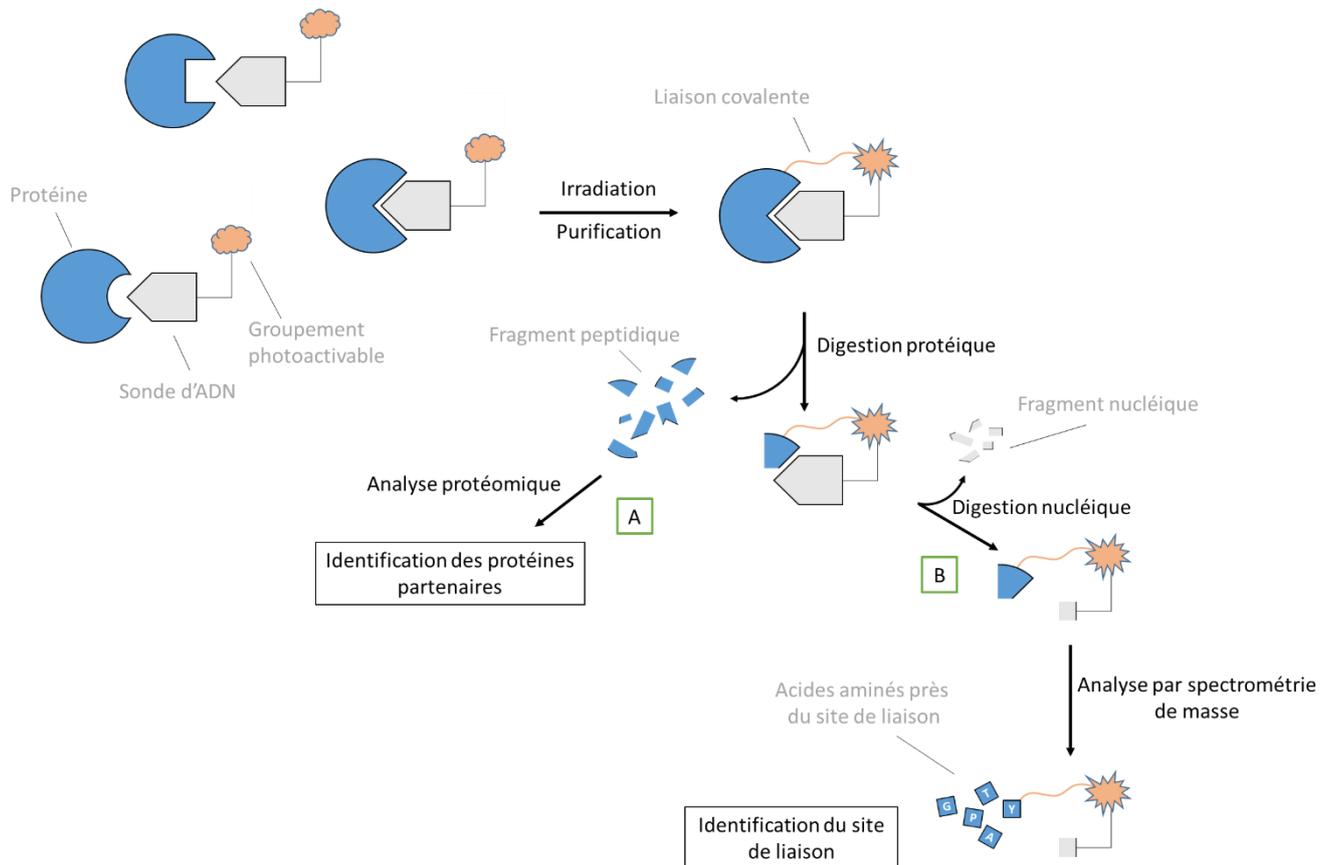


Figure 40 : Principe du photomarquage d'affinité appliqué à l'identification des protéines partenaires de la sonde (voie A), ici ADN, et à l'identification du site de liaison (voie B), par spectrométrie de masse.

Des fonctions supplémentaires peuvent être ajoutées sur les sondes photoactivables pour mieux exploiter la formation de ce lien covalent. L'ajout d'un groupement fluorescent ou d'un isotope radioactif peut faciliter la détection et l'identification des complexes covalents par exemple sur un gel d'électrophorèse. L'ajout d'une biotine sur la sonde permet, grâce à l'interaction biotine/streptavidine, d'enrichir/concentrer les complexes covalents formés ou de les purifier dans des milieux biologiques complexes. Le PAL devient alors une technique d'enrichissement.

Le couplage du PAL à des techniques d'analyses notamment la spectrométrie de masse permet d'accéder à plusieurs types d'informations :

- d'une part, elle permet d'obtenir des informations sur le site d'interaction de la protéine ou des biomolécules piégées par la sonde<sup>222,223</sup> (Figure 40, voie B). Golden et ses collègues ont démontré dès 1999 l'utilité du couplage du PAL à la spectrométrie de masse pour caractériser les complexes ADN-protéines<sup>222</sup>. En substituant la 5-bromo-2'-désoxyuridine à la thymidine au sein d'une molécule d'ADN, ils ont réalisé le pontage *in vitro* d'une molécule d'ADN simple brin au facteur de croissance de fibroblastes basique (bFGF)<sub>155</sub> qui est un régulateur important de protéines des cellules en

prolifération et en différenciation. Le complexe ADN-bFGF<sub>155</sub> a ensuite été digéré par la trypsine, et a été purifié et isolé par gel d'électrophorèse. Avant d'être à nouveau digéré avec de la phosphodiesterase/phosphatase alcaline. La digestion successive du complexe par la trypsine et par ce duo d'endonucléases permet de réduire le complexe à un peptide modifié par un nucléoside. Leur échantillon a ensuite été analysé en LC-MS. La comparaison des spectres de masse pour bFGF<sub>155</sub> digéré et le complexe digéré révèle un seul fragment unique caractéristique du complexe. Le séquençage MS/MS de ce fragment a confirmé celui-ci comme étant le site de liaison covalente et a permis d'identifier la séquence principale de la poche de liaison de bFGF<sub>155</sub>.

- d'autre part, elle peut permettre l'identification de protéines interagissant avec la molécule d'intérêt<sup>224,205</sup> (Figure 40, voie A). En 2011, le groupe de Thomas Carell a utilisé le PAL pour identifier des protéines affines d'une base modifiée de l'ADN, la 8-oxo-désoxyguanine (8-oxoG), par spectrométrie de masse<sup>205</sup>. Au cours de ces travaux, ils ont développé plusieurs sondes d'ADN photochimique basées soit sur un groupement diazirine, soit sur un groupement benzophénone, soit sur des dérivés halogénés de bases de l'ADN, qu'ils ont ensuite comparés. Ces sondes doubles brins d'ADN sont constituées d'un brin possédant la 8-oxoG et d'un autre brin possédant un agent de photomarquage. Pour pouvoir comparer leurs différentes sondes, les auteurs ont travaillé sur des extraits bactériens sur-exprimant la protéine *Lactococcus lactis* formamidopyrimidine glycolase (L/Fpg), protéine qui va spécifiquement reconnaître la 8-oxoG. Une fois leurs différentes sondes incubées avec ces extraits, ils irradient leur système avant de l'analyser sur un gel d'électrophorèse dénaturant. Leur étude comparative de différentes sondes montre que les sondes ADN possédant un groupement benzophénone piègent un plus grand nombre d'interaction non-spécifiques par rapport aux autres sondes, qui présentent un ratio signal sur bruit favorable entre protéines spécifiques et non spécifiques. Ils ont ensuite réalisé une seconde expérience avec une sonde ADN possédant un groupement diazirine et des extraits nucléaires de cellules HCT116 dans le but d'identifier des protéines partenaires de la 8-oxoG. Après l'irradiation de ce nouveau système, leurs échantillons ont été analysés par gel d'électrophorèse dénaturant avant que les bandes correspondant à des complexes ADN-Protéines ne soient digérées à la trypsine et analysées en spectrométrie de masse. Cette étude protéomique a permis d'identifier plusieurs protéines connues pour jouer un rôle dans les mécanismes de réparation de l'ADN, notamment le BER et notamment

hOGG1, une glycosylase qui excise spécifiquement la 8-oxoG<sup>225</sup>. En résumé, l'équipe de Thomas Carell a développé un système basé sur le photomarquage d'affinité permettant l'identification de protéines spécifiques de la 8-oxoG.

Cette dernière étude nous a inspirés dans la conception d'une stratégie permettant de capturer et d'identifier des protéines spécifiques de modifications épigénétiques de l'ADN

## Projet de thèse

Depuis la découverte en 2009 de la 5-hydroxyméthylcytosine en quantité non négligeable chez les mammifères, ainsi que des autres formes oxydées de la 5-méthylcytosine, la communauté scientifique s'est vivement intéressée au rôle biologique de ces bases modifiées : font-elles parties d'une voie de déméthylation de l'ADN ou sont-elles des marques épigénétiques à part entière ? Au cours des dernières années certains éléments de réponse ont été proposés comme nous avons pu le voir dans les chapitres précédents, mais beaucoup de questions restent en suspens : Si la 5hmC, la 5fC et la 5caC sont des marques épigénétiques à part entière et non des intermédiaires de déméthylation, quelles sont les mécanismes qui permettent leurs régulations et leurs maintiens ? Si la 5caC permet la déméthylation de l'ADN, quelle est la décarboxylase permettant de catalyser cette réaction ?

Comme nous l'avons vu, jusqu'à présent, peu de protéines ont été identifiées comme étant spécifiques de ces modifications. Lorsque que j'ai commencé ma thèse, et toujours actuellement, seules deux études proposaient la caractérisation de protéines partenaires de ces dérivés oxydés de la 5mC. L'identification de telles protéines permettrait une meilleure compréhension du rôle de ce processus biologique récemment caractérisé.

Dans ce contexte, mon projet de thèse vise à développer une méthode basée sur le photomarquage d'affinité pour enrichir, identifier et caractériser des protéines interagissant avec les formes oxydées de la 5mC et à visée plus lointaine avec d'autres modifications de l'ADN.

Mon projet se décompose en deux axes :

- Le premier correspond à la conception et l'évaluation de différentes sondes oligonucléotiques photoactivables et la mise au point des conditions permettant de photomarquer des protéines spécifiques au sein d'un mélange protéique complexe. Le point de départ de cet axe s'appuie sur les travaux de thèse de Tom Gérard-Hirne, mon prédécesseur sur ce projet.
- Le second axe porte sur l'identification et la caractérisation des protéines photomarquées à l'aide de nos sondes par spectrométrie de masse. Deux approches ont été étudiées :
  - Une approche moléculaire qui permet la caractérisation de la liaison covalente établie entre une sonde photoactivable et la protéine capturée. Cette approche

nous a permis de mieux comprendre la photochimie de l'agent photoactivable 4-thiothymine.

- Une approche protéomique, plus globale, qui permet l'identification des partenaires d'interaction des modifications épigénétique de l'ADN.

La Figure 41 résume mon projet de thèse et les travaux qui seront présentés dans les trois chapitres suivants.

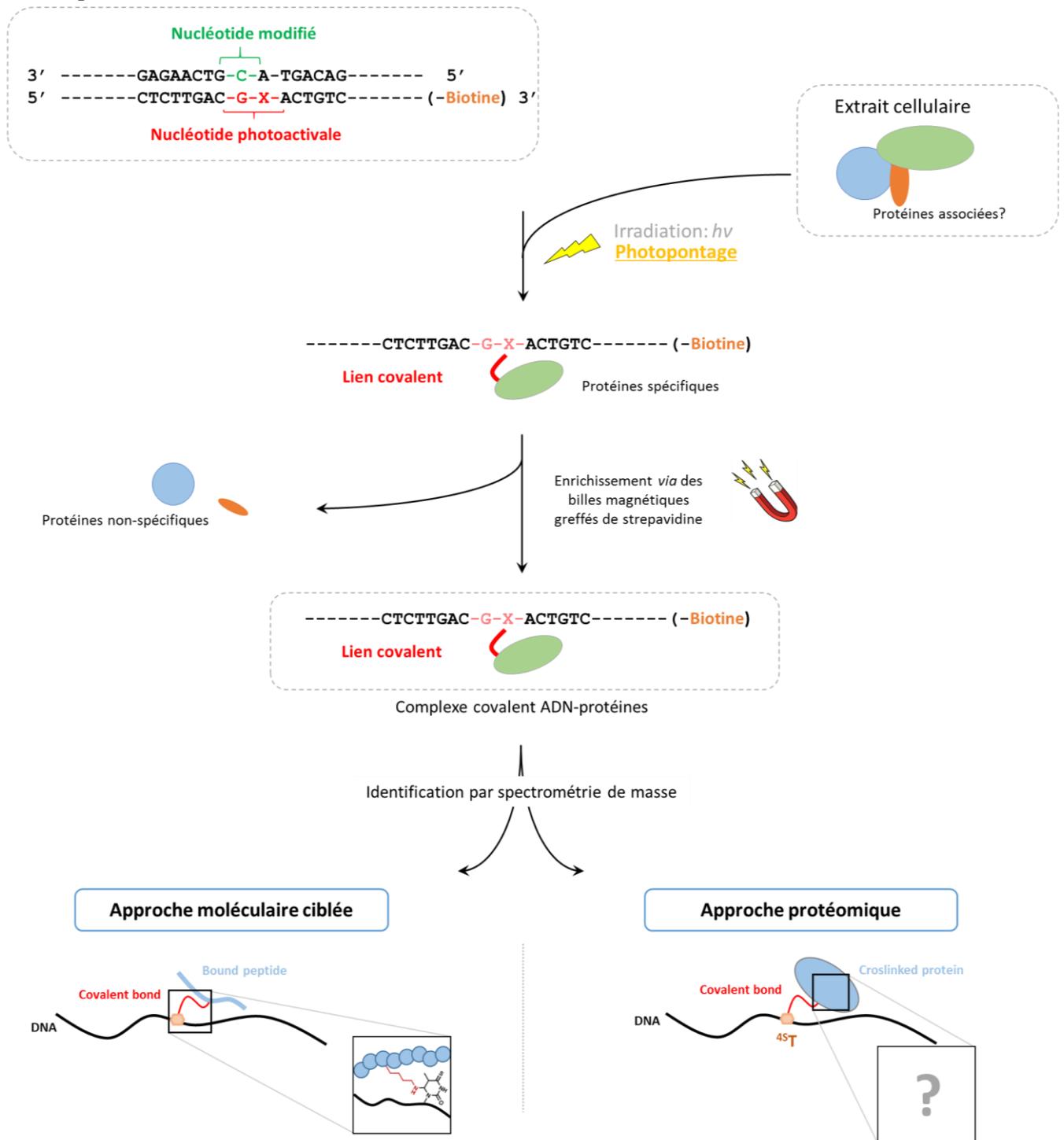


Figure 41 : Schéma récapitulatif du projet de thèse.

## **Partie résultats**

## 4 Conception et évaluation de différentes sondes oligonucléotidiques photoactivables et mise au point des conditions de photomarquage

Dans ce chapitre, nous décrivons l'étude qui nous a permis de définir et de mettre au point les conditions pour (i) photomarquer efficacement des protéines affines pour un oligonucléotide synthétique, simple ou double brin, comportant un motif photoactivable dans sa séquence, (ii) enrichir les protéines ainsi piégées.

Une étude préliminaire a été initiée par Tom Gérard-Hirne au cours de sa thèse et nous avons poursuivi ce travail. Après un rappel des principaux résultats précédemment obtenus, nous présenterons notre contribution à ce travail de mise au point.

### Synthèse de sondes ODN photoactivables

4.1 La conception et la synthèse des sondes ODN photoactivables est un travail antérieur à mon arrivée au laboratoire qui a été réalisé au cours de la thèse de Tom Gérard-Hirne. Dans ce paragraphe, je décrirai ses travaux ainsi que la stratégie mise en place pour synthétiser des oligonucléotides photoactivables et des oligonucléotides ayant une modification épigénétique. J'ai, pour ma part, été amené à reproduire la synthèse de certaines des sondes décrites.

#### 4.1.1 Conception des sondes

La séquence sur laquelle nous avons effectué cette étude préliminaire est présentée dans la Figure 42. Il s'agit d'un duplexe synthétique constitué de 30 paires de bases, donc suffisamment long pour permettre aux protéines de s'y associer<sup>226</sup>. La modification épigénétique sur laquelle nous avons axé notre étude est la 5-hydroxyméthylcytosine (5hmC) et nous avons donc incorporé cette base au cœur de la séquence au niveau d'un dinucléotide CpG (en vert sur la Figure 42).



Figure 42 : Séquence utilisée pour les expériences de photomarquage avec en vert le motif CpG et en rouge, les deux positions d'incorporation des agents photoactivables.

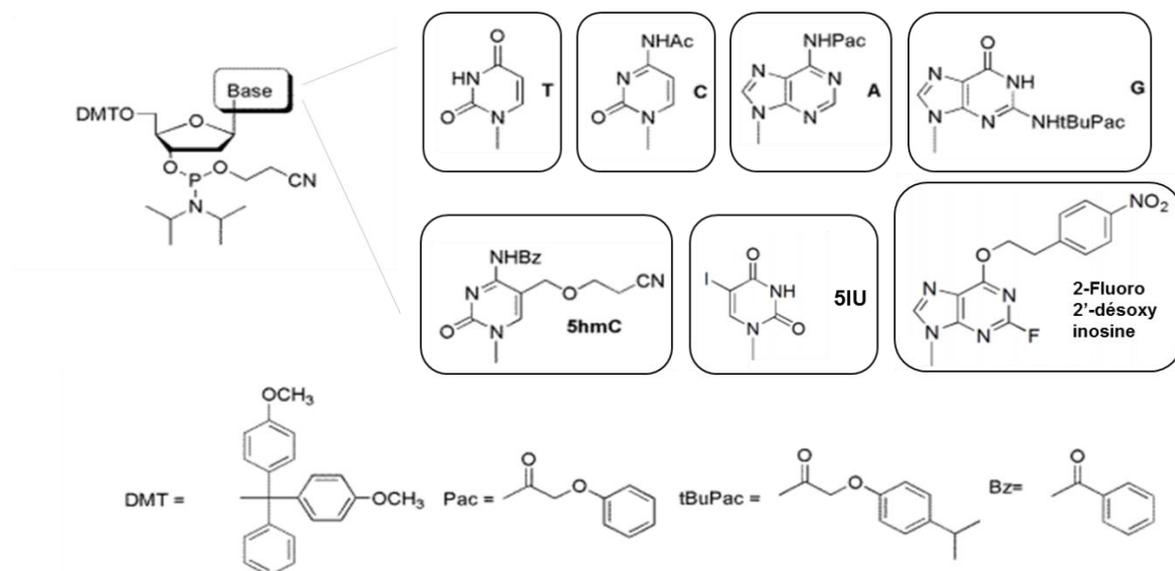
En effet, la 5hmC provient d'une oxydation de la 5mC, qui elle-même provient de la méthylation de la C, or la méthylation de l'ADN a principalement lieu au niveau de dinucléotides CpG.

Les groupements photoactivables ont été introduits sur le brin complémentaire du brin comportant la modification épigénétique soit sur une thymine soit sur une guanine au centre de la séquence, en fonction de la nature du groupement choisi (en rouge sur la Figure 42). C'est ce brin complémentaire qui, après photomarquage sera finalement lié de façon covalente aux protéines affines pour la modification épigénétique de l'ADN. Pour faciliter l'enrichissement des complexes ADN-protéines covalents, une biotine a été fixée à l'extrémité 3' de l'oligonucléotide comportant le groupement photoactivable. Cela permettra d'utiliser l'affinité entre la biotine et la streptavidine pour isoler les complexes d'intérêt au sein de mélanges complexes comme des extraits cellulaires.

#### *4.1.2 Synthèse des séquences oligonucléotidiques modifiées*

Tous les oligonucléotides ont été synthétisés par synthèse chimique sur support solide en collaboration avec l'équipe du Dr Rachid Benhida (Institut de Chimie de Nice, équipe .Molécules Bioactives), à l'exception des oligonucléotides comportant la 4-thiothymine, obtenus commercialement chez Eurogentec. Les synthons phosphoramidites des nucléosides modifiés sont commerciaux et ont été incorporés au cours de la synthèse automatique. Pour introduire certains agents de photomarquage, nous avons procédé à une fonctionnalisation post-synthétique par la méthode du nucléoside convertible<sup>219</sup> précédemment décrite dans l'introduction, paragraphe 3.3.2.2, p.68. La Figure 43 présente les synthons phosphoramidites utilisés dans ces synthèses.

• Synthons phosphoramidite



• Etiquette biotine

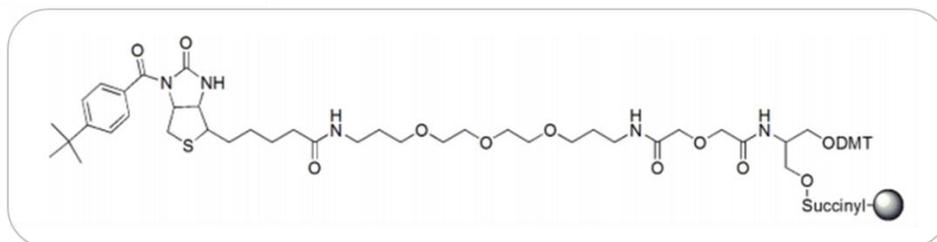


Figure 43 : Structure d'un synthon phosphoramidite, des différentes bases modifiées, des groupes protecteurs et structure de l'étiquette biotine utilisés dans la synthèse sur support solide

4.1.2.1 La synthèse d'oligonucléotides sur support solide

Le principe de la synthèse d'oligonucléotides sur support solide<sup>227,228</sup> est présenté Figure 44. La synthèse s'effectue de l'extrémité 3' vers l'extrémité 5'. L'introduction de l'étiquette biotine est réalisée par l'utilisation de colonnes préchargées avec une unité biotine reliée à un bras espaceur de type polyéthylène glycol (PEG) commercialisée par Glen Research (Figure 43). La position N<sup>1</sup> du cycle uréido de la biotine est protégée avec un groupe *t*-butylbenzoyl, ce qui permet d'éviter des réactions parasites lors de la synthèse automatique. Ce groupement peut être déprotégé dans les conditions de clivage standard. Le bras comporte un motif de type sérinol et l'alcool primaire est protégé par un groupement DMT. Ainsi, il va se comporter comme un phosphoramidite standard et comme le premier « nucléotide » de la synthèse en phase solide. Les nucléotides suivants sont ensuite introduits sous forme de synthons phosphoramidites protégés selon les 5 étapes du cycle présenté Figure 44 :

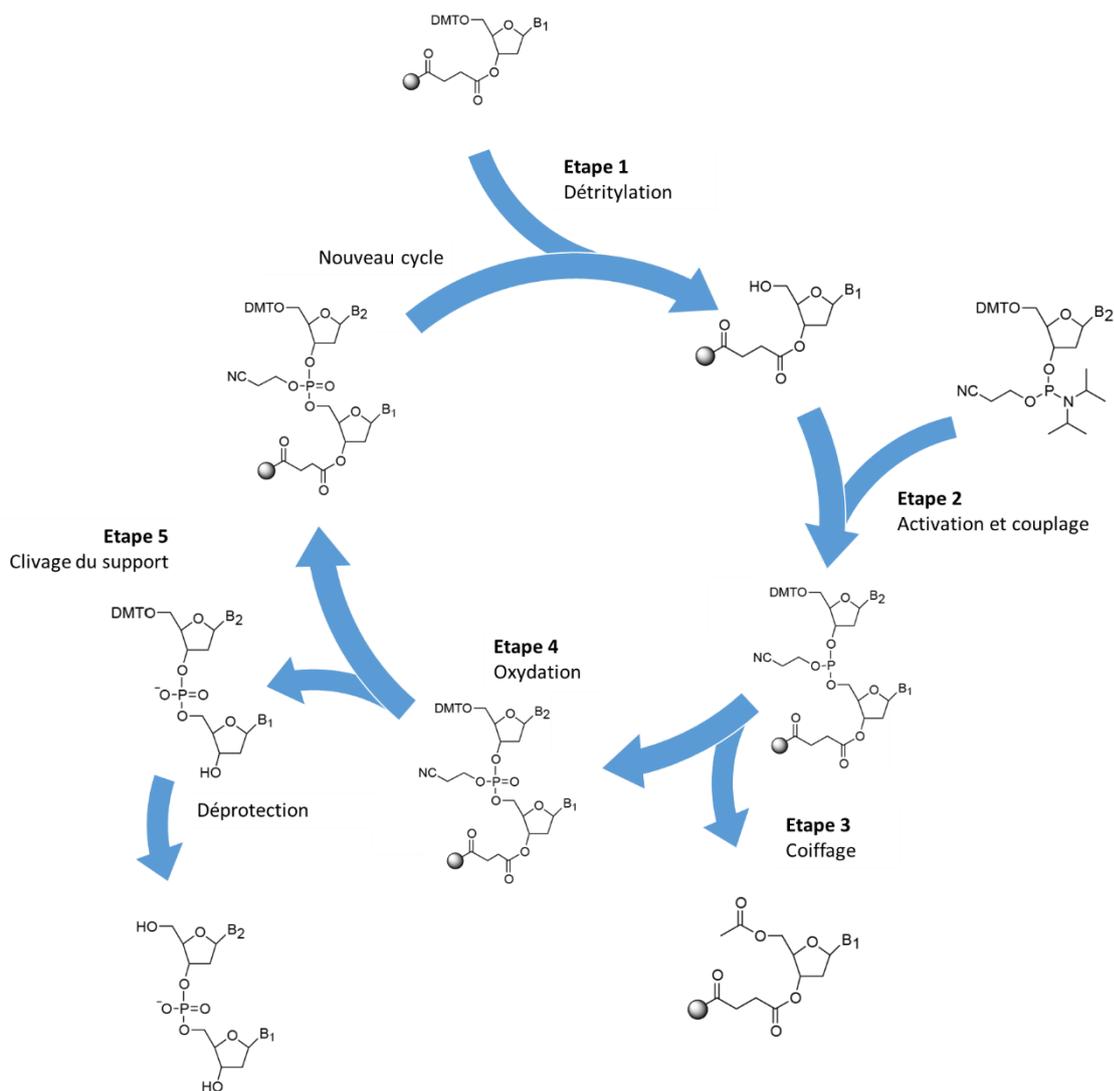


Figure 44: Différentes étapes de la synthèse automatique sur phase solide d'oligonucléotides

1. **Détritylation :** La première étape consiste à cliver le groupement protecteur diméthoxytrityle (DMT) en 5' du nucléoside fixé sur le support solide (dans notre cas, il est sur le composé biotinylé). Cette déprotection, effectuée par un traitement avec une solution d'acide trichloroacétique à 3% dans le dichlorométhane, conduit au dérivé OH-5' libre prêt à être couplé avec le synthon phosphoramidite lors de l'étape suivante.
2. **Activation et couplage :** Un intermédiaire tétrazolyl phosphoramidite activé est créé par addition de benzylthiotétrazole en même temps que le synthon phosphoramidite, permettant de former un meilleur groupement partant au niveau de l'atome de phosphore, qui sera alors substitué par l'hydroxyle en 5' du nucléoside fixé sur le support. Un lien phosphite internucléosidique sera ainsi formé.

3. **Coiffage** : Cette étape consiste en une acétylation (en présence d'anhydride phénoxyacétique et de *N*-méthylimidazole) des fonctions alcools en position 5' qui n'auraient pas réagi au cours de l'étape de couplage. Ces nucléosides ainsi acétylés, ne pourront plus participer aux couplages suivants et on évite ainsi la synthèse d'oligonucléotides plus courts (délétion d'un ou plusieurs nucléotides), produits secondaires de ces synthèses sur support solide qui compliquent ensuite la purification de l'oligonucléotide d'intérêt.
  
4. **Oxydation** : Le lien internucléosidique formé lors de la première étape est un lien instable de type phosphite triester qui est susceptible d'être rompu en milieu acide ou basique, c'est pourquoi, l'atome de phosphore est oxydé en phosphate triester pentavalent, plus stable, à l'aide d'une solution aqueuse d'iode immédiatement après l'étape de coiffage.
  - Le cycle est répété jusqu'à ce que la séquence de l'oligonucléotide soit complète.
  
5. **Clivage** : Après l'ajout du dernier nucléotide, l'oligonucléotide formé est clivé du support solide en condition alcaline (NH<sub>4</sub>OH), qui permet aussi la déprotection des bases et des phosphates. Les ODN sont purifiés par HPLC, puis, le groupement DMT est ensuite clivé par un traitement acide.

Les oligonucléotides contenant la 5mC, la 5hmC, la 5IU et le nucléoside convertible 2-Fluoro-2'-désoxyinosine (Figure 43) ont été synthétisés par Tom Gérard-Hirne. J'ai pour ma part été amené à resynthétiser ce dernier qui permettra par la suite l'introduction des groupements photoactivables benzophénone et diazirine.

#### 4.1.2.2 Incorporation des agents photoactivables benzophénone et diazirine

L'utilisation de la méthode du nucléoside convertible nécessite l'emploi d'une amine primaire peu encombrée pour faciliter la fonctionnalisation post-synthétique. Nous sommes partis d'une benzophénone et d'une trifluorométhylphényldiazirine comportant une fonction acide carboxylique sur le cycle aromatique en position para sur laquelle nous avons effectué un couplage de type peptidique avec un espaceur de type diamine. La synthèse de ces composés est présentée Figure 45.

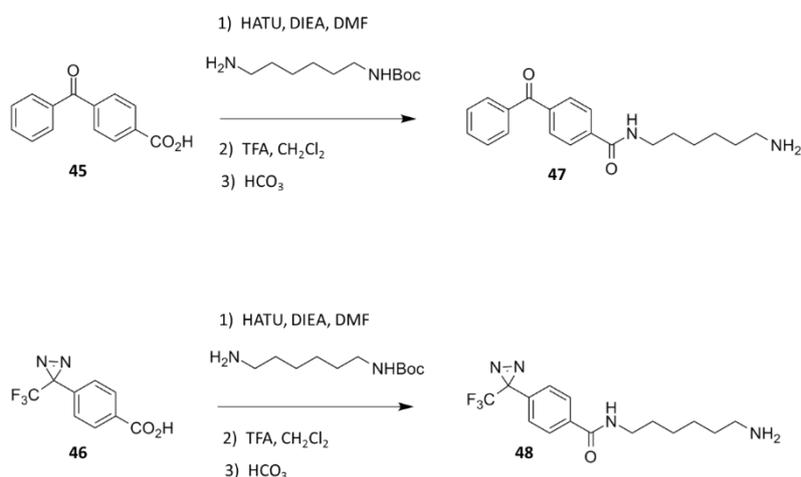


Figure 45: Synthèse des amines comportant un groupement photoactivable.

Dans un premier temps, la 1,6-hexyldiamine monoprotégée par un groupement Boc est couplée avec l'acide 4-benzoylbenzoïque **45** ou l'acide 4-[3-(trifluorométhyl)-diazirin-3-yl] benzoïque **46** préalablement activés par de l'1-[Bis(diméthylamino)méthylène]-1H-1,2,3-triazolo[4,5-b]pyrimidinium-3-oxyle-hexafluorophosphate (HATU) en présence de diéthylamine (DIEA). Dans un second temps, une étape de déprotection du groupement Boc en présence d'acide trifluoroacétique est réalisée. Après neutralisation, on obtient les dérivés benzophénone **47** et diazirine **48** fonctionnalisés avec un bras espaceur comportant une amine primaire.

Le principe de la fonctionnalisation est le suivant : l'oligonucléotide fixé sur son support solide et comportant la 2-Fluoro-2'-désoxyinosine, est traité par l'amine **47** ou **48** en solution dans le DMSO (Figure 46).

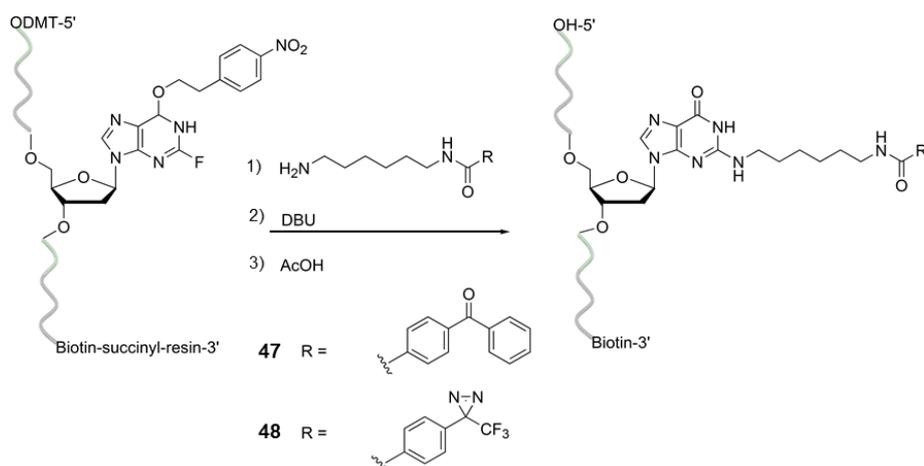


Figure 46 : Introduction de groupement photoactivable benzophénone et diazirine par la méthode du nucléoside convertible.

Les colonnes sont incubées pendant 48h à 55°C, en effectuant des passages réguliers de réactifs à l'aide de deux seringues placées aux extrémités de la colonne comportant le support solide. Puis, après lavage avec du DMSO, puis de l'acétonitrile, les colonnes sont traitées avec une

solution de 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU) dans l'acétonitrile qui permet la déprotection du groupement *para*-nitrophényléthyl de la 2-Fluoro-2'-désoxyinosine. La suite de la synthèse est classique, avec tout d'abord une étape de clivage/déprotection de la séquence oligonucléotidique avec une solution d'ammoniaque concentrée pendant la nuit, à température ambiante. Puis, après purification de l'ODN fonctionnalisé, celui-ci est détritylé en présence d'acide acétique.

Au cours de ma thèse, j'ai été amené à reproduire cette synthèse mise au point par Tom Gérard-Hirne afin d'obtenir un nouveau lot d'oligonucléotide contenant un groupement diazine. Cet ODN a été obtenu avec un rendement estimé à 85 % par comparaison des aires de pics de l'ODN non fonctionnalisé et de l'ODN fonctionnalisé en HPLC. La masse de ce dernier a été vérifiée par spectrométrie de masse MALDI-TOF (Figure 47).

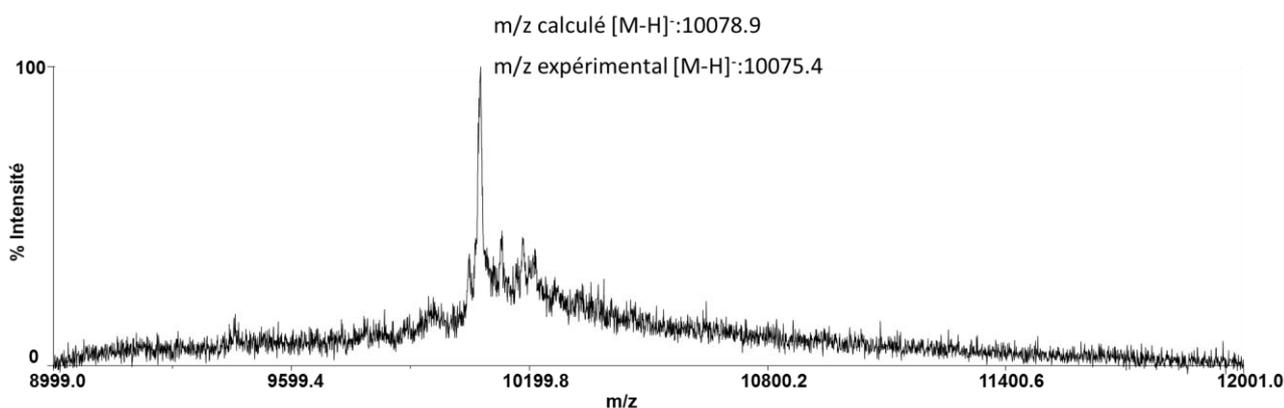


Figure 47 : Caractérisation de l'ODN-Dia par spectrométrie de masse MALDI-TOF, en mode linéaire, ions négatifs, matrice 3-HPA, 10 pmol d'ADN

### 4.1.3 Récapitulatif des oligonucléotides utilisés

La nomenclature suivante a été adoptée pour qualifier les différents ODN modifiés utilisés dans cette étude

- Les oligonucléotides simples brins sont nommés par leur modification c'est-à-dire 4<sup>ST</sup> pour 4-thiothymidine, 5IU pour 5-iodouridine, 5hmC pour 5-hydroxyméthylcytosine, 5mC pour 5-méthylcytosine, Bzp pour benzophénone et Dia pour diazine. Par exemple, un oligonucléotide comportant l'agent photoactivable diazine est nommé **ODN-Dia**. L'ODN ne comportant pas d'agent photoactivable est nommé **ODN-Ctr** (pour contrôle). L'ODN ne comportant pas de modification épigénétique est nommé **ODN-C** (pour cytosine).



## Comparaison et évaluation des sondes ODN photoactivables simple brin

Les différentes sondes ODN photoactivables avaient été testées par et pendant la thèse de Tom Gérard-Hirne au cours d'une étude préliminaire en utilisant les ODN simple brin comportant les motifs photoactivables mis en présence d'une protéine de liaison spécifique à l'ADN simple brin. Nous avons poursuivi cette étude en cherchant à caractériser les complexes ODN-protéine formés dans des systèmes où la sonde est mise en présence d'une protéine unique purifiée mais également en présence d'extraits cellulaires surexprimant ou pas la protéine utilisée comme modèle d'étude.

### 4.2.1 La hRPA comme protéine modèle pour les sondes ADN simple brin.

La protéine RPA humaine (hRPA, human Replication Protein A) est une protéine de liaison à l'ADN simple brin très conservée chez les organismes eucaryotes<sup>229</sup>. hRPA est l'un des acteurs clés de divers processus essentiels du métabolisme de l'ADN, comme la réplication, la recombinaison, la réparation de l'ADN et le maintien des télomères<sup>229,230,231</sup>. La protéine hRPA est un hétérotrimère composé de trois sous-unités : hRPA1, hRPA2 et hRPA3, de masses moléculaires respectives 70, 32 et 14 kDa<sup>232</sup>. Chacune de ces sous-unités possède un ou plusieurs motifs OB-fold (*Oligonucleotide Binding Domain*) nommés DBD A-F qui permettent, soit la liaison à l'ADN simple brin, quelle que soit la séquence de celui-ci, soit l'interaction avec d'autres protéines afin de réguler leur activité (Figure 49).

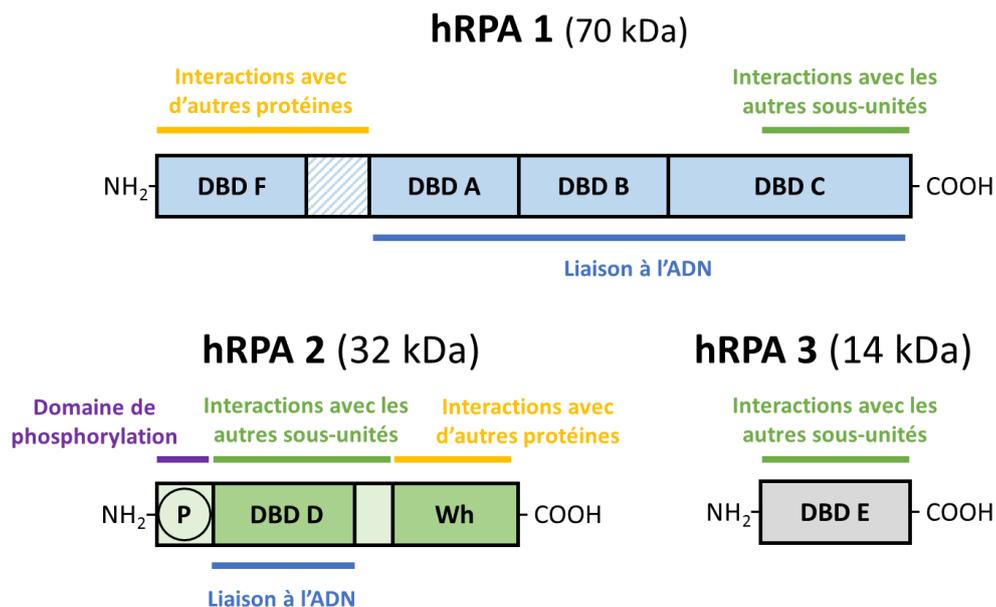


Figure 49: Schéma de la protéine hRPA, ses différentes sous-unités et ses domaines structuraux et fonctionnels.

Les domaines A et B correspondent aux domaines ayant la plus haute affinité pour l'ADN simple brin, alors que les domaines C et D montrent une affinité plus faible<sup>232, 233</sup>. Les domaines E et F ne semblent pas posséder de capacité de liaison à l'ADN, bien qu'une expérience de photomarquage ait montré l'existence d'une interaction de hRPA3 (domaine E) et de l'ADN<sup>234</sup>.

#### 4.2.1.1 Evaluation de l'affinité de hRPA pour les différents ODN par EMSA

La capacité de hRPA à lier chaque sonde ODN a d'abord été évaluée par des essais de déplacement de mobilité électrophorétique (EMSA) sur des gels de polyacrylamide en condition native (Figure 50).

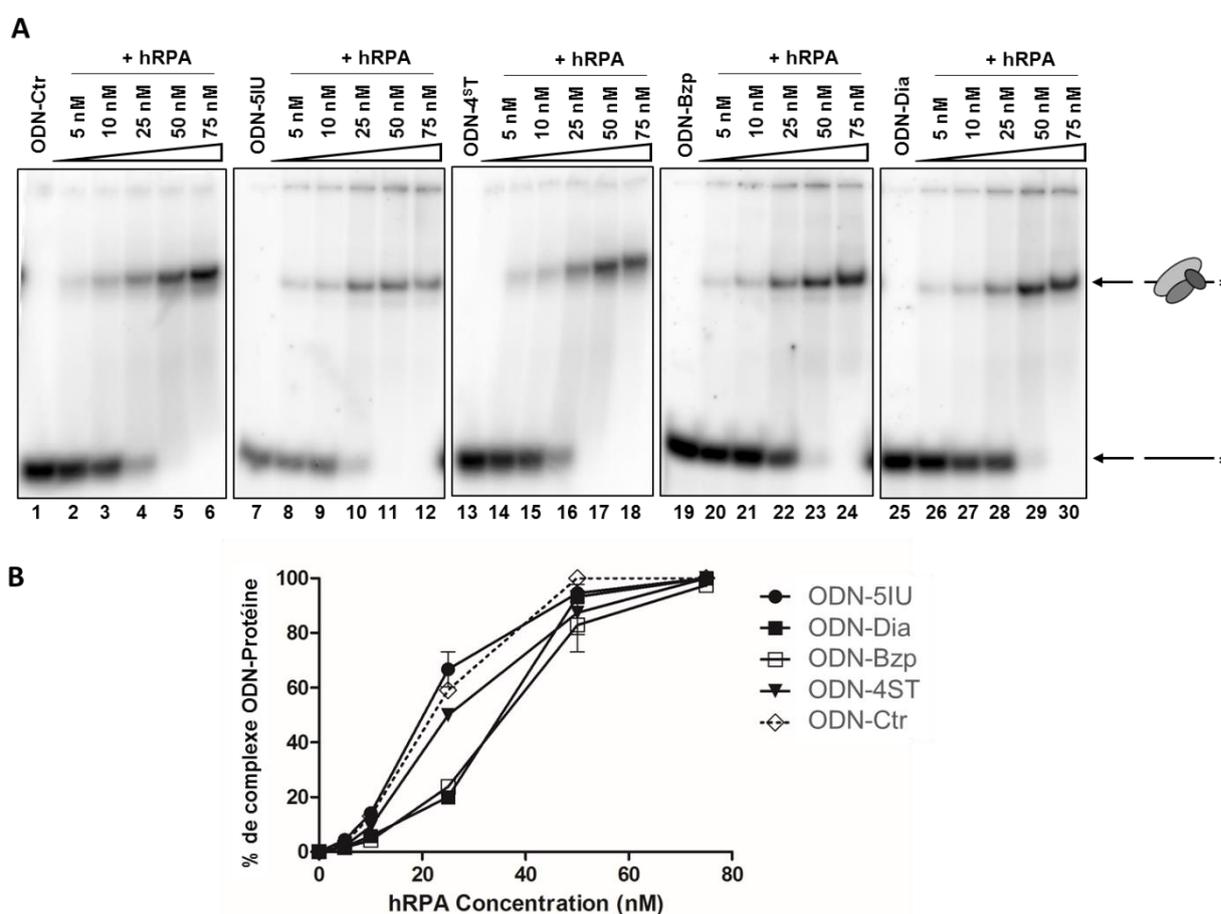


Figure 50 : A. Analyse par EMSA de la liaison entre hRPA purifiée et les ODN simple brin photoactivables synthétisés. La représentation schématique de l'ODN marqué (— \*) est indiquée à droite du gel et les trois ovales représentent les trois sous-unités de hRPA liées à l'ODN : RPA1 (gris clair), RPA2 (gris foncé) RPA3 (noir). B. Analyse quantitative des gels ci-dessus. Les expériences ont été reproduites deux fois pour déterminer la C50, concentration en hRPA pour laquelle 50 % du complexe est formé.

Les ODN ont été marqués au <sup>32</sup>P à leur extrémité 5' avant d'être incubés pendant 15 minutes avec des concentrations croissantes de hRPA (0,5 à 7,5 équivalents). Ce marquage radioactif permet de visualiser les ODN sur gel d'électrophorèse par autoradiographie. Les complexes ODN-hRPA correspondent à une bande dont la migration est retardée par rapport à celle de

l'oligonucléotide non lié (Figure 50A). Le pourcentage de complexation a ensuite été calculé pour chaque concentration en comparant l'intensité de la bande correspondante au complexe ODN-hRPA par rapport à l'intensité totale (ODN non lié et complexé). Les concentrations d'hRPA requises pour former 50 % de complexe ODN-protéine ont été estimées à  $19,5 \pm 0,7$  nM pour l'ODN-5IU ;  $25,1 \pm 1,2$  nM pour l'ODN-4<sup>S</sup>T ;  $36 \pm 2,8$  nM pour l'ODN-Bzp ;  $35,5 \pm 0,7$  nM pour l'ODN-Dia et  $21,9$  nM pour l'ODN-Ctr (Figure 50B). Ces résultats montrent que hRPA lie les ODN modifiés et non modifiés avec une efficacité comparable. Toutefois, l'affinité de hRPA est légèrement moindre pour les sondes présentant une base modifiée volumineuse (ODN-Bzp et ODN-Dia). De plus, cette expérience nous a permis d'établir les conditions de liaison où toutes les sondes ODN sont complexées par hRPA : soit 7,5 équivalents de hRPA. Nous avons conservé ces conditions pour évaluer les propriétés de photomarquage des différentes sondes.

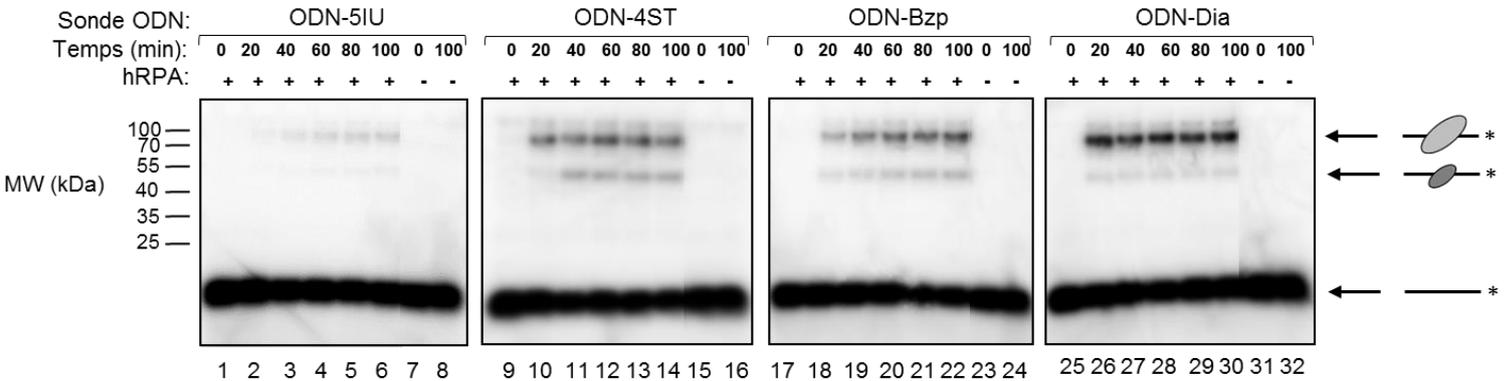
#### 4.2.1.2 Évaluation de l'efficacité de photomarquage des différents ODN avec hRPA

Les différentes sondes oligonucléotidiques comportent des agents photoactivables ayant des propriétés photophysiques et photochimiques différentes, comme présentées dans l'introduction de ce manuscrit, chapitre III. Afin d'évaluer leur comportement dans les réactions de photomarquage et l'efficacité de celles-ci vis-à-vis d'hRPA, une étude cinétique a été réalisée au cours de la thèse de Tom Gérard-Hirne.

Les différentes sondes ont été placées dans les conditions de complexation définies ci-dessus, où 100 % de la sonde oligonucléotide est liée à hRPA. Ainsi, les ODN simples brins radiomarqués ont été pré-incubés avec 7,5 équivalents d'hRPA pendant 15 minutes pour permettre la formation du complexe à température ambiante. Les échantillons ont ensuite été irradiés à 365 nm avec une lampe UV pendant 100 min sur un bloc de glace pour éviter un échauffement excessif de l'échantillon. Toutes les 20 minutes, un prélèvement a été effectué et

mis à l'abri de la lumière à 4 °C. Les échantillons ont ensuite été déposés sur un gel de polyacrylamide SDS-PAGE dénaturant à 10 % et analysés par autoradiographie (Figure 51A).

**A**



**B**

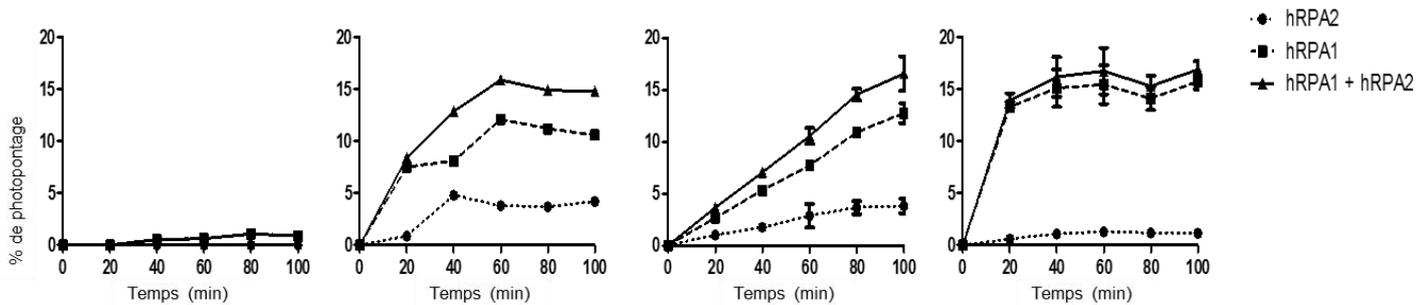


Figure 51 :A. Analyse par gel dénaturant de la cinétique de photomarquage entre hRPA purifiée et les différentes sondes ODN après différents temps d'irradiation. La représentation schématique de l'ODN marqué (-\*) est indiquée à droite du gel et les ovales représentent hRPA1 (gris clair) ou hRPA2 (gris foncé) lié covalamment à l'ODN marqué. B. Analyse quantitative des gels ci-dessus.

Toutes les sondes ODN pontent hRPA puisqu'on observe deux bandes de plus hautes masses moléculaires comparées à la bande correspondant à l'ODN libre. La masse moléculaire apparente des espèces pontées est proche de 45 kDa et 80 kDa, ce qui est cohérent avec les sous-unités hRPA2 (gris ovale) et hRPA1 (gris clair) dont les masses sont de 32 kDa et 70 kDa. hRPA3 n'est pas pontée par aucune sonde. Cela est dû à la position des agents photoactivables en milieu de séquence alors qu'hRPA3 se lie vers l'extrémité 3' sur des ODN courts<sup>232,233</sup>, ainsi que de la faible affinité de hRPA3 pour l'ADN comparé aux deux autres sous unités. Des expériences contrôles en présence des ODN photoactivables irradiés sans hRPA (Figure 51A, piste 8, 16, 24, 32) ou avec un ODN sans agent photoactivable irradié en présence de hRPA (gel non présenté) ont été effectuées. Ces contrôles ne donnent pas lieu à des bandes de plus hautes masses moléculaires, montrant ainsi qu'aucun complexe covalent ne se forme si la protéine est absente ou si l'ODN ne comporte pas le groupement photoactivable. On en déduit que la formation des bandes retardées est bien le résultat de l'activation des groupements photoactivables qui réagissent avec les différentes sous-unités de la protéine.

La quantification de la réaction de photomarquage (Figure 51B) a été réalisée en comparant l'intensité du signal des bandes de l'ODN libre avec celles des bandes correspondant à des complexes ODN-protéines. On constate que la sonde ODN-5IU est peu efficace avec seulement 1 % de pontage alors que les trois autres sondes ont un rendement de pontage similaire (entre 15 et 17 %) après 100 minutes d'irradiation. Ce faible rendement de l'ODN-5IU s'explique du fait de nos conditions d'irradiation. En effet, nous avons utilisé les mêmes conditions d'irradiation pour toutes les sondes, c'est-à-dire l'application d'un rayonnement UV à 365 nm, ce qui est un peu éloigné de la longueur d'onde d'excitation maximale de la 5IU (300 nm). Pour obtenir un meilleur rendement avec cette sonde, il serait souhaitable d'utiliser une autre lampe de longueur d'onde plus faible. Des essais ont été effectués par Tom Gérard-Hirne avec une LED à 300 nm qui a une plage d'irradiation plus étroite mais du fait de sa faible puissance, l'irradiation avec ce type de lampe n'a pas permis d'obtenir des résultats satisfaisants et nous avons donc laissé de côté ces sondes pas suffisamment efficaces avec le matériel d'irradiation dont nous disposions au laboratoire.

Concernant la cinétique de la réaction de photomarquage, nous constatons qu'elle varie d'un ODN à un autre. En effet, la cinétique avec la sonde 4<sup>ST</sup> montre un effet de saturation après 60 minutes d'irradiation, alors que le pontage maximal est atteint plus rapidement, à seulement 20 minutes d'irradiation, avec la sonde de diazirine. Contrairement à ces deux sondes, le rendement de pontage obtenu avec l'ODN contenant de la benzophénone est relativement linéaire au cours du temps et la formation de photoadduits n'arrive pas à saturation après 100 min d'irradiation. Ces résultats sont cohérents avec les mécanismes d'action des différentes sondes photoactivables, préalablement décrits dans le chapitre III de l'introduction. La formation d'un carbène, très réactif, dans le cas de l'ODN-Dia explique le temps de réaction très rapide alors que l'augmentation progressive du rendement de pontage dans le cas de la sonde benzophénone s'explique par la réversibilité du triplet diradical de cet agent photoactivable.

On constate aussi que les principales espèces pontées ont été obtenues avec la sous-unité hRPA1 et que les rapports hRPA1/hRPA2 pontées sont différents d'une sonde à l'autre ce qui pourrait s'expliquer par une orientation différente de l'agent de photomarquage par rapport à chaque sous-unité protéique. Ainsi, on observe un rapport hRPA1/hRPA2 de 2,5/1, 3,3/1 et 13,7/1 respectivement pour l'ODN-4<sup>ST</sup>, l'ODN-Bzp et l'ODN-Dia. Ces résultats montrent que, dans nos conditions, toutes ces sondes sont capables de ponter hRPA avec une efficacité similaire, à l'exception de l'ODN-5IU, que nous avons laissé de côté du fait de sa faible efficacité de pontage dans nos conditions. L'ODN-Dia se distingue toutefois par un temps de réaction

particulièrement court ce qui est un atout dans ce type d'approche car plus on laisse un échantillon sous rayonnement UV, plus celui monte en température, ce qui peut causer la dénaturation de certaines protéines ainsi que de la sonde elle-même. De plus, un temps de réaction court permet le pontage de protéines plus spécifiques puisque plus la réaction est lente, plus on augmente la probabilité de piéger des interactions non spécifiques.

#### 4.2.1.3 Identification de hRPA dans les complexes covalents formés par spectrométrie de masse.

Nous avons cherché à vérifier par spectrométrie de masse que les complexes retardés observés sur gel correspondaient bien à des complexes ODN-hRPA. Pour cela, nous avons reproduit l'expérience de photomarquage entre hRPA purifiée et l'ODN-Dia en gardant les mêmes proportions mais à une échelle plus élevée (1000 pmol de sonde, 7500 pmol de hRPA, 800  $\mu$ L). Nous avons choisi cette sonde car elle présente les meilleurs rendements de photomarquage pour hRPA1. Après irradiation de la hRPA en présence de l'ODN-Dia, l'échantillon est déposé sur un gel SDS-PAGE, révélé au bleu de Coomassie et certaines bandes d'intérêt sont analysées par spectrométrie de masse MALDI-TOF (Figure 52).

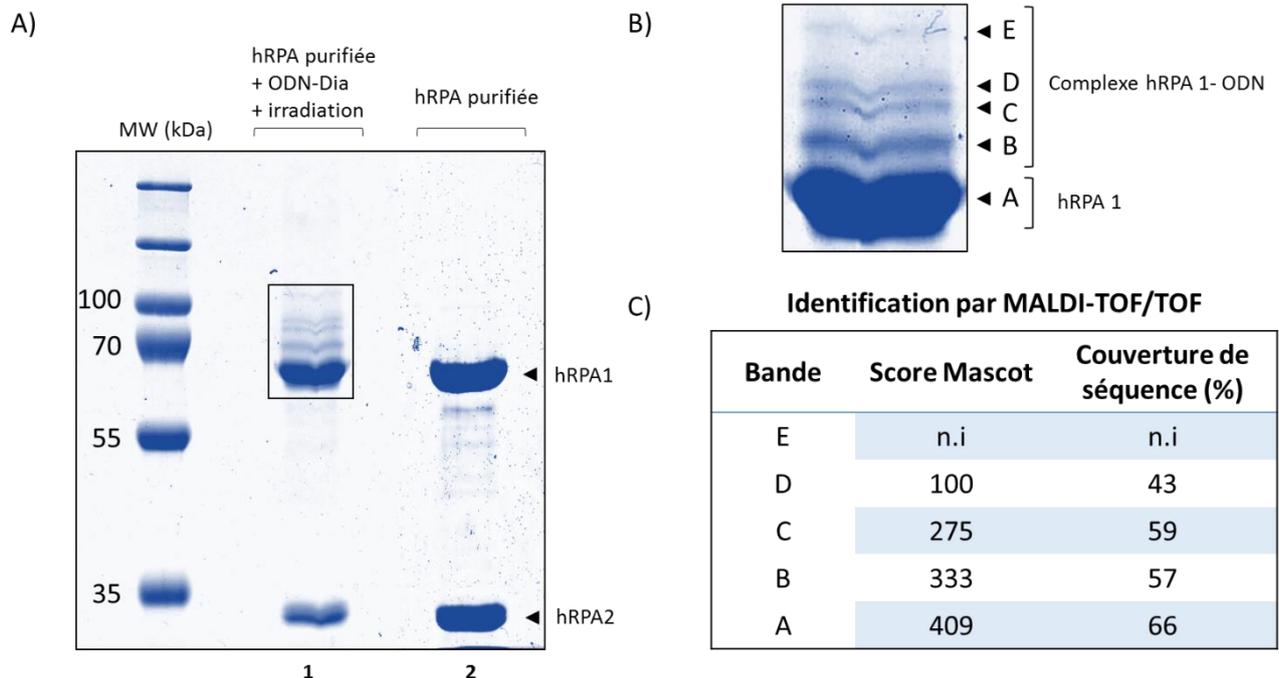


Figure 52 : A) Gel d'électrophorèse dénaturant SDS-PAGE de l'expérience de photomarquage entre hRPA purifiée et l'ODN-Dia (piste 1), coloré au bleu de Coomassie. La piste 2 représente un contrôle, et correspond à la protéine hRPA purifiée. B) Grossissement de la piste 1 de la Figure 52A entre 60 et 100 kDa, zone où se trouvent les complexes potentiels ODN-Dia hRPA1. A, B, C, D et E correspondent à la position des zones excisées pour la digestion à la trypsine et à l'identification par spectrométrie de masse MALDI-TOF / TOF. C) Tableau correspondant aux scores et à la couverture de séquences obtenus avec Mascot après analyse MALDI-TOF / TOF et une recherche en base de données (Swissprot) pour l'identification de la protéine hRPA1 dans les bandes A à E.

La Figure 52A présente le gel SDS-PAGE résultant de notre expérience de photomarquage. La piste 2 correspond à un contrôle avec hRPA purifiée, on y observe deux bandes majoritaires correspondant respectivement à hRPA1 (70 kDa) et à hRPA2 (35 kDa). Par comparaison avec la piste 1, qui représente l'expérience de photomarquage avec l'ODN-Dia, on repère une bande A (Figure 52A et B) qui correspond à hRPA1 non pontée, ainsi que des bandes (B, C, D et E, sur la Figure 52B) de plus hautes masses moléculaires. Ces bandes peuvent donc correspondre à un complexe covalent ODN-hRPA1. Par contre, on n'observe pas de bandes de plus haut masse moléculaire comparé à la bande représentant hRPA2. Ce résultat est en accord avec le photomarquage de l'ODN-Dia vis-à-vis de hRPA, puisque nous avons observé que le pourcentage de rendement de pontage était très faible (<2 %) (Figure 51).

Les bandes A, B, C, D et E sur la Figure 52B ont été excisées du gel, on parle alors de zones, et ont été digérées à la trypsine (après réduction et alkylation des cystéines) selon un protocole de protéomique classique dans le but de réaliser une analyse de type empreinte peptidique (PMF, Peptide Mass Fingerprint) après digestion trypsine des protéines et analyse des peptides résultants après dessalage/concentration, par spectrométrie de masse MALDI-TOF / TOF. Une recherche en base de données, à l'aide du logiciel Mascot®, a permis d'identifier hRPA1 dans les zones A, B, C et D avec un meilleur score d'identification Mascot (409) et une meilleure couverture de séquence (66 %) pour la zone A (Figure 52C). On observe que plus l'on s'éloigne de la zone correspondant à hRPA1, plus le score diminue, traduisant ainsi une diminution de la quantité de protéine et donc de complexe. Ces résultats nous confirment la présence de hRPA1 autour des masses moléculaires correspondant au complexe ODN-Dia-hRPA1 (80 kDa) après irradiation observés dans les gels radioactifs précédents.

Nous ne nous attendions pas à observer plusieurs bandes de plus haute masse moléculaire. En effet, le pontage de l'ODN-Dia sur hRPA1 forme un photoadduit dont la masse est théoriquement proche ou égale à la masse de l'ODN-Dia ajoutée de la masse de hRPA1. Les photoadduits, ayant la même masse, devraient migrer de la même façon sur le gel d'électrophorèse dénaturant ce qui n'est pas le cas. Cette différence de migration peut s'expliquer par un pontage de l'ODN-Dia à différentes positions de hRPA1. Le groupement photoactivable se trouve au milieu de la séquence ADN, le lien covalent s'effectue entre cette position fixe de l'ADN et une position qui peut varier sur la protéine (d'une extrémité à l'autre).

Une fois dénaturée, les complexes peuvent avoir des réticulations différentes, par exemple en « T » ou en « H », selon la position du pontage sur la protéine (Figure 53).

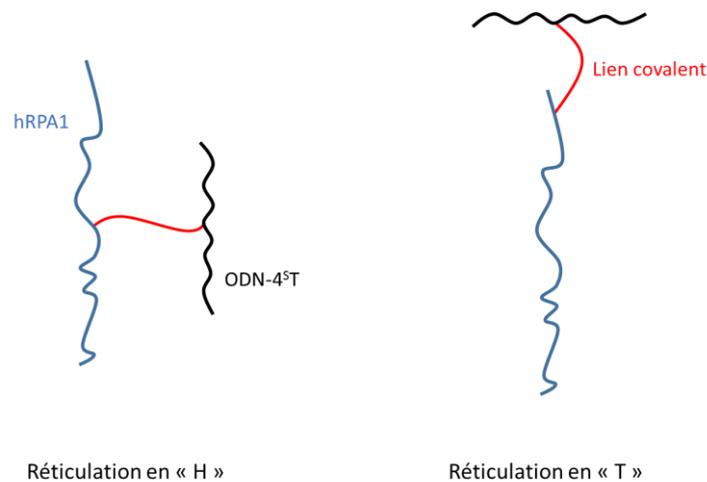


Figure 53 : Schéma de deux réticulations possibles d'un complexe ODN-Dia-hRPA1 après dénaturation

Les complexes covalents n'ayant pas la même réticulation spatiale, ils ne vont pas migrer de la même manière au sein du maillage que forme le gel d'électrophorèse SDS-PAGE. Ainsi chaque bande distincte correspond à un pontage singulier sur la protéine. Il est intéressant de noter que l'intensité du marquage de hRPA1 décroît avec l'augmentation de la masse moléculaire apparente de la bande. Ces bandes n'étaient pas observables lors de la plupart des expériences réalisées en radiomarquage (Figure 51, pistes 10 à 14 et Figure 55, piste 3'). Cela est probablement dû à l'utilisation d'un maillage différent pour le gel d'électrophorèse dénaturant qui rapproche les bandes comprises entre 70 et 150 kDa.

