

## **Thèse**

présentée en vue de l'obtention du titre de

### **Docteur de l'Université de Strasbourg**

Discipline : Sciences du vivant

Mention : Biochimie et biophysique structurales

par

**Nada LALLOUS**

### **Etude structurale et fonctionnelle des modules de reconnaissance des marques épigénétiques dans la protéine humaine UHRF1**

Présentée le 06 janvier 2010 devant le jury composé de :

Dr Jean-Pierre SAMAMA	Directeur de thèse
Dr Catherine BIRCK	Encadrant
Dr Yves BOURNE	Rapporteur externe
Dr Matthieu REFREGIERS	Rapporteur externe
Pr Bruno KIEFFER	Rapporteur interne
Pr Pierre OUDET	Membre invité

Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)



« Vous êtes une flûte à travers le cœur de laquelle le murmure des heures se transforme en musique. Aimer la vie à travers le travail, c'est être intime avec le secret le plus intérieur de la vie. Tout travail est vide sauf quand l'amour est là, parce que le travail, c'est l'amour rendu visible . »

Kahlil Gibran

A mes parents Antoinette et Jihad

A mes frères Fadi et Salim

A mes grands-parents



## Remerciements

Cette thèse a été réalisée dans l'Institut de Génomique et de Biologie Moléculaire et Cellulaire. Je tiens à remercier Dr. Dino MORAS et Dr. Patrick SCHULTZ de m'avoir donné l'opportunité de travailler dans cet institut renommé et au sein du département de Biologie et Génomique Structurale.

Je tiens à remercier mon directeur de thèse Jean-Pierre SAMAMA de m'avoir chaleureusement accueillie au sein de son équipe pour mon stage de Master 2 et qui m'a initié à la biologie structurale et m'a donné l'occasion de travailler sur un projet si passionnant. Je n'oublierai jamais ton mélange Arak-café ☺.

Je suis également très reconnaissante envers Catherine BIRCK qui a dirigé ce travail durant les deux dernières années. L'aboutissement de ce travail doit beaucoup à ta disponibilité, ta rigueur scientifique et aux nombreuses discussions que nous avons partagées durant ces quelques années. Avec toi, j'ai découvert le monde de la biophysique et je te contacterai sûrement pour te poser des questions sur Sedfit, sedphat, BUNCH ...☺.

J'exprime mes remerciements au Professeur Bruno KIEFFER qui a présidé le jury lors de la soutenance mais aussi aux Docteurs Yves BOURNE et Matthieu REFREGIERS, et au Professeur Pierre OUDET pour l'honneur qu'ils m'ont accordé en acceptant de juger et d'examiner ce travail et pour les discussions et les ouvertures scientifiques que nous avons partagées durant la soutenance.

Mes sincères remerciements vont à Bénédicte DELAGOUTTE qui a accompagné mes premiers pas dans le monde de la recherche. Ton soutien scientifique et moral m'a été d'une grande importance. Mes meilleurs vœux pour le nouveau chemin où tu t'es engagée.

Nicolas HAUMESSER, un très grand merci pour tout le support technique et les conseils quotidiens que tu m'as accordés durant mon stage de Master et ma thèse. Si un jour tu as besoin d'un coup de main pour aliquoter des bactéries compétentes, n'hésite pas à m'appeler.

Ally, même si nous n'avons pas pu convaincre Jean-Pierre de nous acheter une Nintendo, nous avons passé des moments agréables non seulement au labo mais aussi à la taverne des serruriers. Grand merci pour ton aide scientifique mais surtout pour les super gâteaux que tu fais.

Ma gratitude va aussi aux membres de la Plateforme de Biologie et Génomique Structurale : Didier Busso, Edo (je te mets à contre cœur ☺), Loubna, Matthieu et Pierre. Je vous remercie pour votre disponibilité et votre aide durant les clonages, l'expression et la cristallisation.

Je remercie toutes les personnes qui m'ont aidée à réaliser ce travail en me donnant la formation nécessaire pour réaliser certaines expériences ou en réalisant d'autres: Pascal EBERLING (Synthèse peptidique), Pierre LEGRAND (Diffraction des rayons X), Eric ENNIFAR et Clarisse MAEHLING (ITC), Yves NOMINE (DC), Dimitri SVERGUN et Javier PEREZ (SAXS), Adeline PAGE, Frank RUFFENACH, Sarah SANGLIER et Cédric ATMANENE (SM), Mark RUFF et Sylvia EILER (thermofluor).

Un grand merci à Jean-Claude THIERRY pour les précieux conseils durant nos trajets de tram.

Je remercie également tous les membres du département de Biologie et Génomique Structurale et les services communs de l'IGBMC.