L'analyse des différentes bandes retardées correspondant à des complexes covalents ADN-hRPA1, nous a aussi permis d'obtenir les couvertures de séquences identifiant la sous unité hRPA1 : les peptides identifiés dans chaque zone sont indiqués sur la séquence de hRPA1 de la façon suivante (Figure 12B) : zone A (protéine non pontée) séquences en gras, zone B, C et D (protéine pontée), séquences soulignées en noir, surlignées en gris et souligné en vert respectivement. Lorsque l'on superpose ces couvertures de séquences, on observe que certaines zones ne sont jamais identifiées par spectrométrie de masse que ce soit pour la protéine purifiée ou pour hRPA1 pontée. Ces zones sont représentées sur la Figure 54B, en police «non modifiée». Une zone n'est plus reconnue lorsque hRPA1 est pontée, il s'agit de la zone

entourée en rouge sur la Figure 54B, correspondant à la séquence « ATAFNEQVDKFFPLIEVNK ».

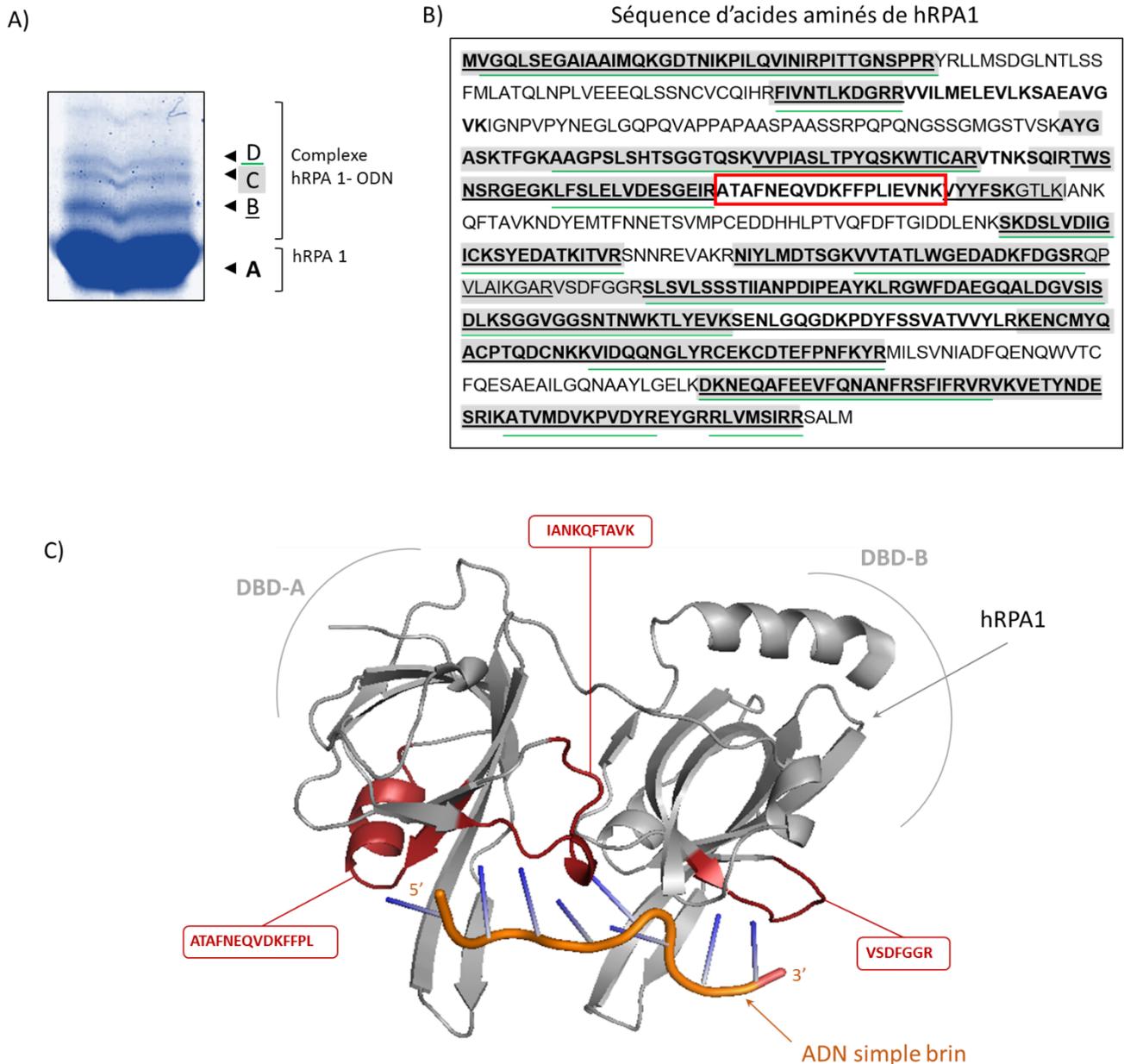


Figure 54 : A) Agrandissement de la piste 1 de la Figure 52A entre 60 et 100 kDa, zone où se trouvent les potentiels complexes ODN-Dia-hRPA1. A, B, C, D et E correspondent à la position des bandes excisées pour la digestion à la trypsine et à l'identification par spectrométrie de masse MALDI-TOF / TOF. B) Séquence d'acide aminés de hRPA1 où sont représentées les couvertures de séquence identifiées en spectrométrie de masse de la bande A (en gras), de la bande B (souligné), de la bande C (surligné gris) et de la bande D (souligné vert). C) Représentation 3D de hRPA1 lié à un ADN simple brin (PBD : LJMC) avec en rouge les régions d'interaction avec l'ADN d'après Bochkarev et al.<sup>235</sup>

D'après la structure cristalline (Figure 54C) représentant un modèle d'interaction entre hRPA1 et un brin d'ADN (8 nucléotides)<sup>235</sup>, ce peptide correspond à l'un des trois peptides interagissant avec l'ADN (en rouge) : ATAFNEQVDKFFPL, IANKQFTAVK et VSDFGGR ; les deux premiers peptides appartenant au domaine A et le dernier au domaine B. Dans notre cas, la « disparition » après photomarquage du peptide ATAFNEQVDKFFPL identifié dans la zone A correspondant à hRPA1 non liée, peut s'expliquer par deux hypothèses :

- L'oligonucléotide lié de façon covalente à hRPA1 va jouer un rôle d'écran lors de la digestion trypsique en empêchant la digestion enzymatique des sites proches du pont covalent.
- L'oligonucléotide n'empêche pas la digestion à la trypsine de hRPA1. Le peptide lié de façon covalente à l'ADN ne se comporte plus comme un peptide lors de son analyse par spectrométrie de masse, il a des nouvelles propriétés physico-chimiques, il est donc difficilement analysable dans des conditions d'analyse en spectrométrie de masse optimisées pour des peptides et n'apparaît pas lors de l'analyse. Pour éviter ce phénomène, une digestion de l'ADN avant l'analyse des peptides devra être envisagée.

Cette étude nous a permis d'identifier de façon indirecte un peptide qui se lie à l'ADN au sein de la sous unité 1 de hRPA, reproduisant ainsi des résultats obtenus par un autre groupe. En effet, lors d'une expérience similaire, Doneanu *et al.* ont utilisé un oligonucléotide constitué de trente thymines pour permettre de caractériser les domaines de liaison de l'ADN au sein de hRPA<sup>236</sup>. Ils ont montré par une méthode de photomarquage couplée à une analyse en spectrométrie de masse que les peptides **ATAF**NEQVDKFFPL, IANK**Q**ETAVK et VSD**F**GGGR permettent la liaison avec l'ADN, notamment via trois phénylalanines (en gras et soulignées) qui pourraient interagir avec les bases de l'ADN via des interactions de type de  $\pi$ -stacking l'ADN. Notre étude permet la visualisation d'une empreinte peptidique.

#### 4.2.2 *Évaluation de l'efficacité de photomarquage des différentes sondes ODN dans un mélange protéique*

Dans un second temps, nous avons évalué la capacité des sondes ODN à ponter spécifiquement hRPA dans un mélange cellulaire complexe. Pour ce faire, nous avons irradié (365 nm, 100 min, 4 °C) les différents ODN radiomarqués préalablement incubés pendant 45 minutes en présence de lysats cellulaires obtenus à partir d'une souche d'*E. coli* surexprimant hRPA. Les échantillons ont ensuite été séparés sur gel d'électrophorèse SDS-PAGE dénaturant à 15 % et analysés comme précédemment par autoradiographie (Figure 55B). La Figure 13A présente un échantillon du lysat cellulaire analysé sur un gel SDS-PAGE réalisé en parallèle et coloré au bleu de Coomassie afin d'illustrer la complexité et la densité des protéines mises en présence des sondes ODN.

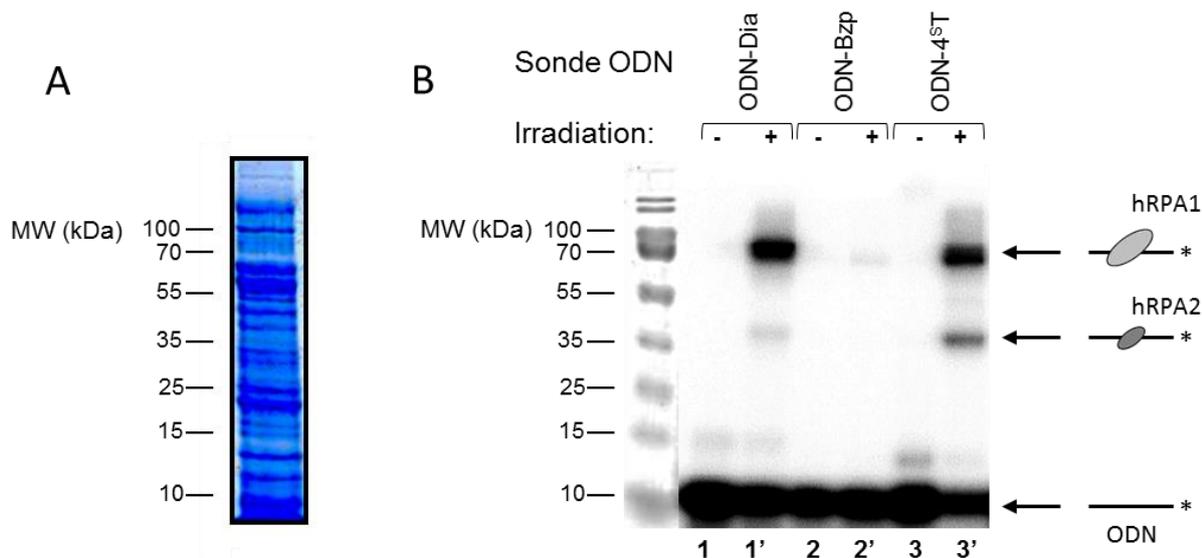


Figure 55 : A. Gel d'électrophorèse SDS-PAGE d'un échantillon de lysat cellulaire obtenu à partir d'*E. coli* surexprimant hRPA (coloration au bleu de Coomassie). B. Analyse par gel d'électrophorèse SDS-PAGE des sondes ODN après irradiation en présence d'un lysat cellulaire surexprimant hRPA. La représentation schématique de l'ODN marqué (—\*) est indiquée à droite du gel et les ovales représentent hRPA1 (gris clair) ou hRPA2 (gris foncé) liées covalamment à l'ODN marqué.

La Figure 55B montre le résultat des différentes expériences réalisées sans (pistes 1, 2, 3) et avec (pistes 1', 2', 3') irradiation des différentes sondes ODN. Nous observons dans tous les cas une bande de plus haute masse moléculaire autour de 80 kDa (pistes 1', 2', 3') similaire à celle observée en présence de hRPA purifiée et correspondant à l'espèce pontée ODN-hRPA1. Par ailleurs, dans les pistes 1' et 3' (sonde ODN-Dia et ODN 4<sup>ST</sup> respectivement), on observe également une bande retardée dont la masse correspond au complexe ODN-hRPA2. Les rendements de photomarquage sont élevés dans le cas des sondes ODN-Dia et ODN 4<sup>ST</sup> (24,5 % et 41,5 % respectivement) mais étonnamment faible en présence de la sonde ODN-Bzp par comparaison à ce que nous avons observé dans un système simple comportant uniquement hRPA pure. Il est intéressant de noter que les rapports de rendement de pontage entre hRPA1 et hRPA2 obtenus avec les sondes ODN-Dia et ODN 4<sup>ST</sup>, respectivement, 16,4/1 et 2,9/1, sont similaires à ceux obtenus avec la protéine pure. Ce résultat peut, là encore, s'expliquer par une orientation différente de la protéine hRPA par rapport à la photosonde.

A l'issue de ces expériences, nous observons donc que seules les sondes ODN-Dia et ODN 4<sup>ST</sup> sont capables de photoponter efficacement et spécifiquement la protéine hRPA surexprimée au sein d'un mélange de protéines. Malgré la grande quantité de protéines présentes dans le lysat, nous n'avons pas observé d'autres protéines piégées par ces sondes. La sonde ODN-4<sup>ST</sup> présente les meilleurs rendements de photomarquage en présence des extraits cellulaires et nous travaillerons par la suite essentiellement avec cette sonde.

#### 4.2.2.1 Identification de hRPA dans les complexes covalents formés

Afin de nous assurer que les complexes covalents observés sous forme de bandes retardées correspondent à des complexes entre la sonde ODN et hRPA, nous avons cherché à identifier hRPA1 au sein des complexes covalents par deux méthodes : d'une part par Western Blot et d'autre part, par une analyse en spectrométrie de masse. Pour cela, nous avons reproduit l'expérience de photomarquage sur lysat cellulaire d'*E. coli* surexprimant hRPA avec la sonde ODN-4<sup>S</sup>T.

##### 4.2.2.1.1 Identification de hRPA1 par Western Blot.

Après incubation de l'ODN-4<sup>S</sup>T (25 μM) en présence d'un lysat cellulaire (2 mg/mL) d'*E. coli* surexprimant hRPA, les échantillons ont ensuite été soit irradiés sur un bloc de glace soit placés dans le noir à 4 °C avant d'être séparés sur un gel d'électrophorèse dénaturant SDS-PAGE 12,5 %. Le gel a été transféré sur une membrane de PVDF pour permettre la détection de la sous-unité hRPA1 à l'aide d'un anticorps dirigé contre cette protéine (Figure 56).

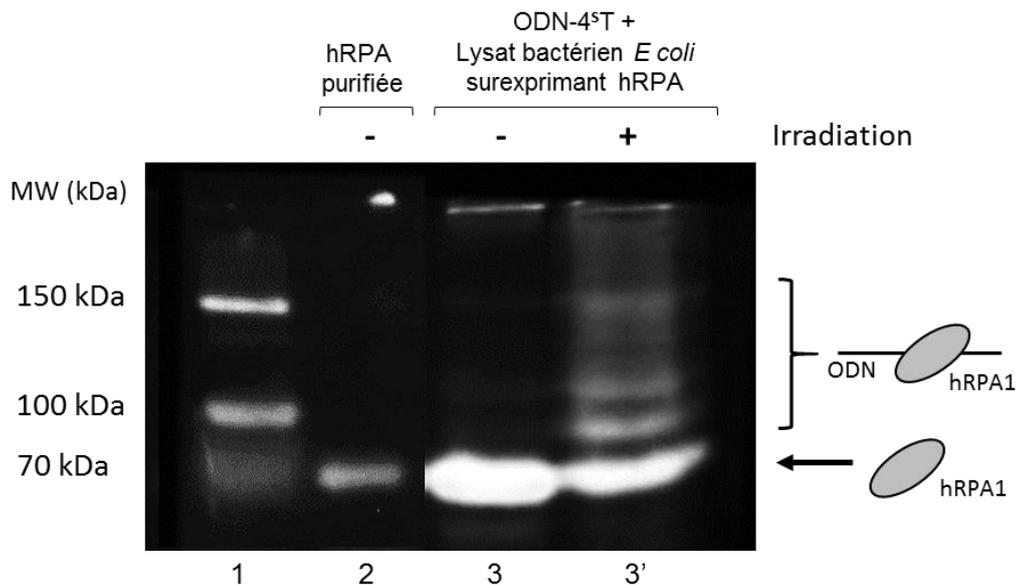
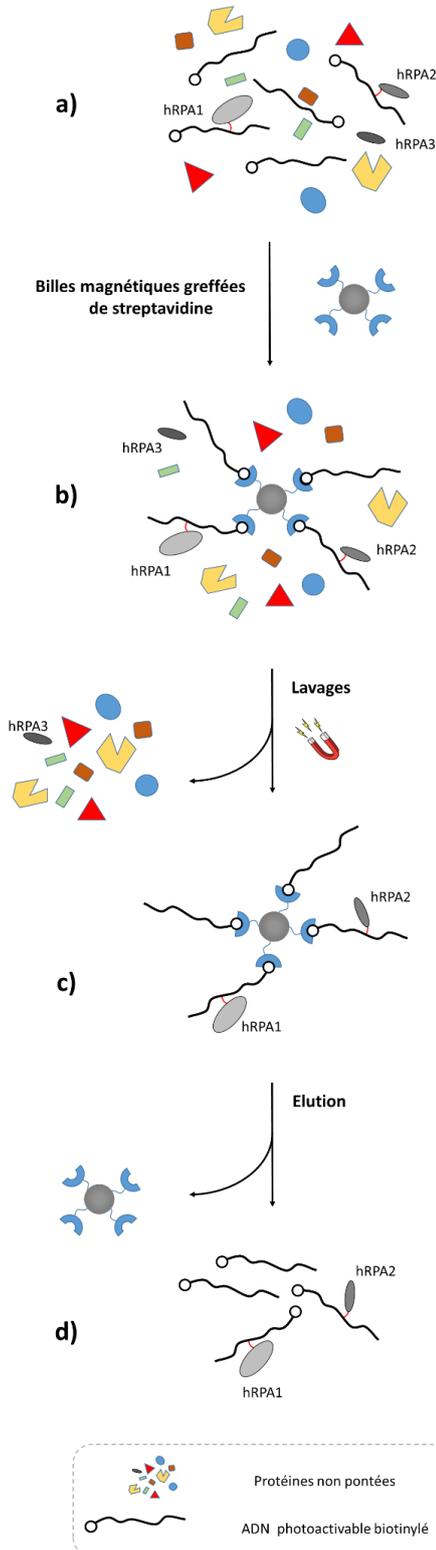


Figure 56 : Analyse par Western Blot des complexes ODN-4<sup>S</sup>T-hRPA1 obtenus par photomarquage de l'ODN-4<sup>S</sup>T en présence de lysat cellulaire d'*E. coli* surexprimant hRPA.

La piste 2 de la Figure 56 correspond à la protéine hRPA1 purifiée et sert de référence dans l'identification de cette sous-unité avant photomarquage. On observe que hRPA1 est bien détectée dans les extraits et identifiée sur les pistes 3 et 3' aux alentours de 70 kDa. Après irradiation, on observe l'apparition de bande de plus haute masse moléculaire (piste 3') comparée à l'expérience sans irradiation (piste 3). Ces bandes correspondent à des complexes ODN-hRPA1 confirmant sans ambiguïté que hRPA1 est bien la protéine pontée parmi toutes

les protéines présentes dans l'extrait cellulaire d'*E. coli*. De façon intéressante, on remarque que ces bandes de plus haut poids moléculaire correspondent aux complexes ODN-hRPA1 déjà observés en bleu de Coomassie (Figure 52B) lors de notre expérience en présence de hRPA purifiée.

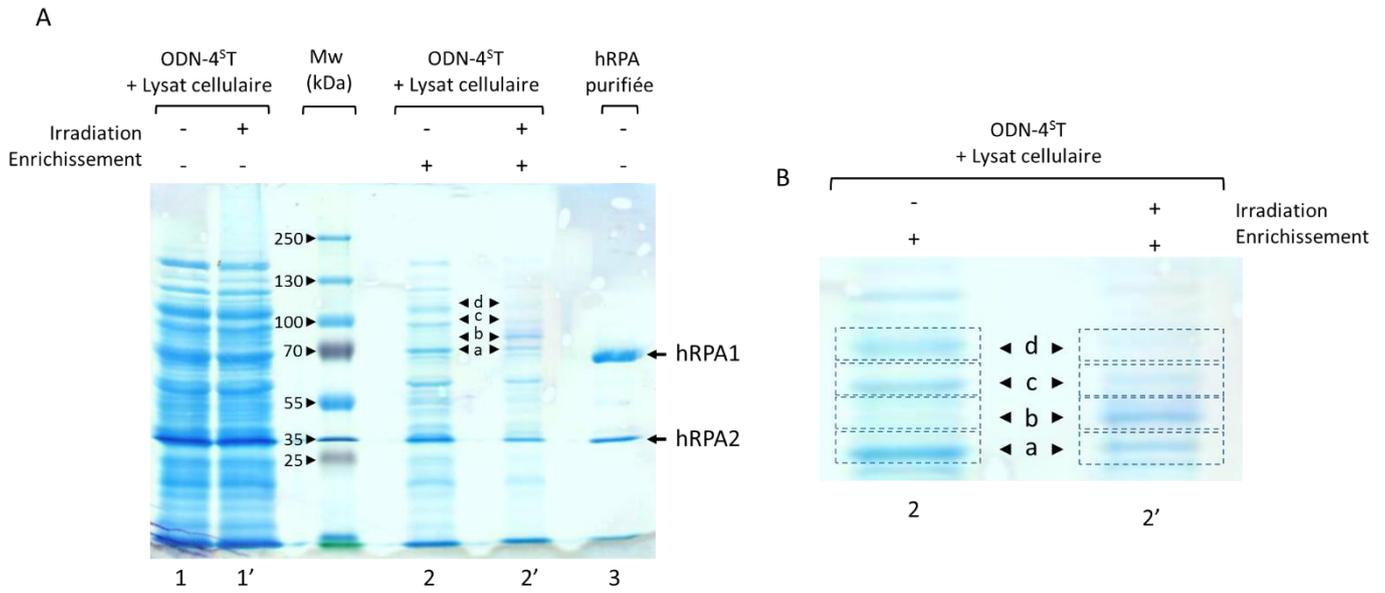
#### 4.2.2.1.2 Identification de hRPA1 par spectrométrie de masse



L'identification des complexes d'intérêt par spectrométrie de masse MALDI-TOF nécessite une étape d'enrichissement des complexes ODN-4<sup>ST</sup>-hRPA avant la séparation sur gel d'électrophorèse dénaturant. L'ODN-4<sup>ST</sup> étant fonctionnalisé en 3' par un groupement biotine, l'interaction de ce dernier avec des billes magnétiques greffées à la streptavidine va permettre l'enrichissement des oligonucléotides liés ou non à une protéine (Figure 57).

En nous basant sur le rendement de photomarquage obtenu avec la sonde ODN-4<sup>ST</sup>, la quantité d'oligonucléotide a été optimisée pour pouvoir visualiser les complexes ODN-4<sup>ST</sup>-hRPA après enrichissement sur gel dénaturant, c'est-à-dire obtenir environ 1 µg de complexe. Après l'irradiation de l'ODN-4<sup>ST</sup> (1000 pmol) en présence d'un lysat cellulaire d'*E. coli* surexprimant hRPA (Figure 57a), des billes magnétiques greffées par de la streptavidine sont ajoutées au milieu. L'affinité entre la biotine et la streptavidine permet la fixation des oligonucléotides (photopontés ou pas) sur la bille magnétique (Figure 57b). L'interaction forte entre la biotine et la streptavidine, ainsi que la liaison covalente entre l'ODN-4<sup>ST</sup> et une sous-unité de hRPA permet de réaliser des lavages en conditions dénaturantes (50 mM Tris-HCl pH= 7.4; 0.1 mg/mL BSA; 0.1 % SDS) à l'aide d'un aimant. Ces lavages éliminent l'excès en protéines non liées à l'ADN et aussi les interactions non covalentes (Figure 57c). Enfin, une étape d'éluion, dans notre cas par chauffage thermique (90 °C, 5 min), permet de libérer l'ODN lié ou non à une protéine. Cet éluât est ensuite séparé sur un gel d'électrophorèse dénaturant (Figure 58A).

Figure 57 : Principe de l'enrichissement des complexes ODN-protéines par purification d'affinité biotine-streptavidine. Lors de l'étape a) la protéine hRPA est représentée avec ses sous-unités séparés dans le milieu pour faciliter la compréhension du schéma.



**C**

**Identification par MALDI-TOF/TOF**

Bande - hv	Score Mascot	Couverture de séquence (%)	Bande + hv	Score Mascot	Couverture de séquence (%)
a	n.i	n.i	a'	100	57
b	77	63	b'	289	67
c	n.i	n.i	c'	144	52
d	n.i	n.i	d'	n.i	n.i

Figure 58 : A) Gel d'électrophorèse dénaturant SDS-PAGE de l'expérience de photomarquage entre l'ODN-4<sup>5</sup>T et le lysat cellulaire d'*E. coli* surexprimant hRPA (coloration au bleu de Coomassie). La piste 3 représente une expérience contrôle, et correspond à la protéine hRPA purifiée. a, b, c et d correspondent à la position des zones excisées pour la digestion à la trypsine et à l'identification par spectrométrie de masse MALDI-TOF/TOF. B) Agrandissement des pistes 2 et 2' entre 70 et environ 140 kDa, où sont représentées les zones excisées. C) Tableau correspondant aux scores Mascot et à la couverture de séquence après l'analyse MALDI-TOF/TOF et une recherche en base de données pour l'identification de la protéine hRPA1 dans les bandes a à d, sans (a, b, c et d) et avec (a', b', c' et d') irradiation.

Les pistes 1, 1' et 2, 2' représentent la même expérience respectivement avant et après enrichissement. En comparant ces deux paires de pistes, on observe une diminution de la complexité du système après enrichissement (moins de protéines). Cet enrichissement va rendre l'analyse de l'échantillon plus simple, que ce soit par spectrométrie de masse MALDI-TOF ou par observation du gel. La piste 3 qui correspond à la migration de la protéine hRPA purifiée permet un repérage sommaire des sous unités de cette protéine dans les autres pistes du gel. Ainsi, dans la zone a (Figure 58B), on trouve une bande présente sur les pistes 2 et 2', qui pourrait correspondre à la sous-unité hRPA1. La bande b n'est visible que dans l'échantillon irradié (piste 2') et pourrait donc correspondre à un complexe covalent ODN-hRPA1. Cette hypothèse a été vérifiée pour 4 zones de protéines (Figure 58B : a, b, c, d) dans les deux

conditions (+/- hv), à l'aide d'une analyse de type empreinte peptidique (PMF, Peptide Mass Fingerprint) réalisée par MALDI-TOF/TOF, comme précédemment décrit après digestion à la trypsine. Les zones c et d permettent de détecter des complexes ayant été encore plus retardés, potentiellement à cause d'une réticulation différente. Dans l'échantillon irradié, l'identification de hRPA1 dans les zones a', b' et c' a été obtenue avec un meilleur score donné par Mascot® (289) et une meilleure couverture de séquence (67 %) pour la bande b'. Dans le contrôle sans irradiation (piste 2), hRPA1 a été identifiée uniquement dans la zone b, avec un faible « score » (77). Ces résultats nous confirment la présence de hRPA1 autour des masses moléculaires correspondant au complexe ODN-4<sup>S</sup>T-hRPA1 (80 kDa) après irradiation. L'absence de hRPA1 au sein de la zone a de l'expérience sans irradiation établit l'efficacité de l'étape d'enrichissement de notre protocole, où le lavage répétitif des billes magnétiques permet d'éliminer la majorité des interactions non covalentes entre hRPA1 et l'ODN-4<sup>S</sup>T. Il existe toutefois un peu de liaison non covalente puisque la protéine est identifiée même en absence d'irradiation (zone b, piste 2) mais avec des scores beaucoup plus faibles ce qui montre l'intérêt du photomarquage dans ce type d'approche par rapport à des expériences de pull-down classiques.

#### 4.2.3 *Photomarquage avec la sonde ODN-4<sup>S</sup>T dans un extrait cellulaire eucaryote*

Nous avons montré que deux de nos sondes ODN simple brin, l'ODN-Dia et l'ODN-4<sup>S</sup>T, étaient efficaces pour capturer la protéine de liaison de l'ADN simple brin hRPA, choisie comme modèle dans un milieu protéique complexe surexprimant cette protéine (lysate bactérien). Nous avons alors voulu nous rapprocher davantage des conditions dans lesquelles nous serons par la suite amenés à travailler, à savoir un protéome eucaryote issu de cellules ne surexprimant aucune protéine particulière.

Pour cela, nous avons incubé l'ODN-4<sup>S</sup>T radiomarqué (2,5 nM) avec un extrait cellulaire de cellule HEK 293 (Humain Embryonic Kidney) (1 mg/mL). Ces cellules proviennent d'une lignée cellulaire humaine spécifique dérivée d'une cellule de rein embryonnaire cultivée dans une culture tissulaire<sup>237</sup>. La Figure 59A présente un échantillon du lysate cellulaire analysé sur un gel SDS-PAGE réalisé en parallèle et coloré au bleu de Coomassie afin d'illustrer la complexité des protéines mises en présence de la sonde ODN-4<sup>S</sup>T. Après incubation de ce lysat

pendant 45 min à température ambiante avec la sonde, l'échantillon a été irradié pendant 60 min (365 nm), avant d'être analysé sur un gel SDS-PAGE (Figure 59B).

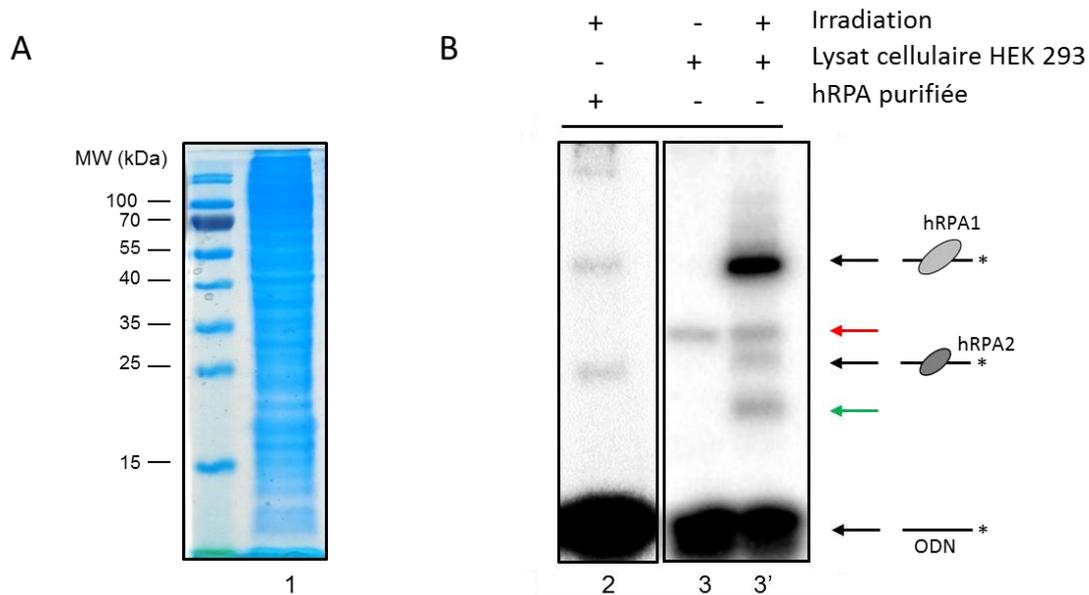


Figure 59 : A. Gel d'électrophorèse dénaturant 15 % où les protéines présentes dans le lysat cellulaire obtenu à partir des cellules HEK 293 sont colorées au bleu de Coomassie. B. Analyse par gel d'électrophorèse SDS-PAGE de la sonde ODN-4<sup>ST</sup> après irradiation en présence d'un lysat cellulaire de cellules HEK 293 (autoradiographie). La représentation schématique de l'ODN marqué (\*) est indiquée à droite du gel et les ovals représentent hRPA1 (gris clair) ou hRPA2 (gris foncé) lié covalamment à l'ODN marqué. Les cadres noirs signifient que les images proviennent de gels différents mais réalisés dans les mêmes conditions

La Figure 59B représente le résultat de notre expérience et de deux contrôles. Le premier contrôle est présenté piste 2, il correspond à une irradiation de l'ODN-4<sup>ST</sup> en présence d'hRPA purifiée. Ainsi, les bandes de plus haute masse moléculaire observées sur cette piste correspondent aux complexes ODN-4<sup>ST</sup>-hRPA1 et ODN-4<sup>ST</sup>-hRPA2 (flèches noires) et permettent de repérer facilement les sous unités hRPA1 et 2 pontées. Dans l'expérience après irradiation (piste 3'), seules quatre bandes discrètes (piste 3', flèches noires, rouge et verte) s'observent avec un rendement de photomarquage élevé (29 %). La bande majoritaire (23 % de photomarquage) indiqué par la flèche noire la plus haute peut correspondre au pontage d'hRPA1 natif présent dans l'extrait cellulaire, car elle migre comme le complexe ODN-4<sup>ST</sup>-hRPA1 observé sur la piste 2. De la même façon, la bande indiquée par la seconde flèche noire (1,3 % de photomarquage) peut correspondre à un complexe ODN-4<sup>ST</sup>-hRPA2. Les masses moléculaires apparentes des bandes pointées par les flèches rouge et verte (45-50 kDa, 2 % et 30-35 kDa, 2,3 % respectivement), sont cohérentes avec le pontage de l'ODN-4<sup>ST</sup> avec les protéines hSSB1 et hSSB2 (36 kDa et 22 kDa, respectivement) qui sont deux autres protéines affines pour l'ADN simple brin présentes dans les cellules eucaryotes<sup>238</sup>. De façon étonnante, la bande caractérisée par la flèche rouge est également observée dans le contrôle sans irradiation. Cette bande a été observée plusieurs fois suggérant qu'il ne s'agit pas d'un artefact.

De plus, lorsque l'on soumet l'échantillon non irradié (celui de la piste 3) à une digestion à la protéinase K, cette bande disparaît ce qui suggère que celle-ci correspond à un complexe ODN-protéine. Ce complexe est lié par une interaction non covalente forte puisqu'il résiste à la dénaturation des échantillons (90 °C, 5 min). L'identification de ce complexe n'a pas pu être réalisée.

De la même manière que pour les expériences avec hRPA pure, nous avons calculés les ratios hRPA1/hRPA2 pontées. De façon intéressante, ce rapport est différent lorsque l'ODN-4<sup>ST</sup> est irradié lors de cette expérience. En effet, cette sonde, qui donne lieu à un ratio de 2,5 /1 en présence de hRPA pure ou d'extraits d'*E. Coli*, donne un rapport de 17,7/1 en présence d'extraits cellulaires eucaryotes natifs. Cela peut s'expliquer par une conformation différente du complexe ODN-hRPA en présence des partenaires protéiques natifs de cette protéine. hRPA interagit avec de nombreuses protéines de traitement de l'ADN, notamment le domaine N-terminal de RPA1 et le domaine C-terminal de hRPA2<sup>239</sup>.

#### 4.2.3.1 Identification de hRPA dans les complexes covalents formés

Afin de vérifier que le complexe le plus intense observé dans la piste 3' (qui constitue 80% des protéines totales photopontées) correspond bien au complexe ODN-4<sup>ST</sup>-hRPA1, nous avons effectué une analyse Western Blot avec un anticorps dirigé contre hRPA1 (Figure 60A).

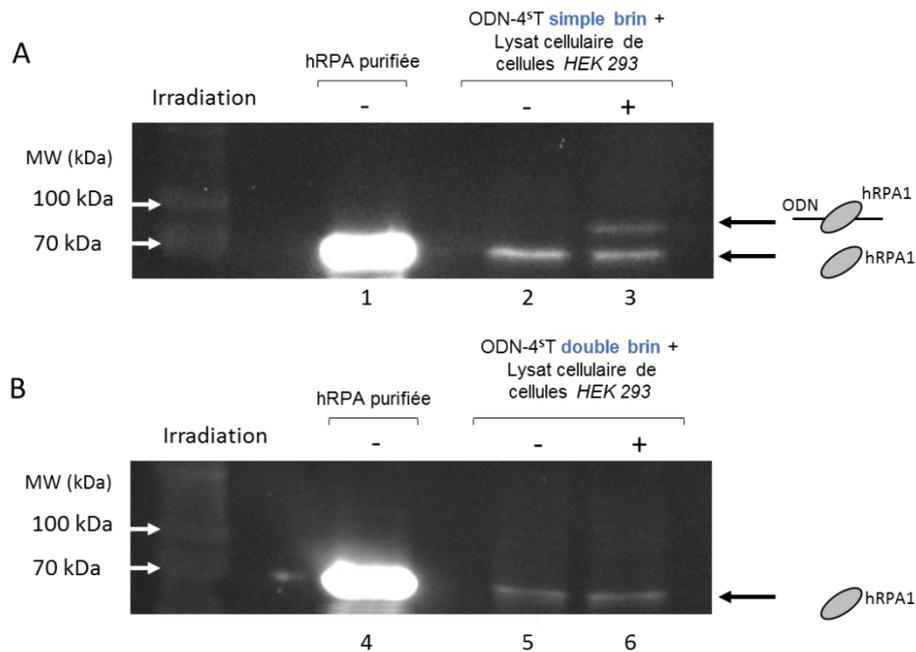


Figure 60 : Analyse des complexes ODN-4<sup>ST</sup>-hRPA1 obtenu par photomarquage de l'ODN-4<sup>ST</sup> simple brin (A) et double brin (B) et du lysat cellulaire de cellule HEK 293 par Western Blot avec un anticorps dirigé contre hRPA1.

Sur cette figure, trois pistes sont représentées. La première, 1, correspond au marquage par l'anticorps de hRPA1 purifiée. La bande identifiée sur cette piste sert de référence pour le repérage de hRPA1 non pontée au niveau des deux autres pistes, qui correspondent respectivement à l'expérience de photomarquage sans et avec irradiation au sein d'un extrait de cellules HEK 293. Sans irradiation (2), une seule bande est révélée et correspond à hRPA1 non pontée. Après irradiation (3), on observe une bande supplémentaire, située à une masse moléculaire plus élevée et dont la masse correspond à une complexe hRPA1-ODN-4<sup>ST</sup>T (80 kDa), confirmant ainsi sans ambiguïté que hRPA1 est la protéine pontée par l'ODN-4<sup>ST</sup>T. L'absence d'autres bandes correspondant à des complexes hRPA1-ODN-4<sup>ST</sup>T de réticulations différentes peut s'expliquer par l'absence d'un excès de hRPA. La quantification des bandes correspondant à hRPA1 sur le Western Blot permet de déterminer un rendement de photomarquage élevé de 40 %.

#### 4.2.3.2 Détermination du caractère sélectif de nos sondes modifiées simple brin

Nous avons aussi réalisé une expérience en remplaçant la sonde ODN-4<sup>ST</sup>T simple brin par une sonde ODN-4<sup>ST</sup>T double brin (ODN-4<sup>ST</sup>T/C), dans le but de vérifier la spécificité de nos sondes simple brin pour des protéines affines pour l'ADN monocaténaire (comme hRPA). L'incubation et l'irradiation de la sonde ODN-4<sup>ST</sup>T/C avec un extrait cellulaire de cellules HEK 293 ont été réalisées comme précédemment. L'expérience a ensuite été analysée à l'aide d'un Western Blot (Figure 60B). Dans ce cas, on n'observe pas de complexe ODN-hRPA1 ponté après irradiation (pistes de droite), indiquant que la sonde ODN-4<sup>ST</sup>T permet bien la reconnaissance spécifique de hRPA1, qui est une protéine affine pour l'ADN monocaténaire alors que la sonde double brin ne le permet pas.

Tom Gérard-Hirne, avait réalisé au cours de sa thèse, une autre expérience sur différents lysats bactériens démontrant la spécificité des sondes simple brin vis-à-vis des protéines affines pour l'ADN monocaténaire. L'ODN-Dia a été incubé et irradié en présence de trois lysats bactériens différents : un lysat cellulaire d'*E. coli* non transformé (hRPA-), un lysat cellulaire d'*E. coli*, surexprimant la protéine hRPA (hRPA-WT) et un lysat cellulaire d'*E. coli*, surexprimant une protéine hRPA mutée (hRPA-D228Y). Cette mutation est située dans la sous-unité hRPA1 et est connue pour réduire l'affinité de l'hRPA pour l'ADN simple brin<sup>240</sup>. Les résultats obtenus après irradiation en présence de la même quantité de protéines totales et séparation sur un gel SDS-PAGE dénaturant à 10 % sont présentés Figure 61.

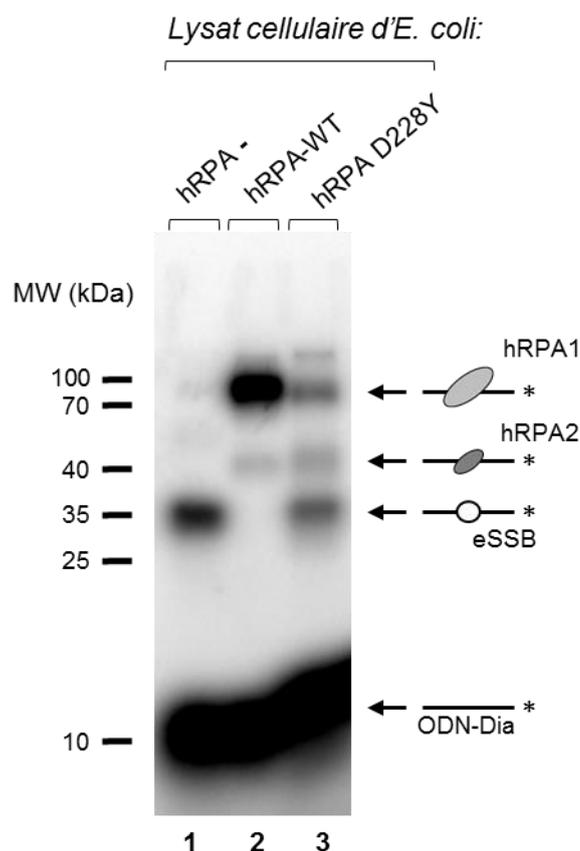


Figure 61 : Analyse du gel dénaturant de l'expérience de photomarquage entre l'ODN-Dia et un lysat cellulaire d'*E. coli* (piste 1), et un lysat cellulaire d'*E. coli* surexprimant hRPA (piste 2) et un lysat cellulaire d'*E. coli* surexprimant hRP-D228Y (piste 3)

Une fois la sonde irradiée en présence de lysat cellulaire d'*E. coli* non transformée, une bande principale est détectée à environ 35 kDa (piste 1). Cette bande correspond très probablement à l'ODN-Dia ponté à la protéine de liaison à l'ADN monocaténaire d'*E. coli* (eSSB) de 19 kDa (rendement de pontage de 19 %). L'identification de cette bande n'a pas été réalisée par Western blot ou spectrométrie de masse mais une expérience similaire réalisée avec la eSSB purifiée après irradiation avec la même sonde montre que la eSSB photopontée migre à la même distance que la bande observée dans la piste 1. Lorsque l'ODN-Dia est irradié en présence de lysat cellulaire provenant de la souche hRPA-WT, cette bande de 35 kDa disparaît en faveur d'autres bandes correspondant à l'hRPA1 et à hRPA2 ponté (30 %) (piste 2). Ces résultats montrent que les sous-unités d'hRPA surexprimées sont en compétition avec la eSSB et empêchent la formation du complexe ODN-Dia-eSSB. Avec les extraits de protéines issus de la souche hRPA-D228Y, la bande correspondant à eSSB est partiellement restaurée (9 % de pontage) tandis que les deux bandes (5 % de pontage) correspondant à hRPA1 et à hRPA2 sont réduites (environ 6 fois moins intenses que dans le cas de la souche hRPA-WT) (piste 3). En outre, hRPA1 est ponté dans une efficacité moindre avec hRPA-D228Y puisque la valeur du rapport de pontage RPA1 / RPA2 a diminué de 13/1 à 4/1. Cette expérience montre bien la compétitivité

entre les protéines affines pour l'ADN simple brin et aucune autre vis-à-vis de la sonde. Ainsi, ces résultats confortent également ce que nous avons pu observer, à savoir que ces sondes sont spécifiques pour les protéines de liaison à l'ADN monocaténaire (SSB) présentes dans un mélange complexe de protéines.

#### 4.2.4 *Conclusion sur la comparaison et l'évaluation des sondes ODN photoactivables simple brin*

La mise au point des différentes sondes photoactivables a été réalisé par Tom Gérard-Hirne. A la fin de sa thèse, il avait initié une étude comparative de différentes sondes avec la protéine hRPA comme modèle d'étude. Cette étude a été complétée et en voici les principales conclusions :

- Cette étude a permis l'établissement de conditions pour photomarquer efficacement des protéines.
- Les sondes ODN-4<sup>ST</sup> et ODN-Dia sont les meilleurs agents de photomarquage dans nos conditions pour l'étude de l'interaction avec une protéine pure ou avec un mélange plus complexe de protéines cellulaires.
- Les expériences de photomarquage sur extraits cellulaires humains ont démontré la puissance de la méthode et le degré élevé de spécificité des sondes photoactivables car peu de protéines ont été piégées, avec pour pontage majoritaire la SSB humaine, l'hRPA.
- L'enrichissement des espèces pontées lors du photomarquage dans un mélange complexe de protéines à l'aide du groupement biotine est efficace, ce qui permet l'identification des protéines piégées par des techniques de protéomique.

De façon intéressante, nous avons aussi montré dans cette étude que la 4-thiothymine, largement utilisée comme sondes intrinsèques de la photo-affinité des structures d'acides nucléiques, est également un agent puissant pour étudier les interactions ADN-protéine et a l'avantage de pouvoir être facilement incorporé soit par synthèse chimique à des positions précises dans la séquence mais aussi enzymatiquement par des ADN polymérases. Au-delà de son utilité pour étudier les protéines connues, comme ici, hRPA, nous avons également démontré en utilisant des sondes bi-fonctionnelles contenant un fragment de biotine, que des sondes contenant de la 4-thiothymine peuvent être utilisées pour piéger de façon sélective les SSB dans un protéome natif. Bien que cela ne soit pas l'objectif premier de cette étude, celle-

ci nous a permis de mettre au point des outils pour ponter covalentement des SSB dans des extraits cellulaires qui peuvent s'appliquer à l'identification de SSB inconnues d'organisme encore mal caractérisés comme le tardigrade, organisme d'une résistance rare<sup>241</sup> qui fait l'objet d'une thématique de recherche au MNHN.

Enfin, cette étude peut être étendue à des sondes photoactivables double brin avec des séquences spécifiques et des modifications telles que des modifications ou des lésions épigénétiques comme décrit précédemment<sup>242,243</sup> pour identifier des protéines affines ou de réparation. Dans la partie suivante, nous allons vérifier la capacité de nos sondes ODN sous forme double brin à capturer des protéines spécifiques de l'ADN bicaténaire.

### Comparaison et évaluation des sondes ODN photoactivables double brin

4.3 L'étude précédente dans un contexte simple brin nous a permis de mettre au point le protocole de photomarquage et d'enrichissement. Cependant, les sondes que nous avons conçues sont destinées à photoponter des protéines de liaison à une modification épigénétique de l'ADN et il était donc nécessaire d'étendre l'étude préliminaire dans un contexte double brin. Nous avons donc évalué la capacité de nos sondes ODN, ici sous forme de duplexe, à photoponter des protéines connues pour se lier à l'ADN double brin. Pour ce faire, nous avons travaillé successivement avec deux protéines modèles, une ADN méthyltransférase et une histone. Avant mon arrivée, la stabilité de duplexes constitués des brins photoactivables appariés à leur brin complémentaire avait été évaluée par des mesures de température de fusion de l'ADN ( $T_m$ ) (Tableau 1) afin de vérifier si les groupements photochimiques introduits dans les séquences ne contribuent pas à déstabiliser la double hélice.

Duplexe <sup>a</sup>	$T_m$ (°C)
ODN-Ctr/C	73,9
ODN-Bzp/C	72
ODN-Dia/C	73,8
ODN-4 <sup>S</sup> T/C	73,4

<sup>a</sup>pH 7,4, tamponné (10 mM Tri, 150 mM NaCl)

Tableau 1 : Température de fusion de différents duplexes d'ADN dont un brin contient un groupement photoactivable.

Lorsque l'on compare les températures de fusion ( $T_m$ ) des différents duplexes possédant des groupements photoactivables à un duplexe contrôle de même séquence sans modification, on observe peu de variation, moins de 0,6 °C entre les duplexes Dia et 4<sup>ST</sup>. Cependant, l'introduction d'une benzophénone induit un léger effet déstabilisateur puisque l'on observe une diminution de 1,9 °C de la  $T_m$  du duplex Bzp par rapport au duplex contrôle. Globalement, les modifications introduites dans les oligonucléotides ne modifient pas la stabilité des duplexes ce qui en fait des bons candidats pour des applications de photomarquage dans un contexte double brin.

#### 4.3.1 Etudes des complexes duplex ODN-DNMT3a

La première protéine modèle sur laquelle nous avons effectué notre étude correspond au domaine catalytique de la *DNA methyltransferase 3a* murine (DNMT3a) composée des acides aminés 608 à 908 (31 kDa), qui avait déjà été étudiée au laboratoire. Le choix de cette protéine n'est pas anodin puisqu'il s'agit d'une enzyme capable de méthyler la séquence que nous avons choisie au niveau du dinucléotide CpG, comme décrit dans l'introduction paragraphe 1.2.1, p.17 ce qui en fait donc une protéine relativement spécifique de nos sondes ODN bicaténares.

Des travaux ultérieurs à mon arrivée au laboratoire avaient établi les meilleures conditions de complexation entre les duplexes et la DNMT3a. Il faut entre 300 et 500 équivalents de DNMT3a pour effectuer une complexation des duplexes à 100 %. Afin de tester la capacité des différents duplexes photoactivables à ponter DNMT3a, les duplexes radiomarqués (5nM) ont été incubés avec de la DNMT3a (500 équivalents, 2,5 µM) en présence de son cofacteur enzymatique, la *S*-adénosyl-L-méthionine (AdoMet) pendant 1h à 37 °C. Les échantillons ont ensuite été irradiés 100 min à 365 nm avant d'être analysés à l'aide d'un gel dénaturant SDS-PAGE (Figure 62).

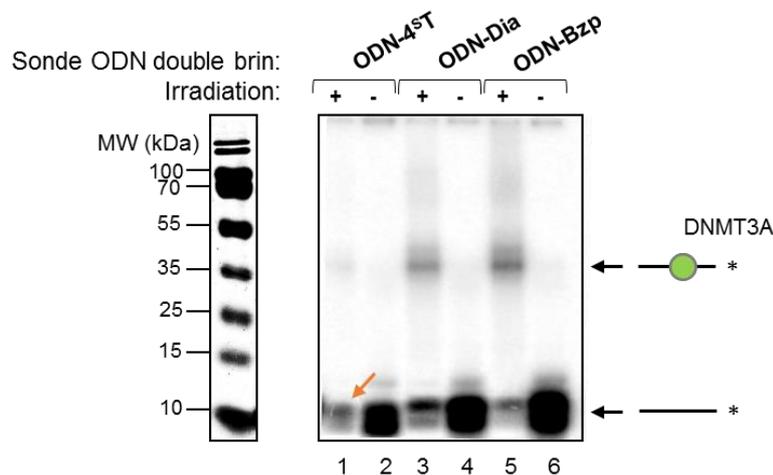


Figure 62 : Analyse par gel d'électrophorèse dénaturant 15 % de la capacité des sondes ODN à lier DNMT3a purifiée. La représentation schématique de l'ODN marqué (◻\*) est indiquée à droite du gel et le cercle vert représente DNMT3a liée covalamment à l'ODN marqué. La flèche orange pointe la bande correspondant à l'ODN-4<sup>ST</sup> n'ayant pas réagi.

L'analyse des échantillons sur gel dénaturant permet d'observer une bande de plus haute masse moléculaire correspondant à la formation d'un complexe entre les sondes ODN double brin et l'enzyme, tout particulièrement pour le duplex ODN-Dia (21 %) et le duplex ODN-Bzp (27,3 %) (piste 3 et 5). Le duplex ODN-4<sup>ST</sup>, ne pont pas l'enzyme avec une grande efficacité (6 %). Cependant, si l'on compare la bande de l'ODN-4<sup>ST</sup> libre après irradiation (piste 1) et celui de l'ODN-4<sup>ST</sup> libre sans irradiation (piste 2), on constate que l'ODN-4<sup>ST</sup> a réagi au cours de l'irradiation puisque la bande est moins intense (de 51%). Ce résultat est contradictoire avec la faible intensité de la bande correspondant au complexe ODN-4<sup>ST</sup>-DNMT3a. Pour expliquer ce phénomène, nous avons émis l'hypothèse qu'une fois le complexe ODN-4<sup>ST</sup>-DNMT3a formé, celui-ci va s'agréger, et ne sera plus capable d'entrer dans le maillage du gel SDS-PAGE, la majorité du complexe se trouverait alors dans le puits du gel ou dans la zone de *stacking*. Cette hypothèse est basée sur le fait que la protéine DNMT3a est capable de s'oligomériser autour d'une molécule de l'ADN<sup>244</sup> et que les oligonucléotides contenant de la 4<sup>ST</sup> ont tendance à s'agréger empiriquement.

En conclusion, les duplexes ODN-Dia, Bzp, mais aussi 4<sup>ST</sup> sont efficaces pour capturer DNMT3a dans ces conditions bien que le cas de la 4<sup>ST</sup> soit particulier dans ce contexte. Malgré le bon rendement de photomarquage observé avec cette protéine, la nécessité d'en utiliser en très large excès pour observer le photomarquage, ainsi que sa capacité à former des agrégats nous a conduits à envisager de travailler en parallèle sur une autre protéine modèle afin de pouvoir mettre au point les conditions de traitement des échantillons en vue de leur analyse en spectrométrie de masse. Au laboratoire, nous disposons de larges quantités d'histones et nous avons donc effectué nos mises au point sur l'histone H3.

### 4.3.2 *Etude des complexes duplex ODN-histone H3*

#### 4.3.2.1 *Photomarquage de l'histone H3 par les différentes sondes.*

L'histone H3 fait partie des 5 protéines histones impliquées dans la structure de la chromatine chez les cellules eucaryotes<sup>245</sup>. Il s'agit d'une protéine très conservée de 15kDa possédant une teneur élevée en arginine / lysine, lui permettant une interaction optimale avec l'ADN par liaison électrostatique.

De façon similaire aux expériences menées avec DNMT3a, nous avons vérifié la capacité des duplexes photoactivables radiomarqués au  $^{32}\text{P}$  (5 nM), à photoponter l'histone H3 (10 eq., 50 nM). La Figure 63 présente le résultat de ces expériences de photomarquage.

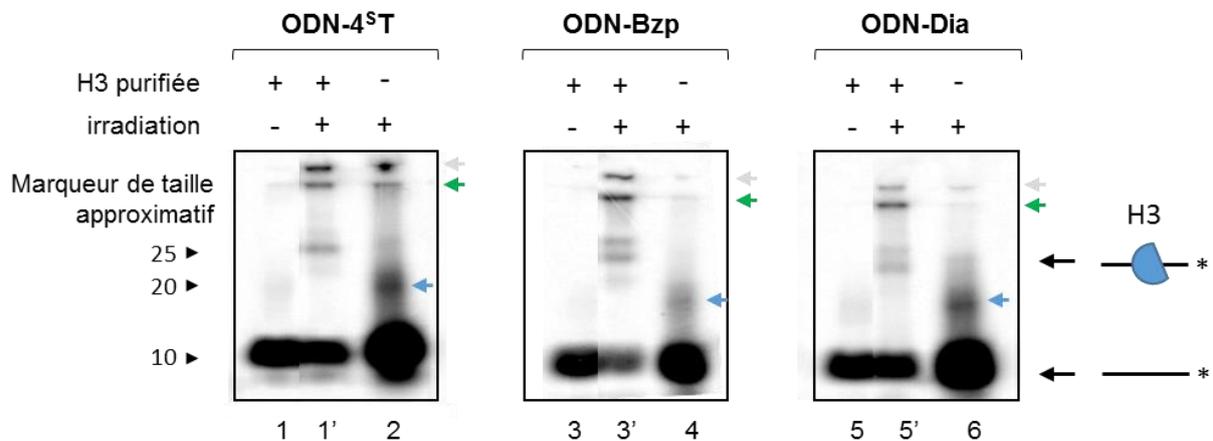


Figure 63 : Analyse par gel d'électrophorèse dénaturant 17% du photomarquage entre les sondes duplexes ODN et H3 purifiée (autoradiographie). La représentation schématique de l'ODN marqué (-\*) est indiquée à droite du gel, le demi-cercle bleu représente H3 lié covalamment à l'ODN marqué, la flèche grise représente les puits, la flèche verte le *stacking* et la flèche bleue le complexe d'intérêt

L'observation des pistes correspondant à l'expérience avec irradiation (1', 3' et 5') montre que toutes les sondes double brin semblent ponter l'histone H3 avec des rendements comparables. Cependant, on observe des bandes de différentes intensités en fonction de la sonde ODN utilisée au niveau du puits du gel (flèche grise) et après la zone de *stacking* (flèche verte). Ces bandes correspondent à des agrégats qui sont retenus par le maillage du gel. Les histones ont en effet tendance à s'agréger facilement. En réalisant une approximation du pontage total (agrégats compris), on peut estimer que les duplexes 4<sup>ST</sup> et Bzp sont les plus efficaces, vient ensuite la sonde possédant une diazirine. Il est intéressant de noter que les duplexes photoactivables irradiés sans protéine donnent lieu à la formation d'un complexe de plus haute masse moléculaire qui représente probablement un dimère d'oligonucléotide entre l'ODN simple brin photoactivable et son brin complémentaire. Ce complexe présent en faible quantité en l'absence de protéine, ne s'observe pas en présence de protéines, ce qui montre que les duplexes photoactivables ont une meilleure tendance à photoponter la protéine associée.

#### 4.3.2.2 Identification par spectrométrie de masse des complexes ODN-H3

Similairement à l'identification que nous avons réalisée sur les complexes ODN-Dia-hRPA, nous avons vérifié par spectrométrie de masse que les complexes observés sur la Figure 63 correspondaient bien à des complexes ODN-H3. Pour ce faire, 3 équivalents de l'histone H3 ont été incubés avec du duplexe ODN-4<sup>ST</sup> (100 pmol) pour reproduire le photomarquage de

H3. Le gel dénaturant (coloré au bleu de Coomassie) résultant de cette expérience est présenté sur la Figure 64.

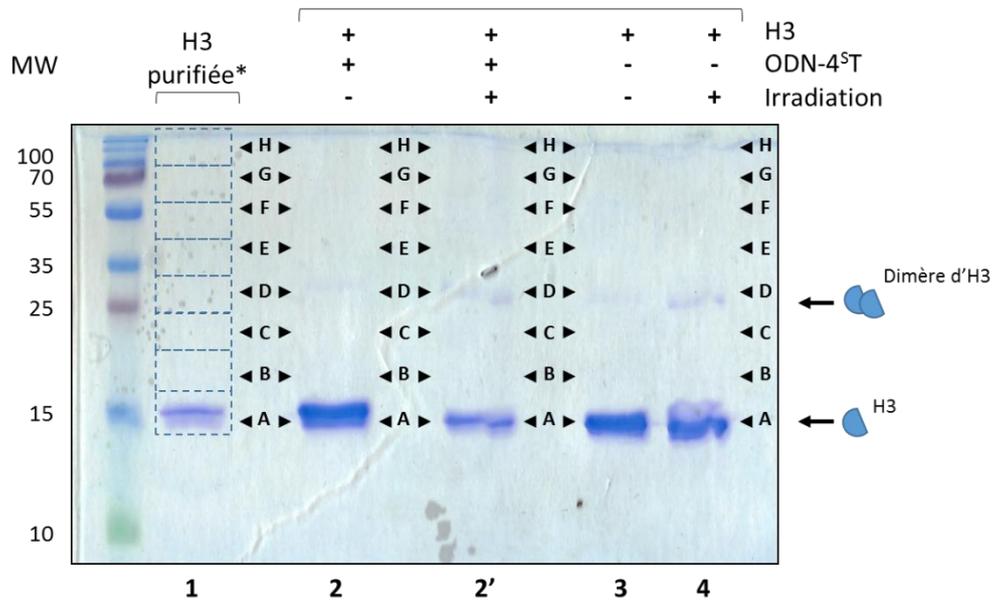


Figure 64 : Gel d'électrophorèse dénaturant SDS-PAGE de l'expérience de photomarquage du duplex ODN-4<sup>ST</sup> avec H3 (coloration au bleu de Coomassie). La bande correspondant à la masse moléculaire de H3 est indiquée, ainsi que celle pouvant correspondre à un dimère d'H3. Les pistes 1 et 3 représentent des expériences contrôles avec la protéine H3 purifiée ayant (piste 3) ou non (piste 1) suivi les mêmes conditions d'expériences. La piste 4 correspond à la protéine H3 irradiée. A, B, C, D, E, F, G et H correspondent à la position des zones excisées pour la digestion à la trypsine et à l'identification par spectrométrie de masse MALDI-TOF / TOF. Le patron d'excision des zones est uniquement représenté sur la piste 1 pour plus de clarté.

On observe peu de bandes de plus haute masse moléculaire par rapport aux bandes correspondant à l'histone H3 (autour de 15 kDa), mis à part une bande qui peut correspondre à un dimère d'histone H3 (autour de 30 kDa) sur la Figure 64, piste 2 à 4. En effet, l'histone H3 peut se dimériser en formant une liaison de type pont disulfure entre les sulfates des cystéines à la position 110 de deux chaînes polypeptidiques H3<sup>246</sup>. Il est toutefois surprenant que le pont disulfure de ce dimère résiste à l'étape de dénaturation de nos échantillons avant leur dépôt sur le gel, c'est-à-dire 5 min à 90 °C en milieu réducteur ( $\beta$ -mercaptoéthanol). L'absence de bande observable en coloration bleu de Coomassie, représentant un complexe ODN-H3, nous a poussés à analyser sans *a priori*, toute la partie supérieure à 15 kDa par spectrométrie de masse. Pour ce faire, nous avons découpé chaque piste en 8 bandes d'environ 0,5 cm au-delà de la masse de 15 kDa, soit de la bande A à la bande H (Figure 64).

#### 4.3.2.2.1 Digestion de la protéine H3 à la trypsine

La digestion trypsique de l'histone H3 produit des fragments peptidiques de petites masses (< 700 Da) du fait du grand nombre de lysines et d'arginines présentes dans la séquence d'acides aminés de cette protéine (pour rappel, la trypsine hydrolyse la liaison peptidique en C-terminal des arginines et lysines) comme on peut le voir sur la Figure 66A qui représente les sites de coupure de la trypsine sur la séquence de la protéine H3.

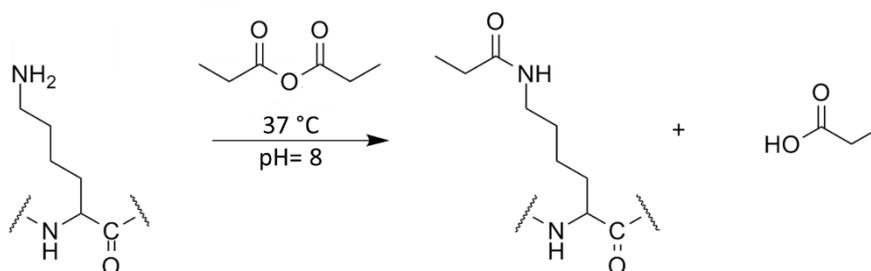


Figure 65 : Réaction de propionylation de la chaîne latérale d'une lysine

Un tableau récapitulatif des fragments théoriques obtenus après une digestion à la trypsine est présenté Annexe 1, p.192. Le principal inconvénient de générer des fragments de faible masse est que leur analyse par spectrométrie de masse est plus difficile pour différentes raisons : 1) en MALDI, leur signal peut être masqué par les signaux abondants de la matrice MALDI, 2) en ESI les signaux peuvent aussi être dans la région des signaux de solvant, 3) ces peptides sont difficiles à purifier par les techniques de séparation de type C18 puisqu'ils sont très hydrophiles ce qui ne permet pas de couplage avec des colonnes de phase inverse en nanoLC-ESI-MS/MS et 4) la recherche en banque de données est compliquée à cause du grand nombre de faux positifs générés. Dans le cas de l'analyse MALDI de nos échantillons digérés par la trypsine (matrice HCCA, acide alpha-Cyano-4-hydroxycinnamique) nous avons des difficultés à identifier la protéine H3 pure après digestion trypsique classique. En effet, la sélection de l'ion précurseur pour la fragmentation MALDI-TOF/TOF d'ions de faible masse conduit à sélectionner dans la fenêtre de sélection des ions de matrice qui viennent interférer dans les spectres MS/MS. Ainsi, les ions fragments ont une intensité proche du bruit de fond, ne permettant pas de réaliser une recherche en banque de données en utilisant les informations MS et MS/MS. Pour améliorer l'identification de H3, nous avons utilisé un protocole de propionylation de H3 comme décrit dans la littérature<sup>247,248</sup>. La réaction de propionylation des fonctions amines primaires est présentée Figure 65.

Ainsi, les résidus lysine sont modifiés chimiquement et ne sont plus reconnus par la trypsine qui agit alors en hydrolysant en C-terminal des résidus Arginine comme l'enzyme Arg C mais avec l'efficacité et la reproductibilité de la trypsine. Nous avons en parallèle également essayé de mettre en place un protocole de digestion avec l'enzyme Arg C mais sans succès et le prix élevé de l'enzyme Arg C nous a plutôt encouragés à optimiser le protocole de propionylation.

Dans ces deux cas, les coupures au sein de la séquence d'acides aminés de H3 se situent après les arginines comme on peut l'observer sur la Figure 66B, ce qui permet d'obtenir des fragments peptidiques de masse plus élevée et donc plus hydrophobes. Deux spectres de masse MALDI-TOF sont présentés en Annexe 2 (p.193) et permettent de comparer les fragments obtenus après digestion de H3 à la trypsine et après application du protocole de propionylation.



#### 4.3.2.2.2 Digestion de H3 par le protocole de propionylation couplé à la trypsine

Les zones A à H de toutes les pistes du gel de la Figure 64 ont été coupées en petits cubes d'1 mm<sup>3</sup>, lavées et séchées avant d'être placées dans une solution, appelée de propionylation. Cette solution composée d'un ratio 7/3 d'anhydride propionique et d'isopropanol dans 50 mM de NH<sub>4</sub>HCO<sub>3</sub> est maintenue à pH 8 par l'ajout d'une solution de NH<sub>4</sub>OH 28 % pendant les 15 minutes d'incubation à 37 °C pour que les amines primaires soient sous leur forme basique (non protonées) et capables d'attaquer le carbonyle de la fonction anhydride. En effet, le pH diminue au cours de cette réaction du fait de l'hydrolyse de l'anhydride propionique, réaction qui conduit à la production d'acide propionique qui acidifie le milieu rapidement et pourrait conduire à une réaction partielle. Il est donc important pendant le protocole de vérifier le pH et de l'ajuster en cours de réaction. Il faut cependant faire attention à ne pas dépasser pH 8-9 pour ne pas marquer d'autres résidus. Si le pH>10, le marquage de résidus ayant des pKa élevés est possible.

Une fois la réaction de propionylation réalisée, les bouts de gel sont lavés et « essorés » au maximum pour éliminer le réactif en excès. Cette étape est difficile et l'« essorage » des gels par ajout d'acétonitrile nécessite deux voire trois lavages en acétonitrile pure pour obtenir des gels essorés capable d'imbiber de nouveau les réactifs d'une seconde étape de propionylation. Cette deuxième étape est réalisée pour assurer un maximum de conversion des fonctions amines en propionylamines (> 95 %) <sup>249</sup>. La digestion à la trypsine peut ensuite être réalisée. Une réaction de propionylation est enfin réalisée en solution sur les peptides produits par la trypsine pour marquer les fonctions amines terminales libérées par la digestion tryptique. Après cette dernière étape, les digestats propionylés sont acidifiés et purifiés par microcolonne de dessalage préparées manuellement avec une résine de type POROS oligoR3 et analysés en spectrométrie de masse MALDI-TOF /TOF. La recherche en banques de données est réalisée *via* Mascot en tenant compte de cette modification en tant que modification fixe. Une représentation schématique de l'identification de la protéine H3 au sein du gel de la Figure 64 est présentée sur la Figure 67A.

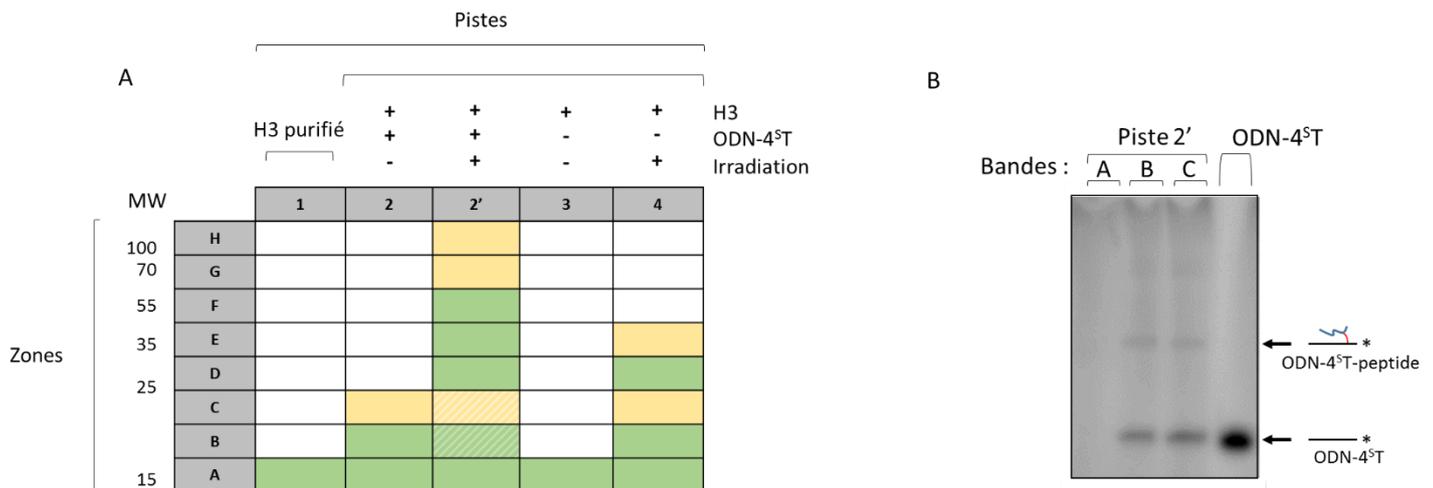


Figure 67 : A) Identification de la protéine H3 par spectrométrie de masse sur le gel présenté sur la Figure 64. Les zones vertes correspondent à une identification de la protéine H3 avec des scores Mascot élevés, les zones jaunes à une identification plus difficile et les zones blanches correspondent à des bandes de gel dans lesquelles la protéine H3 n'est pas identifiée. Les zones hachurées correspondent aux zones où de l'ADN a été détecté. B) Gel dénaturant correspondant au radiomarquage des digestats provenant des zones A, B et C de la piste 2' de la Figure 64

Les deux contrôles correspondant à H3 purifiée sont représentés par les pistes 1 et 3 (la piste 3 correspondant à H3 dans les mêmes conditions d'analyse que les autres pistes). Sur ces pistes, H3 est uniquement identifiée au niveau de la zone A ce qui est en accord avec la migration de cette protéine purifiée. Par contre, lorsque cette protéine est irradiée, elle est identifiée de la zone A à la zone E comme on peut le voir sur la piste 4 et semble donc « traîner » sur le gel. L'irradiation semble favoriser la formation du dimère d'H3 (bande visible à 30 kDa sur la Figure 64, piste 4) et « ralentir » la migration de H3. Un « ralentissement » de la migration est aussi visible lorsque la protéine H3 est mise en présence de l'ODN-4<sup>ST</sup> sans irradiation puisque la protéine est identifiée dans les zones A à C (piste 2). Ceci pourrait être dû à des interactions non covalentes entre l'ODN-4<sup>ST</sup> et H3 qui seraient conservées sur gel dénaturant. Ces deux phénomènes de retard sur gel se retrouvent sur la piste 2' avec irradiation et en présence de l'ODN-4<sup>ST</sup>, où la protéine est identifiée sur tout l'ensemble de la piste. Cela peut s'expliquer par le fait que l'ODN-4<sup>ST</sup> peut se lier de façon covalente soit à H3 soit au dimère d'H3. La formation d'une liaison covalente avec le dimère (soit 10 kDa d'ODN + 30 kDa de dimère) explique l'identification de la protéine dans les bandes E et F de la piste 2' (respectivement à 40 et 55 kDa). Plus généralement, H3 est une protéine qui a tendance à former des agrégats, qui peuvent expliquer la présence de cette protéine dans les zones G et H de la piste 2'.

L'identification de la protéine H3 dans des zones de masse moléculaire élevées, jusqu'à 55 kDa pour la bande F de la piste 2' implique la formation de complexe covalent entre l'ODN-4<sup>ST</sup> et H3. Nous avons aussi vérifié ce résultat en radiomarquant au <sup>32</sup>P les échantillons de peptides issus des bandes digérées à la trypsine de la piste 2' (zones 2A à 2H). Lors de ce radiomarquage l'ODN-4<sup>ST</sup>, tout comme son brin complémentaire, s'il est toujours présent dans l'échantillon,

sont radiomarqués. Le principe de cette expérience est le suivant : si l'ODN-4<sup>S</sup>T est lié de façon covalente à un peptide provenant d'un digestat d'H3, il va migrer plus lentement que l'ODN libre. De plus, si une bande du gel contient de l'ODN-4<sup>S</sup>T celui-ci sera détecté par autoradiographie révélant ainsi la présence ou non de complexe ODN-4<sup>S</sup>T -H3. Le résultat de cette expérience est présenté sur la Figure 67B où seuls les radiomarquages des zones A, B et C de la piste 2' sont montrés. Aucune bande n'est observée sur la piste A, il n'y a donc pas d'ODN-4<sup>S</sup>T dans cette zone. Par contre, on observe deux bandes sur les pistes B et C. La bande la plus basse correspond à l'ODN-4<sup>S</sup>T ou à son brin complémentaire (les deux ayant la même masse moléculaire de 10 kDa), et la bande plus haute, « retardée » correspond à un complexe ODN-4<sup>S</sup>T-peptide provenant de la digestion d'H3. Pour les zones D à H, nous n'avons pas pu détecter d'ODN en radioactivité (non montré), ce qui pourrait s'expliquer par une trop faible quantité de matière en ODN pour pouvoir être détectée selon notre approche.

#### 4.3.3 Photomarquage avec la sonde duplexe ODN-4<sup>S</sup>T dans un mélange protéique complexe.

De façon similaire aux expériences que nous avons montrées avec l'ODN-4<sup>S</sup>T simple brin et hRPA, nous avons tenté de capturer l'histone H3 ainsi que d'autres protéines affines pour l'ADN double brin au sein d'un milieu complexe. Le duplexe ODN-4<sup>S</sup>T radiomarqué a été incubé avec un extrait cellulaire de cellules HEK 293, l'expérience de photomarquage a ensuite été réalisée puis les échantillons ont été déposés sur un gel dénaturant (Figure 68).

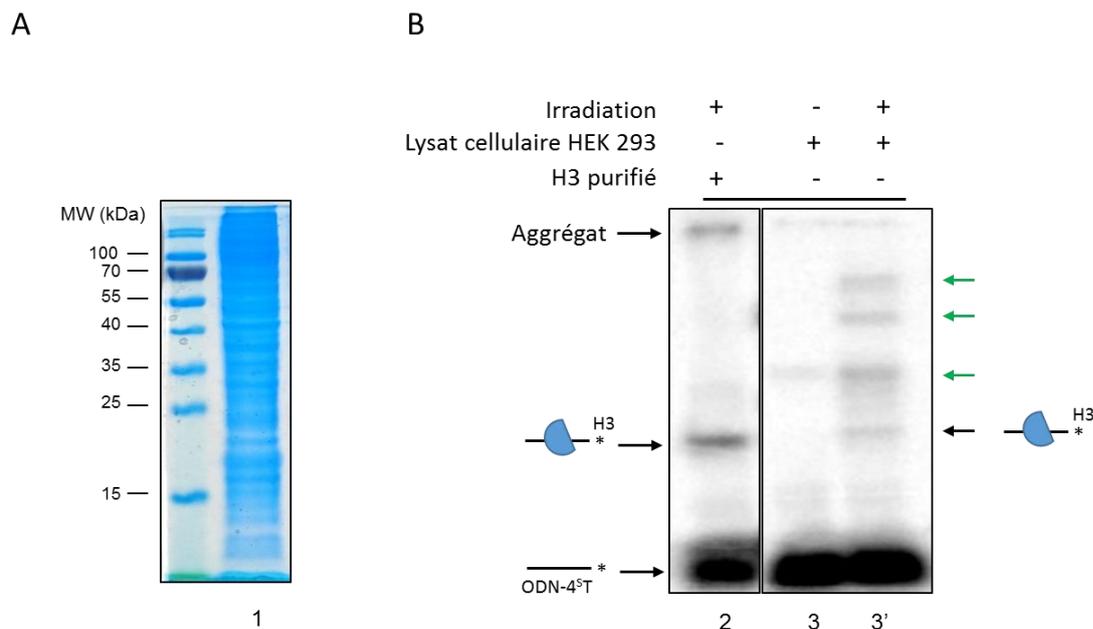


Figure 68 : A. Gel d'électrophorèse dénaturant 15 % où les protéines présentes dans le lysat cellulaire obtenu à partir des cellules HEK 293 sont colorées au bleu de Coomassie. B. Analyse par gel d'électrophorèse SDS-PAGE 15% de la sonde ODN-4<sup>S</sup>T après irradiation en présence d'un lysat cellulaire de cellules HEK 293 (autoradiographie). La représentation schématique de l'ODN marqué (-\*) est indiquée à droite du gel et le demi-cercle représente H3 (bleu) lié covalamment à l'ODN marqué.

Sur la piste correspondant à l'expérience de photomarquage (Figure 68B, piste 3'), on observe plusieurs bandes de masse moléculaire supérieure à l'ODN-4<sup>S</sup>T libre. L'une de ces bandes vers 25 kDa pourrait correspondre au photomarquage de la protéine H3. En effet, cette bande migre à la même masse moléculaire que le complexe ODN-4<sup>S</sup>T-protéine H3 comme on peut le constater par comparaison avec la piste 2 qui correspond au photomarquage avec la protéine H3 purifiée. On peut aussi noter la présence d'une bande dans la piste 3 correspondant à l'expérience sans irradiation et dans la piste 3' (flèche rouge). Cette bande pourrait correspondre à une protéine liée fortement à l'ADN de façon non covalente. Les deux bandes indiquées par les flèches vertes sur la piste 3' n'ont pas pu être identifiées, mais ont des masses apparentes de 40 et 60 kDa, respectivement. La reproduction de cette manipulation dans le but d'identifier les complexes ODN-protéines détectés a été réalisée, nous nous sommes heurtées à des difficultés lors de l'enrichissement de ceux-ci pour leur analyse par spectrométrie de masse. L'identification de ces complexes nécessite donc l'optimisation de l'étape d'enrichissement des complexes au sein d'extraits cellulaires eucaryotes, étape qui n'a pas pu être réalisée avant l'écriture de ce manuscrit.

#### *4.3.4 Conclusion sur la comparaison et l'évaluation des sondes photoactivables double brin*

L'étude de la capacité de nos sondes photoactivables à capturer des protéines affines pour l'ADN double brin s'est accompagnée d'une difficulté à définir une protéine modèle pour mettre au point nos conditions de photomarquage. En effet, la DNMT3a n'est pas assez affine, et son utilisation nécessite des quantités trop importantes en protéine pour que l'étude soit ergonomique. L'histone H3, quant à elle, présente des problèmes d'agrégation et d'analyse en spectrométrie de masse du fait de sa haute teneur en acides aminés basiques. Dans ces conditions, il est difficile d'obtenir des conclusions aussi claires que dans l'étude dans un contexte simple brin. Toutefois, certains points ressortent :

- Les sondes photoactivables 4<sup>S</sup>T, Bzp et Dia pontent les protéines modèles avec quasiment la même activité, aucune ne se démarque des autres.
- Les expériences de photomarquage sur extraits cellulaires humains ont démontré la capacité de la méthode à ponter des protéines avec un degré élevé de spécificité des sondes photoactivables car peu de protéines ont été piégées notamment la protéine modèle H3.

Cette étude démontre aussi la versatilité des agents de photomarquage, comme on peut le voir avec l'ODN-Bzp, qui n'était pas efficace sur le modèle simple brin alors qu'il est efficace dans un contexte double brin.

Pour compléter cette étude, il faudrait envisager l'utilisation d'autres protéines modèles. Il serait intéressant de travailler sur des protéines ayant déjà été caractérisées comme affines pour les modifications épigénétiques de l'ADN. Cela permettrait la comparaison des sondes photoactivables possédant une 5mC ou une 5hmC, et la mise en place de conditions de photomarquage plus adaptées pour notre système de piégeage. On peut notamment citer les protéines UHRF1 et TET. L'utilisation de protéines modèles plus adaptées permettrait aussi d'optimiser les conditions d'enrichissement des espèces pontées lors du photomarquage dans un mélange complexe de protéines à l'aide du groupement biotine. Cet enrichissement est plus difficile sur des extraits cellulaires eucaryotes, du fait de leur plus grande complexité.

A la fin de cette partie, trois de nos sondes sont efficaces et pour la suite, nous nous sommes focalisés sur la 4-thiothymine (4<sup>S</sup>T) qui comporte un certain nombre d'avantages :

- Sa structure analogue aux nucléobases qui va limiter les modifications structurelles de notre molécule d'ADN.
- Son caractère « *zero-length* » où seules les protéines en interaction directe avec l'ADN peuvent être photomarquées.
- Sa photoréactivité. Toutes nos expériences précédentes ont montré que la 4<sup>S</sup>T est réactive que le milieu soit complexe ou non, soit dans différentes conditions.
- Son accessibilité. En effet, la 4<sup>S</sup>T est commercialisée sous forme de triphosphate et ne nécessite pas d'étape de synthèse organique supplémentaire.

Dans une seconde partie de ce travail, nous nous sommes intéressés à essayer de caractériser au niveau moléculaire ces complexes ODN-protéines par spectrométrie de masse dans le but d'obtenir à terme plus d'information sur le site d'interaction de l'ODN avec une protéine d'intérêt.

## 5 Caractérisation au niveau moléculaire de complexes ODN-protéines par spectrométrie de masse MALDI-TOF

### Avant-propos

Dans le chapitre précédent, nous avons mis au point des outils permettant de créer des complexes covalents ADN-protéines et de les purifier. Pour pouvoir aller plus loin avec ces ODN photoactivables, nous avons mené une étude méthodologique destinée à caractériser au niveau moléculaire les complexes photopontés ADN-protéine. En effet, la caractérisation précise du site d'interaction permettrait 1) de comprendre quels sont les résidus mis en jeu dans l'interaction ADN-protéine et 2) d'obtenir des informations sur la réactivité non décrite à ce jour, de la photosonde 4-thiothymine (4<sup>S</sup>T) vis-à-vis des acides aminés.

Afin de simplifier cette étude nous avons d'abord mis au point la méthodologie non pas sur des complexes ADN-protéine mais sur un modèle ADN-peptide. Nous avons utilisé un peptide précédemment étudié au laboratoire correspondant au domaine *N*-terminal de l'histone H3 humaine<sup>250</sup> et nous avons fait l'hypothèse d'une interaction électrostatique potentielle (à défaut d'être spécifique) entre cette séquence peptidique et la séquence oligonucléotidique 30-mère photoactivable contenant la 4<sup>S</sup>T. En parallèle de ce modèle, nous avons aussi travaillé sur les complexes ADN-protéine après avoir effectué une étape de digestion à la trypsine de la protéine permettant ainsi une analyse par spectrométrie de masse comme dans une étude protéomique classique.

Bien qu'il existe de nombreux exemples dans lesquels la spectrométrie de masse a été appliquée avec succès à l'étude des complexes acide nucléique-protéine/peptide en MALDI<sup>251,252,253</sup>, cela reste une tâche difficile puisqu'un certain nombre de défis analytiques doivent être surmontés. En effet les faibles rendements de photomarquage impliquent de recourir à des méthodes d'enrichissement pour augmenter la quantité relative de matériel hétéro-conjugué à analyser afin de permettre la détection de cette espèce minoritaire par spectrométrie de masse. Plusieurs études ont montré que l'analyse de produit « brut » de photomarquage à l'aide de conditions standards d'analyse MALDI ne permettait pas la détection des complexes ADN-peptide, seuls les ADN ou les peptides n'ayant pas réagi étaient détectés<sup>254,255,256</sup>. Par ailleurs, dans un complexe ADN-peptide, la présence à la fois d'un long squelette de phosphates chargés négativement (groupements phosphates des liaisons phosphodiesters, pKa ≈ 1,5 à 2) et d'une partie peptidique généralement chargée positivement en conditions acides ou neutres peut provoquer la suppression de l'ion correspondant au complexe<sup>252,257</sup>. Par conséquent, l'efficacité

de l'ionisation des espèces hétéro-conjuguées est réduite par rapport aux constituants isolés ADN ou peptide. Un grand excès d'espèces peptidiques et oligonucléotidiques n'ayant pas réagi peut induire une suppression de l'ionisation, compromettant ainsi la détection des complexes<sup>252,258</sup>. De plus, les oligonucléotides chargés négativement sont souvent prompts à former des adduits avec les différents sels présents dans les préparations biologiques (eau, tampons, surfaces métalliques, poussières, tubes en verre etc...)<sup>259</sup>. Il est important de réduire au maximum cette cationisation pour l'analyse en spectrométrie de masse. En effet, si l'on combine un faible rendement d'ionisation des hétéroconjugués ADN/peptide avec la dilution du signal par cationisation *via* différents sels comme le sodium ou le potassium, l'intensité du signal diminue considérablement pour ne plus sortir du bruit de fond. Il a été montré dans la littérature, entre autre par Nordhoof *et al.* que l'on pouvait échanger les métaux alcalins sous forme cationique tels que Na<sup>+</sup> et K<sup>+</sup> par des ions ammonium<sup>260</sup>.

De plus, les espèces oligonucléotidiques et peptidiques ou protéiques nécessitent des conditions différentes pour l'ionisation en MALDI. L'ADN et les oligonucléotides possèdent une forte polarité, la plus élevée parmi les biomolécules. En effet, avec le long squelette phosphaté qui a tendance à se déprotoner très facilement et les bases hétérocycliques azotées, plutôt basiques (dans nos conditions d'analyse à pH=3,8), et donc facilement protonables, les interactions électrostatiques entre molécules sont nombreuses et fortes. Il est ainsi important de pouvoir isoler les molécules uniques en rompant les liaisons intermoléculaires électrostatiques pour leur analyse en MALDI, grâce en particulier à l'incorporation efficace de molécules uniques dans la matrice. Les protéines ou peptides sont souvent plus simples à analyser puisque les interactions mises en jeu sont plus souvent de type liaisons hydrogènes donc plus faibles. Il est ainsi plus aisé d'analyser ces biomolécules uniques en spectrométrie de masse. Les conditions d'ionisation et d'analyse de ces biomolécules sont différentes : les oligonucléotides sont ionisés en utilisant des conditions neutres ou légèrement alcalines et analysés en mode ions négatifs, alors que les peptides sont ionisés en conditions acides et analysés en mode ions positifs<sup>261</sup>. De ce fait, la question se pose de savoir quel mode d'ionisation (ions positifs ou négatifs) et quelle matrice utiliser pour un complexe unique constitué de ces deux différentes classes de biomolécules en MALDI. Jensen *et al.*, ont répertorié les matrices communément utilisées pour l'une ou l'autre de ces biomolécules<sup>252</sup> (Tableau 2).

Matrice	Peptide et protéine	Oligo-nucléotide	Réf.
Acide sinapinique	+++++	+	13
Acide 4-hydroxy- $\alpha$ -cyano cinnamique (HCCA)	+++++	+	14
Acide 2,5-dihydroxybenzoïque (DHB)	++++	++	15
Acide férulique	+++	+++	13
2,4,6-Trihydroxyacetophénone	++	+++++	
Acide 3-hydroxypicolinique (3HPA)	+	+++++	16
<b>L'adéquation générale d'une matrice donnée pour une classe de composé donnée est indiquée subjectivement par une échelle qualitative qui varie de + (faible) à +++++ (forte).</b>			

Tableau 2: Matrices communément utilisés en analyse MALDI de protéines, peptides et oligonucléotides d'après Jensen et al, 1996<sup>252</sup>

Ces données indiquent qu'aucune des matrices classiquement utilisées ne permet une ionisation optimale à la fois des oligonucléotides et des peptides. Quelle que soit la matrice MALDI, l'analyse va tendre à diminuer la détection de l'un des partenaires d'interaction. De façon logique, on peut supposer qu'une fois qu'un complexe ADN-peptide ou ADN-protéine est formé, le constituant qui présente la masse la plus élevée l'emporte en termes de compétition à l'ionisation et détermine les conditions d'analyse. Par exemple, il a été montré que pour les complexes avec une entité oligonucléotidique plus longue, l'analyse en mode négatif donne les meilleurs résultats<sup>257,262,263</sup>. Nous supposons donc dans le cas de complexes ADN-protéine (dans notre cas  $MM_{ADN} < MM_{protéine}$ ) que ces complexes se comporteront plutôt comme une protéine alors que dans le cas de complexes ADN-peptides (dans notre cas  $MM_{ADN} > MM_{peptide}$ ), ceux-ci se comporteront plutôt comme des oligonucléotides. Le mode négatif peut s'avérer un peu délicat à mettre en œuvre et les instruments que nous avons utilisés n'étaient pas optimisés pour ce type d'analyse. En effet, l'expertise du laboratoire concerne principalement l'analyse des peptides et protéines en mode ions positifs et nous avons donc dû mettre au point les conditions d'ionisation et de détection des oligonucléotides par MALDI-TOF en mode négatif.

Pour résumer, il est nécessaire d'enrichir et de dessaler les hétéroconjugués ADN-peptides ou protéines pour améliorer leur détection par spectrométrie de masse. Dans le cas d'une ionisation MALDI, le choix de la matrice est important. Les paramètres instrumentaux du spectromètre de masse doivent être optimisés quasiment au cas par cas en fonction de la nature physicochimique et de la masse des hétéroconjugués. Bien qu'il existe de nombreuses études où des complexes ADN-peptide ont été caractérisés par spectrométrie de masse après

photomarquage, chaque étude a été unique en ce qui concerne l'agent photoactivable utilisé, la stratégie d'enrichissement, la mise en place des paramètres instrumentaux. Il n'existe donc pas à ce jour, de protocole unique pour l'analyse des hétéroconjugués ADN-peptides.

### **Etude des acides nucléiques par spectrométrie de masse MALDI-TOF**

5.2 De nombreuses matrices pour l'analyse MALDI des acides nucléiques ont été examinées dans la littérature. Les premières études MALDI avec des acides nucléiques impliquaient principalement l'essai de matrices classiquement utilisées pour l'étude des protéines. Après la découverte de l'acide nicotinique pour l'analyse des protéines, ce composé s'est révélé efficace pour l'analyse de petits oligonucléotides (deux à six bases) ainsi que pour un ARN de transfert de levure beaucoup plus long (environ 24 kDa)<sup>264</sup>. Plusieurs autres matrices utilisées pour l'analyse des protéines, y compris l'acide sinapinique (SA) et l'acide 2,5-dihydroxybenzoïque (DHB), ont également permis de générer des ions à partir de divers oligonucléotides. Cependant, l'utilisation de ces matrices a généralement été limitée aux homopolymères de la thymine (polydT) et à des échantillons constitués de bases mixtes relativement courts (moins de 10 nucléotides)<sup>265,266,267</sup>. En 1993, Pielec *et al.* ont découvert que la 2,4,6-trihydroxyacétophénone (THAP) permettait l'analyse d'oligonucléotides naturels et modifiés de 2 à 24 nucléotides<sup>268</sup>. De même, Currie et Yates ont testé environ 30 matrices différentes et ont constaté que le 3-hydroxy-4-méthoxybenzaldéhyde permettait l'analyse de grands oligomères composés de bases mixtes, y compris un oligonucléotide simple brin de 60 nucléotides de masse moléculaire supérieure à 18 kDa<sup>269</sup>. La même année, Schneider et Chait testent 48 matrices différentes<sup>270</sup>. Parmi elles, 20 ont permis l'analyse d'un oligonucléotide comprenant 18 thymines, mais aucun signal d'ion moléculaire n'a pu être observé dans le cas d'un oligonucléotide constitué de 18 guanines, quelle que soit la matrice utilisée.

Lors de travaux effectués au début des années 90, seules trois matrices : la 2,4,6-trihydroxyacétophénone (THAP), le 3-hydroxy-4-méthoxybenzaldéhyde et l'acide 3-hydroxypicolinique (3-HPA) se sont montrées suffisamment robustes pour l'analyse d'oligonucléotides. Toutefois, il faut garder à l'esprit que dans la plupart de ces études, les séquences d'oligonucléotides étudiées étaient relativement courtes (inférieures à 30 nucléotides). Typiquement, plus un oligonucléotide est long, plus il est difficile à caractériser. Cela est dû, entre autre, à la tendance du squelette phosphodiester de l'ADN à se fragmenter lors du processus MALDI<sup>271</sup>. Ce processus de fragmentation est engendré par la protonation

des bases. Plus l'affinité protonique (AP) de la base est élevée plus la fragmentation est importante. Ainsi les polydT sont les oligonucléotides (oligo dT) les plus stables avec une  $AP(dT) = 224,9 \text{ kcal/mol} < AP(dG) = 234 = AP(dA) = 233 \text{ kcal/mol} = AP(dC) = 233 \text{ kcal/mol}$ <sup>272</sup>.

Actuellement, la matrice qui permet l'étude des oligonucléotides les plus longs est l'acide 3-hydroxypicolinique (3-HPA) (Figure 69), qui a été utilisé efficacement pour l'analyse d'oligonucléotides à bases mixtes allant jusqu'à 89 nucléotides<sup>273,274</sup>. Introduite pour la première fois par le groupe de Becker en 1993, la matrice 3-HPA est ainsi la matrice la plus utilisée pour l'analyse de molécules d'ADN. Il s'agit d'une matrice sous forme de cristaux solides qui présente une faible pression de vapeur et des chromophores UV permettant l'étape de désorption nécessaire à l'expérience MALDI.

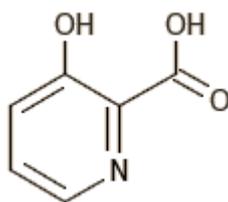


Figure 69 : Structure de l'acide 3-hydroxypicolinique (3-HPA) utilisé comme matrice MALDI

Un inconvénient de la matrice 3-HPA est l'hétérogénéité de cristallisation des dépôts. En effet, les analytes ne se dispersent pas à travers la matrice de manière homogène, les cristaux sont hétérogènes non seulement en terme de répartition au sein d'un même dépôt mais aussi entre deux dépôts différents. Cela conduit à une faible reproductibilité des analyses et un temps d'analyse très long et difficilement automatisable puisqu'il est nécessaire de trouver des « points chauds » (ou « hot spot » en anglais) riches en analytes pour observer un signal en MALDI-TOF.

5.3

### Mise au point des expériences MALDI pour l'analyse des oligonucléotides avec la matrice 3-HPA

#### 5.3.1 Optimisation des dépôts : mise au point du ratio matrice : échantillon

Dans un premier temps, la mise au point des conditions de cristallisation a été réalisée avec deux oligonucléotides polydT contenant 21 (ODN 21 T) ou 31 dT (ODN 31 T). Après plusieurs essais nous sommes arrivés à déterminer des conditions optimisées pour l'analyse MALDI-TOF de ces oligodT qui nous ont servi par la suite pour étalonner le spectromètre de masse. Un

point important que nous avons déjà évoqué est la grande hétérogénéité des dépôts obtenus avec la 3-HPA. Pour maximiser la probabilité d'obtenir des cristaux conduisant à des signaux en MALDI, nous avons augmenté le ratio des volumes matrice : analyte à un ratio (9 : 1). A partir de ce mélange, nous avons réalisé non pas un mais plusieurs dépôts de 1 à 2  $\mu\text{L}$  sur la cible MALDI. Ces optimisations ont permis d'améliorer la détection de l'ODN 31 T, environ un dépôt sur 6 conduisait à un signal détectable en MALDI-TOF.

Il a été montré dans la littérature, que ce problème d'hétérogénéité pouvait être contourné en réduisant le diamètre du dépôt à celui du laser sur la cible<sup>275</sup>. De tels dépôts ne peuvent pas être réalisés avec des pipettes standard. Nous nous sommes donc contentés de limiter le diamètre des dépôts par simple nettoyage préalable des cibles MALDI avec du metal-polish. Cette pâte contient des grains très fins qui conduisent à abraser la surface de la cible MALDI, éliminant les restes de dépôts de matrice et la couche d'oxyde de surface créant ainsi une surface vierge. La rugosité de la surface métallique devient donc plus faible, amenant à une surface de nature plus hydrophobe. Ainsi, les dépôts s'étalent beaucoup moins sur la surface de la cible, conduisant à des surfaces de contact plus faibles. Après évaporation du solvant le diamètre des dépôts est ainsi minimisé.

Comme mentionné précédemment, une faible résolution du signal (élargissement des pics) peut être due à la présence de sels ou de cations métalliques dans les solutions. Nous avons donc travaillé dans l'eau doublement distillée en ajoutant des additifs tels que du citrate de diammonium ou des résines Dowex échangeuses de cations en évitant au maximum de travailler dans du verre.

### 5.3.2 Optimisation des dépôts : dessalage des oligonucléotides

#### 5.3.2.1 Le citrate de diammonium comme additif dans les dépôts MALDI

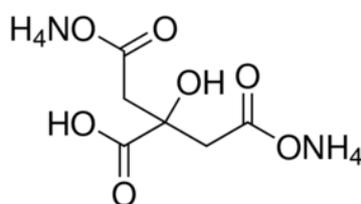


Figure 70 : Formule développée du citrate de diammonium

Les oligonucléotides chargés négativement du fait de leurs groupements phosphates ont une forte tendance à former des liaisons électrostatiques avec des cations métalliques présents dans les milieux *in vitro* ou *in vivo*. Les ions ainsi obtenus et analysés en mode ions négatifs peuvent

être de différentes natures et avoir des rapports m/z différents et déplacés vers les plus hautes masses :  $[M-H]^-$ ,  $[M-2H+Na]^-$ ,  $[M-2H+K]^-$  ainsi que des combinaisons de ces deux cations. La présence de ces adduits cationiques élargit, dilue et écrase le signal obtenu en spectrométrie de masse. Cela conduit à une détection plus difficile (rapport signal/bruit faible), un élargissement du signal qui conduit à une précision en masse et une résolution du signal abaissées. L'ajout de citrate de diammonium (Figure 70) réduit la production d'adduits sur l'ADN par échange entre les cations métalliques ( $Na^+$ ,  $K^+$ ) et les ions ammonium ( $NH_4^+$ ). Il y a ensuite dissociation des paires d'ions ammonium phosphate au moment de la désorption en MALDI. Contrairement à un ion  $Na^+$  ou  $K^+$ , le  $NH_4^+$  peut transférer un proton au groupement phosphate pour former l'acide libre et une molécule d'ammoniac, améliorant ainsi la résolution des signaux obtenus pour des oligonucléotides<sup>276</sup>. Ce processus est représenté par l'équation suivante :



où n représente le nombre de charges et indique que cet échange a lieu sur des ions positifs ou négatifs<sup>272</sup>.

Nous avons cherché à optimiser la quantité de citrate de diammonium dans les dépôts pour réduire les adduits tout en conservant des conditions optimales pour la cristallisation de la matrice 3-HPA. Nous avons fait varier la quantité de citrate de diammonium ajoutée entre 1,4 mg/mL et 2,2 mg/mL dans la solution de matrice : échantillon (9 : 1). La variation de concentration en citrate de diammonium dans les dépôts n'a pas conduit à un effet observable sur les spectres MALDI. Nous avons donc cherché une autre façon d'éliminer les cations métalliques en utilisant une résine échangeuse de cations.

### 5.3.2.2 Utilisation de billes échangeuses de cations : résine Dowex

La résine commerciale Dowex (50 wx 8-400) est un échangeur cationique fort qui permet l'échange des ions  $K^+$  et  $Na^+$  par des  $H^+$ . Le groupe fonctionnel acide sulfonique est attaché à un polymère de divinylbenzène-styrène. L'activation de la résine se fait par une solution d'acide chlorhydrique (2N) pour charger la résine en protons. La résine activée est incubée avec l'échantillon d'ADN et une fois l'échange réalisé, le surnageant est prélevé et préparé avec la matrice MALDI avant analyse.

Nous avons choisi de tester la présence des billes Dowex directement dans les dépôts MALDI pour éviter les échanges avec les cations métalliques de la surface de la cible. Qualitativement, nous avons remarqué une plus grande probabilité de produire un signal d'oligonucléotide avec peu d'adduits cations (signal plus intense et plus résolu). Peut-être *via* un double effet : l'échange des cations contre des protons et la concentration de l'échantillon sur une surface plus réduite, dans les zones du dépôt ne présentant pas de billes. La réduction de la surface de l'échantillon pose toutefois quelques difficultés puisqu'il y est empiriquement plus difficile d'y obtenir des cristaux. Ces billes sont incorporées dans les cristaux et peuvent être analysées sous vide, dans la source MALDI (Figure 71).



Figure 71: Dépôt MALDI présentant des cristaux de 3HPA en présence de billes Dowex (flèche blanche)

Les deux techniques de dessalage présentées sont relativement équivalentes quant à l'effet observé sur les spectres MALDI. Nous avons donc choisi d'utiliser préférentiellement la méthode de l'ajout de citrate de diammonium dans la matrice pour sa facilité de mise en œuvre.

La reproductibilité de cristallisation des cibles n'étant toujours pas optimale, nous avons mis au point un protocole d'activation de la cristallisation permettant l'obtention de cristaux conduisant à un signal détectable en MALDI-TOF.

### 5.3.3 *Optimisation des dépôts : mise au point d'un protocole de cristallisation de la matrice 3-HPA en présence de citrate de diammonium*

Pour ce faire, nous avons utilisé la technique de l'amorçage de cristallisation que l'on retrouve aussi bien dans des techniques de purification chimique comme la re-cristallisation, que dans des méthodes de cristallisation sélective où l'introduction d'un cristal permet le contrôle des futurs cristaux formés. Les étapes de ce protocole sont présentées sur la Figure 72. Après avoir réalisé le dépôt sur la cible, il faut attendre le début de cristallisation de la matrice 3-HPA, sur la périphérie du dépôt pour prélever un très petit volume qui va pouvoir cristalliser plus rapidement du fait du très faible diamètre de cette gouttelette, comme indiqué par Nordhoff *et al.*<sup>275</sup>(Figure 72, images A et B). Ce cristal est ensuite réinjecté dans le dépôt initial et amorce

la cristallisation au point d'insertion comme présenté sur l'image C. L'évaporation complète du solvant conduit à un cristal contenant des aiguilles bien dessinées comme présenté sur l'image D.

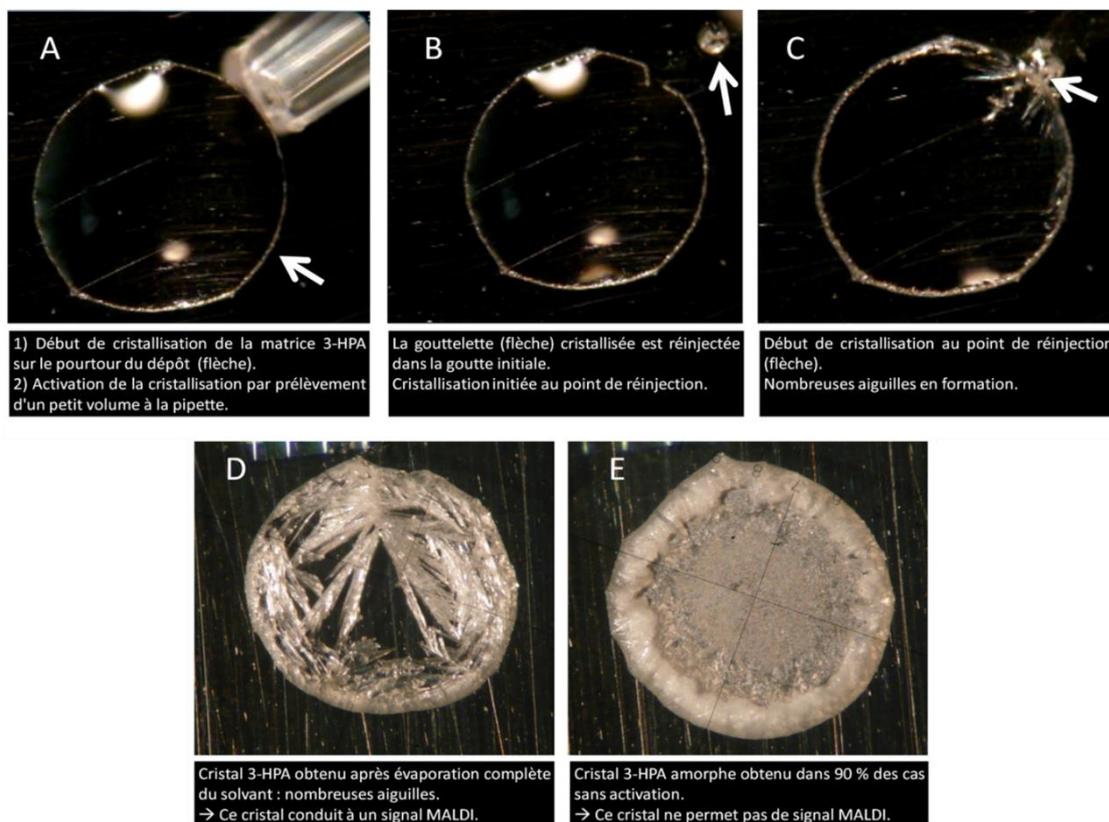


Figure 72 : A, B, C et D) Etapes de l'amorçage de la cristallisation ; E) exemple de cristal amorphe non activé.

L'amorçage de la cristallisation, bien que simple nécessite une certaine dextérité pour sa réalisation. Les étapes du protocole sont très dépendantes des conditions de température à laquelle les dépôts sont effectués sur la cible MALDI. Les cristaux sont en effet réalisés sous microscope et la lampe chauffée la cible. Dans ce cas, le solvant présent au sein du dépôt s'évapore trop rapidement, conduisant à un cristal amorphe ne permettant pas l'obtention d'un signal en MALDI (Figure 72, image E). La cible MALDI préalablement abrasée au metal-polish est ainsi placée à 4 °C avant tout dépôt, ce qui permet d'éviter ce phénomène lors de la préparation des dépôts.

L'amorçage de cristallisation combiné à une solution de matrice contenant du citrate de diammonium à 2,2 mg/mL, permet de réaliser des dépôts analysables dans 9 cas sur 10 (au lieu de 1 sur 6 sans amorçage). Les spectres MALDI-TOF linéaire, mode ions négatifs obtenus sont présentés sur la Figure 73. Ces spectres ont été réalisés avec un étalonnage par défaut pour la mise au point. Par la suite l'étalonnage a été réalisé en externe proche en utilisant un

oligonucléotide polyT 31-mère (ODN 31 T) comme étalon. L'effet de l'amorçage de cristallisation est spectaculaire sur l'incorporation des oligonucléotides dans la matrice 3-HPA comme présenté sur le spectre B en comparaison avec le spectre A réalisé sans amorçage.

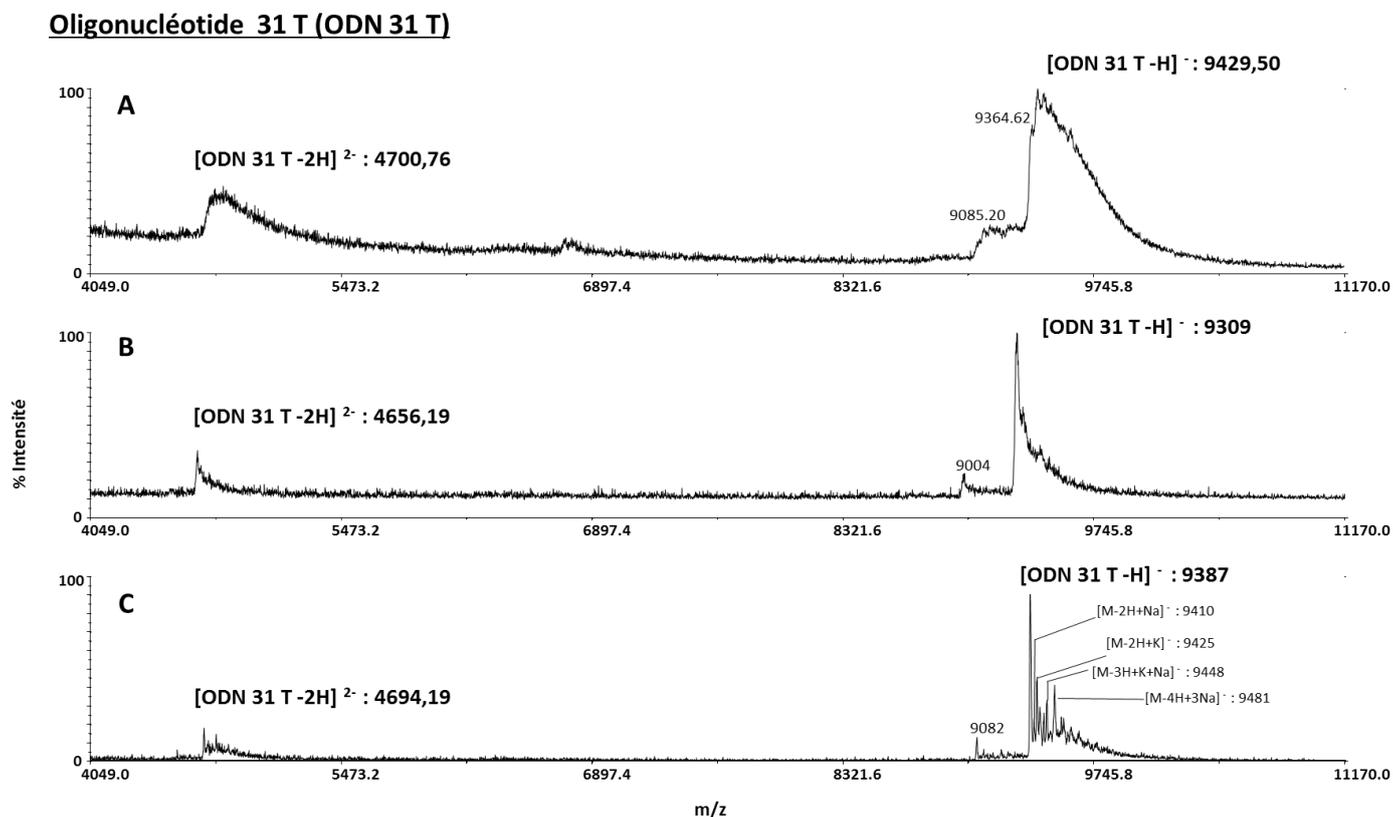


Figure 73 : Spectres MALDI-TOF en mode linéaire, ions négatifs de l'ODN 31 T de masse moléculaire théorique moyenne  $MM = 9368,02$  Da, étalonnage par défaut ; A) mélange 9:1 3HPA : l'ODN 31 T (15 à 20 pmol) avec citrate de diammonium ; B) mélange 9:1 3HPA : l'ODN 31 T (15 à 20 pmol) avec citrate de diammonium et amorçage de la cristallisation ; C) mélange 9:1 3HPA : l'ODN 31 T (15 à 20 pmol) avec citrate de diammonium, amorçage de la cristallisation dans les conditions d'analyse optimisées.

### 5.3.4 Optimisation des paramètres d'acquisition du spectromètre de masse MALDI-TOF (Voyager-DE Pro, Applied Biosystems)

L'appareil MALDI-TOF Voyager-DE™ Pro de la plateforme de spectrométrie de masse et protéomique de l'Institut de Biologie Paris Seine (Resp. L. Matheron, G. Bolbach) est la plupart du temps utilisé pour l'analyse de peptides ou de protéines en mode ions positifs réflecteur ou linéaire, avec des matrices adaptées à ces biomolécules, en particulier l'hydroxy- $\alpha$ -cyano-cinnamique acide (HCCA) ou l'acide 2,5-dihydroxybenzoïc (DHB). Les réglages étaient donc optimisés pour ces analyses et ont nécessité quelques modifications pour la détection des oligonucléotides en mode linéaire, ions négatifs avec la matrice 3-HPA. La résolution des signaux a pu être améliorée en optimisant en particulier le temps de retard Td et la valeur du coefficient G qui fait varier le champ E appliqué aux ions, dans la première partie de la source,

entre la cible et la grille. Ces paramètres Td et G dépendent de la masse m de l'ion, de la géométrie de l'appareil (source, région sans champ) et de la matrice utilisée ( $v_0$  axiale).

En mode linéaire, pour des ions de masse comprise entre 9 et 12 kDa, nous nous sommes placés dans les conditions optimales décrites pour cet appareil c'est-à-dire Td = 150 ns, G = 92 %. La valeur de tension appliquée sur le guide d'ions a été fixée à 0,05 % de la tension d'accélération et permet une transmission efficace des ions jusqu'au détecteur.

Des essais d'analyse ont aussi été réalisés sur un autre appareil le MALDI-TOF/TOF (4700, Proteomics, Applied Biosystems) mais n'ont pas permis l'obtention de signaux aussi résolus. Cet appareil est en effet plutôt dédié à l'analyse des peptides en mode MS et MS/MS et les étapes de réglages auraient nécessité de très grandes quantités de standard d'oligonucléotides que nous n'avions pas à disposition. Etant donné que les signaux optimisés sur l'appareil DE-Pro étaient satisfaisants, nous avons poursuivi avec celui-ci.

Les conditions de cristallisation et les paramètres d'acquisition optimisés ont permis d'améliorer la résolution comme observé sur le spectre C de la Figure 73. Sur ce spectre (15 à 20 pmol d'ODN 31 T déposés sur cible), la molécule déprotonée [ODN 31 T-H]<sup>-</sup> est observée à m/z 9386 avec ses adduits sodium (9410 m/z) et potassium (9425 m/z) bien résolus. L'ion doublement déprotoné [ODN 31 T-2H]<sup>2-</sup> est aussi observé à m/z 4694. L'ion à m/z 9082 qui correspond à une perte de 305 u. pourrait s'expliquer par la perte de la thymidine monophosphate en 3' ou 5' de l'ODN 31 T dans les conditions d'ionisation MALDI.

L'oligonucléotide photoactivable 4<sup>S</sup>T 30-mère utilisé dans notre étude (formule brute : C<sub>312</sub>H<sub>412</sub>N<sub>100</sub>O<sub>194</sub>P<sub>30</sub>S, MM = 9655,70) a été analysé dans ces mêmes conditions en déposant entre 15 et 20 pmol d'ADN sur la cible. On obtient sa masse moléculaire à partir de la masse de l'ion déprotoné [ODN 4<sup>S</sup>T-H]<sup>-</sup> à m/z 9660, avec une précision de 450 ppm, ce qui est attendu par rapport aux performances de l'appareil MALDI-TOF DE Pro utilisé en mode linéaire. L'optimisation de ces paramètres a permis d'augmenter la résolution d'un facteur 3 en passant de R = 330 à R = 1000 pour une masse de 10000 Da en mode linéaire, avec la matrice 3-HPA (Figure 74, A et B). Comme avec l'ODN 31 T on observe des ions qui pourraient correspondre à la perte d'une thymidine monophosphate en 3' de cet oligonucléotide (ion à m/z 9355), ainsi

qu'à la perte d'une des bases de type adénine correspondant à une perte de 135 u. (ion à  $m/z$  9526).

### Oligonucléotide 4<sup>S</sup>T

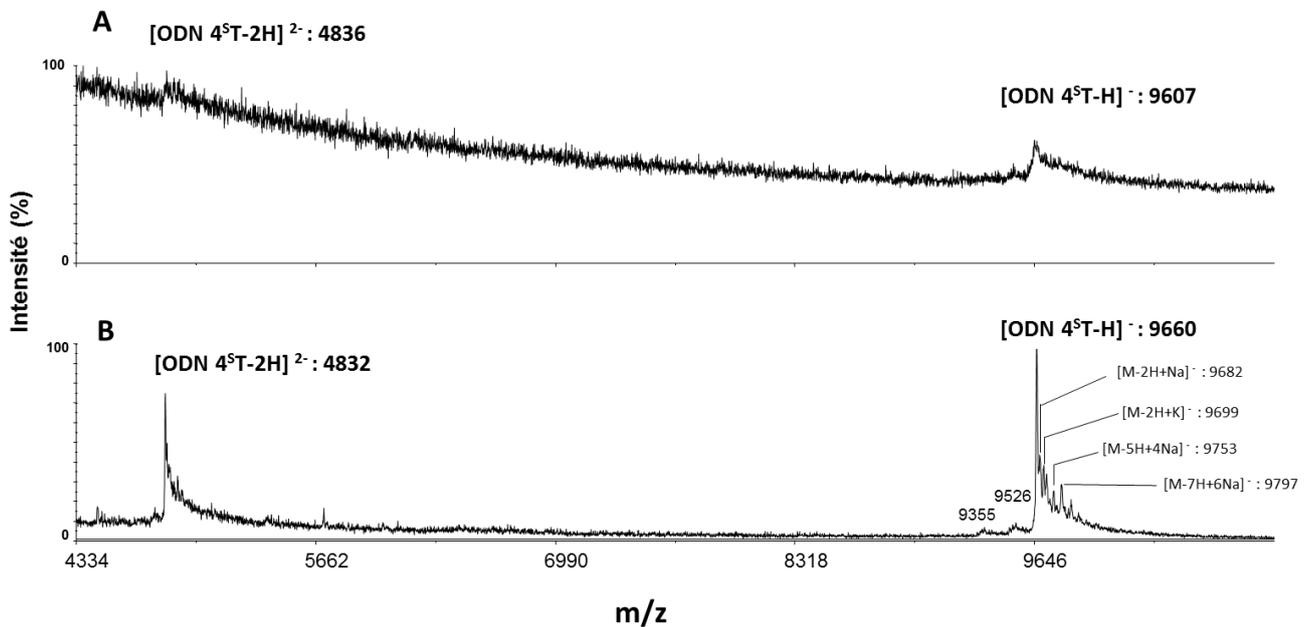


Figure 74 : Spectres MALDI-TOF en mode linéaire, ions négatifs de l'oligonucléotide 4<sup>S</sup>T de masse moléculaire théorique moyenne  $MM = 9655,70$  Da, étalonnage réalisé avec l'oligonucléotide ODN 31 T ; A) mélange 9:1 3HPA : ODN-4<sup>S</sup>T (15 à 20 pmol) avec citrate de diammonium ; B) mélange 9:1 3-HPA : ODN-4<sup>S</sup>T avec citrate de diammonium, amorçage de la cristallisation dans les conditions d'analyse optimisées.

Il est nécessaire ici de noter que la limite de sensibilité induite par la technique de dépôts est relativement élevée. Nous n'arrivons pas à obtenir un signal exploitable en MALDI-TOF en dessous de 10 pmol d'ADN sur la cible. Avec les acides nucléiques la sensibilité de détection atteinte est généralement plus faible qu'avec les protéines ou peptides. Nous avons dit précédemment que les acides nucléiques ou ADN étaient plus difficiles à étudier sous forme de molécules uniques du fait de leur polarité (en particulier les phosphates chargés négativement). Ces phosphates peuvent facilement former des liaisons intermoléculaires avec les bases ou avec les cations du milieu, conduisant à l'impossibilité de détecter les molécules uniques. Ceci n'est pas le cas avec les protéines ou peptides qui forment plutôt des liaisons hydrogène faibles entre eux et donc facilement rompues lors de l'analyse.

Pour résumer, nous avons mis au point des conditions permettant de détecter des ions d'oligonucléotides purs de masse moléculaire de l'ordre de 10 kDa de façon reproductible et robuste. La sensibilité de détection reste plus faible que pour les peptides ou protéines puisque les quantités nécessaires d'ADN pour obtenir ce signal sont relativement élevées (au moins 10 pmol d'ADN alors qu'un peptide est détectable à partir de 0,1 pmol). Ceci devra être gardé à

l'esprit dans les expériences d'irradiation sachant que les rendements de photomarquage sont compris entre 10 et 30 % et que le rendement d'ionisation de l'hétéroconjugué devrait être inférieur à celui de l'oligonucléotide  $4^S$ T pure. Dans le paragraphe suivant nous abordons la transposition de ces optimisations à l'analyse des hétéroconjugués oligonucléotide/peptides obtenus par photomarquage.

### **Analyse de complexes ADN-protéines**

5.4 Pour rappel, notre expérience de photomarquage se déroule de la manière suivante : une sonde oligonucléotidique simple ou double brin (30-mère, ~10 kDa) possédant un agent photoactivable est incubé en présence de protéines (rapport oligonucléotide : protéine de 1 : 10). Les protéines affines pour la sonde peuvent se lier à cet oligonucléotide et cette liaison est rendue covalente lors d'une phase d'irradiation UV.

#### *5.4.1 Identification des complexes ODN-protéines modèles*

Les complexes covalents ODN-protéines, que nous obtenons après irradiation, peuvent être assimilés à une modification de la protéine par l'ajout d'un oligonucléotide, ou l'inverse. Pour analyser ces complexes, deux voies ont été envisagées : 1) analyse du complexe entier ou 2) analyse du complexe après digestion trypsine de la protéine avec ou sans dégradation de l'oligonucléotide (enzymatique ou chimique) par spectrométrie de masse MALDI-TOF.

L'analyse MALDI-TOF (voie 1) (linéaire, ions positifs ou négatifs) du complexe entier ODN-hRPA n'a jamais permis de détecter de complexe, quelle que soit la matrice utilisée : HCCA, DHB ou 3-HPA dans nos conditions de cristallisation optimisées. Nous avons donc travaillé sur la voie 2 d'analyse dont le protocole implique un enrichissement des complexes oligonucléotide-protéine à l'aide de billes magnétiques greffées de streptavidine permettant d'éliminer la protéine libre en excès. Cette solution enrichie est séparée sur gel SDS PAGE dénaturant. Les bandes de protéine obtenues sont digérées à la trypsine *via* un protocole classique de protéomique. Ce digestat peut donc être analysé en MALDI en utilisant deux matrices différentes, HCCA pour l'analyse des peptides en mode réflecteur ions positifs et 3-HPA pour l'oligonucléotide en mode linéaire ions négatifs.

Lors de la préparation de ces échantillons, nous avons remarqué que les cristaux de matrice 3-HPA étaient plus difficiles à obtenir. Il est possible que la présence des peptides libres en excès,

empêche la cristallisation correcte de la matrice, réduisant ainsi les chances d'observer l'oligonucléotide.

Dans cette expérience, le contrôle a été réalisé en incubant hRPA avec la sonde ODN 4<sup>ST</sup> sans irradiation. Les deux échantillons (avec et sans irradiation) ont ensuite été traités exactement de la même façon, des bandes de gel ont été découpées dans les deux pistes, en parallèle dans la zone du gel correspondant à des hétéroconjugués potentiels avec la sous-unité hRPA1, c'est-à-dire dans la zone de MM supérieure à 70 kDa.

Les bandes de gel ont été digérées par la trypsine puis le digestat obtenu a été précipité dans l'éthanol à froid en présence de NaCl pour éliminer les peptides libres en excès et récupérer l'ODN dans le précipité. L'analyse a été réalisée par MALDI-TOF en mode linéaire ions négatifs avec la matrice 3-HPA et les spectres obtenus sont présentés sur la Figure 75.

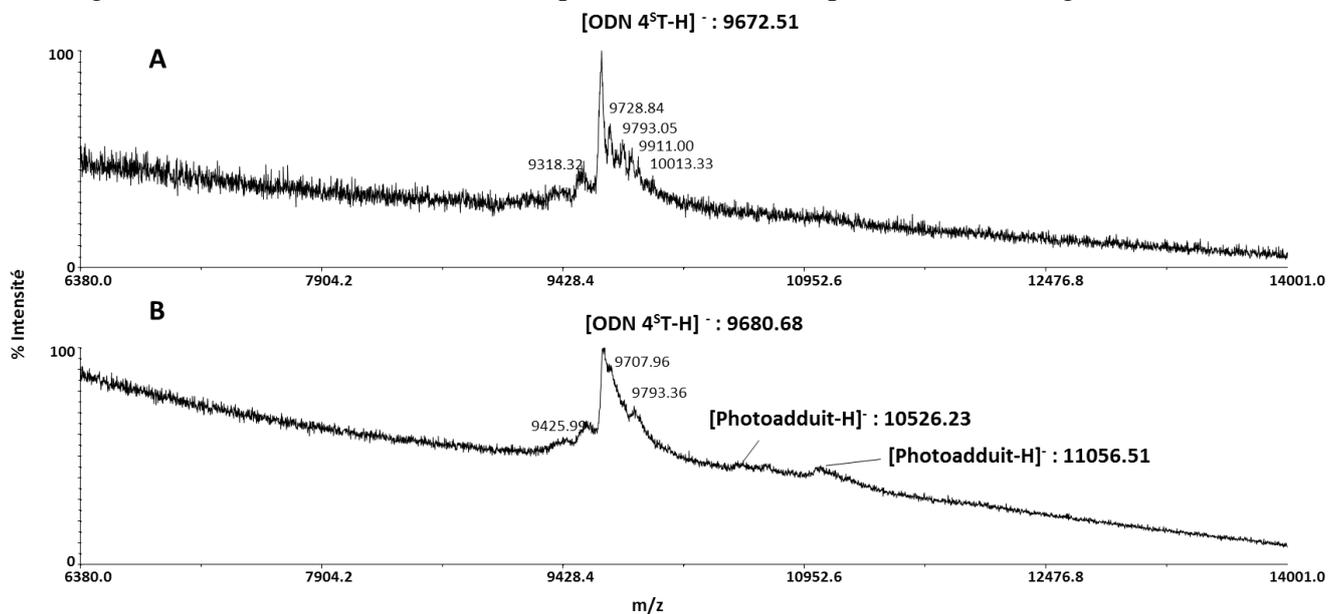


Figure 75 : Spectres MALDI-TOF en mode linéaire, ions négatifs, mélange 9:1 3-HPA : Echantillon (plus de 10 pmol de sonde ADN estimée) avec citrate de diammonium amorçage de la cristallisation dans les conditions d'analyse optimisées, étalonnage réalisé avec l'ODN 31 T. Analyse des complexes ODN 4<sup>ST</sup>-hRPA1, obtenus après digestion sur gel à la trypsine, zone du gel correspondant à des hétéroconjugués potentiels avec la sous-unité hRPA1, c'est-à-dire dans la zone de MM supérieure à 70 kDa sans (A) et avec irradiation (B).

L'ODN libre est observé à m/z 9672 dans le contrôle (A) et vers m/z 9680 dans l'expérience avec irradiation. Ces masses sont peu précises et encore moins dans le cas de l'échantillon irradié. L'élargissement des signaux est dû à différents paramètres 1) l'utilisation de chlorure de sodium dans le processus de précipitation de l'ODN qui apporte des cations métalliques dans le milieu et 2) l'irradiation UV qui détériore en partie l'ODN. Nous avons pu observer qu'en fonction du temps d'irradiation, les signaux de l'ODN libre s'élargissent conduisant à une perte de résolution et de précision en masse (données non présentées). Après irradiation, on observe l'apparition de deux pics larges de m/z supérieur à celui de l'ODN seul aux alentours de m/z

~11056 et  $m/z$  ~10526, absents du spectre contrôle. Ces deux pics sont donc spécifiques de l'irradiation et correspondent à des complexes ODN-peptides dans lesquels les peptides proviennent de la sous unité hRPA1. Bien que cette expérience soit encourageante, nous n'avons pas réussi à la reproduire.

Au vu de la difficulté que nous avons rencontré à isoler et à caractériser un complexe ADN-peptide provenant d'une digestion à la trypsine, nous avons décidé de travailler sur un complexe modèle, plus simple, constitué de la sonde ODN 4<sup>ST</sup> et d'un peptide que nous nommerons ARTK dans la suite du manuscrit, en lien avec les 4 premiers résidus de sa séquence. Ce type de complexes ADN-peptide a déjà été rapporté dans la littérature<sup>277,278,279</sup>.

#### 5.4.2 *Etude d'un complexe modèle : ADN-peptide (ADN-ARTK)*

Le peptide ARTK (Figure 76) synthétisé préalablement au laboratoire<sup>250</sup> a été choisi comme peptide modèle pour mimer l'interaction ADN-protéine. Ce peptide, issu d'une histone, protéine interagissant avec l'ADN, est très chargé positivement à pH physiologique ce qui devrait favoriser une interaction électrostatique avec l'ADN chargé négativement (phosphates).

##### 5.4.2.1 *Structure du peptide ARTK*

Ce peptide ARTK correspond aux 20 premiers résidus de la partie *N*-terminale de la protéine H3.1 humaine (Figure 76). Son extrémité *C*-terminale est amidée pour éviter sa dégradation enzymatique, un groupement biotine a été ajouté sur la chaîne latérale de la lysine 18. Cette étiquette biotine permet une éventuelle purification de ce peptide dans un mélange complexe. Le peptide ARTK (MM = 2294,28 Da) est chargé positivement à pH physiologique du fait de ses nombreux acides aminés basiques.

**Séquence de la Protéine H3 humaine (masse moléculaire moyenne : 15273,35 u.)**

**MARTKOTARKSTGGKAPRKQ**LATKAARKSAPSTGGVKKPHRYRPGTVALREIRRYQK  
STELLIRKLPFQRLVREIAQDFKTDLRFQSAALQEAAYLVGLFEDTNLCAIHAQR  
VTIMPKDIQLARRIGERA

**Peptide ARTK = Domaine N-terminal de la protéine H3 (masse moléculaire : 2294,28 u.)**

ARTKQTARKSTGGKAPRKQ-CONH<sub>2</sub>

(CH<sub>2</sub>)<sub>4</sub>

NH- Biotine

Figure 76 : Séquences de la protéine histone humaine H3 et du peptide ARTK. ARTK est biotinylé sur la fonction amine primaire de la chaîne latérale de la lysine en position 18.

**5.4.2.2 Test de la liaison ARTK-ADN : identification de la formation d'un complexe ARTK-ADN sur gel**

Nous avons dans un premier temps validé ce modèle en vérifiant qu'il existait bien une interaction entre l'ODN 4<sup>ST</sup> et le peptide ARTK. Pour ce faire, nous avons réalisé une expérience de photomarquage avec la sonde ODN 4<sup>ST</sup> radiomarquée au <sup>32</sup>P et 10 équivalents de peptide ARTK, pour favoriser cette interaction. Le gel SDS PAGE dénaturant résultant de cette expérience est présenté Figure 77.

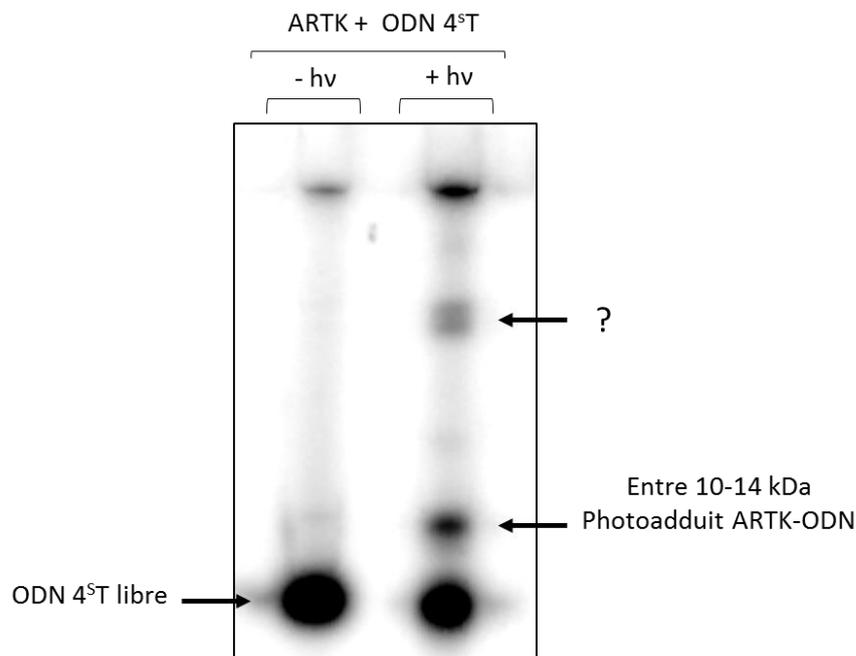


Figure 77 : Autoradiographie du gel SDS-PAGE résultant de l'expérience de photomarquage entre l'ODN-4<sup>ST</sup> radiomarqué et le peptide ARTK sans (piste de gauche) et avec irradiation (piste de droite).

Sur ce gel, la piste de gauche correspond à l'expérience contrôle réalisée sans irradiation, la piste de droite à l'expérience avec irradiation. Après irradiation, on observe de nouvelles bandes de plus haut poids moléculaire que l'ODN. La bande la plus intense se situe à une distance de l'ODN libre correspondant à une masse supplémentaire d'environ 2 kDa, et correspond logiquement à un complexe covalent ODN-ARTK, puisque la masse de ce complexe est attendue à 12 kDa. On estime le rendement de photomarquage à 16%. Les autres bandes moins intenses présentes sur cette piste n'ont pas pu être définies. Elles se situent entre 60 et 100 kDa. L'hypothèse la plus probable serait qu'elle soit le résultat d'une agrégation peptides/ODN fortement liés électrostatiquement.

Le peptide modèle ARTK peut donc se lier à l'ODN  $4^{\text{ST}}$ , et ce lien peut être rendu covalent après irradiation. Le complexe ODN  $4^{\text{ST}}$ -ARTK peut donc servir de modèle pour la mise au point de la détection d'hétéroconjugués ODN-peptide en spectrométrie de masse MALDI-TOF.

#### *5.4.3 Identification du complexe ODN-ARTK*

Le complexe est réalisé de la façon suivante : l'ODN  $4^{\text{ST}}$  est incubé avec 10 équivalents de peptide ARTK avant d'être irradié 60 min à 365 nm dans l'eau bi-distillée. Ces conditions ont été choisies après avoir testé différents ratio ODN : peptide (1 : 1, 1 : 10 et 1 : 50) et plusieurs temps d'irradiation (de 20 à 120 min) afin d'obtenir le meilleur rapport signal-sur-bruit pour l'hétéroconjugué. Nous avons donc choisi de travailler avec un excès de peptide de 10 équivalents et un temps d'irradiation de 60 min.

Une fois le photomarquage effectué, une étape de précipitation de l'ODN est réalisée pour enrichir le complexe ODN-ARTK en éliminant l'excès de peptide. L'échantillon est ensuite analysé par MALDI-TOF en mode linéaire ions négatifs dans les conditions de cristallisation optimisées avec la matrice 3-HPA. Les spectres correspondant au contrôle sans irradiation et à l'expérience sont présentés sur la Figure 78A et B, respectivement.

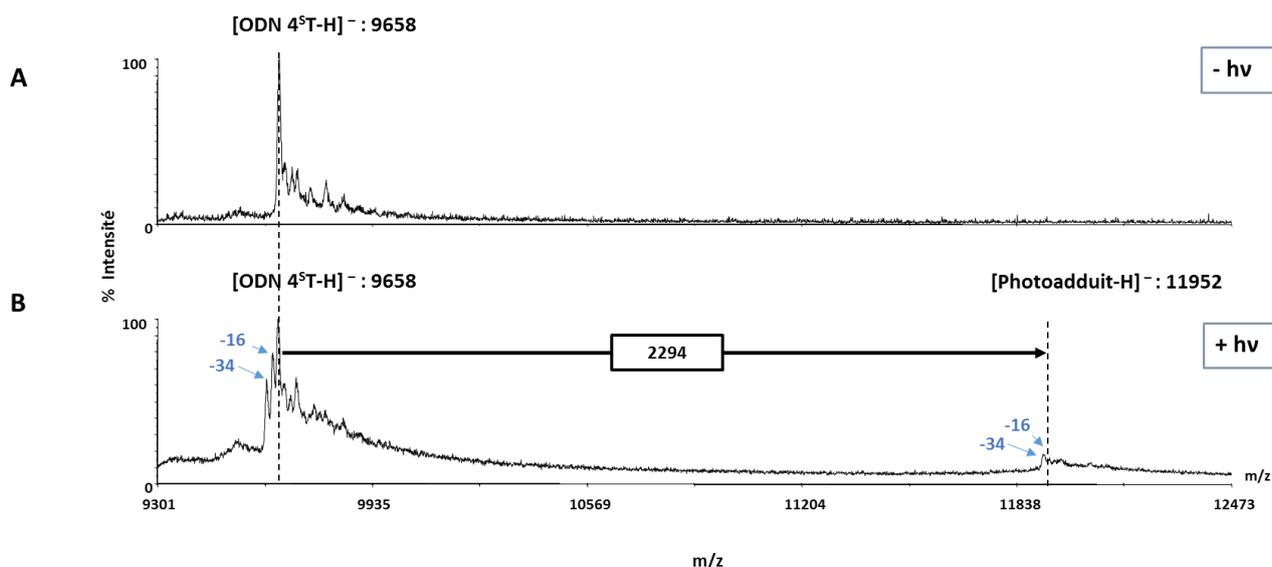


Figure 78 : Spectres MALDI-TOF en mode linéaire ions négatif du mélange ADN  $4^{\text{ST}}$ , peptide ARTK dans un ratio 1 pour 10 sans (A) ou avec irradiation (B). La matrice utilisée est 3-HPA dans nos conditions optimisées. Chaque spectre correspond à l'analyse d'au minimum 10 pmol d'ADN totale (lié et non lié).

Plusieurs phénomènes peuvent être observés après irradiation de notre échantillon. Le plus intéressant est l'apparition d'un pic à  $m/z$  11952 qui correspond au photoadduit ODN  $4^{\text{ST}}$ -ARTK. Ce pic est assez large et peu intense. D'après les conditions d'analyse qui ne permettent de détecter l'hétéroconjugué qu'en mode ions négatifs avec la matrice 3-HPA, on peut faire l'hypothèse que l'ionisation de l'hétéroconjugué se fait sur la partie oligonucléotide. Il est alors possible de comparer l'aire des signaux obtenus pour l'oligonucléotide  $4^{\text{ST}}$  libre et pour l'hétéroconjugué. On remarque alors que l'aire de ce pic représente environ 10 % de l'aire de l'ODN  $4^{\text{ST}}$  totale (c'est-à-dire : aire ADN  $4^{\text{ST}}$  libre + aire photoadduit), ce qui est comparable aux rendements de photomarquage précédemment observés pour cette sonde. Par ailleurs, la réaction de photomarquage semble se faire sans perte de masse puisque la différence de masse observée entre l'ODN libre et le complexe correspond à la masse du peptide ARTK aux imprécisions près de la mesure des masses en mode linéaire (largeur du pic de l'hétéroconjugué). Enfin, nous observons aussi deux pertes de masse : une de 16 u. et une autre de 34 u. à la fois sur l'ODN seul et complexé. Nous reviendrons sur ces pertes de masses et sur la photochimie de la sonde  $4^{\text{ST}}$  dans le paragraphe 5.5 et sur la Figure 81.

Nous avons réalisé en parallèle la même expérience avec l'ODN-Dia possédant une diazirine comme sonde photoactivable.

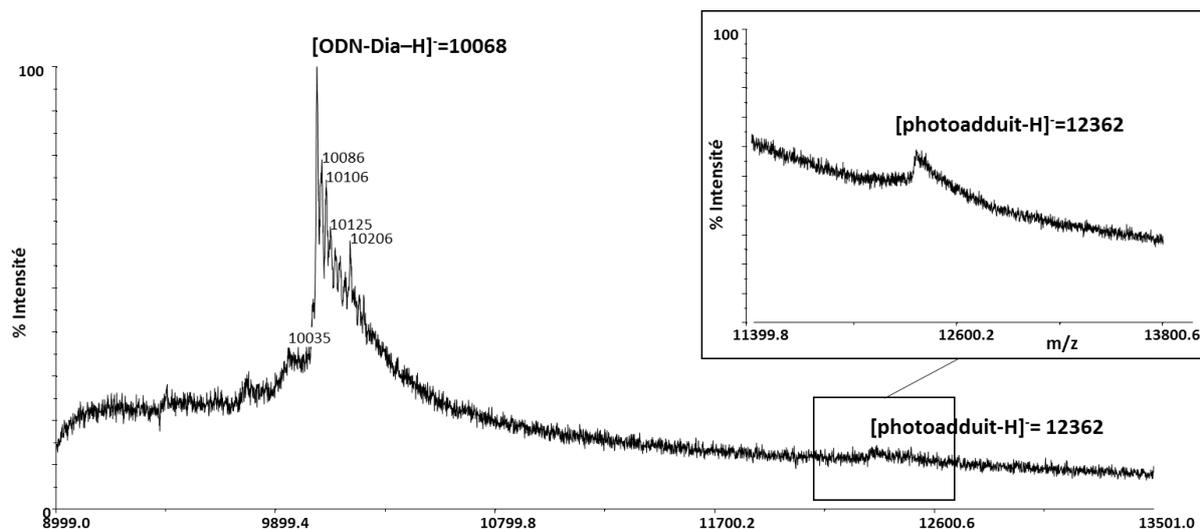


Figure 79 : Spectre MALDI-TOF en mode linéaire ions négatif du mélange ODN-Dia, peptide ARTK dans un ratio 1 pour 10, irradié. La matrice utilisée est 3-HPA dans nos conditions optimisées. Le spectre correspond à l'analyse d'au minimum 10 pmol d'ADN totale (lié et non lié).

Le spectre MALDI-TOF mode linéaire, ions négatifs correspondant (Figure 79) présente deux pics : l'ODN-Dia à m/z 10068 et le complexe covalent ODN-Dia-ARTK à m/z 12362. Cette différence de masse correspond là encore à la masse du peptide ARTK aux imprécisions près, du fait de la largeur du pic correspondant à l'hétéroconjugué. Le signal du complexe est encore moins intense et moins défini que dans le cas du complexes ODN 4<sup>S</sup>T-ARTK. Cela peut traduire un rendement de photomarquage plus faible qui pourrait s'expliquer par le fait que l'ODN-4<sup>S</sup>T, analogue de base, mime exactement l'encombrement de la base T qu'elle remplaceFigure 43. Dans le cas de l'ODN-Dia, la sonde diazirine, placée au bout d'un bras espaceur en position 2 d'une guanine est plus encombrante et pourrait déstabiliser le complexe. La photochimie de la diazirine vis-à-vis des peptides/protéines étant déjà bien documentée dans la littérature<sup>166</sup>, il nous a semblé intéressant d'étudier le système ODN 4<sup>S</sup>T-ARTK afin de caractériser la

5.5 photochimie de cette sonde envers les protéines.

## Caractérisation de la liaison ODN 4<sup>S</sup>T-ARTK

### 5.5.1 Le modèle ODN 4<sup>S</sup>T-ARTK

Pour caractériser le site de liaison du complexe ODN 4<sup>S</sup>T-ARTK, nous avons choisi de dégrader la composante ODN du complexe pour faciliter l'analyse par spectrométrie de masse. Réduire la masse du complexe est en effet intéressante pour deux raisons : 1) on peut faire une analyse

en mode réflecteur qui permet d'avoir une résolution isotopique et une meilleure précision dans la détermination des masses et 2) on s'affranchit de la partie ODN et on peut donc utiliser une matrice MALDI adaptée aux peptides qui est plus facile à utiliser pour la préparation d'échantillon.

La dégradation de l'ADN des complexes modèles a été réalisée selon deux protocoles différents : l'un chimique, l'autre biochimique. La voie biochimique est effectuée à l'aide de la benzonase : endonucléase qui hydrolyse l'ADN. Cette dégradation peut conduire à de petits fragments de 3 à 5 bases voire une dégradation totale. Pour ce qui est de la méthode chimique, il s'agit d'une réaction drastique d'hydrolyse acide par de l'acide fluorhydrique (HF)<sup>280</sup>. Cette approche conduit à la dégradation totale de l'oligonucléotide sans dégrader les peptides.

Après dégradation de l'ODN des complexes ODN 4<sup>ST</sup>-ARTK, nous espérons donc obtenir le peptide ARTK modifié par un nucléotide de l'ODN ou par un petit fragment de celui-ci. La localisation de cette modification permettrait de déterminer le site de liaison covalente entre nos deux partenaires d'interaction. Lors de la réalisation de ces expériences, seul le peptide ARTK a été observé sous sa forme classique et oxydée à +16 u. par MALDI-TOF mode réflecteur, ions positifs avec la matrice HCCA, de façon reproductible (n > 3) et quel que soit le mode de dégradation de l'ADN (Figure 80B).

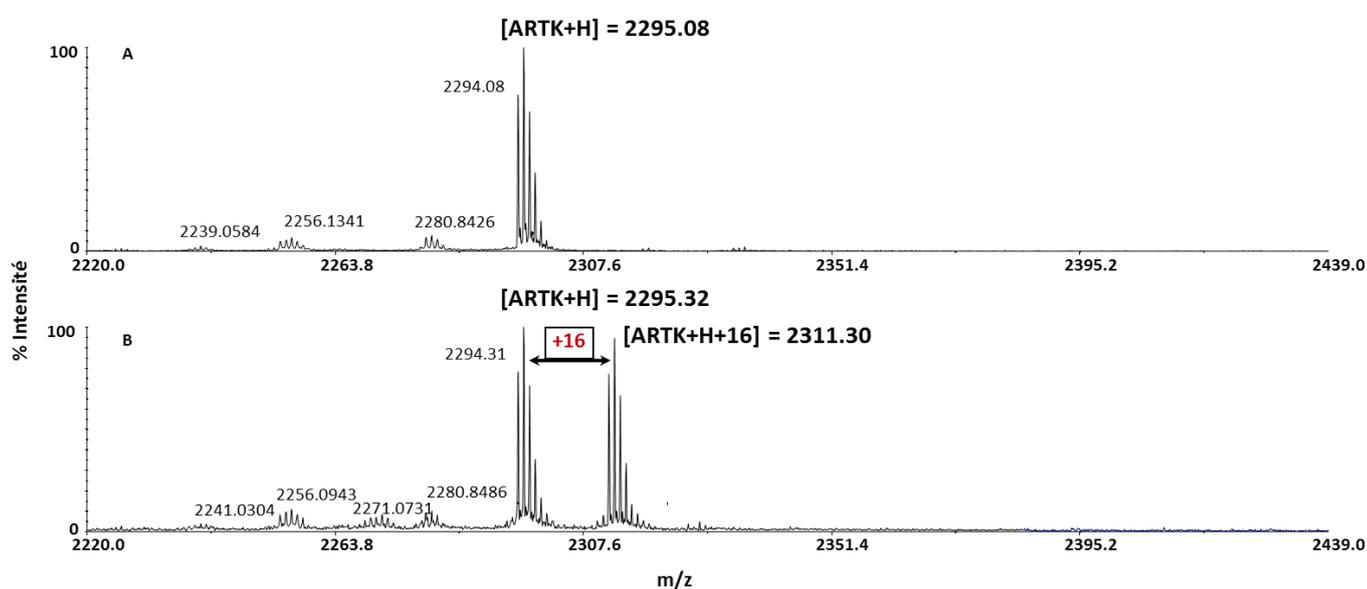


Figure 80 : Spectres MALDI-TOF en mode réflecteur, ions positifs, matrice HCCA (ratio 1 : 1, matrice : échantillon) des digestes à la benzonase des mélanges ODN 4<sup>ST</sup>/peptide ARTK sans (A) ou avec irradiation (B).

Pour savoir si cette modification est spécifique de la liaison avec l'ODN, nous avons réalisé un deuxième contrôle en irradiant le peptide ARTK en absence d'ODN (spectre non présenté). Dans ce cas, le peptide n'est pas modifié, on n'observe pas le deuxième pic à +16 u.

L'irradiation UV dans l'eau bi-distillée en l'absence de l'agent photoactivable ne conduit pas à l'oxydation du peptide. Cette modification (+16 u.) semble donc liée à la présence à la fois de l'ADN, du peptide et de l'irradiation. Cependant, ce contrôle ne nous permet pas de dire si la modification du peptide est due à l'interaction avec l'ODN ou si elle est seulement due à la présence de la photosonde activée par les UV dans le milieu réactionnel.

La caractérisation du site de cette modification a été réalisée par fragmentation CID haute énergie en MALDI-TOF/TOF en sélectionnant le peptide oxydé comme ion parent ( $m/z$  2311,3). Cette modification concerne donc la proline 16 ou l'arginine 17, du peptide ARTK. Nous n'avons pas pu aller plus loin dans la précision de la détermination du site de modification. Il pourrait s'agir d'une arginine ou proline oxydée. Une hypothèse envisageable pour expliquer que le peptide ne contient pas une modification de masse plus importante serait que les mécanismes induits par la dégradation de l'ODN détériorent aussi la liaison établie entre la 4<sup>ST</sup> et le peptide ARTK, et ce, aussi bien par voie chimique qu'enzymatique. Cette hypothèse est difficilement confirmable puisque la nature de la liaison n'est pas connue. En effet, le mécanisme ou les mécanismes par lequel la 4<sup>ST</sup> établit des liaisons covalentes avec des protéines n'est pas documenté dans la littérature, bien qu'il soit connu pour des liaisons avec les acides nucléiques<sup>197</sup>. On peut aussi émettre l'hypothèse que l'oxydation du peptide ARTK soit un effet indirect de l'irradiation de l'ODN-4<sup>ST</sup>, qui une fois excité pourrait accélérer la formation d'hydroxyles radicalaires à partir de molécules d'eau. Ces radicaux viendraient alors réagir sur le peptide le rendant ainsi oxydé.

Pour résumer, cette modification conduisant à un incrément de masse de 16 u. par rapport à la masse initiale du peptide ARTK n'est pas suffisamment spécifique pour émettre une hypothèse de réactivité de la sonde 4<sup>ST</sup>.

Dans le paragraphe suivant nous revenons sur l'analyse des spectres MALDI-TOF des complexes entiers ODN 4<sup>ST</sup>-ARTK en nous intéressant aux pertes de masse de 16 et 34 u. observées précédemment en MALDI-TOF mode linéaire ions négatifs (Figure 78) et présentées plus en détail Figure 81.

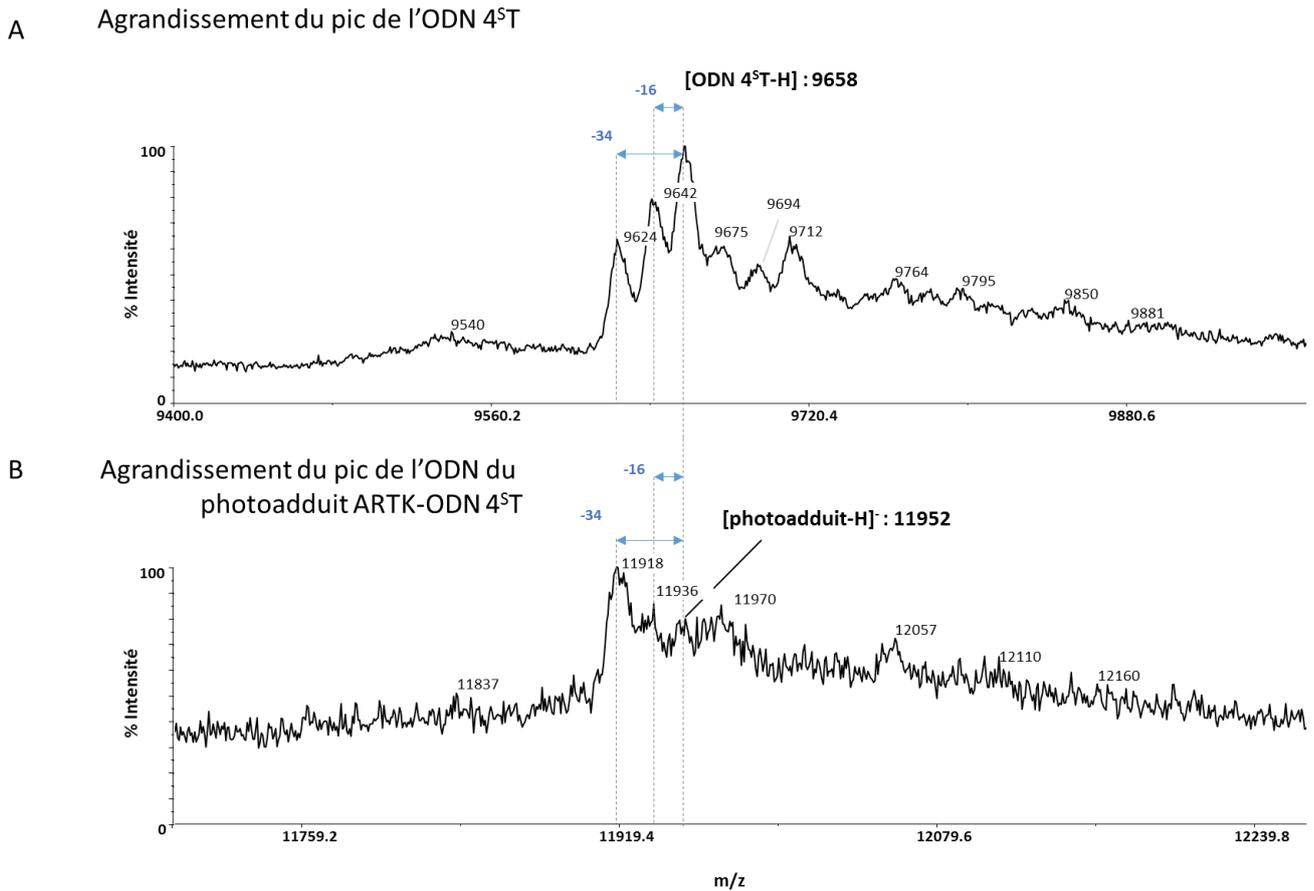


Figure 81 : Agrandissement du signal correspondant à l'ODN 4<sup>S</sup>T libre (A) et du signal correspondant au photoadduit ODN 4<sup>S</sup>T-ARTK (B) du spectre MALDI-TOF en mode linéaire ions négatif du mélange ODN 4<sup>S</sup>T, peptide ARTK dans un ratio 1 pour 10, irradié (Figure 78B). Les traits en pointillé permettent de faire la corrélation entre les deux signaux.

Les spectres présentés Figure 78B et Figure 81 nous donnent deux informations importantes :

- 1) la liaison entre l'ODN 4<sup>S</sup>T et le peptide ARTK conduit à un complexe dont la masse est la somme de la masse d'ARTK et de la masse de l'ODN 4<sup>S</sup>T, 2) l'ODN 4<sup>S</sup>T, après irradiation est altéré et peut perdre 16 et/ou 34 u. aussi bien à partir de l'ODN libre que du complexe ODN 4<sup>S</sup>T -ARTK même si le signal est de plus faible qualité. Pour savoir si la base 4<sup>S</sup>T est impliquée dans ces pertes de masse nous avons reproduit l'expérience de photomarquage entre l'ODN 4<sup>S</sup>T et le peptide ARTK en réalisant une expérience contrôle en parallèle. Dans ce contrôle l'ODN 4<sup>S</sup>T (MM = 9655,7) est remplacé par la même séquence d'ADN contenant une thymine classique (nommé ODN T, MM = 9639,7) à la place de la 4<sup>S</sup>T. Le résultat de cette expérience est présenté Figure 82, où seuls les pics représentant les ADN libres sont indiqués pour plus de clarté. Ces spectres ont été obtenus de manière reproductible et les signaux obtenus pour les trois expériences sont comparables en termes d'intensité.

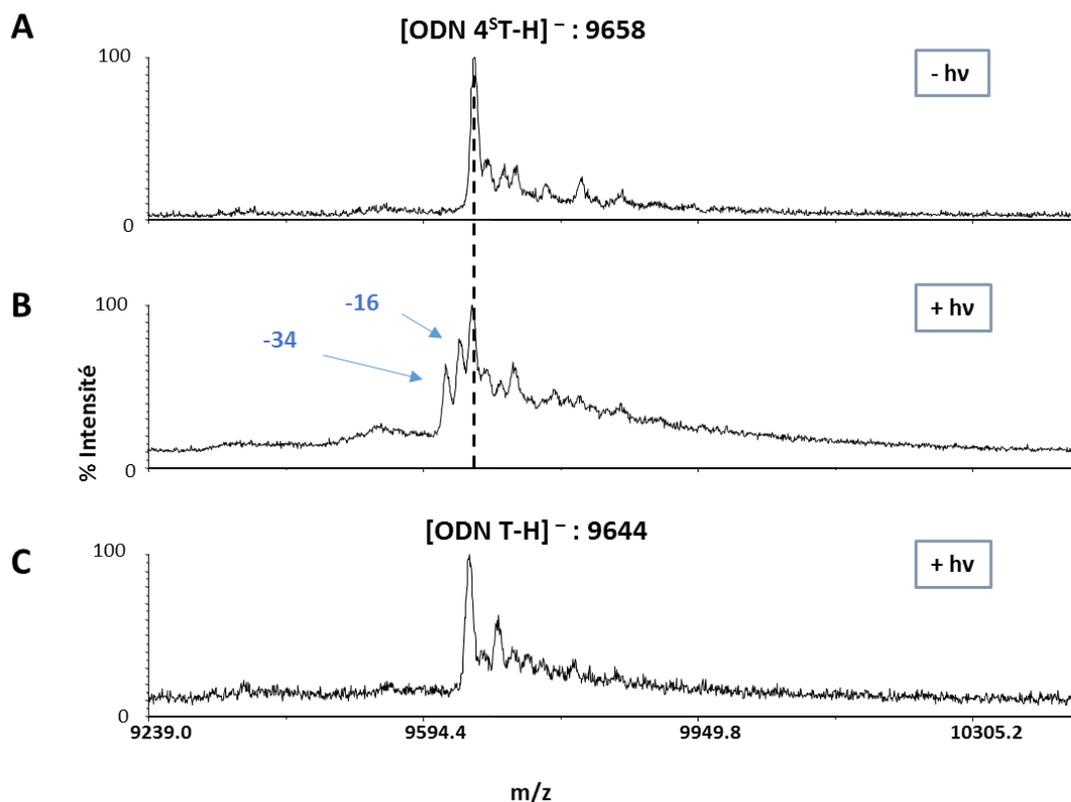


Figure 82 : Spectres MALDI-TOF mode linéaire ions négatifs avec la matrice 3-HPA dans nos conditions optimisées, représentant différentes expériences de photomarquage : A) Mélange ODN 4<sup>S</sup>T/peptide ARTK sans irradiation ; B) Mélange ODN 4<sup>S</sup>T/peptide ARTK avec irradiation ; C) Mélange ODN T/peptide ARTK avec irradiation

L'expérience de photomarquage sans irradiation (Figure 82A) présente un spectre d'oligonucléotide classique où l'on observe l'ADN 4<sup>S</sup>T libre (m/z 9658) avec ses adduits cationiques. Après irradiation (Figure 82B), le signal de cet oligonucléotide est altéré. On observe un élargissement du pic et l'apparition des deux pics correspondant à des pertes de masse de 16 et 34 u. Ces pics ne s'observent pas lors de l'irradiation de notre expérience contrôle avec l'oligonucléotide dans sa version non-thiolée (pic à m/z 9644) (Figure 82C). En effet, on constate un spectre quasiment non altéré, proche du spectre A). On peut donc en déduire que les pertes de masse de 16 et 34 u. sont caractéristiques de la base 4<sup>S</sup>T après irradiation UV. La perte de 16 u. peut s'expliquer par le remplacement du soufre en position 4 par l'oxygène. En effet, lorsque la base 4<sup>S</sup>T est soumise à un rayonnement UV, elle peut réagir avec les molécules de dioxygène présentes en excès en solution dans les tampons ou milieux biologiques. Au cours de cette réaction qui s'apparente à la photochimie des thiocétone, l'état excité triplet T1 va réagir avec une molécule de dioxygène sous forme de singulet <sup>1</sup>O<sub>2</sub>. Cette réaction permet la formation d'un intermédiaire de type peroxy-thio-carbonyle qui va, par une réaction intramoléculaire, conduire à la formation d'une cétone en libérant du monoxyde de soufre<sup>281</sup>, soit la perte du soufre (32 u.) suivie de l'ajout d'un atome d'oxygène (16 u.), pour

une perte totale de 16 u. Un mécanisme analogue est ainsi proposé dans le cas de la 4<sup>ST</sup> sur la Figure 83.

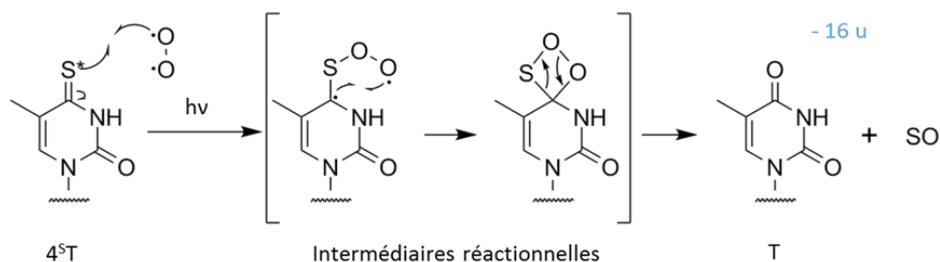


Figure 83 : Mécanisme analogue à celui décrit par Sanchez-Arroyo *et al.* pour la réaction d'un singulet <sup>1</sup>O<sub>2</sub> sur un fonction thiocétone appliqué à la 4-thiothymine.

L'obtention de la cétone peut aussi s'effectuer par réaction du triplet avec une molécule d'eau au cours d'une réaction d'hydrolyse de la thiocétone<sup>282</sup> qui conduit à l'ajout d'une molécule d'eau (18 u.) et à la perte d'un groupement SH<sub>2</sub> (34 u.), soit une perte totale de 16 u. (Figure 84).

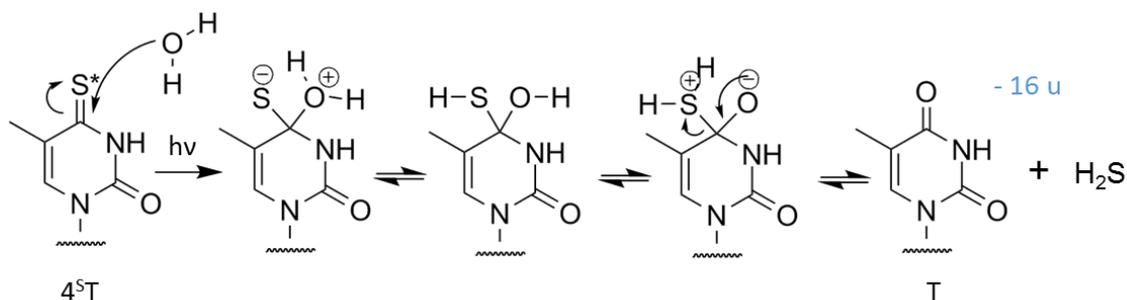


Figure 84 : Réaction d'une molécule d'H<sub>2</sub>O sur la 4-thiothymine excitée.

Ces deux réactions conduisent à la formation d'une thymine (T), rendant la sonde inactive.

La perte de 34 u. est moins évidente à expliquer. Dans une étude de fragmentation des ions protonés de l'uracile et ses dérivés, par dissociation induite par collision (CID), Nelson *et al.* ont montré que l'uracile pouvait perdre entre autre, H<sub>2</sub>O à partir de l'oxygène en position 2 ou 4 de façon équiprobable. De la même façon, ils ont observé que les dérivés thiolés de cette base étaient capables de perdre H<sub>2</sub>S<sup>283</sup>. Dans cet article, il n'est pas mentionné de mécanisme réactionnel permettant d'expliquer cette perte. Par ailleurs, en examinant le spectre ESI de la banque de données MassBank (LC-ESI-Q, mode négatif, MS 30V) de la 2-mercaptopurine (WA002771), une autre base thiolée utilisée en photomarquage, on observe un fragment majoritaire correspondant à une perte de 34 u. qui pourrait être attribuée à l'élimination de H<sub>2</sub>S. On pourrait supposer que dans notre cas, l'énergie apportée par l'irradiation UV subie par l'échantillon en amont de l'analyse MALDI peut fragiliser les liaisons et conduire, en mode MS simple à une fragmentation source observée initialement en mode tandem c'est-à-dire en

donnant une énergie supplémentaire aux ions. L'hypothèse qu'une énergie laser supérieure au seuil de production des ions en MALDI ( $\lambda$  du laser = 337 nm) soit la cause des pertes observées ne peut pas être retenue puisque dans le cas du spectre A) dans lequel il n'y a pas eu d'irradiation UV de l'échantillon au préalable, les pertes de masse de 16 et 34 u. ne sont pas observées en MALDI.

La perte de 34 u. peut aussi s'expliquer par la perte d'un groupement  $\text{H}_2\text{S}$  lors d'une réaction intramoléculaire. Ceci a été proposé dans plusieurs mécanismes photochimiques de la  $4^{\text{S}}\text{T}$ <sup>197</sup>, où l'on a fait réagir des dinucléotides de type  $4^{\text{S}}\text{T-T}$  ou  $\text{T-}4^{\text{S}}\text{T}$ . Ces mécanismes ne sont pas transposables à notre oligonucléotide contenant de la  $4^{\text{S}}\text{T}$  puisque ce nucléotide modifié est encadré par une guanine en 5' et une adénine en 3' (Figure 85).

5' - TTG CAC TCT CTT GAC G-4<sup>S</sup>T-A CTG TCC AGC TTC-Biotine -3'

Figure 85 : Séquence de l'oligonucléotide contenant un 4-thiothymine, dont ces proches voisins sont soulignés.

Saintomé *et al.* ont montré qu'un analogue de 4-thiothymidine, ici la  $N^3$ -méthyl-4-thiothymidine, réagit avec une adénosine pour former une adénosine  $N^6$ -formamidopyrimidine après une phase d'irradiation<sup>284</sup> comme on peut le voir sur la Figure 86.

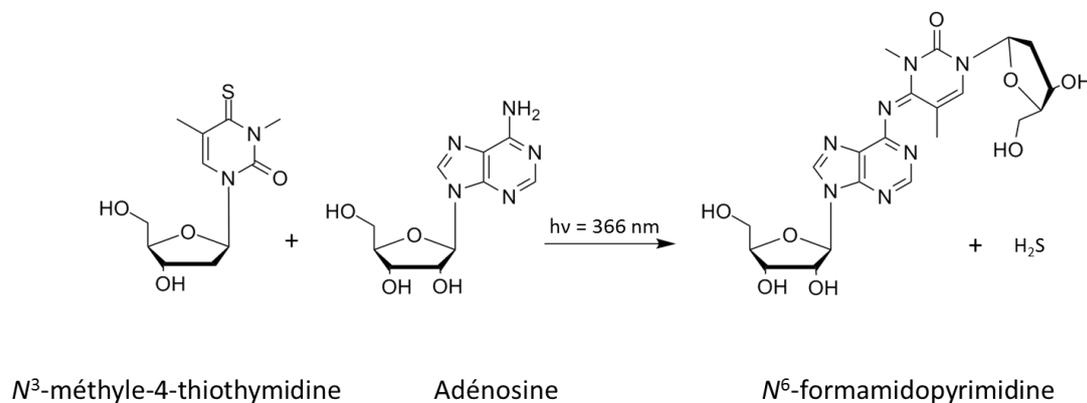


Figure 86 : Réaction entre la  $N^3$ -méthyle-4-thiothymidine et une adénosine sous irradiation qui entraîne la perte d'un groupement  $\text{H}_2\text{S}$  pour former une adénosine  $N^6$ -formamidopyrimidine. D'après Saintomé *et al.*

Dans cette réaction, la  $N^3$ -méthyle-4-thiothymidine est attaquée par la fonction amine primaire de la base adénine entraînant la perte d'une molécule  $\text{H}_2\text{S}$  pour former une imine. La description de cette réaction, nous permet de proposer un mécanisme réactionnel expliquant la perte de 34u. avec la 4-thiothymine (Figure 87).

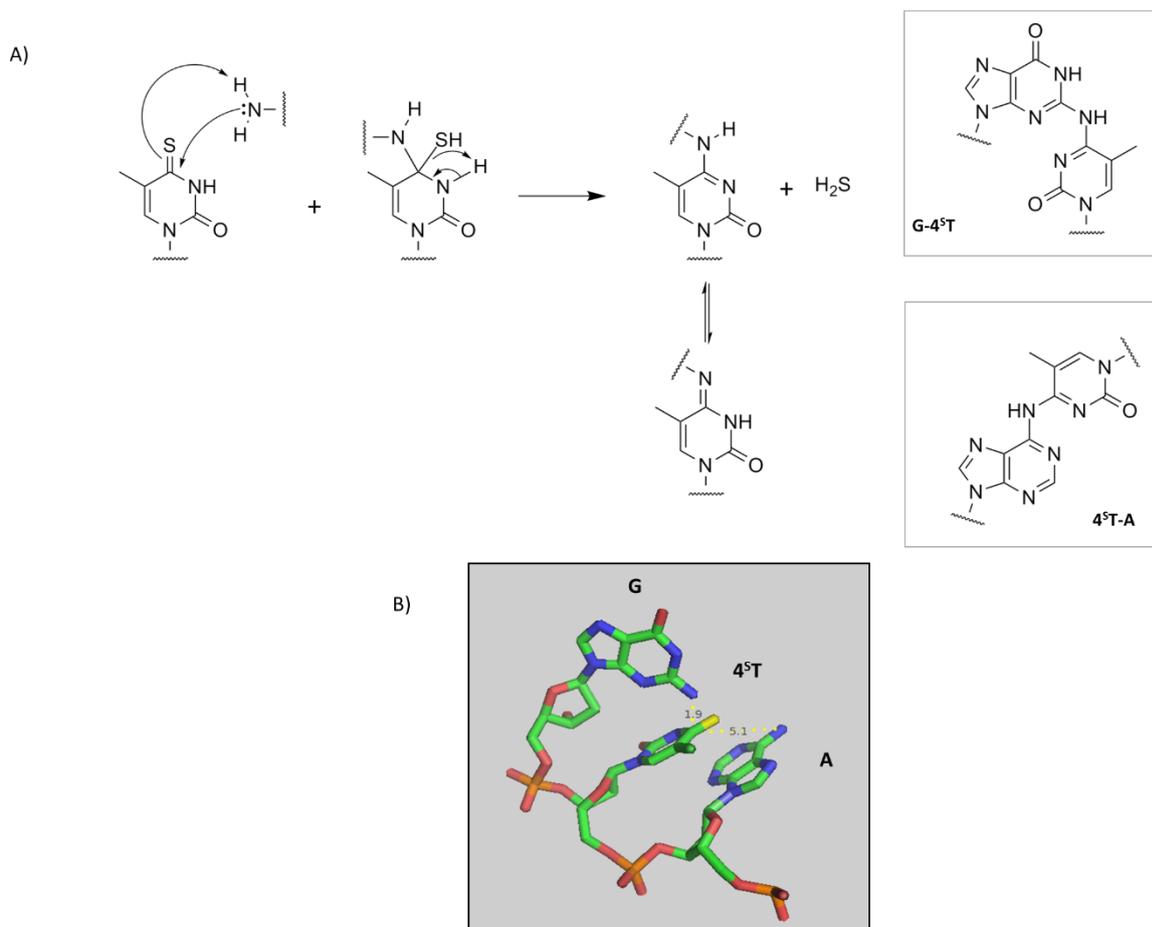


Figure 87 : A) Proposition de mécanisme entre la 4-thiothymine et une guanine ( $G-4^S T$ ) ou une adénine  $4^S T-A$  conduisant à une perte de 34u. B) Modèle en 3D du tri-nucléotide  $G-4^S T-A$  où sont représentées les distances entre l'amine de la guanine et la thiocétone de la  $4^S T$  et l'amine de l'adénine et la thiocétone de la  $4^S T$

L'irradiation permet l'excitation de la thiocétone de la  $4^S T$  qui est alors attaqué par une amine. Après des étapes de prototropie, une étape d'élimination permet de libérer une molécule de  $H_2S$  formant ainsi un dinucléotide lié covalent par les bases entre une guanine et la  $4^S T$  ( $G-4^S T$ ) ou entre la  $4^S T$  et une adénine ( $4^S T-A$ ). La nature du lien formé de type « imine » crée une contrainte qui éloigne les deux sucres des nucléosides l'un de l'autre dans le cas de  $4^S T-A$ . Cet éloignement rend peu probable la formation de ce lien au sein notre séquence. De plus, la visualisation en 3 dimensions du tri-nucléotide  $G-4^S T-A$  à l'aide du logiciel PyMol, nous a permis d'observer, par rotation des liaisons simples, que l'amine de la guanine est plus proche de la thiocétone de la  $4^S T$  (1,9 Å) que celle de l'adénine (5,1 Å) comme on peut le voir sur la Figure 87B. Il donc plus probable que la perte de 34 u. soit initiée par une réaction intramoléculaire entre la guanine et la  $4^S T$  voisine.

L'observation de ces deux pertes de 16 et 34 u. à partir de l'ion correspondant au complexe ODN 4<sup>S</sup>T-ARTK sur le spectre de masse B, Figure 78, Figure 81 et Figure 88C, indique que l'atome de soufre n'est pas directement impliqué dans la liaison covalente avec le peptide. Cependant, la présence du soufre est nécessaire à la réaction photochimique puisqu'aucun complexe n'a jamais été constaté par irradiation de l'ODN contenant une thymine à la place de le 4-thiothymine (Figure 82C).

### 5.5.2 Les modèles ODN 4<sup>S</sup>T-acides aminés

Afin d'aller plus loin dans l'élucidation du mécanisme réactionnel de la base 4<sup>S</sup>T, nous avons choisi de simplifier le système d'interaction en le limitant à l'interaction de l'ODN 4<sup>S</sup>T avec des acides aminés. Il s'agit de la même expérience de photomarquage réalisée avec différents acides aminés caractéristiques : la lysine (K) qui est basique, le tryptophane (W) qui est aromatique et hydrophobe et la leucine (L) hydrophobe mais sans fonction réactive. Les acides aminés utilisés sont protégés en *N*-terminal pour ne laisser que les chaînes latérales accessibles, comme s'ils étaient à l'intérieur d'un peptide ou d'une protéine. La synthèse peptidique étant un des fers de lance du laboratoire, nous avons choisi des acides aminés protégés sur leur amine en position  $\alpha$  par un groupement protecteur Fmoc (fluorénylméthoxycarbonyl), Fmoc-Lys-OH, Fmoc-Trp-OH et Fmoc-Leu-OH, puisqu'ils étaient disponibles au laboratoire. Cette protection augmente la masse du photoadduit produit (comparativement à la masse obtenue avec l'acide aminé libre) permettant qu'il ne soit pas superposé au pic de l'ion correspondant à l'ODN libre. Elle a par contre l'inconvénient majeur d'augmenter l'hydrophobie des acides aminés, ce qui diminue leur seuil de solubilité. Les problèmes de solubilité associés au faible rendement de photomarquage nous ont conduits à utiliser un ratio ODN 4<sup>S</sup>T : acide aminé de 1 : 1000 dans nos expériences.

Seuls les spectres correspondant à l'expérience avec la Fmoc-Lys-OH sont présentés (Figure 88).

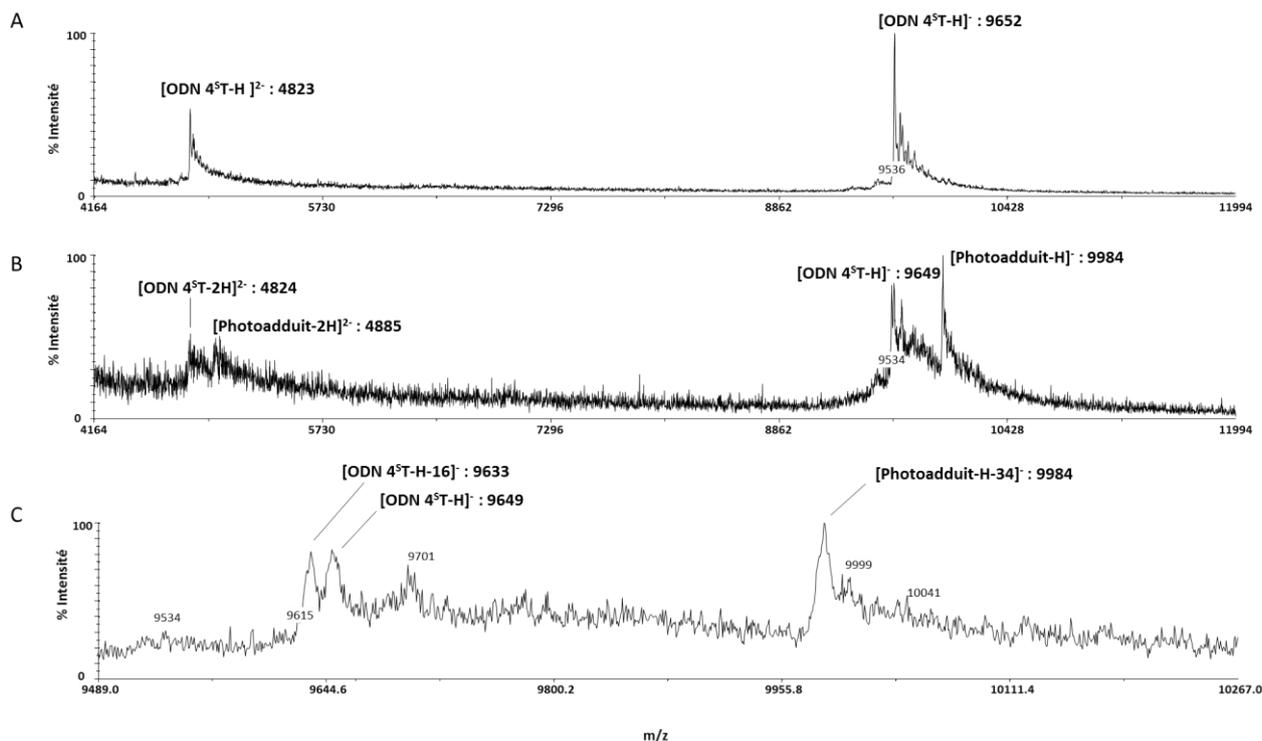


Figure 88 : Spectres MALDI-TOF mode linéaire ions négatifs avec la matrice 3-HPA dans nos conditions optimisées, représentant différentes expériences de photomarquage entre l'ADN-4<sup>ST</sup> et de la Fmoc-Lys-OH ( MM exacte : 366 u.). A) Expérience contrôle sans irradiation ; B) Expérience avec irradiation ; C) Agrandissement de la zone 9489 à 10267 m/z du spectre B correspond aux signaux de l'ODN-4<sup>ST</sup> et du photoadduit ODN 4<sup>ST</sup>-Fmoc-Lys-OH

Dans cette expérience, nous n'observons pas de pic correspondant à la masse théorique du photoadduit (10018 m/z). En effet, seul un pic dont la masse correspond à la masse du photoadduit ayant perdu 34u. est observable (Figure 88C). L'absence d'un pic correspondant à la masse du photoadduit, ainsi que l'absence d'un pic correspondant à la masse du photoadduit retranchée de 16u. suggère un mécanisme photochimique différent de celui que l'on peut observer avec le peptide ARTK. On peut supposer un mécanisme de type addition 1,2 sur la thiocétone qui s'accompagne d'une perte d'une molécule de H<sub>2</sub>S (34u.) comme nous avons pu le décrire dans l'introduction, paragraphe 3.2.3.2.1, p.63 ainsi que sur la Figure 87A.

Aucun complexe n'a été détecté avec Fmoc-Leu-OH et Fmoc-Trp-OH. Nous avons aussi essayé avec une histidine non protégée et un complexe a pu être observé (non présenté). Il semble donc qu'une charge positive de l'acide aminé soit nécessaire pour l'interaction avec l'ODN 4<sup>ST</sup>. La sonde 4<sup>ST</sup> semble sélective pour les acides aminés basiques. Il faudrait étendre cette étude à des exemples plus nombreux pour augmenter la statistique et pouvoir conclure quant à la sélectivité de la sonde. Ceci est toutefois en accord avec les résultats obtenus dans l'expérience de photomarquage avec hRPA suivie de digestion sur gel qui nous a conduit à observer une modification de +16 u. sur le peptide ARTK sur le résidu Arginine 16 ou Proline 17 si l'on

suppose que cet adduit est représentatif d'une interaction physique entre l'ODN 4<sup>S</sup>T et le peptide. Pour aller plus loin et essayer de faire une étude systématique pour déterminer la sélectivité de la sonde 4<sup>S</sup>T, nous avons essayé de reproduire cette étude sur des tripeptides de séquence GXG-CONH<sub>2</sub> en faisant varier X (X = un des 20 acides aminés) mais nous n'avons jamais pu observer de complexes dans nos conditions avec ces petits peptides.

### 5.5.3 Photochimie de la 4<sup>S</sup>T : proposition de mécanisme vis-à-vis d'un peptide

Au cours de ces travaux nous avons montré que la sonde 4<sup>S</sup>T présente plusieurs caractéristiques réactionnelles :

- Une fois la liaison covalente établie, la masse du complexe ODN 4<sup>S</sup>T-peptide ARTK, est égale à la somme de la masse de l'ODN 4<sup>S</sup>T et de la masse du peptide ARTK. La réaction de photochimie de la 4<sup>S</sup>T sur un peptide n'entraîne donc pas de perte de groupement chimique. Elle ne semble pas non plus engendrer de sous-produit.
- L'atome de soufre, indispensable à la réaction de photomarquage n'est pas directement lié sur le peptide ARTK. Le lien covalent n'est donc pas de type S-X. Dans le cas d'une interaction avec un peptide, il est peu probable que le lien covalent s'effectue sur la position 4 de la 4<sup>S</sup>T au niveau de la thiocétone, puisqu'une attaque nucléophile à cette position ne permettrait pas d'engendrer la perte de 16 u.
- La sonde 4<sup>S</sup>T semble avoir de l'affinité pour les résidus basiques. Cette affirmation est à nuancer au vu du faible nombre d'acides aminés que nous avons pu tester, elle reste cependant très informative. Cette sélectivité permet d'expliquer la formation d'une liaison covalente entre notre ODN 4<sup>S</sup>T et le peptide ARTK qui contient 4 lysines, 3 arginines qui sont toutes les deux des acides aminés basiques.

En tenant compte de ces résultats et par analogie avec les mécanismes photochimiques décrits (introduction, paragraphe 3.2.3.2.1, p. 63 et 3.2.3.2.2, p. 65) entre la 4<sup>S</sup>U et les bases de l'ADN (thymine dans nos exemples), nous pouvons émettre l'hypothèse qu'un mécanisme similaire à l'addition 1,4 est probable dans le cas d'une réaction avec la fonction amine d'une lysine (nucléophile) au sein d'un peptide. L'addition 1,2, quant à elle, est possible comme nous avons pu le voir sur les expériences de réactivité sur les acides aminés, mais elle n'explique pas la

perte de 16u. que nous observons dans le cas de la réactivité avec une peptide. Il est aussi possible que ces deux mécanismes aient lieu en même temps. Pour ce qui est de la réaction de cycloaddition [2+2], elle est fortement improbable avec des acides aminés puisqu'elle nécessite la présence d'une double liaison « activée » par un groupement électroattracteur proche, comme dans le cas de la thymine<sup>197</sup>.

Seule l'addition 1,4 semble valider tous nos critères pour la réactivité de la 4<sup>S</sup>T vis-à-vis d'un peptide. En corrélant ces informations au premier mécanisme proposé par Ito *et al.*<sup>198</sup> avec la thiouracile, nous proposons le mécanisme photochimique suivant (Figure 89) :

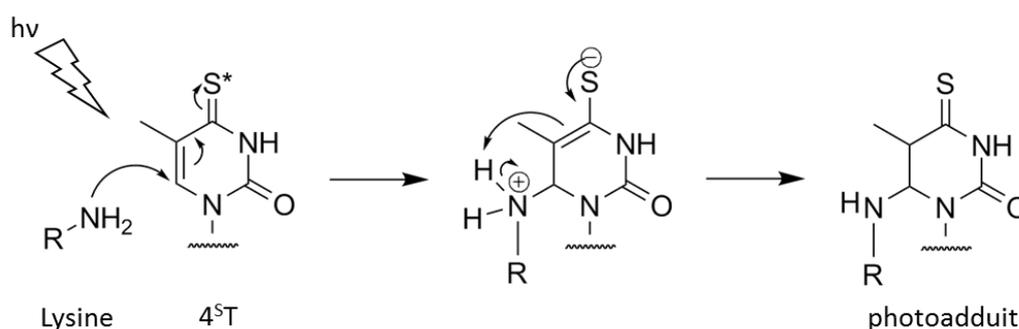


Figure 89 : Schéma du mécanisme proposé de la photochimie de la 4<sup>S</sup>T vis-à-vis d'une lysine au sein d'un peptide.

L'activation de la 4<sup>S</sup>T va rendre possible une réaction de type addition 1,4 au niveau de la position 6 de la base nucléique. Un nucléophile proche, ici la fonction amine libre de la chaîne latérale d'une lysine, peut attaquer la position 6 activée de la 4<sup>S</sup>T pour former le photoadduit proposé Figure 89. Cette addition est réalisée sans perte, ni ajout de masse. De plus, la liaison n'implique pas directement le soufre ce qui est en accord avec nos résultats et hypothèses. Dans ce mécanisme, une seconde activation UV de la 4<sup>S</sup>T permet une réaction intramoléculaire de cette base avec une adénine ou guanine proche, provoquant ainsi la perte de 34 u.

## Conclusion de l'étude méthodologique de complexes photomarqués

L'étude méthodologique destinée à caractériser au niveau moléculaire les complexes photopontés ADN-protéines que nous avons menée ne nous a pas permis d'identifier avec exactitude les résidus mis en jeu lors de l'interaction d'une molécule d'ODN avec nos protéines modèles. Toutefois, cette étude a permis deux choses :

1. L'établissement d'un protocole de cristallisation des dépôts MALDI qui permet d'analyser des oligonucléotides par spectrométrie de masse MALDI-TOF de façon reproductible. L'utilisation de ce protocole permet aussi d'analyser directement des complexes ODN-protéines, sans étape de digestion préalable de l'ODN ou de la protéine liée à celui-ci. Nous espérons pouvoir utiliser ce protocole pour déterminer les séquences de protéines en interaction avec une molécule d'ODN.
2. L'obtention d'information sur la réactivité, non décrite à ce jour, de la photosonde  $4^{\text{ST}}$  vis-à-vis d'un peptide. Nous avons vu au cours de cette étude que la  $4^{\text{ST}}$  réagissait de façon préférentielle avec les lysines. Cette réaction photochimique se réalise sans perte de masse dans le cas d'une réaction sur un peptide et le photoadduit produit a une masse correspondant à la somme de la masse de l'ODN- $4^{\text{ST}}$  et celle de la protéine/peptide en interaction (masse de l'ODN + masse du partenaire = masse du photoadduit). Et finalement, que le photoadduit obtenu pouvait perdre 16 et/ou 34 u.

Ces derniers résultats sont intéressants dans l'optique d'une analyse protéomique qui utiliserait la sonde  $4^{\text{ST}}$ . En effet, l'identification des protéines par spectrométrie de masse s'effectue par une recherche en banque de données où des incréments de masse provenant de modification chimique ou biochimique peuvent être implémentés. C'est-à-dire qu'une protéine liée à l'ODN- $4^{\text{ST}}$  pourrait être identifiée par incrémentation de la masse de cette protéine par la masse exacte de l'ODN- $4^{\text{ST}}$  ou d'un fragment de celui-ci. De plus, on pourrait spécifier au moteur de recherche que la modification est située sur une lysine pour améliorer l'identification protéique. Globalement, une meilleure compréhension des mécanismes de photomarquage permet une identification plus fine des partenaires d'interaction et du site de cette dernière.

Une caractérisation plus poussée de la photochimie de la  $4^{\text{ST}}$  est en cours d'élaboration au laboratoire des Biomolécules. Celle-ci se baserait sur l'étude de l'interaction de la  $4^{\text{ST}}$ , non incorporé dans un oligonucléotide, avec tous les acides aminés par spectrométrie de masse et par RMN avec ou sans irradiation. De même, des expériences de spectrométrie de masse en

ESI, en utilisant une introduction d'échantillon en source nanospray pseudo-dynamique pour la détection des complexes est en cours et nécessite encore certaines optimisations, en particulier pour la préparation d'échantillon.

Dans le chapitre suivant, nous allons nous intéresser à la deuxième approche en spectrométrie de masse : l'analyse protéomique globale, qui permet l'identification de protéines partenaires de modifications épigénétiques tel que les 5mC ou la 5hmC.

## 6 Identification de protéines de liaisons aux modifications épigénétiques de l'ADN en présence des sondes ODN.

Ayant en main tous les outils et les protocoles pour mettre en œuvre la stratégie PAL pour identifier des protéines de liaison aux modifications épigénétiques de l'ADN - plus spécifiquement à la 5-hydroxyméthylcytosine -, nous avons donc effectué quelques essais préliminaires de photomarquage en présence de nos sondes duplexes ODN et de différents extraits cellulaires.

### Essais préliminaires

#### 6.1.1 Essai préliminaires sur un extrait protéique de cellules HeLa

6.1

Les tous premiers essais ont été réalisés par mon prédécesseur Tom Gérard-Hirne à la fin de sa thèse en présence d'extraits de cellules HeLa. Lors de cette première étude, il a comparé deux sondes ODN double brin différentes : l'une contenant la modification 5hmC et l'autre, dite contrôle, contenant une cytosine. Dans les deux cas, le brin complémentaire comportait l'agent de photomarquage diazirine (ODN-Dia) radiomarqué. Après incubation de ces deux sondes (2 nM) avec les extraits nucléaires HeLa (2 mg/mL), une phase d'irradiation à 365 nm a été appliquée avant le dépôt de ces deux échantillons sur gel d'électrophorèse dénaturant (Figure 90).

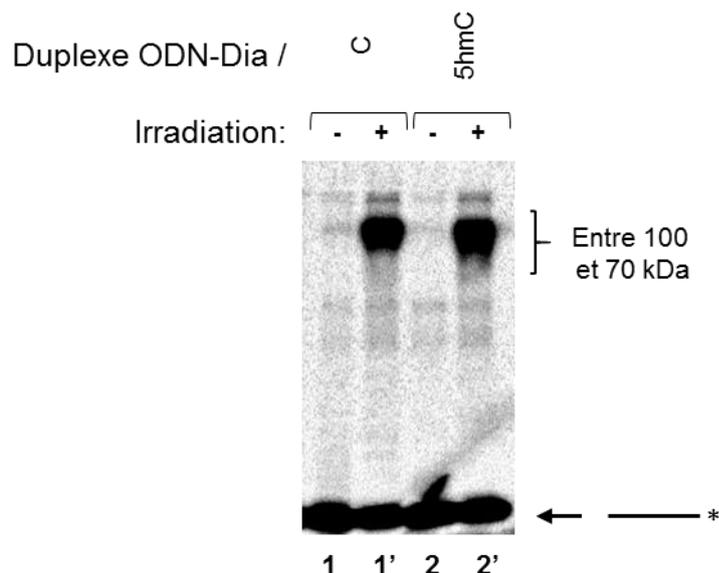


Figure 90 : Photomarquage d'extraits nucléaires de cellules HeLa par les sondes ODN-Dia/C et ODN-Dia/5hmC et séparation sur gel dénaturant d'acrylamide.

Cette sonde permet le photomarquage de protéines puisqu'en l'absence d'irradiation, aucune bande retardée n'est détectée. En revanche, la comparaison des profils de photomarquage sur gel, ne permet pas de mettre en évidence de différences entre les deux duplexes. En effet, leurs profils sont similaires, aucune bande ne se démarque en comparant la piste 1' et 2' de la Figure 90. Cette similarité peut s'expliquer par le fait que ces cellules ne contiennent que de faibles niveaux de 5hmC dans leur génome comme on peut le voir sur la Figure 91, tirée de l'étude de Li et Liu<sup>285</sup>.

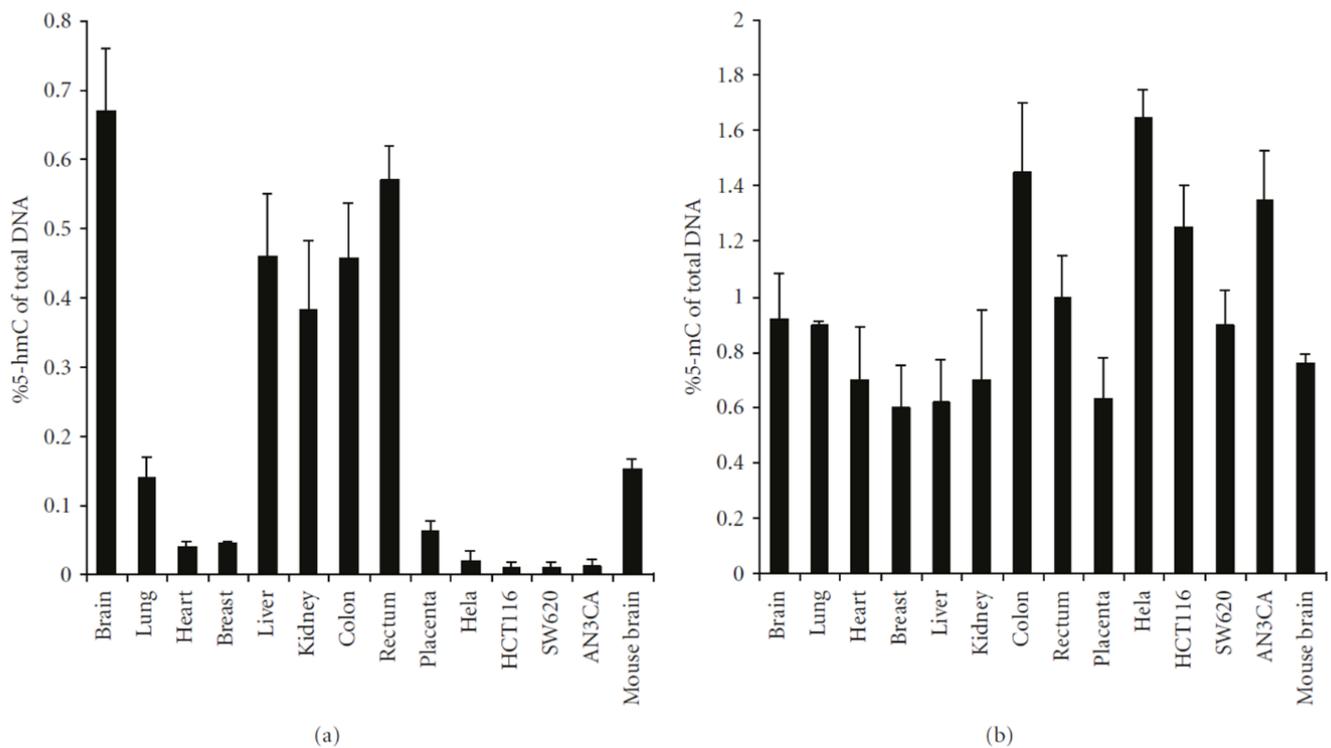


Figure 91 : Quantification des bases 5hmC (a) et 5mC (b) dans l'ADN génomique isolé à partir de tissus humains. Les données sont des valeurs moyennes  $\pm$  écart-type de 3 essais différents d'après Li et Liu (2011).

Au cours de leur étude, Li et Liu ont quantifié les taux de 5hmC et 5mC dans différents tissus humains, ainsi que dans différentes lignées cellulaires cancéreuses humaines en utilisant un dosage immunologique de ces bases modifiées<sup>285</sup>. Le taux de 5mC est relativement stable et élevé au sein des tissus (Figure 91b) (compris entre 0,6 et 1 % de l'ADN total avec une exception pour les cellules du colon à 1,4 %) ainsi qu'au niveau des cellules cancéreuses où on la retrouve particulièrement dans les cellules HeLa (1,6 %) et AN3CA (cellules cancéreuses de l'endomètre) (1,4 %). Les taux de 5hmC sont plus faibles que ceux de 5mC pour toutes les cellules testées dans leur étude (Figure 91a et b). Si l'on s'intéresse aux taux de 5hmC au sein de cellules cancéreuses comme les HeLa (Figure 91a), on observe que cette modification de l'ADN est peu représentée avec moins de 0,1 % de l'ADN total. Ainsi, le choix d'utiliser des

extraits protéiques provenant de cellules HeLa pour caractériser des protéines affines pour la 5hmC n'est pas forcément le meilleur. Globalement, on observe que les taux de 5hmC sont plus élevés au sein des tissus cellulaires et tout particulièrement dans les tissus de cerveau (0,65 %). Afin d'essayer d'obtenir des résultats plus satisfaisants, nous avons reproduit ces expériences mais cette fois en présence d'extraits nucléaires de cerveaux de rats, qui présentent un fort taux de 5hmC<sup>139,286</sup>.

### 6.1.2 Essai préliminaires sur un extrait protéique nucléaire de cerveau de rat

Les extraits nucléaires ont été préparés à partir de cerveaux entiers avec le kit CellLytic™ NUCLEAR™ Extraction, de Sigma-Aldrich. Pour rester dans les mêmes conditions que celles utilisées par Tom Gérard-Hirne, et ainsi pouvoir comparer les résultats obtenus, nous avons utilisé des ODN double brin photoactivables contenant une sonde de type diazirine. Les duplexes ODN-Dia/C, ODN-Dia/5mC, ODN-Dia/5hmC et ODN-Ctr/5hmC (Figure 92A) ont été incubés (2 nM) avec un extrait nucléaire de cerveau de rat (1 mg/mL) avant d'être irradiés à 365 nm pendant 100 min. Les échantillons de cette expérience ont ensuite été déposés sur un gel dénaturant SDS-PAGE. Le gel est présenté sur la Figure 92B.

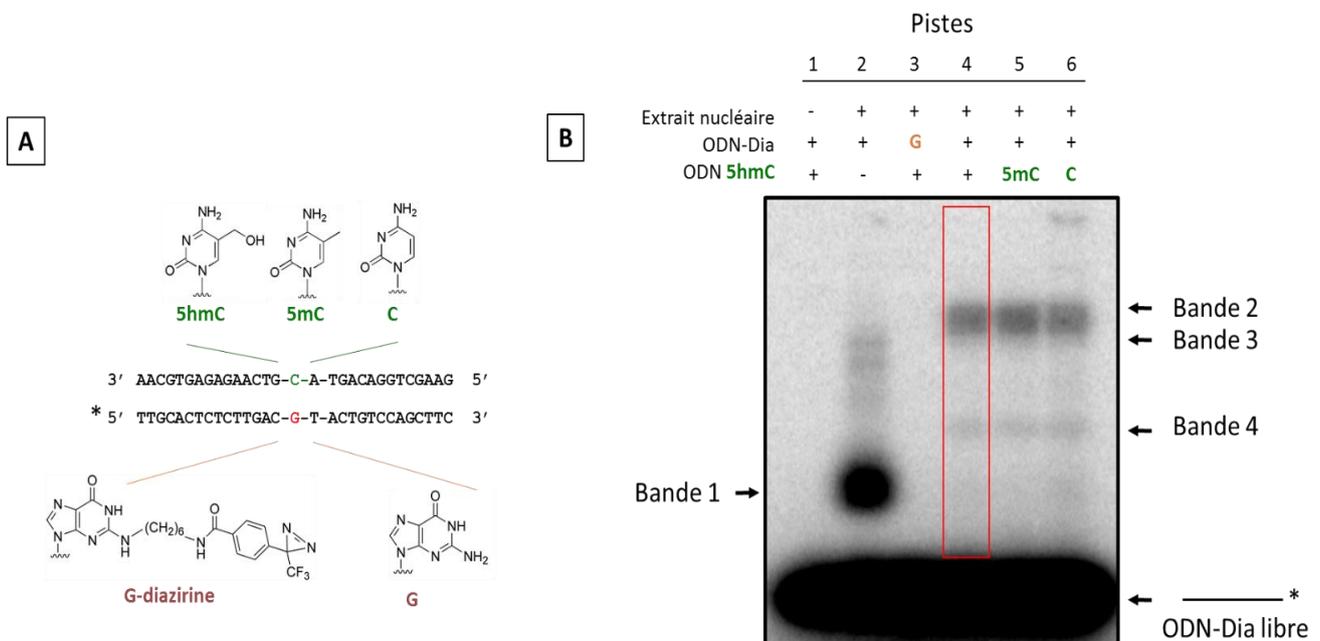


Figure 92 : A) Schéma des ODN-Dia/C, ODN-Dia/5mC, ODN-Dia/5hmC et ODN-Ctr/5hmC. B) Photomarquage d'extraits nucléaires de cerveau de rat par les sondes ODN-Dia/C, ODN-Dia/5mC, ODN-Dia/5hmC et ODN-G/5hmC et séparation sur gel dénaturant d'acrylamide. La représentation schématique de l'ODN marqué (—\*) est indiquée à droite du gel.

Les trois premières pistes de ce gel correspondent à des expériences contrôles : la piste 1 représente l'ODN-Dia/5hmC sans extrait nucléaire et permet d'assimiler toutes bandes visibles sur cette piste à des bandes propres au duplex ODN ; la piste 2 correspond à l'expérience de photomarquage réalisée avec l'ODN-Dia simple brin et met en évidence toutes les bandes qui pourraient être causées par la dénaturation du duplex et induire un photomarquage de l'ODN simple brin ; enfin, la piste 3 correspond à l'ODN-Ctr/5hmC dans lequel le duplexe ne comporte pas de sonde photoactivable. Les pistes 4, 5 et 6 correspondent à l'expérience de photomarquage réalisée avec, respectivement, l'ODN-Dia/5hmC, ODN-Dia/5mC et ODN-Dia/C. Seule la bande correspondant à l'ODN radiomarké libre est visible sur les pistes 1 et 3, on en déduit qu'aucun photomarquage n'a lieu au sein du duplexe (piste 1) et sans agent photoactivable (piste 3). Plusieurs bandes de plus hautes masses moléculaires sont détectables au niveau de la piste 2, une majoritaire (bande 1) et deux autres moins intenses à de plus hautes masses moléculaires montrant ainsi un photomarquage entre l'ODN-Dia et plusieurs protéines de masses différentes. Plusieurs bandes spécifiques du photomarquage sont détectées sur les pistes 4, 5 et 6, montrant un photomarquage efficaces de protéines mais aucune différence n'est notable entre nos différentes sondes. On peut toutefois remarquer que le profil des pistes 4, 5 et 6 est différent de celui de la piste 2, montrant ainsi que l'ODN-Dia ne photomarque pas les mêmes protéines qu'il soit sous forme d'un simple brin ou d'un duplexe, observation que nous avons déjà faite sur d'autres types d'extraits (Chapitre IV). L'absence de différences entre l'ODN-Dia/5hmC (piste 4), ODN-Dia/5mC (piste 5) et ODN-Dia/C (piste 6) est décevante mais, est à relativiser. En effet, l'analyse par gel d'acrylamide couplée à une détection radioactive est très sensible mais il est probable que les bandes détectées sur le gel correspondent à des mélanges de différentes protéines et des différences de profils sur gel ne sont pas nécessairement évidentes à détecter.

### *6.1.3 Bilan des essais préliminaires*

Les deux expériences de photomarquage réalisées sur des extraits cellulaires différents se concluent d'une manière similaire : nous arrivons à photomarquer des protéines affines nos sondes double brin, mais aucune différence n'est observable sur les gels d'électrophorèse entre une sonde possédant une 5-hydroxyméthylcytosine et une sonde contrôle comportant une cytosine. Seule, la spectrométrie de masse, pourra permettre de caractériser des différences. Nous avons tenté de reproduire ces expériences dans le but de caractériser certaines des bandes correspondants à des complexes photomarkés par spectrométrie de masse MALDI-TOF.

Malheureusement, nous nous sommes retrouvés confronté à une limite liée à la quantité de matériel nécessaire pour obtenir des résultats satisfaisants :

- En effet, si le rendement de photomarquage de nos sondes s'évalue autour de 10 %, dans nos conditions, seuls 2 fmol d'ODN sont liées à des protéines au cours de nos expériences utilisant des sondes radiomarquées ce qui correspond à des quantités de protéines inférieures au ng. Ces quantités sont impossibles à analyser par spectrométrie de masse MALDI-TOF selon les protocoles que nous avons décrits précédemment.

Par ailleurs, l'analyse protéomique sur extraits nécessite l'enrichissement des complexes ODN-protéines. En effet, bien que cela ne se remarque pas sur nos gels d'électrophorèses précédents, où seules les molécules d'ODN radiomarquées sont visibles, chaque piste de ces gels contient un grand nombre de protéines, comme nous avons pu le voir sur Figure 59 avec des extraits de cellules HEK293. Une étape d'enrichissement des complexes ODN-protéines à l'aide de billes greffées de streptavidine est donc nécessaire. Nous avons mis en œuvre ces procédures d'enrichissement à l'aide des conditions que nous avons mis au point pour les extraits bactériens surexprimant hRPA. Nous ne sommes jamais parvenus à enrichir efficacement les complexes formés au sein d'extrait de cellules eucaryotes. Les complexes n'ont jamais été détectés, rendant ainsi leurs identifications impossibles. Ce résultat peut être dû à la complexité plus élevée de l'extrait et/ou une quantité trop faible d'oligonucléotide photoactivable par rapport à l'ensemble des biomolécules

Ces difficultés rencontrées nous ont mis face à un choix:

1. Nous pouvions continuer le même système en synthétisant de plus grandes quantités d'oligonucléotides, et en optimisant nos conditions d'enrichissement à l'aide de l'interaction biotine/streptavidine.
2. Nous pouvions changer de système en nous appuyant sur les deux études qui ont permis l'identification des protéines partenaires des dérivés oxydés de la 5-méthylcytosine<sup>99,126</sup> dont nous avons parlé en introduction (voir introduction P xx). C'est-à-dire utiliser nos oligonucléotides modifiés pour réaliser un enrichissement protéique *via* un pull-down.

Nous avons pris la décision de changer de système et d'utiliser une méthode de pull-down couplée à du photomarquage basé sur l'interaction biotine/streptavidine. Ce choix a été favorisé par le fait qu'un tel système avait déjà été développé dans notre laboratoire par Vivien Berthelot

sur des molécules d'ADN longues (>150 pb)<sup>287</sup>, me permettant de gagner un temps précieux en fin de thèse. Bien qu'il soit possible de transposer le système de pull-down développé par Vivien Berthelot à nos oligonucléotides modifiés (30 mères), cela peut poser un inconvénient majeur. En effet, en raison de la capacité d'une molécule de streptavidine à lier quatre molécules de biotine, la gêne stérique engendrée par la proximité des oligonucléotides modifiés peut empêcher l'interaction d'une protéine externe avec ceux-ci. Pour éviter ce phénomène, des oligonucléotides de 125 à 425 pb sont conseillés<sup>288</sup>. L'utilisation de sondes d'ADN plus longue permet aussi de les synthétiser en grande quantité par voie enzymatique à l'aide d'une PCR (*polymerase chain reaction*). Nous avons choisi de réaliser un pull-down avec des sondes d'ADN plus longues.

### **Conception de sondes ADN « boucles » photoactivables**

6.2 Les travaux de Iurlaro *et al.* et de Spruijt *et al.*, comme décrit dans la partie introduction (p.41), ont permis l'identification de différentes protéines partenaires des dérivés oxydés de la 5-méthylcytosine<sup>99,126</sup>. Pour rappel, ces deux études se basent sur un système de pull-down pour enrichir leur système en protéine d'intérêt. C'est-à-dire qu'ils ont fixé de l'ADN possédant plusieurs exemplaires d'une forme oxydée de la 5mC sur un support (billes magnétiques), avant de charger différents extraits cellulaires sur ce support greffé. Le partenaire d'interaction se fixe sur l'ADN. Le support est ensuite lavé pour éliminer toutes les interactions non spécifiques. La dernière étape de ce système consiste en un relargage soit du déterminant d'affinité, soit du partenaire d'interaction à l'aide de conditions d'élution drastiques. Comme nous l'avons vu lors de nos travaux de mise au point, l'incorporation d'un agent de photomarquage peut s'avérer très intéressant pour améliorer l'enrichissement des protéines partenaires et faciliter leur identification par spectrométrie de masse. Il nous a ainsi semblé intéressant d'associer cette approche à l'approche de pull-down utilisée dans les précédents travaux en y apportant quelques modifications.

#### *6.2.1 Un système de pull-down « boucle »*

Ainsi, nous avons choisi d'utiliser un système dit de pull-down « boucle » qui se différencie du système de pull-down classique par le fait que les deux extrémités du duplexe d'ADN sont fixées sur le support solide. Ce système a déjà été utilisé dans notre laboratoire au cours de la thèse de Vivien Berthelot. Ses travaux portaient sur l'identification de protéines reconnaissant

spécifiquement des cassures de l'ADN, c'est-à-dire des extrémités libres<sup>287</sup>. Pour se faire, Berthelot *et al.* ont conçu deux phases chromatographiques différentes, représentées Figure 93.

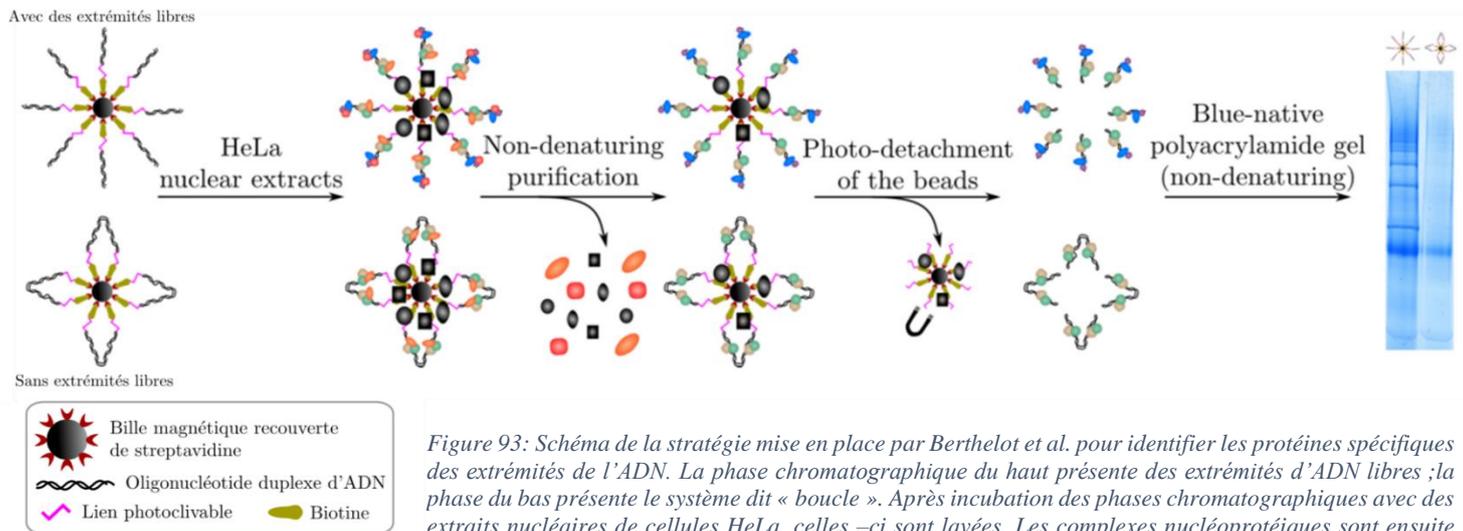


Figure 93: Schéma de la stratégie mise en place par Berthelot *et al.* pour identifier les protéines spécifiques des extrémités de l'ADN. La phase chromatographique du haut présente des extrémités d'ADN libres ; la phase du bas présente le système dit « boucle ». Après incubation des phases chromatographiques avec des extraits nucléaires de cellules HeLa, celles-ci sont lavées. Les complexes nucléoprotéiques sont ensuite détachés des billes par irradiation du lien photo-clivable représenté en rose. Le matériel purifié est déposé sur un gel d'électrophorèse et les bandes résultantes sont excisées et leurs contenus identifiés par LC-MS/MS après digestion trypsique. Figure adaptée de Berthelot *et al.*

La première phase chromatographique (en haut à gauche) correspond à un système avec extrémités libres alors que la deuxième (en bas à gauche) n'en possède pas, l'ADN y est accroché sous forme de boucle. Sur ces phases, le déterminant d'affinité (ADN double brin) est fixé sur des billes magnétiques recouvertes de streptavidine à l'aide de l'interaction biotine/streptavidine. Le groupement biotine et la séquence d'ADN d'étude sont séparés par un groupement photo-clivable, excitable à 365 nm, qui va permettre l'éluion du déterminant d'affinité et de ses protéines liées, des billes magnétiques. Le système à extrémités d'ADN libres a été conçu pour purifier des complexes nucléoprotéiques spécifiques des extrémités de l'ADN tandis que l'autre a été utilisé comme une phase de contrôle qui recrute les protéines présentant une affinité générique pour l'ADN double brin. Leur expérience est détaillée sur la Figure 93. La comparaison des protéines purifiées sur ces deux phases après leurs identifications et leurs quantifications par spectrométrie de masse a permis à Berthelot *et al.* d'identifier les protéines spécifiques des extrémités de l'ADN.

Nous avons décidé de nous inspirer de ces travaux, plus particulièrement de la phase chromatographique contrôle, pour établir notre système d'enrichissement des protéines partenaires de la 5hmC. En effet, cette phase présente deux avantages principaux :

1. Elle permet de s'affranchir de la présence de protéines spécifiques des extrémités de l'ADN. En effet, les travaux de Vivien Berthelot ont montré que lorsque l'ADN ne

présentait pas d'extrémités libres au sein d'un pull-down, celui-ci ne recrutait pas de protéines spécifiques de ces extrémités. Ces protéines n'ont qu'un intérêt limité dans le cadre de notre étude puisque la majorité des modifications épigénétiques de l'ADN se situe au niveau de régions promoteur de gène, eux-mêmes au cœur de la molécule d'ADN<sup>289</sup>.

2. Les conditions de préparation de la phase chromatographique et celles d'incubation des extraits nucléaires ont déjà été optimisées au cours de la thèse de Vivien Berthelot. Ainsi, seules les conditions d'élution et d'analyse seront à mettre au point lors de la réalisation de notre expérience de pull-down.

### 6.2.2 *Un système de pull-down « boucle » photoactivable*

La deuxième modification que nous avons réalisée par rapport à un système de pull-down classique est l'ajout d'agent de photomarquage au sein du déterminant d'affinité. Ainsi, il sera possible d'effectuer des étapes de lavages plus drastiques car les protéines d'intérêt seront liées covalamment. Par ailleurs, comme nous l'avons déjà mentionné, l'utilisation d'agents de photomarquage permet aussi la capture de protéines dites « transitoires » qui ne sont pas enrichies sur les pull-down classiques, leurs interactions avec l'ADN étant de trop courte durée pour qu'elles soient enrichies par cette technique.

La phase chromatographique « boucle » décrite par Berthelot *et al.* comporte un espaceur photoclivable entre la biotine et l'ADN. Dans notre cas, l'utilisation du groupement photoclivable à 365 nm pour l'élution du déterminant d'affinité n'est pas compatible avec l'emploi des agents de photomarquage utilisés dans notre étude puisqu'ils sont activés à la même longueur d'onde. Afin de faciliter l'élution du déterminant d'affinité des billes magnétiques, nous avons envisagé d'utiliser la desthiobiotine à place de la biotine à l'extrémité de nos sondes. La desthiobiotine est un analogue de la biotine qui se lie moins fortement à la

streptavidine<sup>290</sup> (Kd entre  $10^{-11}$  et  $10^{-13}$  M au lieu de  $10^{-14}$  pour la biotine) et est facilement déplacée par labiotine ou par l'application de conditions peu stringentes<sup>291</sup>.

Pour résumer, le système de piégeage des protéines affines pour des modifications épigénétiques de l'ADN à l'aide d'un pull-down couplé au photomarquage est représenté schématiquement sur la Figure 94. Celui-ci nous permet d'éliminer les protéines affines pour les extrémités libres de l'ADN et d'enrichir plus spécifiquement les protéines partenaires de modifications introduites dans notre séquence d'ADN.

Dans un premier temps, la phase chromatographique sans extrémités libre est incubée avec un extrait cellulaire. Pendant cette incubation les protéines affines pour l'ADN et pour ses différentes modifications vont venir se lier à celui-ci. Ce lien va être rendu covalent au cours d'une phase d'irradiation. Le lien covalent formé entre l'ADN et les protéines affines va être mis à contribution lors de la réalisation de lavage en conditions « drastiques » permettant d'éliminer les protéines non affines de notre système. Ces lavages sont effectués comme précédemment décrit, grâce à un aimant qui permet de retenir les billes magnétiques tout en aspirant et éliminant les solutions de lavage. On enrichit donc ce système en protéines d'intérêt. Celles-ci vont pouvoir être identifiées par spectrométrie de masse.

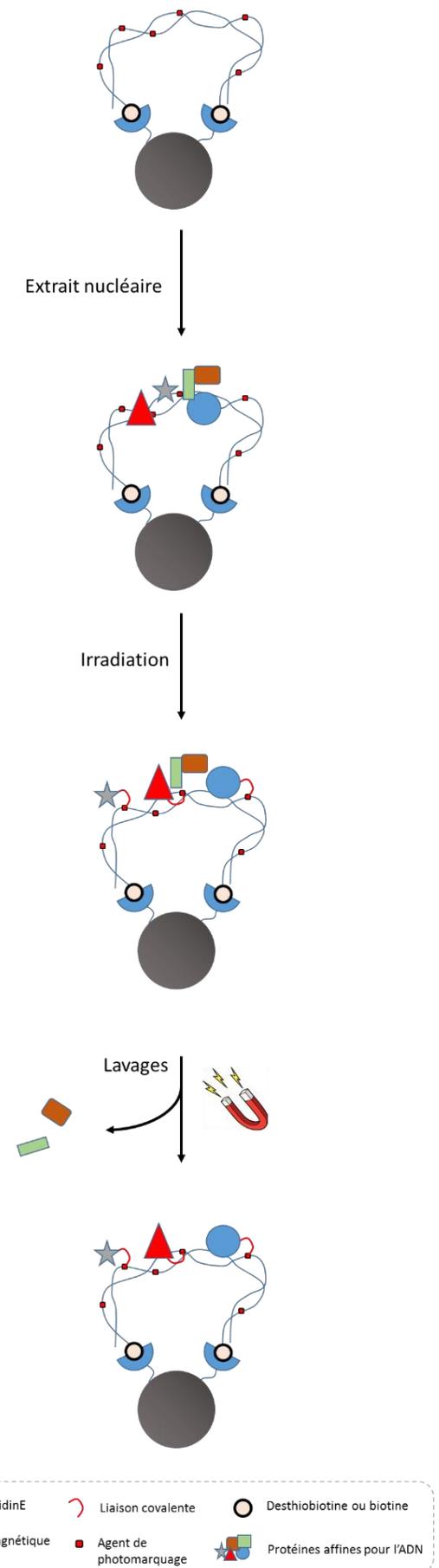


Figure 94 : Principe de l'expérience de pull-down « boucle » couplée à du photomarquage.

### 6.2.3 Préparation des sondes ADN

L'utilisation d'une phase chromatographique « boucle » nécessite l'utilisation d'une molécule d'ADN assez longue d'au moins 100-150 pb pour que celle-ci puisse se replier sur elle-même<sup>292</sup>. Dans cette optique, nous n'avons pas pu utiliser les séquences d'ADN que nous avons déjà à disposition (30-mères pour rappel), nous en avons donc préparé de nouvelles par amplification de séquence via la technique de PCR (*polymerase chain reaction*).

#### 6.2.3.1 Choix de la séquence d'ADN

La séquence d'ADN que nous avons choisie pour réaliser les différentes expériences de pull-down est une fraction de la région *promotor* du gène codant le récepteur de l'acide rétinoïque humain beta (RAR  $\beta$ ), qui a été déjà étudiée au sein du laboratoire. RAR $\beta$  est un membre de la super-famille des récepteurs nucléaires et joue un rôle clé dans la modulation des effets de l'acide rétinoïque (RA) sur la croissance et la différenciation des cellules<sup>293</sup>. Il a été montré que la régulation négative de l'expression de RAR $\beta$  dans de nombreuses lignées cellulaires malignes<sup>294</sup>, était due à la méthylation de la séquence d'ADN codant ce gène<sup>295</sup>. Notamment, une hyperméthylation du promoteur RAR $\beta$  est fréquemment signalée dans le cancer du sein<sup>296</sup>. Widschwendter *et al.* ont déterminé le profil de méthylation de cette séquence d'ADN au sein de lignée cellulaire cancéreuse différentes par un séquençage au bisulfite<sup>297</sup>. Ce profil est présenté sur la Figure 95.



Figure 95 : Séquence d'ADN de la région promotrice du gène de RAR $\beta$  du nucléotide -123 au nucléotide 278. Le site de début de transcription est indiqué par une flèche noire. Les deux éléments RAR et la TATA box sont soulignés. Les positions des 5mC sont indiquées par des cercles pleins ou des cercles avec une étoile, ce qui correspond respectivement à une détection dans une ou deux lignées cellulaires. Les deux encadrés orange ainsi que le surlignage de la même couleur correspondent à notre séquence d'étude. Les encadrés représentent les amorces que nous avons utilisées pour isoler et amplifier cette séquence appelée tRAR $\beta$ , pour template RAR $\beta$ . D'après Widschwendter *et al.*

Sur ce profil, on observe que toutes les méthylation de l'ADN ont lieu au niveau de motifs CpG, et que la majorité d'entre eux se trouve entre les nucléotides 82 et 237. Cette séquence ayant des îlots CpG pouvant être méthylés et étant impliqué dans la régulation de l'expression de gènes, elle nous a parue judicieuse pour la réalisation de notre étude. A l'aide du logiciel

IDT SciTools®, qui permet la conception d'amorces efficaces pour l'amplification par PCR d'une séquence donnée d'ADN, nous avons défini notre séquence d'étude (Figure 95). Celle-ci mesure 202 paires de base, du nucléotide 12 au nucléotide 214, et contient 11 motifs CpG. Pour la suite de ce manuscrit, il s'agira de notre séquence de travail que nous appellerons *template* RAR $\beta$  (tRAR $\beta$ ).

### 6.2.3.2 Synthèse des déterminants d'affinité

Contrairement aux séquences 30-mères utilisées précédemment, la synthèse des différents oligonucléotides duplexes d'ADN servant de déterminants d'affinité a été effectuée à l'aide de la technique d'amplification en chaîne par polymérase (PCR). Cette technique permet de reproduire fidèlement et en grande quantité une séquence d'ADN tout en permettant l'introduction de modifications sous forme de nucléosides triphosphate modifiés.

Dans un premier temps, nous avons isolé et amplifié tRAR $\beta$  à partir d'ADN génomique provenant de cellules HCT116. Pour cela, deux amorces délimitant notre séquence de travail (sens et anti sens) ont été ajoutées à 30 ng d'ADN génomique en présence d'une polymérase (Phusion®) et de dNTP pour réaliser une PCR en gradient de température. Il s'agit d'une technique qui en plus d'amplifier la séquence donnée, permet d'expérimenter plusieurs températures d'appariement de l'ADN aidant ainsi à la mise au point du protocole de PCR, notamment de l'étape d'hybridation. Le résultat de cette première PCR a ensuite été analysé sur un gel d'agarose (Figure 96).

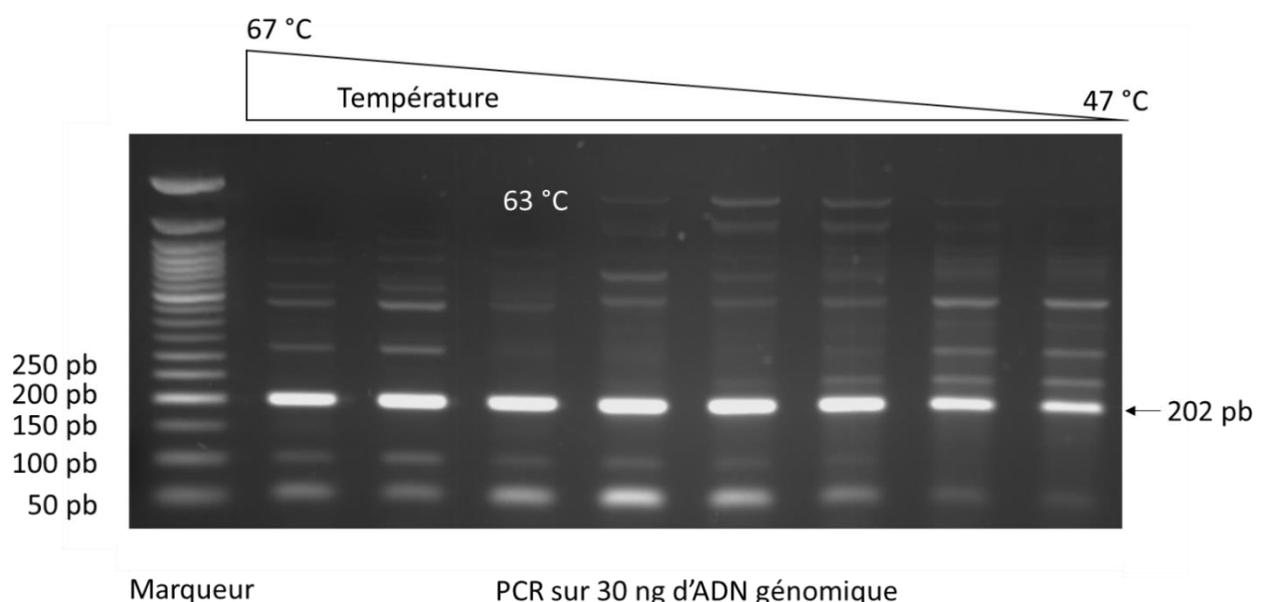
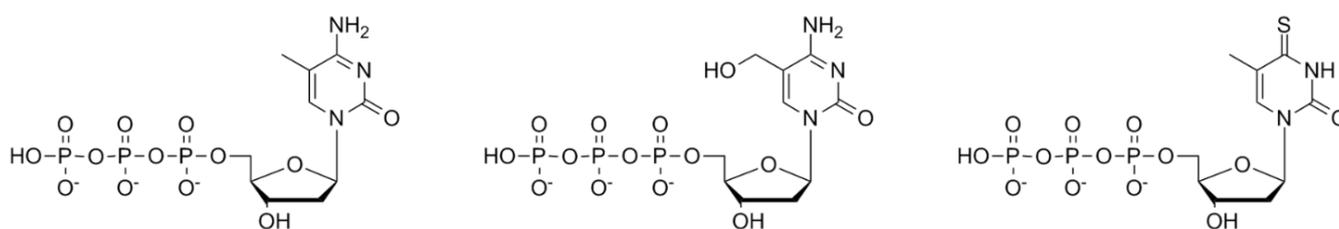


Figure 96 : Détermination expérimentale de la température d'appariement optimale en utilisant un gradient de 47 à 67 ° C. Présentation des résultats sous forme de gel d'agarose 1%

L'expérience de PCR a été effectuée sur un gradient de température étalonné de 47 à 67 °C. Chaque piste du gel présenté sur la *Figure 96* correspond à une température d'hybridation donnée. On observe sur chaque piste de ce gel, une bande autour de ~200 pb, qui correspond à notre séquence RARβ (202 pb). La réaction de PCR a donc bien amplifié notre séquence d'intérêt quelle que soit la température d'hybridation appliquée. L'efficacité de la réaction de PCR est évaluée par observation de 1) l'intensité de la bande correspondant à la taille de tRARβ, et du 2) nombre de bandes visibles sur une piste ainsi que leurs proximités avec la bande d'intérêt. En effet, plus il y aura de bandes apparentes sur une piste, plus la réaction de PCR aura amplifié de séquences d'ADN différentes, réduisant ainsi le rendement de la réaction pour notre séquence d'intérêt. Par comparaison visuelle de ces deux paramètres, on observe que la troisième piste en partant de la gauche correspond à la piste ayant le moins de bandes, et présentant une forte intensité pour la bande située à ~200 pb. Cette piste correspond à la réaction de PCR avec une température d'appariements à 63 °C, température (et conditions) que nous avons gardées pour toutes nos synthèses de tRARβ par PCR. La bande correspondant à tRARβ a ensuite été excisée et une fois l'ADN extrait de cette bande, nous avons réalisé un séquençage de celui-ci pour nous assurer de l'amplification fidèle de notre séquence. Les résultats du séquençage sont conformes à la séquence attendue.

Toutes les modifications chimiques de l'ADN ont ensuite été incorporées sous forme de nucléosides triphosphates commerciaux (*Figure 97*) lors de nouvelles réactions de PCR.



*Figure 97 : Représentation des trois nucléosides triphosphates incorporés au sein de notre déterminant d'affinité. De gauche à droite, la forme triphosphate de la 5mC, celle de 5hmC et de la 4<sup>ST</sup>.*

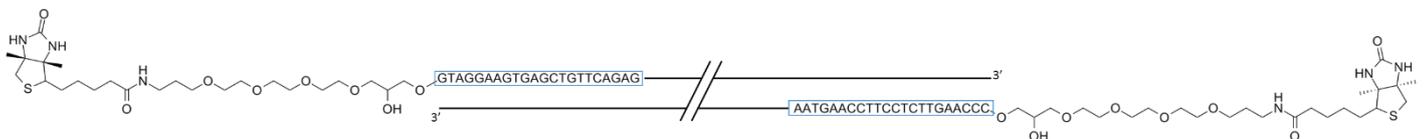
La 5mC et la 5hmC ont été incorporées de façon uniforme dans la séquence, c'est-à-dire qu'à l'exception des amorces, toutes les cytosines de tRARβ sont remplacées par soit de la 5mC soit par de la 5hmC. Cela représente 103 positions méthylées ou hydroxyméthylées, le déterminant d'affinité est donc sous une pseudo-forme d'hyper-méthylation ou d'hyper-hydroxyméthylation. Pour ce qui est de l'agent de photomarquage, notre choix s'est porté sur la 4-thiothymidine et non plus sur la diazirine pour les raisons suivantes :

- Sa structure analogue aux nucléobases qui va limiter les modifications structurales de notre sonde d'ADN.
- Son caractère « *zero-length* » qui implique que seules les protéines en interaction directe avec l'ADN peuvent être photomarquées.
- Sa photoréactivité : toutes nos expériences précédentes ont montré que la 4<sup>ST</sup> est réactive que le milieu soit complexe ou non, et que les rendements de photomarquage sont bons.
- Son accessibilité : en effet, la 4<sup>ST</sup> est commercialisée sous forme de nucléoside triphosphate et ne nécessite pas d'étapes de synthèse organique supplémentaires.

Afin de perturber au minimum l'appariement de notre double brin d'ADN et de garder la 4<sup>ST</sup> comme une modification mineure par comparaison avec la 5mC ou à la 5hmC, seules 5 % des thymines de notre séquence d'ADN ont été remplacées par des 4<sup>ST</sup>. Pour cela, la solution de dTTP introduite lors de la réaction de PCR est constituée de 95 % de dTTP et de 5 % de d4<sup>ST</sup>TTP. Cela représente une moyenne de 3,9 d'incorporations de 4<sup>ST</sup> par déterminant d'affinité.

Les amorces sens et antisens ont été fonctionnalisées par un groupement biotine ou desthiobiotine. Ainsi, au cours de l'amplification, le fragment d'ADN produit est fonctionnalisé à ses deux extrémités par ces marqueurs d'affinité (Figure 98) et les amplicons seront donc capables de se lier aux billes recouvertes de streptavidine pour former la phase chromatographique.

A



B

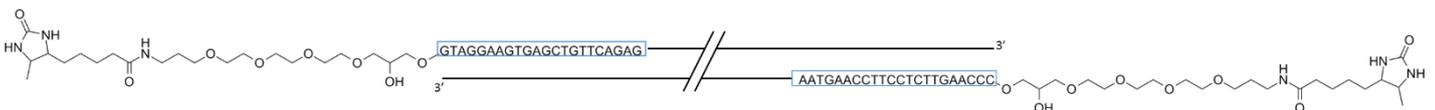


Figure 98 : Schéma de déterminants d'affinité sous leurs formes biotinylées (A) ou desthiobiotinylées (B)

Au terme de la PCR, il est possible que des amorces n'ayant pas réagi soient encore présentes dans le milieu réactionnel. Leur présence au cours de la fixation des produits de PCR sur les billes magnétiques greffées de streptavidine, représente un inconvénient puisque les amorces biotinylées joueraient le rôle de compétiteur pour nos oligonucléotides duplexes d'ADN. De plus, ces amorces bien que courtes (22 paires de bases), pourraient recruter des protéines spécifiques de l'ADN simple brin, ce qui engendrait un plus grand nombre de protéines non spécifiques dans nos échantillons. Pour ces raisons, les amorces résiduelles présentes dans les produits de PCR sont éliminées à l'aide d'un gel d'électrophorèse d'agarose 1 %.

Par cette méthode, nous avons préparé 12 déterminants d'affinité différents. Les ADN ne possédant pas de modification seront notés C, pour cytosine, ceux ayant des 5mC seront notés mC et enfin ceux ayant des 5hmC seront notés H. La présence de  $4^S$ T dans la séquence sera révélée par l'ajout du chiffre 4 et la fonctionnalisation des extrémités par une biotine ou une desthiobiotine sera respectivement indiquée par B ou DB. Chaque modification épigénétique de l'ADN (C, 5mC et 5hmC) va être représentée par 4 molécules d'ADN :

1. Sans agent de photomarquage sous forme biotinylée (CB, mCB et HB)
2. Sans agent de photomarquage sous forme desthiobiotinylée (CDB, mCDB et HDB)
3. Avec agent de photomarquage sous forme biotinylée (C4B, mC4B et H4B)
4. Avec agent de photomarquage sous forme desthiobiotinylée (C4DB, mC4DB et H4DB)

Les formes 1 et 2 seront utilisées pour réaliser un pull-down dit « classique » dans le but de pouvoir le comparer à notre pull-down couplé à du photomarquage utilisant les formes 3 et 4.

### 6.2.3.3 Préparation des phases chromatographiques « boucle »

Les phases chromatographiques « boucles » sont préparées par incubation de billes magnétiques recouvertes de streptavidine et des oligonucléotides duplexes d'ADN produits par PCR dans une solution tampon. L'incubation se déroule sur une nuit à 4 °C sous faible agitation. Après cette étape, les phases chromatographiques sont lavées deux fois avec du tampon d'incubation pour éliminer tous les oligonucléotides qui se seraient liés aux billes de manière non spécifique, c'est-à-dire pas par l'interaction (desthio)biotine/streptavidine. Les phases chromatographiques sont alors prêtes pour l'expérience de pull-down.

## Mise en place de l'analyse protéomique des pull-down « boucle »

### 6.3.1 Mise en place des conditions d'élution

6.3 Avant de réaliser l'expérience de pull-down qui nous permettra d'identifier et de quantifier les protéines affines pour les modifications épigénétiques de l'ADN, nous avons effectué une expérience de pull-down avec les phases chromatographiques ne possédant pas de modifications épigénétiques pour tenter de déterminer les meilleures conditions d'élutions des protéines de notre phase chromatographique.

Nous avons effectué cette mise au point en incubant des phases chromatographiques CB, CDB et C4DB avec 800 µg d'extraits nucléaires commerciaux provenant de cellules HeLa. Après une incubation de 10 min à 30 °C, la phase chromatographique C4DB est irradiée 60 min à 365 nm, alors que les autres phases sont placées à 4 °C. Les suspensions [billes + protéines] sont alors lavées quatre fois avec du tampon Ku (40 mM HEPES; 5 mM MgCl<sub>2</sub>; 60 mM KCl; 0.4 mM EDTA; 37 mM glycérol; 0.01 % NP-40; 0.5 mM DTT; pH= 7.4) pour éliminer toutes les protéines fixées de manière non spécifiques. Les phases sont ensuite lavées plusieurs fois à l'eau distillée avant d'être éluées selon 3 protocoles différents :

- 1) Elution thermique : les échantillons sont chauffés 10 min à 95°C. Dans ces conditions, toutes les protéines fixées sur la phase chromatographique vont être dénaturées et relarguées des billes, qu'elles soient spécifiques ou non.
- 2) Elution par compétition avec la biotine : les échantillons sont placés dans une solution de 10 µM de biotine, à 37 °C pendant 10 min sous faible agitation. La biotine rentre en compétition avec le déterminant d'affinité lié à la streptavidine via une biotine ou une desthiobiotine, permettant ainsi son élution, plus efficacement pour les déterminants ayant des desthiobiotines à leurs extrémités.
- 3) Elution en conditions acides : les échantillons sont placés dans une solution de 0.1 M glycine-HCl pH= 2.8, pendant 10 min à 37 °C sous faible agitation. Ces conditions acides « douces » permettent de libérer les groupements biotines sans dénaturer la streptavidine. Il y a ainsi moins de risques qu'elle contamine nos échantillons protéiques.

Les éluats des trois phases chromatographiques (CB, CDB et C4DB après irradiation) sont ensuite dessalés par passage sur un gel d'électrophorèse dénaturant sans effectuer de migration séparative avant d'être digérés dans le gel par la trypsine. En parallèle, nous avons réalisé un lot d'expériences avec CB, CDB et C4DB sans élution en digérant les protéines directement sur

les billes magnétiques. Les digestats obtenus de toutes nos expériences ont été analysés par nanoLC ESI-MS/MS. Pour des raisons techniques de panne d'appareil, cette analyse n'a pas pu être réalisée sur la plateforme de spectrométrie de masse de l'UPMC comme les autres analyses de spectrométrie de masse réalisées dans ce travail de thèse. Cette analyse a été effectuée sur la plateforme de spectrométrie de masse de l'institut Curie dirigée par Damarys Loew. Les identifications de protéines en banques de données (Swissprot 2017) ont été effectuées à l'aide du logiciel *Proteome Discoverer* (version 2.1) en utilisant le moteur de recherche Mascot. Les résultats de ces analyses, en termes de nombre de protéines identifiées, sont résumés dans le Tableau 3.

	CB		CDB		C4DB	
	Nb de protéines identifiées	Nb de peptides identifiés pour la streptavidine	Nb de protéines identifiées	Nb de peptides identifiés pour la streptavidine	Nb de protéines identifiées	Nb de peptides identifiés pour la streptavidine
Elution thermique	816	13	823	12	768	12
Elution acide	1238	3	1183	3	1321	3
Elution biotine	297	5	665	5	906	5
Digestion sur billes	2193	7	2218	6	2258	6

Tableau 3 : tableau récapitulatif du nombre de protéines identifiées par spectrométrie de masse ainsi que du nombre de peptides ayant permis l'identification de la streptavidine pour les 12 expériences réalisées dans le but d'optimiser les conditions d'élution de notre système.

L'évaluation des méthodes d'élution que nous avons testées a été effectuée selon deux paramètres :

- Le nombre de protéines identifiées par spectrométrie de masse. Plus un échantillon va contenir de protéines, plus celui-ci a de chance de contenir des protéines faiblement abondantes.
- Le nombre de peptides identifiés permettant l'identification de la streptavidine. En effet, si cette protéine est éluee, cela signifie que toutes les protéines sont relarguées dans les éluats, c'est-à-dire à la fois les protéines spécifiques de l'interaction biotine ou desthiobiotine/streptavidine (= protéines d'intérêt) mais aussi les protéines adsorbées de façon non spécifique sur les billes. Le nombre de peptides identifiés pour la streptavidine donne une idée de son abondance dans l'échantillon. Ainsi, plus on identifie de peptides appartenant à la streptavidine, plus l'échantillon contient de protéines non spécifiques.

Si l'on observe les résultats présentés dans le Tableau 3, seule l'élution thermique présente un nombre élevé de peptides permettant l'identification de la streptavidine (12), indiquant ainsi une grande proportion de protéines non spécifiques. Avec respectivement 3, 5 et 6 peptides, les conditions acides, biotine et digestion sur billes correspondent aux conditions ayant une meilleure proportion supposée de protéines d'intérêt par rapport aux protéines non spécifiques. Pour ce qui est du nombre de protéines par échantillon, seules deux conditions donnent accès à un nombre élevé de protéines identifiées : l'élution acide (~1100-1300 protéines) et la digestion trypsique sur billes (~2100-2300 protéines). Dans le cas de l'élution à la biotine, le nombre de protéines identifiées est faible et cette condition ne sera donc pas retenue. En revanche, il est intéressant de noter que dans cette condition d'élution, la quantité de protéines identifiées augmente entre les pull-down avec CB, CDB et C4DB (échantillon réalisé avec irradiation). Cela s'explique par le fait que la biotine entre plus facilement en compétition avec l'ADN desthiobiotinylé qu'avec l'ADN biotinylé. L'ajout de la 4<sup>S</sup>T (C4DB), permet aussi d'augmenter le nombre de protéines identifiées. Cette augmentation est due à l'établissement de liens covalents entre les protéines et C4DB lors de la phase d'irradiation. Ce résultat est encourageant puisqu'il prouve que notre système est fonctionnel et permet d'enrichir des protéines de plus faible affinité pour l'ADN, qui sont éliminées par les lavages stringents en l'absence de photomarquage.

En fonction de ces résultats, nous avons décidé d'utiliser la digestion trypsique sur billes pour la suite de nos expériences. En effet, cette préparation d'échantillon permet l'identification d'un plus grand nombre de protéines. Ceci est probablement dû au fait que, comparativement aux autres protocoles, elle ne nécessite pas d'étape supplémentaire de dessalage des échantillons en fin de pull-down sur gel d'électrophorèse et permet une digestion trypsine d'efficacité quasi équivalente à une digestion en solution. Pour des raisons techniques, seules les molécules d'ADN comportant des desthiobiotines ont été réalisées pour la formation des pull-down « boucle » de la suite de cette étude. Leurs analogues comportant des biotines ne font pas partie de ce manuscrit mais seront réalisés au sein de notre laboratoire.

### 6.3.2 Les différentes phases chromatographiques utilisées

Afin de tester notre pull-down « boucle » couplé à du photomarquage, nous avons réalisé sept phases chromatographiques différentes (Figure 99) formées comme précédemment décrit.

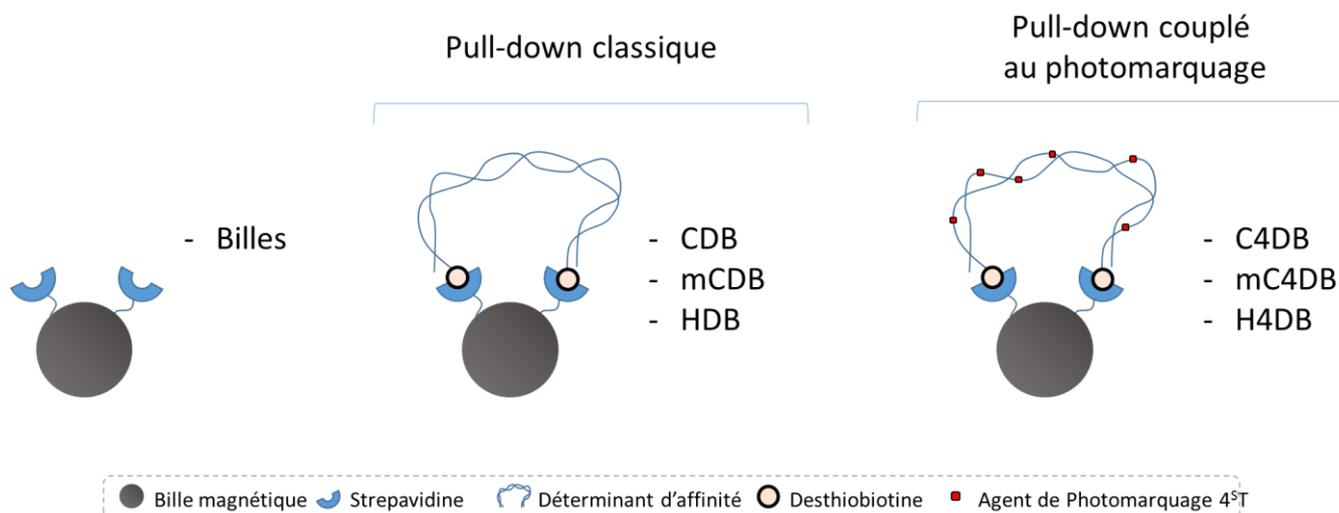


Figure 99: Représentation schématique des phases chromatographiques utilisées pour le pull-down.

La première correspond aux billes magnétiques recouvertes de streptavidine. Cette phase sert de contrôle et permet d'identifier les protéines non spécifiques aux différents déterminants d'affinité (ADN). Trois phases ont été réalisées pour effectuer un pull-down sans photomarquage (classique) avec ou sans modifications épigénétiques : CDB, mCDB et HDB. Ces trois phases chromatographiques serviront aussi de référence pour le pull-down avec photomarquage réalisé avec C4DB, mC4DB et H4DB.

Toutes les phases chromatographiques ont été incubées 10 min à 30 °C avec 800 µg d'extraits nucléaires provenant de cellules HeLa. Après incubation, les phases chromatographiques possédant la  $4^5T$  sont irradiées 60 min à 365 nm sur un bloc de glace, alors que les autres phases sont placées à 4 °C pendant cette étape. Les différentes phases [billes + protéines] sont alors lavées quatre fois avec du tampon Ku (40 mM HEPES; 5 mM  $MgCl_2$ ; 60 mM KCl; 0.4 mM EDTA; 37 mM glycerol; 0.01 % NP40; 0.5 mM DTT; pH= 7.4) pour éliminer toutes les protéines fixées de manière non spécifique. Les billes sont finalement lavées plusieurs fois à l'eau distillée en diminuant progressivement le volume de lavage jusqu'à l'obtention d'un volume final de 40 µL avant que les échantillons ne soient digérés à la trypsine puis analysés par nanoLC ESI-MS/MS. Toutes les expériences et analyses ont été réalisées en 5 exemplaires.

### 6.3.3 Description de la méthode de quantification « label free » des protéines par nanoLC ESI-MS/MS utilisée

L'approche de quantification par spectrométrie de masse que nous avons utilisée dans notre étude est une approche « *label-free* ». Nous avons choisi ce type d'approche car elle est facile à mettre en place, permet théoriquement de comparer sur un nombre illimité d'échantillons et permet de travailler sur n'importe quel type d'extrait cellulaire sans marquage chimique ou métabolique préalable de ceux-ci. Ce dernier point est intéressant pour l'application, à terme, de notre système puisque l'identification de protéines spécifiques des bases oxydées de la 5mC pourra se baser sur différents types cellulaires présentant des taux élevés en modifications épigénétiques comme les cellules embryonnaires ou les cellules de cerveau.

L'analyse quantitative des protéines enrichies au cours des différents pull-down a été effectuée à l'aide du logiciel développé par l'Institut Curie myProMS<sup>298</sup> v3.0, qui effectue à la fois la validation des identifications obtenues via des moteurs de recherche, l'évaluation du taux de faux positif (FDR *false discovery rate*), la quantification des protéines à l'aide du logiciel MassChroQ<sup>299</sup>, l'analyse statistique et la visualisation des données. Ceci est réalisé en deux étapes.

1. **Une analyse qualitative** qui permet l'identification des protéines enrichies à l'aide du logiciel *Proteome Discoverer* (Sequest). La banque de données utilisée est la banque humaine SwissProt (2017). Pour cette analyse qualitative, une protéine a été considérée comme présente si au moins trois peptides de cette protéine étaient détectés avec un FDR de 1 %.
2. **Une analyse quantitative des protéines enrichies.** Cette quantification a été effectuée selon une stratégie « *label free* », par la méthode de l'extraction du chromatogramme ionique ou *eXtracted Ion Chromatogram* (XIC) via le logiciel MassChroQ.

Le XIC est une méthode de quantification basé sur l'intensité du signal MS des peptides caractéristiques d'une protéine. Bondarenko *et al.* ont montré qu'il y avait une excellente linéarité sur quatre ordres de grandeur entre l'intensité du signal dû à un peptide et sa concentration<sup>300</sup>. Le XIC correspond à la production d'un chromatogramme d'ion extrait qui va permettre de connaître la contribution d'un ion (un m/z donné) et donc d'un peptide donné au signal MS total. A titre d'exemple, la Figure 100 représente le TIC (courant ionique total) correspondant à l'analyse d'un des éluats provenant d'un pull-down (A) ainsi que le XIC pour un peptide de m/z donné (B).

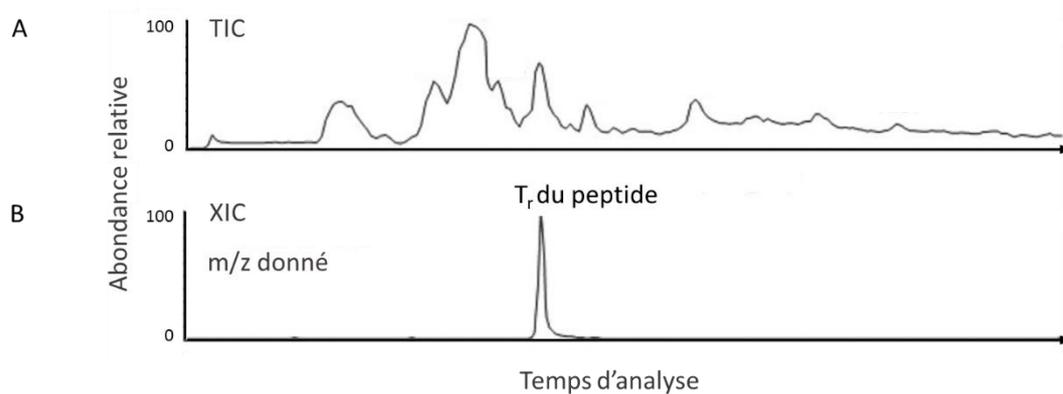


Figure 100 : Représentation d'un chromatogramme correspondant à un peptide extrait par la méthode XIC (B) par rapport au chromatogramme correspondant à l'analyse de l'échantillon totale.

L'aire du pic du XIC (B) peut être utilisée pour une quantification relative de ce peptide dans l'éluat considéré. Pour se faire, la méthode du XIC est appliquée sur tous les peptides identifiés comme appartenant à une protéine et la quantité relative de cette protéine est calculée à l'aide de la moyenne géométrique des aires des pics des XIC obtenus. Cette approche est ensuite reproduite sur toutes les protéines identifiées au sein d'un échantillon, permettant ainsi d'obtenir la quantité relative de toutes les protéines pour chaque échantillon provenant, dans notre cas, des différentes phases chromatographiques utilisées. La quantification étant relative, elle se fera toujours entre plusieurs échantillons. La comparaison de la quantité d'une protéine au sein de deux échantillons différents correspond au rapport entre les moyennes géométriques des différents peptides de cette protéine. Ce rapport est appelé FC (*Fold change*). Un enrichissement d'une protéine dans une analyse par rapport à une autre est significatif à partir d'un FC de 1,5, c'est-à-dire que cette protéine est 1,5 fois plus présente dans une condition que dans l'autre. Ainsi, une protéine est enrichie d'une expérience à l'autre uniquement si son  $|FC| \geq 1,5$  et sa  $p\text{-value} < 0,05$ . Dans la suite de ce manuscrit, seules les protéines validant ses critères seront présentées.

Il est important de noter que la quantification par la méthode des XIC dépend des paramètres expérimentaux tels que les critères d'exclusion dynamique des ions précurseurs (ions correspondant des peptides), la largeur des pics chromatographiques et la vitesse de scan du spectromètre de masse. Dans la mesure où en quantification *label-free* tous les échantillons à comparer sont analysés séparément, il faut veiller à ce que les conditions d'analyses LC-MS/MS soient stables et comparables pour tous ces échantillons (Reproductibilité des profils chromatographiques, stabilité de la source d'ionisation et de la détection). Dans notre cas, l'analyse de nos 35 échantillons a été effectuée à la suite, sur le même appareil et dans les mêmes conditions pour éviter au maximum les biais expérimentaux.

#### 6.3.4 Validation des phases chromatographiques pour l'enrichissement de protéines affines pour l'ADN

Avant de pouvoir identifier les protéines spécifiques de la 5mC et de la 5hmC, nous nous sommes assurés que nos phases chromatographiques permettaient bien d'enrichir des protéines affines pour l'ADN. Nous avons effectué la quantification des protéines enrichies avec la phase chromatographique CDB comparée à celles obtenues avec la phase contrôle ne contenant que les billes greffées à la streptavidine. Le résultat de cette quantification est présenté sur la Figure 101.

Cette figure correspond à un Volcano plot où chaque protéine identifiée est représentée par un point bleu dont la taille est proportionnelle au nombre de peptides utilisés pour l'identification de la protéine en question. Les protéines se trouvant dans la zone rouge ( $FC < -1,5$  ;  $p\text{-value} < 0,05$ ) correspondent aux protéines enrichies par la phase Billes alors que celles se trouvant dans la zone jaune ( $FC > 1,5$  ;  $p\text{-value} < 0,05$ ) correspondent aux protéines enrichies par la phase CDB. Ces protéines sont donc enrichies en présence d'ADN. Les protéines se trouvant hors de ces zones correspondent à des protéines non enrichies que l'on retrouve dans les mêmes quantités sur la phase CDB que sur la phase Billes.

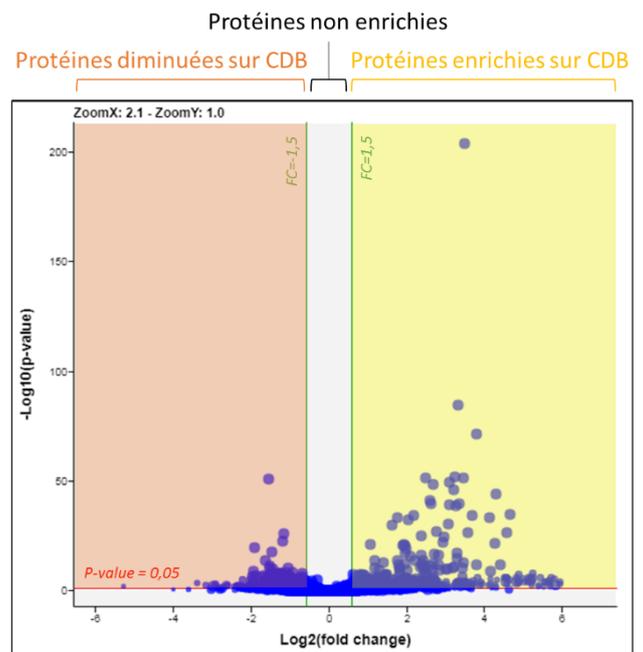


Figure 101 : Volcano plot représentant la quantification des protéines enrichies par la phase chromatographiques CDB par rapport aux protéines enrichies par la phase chromatographique Billes.

Dans le but de savoir si les protéines enrichies par la phase CDB sont spécifiques ou en lien avec l'ADN, nous avons réalisé un enrichissement d'annotations GO (Gene Ontology) sur celles-ci. Cette analyse permet d'identifier les fonctions moléculaires effectuées par ces protéines et sont représentées sous forme d'arbre sur la Figure 102 .

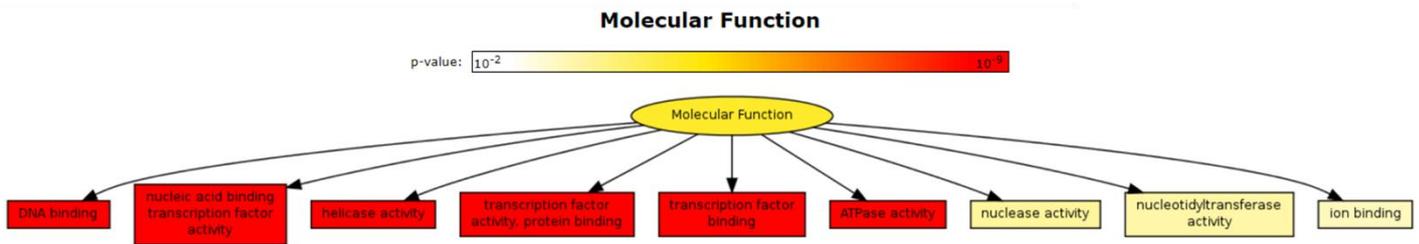


Figure 102 : Résultat de l'enrichissement des annotations GO effectuées sur les protéines enrichies par la phase CDB par rapport à la phase Billes. La valeur de la p-value correspond à la probabilité que la fonction décrite ne soit pas effectuée. Plus la p-value est basse, plus la fonction moléculaire est représentée.

Sur l'ensemble des protéines enrichies par la phase chromatographique CDB, la majorité est impliquée dans une fonction en lien avec l'ADN (encadrés rouge), on peut citer par exemple l'activité hélicase qui permet d'ouvrir la double hélice d'ADN. Nos phases chromatographiques constituées d'ADN, permettent bien d'enrichir des protéines affines pour l'ADN.

## 6.4 L'analyse protéomique des pull-down « boucle »

### 6.4.1 Méthode employée pour l'analyse des résultats obtenus

L'identification des protéines affines et spécifiques de la 5mC et de la 5hmC nécessite une analyse méticuleuse des résultats obtenus. Pour une meilleure compréhension de cette analyse, celle-ci est résumée ci-dessous (Figure 103).

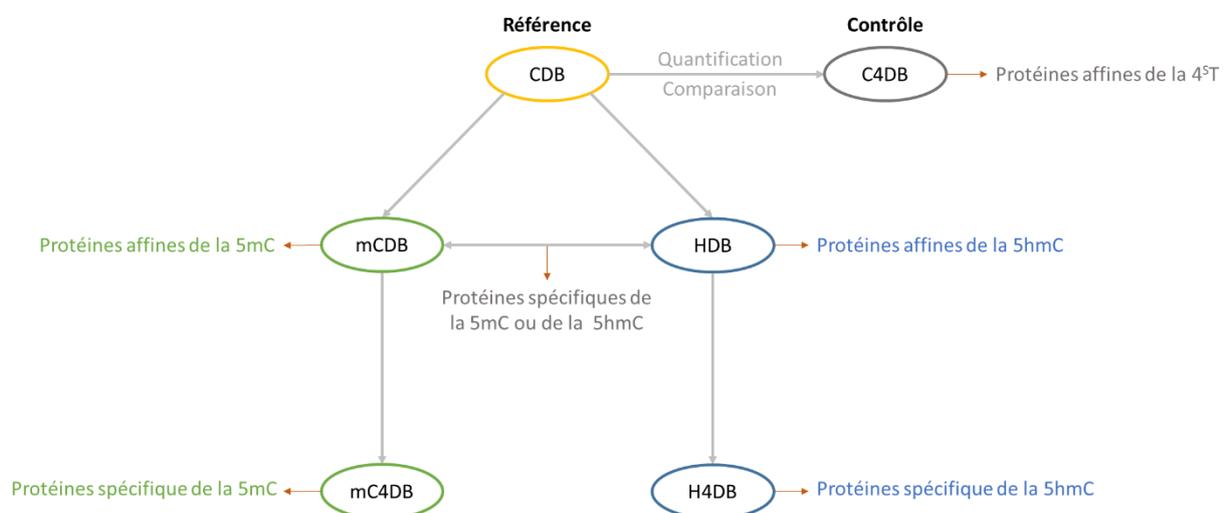


Figure 103 : Représentation schématique de l'analyse des données obtenues par spectrométrie de masse dans le but d'identifier les protéines spécifiques de la 5mC et/ou de la 5hmC. Les ovales correspondent aux expériences de pull-down réalisées avec les différentes phases chromatographiques.

Le pull-down réalisé par la phase chromatographique CDB sert de référence et permet théoriquement d'enrichir toutes les protéines affines pour l'ADN double brin. La comparaison de cet enrichissement de référence avec celui obtenu avec les phases :

- mCDB va permettre de déterminer les protéines affines pour la 5mC.
- HDB va permettre de déterminer les protéines affines pour la 5hmC.
- C4DB va permettre de déterminer les protéines affines pour la 4<sup>ST</sup> et sensible à nos conditions d'irradiation.

La comparaison des pull-downs HDB et mCDB permettra l'identification des protéines sélectives d'une modification épigénétique par rapport à l'autre.

Enfin, les pull-downs effectués avec irradiation (ceux qui contiennent la sonde 4<sup>ST</sup>, c'est-à-dire mC4DB et H4DB) seront comparés aux pull-down correspondants sans sonde et sans irradiation (mCDB et HDB) pour voir l'effet du photomarquage covalent sur les protéines partenaires de l'ADN identifiées.

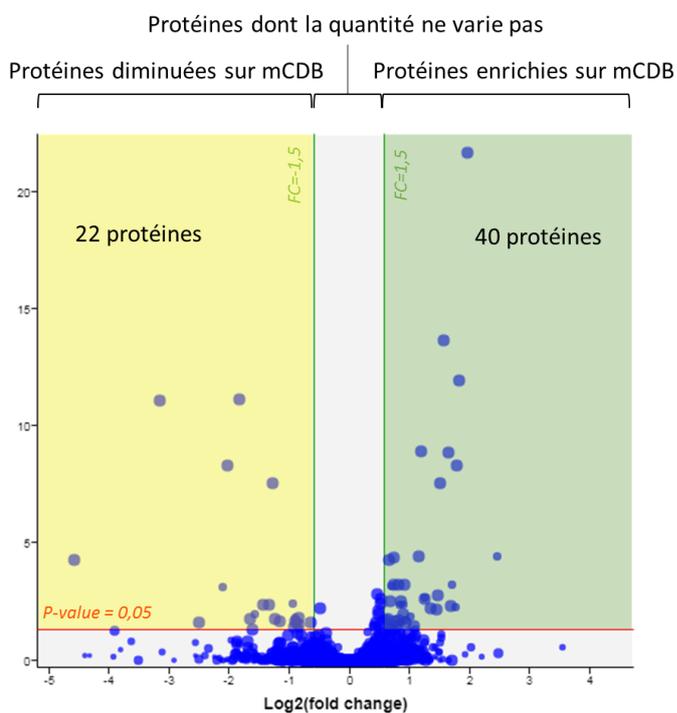
## 6.4.2 Le pull-down classique pour identifier les protéines affines de la 5mC et de la 5hmC

### 6.4.2.1 Enrichissement des protéines affines pour la 5mC

L'identification des protéines affines pour la 5mC est réalisée à l'aide de la quantification des protéines enrichies par le pull-down mCDB comparées aux protéines identifiées dans le pull-down de référence, CDB. Le résultat de cette analyse est présenté sur la Figure 104.

Protéines dont la quantité **diminue** en présence de 5mC

MSH3
UBTF
BEND3
KHSRP
RB1
TFAP4
FUBP1
PURB
HELB
MAPRE1
DEAF1
TFAP2A
TAF1A
EZR
ARPC2
ENO1
NEIL1
MYC
NSUN3
SCIMP
LINC01600
DISP2



Protéines dont la quantité **augmente** en présence de 5mC

BAZ1A	LCORL
SIN3A	STAG1
SMARCA5	HDAC3
KMT2C	MYRF
PDS5B	ASXL1
UHRF1	FOXO3
ZHX2	KLK13
ZFHX3	KCNJ15
SMARCA1	PCBD1
ARID4B	L3MBTL4
FOKK1	TG
CHAF1A	OLAH
FOKK2	UMOD
ZHX3	TRIM34
ZHX1	CDH15
ARID4A	CBX6
FOXP1	AK9
MAML1	
FOXC1	
MECP2	
HNF1B	
TLK1	
UHRF2	

Figure 104 : Volcano plot représentant la quantification des protéines enrichies par le pull-down mCDB par rapport aux protéines enrichies par le pull-down CDB. La liste des protéines enrichies par la phase mCDB ou CDB sont représentées sur chaque côté du graphique. Les protéines qui ont déjà été caractérisées comme étant affines pour la 5mC et/ou la 5hmC ont été hachurées sur les listes.

Cette expérience a permis l'identification de 5968 protéines. Les protéines dont la quantité ne varie pas d'un pull-down à l'autre ( $|FC| < 1,5$  et une  $p\text{-value} > 0,05$ ) sont des protéines ne présentant pas d'affinité particulière pour la 5mC. Parmi les 5968 protéines identifiées, 40 sont enrichies sur le pull-down mCDB (zone verte,  $FC > 1,5$ ,  $p\text{-value} < 0,05$ ), et 22 sont présentes en quantité plus faible (zone jaune,  $FC < -1,5$ ,  $p\text{-value} < 0,05$ ). Ces différences d'enrichissement pour ces protéines traduisent l'affinité de ces dernières envers la 5mC :

1. Les protéines dont la quantité est plus importante lors du pull-down mCDB (liste verte) correspondent à des protéines affines pour la 5mC. Lors des pull-down, ces protéines sont liées plus fortement à l'ADN méthylé (mCDB) qu'à celui non méthylé (CDB), et

résistent donc mieux aux étapes de lavages, elles se trouvent alors enrichies. Il est possible que ces protéines enrichies ne soient pas directement liées à l'ADN mais fassent partie d'un complexe qui interagit avec ce dernier.

2. Les protéines dont la quantité est diminuée dans le pull-down effectué avec mCDB comparativement au pull-down effectué avec CDB (liste jaune) correspondent à des protéines moins affines pour la 5mC que pour la C.
3. La diminution de la quantité d'une protéine lors du pull-down mCDB pourrait aussi traduire le caractère transitoire d'une protéine affine pour la 5mC. En effet, lors du pull-down CDB, cette protéine ne reconnaît pas de 5mC et peut rester « parcourir » l'ADN dans le but d'en trouver. Dans ce cas, la protéine « transitoire » reste liée à l'ADN, ce qui permet son enrichissement. *A contrario*, lors du pull-down mCDB, la protéine « transitoire » repère plus rapidement la 5mC. Une fois son rôle terminé, on peut par exemple citer le rôle d'une enzyme de réparation de l'ADN, cette protéine se détache de l'ADN. Le temps de présence de cette protéine sur le déterminant d'affinité est donc plus court, induisant un enrichissement moindre.
4. Il est aussi possible que la diminution de la quantité d'une protéine soit due à la différence d'affinité avec une autre protéine affine pour la 5mC qui lui empêcherait d'avoir accès à l'ADN. La diminution serait dans ce cas-là due à une compétition d'affinité entre protéines.

Au sein de ces 62 protéines, 14 ont déjà été identifiées comme des protéines interagissant avec des modifications épigénétique de la cytosine par Spruijt *et al*<sup>99</sup>, et Iurlaro<sup>126</sup> *et al* et sont présentées sur la Figure 104 par des hachures. Les autres protéines pourraient correspondre à l'identification de nouvelles protéines interagissant avec la 5mC, ce qu'il faudra confirmer dans la suite de ce travail. On observe que l'expérience de pull-down avec la 5mC permet l'enrichissement de protéines telles que UHRF1 ou MeCP2 qui ont déjà été décrites comme étant sélectives de la 5mC (introduction p.19). L'enrichissement de ces 2 protéines nous conforte sur la pertinence de notre système et renforce la validité de nos résultats et de leur interprétation. Il est intéressant de noter que la protéine NEIL1 est moins enrichie pour la 5mC que pour la C. Pour rappel, dans l'étude de Spruijt *et al*, elle a été caractérisée comme sélective des formes oxydées de la 5mC, notamment la 5hmC<sup>99</sup>.

### 6.4.2.2 Enrichissement des protéines affines pour la 5hmC

De la même façon que pour les protéines enrichies en présence de la 5mC, les protéines interagissant sélectivement avec la 5hmC ont été identifiées par comparaison de leur enrichissement sur la phase chromatographique HDB par rapport à la phase CDB (Figure 105).

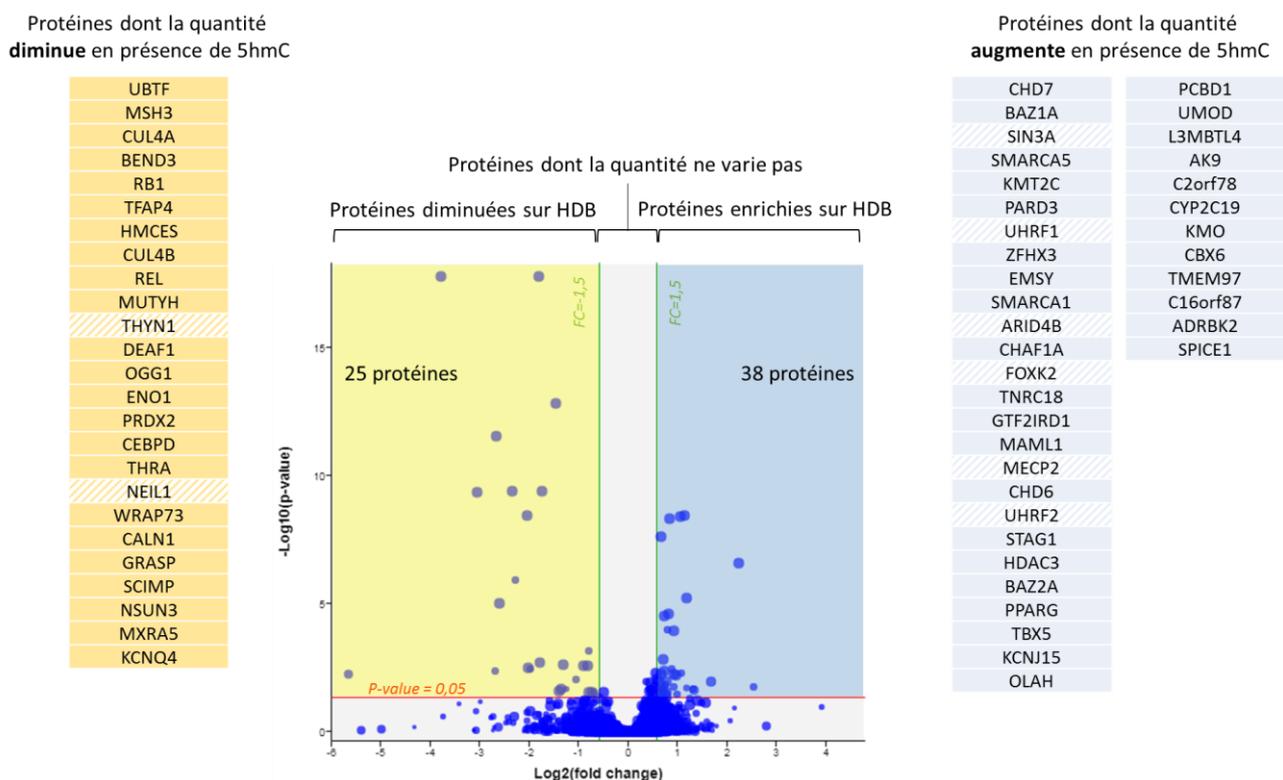


Figure 105 : Volcano plot représentant la quantification des protéines enrichies par le pull down HDB par rapport aux protéines enrichies par le pull-down CBD. La liste des protéines enrichies par la phase HB ou CDB sont représentées sur chaque côté du graphique. Les protéines qui ont déjà été caractérisées comme étant affines pour la 5mC et/ou la 5hmC ont été hachurées sur les listes.

Lors de cette comparaison, 63 protéines, sur 5980 protéines détectées, sont enrichies différemment lors des deux pull-down. 38 sont présentes en plus grande quantité sur le pull-down HDB, et correspondent à des protéines affines de la 5hmC. Cela est d'ailleurs confirmé par la présence de UHRF2, MeCP2 ou encore UHRF1 parmi ces protéines (liste bleu). Nous avons déjà décrit le rôle de ces protéines dans l'introduction, page 19. Par ailleurs, on observe aussi 25 protéines dont la quantité est diminuée en présence de 5hmC. Ces protéines correspondent soit à des protéines moins affines pour cette modification soit à des protéines transitoires comme précédemment expliqué. Cette dernière hypothèse est encouragée par la présence des protéines NEIL1, THYN1 et HMCES (« *Embryonic stem cell-specific 5-hydroxymethylcytosine-binding protein* ») qui sont décrites dans la littérature comme étant sélectives de la 5hmC<sup>99,301</sup>.

### 6.4.2.3 Identification des protéines sélectives de la 5mC et de la 5hmC

La comparaison des pull-down mCDB et HDB par rapport au pull-down CDB a permis de dégager une liste de 56 protéines (présentée en Annexe 3, p.194) qui sont soit enrichies en présence de 5mC (par comparaison à la C, 18), soit enrichies en présence de 5hmC (par comparaison à la C, 16) ou bien affines pour les deux (22 protéines). Pour identifier la spécificité de ces protéines envers la 5mC ou la 5hmC, la comparaison des pull-down HDB et mCDB a été réalisée. Cette comparaison est réduite aux 56 protéines pour éviter les faux positifs qui correspondraient aux protéines ayant moins d'affinité pour la 5mC ou la 5hmC **par rapport à la C**. Par exemple : une protéine X peu affine pour la 5hmC est présente en faible quantité lors du pull-down HDB, alors qu'elle est présente dans des quantités « normales » lors des pull-down CDB et mCDB. Lors de la comparaison du pull-down HDB par rapport au pull-down mCDB, X va paraître diminué sur le pull-down HDB. On pourrait en déduire que X est enrichie en présence de 5mC par rapport à la séquence 5hmC alors qu'en réalité, il s'agit d'une protéine peu affine pour la 5hmC que l'on retrouve associée aussi bien en présence de 5mC que de C. L'utilisation de la liste de 56 protéines affines pour la 5mC et pour la 5hmC par rapport à C permet de distinguer les protéines plus affines pour la 5mC (par exemple) par rapport à la 5hmC et par rapport à la C. Le résultat de la comparaison des pull-down HDB et mCDB est présenté sur la Figure 106.

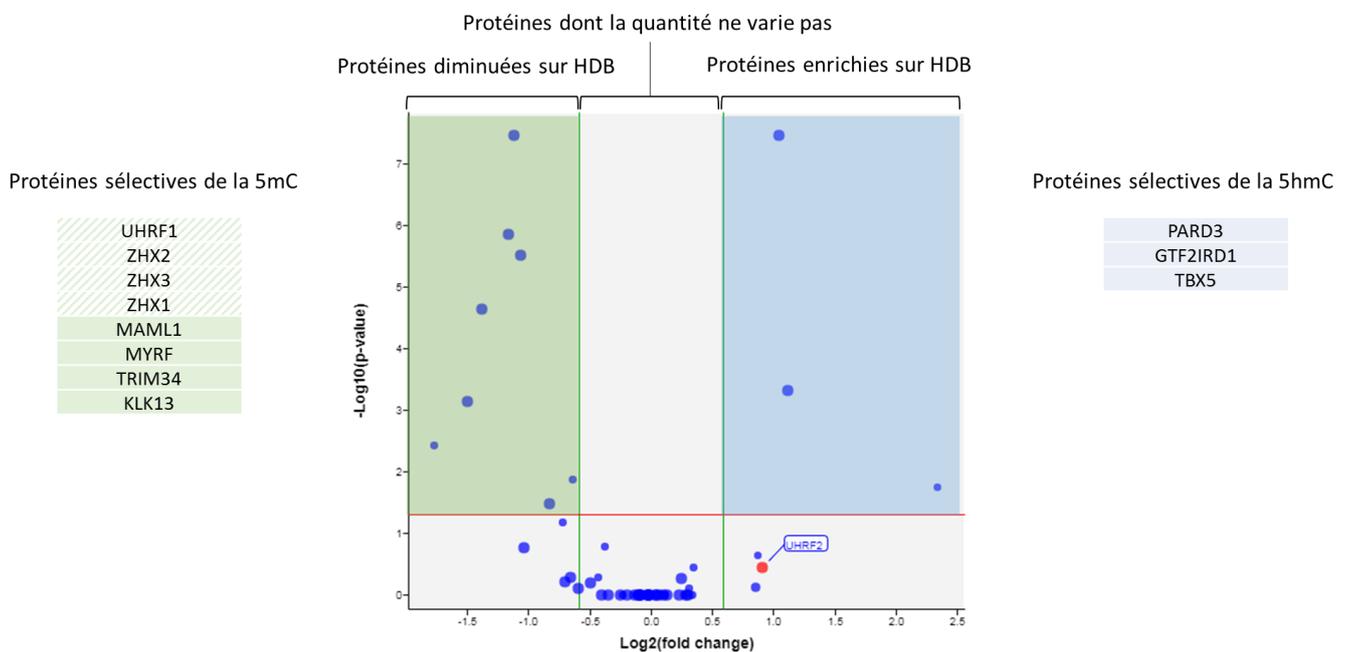


Figure 106 : Volcano plot représentant la quantification des 56 protéines affines pour la 5mC et pour la 5hmC par rapport à C, par le pull-down HDB par rapport pull-down mCDB. La liste des protéines enrichies ou diminuées par la phase HDB par rapport à la phase mCDB sont représentées sur chaque côté du graphique. Les protéines qui ont déjà été caractérisées comme étant affines pour des modifications épigénétiques de la cytosine ont été hachurées sur les listes. .

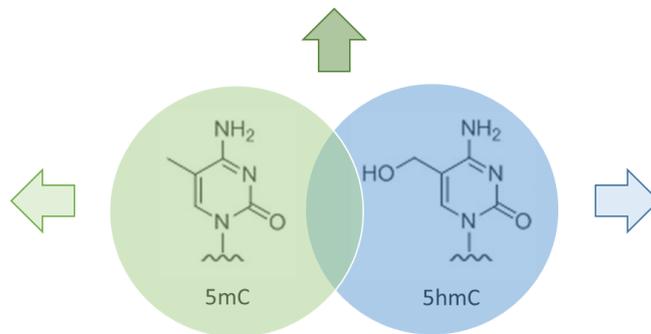
Parmi les 56 protéines comparées, seulement 11 présentent des quantités différentes sur les deux pull-down dont 3 sont enrichies sur HDB et 8 sur mCDB. Cela signifie que les 45 autres protéines de la liste sont affines pour ces deux modifications épigénétiques sans spécificité de la 5mC par rapport à la 5hmC et inversement. Les 8 protéines enrichies sur mCDB correspondent à des protéines sélectives de la 5mC et les 3 enrichies sur HDB sont sélectives de la 5hmC. Ces résultats sont présentés sur la Figure 107 et seront discutés ultérieurement, à la fin de cette partie.

45 Protéines affines de la 5mC et de la 5hmC

BAZ1A	MECP2	
SIN3A	UHRF2	
SMARCA5	STAG1	
KMT2C	UMOD	
ZFX3	KCNJ15	
SMARCA1	L3MBTL4	
ARID4B	AK9	
CHAF1A	HDAC3	
FOXK2	CBX6	
OLA1	TMEM97	PPARG
FOXK1	C16orf87	EMSY
PDS5B	ADRBK2	TNRC18
ARID4A	SPICE1	HNF1B
FOXP1	C2orf78	TLK1
FOXC1	CYP2C19	ASXL1
CDH15	KMO	FOXO3
CHD6	PCBD1	LCORL
CHD7	BAZ2A	TG

8 Protéines sélectives de la 5mC

UHRF1
ZHX2
ZHX3
ZHX1
MAML1
MYRF
TRIM34
KLK13



3 Protéines sélectives de la 5hmC

PARD3
GTF2IRD1
TBX5

Figure 107 : Représentation schématique des protéines spécifiques de la 5mC et de la 5hmC dont la quantité est enrichie en présence de ces modifications.

L'analyse protéomique de nos différents pull-down nous a permis d'identifier 92 protéines, (56 protéines dont la quantité est enrichie en présence de la 5mC et/ou de la 5hmC et 36 dont la quantité diminue (listes jaunes sur la Figure 104 et la Figure 105)) qui interagissent avec la 5mC et/ou la 5hmC (liste en Annexe 3). Bien que ce nombre paraisse important, il est relativement faible (~1%) comparé aux plus de 7000 protéines identifiées au cours de cette analyse. Dans la suite de notre étude, nous allons évaluer l'apport du photomarquage dans cette analyse protéomique.

#### 6.4.3 *Le pull-down couplé au photomarquage pour identifier les protéines affines de la 5mC et de la 5hmC*

Le pull-down couplé au photomarquage pour identifier les protéines affines de la 5mC et de la 5hmC se base sur la comparaison des pull-down avec et sans irradiation. Cette identification peut s'effectuer de deux façons :

- Par détermination du taux d'enrichissement après irradiation des protéines de la liste de protéines obtenues lors de l'analyse protéomique du pull-down classique (liste des 92 protéines). Cette comparaison permet d'évaluer l'apport du photomarquage au sein de notre analyse.
- Par comparaison des résultats de l'analyse protéomique des deux pull-down avec et sans photomarquage. Les listes de protéines obtenues sont comparées, permettant ainsi de déterminer des protéines sélectivement enrichies dans l'une ou l'autre des analyses.

Nous avons choisi de nous focaliser sur la première approche car elle permet d'observer l'apport du photomarquage par rapport à un pull-down classique. La deuxième approche ne sera pas décrite, mais les conclusions seront présentées. Avant de réaliser ces approches, nous avons réalisé une analyse contrôle pour nous assurer que l'agent de photomarquage,  $4^S\text{T}$ , ne perturbait pas notre système.

Enrichissement de protéines affines pour l'ADN, non spécifiques pour la  $4^S\text{T}$  incorporation de la base modifiée  $4^S\text{T}$  au sein des déterminants d'affinité peut engendrer l'enrichissement de protéines affines de cette base. De même, la réalisation d'une irradiation peut modifier la composition des protéines identifiées au sein d'un pull-down. Afin de nous assurer qu'aucune protéine n'est spécifique de la  $4^S\text{T}$ , ni n'est enrichie par irradiation, nous avons réalisé la quantification des protéines enrichies par la phase C4DB après irradiation par rapport à celles enrichies par la phase CDB (Figure 108).

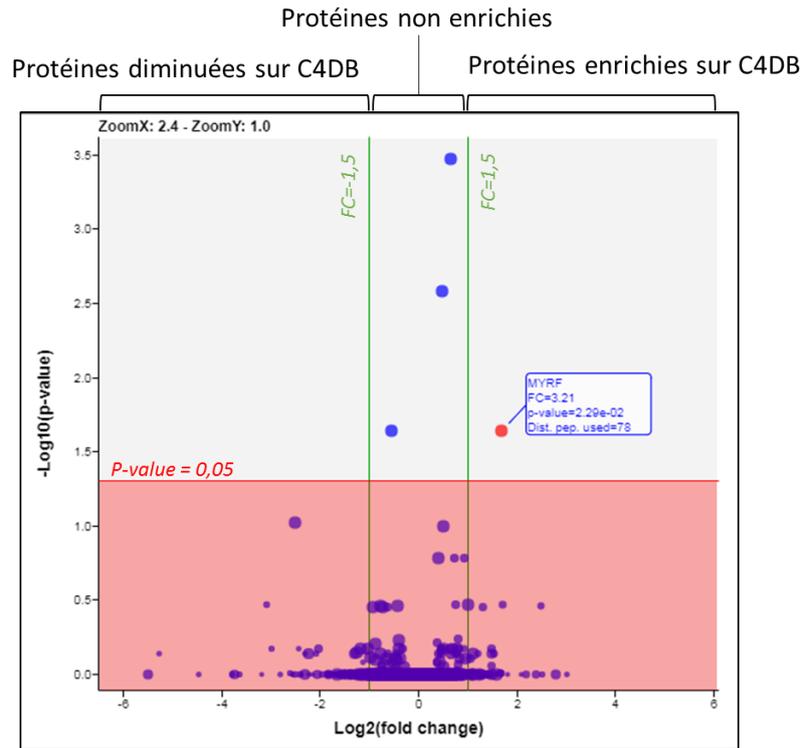


Figure 108 : Volcano plot représentant la quantification des protéines enrichies par la phase chromatographiques C4DB par rapport aux protéines enrichies par la phase chromatographique CDB.

On observe que la majorité des protéines identifiées a une  $p\text{-value} > 0,05$  (zone rouge) et que parmi celles-ci un grand nombre possède une  $p\text{-value} = 1$  ( $-\log(p\text{-value})=0$ ). Cela signifie que la valeur obtenue lors du test statistique, qui permet de valider nos résultats, est proche de l'hypothèse nulle : soit de l'absence d'enrichissement pour toutes les protéines. C'est-à-dire qu'il n'y a pas de variation de quantité pour ces protéines entre les deux pull-down comparés. On peut donc en conclure que les pull-down CDB et C4DB enrichissent les mêmes protéines, et que l'incorporation de  $4^{\text{ST}}$  et l'irradiation n'induisent pas l'enrichissement de protéines qui leurs sont spécifiques à l'exception d'une protéine, MYRF. Il s'agit d'un facteur de transcription qui active spécifiquement l'expression des gènes de myéline<sup>302</sup>. Cette protéine ne sera donc pas pris en compte lors de la comparaison des pull-down possédant des  $4^{\text{ST}}$ .

#### 6.4.3.1 L'apport du photomarquage aux résultats de l'analyse protéomique du pull-down classique

Dans le but d'évaluer l'apport du photomarquage à l'analyse protéomique que nous avons présentée précédemment, nous avons choisi de comparer les pull-down mCDB et HDB à leur analogue possédant la  $4^{\text{ST}}$  (mC4DB et H4DB, respectivement) obtenus après irradiation. Cette comparaison est effectuée uniquement sur les 92 protéines (voir Annexe 3) déterminées comme

interagissant avec la 5mC et/ou la 5hmC. Nous limiter à ces protéines permet d'évaluer directement l'enrichissement de protéines d'intérêt par photomarquage. En principe et dans cette analyse, si une protéine est enrichie lors du pull-down avec photomarquage, celle-ci est sélective de la modification épigénétique présente sur ce dernier.

La comparaison du pull-down mC4DB par rapport à mCDB et du pull-down H4DB par rapport à HDB est présentée sous forme de tableau (Tableau 4), où seules les protéines enrichies par le photomarquage sont représentées.

Protéines	Protéines enrichies sur mC4DB/mCDB	Protéines enrichies par H4DB/HDB
CALN1	↑	↑
CUL4A	-	↑
CUL4B	-	↑
GRASP	-	↑
HMCE5	-	↑
KCNQ4	-	↑
MSH3	-	↑
MUTYH	-	↑
MXRA5	-	↑
NEIL1	↑	↑
NSUN3	↑	↑
OGG1	-	↑
PCBD1	↑	-
PURB	-	↑
SCIMP	↑	↑
THYN1	-	↑
UHRF1	↑	↑
WRAP73	↑	↑

Tableau 4 : Tableau présentant les protéines dont le FC > 1,5 ; p-value < 0,05 lors que la quantification des protéines résultants du pull-down mC4DB par rapport à mCDB et du pull-down H4DB par rapport à HDB sur les 92 protéines obtenues par l'analyse protéomique du pull-down classique.

Le photomarquage a permis d'enrichir 18 protéines sur les 92 comparées. Parmi elles, 11 sont sélectivement enrichies en présence de 5hmC, 6 sont enrichies à la fois en présence de 5mC et de 5hmC et une seule protéine est enrichie en présence de 5mC. Ces protéines sont donc respectivement sélectives de la 5hmC, des deux et de la 5mC. Ces résultats sont présentés sur la Figure 109.

Protéines affines de la 5mC et de la 5hmC

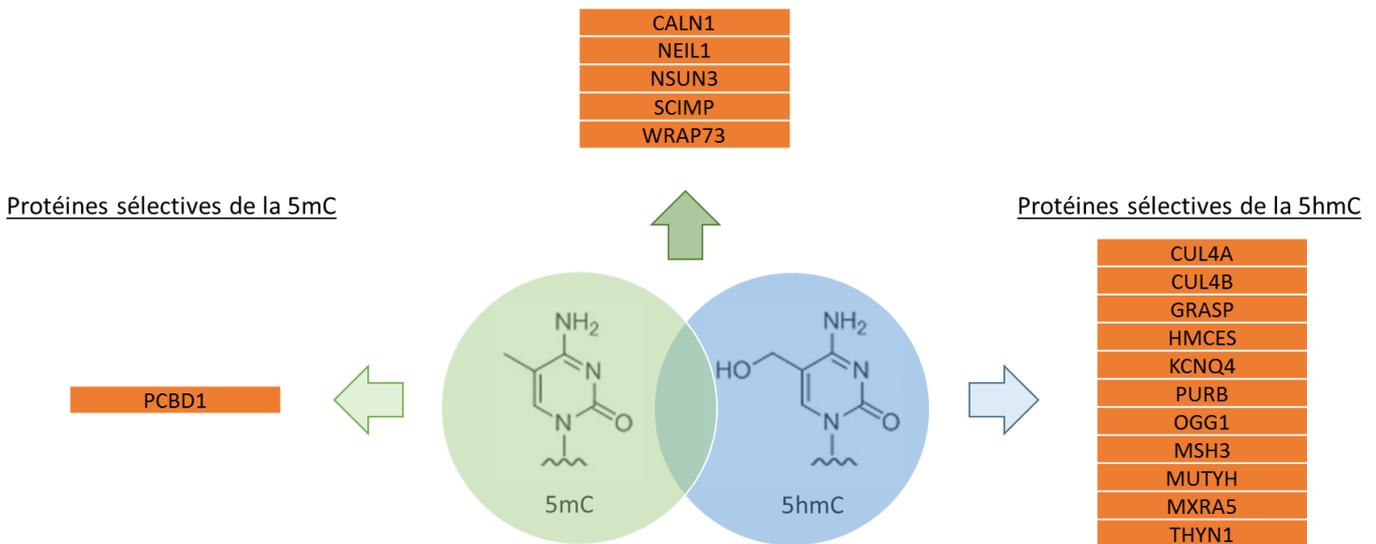


Figure 109 : Représentation schématique des protéines transitoires spécifiques de la 5mC et de la 5mC déterminées par photomarquage liée au pull-down.

Il est intéressant de noter que sur les 18 protéines enrichies, 16 (toutes sauf UHRF1 et PCBD1) sont des protéines dont la quantité diminue en présence de modification épigénétique (en jaune sur la Figure 104 et la Figure 105). Nous émettons l'hypothèse que l'enrichissement de ces protéines lors du photomarquage pourrait traduire le caractère de liaison transitoire de celles-ci. En effet, lorsqu'une protéine se lie de façon transitoire, son temps de présence sur l'ADN va être court, elle est donc difficile à enrichir puisqu'elle est en échange constant avec le milieu. Elle se trouve diluée lors des lavages des phases chromatographiques. Si l'on ponte cette protéine par irradiation, son temps de résidence sur l'ADN est alors infini, cette protéine est facilement enrichie dans les mêmes conditions de lavage des pull-down. Par un raisonnement similaire, on peut déduire que les protéines non transitoires ne sont pas enrichies avec le photomarquage car le temps de résidence de ces protéines sur l'ADN est similaire dans les deux cas de figures. La Figure 110 représente un schéma explicatif de ces phénomènes.

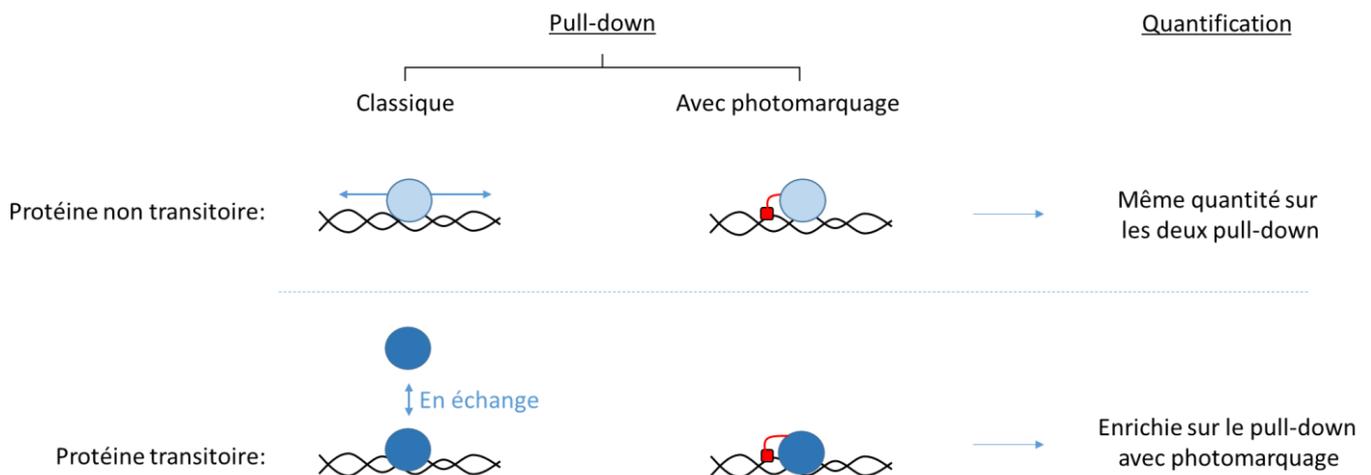


Figure 110 : Schéma explicatif des résultats obtenus en comparant les pull-down avec et sans photomarquage dans le cas de protéines non transitoires et transitoires. Dans le cas du pull-down classique, les protéines non transitoires parcourent l'ADN ou s'y fixent, alors que les protéines transitoires vont être en échange entre le milieu et l'ADN. En effet, elles vont parcourir l'ADN, reconnaître une modification, effectuer leur rôle enzymatique et se détacher de l'ADN. La photomarquage permet de piéger ces deux classes de protéines sur l'ADN.

Ainsi, le système de pull-down couplé au photomarquage permet de révéler les protéines à caractère transitoire et spécifique de l'ADN possédant des modifications épigénétiques. Ce résultat est difficile à confirmer dans la littérature puisque peu d'études s'intéressent à cet aspect (la caractérisation du caractère transitoire des protéines) bien qu'il existe toutefois des techniques développées à cette escient<sup>303</sup>. Toutefois, si l'on observe la fonction de certaines des protéines de cette liste, par exemple NEIL1, on peut justifier nos résultats. NEIL1 est une enzyme impliquée dans la réparation de l'ADN comme nous avons pu le voir dans l'introduction (p. 32). Son rôle est de reconnaître un mésappariement spécifique de l'ADN et de réaliser la coupure de la base mésappariée. Elle va donc reconnaître l'ADN puis une fois sa fonction enzymatique réalisée, se détacher de celui-ci. On peut alors supposer qu'en présence de la base modifiée, cette protéine va avoir un temps de résidence court sur l'ADN et donc qu'elle aura un caractère transitoire, caractère qui est retranscrit par nos résultats.

### Comparaison des résultats de l'analyse protéomique des pull-down classique et avec photomarquage

Nous avons aussi réalisé la comparaison des listes de protéines obtenues par l'analyse protéomique des deux pull-down : classique et avec photomarquage. Cette comparaison n'est pas présentée dans ce manuscrit pour éviter toute confusion avec les analyses précédentes. Elle est par contre résumée dans le paragraphe suivant.

Le photomarquage permet d'augmenter la quantité de la plupart des protéines identifiées comme affines de la 5mC et/de la 5hmC par rapport au pull-down classique, on peut par exemple citer les protéines UHRF1, UHRF2 ou encore les protéines ZHX. Toutefois, on n'identifie pas de nouvelles protéines affines pour la 5mC ou la 5hmC comme nous l'espérons. Ce résultat est à nuancer puisque nous avons utilisé des conditions de lavage, certes dénaturantes, mais pas drastiques lors des pull-down comme nous aurions pu le faire avec notre système avec photomarquage. De nouvelles expériences utilisant des conditions de lavages plus drastiques des phases chromatographiques après irradiation devrait permettre un enrichissement plus sélectif, qui permettrait l'identification de nouvelles protéines spécifiques des modifications épigénétique de l'ADN.

### **Conclusion de l'analyse protéomique**

- 6.6 L'utilisation du pull-down couplé au photomarquage nous a permis d'identifier des protéines partenaires de la 5mC et de la 5hmC. Le bilan de notre analyse protéomique est résumé dans la Figure 111.

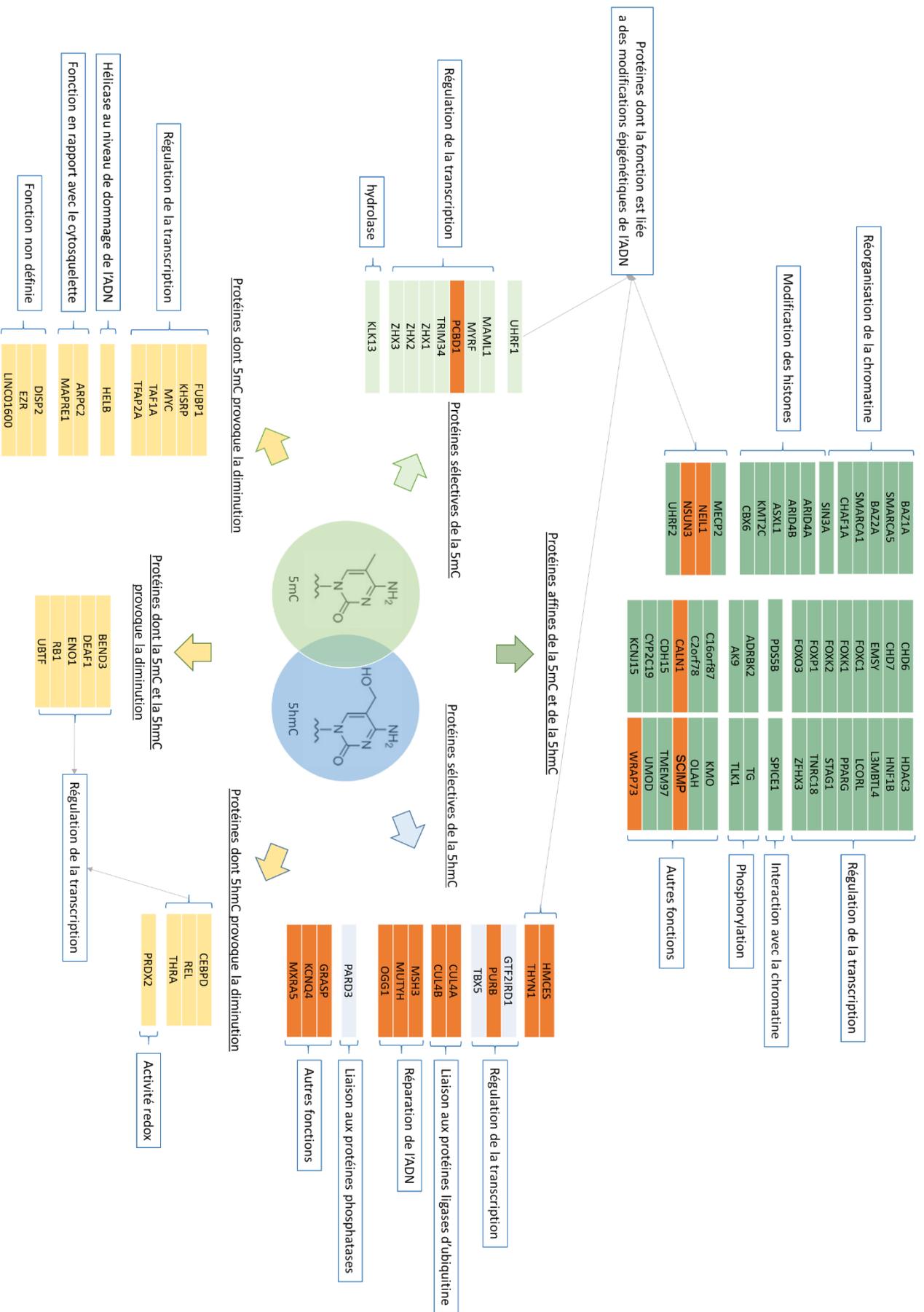


Figure 111 : Schéma bilan de toutes les protéines interagissant avec la 5mC ou la 5hmC. Les protéines surlignées en orange ont été identifiées comme des protéines transitoires. Les fonctions de toutes les protéines ont été répertoriées et sont représentées dans les encadrés bleus.

Pour faciliter la discussion et permettre une meilleure appréhension de ces résultats, les fonctions principales associées aux protéines identifiées comme affines de la 5mC et de la 5hmC ont été représentées. La détermination des fonctions des protéines a été réalisée à l'aide de la Gene Ontology ainsi que de la base de donnée EpiFactors<sup>304</sup>, qui recense les protéines en lien avec la régulation épigénétique.

Notre étude permet de distinguer 6 groupes de protéines : les protéines sélectives de la 5mC, les protéines sélectives de la 5hmC et les protéines affines pour les deux modifications. Parmi tous ces groupes une fonctionnalité ressort, beaucoup de protéines interviennent dans la régulation de la transcription. Cette fonctionnalité est en accord avec le rôle joué par la méthylation de l'ADN, c'est-à-dire réguler l'expression des gènes<sup>305</sup>. D'autres fonctions comme la réorganisation de la chromatine, l'interaction avec la chromatine<sup>306</sup> ou encore la modification des histones<sup>307</sup> sont bien en lien avec la régulation épigénétique. Ces résultats rationnels permettent de valider notre approche.

On observe que parmi les protéines sélectives de la 5hmC, certaines se lient aux ligases d'ubiquitine (CUL4A et CUL4B). Ce résultat est intéressant puisque que l'ubiquitinylation des histones permet la reconnaissance de la 5mC par la protéine UHRF1<sup>46</sup> et ce qui pourrait laisser envisager un mécanisme similaire pour la reconnaissance de la 5hmC par UHRF2. L'étude des protéines CUL4A et B pourrait se révéler intéressante dans la compréhension de l'ubiquitinylation et son rôle dans la reconnaissance des modifications épigénétiques de l'ADN. Parmi les protéines sélectives de la 5hmC, on observe aussi des protéines liées aux mécanismes de réparation de l'ADN, comme OGG1, MUTYH, MSH3 ou encore NEIL1 qui suggère le rôle d'intermédiaire de cette base dans une voie de déméthylation de l'ADN. En effet, OGG1<sup>308</sup> tout comme MUTYH<sup>309</sup> sont caractérisées comme des protéines réparant les dommages oxydatifs de l'ADN.

Pour conclure, le but de cette dernière partie était de développer une expérience pull-down couplée au photomarquage qui permette l'identification de protéines partenaires de modifications de l'ADN. Nous avons testé notre système sur deux modifications, la 5mC et la 5hmC. Dans les deux cas nous avons été en mesure d'identifier des protéines qui leurs sont sélectives, et de caractériser les fonctions de celles-ci. Fonctions qui sont en accord avec le rôle de ces deux modifications épigénétiques. Nous avons aussi montré que le photomarquage pouvait permettre de révéler la nature transitoire de certaines protéines. Ce résultat est

intéressant car il permet la définition à la fois de la sélectivité d'une protéine envers un substrat et du caractère transitoire de cette protéine.

Bien que l'expérience de pull-down couplée au photomarquage permette l'obtention d'un grand nombre d'informations, celui-ci est perfectible. Plusieurs améliorations sont envisageables et donneront lieu à des expériences supplémentaires:

- Lors de ces premières expériences, nous recherchions à comparer l'apport du photomarquage au sein d'une analyse protéomique découlant d'un pull-down. Dans un second temps, il aurait été intéressant de chercher à optimiser les conditions de lavages et notamment d'utiliser des conditions plus drastiques qui nous auraient permis de mieux exploiter toutes les possibilités permises par le photomarquage.
- Lors de notre étude, nous nous sommes limités à 5 % d'agent de photomarquage  $4^S\text{T}$  de façon arbitraire sans avoir eu le temps de tester d'autres pourcentages. Nous avons vu que la  $4^S\text{T}$  ne recrutait pas par elle-même de protéine et il serait intéressant d'en incorporer plus dans la séquence afin d'augmenter les chances de photomarquer des protéines.
- Une autre amélioration envisageable provient du développement de méthodes de plus en plus performantes d'analyse en spectrométrie de masse. On peut notamment citer la méthode SWATH-MS<sup>310</sup> (*Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra*). Contrairement aux méthodes comme le comptage spectral ou le XIC de notre étude qui se base sur un mode de fragmentation DDA (Data Dependant Acquisition), où les fragmentations des ions peptidiques sont déclenchées à partir des informations obtenues sur un scan MS (fragmentation des 20 peptides les plus intenses d'un scan MS par exemple), le SWATH est une méthode d'acquisition DIA (Data Independant Acquisition). C'est-à-dire que la fragmentation des peptides se fait sans sélection préalable du précurseur. Selon cette méthode, des spectres MS/MS sont acquis sur des plages de masse de précurseurs de 25 Da sélectionnées de façon consécutive, sur une gamme de masse 400-1200 m/z par exemple. Ce mode d'acquisition SWATH permet alors l'enregistrement de données MS/MS de n'importe quel peptide détectable, ce qui entraîne une faible perte d'informations sur les peptides liée à la sélection des précurseurs. A partir d'une même expérience, le mode d'acquisition SWATH conduit à l'obtention d'une quantité d'informations plus importante qu'en DDA, et permet donc d'obtenir une très bonne couverture analytique. La quantification par cette méthode se fait par une méthode analogue des XIC, mais

cette fois-directement sur les ions fragments<sup>311</sup>. L'utilisation d'une telle méthode d'analyse en spectrométrie de masse des protéines enrichies sur le pull-down couplé au photomarquage permettrait l'identification des protéines peu exprimées, ou présentes en quantité infime. De plus, nous pourrions améliorer l'identification de ces protéines en incorporant les données que nous avons obtenues lors de l'appréhension de la photochimie de la 4<sup>S</sup>T.

## Conclusion générale et perspective

Mon projet de thèse s'est articulé autour d'une question :

### **Comment identifier des protéines partenaires des bases épigénétiques peu abondantes de l'ADN ?**

Pour répondre à cette question, nous avons développé une méthode basée sur le photomarquage d'affinité couplé à la spectrométrie de masse utilisant des sondes d'ADN photoactivables. Mes travaux se sont déroulés en plusieurs parties:

- 1) L'évaluation de sondes oligonucléotidiques photoactivables obtenues par voie chimique.

Lors de la mise en œuvre d'une approche utilisant le photomarquage, il est courant de tester les propriétés de plusieurs agents photoactivables incorporés dans le déterminant d'affinité afin de déterminer le plus approprié pour l'étude. Au cours de cette partie, nous avons comparé les propriétés de photomarquage de 4 groupements photoactivables : la benzophénone, la diazirine, la 4-thiothymidine et la 5-iodouracile. Cette comparaison a été réalisée dans un contexte simple brin à l'aide d'une protéine modèle hRPA, et dans un contexte double brin à l'aide des protéines modèles DNMT3a et H3. Trois conclusions ressortent de cette étude :

- Les sondes oligonucléotidiques photoactivables permettent le photomarquage de protéines spécifiques dans un milieu protéique complexe dans un contexte simple et double brin d'ADN.
- Les sondes oligonucléotidiques photoactivables simple brin permettent de piéger spécifiquement les SSB et même si ce n'était pas l'objectif premier de ces travaux, nous avons mis au point des outils intéressants pour caractériser des SSB dans des extraits cellulaires de divers organismes.
- Les sondes possédant les agents de photomarquage 4-thiothymidine et diazirine présentent les meilleurs rendements de photomarquage. La diazirine est souvent décrite comme particulièrement efficace et avantageuse dans ce type d'approche. Notre étude a permis de montrer que la 4-thiothymidine n'est pas en reste bien qu'elle soit peu utilisée pour ce type d'étude. Outre les bons rendements de photomarquage, elle est facile à incorporer dans l'ADN par voie chimique ou enzymatique, elle constitue un mime de base ne perturbant pas la

structure de l'ADN, et comme nous avons pu le voir dans la dernière partie, elle ne recrute pas par elle-même de protéines spécifiques. Il s'agit donc d'un outil à ne pas négliger lors de la conception de sondes d'ADN photoactivables.

Ces travaux ouvrent la voie à plusieurs études. La première que l'on peut envisager est la recherche de nouvelles protéines spécifiques de l'ADN simple brin au sein d'organismes connus ou encore peu caractérisés. On peut notamment citer les tardigrades<sup>241</sup> qui sont une thématique de recherche émergente du Muséum Nationale d'Histoire Naturelle. Une autre étude envisageable est l'optimisation des conditions d'enrichissement des complexes ODN-protéines obtenus après photomarquage au sein d'extrait de cellules eucaryotes par l'interaction biotine, streptavidine. La mise au point de ces conditions permettrait une meilleure caractérisation de ces complexes et d'identifier de nouvelles protéines.

## 2) La caractérisation au niveau moléculaire de complexes ODN-protéines par spectrométrie de masse.

Dans cette seconde partie, nous avons essayé de caractériser au niveau moléculaire les complexes ODN-protéines, formés par les sondes photoactivables, par spectrométrie de masse dans le but d'obtenir à terme plus d'informations sur le site d'interaction de ces protéines vis-à-vis de l'ODN. Bien que nous n'ayons pas réussi à remplir cet objectif, cette étude nous a permis :

- De développer un protocole de cristallisation de la matrice 3-HPA en MALDI qui permet une analyse reproductible d'un oligonucléotide de longue taille et d'oligonucléotide lié de façon covalente à un peptide.
- D'appréhender la photochimie de la 4-thiothymine envers les acides aminés, chose qui n'a jamais été décrit dans la littérature. La compréhension de ce mécanisme photochimique permet l'implémentation des modifications propres à la 4<sup>S</sup>T lors de l'identification de protéines liées de façon covalente à celle-ci par recherche en banque de données. A terme, Cela aidera à caractériser des protéines photomarquées par ce type de sonde.

L'étude des complexes ODN-protéines par spectrométrie peut être compléter par l'utilisation d'un système d'ionisation ESI couplé à une analyse en tandem (MS/MS) de l'échantillon. Cette analyse permettra d'identifier le site de la liaison covalente établie lors du photomarquage. Une

caractérisation de la photochimie de la  $4^S\text{T}$  est aussi une perspective en cours d'élaboration au laboratoire des Biomolécules. Pour réaliser cette étude, l'interaction de la  $4^S\text{T}$ , non incorporée dans un oligonucléotide, avec tous les acides aminés, sera étudiée par spectrométrie de masse et par RMN avec ou sans irradiation.

### 3) L'identification des protéines affines pour la 5mC et de la 5hmC par un pull-down couplé au photomarquage

L'identification des protéines affines pour la 5mC et de la 5hmC a été réalisée en deux temps. Dans un premier temps, nous avons tenté de photomarquer ces protéines à l'aide des sondes photoactivables développées dans la première partie de mes travaux. Malheureusement, bien que nous arrivions à observer des photomarquages, ces expériences étaient limitées par les quantités de matière de sondes à disposition et par une méthode d'enrichissement de complexe obtenu à l'aide de l'interaction biotine/streptavidine qui n'était pas optimale. Dans un second temps, nous avons fait le choix de changer de système d'enrichissement de protéines en utilisant un système de pull-down déjà en place au sein de notre laboratoire au MNHN. Nous avons agrémenté ce système par l'agent de photomarquage  $4^S\text{T}$ . L'analyse protéomique des pull-down réalisés avec les sondes contenant de la 5mC et de la 5hmC sans ou avec photomarquage nous a permis d'établir les conclusions suivantes :

- Le pull-down classique permet d'enrichir des protéines déjà connues comme étant sélectives de la 5mC et de la 5hmC. Il a aussi permis l'identification de protéines sélectives comme des facteurs de transcription.
- Le photomarquage couplé au pull-down permet d'augmenter la quantité de protéines enrichies.
- Le photomarquage couplé au pull-down permet de révéler le caractère transitoire de certaines protéines affines de la 5mC et de la 5hmC.

L'une des perspectives de cette étude serait la validation du rôle de certaines protéines que nous avons identifiées comme affines soit sur la 5mC soit sur le 5hmC. Par exemple, la protéine HMCES est une protéine dont la fonction n'a jamais été validée vis-à-vis de la 5hmC. En effet, la fonction de HMCES lui a été attribuée par similarité avec des protéines homologues dans d'autres organismes<sup>301</sup>. Cette validation pourrait être effectuée par des études de « knock-down » ou encore des tests enzymatiques *in vitro*. Une autre perspective envisageable de ce projet est l'amélioration du pull-down couplé au photomarquage. Deux axes principaux sont à

envisager comme nous l'avons déjà discuté précédemment, notamment pour répondre à la question posée au début de cette conclusion générale: (1) l'optimisation de l'enrichissement des protéines partenaires par le pull-down couplé au photomarquage et (2) la comparaison de ces enrichissements par des méthodes de quantification plus performantes en spectrométrie de masse comme le SWATH-MS. Parmi les autres perspectives de ce travail, le système de pull-down couplé au photomarquage, que nous avons développé permettra à terme d'effectuer des études sur d'autres bases modifiées de l'ADN comme la 5fC, la 5caC, la base J des trypanosomes<sup>75</sup> ou d'autres modifications représentées sur la Figure 1 de ce manuscrit. En effet, notre système est bien en place et modulable puisque des bases modifiées de l'ADN peuvent y être introduites enzymatiquement, à condition de disposer des nucléotides triphosphates correspondants. On peut aussi envisager d'incorporer des séquences courtes d'ADN obtenues chimiquement, par ligation enzymatique dans une molécule d'ADN plus longue. Cette dernière méthode permettrait ainsi l'incorporation des oligonucléotides 30-mères que nous avons synthétisés au cours de la première partie de ces travaux, en vue de les utiliser dans un pull-down « boucle ».

## Annexe

A

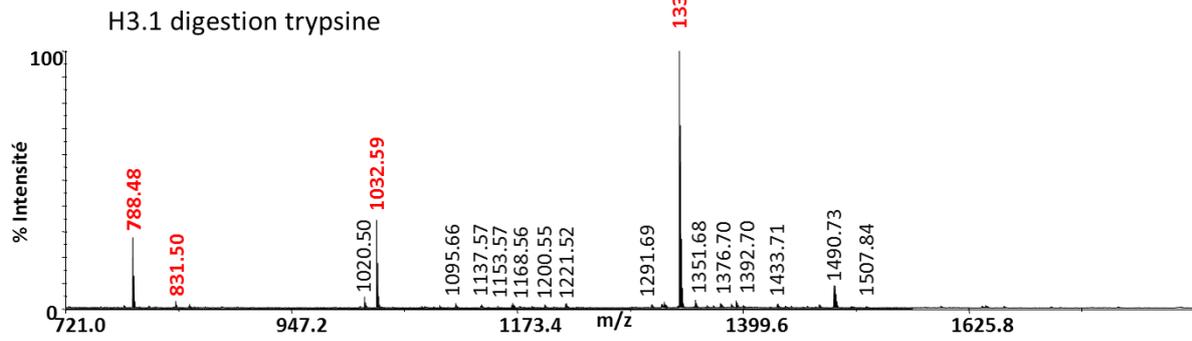
Position of cleavage site	Name of cleaving enzyme(s)	Resulting peptide sequence (see explanations)	Peptide length [aa]	Peptide mass [Da]
3	Trypsin	MAR	3	376.474
5	Trypsin	TK	2	247.294
9	Trypsin	QTAR	4	474.517
10	Trypsin	K	1	146.189
15	Trypsin	STGGK	5	448.476
18	Trypsin	APR	3	342.398
19	Trypsin	K	1	146.189
24	Trypsin	QLATK	5	559.663
27	Trypsin	AAR	3	316.360
28	Trypsin	K	1	146.189
37	Trypsin	SAPATGGVK	9	786.883
41	Trypsin	KPHR	4	536.635
50	Trypsin	YRPGTVALR	9	1032.211
53	Trypsin	EIR	3	416.478
54	Trypsin	R	1	174.203
57	Trypsin	YQK	3	437.496
64	Trypsin	STELLIR	7	830.980
65	Trypsin	K	1	146.189
70	Trypsin	LPPQR	5	659.786
73	Trypsin	LVR	3	386.495
80	Trypsin	EIAQDFK	7	849.939
84	Trypsin	TDLR	4	503.556
116	Trypsin	FQSSAVMALQEACEAYLVGLFEDTNLCAIHA	32	3473.979
117	Trypsin	R	1	174.203
123	Trypsin	VTIMPK	6	687.896
129	Trypsin	DIQLAR	6	714.820
130	Trypsin	R	1	174.203
132	Trypsin	IR	2	287.362
135	Trypsin	GER	3	360.370
136	<b>end of sequence</b>	A	1	89.094

B

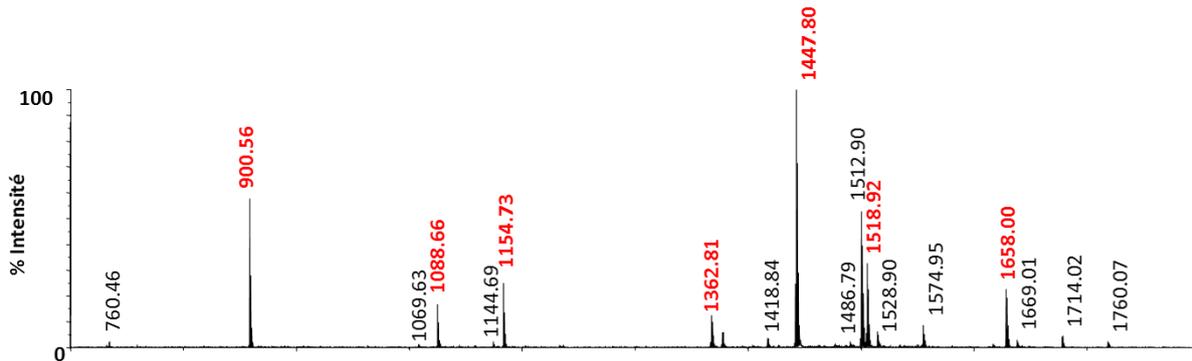
Position of cleavage site	Name of cleaving enzyme(s)	Resulting peptide sequence (see explanations)	Peptide length [aa]	Peptide mass [Da]
3	Arg-C proteinase	MAR	3	376.474
9	Arg-C proteinase	TKQTAR	6	703.797
18	Arg-C proteinase	KSTGGKAPR	9	901.034
27	Arg-C proteinase	KQLATKAAR	9	986.183
41	Arg-C proteinase	KSAPATGGVKKPHR	14	1433.677
43	Arg-C proteinase	YR	2	337.379
50	Arg-C proteinase	PGTVALR	7	712.847
53	Arg-C proteinase	EIR	3	416.478
54	Arg-C proteinase	R	1	174.203
64	Arg-C proteinase	YQKSTELLIR	10	1250.461
70	Arg-C proteinase	KLPPQR	6	787.960
73	Arg-C proteinase	LVR	3	386.495
84	Arg-C proteinase	EIAQDFKTDLR	11	1335.480
117	Arg-C proteinase	FQSSAVMALQEACEAYLVGLFEDTNLCAIHA	33	3630.166
129	Arg-C proteinase	VTIMPKDIQLAR	12	1384.700
130	Arg-C proteinase	R	1	174.203
132	Arg-C proteinase	IR	2	287.362
135	Arg-C proteinase	GER	3	360.370
136	<b>end of sequence</b>	A	1	89.094

Annexe 1 : A) tableau représentant tous les fragments peptidiques obtenus après la digestion trypsique de la protéine H3,1. B) tableau représentant tous les fragments peptidiques obtenus après la digestion à l'Arg-C.

A) digestion trypsique de H3.1



B) propionylation et digestion trypsique de H3.1



Annexe 2 : Spectre de masse obtenu par une analyse MALDI-TOF, mode réflecteur, mode positif, matrice HCCA des fragments peptidiques après digestion trypsique (A) et après le protocole de propionylation couplé à une digestion trypsique (B) de la protéine H3.1. Les masses en rouges correspondent à des masses identifiées comme appartenant à la protéine H3.1. On observe qu'on observe plus de pic correspondant à H3 avec l'utilisation du protocole de propionylation ce qui permet une meilleure identification générale de cette protéine.

Liste des 56 protéines dont la quantité augmente en présence de 5mC et/ou 5hmC
ADRBK2
AK9
ARID4A
ARID4B
ASXL1
BAZ1A
BAZ2A
C16orf87
C2orf78
CBX6
CDH15
CHAF1A
CHD6
CHD7
CYP2C19
EMSY
FOXC1
FO XK1
FO XK2
FOXO3
FOXP1
GTF2IRD1
HDAC3
HNF1B
KCNJ15
KLK13
KMO
KMT2C
L3MBTL4
LCORL
MAML1
MECP2
MYRF
OLAH
PARD3
PCBD1
PDS5B
PPARG
SIN3A
SMARCA1
SMARCA5
SPICE1
STAG1
TBX5
TG
TLK1
TMEM97
TNRC18
TRIM34
UHRF1
UHRF2
UMOD
ZFHX3
ZHX1
ZHX2
ZHX3

Liste des 36 protéines dont la quantité diminue en présence de 5mC et/ou 5hmC
ARPC2
BEND3
CALN1
CEBPD
CUL4A
CUL4B
DEAF1
DISP2
ENO1
EZR
FUBP1
GRASP
HELB
HMCES
KCNQ4
KHSRP
LINC01600
MAPRE1
MSH3
MUTYH
MXRA5
MYC
NEIL1
NSUN3
OGG1
PRDX2
PURB
RB1
REL
SCIMP
TAF1A
TFAP2A
THRA
THYN1
UBTF
WRAP73

Annexe 3 : Listes des 92 protéines dont la quantité varie en présence de 5mC et/ou de la 5hmC

## Matériels et méthodes

**Materials.** Solvents and most of the starting materials were purchased from Sigma Aldrich, TCI and Carlo Erba. All solvents were of analytical grade. Reagents and solvents for oligonucleotides synthesis, as well as dT, Ac-dC, Pac-dA, and dmf-dG phosphoramidites were purchased from Link Technologies. 2-FdI, and 3'-Protected BiotinLC Serinol CPG were purchased from GlenResearch. Unmodified oligonucleotides (ODNs) and 4<sup>S</sup>T containing ODNs were purchased from Eurogentec (Belgium). [ $\gamma$ -<sup>32</sup>P]ATP (3000 Ci/mmol) was purchased from Perkin Elmer and Micro Bio-Spin® 6 columns from Biorad. The magnetic streptavidin-coated beads were from Roche. The genomic DNA used as PCR template was obtained from Gabriel Leberre from HCT 116 cells. The PCR primers endowed with a biotin or desthiobiotin moiety were from Eurogentec. Recombinant hRPA was expressed in *Escherichia coli* BL21 (DE3) strain transformed with the pET11a-hRPA plasmid that permits the co-expression of hRPA1, hRPA2, and hRPA3. hRPA was purified as described previously in <sup>312</sup>. The C-terminal catalytic domain of the murine DMT3a (623-908) was obtained as previously described in <sup>313</sup>. Histone H3.1 human was purchased from New England BioLabs®. Hela nuclear extracts were obtained from Ipracell (Belgium). Nuclear extracts from rat brain were prepared using the CellLytic NuCLEAR Extraction Kit (Sigma). Mouse anti-RPA1 monoclonal antibody was from Millipore. Mouse anti-RPA2 and anti-RPA3 monoclonal antibody was from Novus Biologicals. Trypsin used for digestion was of Gold mass spectrometry grade from Promega, Madison, Wisconsin, USA. Microcolumns used for desalting of peptides were from Applied Biosystems.

**Methods.** <sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded on a Bruker Avance III spectrometer, operating at a <sup>1</sup>H frequency of 500 MHz. Chemical shifts are given in parts per million (ppm) using solvent residual signals as the reference and coupling constants (*J*) are given in Hertz (Hz). All yields reported are unoptimized. Flash chromatography was performed using silica gel Merck 60 with 0.040–0.063  $\mu$ m. Oligonucleotides HPLC purifications were performed on a Waters HPLC apparatus (Waters™ 1525 Binary HPLC Pump with Waters™ 2487 Dual  $\lambda$  Absorbance Detector) using a Xterra Prep RP18 column (7.8\*300 mm, 10  $\mu$ m) and purity was checked using a Waters Xterra MS C18 column (4.6\*250 mm, 5  $\mu$ m). UV spectra were recorded on a Nanodrop-2000c spectrophotometer (Thermo Scientist). High Resolution Mass Spectra (HRMS) were recorded at the Institut Parisien de Chimie Moléculaire (FR 2769) of UPMC (electrospray source). Radiolabelled oligonucleotides, non-crosslinked ODN:protein and

crosslinked ODN-protein complexes were visualized on an Imager Typhoon (GE Healthcare). MALDI-TOF spectra of oligonucleotides and MALDI-TOF/TOF spectra of peptides were recorded at the Institut de Biologie Paris-Seine (Paris, France). NanoLC-MS spectra of proteins and quantification were recorded at the Institut Curie (Paris, France).

## Partie expérimentale du chapitre I

### Chemistry

**[(4-Benzoyl-benzoylamino)hexyl]carbamic acid *tert*-butyl ester.** To a stirred solution of 4-benzoylbenzoic acid (2.2 mmol, 500 mg) in DMF (10 mL) were added DIPEA (1.1 equiv), and HATU (1.1 equiv). The mixture was stirred for 20 min and a solution of *tert*-butyl-*N*-(6-aminohexyl)carbamate (2.2 mol, 556 mg) previously dissolved in DMF (5 mL) in the presence of DIPEA (1.2 equiv) was added to the ester. The resulting mixture was stirred at room temperature in the dark overnight. The solvents were removed under reduced pressure and the crude residue was purified by flash chromatography (CH<sub>2</sub>Cl<sub>2</sub>/MeOH) to give 1 as a yellow solid (825 mg, 88%). <sup>1</sup>H-NMR (CDCl<sub>3</sub>): δ (ppm) 7.91 (dd, *J* = 1.9, *J* = 6.5 Hz, 2H); 7.82 (m, 2H); 7.78 (m, 2H); 7.60 (m, 2H); 7.48 (m, 1H); 6.66 (br, 1H); 3.47 (dd, *J* = 7.0, *J* = 12.7 Hz, 2H); 3.12 (t, *J* = 7.0 Hz, 2H); 1.64 (m, 2H); 1.49 (m, 2H); 1.43 (m, 2H); 1.42 (s, 9H); 1.37 (m, 2H); <sup>13</sup>C-NMR (CDCl<sub>3</sub>): δ (ppm) 196.1; 166.7; 156.4; 140.0; 138.1; 137.1; 132.9; 130.1; 128.5; 127.0; 40.1; 39.7; 30.1; 29.4; 28.5; 26.0; 25.8. HRMS calcd. for C<sub>25</sub>H<sub>32</sub>N<sub>2</sub>O<sub>4</sub>Na [M+Na]<sup>+</sup> at *m/z*: 447.2254, found: 447.2257

***N*-Aminohexyl-4-benzoyl-benzamide.** To a solution of 1 (800 mg, 1.88 mmol) in 10 mL of chloroform was added 4 mL of trifluoroacetic acid. The mixture was stirred at room temperature overnight in the dark. The solution was concentrated under reduced pressure and the crude residue was partitioned between 1 M NaOH and ethyl acetate. The aqueous layer was extracted 3 times with ethyl acetate. The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub> and then concentrated. The residue was purified by flash chromatography (CHCl<sub>3</sub>/MeOH) to give 2 as a white solid (580 mg, 95 %). *R*<sub>f</sub>(CH<sub>2</sub>Cl<sub>2</sub>/MeOH: 9/1) 0.21. <sup>1</sup>H-NMR (DMSO-d<sub>6</sub>): δ (ppm) 8.70 (t, *J* = 5.5 Hz, 1H); 7.99 (dd, *J* = 5.5, *J* = 6.5 Hz, 2H); 7.83 (br, 1H); 7.79 (dd, *J* = 2.0, *J* = 6.6 Hz, 2H); 7.75 (m, 2H); 7.70 (m, 1H); 7.58 (m, 2H); 3.28 (dd, *J* = 7, *J* = 12.7 Hz, 2H); 2.78 (m, 2H); 1.55 (m, 4H); 1.34 (m, 4H); <sup>13</sup>C-NMR (DMSO-d<sub>6</sub>): δ (ppm) 195.4; 165.4; 139.0; 138.0; 136.7; 133.0; 129.7; 129.5; 128.7; 127.3; 39.2; 38.7; 28.9; 27.0; 26.0; 25.5. HRMS calcd for C<sub>20</sub>H<sub>25</sub>N<sub>2</sub>O<sub>2</sub> [M+H]<sup>+</sup> at *m/z* 325.1910, found: 325.1911

***N*-[3-[[4-[3-(trifluoromethyl)-3*H*-diazirin-3-yl]benzoyl]amino]hexylcarbamic acid *tert*-butyl ester.** To a stirred solution of 4-[3-(trifluoromethyl)-3*H*-diazirine-3-yl]benzoic acid (2.17 mmol, 500 mg) in DMF (12 mL) were added DIPEA (1.2 equiv), and HATU (1.2 equiv). The mixture was stirred for 20 min and a solution of *tert*-butyl-*N*-(6-aminohexyl)carbamate previously dissolved in DMF (3 mL) in the presence of DIPEA (1.2 equiv) was added to the ester. The resulting mixture was stirred at room temperature in the dark overnight. The solvents were removed under reduced pressure and the crude residue was purified by flash chromatography (CH<sub>2</sub>Cl<sub>2</sub>/MeOH) to give 3 as a yellow solid (740 mg, 80 %). <sup>1</sup>H-NMR (CDCl<sub>3</sub>): δ (ppm) 7.85 (dd, *J* = 8.60 Hz, 2H); 7.23 (d, *J* = 8.25 Hz, 2H); 6.67 (br s, 1H); 3.43 (m, 2H); 3.11 (t, 2H); 1.61 (m, 2H); 1.48 (m, 2H); 1.42 (s, 9H); 1.36 (m, 4H); <sup>13</sup>C-NMR (CDCl<sub>3</sub>): δ (ppm) 166.3; 156.3; 135.9; 132.1; 127.5; 126.6; 123.0; 79.3; 40.0; 39.6; 30.1; 29.3; 28.5; 28.4; 25.9; 25.7. HRMS calcd. for C<sub>20</sub>H<sub>27</sub>F<sub>3</sub>N<sub>4</sub>NaO<sub>3</sub> [M+Na]<sup>+</sup> at *m/z* 451.1927, found: 451.1931

***N*-(3-aminohexyl)-4-[3-(trifluoromethyl)-3*H*-diazirin-3-yl]benzamide.** To a solution of 3 (700 mg, 1.6 mmol) in 10 mL of chloroform was added 4 mL of trifluoroacetic acid. The mixture was stirred at room temperature overnight in the dark. The solution was concentrated under reduced pressure and the crude residue was partitioned between 1 M NaOH and ethyl acetate. The aqueous layer was extracted 3 times with ethyl acetate. The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub> and then concentrated. The residue was purified by flash chromatography (CHCl<sub>3</sub>/MeOH) to give 4 as an oil (quantitative). *R*<sub>f</sub>(CH<sub>2</sub>Cl<sub>2</sub>/MeOH: 9/1) 0.27. <sup>1</sup>H-NMR (DMSO-d<sub>6</sub>): δ (ppm) 8.64 (t, *J* = 5.6 Hz, 1H) ; 7.95 (d, *J* = 8.5 Hz, 2H); 7.76 (br s, 2H); 7.37 (d, *J* = 8.1 Hz, 2H); 3.26 (m, 2H); 2.76 (m, 2H); 1.51 (m, 4H); 1.31 (m, 4H); <sup>13</sup>C-NMR (DMSO-d<sub>6</sub>): δ (ppm) 165.0; 136.2; 130.0; 128.1; 126.4; 121.8; 39.1; 38.7; 28.8; 28.04; 27.0; 25.9; 25.5; HRMS calcd. for C<sub>15</sub>H<sub>20</sub>F<sub>3</sub>N<sub>4</sub>O [M+H]<sup>+</sup> at *m/z* 329.1583, found: 329.1585

### Oligodeoxyribonucleotides synthesis

ODN synthesis was performed on an Expedite 8900 DNA synthesizer with Multiple oligo synthesis system (both from Applied Biosystem) using the “trityl on” mode and mild phosphoramidite chemistry on a 0.2 μmol scale. The standard DNA assembly protocol “DMTr-on” was used except for the following modifications: 5-Ethylthio-1*H*-tetrazole (ETT) was used as activating agent; Pac-anhydride was used for capping; a longer coupling time was applied to 2-F-dI phosphoramidite (1200s) and to the next base (dC for 2-FdI 600s). Columns were flushed with argon after completion of the synthesis and stored at -20°C before post-synthetic

modification. For the post-synthetic modification of 2-F-dI containing oligonucleotides, columns were treated with a 0.2-0.5 M solution of the desired amine 2 or 4 and 5 equivalents of DIEA for 48-72 h at 55°C. Columns were then deprotected with DBU and washed according to the instructions of the manufacturer. After that, ODNs were cleaved from the solid support and deprotected overnight with concentrated aqueous ammonia at room temperature for 24 h.

### **Oligodeoxyribonucleotides purification**

The DMT-derivatives of the ODNs were lyophilized to dryness and analyzed and purified by RP HPLC. The following gradient system was used: 95% A -(40 min)→ 60 % A / 40 % B -(5 min)→ 100 % B -(5 min)→ 100 % A with A = 0.1 M TEAA pH 7.0 and B = Acetonitrile. After HPLC purification, ODNs were detritylated in 80 % acetic acid in H<sub>2</sub>O for 1 hour at 22 °C. The mixture was evaporated, 1.5 mL of water was added and the aqueous phase was extracted 3 times with Et<sub>2</sub>O (3 × 1.5 mL). ODNs were obtained after further HPLC purification and lyophilization.

### **MALDI-TOF characterization of ODNs.**

A PerSeptive Voyager DE-STR MALDI-TOF mass spectrometer (PerSeptive Biosystems, Framingham, MA, USA), equipped with a 337-nm pulsed nitrogen laser (20 Hz) and an Acqiris® 2 GHz digitizer board, was used for all experiments. Mass spectra were obtained in the linear negative ions mode with the following settings: accelerating voltage 20 kV, grid voltage 91% of accelerating voltage, extraction delay time of 300 ns. The laser intensity was set just above the ion generation threshold to obtain peaks with the highest possible signal-to-noise (S/N) ratio without significant peak broadening. All data were processed using the Data Explorer software package (Applied Biosystems). 3-Hydroxypicolinic acid (3-HPA, used as the matrix for MALDI-TOF experiments, was of the highest grade available and used without further purification) was purchased from Sigma Aldrich Co. ODN-Ctr: calculated  $m/z$  for C<sub>317</sub>H<sub>420</sub>N<sub>102</sub>O<sub>197</sub>P<sub>30</sub>S: 9767.78; found: [M-H]<sup>-</sup> 9774.56. ODN-B: calculated  $m/z$  for C<sub>337</sub>H<sub>441</sub>N<sub>103</sub>O<sub>199</sub>P<sub>30</sub>S: 100074.94; found: [M-H]<sup>-</sup> 10078.19. ODN-D: calculated MM for C<sub>332</sub>H<sub>436</sub>F<sub>3</sub>N<sub>105</sub>O<sub>198</sub>P<sub>30</sub>S: 100078.90; found: [M-H]<sup>-</sup> 10074.40.

### **DNA annealing**

The double stranded DNA probes were obtained by mixing 1 equivalent of the photoactivable single strand DNA (either with benzophenone, diazirine, 5-iodouracile or 4-thiothymidine group) with 1.5 equivalent of the complementary oligonucleotide in an hybridation buffer (20

mM Tris-HCl pH= 7.4; 10 mM LiCl). Samples were heated at 90°C for 5 minutes then allowed to cool down to room temperature (typically overnight) and stored at 4 °C until used.

### **<sup>32</sup>P labelling of ODNs**

The indicated single-stranded DNA was radiolabelled at its 5' end by T4 polynucleotide kinase and [ $\gamma$ -<sup>32</sup>P] ATP according to the manufacture instructions. Unincorporated [ $\gamma$ -<sup>32</sup>P] ATP was removed using a Biospin 6 column (Bio-Rad) equilibrated in TE buffer. Labelled oligonucleotides were checked for purity on a 15 % Acrylamide 7 M Urea denaturing gel and further purified if needed.

### **Electrophoretic Mobility Shift Assay (EMSA)**

For all EMSAs, hRPA was diluted and preincubated (20 min at 4°C) in buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM DTT, 10 % glycerol, 0.2 mg/mL BSA, and 0.1 mM EDTA. Radiolabelled oligonucleotide (10 nM) in reaction buffer (50 mM HEPES, pH 7.9, 0.1 mg/mL BSA, 100 mM KCl and 2 % glycerol) was incubated with increasing amounts of hRPA (5, 10, 25, 50, 75 nM) for 15 min at 20°C (10  $\mu$ L). Samples were then loaded on a non-denaturing 5 % acrylamide gel (acrylamide:bisacrylamide 29:1) in 0.5X TBE buffer. Electrophoresis was performed for 45 min at room temperature. After electrophoresis, the gels were exposed to a phosphorimager screen. After being exposed for at least 10 h, the screen was scanned and the samples in the gel were quantified using ImageQuant version 5.1 or Image J software. Each experiment was reproduced at least three times. Error bars correspond to the standard deviation calculated from at least three independent experiments.

### **Photocrosslinking kinetics**

hRPA was diluted and preincubated (20 min at 4°C) in buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM DTT, 10 % glycerol, 0.2 mg/mL BSA, and 0.1 mM EDTA. hRPA (4 equivalents, final concentration 40 nM) was added to the labelled oligodeoxynucleotide (final concentration 10 nM) in reaction buffer (50 mM HEPES, pH 7.9, 0.1 mg/mL BSA, 100 mM KCl and 2 % glycerol). Samples were incubated 30 min at 20°C and then irradiated up to 2 h in 20 min increments at 4°C with a 365 nm UV lamp (Vilber Lourmat, VL-6.L, 6W). Control samples were incubated at 4°C but not irradiated. Samples were heated at 90°C for 5 min, then loaded on a denaturing 10 % acrylamide gel (acrylamide:bisacrylamide 19:1, 375 mM Tris-HCl pH 8.8, 0.1% SDS) in 1X TG-SDS Buffer. Electrophoresis was performed for 60 min at room temperature. Gel analysis was performed as previously described.

## Photolabeling experiments on:

### ➤ Purified proteins:

- hRPA: Single strand photoactivable oligonucleotide (2.5 nM) were incubated in reaction buffer (50 mM HEPES, pH 7.9, 0.1 mg/mL BSA, 100 mM KCl and 2% glycerol) with 25 nM of hRPA.
- DNMT3a: Duplexes (5 nM) were incubated in reaction buffer (20 mM HEPES pH 7.2, 1 mM EDTA, 50 mM KCl and 50 µg/mL BSA) with 2.5 µM DNMT3a and 3 mM AdoMet.
- H3.1: Duplexes (2.5 nM) were incubated in reaction buffer (50 mM HEPES, pH 7.9, 0.1 mg/mL BSA, 100 mM KCl and 2% glycerol) with 25 nM H3.1.

➤ **Bacterial lysates:** To a solution containing 1 mg/mL of bacterial protein lysates (standard, overexpressing hRPA WT or overexpressing hRPA D228Y: mutant with a reduced affinity for single-stranded DNA (for protocol see Audry et al.<sup>314</sup>) in cell lysis buffer was added 1 nM of the labelled oligodeoxynucleotide.

➤ **Eukaryotic lysates:** To a solution containing 1 mg/mL of HEK 293 protein lysates (50 µL) in hRPA reaction buffer (50 mM HEPES, pH 7.9, 0.1 mg/mL BSA, 100 mM KCl and 2 % glycerol) was added 2 nM of the labelled ODN-4<sup>S</sup>T or the labelled duplex ODN-4<sup>S</sup>T/C.

Samples were then incubated for 40 min at room temperature, under shaking, in the dark and irradiated for 100 min at 4°C with the UV lamp. Samples were heated at 90°C for 5 min before being loaded on denaturing acrylamide gel (depending of the experiment from 10 % to 15 % acrylamide:bisacrylamide 37.5:1, 375 mM Tris-HCl pH 8.8, 0.1% SDS) in 1X TG-SDS buffer. Electrophoresis was performed for 55 to 70 min at room temperature. Gel analysis was performed as previously described.

## Identification of hRPA in the higher mass band on gel

The photolabeling of bacterial lysate overexpressing hRPA and HEK293 extracts experiment, described above, was reproduced with only the 4-thio-thymidine based photoprobe (100 pmol) and with 1 mg/mL of extract (total volume of 40 µL). One of each experiment went through the same protocol as described previously and the resulting gel was transferred to a PVDF membrane using the Trans-Blot<sup>®</sup> Turbo<sup>™</sup> Transfer System for western blot experiment. For the other replicat, a step of enrichment was added using streptavidin coated magnetic beads

before the gel was loaded. Once the gel was run, it was stained using the InstantBlue™ solution from expedeon protein solution for further mass spectrometry proteomic application.

### **Western blot**

The PVDF membrane was washed in phosphate buffered saline, 0.1% Tween (PBST) buffer for 3 x5 min at room temperature on an orbital shaker and then incubated in PBST buffer containing 5% BSA at room temperature for 1 h on an orbital shaker. Finally, membranes were rinsed for 5 min with PBST. The standard ECL western blotting protocol from Amersham Biosciences was used to identify hRPA with antibodies specific for hRPA1 subunit.

### **Magnetic beads enrichment/purification, and release.**

Streptavidin coated magnetic beads were washed and equilibrated 2 times with buffer A (50 mM Tris-HCl pH=7.4; 0.1 mg/mL BSA). Oligodeoxynucleotide-protein complexes in aqueous solution from photolabeling experiments were incubated with the magnetic beads for 45 min at room temperature on an orbital shaker (~20-40 rpm). Using a magnet, the supernatant was discarded and then the beads were washed 2 times with buffer A, 2 times with buffer B (50 mM Tris-HCl pH= 7.4; 0.1 mg/mL BSA; 0.1 % SDS), 2 times with buffer C (50 mM Tris-HCl pH= 7.4; 0.1 mg/mL BSA; 1 M NaCl) and finally with distilled water until the final desired volume is reached by decreasing the washing volume (150 µL then 100 µL, etc.). Complexes were then released from the beads with 10 min incubation at 95°C before being analyzed by a denaturing gel as previously described.

### **MALDI TOF/TOF analysis of crosslinked proteins**

#### ***In gel digestion***

Gel pieces were systematically excised from both irradiated and non-irradiated experiments at the region of interest. Each gel piece was around 2 to 3 mm large and were then cut into 1 mm<sup>3</sup> pieces, washed twice with deionized water/ACN (1/1) and dried in ACN. Proteins were reduced with 10 mM DTT in 100 mM ammonium bicarbonate at 56°C for 45 min and alkylated with 55 mM iodoacetamide in 100 mM ammonium bicarbonate at room temperature for 30 min in the dark. The gels pieces were washed twice with deionized water/ACN (1/1) and dried in ACN. Then gel pieces were soaked in a 50 mM ammonium bicarbonate containing 8 ng/µL of trypsin Gold mass spectrometry grade (Promega, Madison, Wisconsin, USA) for 45 min on ice. After removing residual trypsin, the gel was incubated in 50 mM ammonium bicarbonate at 37°C overnight. Digested peptides were collected in supernatant and acidified in 0.1 % TFA.

## **Propionylation**

For the identification of the H3-ODN complexes, the bands corresponding to the desired photoadduct or in an interesting range mass were excised from the resulting gel of the photolabelling, and placed in a 1.5 mL Eppendorf. The pieces of gel were washed 2 times for 15 min with a 1 to 1 acetonitrile/water solution, and finally washed for 10 min with 100 % acetonitrile. Once all the gel were dried, a solution containing 50  $\mu$ L of 50 mM  $\text{NH}_4\text{HCO}_3$ , 50  $\mu$ L of propionylation solution (7/3 propionic anhydride/isopropanol, fresh) and 5  $\mu$ L  $\text{NH}_4\text{OH}$  28 % was added until the piece of gel were all covered. The samples were then placed 15 min at 37 °C, pH= 8. The solutions were discarded and the gels were washed quickly with 200  $\mu$ L of  $\text{NH}_4\text{HCO}_3$  50 mM and then again for 10 min. The gels were rinsed with 100 % acetonitrile for 1 and 10 min successively and repetitively until dryness. A trypsin digestion is then realized as described above. The digests were kept and a solution of 10  $\mu$ L of  $\text{NH}_4\text{HCO}_3$  100 mM and 1  $\mu$ L  $\text{NH}_4\text{OH}$  28 % were added to each samples. 5  $\mu$ L of a propionylation solution (7/3 propionic anhydride/isopropanol, fresh) were also added. The samples were then placed 15 min at 37 °C, pH= 8. Modified digests were dried (speed vac) and stored at 4 °C.

## ***Desalting and concentration of the peptide digests after enzymatic digestion***

Microcolumns packed with Poros reversed-phase 20R2 (Applied biosystems) resin were prepared using GELoader tips as described previously<sup>315</sup>. The Poros R2 column was washed with 10  $\mu$ L of 0.1 % TFA and the bound peptides were eluted directly onto the MALDI target using 0.5  $\mu$ L of a matrix solution consisting of CHCA (10  $\text{mg}\cdot\text{mL}^{-1}$ ) in ACN :  $\text{H}_2\text{O}$  (1 : 1) containing 0.1 % TFA.

## ***Mass Spectrometry Analysis***

Positive ions MALDI-TOF as well as MALDI-TOF/TOF mass spectra were performed using the Applied Biosystems 4700 Proteomics Analyzer instrument. MALDI-TOF mass spectra of the samples were performed in reflector mode using a focus mass set at 2100 u, near the threshold fluence. MALDI-TOF/TOF experiments were carried out under CID with gas ( $\text{N}_2 \sim 2 \times 10^{-7}$  Torr) with collision energy of 1keV. MS and MS/MS spectra were calibrated using external peptide standards with CHCA as matrix. Tandem mass spectra were calibrated using the angiotensin II peptide. Typically the precursor ion  $[\text{M}+\text{H}]^+$  was selected in a window (-5u,

5u), centered on the first isotope. In this window, it was verified that only the chosen precursor was selected. The laser fluence was adjusted to get unsaturated signal for the selected precursor. Data Explorer 4.6 software was used to analyze the spectra.

### ***Database searching***

The peptide mass fingerprints and MS/MS for the trypsin digests were submitted to a MASCOT sequence query search ([www.matrixscience.com](http://www.matrixscience.com)) after advanced base line correction, noise removal (standard deviation to remove = 2) and peak deisotoping (MASCOT search parameters were: Swissprot, all entries for the taxonomy, oxidized M and carbamidomethyl C as variable modifications, 2 missed cleavages allowed for trypsin, 40 ppm for the peptide mass tolerance and 0.2 Da for the MS/MS tolerance).

### **Radiolabeling of the trypsin digest**

A portion of the trypsin digested band of interest were collected (typically 5  $\mu$ L). The photoadducts, possibly present in the digest, were radiolabeled at the 5' end of the oligonucleotide part by T4 polynucleotide kinase and [ $\gamma$ - $^{32}$ P] ATP according to the manufacture instructions. Unincorporated [ $\gamma$ - $^{32}$ P] ATP was removed using a Biospin 30 column (Bio-Rad) equilibrated in TE buffer. The samples were then loaded on a 15% Acrylamide 7 M Urea denaturing gel. Electrophoresis was performed for 40 min at room temperature. After electrophoresis, the gels were exposed to a phosphorimager screen. After being exposed for at least 10 h (typically overnight), the screen was scanned with the Phosphorimager TYPHOON instrument (Molecular Dynamics).

### **Partie expérimentale de la partie II**

The peptide ARTK was given by Dominique Guianvarc'h and synthesized by Sylvia Scarabelli<sup>250</sup>.

### **Photolabeling of ARTK with radiolabeled probe**

For the photolabelling test of the ARTK peptide, the ODN-4<sup>ST</sup> (5 nM) was incubated with 50 nM ARTK. For the characterization of the ARTK-ODN complex, 100 pmol of ODN-4<sup>ST</sup> was incubated with 1000 pmol of the ARTK peptide. Finally, for the photolabelling assay with

different amino acids, the ODN-4<sup>S</sup>T was incubated with 1000 times more Fmoc amino acid (either leucine, lysine or tryptophane).

Samples were then incubated for 45 min in the dark at room temperature. Samples were then irradiated on ice with the UV lamp (Vilber Lourmat, VL-6.L, 6W) for 60 minutes. Control was incubated for 60 minutes at 4 °C but not irradiated. The samples were then analyzed by MALDI-TOF mass spectrometry. In the case of the photolabelling test with ARTK, samples were loaded on non-denaturing 5% acrylamide gel (acrylamide:bisacrylamide 29:1) in 0.5X TBE buffer. Electrophoresis was performed for 60 min at room temperature. After electrophoresis, the gels were exposed to a phosphorimager screen. After being exposed for at least 10 h (typically overnight), the screen was scanned with the Phosphorimager TYPHOON instrument (Molecular Dynamics).

### **DNA Degradation**

The DNA parts of the photoadducts were degraded with a 47-50 % HF solution overnight. The excess of HF was then evaporated on HF ramp specifically built for this high acidic product. The samples were then snap frozen and lyophilized overnight. Samples were stored at -20 °C. We also degraded the DNA moiety using a DNase (Benzonase® Nuclease, Sigma). In this case, the samples were dilute in a solution containing 20 Mm Tris HCl, pH 8.0, 2 mM MgCl<sub>2</sub>, and 20 mM NaCl. The DNase was then added (20 U), and all the samples were placed at 37 ° for at least 2h, 200 rpm. The samples were kept at - 20 °C until used.

### **Characterization of peptides.**

MALDI-TOF and MALDI-TOF/TOF spectrum were recorded as described above either with the pure peptide, the photoadduct or the DNA degraded photo adduct.

### **MALDI-TOF characterization of ODNs.**

A Voyager DE-PRO MALDI-TOF mass spectrometer (Applied Biosystems SCIEX), equipped with a 337-nm pulsed nitrogen laser (20 Hz) and an Acqiris® 2 GHz digitizer board, was used for all experiments. Mass spectra were obtained in the linear negative ion mode with the following settings: accelerating voltage 25 kV, grid voltage 92% of accelerating voltage, guide wire 0.05 and an extraction delay time of 150 ns. The laser intensity was set just above the ion

generation threshold to obtain peaks with the highest possible signal-to-noise (S/N) ratio without significant peak broadening. All data were processed using the Data Explorer software package (Applied Biosystems). 3-Hydroxypicolinic acid<sup>316</sup> (3-HPA, used as the matrix for MALDI-TOF experiments, was of the highest grade available and used without further purification) was purchased from Sigma Aldrich Co. For each analysis of oligonucleotides, 1  $\mu$ L of sample was diluted in a solution containing 9  $\mu$ L of 3-HPA (40 mg/mL) and 1  $\mu$ L of diammonium citrate (24 mg/mL). After the sample deposit on the MALDI plate, the crystallization was initiated with a tip under microscope, allowing analyzable crystals.

### **Partie expérimentale de la partie III**

#### **Photolabeling of nuclear extracts**

HeLa nuclear extracts were obtained from Santa Cruz Biotechnology. Rat brains were a gracious gift of Laurent Venance (Collège de France). Nuclear extracts from rat brain were prepared using the CellLytic NuCLEAR Extraction Kit (Sigma). Protein concentration was estimated using a Bradford assay. The radiolabeled duplex (final concentration 4 nM), prepared as previously described, were incubated in a solution of 1 mg/mL of nuclear extracts at 4 °C for 45 min. The samples were then irradiated 20 min on ice with a UV lamp (365 nm). Samples were loaded on a denaturing 12.5 % polyacrylamide gel (375 mM Tris-HCl pH=8.8; 0.1 % SDS) and electrophoresis was performed for 45 min at room temperature. After electrophoresis, the gel was exposed to a phosphorimager screen. After being exposed for at least 12 hours, the screen was scanned with the phosphorimager TYPHOON instrument (Molecular Dynamics).

#### **PCR template/probes**

The probes were obtained by PCR amplification of genomic region corresponding to the promoter of Rar $\beta$  gene using Phusion High Fidelity polymerase (Thermoscientific). The reaction were performed as advised by the provider except the annealing temperature that we fixed at 63°C. The primers used in the reaction were (Figure 98):

Rar $\beta$ - Forward (biotinylated or desthiobiotinylated):

5'-GTAGGAAGTGAGCTGTTCAGAG-3'

Rar $\beta$ - Reverse (biotinylated or desthiobiotinylated):

3'-AATGAACCTTCCTCTTGAACCC-5'

To introduce the epigenetic modifications, we used modified nucleotides: dmCTP and d5hmCTP (Zymoresearch). The photoactivable agent was introduced by running the PCR

reaction with a dTTP containing 5 % of d4<sup>S</sup>TTP (Trilink). Each probe was synthesized using the same PCR protocol, only the nucleotide changed depending on the probe we were building.

### **Purification of the PCR products**

All the PCR products were purified using a 1 % agarose gel. 1.5 g of d5 agarose (Dutscher) were dissolved in 150 mL of 0.5 X TBE buffer (Thermo Fischer Scientist) by microwaving the solution for 3 min, 750 watts. Once the solution had cooled down, few droplet of BET were added before casting the gel. The samples were then loaded with around 5% of xyanocyanate fycoll solution. The gel was running 45 min at 100 Volts. The DNA band corresponding to the desired products was excised from the gel. The DNA was then extracted using E.Z.N.A.® Gel Extraction Kit (Omega biotech). The purified DNA obtained were then quantified on a Nanodrop-2000c spectrophotometer (Thermo Scientist).

### **Chromatic phase preparation**

The forward and reverse biotinylated or desthiobiotinylated primers were designed to produce duplex DNA oligonucleotides in a bi-(desthio)-biotinylated form. Before use, the streptavidin coated beads were washed and equilibrated 2 times with TEN 1000 buffer (10 mM Tris-HCl pH= 8; 1 mM EDTA; 1 M NaCl). Preparation of the chromatographic phases was performed by overnight incubation, at 4 °C under gentle rotation, of 7.5 pmol of the double strand DNA probe with 350 µg of streptavidin-coated beads. The functionalized beads were then magnet-sedimented and washed using the TEN 1000 buffer until further use.

### **Pull-down experiments**

For each samples, functionalized beads (described above) were mixed with 800 µg of HeLa nuclear extracts (20 mM HEPES; 100 mM KCl; 220 mM glycerol; 0.5 mM PMSF; 0.5 mM DTT; 0.2 mM EDTA). The mixture was then incubated 10 min at 30°C under gentle agitation. The chromatic phases containing 4<sup>S</sup>T were irradiated for 60 min on ice, whereas the other phases were placed at 4°C in the meantime. The beads were then magnet-sedimented and the pellet was washed four times with 200 µL of the Ku buffer (40 mM HEPES; 5 mM MgCl<sub>2</sub>; 60 mM KCl; 0.4 mM EDTA; 37 mM glycerol; 0.01 % NP40; 0.5 mM DTT; pH= 7.4). The beads were finally washed with distilled water until the final desired volume is reached by decreasing the washing volume (150 µL then 100 µL, until 40 µL). All the experiments were done in quintuplet.

### **Elution and trypsin digestion**

The elution of the probes from the magnetic beads were done with 3 different conditions: biotin based elution, acidic elution and thermal condition. Firstly, the beads were magnet-sedimented and the pellet was dissolved in either 20  $\mu$ L of 10  $\mu$ M biotin solution (biotin based elution), 20  $\mu$ L of 0.1 M glycine-HCl pH= 2.8 (acidic elution) or 20  $\mu$ L of distilled water (thermal elution). Both samples from the acidic and biotin elutions were placed at 37 °C, 200 rpm and 10 min. The samples of the thermal elution were placed at 95°C for 10 min. Once the elution were performed, the beads were magnet-sedimented and the supernatant was kept. All the samples from the elution were purified using a denaturing 10 % polyacrylamide gel. The lane corresponding to the samples were then trypsin digested on gel. Some samples were not eluted allowing a direct trypsin digestion on beads.

### **On gel digestion**

For eluted samples, SDS/PAGE was used without separation as a clean-up step, and only one gel slice was excised. Subsequently, gel slices were washed and proteins were reduced with 10 mM DTT before alkylation with 55 mM iodoacetamide. After washing and shrinking the gel pieces with 100% MeCN, we performed in-gel digestion using trypsin/LysC (Promega) overnight in 25 mM  $\text{NH}_4\text{HCO}_3$  at 30°C. Peptides were then extracted using 60/35/5 MeCN/ $\text{H}_2\text{O}$ / $\text{HCOOH}$  and vacuum concentrated to dryness

### **On bead digestion**

Proteins on magnetic beads were washed twice with 100  $\mu$ L of 25 mM  $\text{NH}_4\text{HCO}_3$  and we performed on-beads digestion with 0.2  $\mu$ g of trypsin/LysC (Promega) for 1 hour in 50  $\mu$ L of 25 mM  $\text{NH}_4\text{HCO}_3$ . Sample were then loaded onto a homemade C18 StageTips for desalting (principle by stacking one 3M Empore SPE Extraction Disk Octadecyl (C18) and beads from SepPak C18 Cartridge Waters into a 200  $\mu$ l micropipette tip). Peptides were eluted using 40/60 MeCN/ $\text{H}_2\text{O}$  + 0.1% formic acid and vacuum concentrated to dryness.

### **Mass spectrometry analysis**

Samples were chromatographically separated using an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled to a TripleTOF<sup>TM</sup> 6600 mass spectrometer (ABSciex) or an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific):

-TripleTOF<sup>TM</sup> analysis: Peptides were loaded onto a C18-reversed phase column (75- $\mu$ m inner diameter  $\times$  2 cm; nanoViper Acclaim PepMap<sup>TM</sup> 100, Thermo Scientific),

separated and MS data acquired using Analyst software. Peptides separation was performed over a linear gradient of 90 min from 1% to 30% acetonitrile (75- $\mu$ m inner diameter  $\times$  50 cm; nanoViper C18, 3  $\mu$ m, 100Å, Acclaim PepMap™ RSLC, Thermo Scientific). The mass spectrometer was operated in DDA top30 high sensitivity mode with 250 and 50 ms acquisition time for MS1 and MS2 scans respectively, and 15 s dynamic exclusion. Rolling collision energy with a collision energy spread of 5 eV was used for fragmentation.

-Orbitrap analysis: Peptides were loaded onto a C18-reversed phase column (75- $\mu$ m inner diameter  $\times$  2 cm; nanoViper Acclaim PepMap™ 100, Thermo Scientific), separated and MS data acquired using Xcalibur software. Peptides separation was performed over a linear gradient of 100 min from 2% to 30% acetonitrile (75- $\mu$ m inner diameter  $\times$  50 cm; nanoViper C18, 2  $\mu$ m, 100Å, Acclaim PepMap™ RSLC, Thermo Scientific). Full-scan MS was acquired in the Orbitrap analyzer with a resolution set to 120,000 and ions from each full scan were HCD fragmented and analyzed in the linear ion trap.

## Identification

For identification the data were searched against the SwissProt *Homo Sapiens* (February 2017) database using Mascot or Sequest through proteome discoverer (version 2.1). Enzyme specificity was set to trypsin and a maximum of two missed cleavage site were allowed. Oxidized methionine, *N*-terminal acetylation, and carbamidomethyl cysteine were set as variable modifications. Maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.6 Da or 0.1 Da for MS/MS peaks. The resulting files were further processed using *myProMS*<sup>298</sup> v3.0. FDR calculation used Percolator and was set to 1% at the peptide level for the whole study. The label free quantification was performed by peptide Extracted Ion Chromatograms (XICs) computed with MassChroQ<sup>299</sup> version 1.2.1. Global MAD normalization was applied on the total signal to correct the XICs for each biological replicate. Protein ratios were computed as the geometrical mean of related peptides. To estimate ratio significance, a two-tailed *t test* was performed with a Benjamini–Hochberg FDR control threshold set to 0.05. GO term enrichment analysis was computed with the GO::TermFinder package through *myProMS* as in Obino *et al.*<sup>317</sup>. Fold change-based GO enrichment analysis was performed as in Kowal *et al.*<sup>318</sup>.

## Bibliographie

1. Watson, J. D. & Crick, F. H. C. The Structure of Dna. *Cold Spring Harb. Symp. Quant. Biol.* **18**, 123–131 (1953).
2. Limbach, P. A., Crain, P. F. & McCloskey, J. A. Summary: the modified nucleosides of RNA. *Nucleic Acids Res.* **22**, 2183–2196 (1994).
3. Sun, W.-J. *et al.* RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data. *Nucleic Acids Res.* **44**, D259–D265 (2016).
4. Grosjean, H. *Nucleic Acids Are Not Boring Long Polymers of Only Four Types of Nucleotides: A Guided Tour.* (Landes Bioscience, 2013).
5. Sood, A. J., Viner, C. & Hoffman, M. M. DNAmoD: the DNA modification database. *bioRxiv* 071712 (2016). doi:10.1101/071712
6. Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195–1214 (2003).
7. Rastogi, R. P., Richa, Kumar, A., Tyagi, M. B. & Sinha, R. P. Molecular Mechanisms of Ultraviolet Radiation-Induced DNA Damage and Repair. *J. Nucleic Acids* **2010**, (2010).
8. Burrows, C. J. & Muller, J. G. Oxidative Nucleobase Modifications Leading to Strand Scission. *Chem. Rev.* **98**, 1109–1152 (1998).
9. Margison, G. P., Koref, S., F, M. & Povey, A. C. Mechanisms of carcinogenicity/chemotherapy by O6-methylguanine. *Mutagenesis* **17**, 483–487 (2002).
10. Hurley, L. H. DNA and its associated processes as targets for cancer therapy. *Nat. Rev. Cancer* **2**, 188–200 (2002).
11. Loechler, E. L., Green, C. L. & Essigmann, J. M. In vivo mutagenesis by O6-methylguanine built into a unique site in a viral genome. *Proc. Natl. Acad. Sci.* **81**, 6271–6275 (1984).
12. Saffhill, R., Margison, G. P. & O'Connor, P. J. Mechanisms of carcinogenesis induced by alkylating agents. *Biochim. Biophys. Acta BBA - Rev. Cancer* **823**, 111–145 (1985).

13. Arber, W. & Dussoix, D. Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J. Mol. Biol.* **5**, 18–36 (1962).
14. Tock, M. R. & Dryden, D. T. The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* **8**, 466–472 (2005).
15. Waddington, C. H. The Epigenotype. *Int. J. Epidemiol.* **41**, 10–13 (2012).
16. Berger, S. L., Kouzarides, T., Shiekhatar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **23**, 781–783 (2009).
17. Lee, J. T. Epigenetic Regulation by Long Noncoding RNAs. *Science* **338**, 1435–1439 (2012).
18. Ramsahoye, B. H. *et al.* Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci.* **97**, 5237–5242 (2000).
19. Feng, S., Jacobsen, S. E. & Reik, W. Epigenetic Reprogramming in Plant and Animal Development. *Science* **330**, 622–627 (2010).
20. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
21. Das, P. M. & Singal, R. DNA Methylation and Cancer. *J. Clin. Oncol.* **22**, 4632–4642 (2004).
22. Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nat. Rev. Cancer* **4**, 143–153 (2004).
23. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683–692 (2007).
24. Shen, L. *et al.* Genome-Wide Profiling of DNA Methylation Reveals a Class of Normally Methylated CpG Island Promoters. *PLOS Genet.* **3**, e181 (2007).
25. Jeltsch, A. Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases. *ChemBioChem* **3**, 274–293 (2002).

26. Gowher, H. & Jeltsch, A. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites. *J. Mol. Biol.* **309**, 1201–1208 (2001).
27. Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H. & Tajima, S. DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction. *J. Biol. Chem.* **279**, 27816–27823 (2004).
28. Gowher, H., Liebert, K., Hermann, A., Xu, G. & Jeltsch, A. Mechanism of Stimulation of Catalytic Activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *J. Biol. Chem.* **280**, 13341–13348 (2005).
29. Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Genome Res.* **14**, 9–25 (1975).
30. Test 'Title' - HollidayAndPugh1975.pdf. Available at: <http://depts.washington.edu/lairdlab/pdfs/HollidayAndPugh1975.pdf>. (Accessed: 3rd May 2017)
31. Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
32. Liang, G. *et al.* Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. *Mol. Cell. Biol.* **22**, 480–491 (2002).
33. Jones, P. A. & Liang, G. Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* **10**, 805–811 (2009).
34. Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).
35. Chuang, L. S.-H. *et al.* Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1. *Science* **277**, 1996–2000 (1997).

36. Schermelleh, L. *et al.* Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.* **35**, 4301–4312 (2007).
37. Sharif, J. *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**, (2007).
38. Bostick, M. *et al.* UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science* **317**, 1760–1764 (2007).
39. Bashtrykov, P., Jankevicius, G., Jurkowska, R. Z., Ragozin, S. & Jeltsch, A. The UHRF1 Protein Stimulates the Activity and Specificity of the Maintenance DNA Methyltransferase DNMT1 by an Allosteric Mechanism. *J. Biol. Chem.* **289**, 4106–4115 (2014).
40. Hashimoto, H., Horton, J. R., Zhang, X. & Cheng, X. UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics* **4**, 8–14 (2009).
41. Liu, X. *et al.* UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat. Commun.* **4**, 1563 (2013).
42. Rajakumara, E. *et al.* PHD Finger Recognition of Unmodified Histone H3R2 Links UHRF1 to Regulation of Euchromatic Gene Expression. *Mol. Cell* **43**, 275–284 (2011).
43. Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. & Shirakawa, M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455**, 818–821 (2008).
44. Hashimoto, H. *et al.* The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* **455**, 826–829 (2008).
45. Hashimoto, H., Horton, J. R., Zhang, X. & Cheng, X. UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics Off. J. DNA Methylation Soc.* **4**, 8–14 (2009).

46. Nishiyama, A. *et al.* Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* **502**, 249–253 (2013).
47. Qin, W. *et al.* DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res.* **25**, 911–929 (2015).
48. Oswald, J. *et al.* Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* **10**, 475–478 (2000).
49. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Embryogenesis: Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502 (2000).
50. Hackett, J. A., Zyllicz, J. J. & Surani, M. A. Parallel mechanisms of epigenetic reprogramming in the germline. *Trends Genet.* **28**, 164–174 (2012).
51. Kangaspeka, S. *et al.* Transient cyclical methylation of promoter DNA. *Nature* **452**, 112–115 (2008).
52. Ohno, R. *et al.* A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. *Development* **140**, 2892–2903 (2013).
53. Bhattacharya, S. K., Ramchandani, S., Cervoni, N. & Szyf, M. A mammalian protein with specific demethylase activity for mCpG DNA. *Nature* **397**, 579–583 (1999).
54. Ng, H.-H. *et al.* MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat. Genet.* **23**, 58–61 (1999).
55. Wu, S. C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**, 607–620 (2010).
56. Jacobs, A. L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1–20 (2012).
57. Jost, J.-P., Siegmann, M., Sun, L. & Leung, R. Mechanisms of DNA Demethylation in Chicken Embryos PURIFICATION AND PROPERTIES OF A 5-METHYLCYTOSINE-DNA GLYCOSYLASE. *J. Biol. Chem.* **270**, 9734–9739 (1995).

58. Zhu, B. *et al.* 5-Methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. *Proc. Natl. Acad. Sci.* **97**, 5135–5139 (2000).
59. Hashimoto, H., Hong, S., Bhagwat, A. S., Zhang, X. & Cheng, X. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* **40**, 10203–10214 (2012).
60. Zhu, B. *et al.* 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucleic Acids Res.* **28**, 4157–4165 (2000).
61. Reardon, J. T. & Sancar, A. Nucleotide Excision Repair. in (ed. Biology, B.-P. in N. A. R. and M.) **79**, 183–235 (Academic Press, 2005).
62. Barreto, G. *et al.* Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445**, 671–675 (2007).
63. Jin, S.-G., Guo, C. & Pfeifer, G. P. GADD45A Does Not Promote DNA Demethylation. *PLOS Genet.* **4**, e1000013 (2008).
64. Schmitz, K.-M. *et al.* TAF12 Recruits Gadd45a and the Nucleotide Excision Repair Complex to the Promoter of rRNA Genes Leading to Active DNA Demethylation. *Mol. Cell* **33**, 344–353 (2009).
65. Le May, N. *et al.* NER Factors Are Recruited to Active Promoters and Facilitate Chromatin Modification for Transcription in the Absence of Exogenous Genotoxic Attack. *Mol. Cell* **38**, 54–66 (2010).
66. Li, Z. *et al.* Gadd45a promotes DNA demethylation through TDG. *Nucleic Acids Res.* **43**, 3986–3997 (2015).

67. Conticello, S. G. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **9**, 229 (2008).
68. Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced Cytidine Deaminase Deaminates 5-Methylcytosine in DNA and Is Expressed in Pluripotent Tissues IMPLICATIONS FOR EPIGENETIC REPROGRAMMING. *J. Biol. Chem.* **279**, 52353–52360 (2004).
69. Kumar, R. *et al.* AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. *Nature* **500**, 89–92 (2013).
70. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042–1047 (2010).
71. Bransteitter, R., Pham, P., Scharff, M. D. & Goodman, M. F. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci.* **100**, 4102–4107 (2003).
72. Nabel, C. S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).
73. Klose, R. J. *et al.* The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* **442**, 312–316 (2006).
74. Falnes, P. Ø., Johansen, R. F. & Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* **419**, 178–182 (2002).
75. Borst, P. & Sabatini, R. Base J: Discovery, Biosynthesis, and Possible Functions. *Annu. Rev. Microbiol.* **62**, 235–251 (2008).
76. Yu, Z. *et al.* The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res.* **35**, 2107–2115 (2007).

77. Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8**, 1698–1710 (2009).
78. Penn, N. W., Suwalaki, R., O'Riley, C., Bojanowski, K. & Yura, R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* **126**, 781–790 (1972).
79. Kriaucionis, S. & Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* **324**, 929–930 (2009).
80. Tahiliani, M. *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324**, 930–935 (2009).
81. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
82. Ito, S. *et al.* Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **333**, 1300–1303 (2011).
83. He, Y.-F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **333**, 1303–1307 (2011).
84. Frauer, C. *et al.* Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain. *PLOS ONE* **6**, e21306 (2011).
85. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40**, 4841–4849 (2012).
86. Valinluck, V. & Sowers, L. C. Endogenous Cytosine Damage Products Alter the Site Selectivity of Human DNA Maintenance Methyltransferase DNMT1. *Cancer Res.* **67**, 946–950 (2007).
87. Ji, D., Lin, K., Song, J. & Wang, Y. Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. *Mol. Biosyst.* **10**, 1749–1752 (2014).

88. Otani, J. *et al.* Cell Cycle-Dependent Turnover of 5-Hydroxymethyl Cytosine in Mouse Embryonic Stem Cells. *PLOS ONE* **8**, e82961 (2013).
89. Inoue, A., Shen, L., Dai, Q., He, C. & Zhang, Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* **21**, 1670–1676 (2011).
90. Liutkevičiūtė, Z., Lukinavičius, G., Masevičius, V., Daujotytė, D. & Klimašauskas, S. Cytosine-5-methyltransferases add aldehydes to DNA. *Nat. Chem. Biol.* **5**, 400–402 (2009).
91. Chen, C.-C., Wang, K.-Y. & Shen, C.-K. J. The Mammalian de Novo DNA Methyltransferases DNMT3A and DNMT3B Are Also DNA 5-Hydroxymethylcytosine Dehydroxymethylases. *J. Biol. Chem.* **287**, 33116–33121 (2012).
92. Smiley, J. A., Kundracik, M., Landfried, D. A., Barnes Sr., V. R. & Axhemi, A. A. Genes of the thymidine salvage pathway: Thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa*. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1723**, 256–264 (2005).
93. Schiesser, S. *et al.* Mechanism and Stem-Cell Activity of 5-Carboxycytosine Decarboxylation Determined by Isotope Tracing. *Angew. Chem. Int. Ed.* **51**, 6516–6520 (2012).
94. Maiti, A. & Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine POTENTIAL IMPLICATIONS FOR ACTIVE DEMETHYLATION OF CpG SITES. *J. Biol. Chem.* **286**, 35334–35338 (2011).
95. Schomacher, L. *et al.* Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nat. Struct. Mol. Biol.* **23**, 116–124 (2016).
96. Song, C.-X. *et al.* Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming. *Cell* **153**, 678–691 (2013).

97. Neri, F. *et al.* Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Rep.* **10**, 674–683 (2015).
98. Weber, A. R. *et al.* Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat. Commun.* **7**, 10806 (2016).
99. Spruijt, C. G. *et al.* Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **152**, 1146–1159 (2013).
100. Bandaru, V., Sunkara, S., Wallace, S. S. & Bond, J. P. A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to Escherichia coli endonuclease VIII. *DNA Repair* **1**, 517–529 (2002).
101. Krokeide, S. Z. *et al.* Human NEIL3 is mainly a monofunctional DNA glycosylase removing spiroimindiohydantoin and guanidinohydantoin. *DNA Repair* **12**, 1159–1164 (2013).
102. Müller, U., Bauer, C., Siegl, M., Rottach, A. & Leonhardt, H. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.* **42**, 8592–8604 (2014).
103. Moréra, S. *et al.* Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA. *Nucleic Acids Res.* **40**, 9917–9926 (2012).
104. Mokkapatil, S. K., Wiederhold, L., Hazra, T. K. & Mitra, S. Stimulation of DNA Glycosylase Activity of OGG1 by NEIL1: Functional Collaboration between Two Human DNA Glycosylases. *Biochemistry (Mosc.)* **43**, 11596–11604 (2004).
105. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **6**, 1049–1055 (2014).
106. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **6**, 1049–1055 (2014).

107. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).
108. Thalhammer, A., Hansen, A. S., El-Sagheer, A. H., Brown, T. & Schofield, C. J. Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation. *Chem. Commun.* **47**, 5325–5327 (2011).
109. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).
110. Su, M. *et al.* 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites. *Angew. Chem. Int. Ed.* **55**, 11797–11800 (2016).
111. Vigneault, F. & Guérin, S. L. Regulation of gene expression: probing DNA–protein interactions in vivo and in vitro. *Expert Rev. Proteomics* **2**, 705–718 (2005).
112. Hégarat, N., François, J.-C. & Praseuth, D. Modern tools for identification of nucleic acid-binding proteins. *Biochimie* **90**, 1265–1272 (2008).
113. Lescuyer, P., Chevallet, M. & Rabilloud, T. L’analyse protéomique : concepts, réalités et perspectives en thérapeutique. *médecine/sciences* **20**, 587–592 (2004).
114. Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat. Protoc.* **2**, 1849–1861 (2007).
115. Fried, M. & Crothers, D. M. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**, 6505–6525 (1981).
116. Garner, M. M. & Revzin, A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res.* **9**, 3047–3060 (1981).
117. Hendrickson, W. Protein-DNA interactions studied by the gel electrophoresis-DNA binding assay. *BioTechniques* **3**, 198–207 (1985).

118. Revzin, A. Gel electrophoresis assays for DNA-protein interactions. *BioTechniques* **7**, 346–355 (1989).
119. Soeno, Y. *et al.* Identification of novel ribonucleo-protein complexes from the brain-specific snoRNA MBII-52. *RNA* **16**, 1293–1300 (2010).
120. Woo, A. J., Dods, J. S., Susanto, E., Ulgiati, D. & Abraham, L. J. A Proteomics Approach for the Identification of DNA Binding Activities Observed in the Electrophoretic Mobility Shift Assay. *Mol. Cell. Proteomics* **1**, 472–478 (2002).
121. Uguru, G. C. *et al.* Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. *Mol. Microbiol.* **58**, 131–150 (2005).
122. Stead, J. A., Keen, J. N. & McDowall, K. J. The Identification of Nucleic Acid-interacting Proteins Using a Simple Proteomics-based Approach That Directly Incorporates the Electrophoretic Mobility Shift Assay. *Mol. Cell. Proteomics* **5**, 1697–1702 (2006).
123. Jin, S.-G., Kadam, S. & Pfeifer, G. P. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.* **38**, e125 (2010).
124. Fatemi, M. & Wade, P. A. MBD family proteins: reading the epigenetic code. *J. Cell Sci.* **119**, 3033–3037 (2006).
125. Zhang, L. *et al.* Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **8**, 328–330 (2012).
126. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).
127. Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J. & Salemme, F. R. Structural Origins of High-Affinity Biotin Binding to Streptavidin. *Science* **243**, 85–88 (1989).

128. Ong, S.-E. & Mann, M. Mass spectrometry–based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252–262 (2005).
129. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389**, 1017–1031 (2007).
130. Strittmatter, E. F., Ferguson, P. L., Tang, K. & Smith, R. D. Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom.* **14**, 980–991 (2003).
131. Washburn, M. P., Wolters, D. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247 (2001).
132. Chen, X., Smith, L. M. & Bradbury, E. M. Site-Specific Mass Tagging with Stable Isotopes in Proteins for Accurate and Efficient Protein Identification. *Anal. Chem.* **72**, 1134–1143 (2000).
133. Ong, S.-E. *et al.* Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
134. Ramus, C., Peredo, A. G. de, Dahout, C., Gallagher, M. & Garin, J. An Optimized Strategy for ICAT Quantification of Membrane Proteins. *Mol. Cell. Proteomics* **5**, 68–78 (2006).
135. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
136. Yao, X., Freas, A., Ramirez, J., Demirev, P. A. & Fenselau, C. Proteolytic <sup>18</sup>O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus. *Anal. Chem.* **73**, 2836–2842 (2001).

137. Goodlett, D. R. *et al.* Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. *Rapid Commun. Mass Spectrom.* **15**, 1214–1221 (2001).
138. Zhou, T. *et al.* Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. *Mol. Cell* **54**, 879–886 (2014).
139. Globisch, D. *et al.* Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. *PLOS ONE* **5**, e15367 (2010).
140. Gavrilov, A., Razin, S. V. & Cavalli, G. In vivo formaldehyde cross-linking: it is time for black box analysis. *Brief. Funct. Genomics* **14**, 163–165 (2015).
141. Fraenkel-Conrat, H., Cooper, M. & Olcott, H. S. The Reaction of Formaldehyde with Proteins. *J. Am. Chem. Soc.* **67**, 950–954 (1945).
142. Metz, B. *et al.* Identification of Formaldehyde-induced Modifications in Proteins REACTIONS WITH MODEL PEPTIDES. *J. Biol. Chem.* **279**, 6235–6243 (2004).
143. Toews, J., Rogalski, J. C., Clark, T. J. & Kast, J. Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions. *Anal. Chim. Acta* **618**, 168–183 (2008).
144. Toews, J., Rogalski, J. C. & Kast, J. Accessibility governs the relative reactivity of basic residues in formaldehyde-induced protein modifications. *Anal. Chim. Acta* **676**, 60–67 (2010).
145. Lu, K. *et al.* Structural Characterization of Formaldehyde-Induced Cross-Links Between Amino Acids and Deoxynucleosides and Their Oligomers. *J. Am. Chem. Soc.* **132**, 3388–3399 (2010).
146. Fullwood, M. J. & Ruan, Y. ChIP-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.* **107**, 30–39 (2009).
147. J, W. & Tk, K. Histone modifications and transcription factor binding on chromatin ChIP-PCR assays. *Methods Mol. Biol. Clifton NJ* **325**, 273–283 (2006).

148. Wang, C. I. *et al.* Chromatin proteins captured by ChIP–mass spectrometry are linked to dosage compensation in *Drosophila*. *Nat. Struct. Mol. Biol.* **20**, 202–209 (2013).
149. Smith, K. C. Dose dependent decrease in extractability of DNA from bacteria following irradiation with ultraviolet light or with visible light plus dye. *Biochem. Biophys. Res. Commun.* **8**, 157–163 (1962).
150. Varghese, A. J. Photochemical Addition of Amino Acids and Related Compounds to Nucleic Acid Constituents. in *Aging, Carcinogenesis, and Radiation Biology* (ed. Smith, K. C.) 207–223 (Springer US, 1976). doi:10.1007/978-1-4757-1662-7\_12
151. Kunkel, G. R. & Martinson, H. G. Histone-DNA interactions within chromatin. Isolation of histones from DNA-histone adducts induced in nuclei by UV light. *Nucleic Acids Res.* **5**, 4263–4272 (1978).
152. Cao, T. M. & Sung, M. T. Ultraviolet light induced preferential crosslinking of histone H3 to DNA in chromatin and nuclei of chicken erythrocytes. *Biochemistry (Mosc.)* **21**, 3419–3427 (1982).
153. Hillel, Z. & Wu, C.-W. Photochemical cross-linking studies on the interaction of *Escherichia coli* RNA polymerase with T7 DNA. *Biochemistry (Mosc.)* **17**, 2954–2961 (1978).
154. Park, C. S., Hillel, Z. & Wu, C.-W. DNA strand specificity in promoter recognition by RNA polymerase. *Nucleic Acids Res.* **8**, 5895–5912 (1980).
155. Shivanna, B. D., Mejillano, M. R., Williams, T. D. & Himes, R. H. Exchangeable GTP binding site of beta-tubulin. Identification of cysteine 12 as the major site of cross-linking by direct photoaffinity labeling. *J. Biol. Chem.* **268**, 127–132 (1993).
156. P. Sinha, R. & Häder, D.-P. UV-induced DNA damage and repair: a review. *Photochem. Photobiol. Sci.* **1**, 225–236 (2002).

157. Fleet, G. W. J., Porter, R. R. & Knowles, J. R. Affinity Labelling of Antibodies with Aryl Nitrene as Reactive Group. *Nature* **224**, 511–512 (1969).
158. Smith, R. A. G. & Knowles, J. R. Aryldiazirines. Potential reagents for photolabeling of biological receptor sites. *J. Am. Chem. Soc.* **95**, 5072–5073 (1973).
159. Galardy, R. E., Craig, L. C. & Printz, M. P. Benzophenone Triplet: a New Photochemical Probe of Biological Ligand-Receptor Interactions. *Nature* **242**, 127–128 (1973).
160. Wieland, T., Dungen, A. V. & Birr, C. über Peptidsynthesen, LI. über Antamanid, IX. Synthese einer antitoxischen Antamanid-Variante mit p-Azidophenylalanin in Stelle 6. *Justus Liebigs Ann. Chem.* **752**, 109–114 (1971).
161. Czarnecki, J. J. Tautomerism of 2-azidoadenine nucleotides Effects on enzyme kinetics and photoaffinity labeling. *Biochim. Biophys. Acta BBA - Gen. Subj.* **800**, 41–51 (1984).
162. Pomerantz, A. H., Rudolph, S. A., Haley, B. E. & Greengard, P. Photoaffinity labeling of a protein kinase from bovine brain with 8-azidoadenosine 3',5'-monophosphate. *Biochemistry (Mosc.)* **14**, 3858–3862 (1975).
163. Bergeron, R. J., Dionis, J. B. & Ingeno, M. J. Synthesis of a parabactin photoaffinity label. *J. Org. Chem.* **52**, 144–149 (1987).
164. Platz, M. S. Comparison of Phenylcarbene and Phenylnitrene. *Acc. Chem. Res.* **28**, 487–492 (1995).
165. Staros, J. V., Bayley, H., Standring, D. N. & Knowles, J. R. Reduction of aryl azides by thiols: Implications for the use of photoaffinity reagents. *Biochem. Biophys. Res. Commun.* **80**, 568–572 (1978).
166. Brunner, J., Senn, H. & Richards, F. M. 3-Trifluoromethyl-3-phenyldiazirine. A new carbene generating group for photolabeling reagents. *J. Biol. Chem.* **255**, 3313–3318 (1980).

167. Platz, M. *et al.* Photolysis of 3-aryl-3-(trifluoromethyl)diazirines: a caveat regarding their use in photoaffinity probes. *Bioconjug. Chem.* **2**, 337–341 (1991).
168. Galardy, R. E., Craig, L. C., Jamieson, J. D. & Printz, M. P. Photoaffinity Labeling of Peptide Hormone Binding Sites. *J. Biol. Chem.* **249**, 3510–3518 (1974).
169. Boyd, N. D., Kage, R., Dumas, J. J., Krause, J. E. & Leeman, S. E. The peptide binding site of the substance P (NK-1) receptor localized by a photoreactive analogue of substance P: presence of a disulfide bond. *Proc. Natl. Acad. Sci.* **93**, 433–437 (1996).
170. Sachon, E., Girault-Lagrange, S., Chassaing, G., Lavielle, S. & Sagan, S. Analogs of Substance P modified at the C-terminus which are both agonist and antagonist of the NK-1 receptor depending on the second messenger pathway. *J. Pept. Res.* **59**, 232–240 (2002).
171. Salisbury, C. M. & Cravatt, B. F. Activity-based probes for proteomic profiling of histone deacetylase complexes. *Proc. Natl. Acad. Sci.* **104**, 1171–1176 (2007).
172. Tate, J. J., Persinger, J. & Bartholomew, B. Survey of four different photoreactive moieties for DNA photoaffinity labeling of yeast RNA polymerase III transcription complexes. *Nucleic Acids Res.* **26**, 1421–1426 (1998).
173. Dorman, G. & Prestwich, G. D. Benzophenone Photophores in Biochemistry. *Biochemistry (Mosc.)* **33**, 5661–5673 (1994).
174. Vodovozova, E. L. Photoaffinity labeling and its application in structural biology. *Biochem. Mosc.* **72**, 1–20 (2007).
175. Geahlen, R. L. & Haley, B. E. Interactions of a photoaffinity analog of GTP with the proteins of microtubules. *Proc. Natl. Acad. Sci.* **74**, 4375–4377 (1977).
176. Evans, R. K., Johnson, J. D. & Haley, B. E. 5-Azido-2'-deoxyuridine 5'-triphosphate: a photoaffinity-labeling reagent and tool for the enzymatic synthesis of photoactive DNA. *Proc. Natl. Acad. Sci.* **83**, 5382–5386 (1986).

177. Hicke, B. J., Willis, M. C., Koch, T. H. & Cech, T. R. Telomeric Protein-DNA Point Contracts Identified by Photo-Crosslinking Using 5-Bromodeoxyuridine. *Biochemistry (Mosc.)* **33**, 3364–3373 (1994).
178. Ogata, R. & Gilbert, W. Contacts between the lac repressor and the thymines in the lac operator. *Proc. Natl. Acad. Sci.* **74**, 4973–4976 (1977).
179. Chen, A., Powell, L. M., Dryden, D. T. F., Murray, N. E. & Brown, T. Tyrosine 27 of the specificity polypeptide of Eco KI can be UV crosslinked to a bromodeoxyuridine-substituted DNA target sequence. *Nucleic Acids Res.* **23**, 1177–1183 (1995).
180. Willis, M. C., Hicke, B. J., Uhlenbeck, O. C., Cech, T. & Koch, T. H. Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science* **262**, 1255–1258 (1993).
181. Jensen, K. B., Atkinson, B. L., Willis, M. C., Koch, T. H. & Gold, L. Using in vitro selection to direct the covalent attachment of human immunodeficiency virus type 1 Rev protein to high-affinity RNA ligands. *Proc. Natl. Acad. Sci.* **92**, 12220–12224 (1995).
182. Malkov, V. A. & Camerini-Otero, R. D. Photocross-links between Single-stranded DNA and Escherichia coli RecA Protein Map to Loops L1 (Amino Acid Residues 157-164) and L2 (Amino Acid Residues 195-209). *J. Biol. Chem.* **270**, 30230–30233 (1995).
183. Norris, C. L., Meisenheimer, K. M. & Koch, T. H. Mechanistic Studies Relevant to Bromouridine-Enhanced Nucleoprotein Photocrosslinking: Possible Involvement of an Excited Tyrosine Residue of the Protein. *Photochem. Photobiol.* **65**, 201–207 (1997).
184. Norris, C. L., Meisenheimer, P. L. & Koch, T. H. Mechanistic Studies of the 5-Iodouracil Chromophore Relevant to Its Use in Nucleoprotein Photo-Cross-Linking. *J. Am. Chem. Soc.* **118**, 5796–5803 (1996).
185. Dietz, T. M. & Koch, T. H. Photochemical Coupling of 5-Bromouracil to Tryptophan, Tyrosine and Histidine, Peptide-Like Derivatives in Aqueous Fluid Solution. *Photochem. Photobiol.* **46**, 971–978 (1987).

186. Dietz, T. M. & Koch, T. H. Photochemical Reduction of 5-Bromouracil by Cysteine Derivatives and Coupling of 5-Bromouracil to Cystine Derivatives. *Photochem. Photobiol.* **49**, 121–129 (1989).
187. Weintraub, H. The Assembly of Newly Replicated DNA into Chromatin. *Cold Spring Harb. Symp. Quant. Biol.* **38**, 247–256 (1974).
188. Blatter, E. E., Ebright, Y. W. & Ebright, R. H. Identification of an amino acid–base contact in the GCN4–DNA complex by bromouracil-mediated photocrosslinking. *Nature* **359**, 650–652 (1992).
189. Gauss, D. H. & Sprinzl, M. Compilation of tRNA sequences. *Nucleic Acids Res.* **9**, r1–r23 (1981).
190. Yaniv, M. & Favre, A. Structure of Transfer RNA: Evidence for Interaction between Two Non-adjacent Nucleotide Residues in tRNA<sup>Val1</sup> from *Escherichia coli*. *Nature* **223**, 1331–1333 (1969).
191. Egli, M. & Saenger, W. *Principles of Nucleic Acid Structure*. (Springer Science & Business Media, 2013).
192. Sintim, H. O. & Kool, E. T. Enhanced Base Pairing and Replication Efficiency of Thiothymidines, Expanded-size Variants of Thymidine. *J. Am. Chem. Soc.* **128**, 396–397 (2006).
193. Somerville, L. *et al.* Structure and Dynamics of Thioguanine-modified Duplex DNA. *J. Biol. Chem.* **278**, 1005–1011 (2003).
194. Bökemeier, W. & Lezius, A. G. On the structure of poly d(A-s 4 T). *Nucleic Acids Res.* **14**, 2241–2249 (1986).
195. Harada, Y., Suzuki, T., Ichimura, T. & Xu, Y.-Z. Triplet Formation of 4-Thiothymidine and Its Photosensitization to Oxygen Studied by Time-Resolved Thermal Lensing Technique. *J. Phys. Chem. B* **111**, 5518–5524 (2007).

196. Ge, J., Feng, L., Yaozhong, X. & Xiaohui, Z. UVA Assisted 4-Thiothymidine for Cancer Treatment. *Prog. Chem.* **28**, 1224–1237 (2016).
197. Favre, A., Saintomé, C., Fourrey, J.-L., Clivio, P. & Laugâa, P. Thionucleobases as intrinsic photoaffinity probes of nucleic acid structure and nucleic acid-protein interactions. *J. Photochem. Photobiol. B* **42**, 109–124 (1998).
198. Ito, S., Saito, I., Nakata, A. & Matsuura, T. Photochemical synthesis of model compounds for nucleic acid-protein adducts. *Nucleic Acids Res.* **1**, s321–s326 (1978).
199. Ito, S., Saito, I., Nakata, A. & Matsuura, T. Photochemical Addition of L-Lysine to 1,3-Dimethyl-4-Thiouracil\*. *Photochem. Photobiol.* **32**, 683–685 (1980).
200. Wlassoff, W. A. *et al.* Synthesis and Characterization of (d)NTP Derivatives Substituted with Residues of Different Photoreagents. *Bioconjug. Chem.* **6**, 352–360 (1995).
201. C. Smith, C., Hollenstein, M. & J. Leumann, C. The synthesis and application of a diazirine-modified uridine analogue for investigating RNA–protein interactions. *RSC Adv.* **4**, 48228–48235 (2014).
202. Ludwig, J. & Eckstein, F. Rapid and efficient synthesis of nucleoside 5'-O-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* **54**, 631–635 (1989).
203. Shah, K., Wu, H. & Rana, T. M. Synthesis of Uridine Phosphoramidite Analogs: Reagents for Site-Specific Incorporation of Photoreactive Sites into RNA Sequences. *Bioconjug. Chem.* **5**, 508–512 (1994).
204. Liebmann, M., Di Pasquale, F. & Marx, A. A New Photoactive Building Block for Investigation of DNA Backbone Interactions: Photoaffinity Labeling of Human DNA Polymerase  $\beta$ . *ChemBioChem* **7**, 1965–1969 (2006).

205. Winnacker, M., Welzmler, V., Strasser, R. & Carell, T. Development of a DNA Photoaffinity Probe for the Analysis of 8-OxodG-Binding Proteins in a Human Proteome. *ChemBioChem* **11**, 1345–1349 (2010).
206. Weisbrod, S. H. & Marx, A. Novel strategies for the site-specific covalent labelling of nucleic acids. *Chem. Commun.* 5675–5685 (2008). doi:10.1039/B809528K
207. Lercher, L., McGouran, J. F., Kessler, B. M., Schofield, C. J. & Davis, B. G. DNA Modification under Mild Conditions by Suzuki–Miyaura Cross-Coupling for the Generation of Functional Probes. *Angew. Chem. Int. Ed.* **52**, 10553–10558 (2013).
208. Beyer, C. & Wagenknecht, H.-A. In situ azide formation and “click” reaction of Nile Red with DNA as an alternative postsynthetic route. *Chem. Commun.* **46**, 2230–2231 (2010).
209. El-Sagheer, A. H. & Brown, T. Click chemistry with DNA. *Chem. Soc. Rev.* **39**, 1388–1405 (2010).
210. Gutmiedl, K., Wirges, C. T., Ehmke, V. & Carell, T. Copper-Free “Click” Modification of DNA via Nitrile Oxide–Norbornene 1,3-Dipolar Cycloaddition. *Org. Lett.* **11**, 2405–2408 (2009).
211. Schoch, J., Wiessler, M. & Jäschke, A. Post-Synthetic Modification of DNA by Inverse-Electron-Demand Diels–Alder Reaction. *J. Am. Chem. Soc.* **132**, 8846–8847 (2010).
212. Raindlová, V., Pohl, R., Šanda, M. & Hocek, M. Direct Polymerase Synthesis of Reactive Aldehyde-Functionalized DNA and Its Conjugation and Staining with Hydrazines. *Angew. Chem. Int. Ed.* **49**, 1064–1066 (2010).
213. Raindlová, V., Pohl, R. & Hocek, M. Synthesis of Aldehyde-Linked Nucleotides and DNA and Their Bioconjugations with Lysine and Peptides through Reductive Amination. *Chem. – Eur. J.* **18**, 4080–4087 (2012).
214. Wang, C. C.-Y., Seo, T. S., Li, Z., Ruparel, H. & Ju, J. Site-Specific Fluorescent Labeling of DNA Using Staudinger Ligation. *Bioconjug. Chem.* **14**, 697–701 (2003).

215. Shigdel, U. K., Zhang, J. & He, C. Diazirine-Based DNA Photo-Cross-Linking Probes for the Study of Protein–DNA Interactions. *Angew. Chem. Int. Ed.* **47**, 90–93 (2008).
216. Rohde, H., Schmalisch, J., Harpaz, Z., Diezmann, F. & Seitz, O. Ascorbate as an Alternative to Thiol Additives in Native Chemical Ligation. *ChemBioChem* **12**, 1396–1400 (2011).
217. Eberhard, H., Diezmann, F. & Seitz, O. DNA as a Molecular Ruler: Interrogation of a Tandem SH2 Domain with Self-Assembled, Bivalent DNA–Peptide Complexes. *Angew. Chem. Int. Ed.* **50**, 4146–4150 (2011).
218. Schlegel, M. K., Hütter, J., Eriksson, M., Lepenies, B. & Seeberger, P. H. Defined Presentation of Carbohydrates on a Duplex DNA Scaffold. *ChemBioChem* **12**, 2791–2800 (2011).
219. MacMillan, A. M. & Verdine, G. L. Synthesis of functionally tethered oligodeoxynucleotides by the convertible nucleoside approach. *J. Org. Chem.* **55**, 5931–5933 (1990).
220. Shigdel, U. K., Zhang, J. & He, C. Diazirine-Based DNA Photo-Cross-Linking Probes for the Study of Protein–DNA Interactions. *Angew. Chem.* **120**, 96–99 (2008).
221. Singh, A., Thornton, E. R. & Westheimer, F. H. The photolysis of diazoacetylchymotrypsin. *J. Biol. Chem.* **237**, 3006–3008 (1962).
222. Golden, M. C., Resing, K. A., Collins, B. D., Willis, M. C. & Koch, T. H. Mass spectral characterization of a protein–nucleic acid photocrosslink. *Protein Sci.* **8**, 2806–2812 (1999).
223. Maltseva, E. A. *et al.* Interaction of nucleotide excision repair factors RPA and XPA with DNA containing bulky photoreactive groups imitating damages. *Biochem. Mosc.* **71**, 270–278 (2006).

224. Robinette, D., Neamati, N., Tomer, K. B. & Borchers, C. H. Photoaffinity labeling combined with mass spectrometric approaches as a tool for structural proteomics. *Expert Rev. Proteomics* **3**, 399–408 (2006).
225. Radicella, J. P., Dherin, C., Desmaze, C., Fox, M. S. & Boiteux, S. Cloning and characterization of hOGG1, a human homolog of the OGG1 gene of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **94**, 8010–8015 (1997).
226. Baumann, P. & West, S. C. DNA end-joining catalyzed by human cell-free extracts. *Proc. Natl. Acad. Sci.* **95**, 14066–14070 (1998).
227. Caruthers, M. H. & Beaucage, S. L. Phosphoramidite compounds and processes. (1983).
228. Beaucage, S. L. Oligodeoxyribonucleotides synthesis. Phosphoramidite approach. *Methods Mol. Biol. Clifton NJ* **20**, 33–61 (1993).
229. Wold, M. S. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66**, 61–92 (1997).
230. Smith, J., Zou, H. & Rothstein, R. Characterization of genetic interactions with RFA1: the role of RPA in DNA replication and telomere maintenance. *Biochimie* **82**, 71–78 (2000).
231. Salas, T. R. *et al.* Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic Acids Res.* **34**, 4857–4865 (2006).
232. Bochkareva, E., Korolev, S., Lees-Miller, S. P. & Bochkarev, A. Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J.* **21**, 1855–1863 (2002).
233. Bochkarev, A., Bochkareva, E., Frappier, L. & Edwards, A. M. The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. *EMBO J.* **18**, 4498–4504 (1999).

234. Salas, T. R., Petrusseva, I., Lavrik, O. & Saintomé, C. Evidence for direct contact between the RPA3 subunit of the human replication protein A and single-stranded DNA. *Nucleic Acids Res.* **37**, 38–46 (2009).
235. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* **385**, 176–181 (1997).
236. Doneanu, C. E., Gafken, P. R., Bennett, S. E. & Barofsky, D. F. Mass Spectrometry of UV-Cross-Linked Protein–Nucleic Acid Complexes: Identification of Amino Acid Residues in the Single-Stranded DNA-Binding Domain of Human Replication Protein A. *Anal. Chem.* **76**, 5667–5676 (2004).
237. Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5. *J. Gen. Virol.* **36**, 59–72 (1977).
238. Richard, D. J. *et al.* Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature* **453**, 677–681 (2008).
239. Feldkamp, M. D., Mason, A. C., Eichman, B. F. & Chazin, W. J. Structural Analysis of Replication Protein A Recruitment of the DNA Damage Response Protein SMARCAL1. *Biochemistry (Mosc.)* **53**, 3052–3061 (2014).
240. Audry, J. *et al.* RPA prevents G-rich structure formation at lagging-strand telomeres to allow maintenance of chromosome ends. *EMBO J.* **34**, 1942–1958 (2015).
241. Jönsson, K. I., Rabbow, E., Schill, R. O., Harms-Ringdahl, M. & Rettberg, P. Tardigrades survive exposure to space in low Earth orbit. *Curr. Biol.* **18**, R729–R731 (2008).

242. Lavrik, O. I. *et al.* Photoaffinity Labeling of Mouse Fibroblast Enzymes by a Base Excision Repair Intermediate EVIDENCE FOR THE ROLE OF POLY(ADP-RIBOSE) POLYMERASE-1 IN DNA REPAIR. *J. Biol. Chem.* **276**, 25541–25548 (2001).
243. Winnacker, M., Welzmler, V., Strasser, R. & Carell, T. Development of a DNA Photoaffinity Probe for the Analysis of 8-OxodG-Binding Proteins in a Human Proteome. *ChemBioChem* **11**, 1345–1349 (2010).
244. Jurkowska, R. Z. *et al.* Oligomerization and Binding of the Dnmt3a DNA Methyltransferase to Parallel DNA Molecules. *J. Biol. Chem.* **286**, 24200–24207 (2011).
245. Bhasin, M., Reinherz, E. L. & Reche, P. A. Recognition and classification of histones using support vector machine. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **13**, 102–112 (2006).
246. Camerini-Otero, R. D. & Felsenfeld, G. Histone H3 disulfide dimers and nucleosome structure. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5519–5523 (1977).
247. Maile, T. M. *et al.* Mass Spectrometric Quantification of Histone Post-translational Modifications by a Hybrid Chemical Labeling Method. *Mol. Cell. Proteomics* **14**, 1148–1158 (2015).
248. Ouvry-Patat, S. A. & Schey, K. L. Characterization of antimicrobial histone sequences and posttranslational modifications by mass spectrometry. *J. Mass Spectrom.* **42**, 664–674 (2007).
249. Garcia, B. A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat. Protoc.* **2**, 933–938 (2007).
250. Guitot, K. *et al.* Label-free measurement of histone lysine methyltransferases activity by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Biochem.* **456**, 25–31 (2014).

251. Bennett, S. E., Jensen, O. N., Barofsky, D. F. & Mosbaugh, D. W. UV-catalyzed cross-linking of Escherichia coli uracil-DNA glycosylase to DNA. Identification of amino acid residues in the single-stranded DNA binding site. *J. Biol. Chem.* **269**, 21870–21879 (1994).
252. Jensen, O. N., Kulkarni, S., Aldrich, J. V. & Barofsky, D. F. Characterization of Peptide-Oligonucleotide Heteroconjugates by Mass Spectrometry. *Nucleic Acids Res.* **24**, 3866–3872 (1996).
253. Urlaub, H., Hartmuth, K. & Lührmann, R. A two-tracked approach to analyze RNA–protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods* **26**, 170–181 (2002).
254. Jensen, O. N. *et al.* Direct observation of UV-crosslinked protein–nucleic acid complexes by matrix-assisted laser desorption ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **7**, 496–501 (1993).
255. Sastry, S. & Ross, B. M. Probing the Interaction of T7 RNA Polymerase with Promoter. *Biochemistry (Mosc.)* **38**, 4972–4981 (1999).
256. Wong, D. L. & Reich, N. O. Identification of Tyrosine 204 as the Photo-Cross-Linking Site in the DNA–EcoRI DNA Methyltransferase Complex by Electrospray Ionization Mass Spectrometry. *Biochemistry (Mosc.)* **39**, 15410–15417 (2000).
257. Steen, H., Petersen, J., Mann, M. & Jensen, O. N. Mass spectrometric analysis of a UV-cross-linked protein–DNA complex: Tryptophans 54 and 88 of E. coli SSB cross-link to DNA. *Protein Sci.* **10**, 1989–2001 (2001).
258. Schmidt, C., Kramer, K. & Urlaub, H. Investigation of protein–RNA interactions by mass spectrometry—Techniques and applications. *J. Proteomics* **75**, 3478–3494 (2012).
259. Huth-Fehre, T., Gosine, J. N., Wu, K. J. & Becker, C. H. Matrix-assisted laser desorption mass spectrometry of oligodeoxythymidylic acids. *Rapid Commun. Mass Spectrom.* **6**, 209–213 (1992).

260. Nordhoff, E. *et al.* Matrix-assisted laser desorption/ionization mass spectrometry of nucleic acids with wavelengths in the ultraviolet and infrared. *Rapid Commun. Mass Spectrom.* **6**, 771–776 (1992).
261. Hoffmann, E. de & Stroobant, V. *Mass Spectrometry: Principles and Applications.* (John Wiley & Sons, 2007).
262. Holz, B. *et al.* Identification of the Binding Site for the Extrahelical Target Base in N 6-Adenine DNA Methyltransferases by Photo-cross-linking with Duplex Oligodeoxyribonucleotides Containing 5-Iodouracil at the Target Position. *J. Biol. Chem.* **274**, 15066–15072 (1999).
263. Kühn-Hölsken, E., Dybkov, O., Sander, B., Lührmann, R. & Urlaub, H. Improved identification of enriched peptide–RNA cross-links from ribonucleoprotein particles (RNPs) by mass spectrometry. *Nucleic Acids Res.* **35**, e95–e95 (2007).
264. Spengler, B., Pan, Y., Cotter, R. J. & Kan, L.-S. Molecular weight determination of underivatized oligodeoxyribonucleotides by positive-ion matrix-assisted ultraviolet laser-desorption mass spectrometry. *Rapid Commun. Mass Spectrom.* **4**, 99–102 (1990).
265. Huth-Fehre, T., Gosine, J. N., Wu, K. J. & Becker, C. H. Matrix-assisted laser desorption mass spectrometry of oligodeoxythymidylic acids. *Rapid Commun. Mass Spectrom.* **6**, 209–213 (1992).
266. Parr, G. R., Fitzgerald, M. C. & Smith, L. M. Matrix-assisted laser desorption/ionization mass spectrometry of synthetic oligodeoxyribonucleotides. *Rapid Commun. Mass Spectrom.* **6**, 369–372 (1992).
267. Tang, K., Allman, S. L., Jones, R. B., Chen, C. H. & Araghi, S. Laser mass spectrometry of polydeoxyribothymidylic acid mixtures. *Rapid Commun. Mass Spectrom.* **7**, 63–66 (1993).

268. Pieves, U., Zürcher, W., Schär, M. & Moser, H. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic Acids Res.* **21**, 3191–3196 (1993).
269. Currie, G. J. & Yates, J. R. Analysis of oligodeoxynucleotides by negative-ion matrix-assisted laser desorption mass spectrometry. *J. Am. Soc. Mass Spectrom.* **4**, 955–963 (1993).
270. Schneider, K. & Chait, B. T. Matrix-assisted laser desorption mass spectrometry of homopolymer oligodeoxyribonucleotides. Influence of base composition on the mass spectrometric response. *Org. Mass Spectrom.* **28**, 1353–1361 (1993).
271. Zhu, L., Parr, G., Fitzgerald, M., Nelson, C. & Smith, L. Oligodeoxynucleotide Fragmentation in MALDI/TOF Mass Spectrometry Using 355-nm Radiation. *J. Am. Chem. Soc.* **117**, 6048–6056 (1995).
272. Murray, K. K. DNA Sequencing by Mass Spectrometry. *J. Mass Spectrom.* **31**, 1203–1215 (1996).
273. Wu, K. J., Shaler, T. A. & Becker, C. H. Time-of-Flight Mass Spectrometry Of Underivatized Single-Stranded DNA Oligomers by Matrix-Assisted Laser Desorption. *Anal. Chem.* **66**, 1637–1645 (1994).
274. Wu, K. J., Steding, A. & Becker, C. H. Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix. *Rapid Commun. Mass Spectrom.* **7**, 142–146 (1993).
275. Nordhoff, E. *et al.* Sample preparation protocols for MALDI-MS of peptides and oligonucleotides using prestructured sample supports. *Int. J. Mass Spectrom.* **226**, 163–180 (2003).

276. Pieleles, U., Zürcher, W., Schär, M. & Moser, H. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic Acids Res.* **21**, 3191–3196 (1993).
277. Flett, F. J., Walton, J. G. A., Mackay, C. L. & Interthal, H. Click Chemistry Generated Model DNA–Peptide Heteroconjugates as Tools for Mass Spectrometry. *Anal. Chem.* **87**, 9595–9599 (2015).
278. Catron, B., Caruso, J. A. & Limbach, P. A. Selective Detection of Peptide-Oligonucleotide Heteroconjugates Utilizing Capillary HPLC-ICPMS. *J. Am. Soc. Mass Spectrom.* **23**, 1053–1061 (2012).
279. Krivos, K. L. & Limbach, P. A. Sequence Analysis of Peptide:Oligonucleotide Heteroconjugates by Electron Capture Dissociation and Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* **21**, 1387–1397 (2010).
280. Catania, J., Keenan, B. C., Margison, G. P. & Fairweather, D. S. Determination of 5-methylcytosine by acid hydrolysis of DNA with hydrofluoric acid. *Anal. Biochem.* **167**, 347–351 (1987).
281. Sánchez-Arroyo, A. J. *et al.* Photochemical Oxidation of Thioketones by Singlet Molecular Oxygen Revisited: Insights into Photoproducts, Kinetics, and Reaction Mechanism. *J. Org. Chem.* **80**, 10575–10584 (2015).
282. Rizzi, V. *et al.* Rose Bengal-photosensitized oxidation of 4-thiothymidine in aqueous medium: evidence for the reaction of the nucleoside with singlet state oxygen. *Phys. Chem. Chem. Phys.* **17**, 26307–26319 (2015).
283. Nelson, C. C. & McCloskey, J. A. Collision-induced dissociation of uracil and its derivatives. *J. Am. Soc. Mass Spectrom.* **5**, 339–349 (1994).

284. Saintomé, C., Clivio, P., Favre, A. & Fourrey, J.-L. Photochemistry of 4-Thiothymine Derivatives in the Presence of N-9-Substituted-Adenine Derivatives: Formation of N-6-Formamidopyrimidines. *J. Org. Chem.* **62**, 8125–8130 (1997).
285. Li, W. & Liu, M. Distribution of 5-Hydroxymethylcytosine in Different Human Tissues. *Journal of Nucleic Acids* (2011). doi:10.4061/2011/870726
286. Münzel, M. *et al.* Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain. *Angew. Chem. Int. Ed.* **49**, 5375–5377 (2010).
287. Berthelot, V. *et al.* The human DNA ends proteome uncovers an unexpected entanglement of functional pathways. *Nucleic Acids Res.* **44**, 4721–4733 (2016).
288. Jutras, B. L., Verma, A. & Stevenson, B. Identification of novel DNA-binding proteins using DNA-affinity chromatography/pull down. *Curr. Protoc. Microbiol.* **Chapter 1**, Unit1F.1 (2012).
289. Wu, H. *et al.* Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* **25**, 679–684 (2011).
290. Publishers, B. S. *Combinational Chemistry & High Throughput Screening*. (Bentham Science Publishers, 1998).
291. Hirsch, J. D. *et al.* Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. *Anal. Biochem.* **308**, 343–357 (2002).
292. Vafabakhsh, R. & Ha, T. Extreme bendability of sub-100 bp long DNA revealed by single molecule cyclization. *Science* **337**, 1097–1101 (2012).
293. Hayashi, K. *et al.* Differential effects of retinoic acid on the growth of isogenic metastatic and non-metastatic breast cancer cell lines and their association with distinct expression of retinoic acid receptor beta isoforms 2 and 4. *Int. J. Oncol.* **22**, 623–629 (2003).

294. Xu, X. C. *et al.* Differential expression of nuclear retinoid receptors in normal, premalignant, and malignant head and neck tissues. *Cancer Res.* **54**, 3580–3587 (1994).
295. Pu, R. T. *et al.* Methylation profiling of benign and malignant breast lesions and its application to cytopathology. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* **16**, 1095–1101 (2003).
296. Feng, W. *et al.* Tumor suppressor genes are frequently methylated in lymph node metastases of breast cancers. *BMC Cancer* **10**, 378 (2010).
297. Widschwendter, M. *et al.* Methylation and Silencing of the Retinoic Acid Receptor- $\beta$ 2 Gene in Breast Cancer. *JNCI J. Natl. Cancer Inst.* **92**, 826–832 (2000).
298. Pouillet, P., Carpentier, S. & Barillot, E. myProMS, a web server for management and validation of mass spectrometry-based proteomic data. *PROTEOMICS* **7**, 2553–2556 (2007).
299. Valot, B., Langella, O., Nano, E. & Zivy, M. MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics* **11**, 3572–3577 (2011).
300. Bondarenko, P. V., Chelius, D. & Shaler, T. A. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **74**, 4741–4749 (2002).
301. Aravind, L., Anand, S. & Iyer, L. M. Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. *Biol. Direct* **8**, 20 (2013).
302. Li, Z., Park, Y. & Marcotte, E. M. A Bacteriophage Tailspike Domain Promotes Self-Cleavage of a Human Membrane-Bound Transcription Factor, the Myelin Regulatory Factor MYRF. *PLOS Biol.* **11**, e1001624 (2013).
303. Ngounou Wetie, A. G. *et al.* Investigation of stable and transient protein–protein interactions: Past, present, and future. *PROTEOMICS* **13**, 538–557 (2013).

304. Medvedeva, Y. A. *et al.* EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database J. Biol. Databases Curation* **2015**, (2015).
305. Siegfried, Z. & Simon, I. DNA methylation and gene expression. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**, 362–371 (2010).
306. Jones, P. L. & Wolffe, A. P. Relationships between chromatin organization and DNA methylation in determining gene expression. *Semin. Cancer Biol.* **9**, 339–347 (1999).
307. Heintzman, N. D. *et al.* Histone Modifications at Human Enhancers Reflect Global Cell Type-Specific Gene Expression. *Nature* **459**, 108–112 (2009).
308. de Souza-Pinto, N. C. *et al.* Repair of 8-oxodeoxyguanosine lesions in mitochondrial dna depends on the oxoguanine dna glycosylase (OGG1) gene and 8-oxoguanine accumulates in the mitochondrial dna of OGG1-defective mice. *Cancer Res.* **61**, 5378–5381 (2001).
309. Ohtsubo, T. *et al.* Identification of human MutY homolog (hMYH) as a repair enzyme for 2-hydroxyadenine in DNA and detection of multiple forms of hMYH located in nuclei and mitochondria. *Nucleic Acids Res.* **28**, 1355–1364 (2000).
310. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics MCP* **11**, O111.016717 (2012).
311. Huang, Q. *et al.* SWATH enables precise label-free quantification on proteome scale. *Proteomics* **15**, 1215–1223 (2015).
312. Gomes, X. V., Henricksen, L. A. & Wold, M. S. Proteolytic Mapping of Human Replication Protein A: Evidence for Multiple Structural Domains and a Conformational Change upon Interaction with Single-Stranded DNA†. *Biochemistry (Mosc.)* **35**, 5586–5595 (1996).

313. Gowher, H. & Jeltsch, A. Molecular Enzymology of the Catalytic Domains of the Dnmt3a and Dnmt3b DNA Methyltransferases. *J. Biol. Chem.* **277**, 20409–20414 (2002).
314. Audry, J. *et al.* RPA prevents G-rich structure formation at lagging-strand telomeres to allow maintenance of chromosome ends. *EMBO J.* n/a-n/a (2015). doi:10.15252/embj.201490773
315. Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R. & Roepstorff, P. Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.* **34**, 105–116 (1999).
316. Carda-Broch, S., Berthod, A. & Armstrong, D. W. Ionic matrices for matrix-assisted laser desorption/ionization time-of-flight detection of DNA oligomers. *Rapid Commun. Mass Spectrom.* **17**, 553–560 (2003).
317. Obino, D. *et al.* Actin nucleation at the centrosome controls lymphocyte polarity. *Nat. Commun.* **7**, ncomms10969 (2016).
318. Kowal, J. *et al.* Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc. Natl. Acad. Sci.* **113**, E968–E977 (2016).

## Table des illustrations

Figure 1 : Présentation des modifications covalentes de l'ADN. En gris sont représentées les bases de l'ADN provenant d'une synthèse chimique (d'après DNAmoD), on ne les retrouve pas naturellement. Les modifications de l'ADN sont ici représentées sous leurs formes abrégées. Graphique tiré de la base de donnée DNAmoD <sup>5</sup> .....	13
Figure 2: Mécanisme enzymatique de la méthylation de l'ADN par les DNMT (ENZ=enzyme).....	16
Figure 3: Schéma de la méthylation de novo et du maintien de la méthylation par les DNMT1, 3A et 3B d'après Jones et Liang <sup>29</sup> . Les boules grises représentent des motifs CpG non méthylés, les noires représentent des motifs CpG méthylés.....	17
Figure 4 : Schéma représentant le nouveau modèle de maintien de la méthylation de l'ADN formulé en 2009 par Jones et Liang <sup>29</sup> . a) Représentation de l'ADN après la méthylation de novo de l'ADN, les DNMT3 étant toujours fixées sur l'ADN au sein du nucléosome. b) Mécanisme de maintien de la méthylation comprenant PCNA, UHRF1 et DNMT1 sur de l'ADN hémiméthylé obtenu après réplication de l'ADN. c) Etape correctrice de méthylation des CpG oubliés lors de l'étape de maintien de la méthylation. ....	18
Figure 5: Les différents domaines de la protéine UHRF1 et leurs activités. D'après Sharif et al. <sup>37</sup> .....	19
Figure 6 : Modèle présentant le recrutement de DNMT1 par l'enzyme UHRF1 d'après Qin, W. et al <sup>47</sup> .....	20
Figure 7 : Schéma représentant les différentes voies de déméthylation de l'ADN d'après Wu et Zhang <sup>49</sup> .....	21
Figure 8 : Schéma des étapes du BER court (d'après une figure de Mariana RUIZ <sup>53</sup> ).....	23
Figure 9 : Réaction de désamination de la 5mC en T par l'AID/APOBEC.....	25
Figure 10 : Réaction de déméthylation de la 1mA et de la 3mC par AlkB .....	26
Figure 11 : Cycle catalytique caractéristique des dioxygénases à Fe (II) dépendantes du 2-oxoglutarate (2OG) adapté de Mantri et al. <sup>82</sup> pour l'oxydation de la 5mC en 5hmC. Les zones entourées en bleu correspondent aux acides aminés du site catalytique des enzymes TET. ....	28
Figure 12: Oxydations successives de la 5mC par les enzymes TET.....	29
Figure 13 : Représentation des trois mécanismes possibles d'élimination directe de la méthylation de l'ADN. ..	30
Figure 14 : Réaction de déshydroxyméthylation des 5hmC par les enzymes DNMT.....	31
Figure 15 : Voie de recyclage de la thymine impliquant des oxydations successives du méthyle de la thymine par la thymine-7-hydroxylase (T7H), Fe(II)/2-oxoglutarate dépendante, et une décarboxylation. ....	31
Figure 16: Rôle de NEIL1/2 d'après Schomacher et al. <sup>95</sup> .....	33
Figure 17 : Schéma bilan des voies de régulation de la méthylation de l'ADN. AM-PD correspond à la dilution passive de la méthylation dépendante de la réplication. AM-AR représente la voie de déméthylation se basant sur le BER. Les traits pleins correspondent aux mécanismes acceptés par la communauté scientifique, ceux en pointillés sont toujours discutés. ....	34
Figure 18 : Affinité des protéines MBD par rapport aux oligomères contenant de la 5mC et de la 5hmC adapté de Jin et al <sup>124</sup> . (A) Liaison de MBD2b, du complexe MBD2b / MBD3L1 (complexe MIRA), de MBD1 et MBD4 (100 ng de chaque protéine) avec des oligonucléotides 76-mères contenant des cytosines modifiées symétriquement, 5mC ou 5hmC dans des sites CpG. Les gels se situant à gauche correspondent à des expériences contrôles sans protéines. ....	38
Figure 19 : Affinité de TDG vis-à-vis d'oligonucléotides 23-mères contenant une cytosine modifiée. (a) EMSA d'une version inactive de TDG (hTDG <sup>cat(N140A)</sup> ) à deux concentrations différentes, et des oligonucléotides contenant soit une cytosine, soit une 5mC, soit une 5hmC (4 nM) ; (b) EMSA de la hTDG <sup>cat(N140A)</sup> avec un oligonucléotide contenant une 5fC ; (c) EMSA de la hTDG <sup>cat(N140A)</sup> avec un oligonucléotide contenant une 5caC. Figure adaptée des travaux de Zhang et al. <sup>126</sup> .....	39
Figure 20 ; Schéma représentant une purification d'affinité de type « DNA pulldown » d'après Iurlaro, M. et al <sup>4</sup> . L'oligonucléotide, possédant ou non des modifications épigénétiques, est fixé à l'aide de son groupement biotine sur des billes magnétiques comportant des streptavidines. Le déterminant d'affinité obtenu est incubé avec un extrait nucléaire, avant d'être lavé. Les protéines affines envers l'oligonucléotide sont ensuite éluées avant d'être analysées par spectrométrie de masse. ....	41
Figure 21 : Schéma représentatif de la méthode SILAC couplée à du pull-down utilisée par Spruijt et al. <sup>100</sup> . La culture cellulaire des cellules souches embryonnaires est réalisée dans deux milieux différents : un « léger » et un « lourd » composé de lysines et d'arginines comportant respectivement 8 et 10 <sup>13</sup> C. Les extraits cellulaires « légers » (jaune) et « lourds » (rouge) sont incubés respectivement avec un ADN non modifié et un ADN modifié au cours d'un pull-down, avant que les protéines enrichies ne soient mélangées avant d'être séparées à l'aide d'un gel d'électrophorèse et digérées à la trypsine pour être analysé par spectrométrie de masse. Le rapport d'intensité des pics correspondant à une même protéine entre sa version « légère » et sa version « lourde » permet de déterminer l'affinité de cette protéine pour la modification de l'ADN. ....	45
Figure 22 : Vue d'ensemble de toutes les protéines identifiées comme étant spécifiques de la 5mC ou de la 5hmC au sein de trois milieux cellulaires différents d'après Spruijt et al. <sup>100</sup> .....	47
Figure 23 : Réaction chimique ayant lieu durant le pontage de biomolécules par le formaldéhyde. ....	50

Figure 24: Groupements photoactivables les plus couramment utilisés: (a) azoture d'aryle; (b) diazirine; (c) benzophénone.....	52
Figure 25 : Principaux mécanismes des réactions des azotures d'aryles activés. Le chemin réactionnel encadré en bleu représente la réaction de pontage désirée. Les autres correspondent à des réactions secondaires. D'après Geurink et al. <sup>163</sup> .....	53
Figure 26 : Mécanismes réactionnels de l'excitation de la trifluorométhylaryldiazirine. Le chemin réactionnel encadré en bleu représente la réaction de pontage désirée. Les autres correspondent à des réactions secondaires. D'après Geurink et al. <sup>163</sup> .....	55
Figure 27 : Principaux mécanismes de réaction de la benzophénone irradiée. Le chemin réactionnel encadré en bleu représente la réaction de pontage désirée. Les autres correspondent à des réactions secondaires. D'après Dorman et al. <sup>173</sup> .....	57
Figure 28 : Groupement photoactivables puriques et pyrimidiques de type arylazides. ....	59
Figure 29 : Différents dérivés halogénés pyrimidiques.....	60
Figure 30 : Réactions photochimiques de la 5BrU avec des mimes peptidiques: i) de tryptophane, ii) de tyrosine et iii) d'histidine d'après Dietz et al <sup>185</sup> .....	61
Figure 31 : Structure des bases 4 <sup>S</sup> T et 6 <sup>S</sup> G.....	62
Figure 32 : Diagramme de Jablonsky de la 4 <sup>S</sup> T représentant ses différents niveaux énergétiques.D'après Harada et al <sup>195</sup> .....	62
Figure 33 : Structure des photoadduits obtenus après irradiation des dinucléotides 4 <sup>S</sup> U-T (haut) et T-4 <sup>S</sup> U (bas).Figure adaptée de Favre et al. <sup>197</sup> .....	64
Figure 34 : Photoréaction de la 1,3-diméthyle-4-thiouracile avec la lysine d'après Ito et al <sup>199</sup> . Première (A) et seconde (B) chimiosélectivité mécanistique. ....	65
Figure 35 : Synthèse du triphosphate U <sup>DZ</sup> TP d'après Smith et al <sup>202</sup> . Les réactifs et les conditions sont les suivantes : (i) 4-pentyn-1-ol, NaH, THF, 0 °C, 2 h, 68%; (ii) Pd(PPh <sub>3</sub> ) <sub>2</sub> Cl <sub>2</sub> , Et <sub>3</sub> N, CuI, DMF, r.t., o.n., 84%. (iii) Ac <sub>2</sub> O, pyridine, DMAP, 0 °C to r.t., 2 h, 71%; (iv) 1% DCAA, CH <sub>2</sub> Cl <sub>2</sub> , 0 °C, 1.5 h, 48%; (v) (1) 2-chloro-1,3,2-benzodioxaphosphorin-4-one, pyridine, dioxane; (2) (nBu <sub>3</sub> NH) <sub>2</sub> H <sub>2</sub> P <sub>2</sub> O <sub>7</sub> , nBu <sub>3</sub> N, DMF; (3) I <sub>2</sub> , pyridine, H <sub>2</sub> O; (4) NH <sub>3</sub> (aq.), H <sub>2</sub> O, 21% (sur les 4 dernières étapes) .....	66
Figure 36: Principe de l'introduction d'un nucléoside photoactivable (PA-(d)NTP) à une position spécifique dans l'ADN .....	67
Figure 37 : Exemple de phosphoramidites fonctionnalisés par un groupement photoactivable au niveau du sucre (A) et au niveau de la base (B). D'après Leibmann et al <sup>205</sup> . et Smith et al. <sup>202</sup> .....	68
Figure 38 : Fonctionnalisation post-synthétique d'une molécule d'ADN halogénée par un groupement possédant une benzophénone selon un couplage de Suzuki-Miyaura. D'après Lercher et al. ....	69
Figure 39 : Fonctionnalisation post-synthétique de la 2-fluoro-2'-désoxyinosine et de la O6-triazolyl-2'-désoxyuridine par la méthode du nucléoside convertible par des diazirines aliphatiques adapté des travaux de Shigdel et al <sup>220</sup> .....	70
Figure 40 : Principe du photomarquage d'affinité appliqué à l'identification des protéines partenaires de la sonde (voie A), ici ADN, et à l'identification du site de liaison (voie B), par spectrométrie de masse. ....	71
Figure 41 : Schéma récapitulatif du projet de thèse. ....	75
Figure 42 : Séquence utilisée pour les expériences de photomarquage avec en vert le motif CpG et en rouge, les deux positions d'incorporation des agents photoactivables. ....	77
Figure 43 : Structure d'un synthon phosphoramidite, des différentes bases modifiées, des groupes protecteurs et structure de l'étiquette biotine utilisés dans la synthèse sur support solide .....	79
Figure 44: Différentes étapes de la synthèse automatique sur phase solide d'oligonucléotides.....	80
Figure 45: Synthèse des amines comportant un groupement photoactivable. ....	82
Figure 46 : Introduction de groupement photoactivable benzophénone et diazirine par la méthode du nucléoside convertible.....	82
Figure 47 : Caractérisation de l'ODN-Dia par spectrométrie de masse MALDI-TOF, en mode linéaire, ions négatifs, matrice 3-HPA, 10 pmol d'ADN.....	83
Figure 48 : Schéma des différentes sondes ODN photoactivables synthétisées. La position et la structure des modifications introduites dans les oligonucléotides sont représentées par des traits. ....	84
Figure 49: Schéma de la protéine hRPA, ses différentes sous-unités et ses domaines structuraux et fonctionnels. ....	85
Figure 50 : A. Analyse par EMSA de la liaison entre hRPA purifiée et les ODN simple brin photoactivables synthétisés. La représentation schématique de l'ODN marqué (___*) est indiquée à droite du gel et les trois ovales représentent les trois sous-unités de hRPA liées à l'ODN : RPA1 (gris clair), RPA2 (gris foncé) RPA3 (noir). B. Analyse quantitative des gels ci-dessus. Les expériences ont été reproduites deux fois pour déterminer la C50, concentration en hRPA pour laquelle 50 % du complexe est formé. ....	86
Figure 51 :A. Analyse par gel dénaturant de la cinétique de photomarquage entre hRPA purifiée et les différentes sondes ODN après différents temps d'irradiation. La représentation schématique de l'ODN marqué (--*) est	

indiquée à droite du gel et les ovales représentent hRPA1 (gris clair) ou hRPA2 (gris foncé) lié covalamment à l'ODN marqué. B. Analyse quantitative des gels ci-dessus.....	88
Figure 52 : A) Gel d'électrophorèse dénaturant SDS-PAGE de l'expérience de photomarquage entre hRPA purifiée et l'ODN-Dia (piste 1), coloré au bleu de Coomassie. La piste 2 représente un contrôle, et correspond à la protéine hRPA purifiée. B) Grossissement de la piste 1 de la Figure 52A entre 60 et 100 kDa, zone où se trouvent les complexes potentiels ODN-Dia hRPA1. A, B, C, D et E correspondent à la position des zones excisées pour la digestion à la trypsine et à l'identification par spectrométrie de masse MALDI-TOF / TOF. C) Tableau correspondant aux scores et à la couverture de séquences obtenus avec Mascot après analyse MALDI-TOF / TOF et une recherche en base de données (Swissprot) pour l'identification de la protéine hRPA1 dans les bandes A à E. ....	90
Figure 53 : Schéma de deux réticulations possibles d'un complexe ODN-Dia-hRPA1 après dénaturation.....	92
Figure 54 : A) Agrandissement de la piste 1 de la Figure 52A entre 60 et 100 kDa, zone où se trouvent les potentiels complexes ODN-Dia-hRPA1. A, B, C, D et E correspondent à la position des bandes excisées pour la digestion à la trypsine et à l'identification par spectrométrie de masse MALDI-TOF / TOF. B) Séquence d'acide aminés de hRPA1 où sont représentées les couvertures de séquence identifiées en spectrométrie de masse de la bande A (en gras), de la bande B (souligné), de la bande C (surligné gris) et de la bande D (souligné vert). C) Représentation 3D de hRPA1 lié à un ADN simple brin (PBD : 1JMC) avec en rouge les régions d'interaction avec l'ADN d'après Bochkarev et al. <sup>235</sup> .....	93
Figure 55 : A. Gel d'électrophorèse SDS-PAGE d'un échantillon de lysat cellulaire obtenu à partir d'E. coli surexprimant hRPA (coloration au bleu de Coomassie). B. Analyse par gel d'électrophorèse SDS-PAGE des sondes ODN après irradiation en présence d'un lysat cellulaire surexprimant hRPA. La représentation schématique de l'ODN marqué ( _*) est indiquée à droite du gel et les ovales représentent hRPA1 (gris clair) ou hRPA2 (gris foncé) liées covalamment à l'ODN marqué.....	95
Figure 56 : Analyse par Western Blot des complexes ODN-4 <sup>ST</sup> -hRPA1 obtenus par photomarquage de l'ODN-4 <sup>ST</sup> en présence de lysat cellulaire d'E. coli surexprimant hRPA. ....	96
Figure 57 : Principe de l'enrichissement des complexes ODN-protéines par purification d'affinité biotine-streptavidine. Lors de l'étape a) la protéine hRPA est représentée avec ses sous unités séparés dans le milieu pour faciliter la compréhension du schéma. ....	97
Figure 58 : A) Gel d'électrophorèse dénaturant SDS-PAGE de l'expérience de photomarquage entre l'ODN-4 <sup>ST</sup> et le lysat cellulaire d'E. coli surexprimant hRPA (coloration au bleu de Coomassie). La piste 3 représente une expérience contrôle, et correspond à la protéine hRPA purifiée. a, b, c et d correspondent à la position des zones excisées pour la digestion à la trypsine et à l'identification par spectrométrie de masse MALDI-TOF/TOF. B) Agrandissement des pistes 2 et 2' entre 70 et environ 140 kDa, où sont représentées les zones excisées. C) Tableau correspondant aux scores Mascot et à la couverture de séquence après l'analyse MALDI-TOF/TOF et une recherche en base de données pour l'identification de la protéine hRPA1 dans les bandes a à d, sans (a, b, c et d) et avec (a', b', c' et d') irradiation. ....	98
Figure 59 : A. Gel d'électrophorèse dénaturant 15 % où les protéines présentes dans le lysat cellulaire obtenu à partir des cellules HEK 293 sont colorées au bleu de Coomassie. B. Analyse par gel d'électrophorèse SDS-PAGE de la sonde ODN-4 <sup>ST</sup> après irradiation en présence d'un lysat cellulaire de cellules HEK 293 (autoradiographie). La représentation schématique de l'ODN marqué ( _*) est indiquée à droite du gel et les ovales représentent hRPA1 (gris clair) ou hRPA2 (gris foncé) lié covalamment à l'ODN marqué. Les cadres noirs signifient que les images proviennent de gels différents mais réalisés dans les mêmes conditions .....	100
Figure 60 : Analyse des complexes ODN-4 <sup>ST</sup> -hRPA1 obtenu par photomarquage de l'ODN-4 <sup>ST</sup> simple brin (A) et double brin (B) et du lysat cellulaire de cellule HEK 293 par Western Blot avec un anticorps dirigé contre hRPA1.....	101
Figure 61 : Analyse du gel dénaturant de l'expérience de photomarquage entre l'ODN-Dia et un lysat cellulaire d'E.coli (piste 1), et un lysat cellulaire d'E.coli surexprimant hRPA (piste 2) et un lysat cellulaire d'E.coli surexprimant hRP-D228Y (piste 3).....	103
Figure 62 : Analyse par gel d'électrophorèse dénaturant 15 % de la capacité des sondes ODN à lier DNMT3a purifiée. La représentation schématique de l'ODN marqué ( _*) est indiquée à droite du gel et le cercle vert représente DNMT3a liée covalamment à l'ODN marqué. La flèche orange pointe la bande correspondant à l'ODN-4 <sup>ST</sup> n'ayant pas réagi. ....	106
Figure 63 : Analyse par gel d'électrophorèse dénaturant 17% du photomarquage entre les sondes duplexes ODN et H3 purifiée (autoradiographie). La représentation schématique de l'ODN marqué (-*) est indiquée à droite du gel, le demi-cercle bleu représente H3 lié covalamment à l'ODN marqué, la flèche grise représente les puits, la flèche verte le stacking et la flèche bleu le complexe d'intérêt.....	108
Figure 64 : Gel d'électrophorèse dénaturant SDS-PAGE de l'expérience de photomarquage du duplex ODN-4 <sup>ST</sup> avec H3 (coloration au bleu de Coomassie). La bande correspondant à la masse moléculaire de H3 est indiquée, ainsi que celle pouvant correspondre à un dimère d'H3. Les pistes 1 et 3 représentent des expériences contrôles avec la protéine H3 purifiée ayant (piste 3) ou non (piste 1) suivi les mêmes conditions d'expériences. La piste 4	

correspond à la protéine H3 irradiée. A, B, C, D, E, F, G et H correspondent à la position des zones excisées pour la digestion à la trypsine et à l'identification par spectrométrie de masse MALDI-TOF / TOF. Le patron d'excision des zones est uniquement représenté sur la piste 1 pour plus de clarté. ....	109
Figure 65 : Réaction de propionylation de la chaîne latérale d'une lysine.....	110
Figure 66 : Représentation de la séquence d'acides aminés de la protéine H3.1 et de ses sites de coupure par la trypsine (A) et par l'ArgC (coupure similaire à la propionylation couplée à la trypsine) (B) d'après le logiciel ExPASy.....	112
Figure 67 : A) Identification de la protéine H3 par spectrométrie de masse sur le gel présenté sur la Figure 64. Les zones vertes correspondent à une identification de la protéine H3 avec des scores Mascot élevés, les zones jaunes à une identification plus difficile et les zones blanches correspondent à des bandes de gel dans lesquelles la protéine H3 n'est pas identifiée. Les zones hachurées correspondent aux zones où de l'ADN a été détecté. B) Gel dénaturant correspondant au radiomarquage des digestats provenant des zones A, B et C de la piste 2' de la Figure 64. ....	114
Figure 68 : A. Gel d'électrophorèse dénaturant 15 % où les protéines présentes dans le lysat cellulaire obtenu à partir des cellules HEK 293 sont colorées au bleu de Coomassie. B. Analyse par gel d'électrophorèse SDS-PAGE 15% de la sonde ODN-4 <sup>ST</sup> après irradiation en présence d'un lysat cellulaire de cellules HEK 293 (autoradiographie). La représentation schématique de l'ODN marqué (-*) est indiquée à droite du gel et le demi-cercle représente H3 (bleu) lié covalamment à l'ODN marqué. ....	115
Figure 69 : Structure de l'acide 3-hydroxypicolinique (3-HPA) utilisé comme matrice MALDI .....	122
Figure 70 : Formule développée du citrate de diammonium.....	123
Figure 71 : Dépôt MALDI présentant des cristaux de 3HPA en présence de billes Dowex (flèche blanche) ....	125
Figure 72 : A, B, C et D) Etapes de l'amorçage de la cristallisation ; E) exemple de cristal amorphe non activé. ....	126
Figure 73 : Spectres MALDI-TOF en mode linéaire, ions négatifs de l'ODN 31 T de masse moléculaire théorique moyenne MM = 9368,02 Da, étalonnage par défaut ; A) mélange 9:1 3HPA : l'ODN 31 T (15 à 20 pmol) avec citrate de diammonium ; B) mélange 9:1 3HPA : l'ODN 31 T (15 à 20 pmol) avec citrate de diammonium et amorçage de la cristallisation ; C) mélange 9:1 3HPA : l'ODN 31 T (15 à 20 pmol) avec citrate de diammonium, amorçage de la cristallisation dans les conditions d'analyse optimisées.....	127
Figure 74 : Spectres MALDI-TOF en mode linéaire, ions négatifs de l'oligonucléotide 4 <sup>ST</sup> de masse moléculaire théorique moyenne MM = 9655,70 Da, étalonnage réalisé avec l'oligonucléotide ODN 31 T ; A) mélange 9:1 3HPA : ODN-4 <sup>ST</sup> (15 à 20 pmol) avec citrate de diammonium ; B) mélange 9:1 3-HPA : ODN-4 <sup>ST</sup> avec citrate de diammonium, amorçage de la cristallisation dans les conditions d'analyse optimisées. ....	129
Figure 75 : Spectres MALDI-TOF en mode linéaire, ions négatifs, mélange 9:1 3-HPA : Echantillon (plus de 10 pmol de sonde ADN estimée) avec citrate de diammonium amorçage de la cristallisation dans les conditions d'analyse optimisées, étalonnage réalisé avec l'ODN 31 T. Analyse des complexes ODN 4 <sup>ST</sup> -hRPA1, obtenus après digestion sur gel à la trypsine, zone du gel correspondant à des hétéroconjugués potentiels avec la sous-unité hRPA1, c'est-à-dire dans la zone de MM supérieure à 70 kDa sans (A) et avec irradiation (B). ....	131
Figure 76 : Séquences de la protéine histone humaine H3 et du peptide ARTK. ARTK est biotinylé sur la fonction amine primaire de la chaîne latérale de la lysine en position 18. ....	133
Figure 77 : Autoradiographie du gel SDS-PAGE résultant de l'expérience de photomarquage entre l'ODN-4 <sup>ST</sup> radiomarqué et le peptide ARTK sans (piste de gauche) et avec irradiation (piste de droite).....	133
Figure 78 : Spectres MALDI-TOF en mode linéaire ions négatif du mélange ADN 4 <sup>ST</sup> , peptide ARTK dans un ratio 1 pour 10 sans (A) ou avec irradiation (B). La matrice utilisée est 3-HPA dans nos conditions optimisées. Chaque spectre correspond à l'analyse d'au minimum 10 pmol d'ADN totale (lié et non lié). ....	135
Figure 79 : Spectre MALDI-TOF en mode linéaire ions négatif du mélange ODN-Dia, peptide ARTK dans un ratio 1 pour 10, irradié. La matrice utilisée est 3-HPA dans nos conditions optimisées. Le spectre correspond à l'analyse d'au minimum 10 pmol d'ADN totale (lié et non lié). ....	136
Figure 80 : Spectres MALDI-TOF en mode réflecteur, ions positifs, matrice HCCA (ratio 1 : 1, matrice : échantillon) des digestes à la benzonase des mélanges ODN 4 <sup>ST</sup> /peptide ARTK sans (A) ou avec irradiation (B). ....	137
Figure 81 : Agrandissement du signal correspondant à l'ODN 4 <sup>ST</sup> libre (A) et du signal correspondant au photoadduit ODN 4 <sup>ST</sup> -ARTK (B) du spectre MALDI-TOF en mode linéaire ions négatif du mélange ODN 4 <sup>ST</sup> , peptide ARTK dans un ratio 1 pour 10, irradié (Figure 78B). Les traits en pointillé permettent de faire la corrélation entre les deux signaux. ....	139
Figure 82 : Spectres MALDI-TOF mode linéaire ions négatifs avec la matrice 3-HPA dans nos conditions optimisées, représentant différentes expériences de photomarquage : A) Mélange ODN 4 <sup>ST</sup> /peptide ARTK sans irradiation ; B) Mélange ODN 4 <sup>ST</sup> /peptide ARTK avec irradiation ; C) Mélange ODN T/ peptide ARTK avec irradiation .....	140
Figure 83 : Mécanisme analogue à celui décrit par Sanchez-Arroyo et al. pour la réaction d'un singulet <sup>1</sup> O <sub>2</sub> sur un fonction thiocétone appliqué à la 4-thiothymine. ....	141

Figure 84 : Réaction d'une molécule d'H <sub>2</sub> O sur la 4-thiothymine excitée.....	141
Figure 85 : Séquence de l'oligonucléotide contenant un 4-thiothymine, dont ces proches voisins sont soulignés. .....	142
Figure 86 : Réaction entre la N <sup>3</sup> -méthyle-4-thiothymine et une adénosine sous irradiation qui entraîne la perte d'un groupement H <sub>2</sub> S pour former une adénosine N <sup>6</sup> -formamidopyrimidine. D'après Saintomé et al.....	142
Figure 87 : A) Proposition de mécanisme entre la 4-thiothymine et une guanine (G-4 <sup>ST</sup> ) ou une adénine 4 <sup>ST</sup> -A conduisant à une perte de 34u. B) Modèle en 3D du tri-nucléotide G-4 <sup>ST</sup> -A où sont représentées les distances entre l'amine de la guanine et la thiocétone de la 4 <sup>ST</sup> et l'amine de l'adénine et la thiocétone de la 4 <sup>ST</sup> .....	143
Figure 88 : Spectres MALDI-TOF mode linéaire ions négatifs avec la matrice 3-HPA dans nos conditions optimisées, représentant différentes expériences de photomarquage entre l'ADN-4 <sup>ST</sup> et de la Fmoc-Lys-OH (MM exacte : 366 u.). A) Expérience contrôle sans irradiation ; B) Expérience avec irradiation ; C) Agrandissement de la zone 9489 à 10267 m/z du spectre B correspond aux signaux de l'ODN-4 <sup>ST</sup> et du photoadduit ODN 4 <sup>ST</sup> - Fmoc-Lys-OH. ....	145
Figure 89 : Schéma du mécanisme proposé de la photochimie de la 4 <sup>ST</sup> vis-à-vis d'une lysine au sein d'un peptide. .....	147
Figure 90 : Photomarquage d'extraits nucléaires de cellules HeLa par les sondes ODN-Dia/C et ODN-Dia/5hmC et séparation sur gel dénaturant d'acrylamide. ....	150
Figure 91 : Quantification des bases 5hmC (a) et 5mC (b) dans l'ADN génomique isolé à partir de tissus humains. Les données sont des valeurs moyennes ± écart-type de 3 essais différents d'après Li et Liu (2011). ....	151
Figure 92 : A) Schéma des ODN-Dia/C, ODN-Dia/5mC, ODN-Dia/5hmC et ODN-Ctr/5hmC. B) Photomarquage d'extraits nucléaires de cerveau de rat par les sondes ODN-Dia/C, ODN-Dia/5mC, ODN-Dia/5hmC et ODN-G/5hmC et séparation sur gel dénaturant d'acrylamide. La représentation schématique de l'ODN marqué (-*) est indiquée à droite du gel. ....	152
Figure 93: Schéma de la stratégie mise en place par Berthelot et al. pour identifier les protéines spécifiques des extrémités de l'ADN. La phase chromatographique du haut présente des extrémités d'ADN libres ; la phase du bas présente le système dit « boucle ». Après incubation des phases chromatographiques avec des extraits nucléaires de cellules HeLa, celles -ci sont lavées. Les complexes nucléoprotéiques sont ensuite détachés des billes par irradiation du lien photoclivable représenté en rose. Le matériel purifié est déposé sur un gel d'électrophorèse et les bandes résultantes sont excisées et leurs contenus identifiés par LC-MS/MS après digestion trypsique. Figure adaptée de Berthelot et al. ....	156
Figure 94 : Principe de l'expérience de pull-down « boucle » couplée à du photomarquage. ....	158
Figure 95 : Séquence d'ADN de la région promotrice du gène de RARβ du nucléotide -123 au nucléotide 278. Le site de début de transcription est indiqué par une flèche noire. Les deux éléments RAR et la TATA box sont soulignés. Les positions des 5mC sont indiquées par des cercles pleins ou des cercles avec une étoile, ce qui correspond respectivement à une détection dans une ou deux lignées cellulaires. Les deux encadrés orange ainsi que le surlignage de la même couleur correspondent à notre séquence d'étude. Les encadrés représentent les amorces que nous avons utilisés pour isoler et amplifier cette séquence appeler tRARβ, pour template RARβ. D'après Widschwendtet et al. ....	159
Figure 96 : Détermination expérimentale de la température d'appariement optimale en utilisant un gradient de 47 à 67 ° C. Présentation des résultats sous forme de gel d'agarose 1%.....	160
Figure 97 : Représentation des trois nucléosides triphosphates incorporés au sein de notre déterminant d'affinité. De gauche à droite, la forme triphosphate de la 5mC, celle de 5hmC et de la 4 <sup>ST</sup> . ....	161
Figure 98 : Schéma de déterminants d'affinité sous leurs formes biotinylées (A) ou desthiobiotinylées (B).....	162
Figure 99: Représentation schématique des phases chromatographiques utilisées pour le pull-down.....	167
Figure 100 : Représentation d'un chromatogramme correspondant à un peptide extrait par la méthode XIC (B) par rapport au chromatogramme correspondant à l'analyse de l'échantillon totale. ....	169
Figure 101 : Volcano plot représentant la quantification des protéines enrichies par la phase chromatographiques CDB par rapport aux protéines enrichies par la phase chromatographique Billes. ....	170
Figure 102 : Résultat de l'enrichissement des annotations GO effectuées sur les protéines enrichies par la phase CDB par rapport à la phase Billes. La valeur de la p-value correspond à la probabilité que la fonction décrite ne soit pas effectuée. Plus la p-value est basse, plus la fonction moléculaire est représentée. ....	171
Figure 103 : Représentation schématique de l'analyse des données obtenues par spectrométrie de masse dans le but d'identifier les protéines spécifiques de la 5mC et/ou de la 5hmC. Les ovales correspondent aux expériences de pull-down réalisées avec les différentes phases chromatographiques. ....	171
Figure 104 : Volcano plot représentant la quantification des protéines enrichies par le pull-down mCDB par rapport aux protéines enrichies par le pull-down CBD. La liste des protéines enrichies par la phase mCDB ou CBD sont représentées sur chaque côté du graphique. Les protéines qui ont déjà été caractérisées comme étant affines pour la 5mC et/ou la 5hmC ont été hachurées sur les listes. ....	173
Figure 105 : Volcano plot représentant la quantification des protéines enrichies par le pull down HDB par rapport aux protéines enrichies par le pull-down CBD. La liste des protéines enrichies par la phase HB ou CBD sont	

représentées sur chaque côté du graphique. Les protéines qui ont déjà été caractérisées comme étant affines pour la 5mC et/ou la 5hmC ont été hachurées sur les listes. ....	175
Figure 106 : Volcano plot représentant la quantification des 56 protéines affines pour la 5mC et pour la 5hmC par rapport à C, par le pull-down HDB par rapport pull-down mCBD. La liste des protéines enrichies ou diminuées par la phase HDB par rapport à la phase mCDB sont représentées sur chaque côté du graphique. Les protéines qui ont déjà été caractérisées comme étant affines pour des modifications épigénétiques de la cytosine ont été hachurées sur les listes. ....	176
Figure 107 : Représentation schématique des protéines spécifiques de la 5mC et de la 5hmC dont la quantité est enrichie en présence de ces modifications. ....	177
Figure 108 : Volcano plot représentant la quantification des protéines enrichies par la phase chromatographiques C4DB par rapport aux protéines enrichies par la phase chromatographique CDB. ....	179
Figure 109 : Représentation schématique des protéines transitoires spécifiques de la 5mC et de la 5mC déterminées par photomarquage liée au pull-down. ....	181
Figure 110 : Schéma explicatif des résultats obtenus en comparant les pull-down avec et sans photomarquage dans le cas de protéines non transitoires et transitoires. Dans le cas du pull-down classique, les protéines non transitoires parcourent l'ADN ou s'y fixent, alors que les protéines transitoires vont être en échange entre le milieu et l'ADN. En effet, elles vont parcourir l'ADN, reconnaître une modification, effectuer leur rôle enzymatique et se détacher de l'ADN. La photomarquage permet de piéger ces deux classes de protéines sur l'ADN. ....	182
Figure 111 : Schéma bilan de toutes les protéines interagissant avec la 5mC ou le 5hmC. Les protéines surlignées en orange ont été identifiées comme des protéines transitoires. Les fonctions de toutes les protéines ont été répertoriées et sont représentées dans les encadrés bleus. ....	184

## Table des tableaux

Tableau 1 : Température de fusion de différents duplexes d'ADN dont un brin contient un groupement photoactivable. ....	105
Tableau 2: Matrices communément utilisés en analyse MALDI de protéines, peptides et oligonucléotides d'après Jensen et al, 1996 <sup>252</sup> .....	120
<i>Tableau 3 : tableau récapitulatif du nombre de protéines identifiées par spectrométrie de masse ainsi que du nombre de peptides ayant permis l'identification de la streptavidine pour les 12 expériences réalisées dans le but d'optimiser les conditions d'éluion de notre système. ....</i>	<i>165</i>
Tableau 4 : Tableau présentant les protéines dont le FC>1,5 ; p-value<0,05 lors que la quantification des protéines résultants du pull-down mC4DB par rapport à mCDB et du pull-down H4DB par rapport à HDB sur les 92 protéines obtenues par l'analyse protéomique du pull-down classique.....	180

Au cours des dernières décennies, la méthylation de l'ADN en position 5 de la cytosine est apparue comme une importante modification épigénétique qui joue un rôle essentiel dans le contrôle spécifique de l'expression des gènes. Cependant, les mécanismes impliqués dans la régulation de la méthylation de l'ADN restent incompris. Des études récentes ont montré que des protéines de type oxydases, nommées TET, peuvent catalyser l'oxydation de la 5-méthylcytosine (5mC) et générer des dérivés oxydés de celle-ci ce qui soulève la question du rôle biologique des formes oxydées de la 5mC. L'identification et la caractérisation des protéines interagissant avec ces formes oxydées devraient permettre une meilleure compréhension de la fonction de ces modifications de l'ADN et de la régulation de la méthylation de l'ADN.

Dans ce projet, nous avons développé des sondes photoactivables basées sur l'ADN pour capturer, isoler et caractériser les protéines associées à ces modifications épigénétiques de l'ADN. Tout d'abord, nous avons conçu et évalué les propriétés de différentes sondes oligonucléotidiques photoactivables. Nous avons ensuite réalisé une étude méthodologique afin de caractériser au niveau moléculaire les photoadduits obtenus par MALDI-TOF. Enfin, nous avons développé une méthode de pull-down couplé à du photomarquage et associée à une analyse protéomique par spectrométrie de masse afin d'identifier les protéines ayant une affinité spécifique pour ces modifications épigénétiques.

**Mots clés :** Photomarquage, Spectrométrie de masse, Epigénétique, Modifications de l'ADN, Pull-down

**Photoaffinity labeling coupled with mass spectrometry to identify epigenetics modifications proteins partners.**

Over the past few decades, DNA methylation at the 5-position of cytosine has emerged as an important epigenetic modification that plays essential roles in the specific control of gene expression. However, the mechanisms involved in the regulation of DNA methylation remain unclear. Recent studies have shown that oxidase proteins, called TETs, can catalyze the oxidation of 5-methylcytosine (5 mC) and generate oxidized derivatives thereof, raising the question of the biological role of the oxidized forms of 5mC. The identification and characterization of proteins interacting with these oxidized forms should allow for a better understanding of the function of these DNA modifications and the regulation of DNA methylation.

In this project, we develop DNA-based photoactivatable probes to capture, isolate and characterize the proteins associated with these epigenetic DNA modifications. First, we designed and evaluated the properties of different photoactivatable oligonucleotide probes. We then carried out a methodological study in order to characterize at the molecular level the obtained photoadducts by MALDI-TOF. Finally, we developed a pull-down method coupled to photolabelling and associated with proteomic analysis by mass spectrometry to identify proteins with specific affinity for these epigenetic changes.

**Keywords:** Photocrosslink, Mass spectrometry, Epigenetic, DNA modifications, Pull-down