Virginie et Pierre, merci beaucoup pour la TEV, là vous serez tranquilles ;)

Mes supers chers amis, tous les habouls et toutes les haboulas, j'ai passé avec vous des moments magnifiques que je n'oublierai jamais. : Les repas de midi tous les jours, les pauses thés, la randonnée (oui ca m'est arrivée une fois), les pique-niques, les soirées thématiques (couscous, pizza, italiennes, etc...), les discussions dans les couloirs et dans les chambres froides, etc. Vous allez me manquer beaucoup : Alex (le fameux Zoro, oups c'est un secret), Marie-Laure, Martin (arrête de lire les remerciements et mange), Pierrot, Justine, Emeline, Dr. W., Massimo, Cori, Sere (quand tu voudras nettoyer la salle de bain à 3h du matin, je serais partante ☺), Chris, Robin, Laura, Matthieu, Sandrine, Edo, Loubna, Marie, Denise (ma super prof d'espagnol), Luc, Emiko, Pierre, Lucille, Isabelle, Ally, Tiphaine, Benoit, Julie, Guillaume, Angelita, Zhana, Morgan, Sankar,...

Je remercie de tout mon cœur mes amis hors-labo avec qui j'ai passé pas mal de temps aussi. J'en cite quelques uns sinon il me faut des pages : Cynthiou, Oli, Gigi et Mimo, Allouni, Bacho, Richy, Daliou, Mahou, Rabih, Manmoun, Dido, Maya, Dandoun, Juliette, Ayman.....

Mes amis du Liban qui ont toujours été à côté par leurs coups de fil, leurs e-mails ou bien leurs messages sur facebook, je vous aime beaucoup.

Un simple merci est insuffisant pour la famille Awad qui m'a accueillie sans même me connaître. Vous êtes ma famille adoptive Strasbourgeoise Bob, Elisabeth, Marie, Mimi, Gab, Albert et Gaël. Je remercie également Eid, Yumiko, Jonathan et Julien pour leur accueil et leur présence à coté de moi.

Guy, ton support m'a permis de venir en France et réaliser ce que j'avais envie de faire. Merci beaucoup.

Mes oncles, le support que vous m'avez accordé est énorme, je vous suis très reconnaissante et je n'oublierai jamais tout ce que vous avez fait pour moi. Merci également au reste de ma famille, grands-parents, cousins, tantes, pour votre encouragement et votre amour.

Maman, Papa, Fedo et Solium, sans vous il n'y a rien de tout ça ... Je vous adore.

Magdo, ta présence et ton encouragement continuel durant cette période m'a permis de surmonter les moments les plus sombres. Je n'ai trouvé aucun mot qui soit capable d'exprimer mes sentiments envers toi, je ne sais pas si « je t'aime » pourrait suffire.

Grand merci aussi à toutes les personnes que j'ai oubliées de citer et vous êtes tous les bienvenus chez moi à Madrid ☺.



# Table de Matières

<b>Chapitre I: Introduction .....</b>	<b>1</b>
1. La régulation épigénétique .....	3
1.1. Composition et structure de la chromatine.....	3
1.1.1. Composition de la chromatine .....	3
1.1.2. Structure de la chromatine .....	4
1.1.3. Les différents types de chromatine .....	7
1.2. Les modifications de la chromatine .....	7
1.2.1. La méthylation de l'ADN .....	8
1.2.1.1. Les ADN méthyltransférases.....	8
1.2.1.2. Les îlots CpG.....	9
1.2.1.3. Les modules de reconnaissance des CpG méthylés .....	9
1.2.1.4. La dérégulation de la méthylation de l'ADN .....	10
1.2.2. Les modifications des histones .....	11
1.2.2.1. Les enzymes modifiant les histones .....	12
1.2.2.2. Effets des modifications sur la chromatine .....	12
1.2.2.3. Le code histone.....	13
1.2.2.4. Les lysines méthylées et leur mode de reconnaissance.....	13
1.2.2.5. La dérégulation des modifications d'histones .....	14
1.3. Thérapie épigénétique .....	15
2. La protéine humaine hUHRF1 et la régulation épigénétique.....	16
2.1. La famille UHRF.....	16
2.2. Expression cellulaire de hUHRF1 .....	16
2.3. Les différents rôles de hUHRF1 .....	17
2.3.1. Rôle de hUHRF1 dans la régulation du cycle cellulaire.....	17
2.3.2. Rôle de hUHRF1 dans la cancérogenèse.....	19
2.3.3. Rôle de hUHRF1 dans la régulation épigénétique.....	21
2.4. Architecture de la protéine hUHRF1 .....	22
2.4.1. Les domaines NIRF-N, Tandem tudor et RING de hUHRF1 .....	23
2.4.2. Le domaine SRA de hUHRF1 .....	24
2.4.3. Le domaine PHD de hUHRF1 .....	26
3. Domaines PHD et régulation épigénétique .....	27

3.1. Caractéristiques structurales et fonctionnelles des domaines PHD .....	27
3.1.1. Les caractéristiques structurales .....	27
3.1.2. Les caractéristiques fonctionnelles des PHD.....	29
3.2. Les PHD reconnaissant des marques spécifiques de l'histone H3.....	30
3.2.1. Les PHD reconnaissant la lysine 4 non modifiée .....	30
3.2.2. Les PHD reconnaissant la lysine 4 méthylée.....	32
3.2.3. Les PHD reconnaissant la lysine 9 méthylée.....	35
3.2.4. Comparaison de séquences entre les différentes classes de PHD reconnaissant l'histone H3.....	36
<b>Chapitre II: Matériels et Méthodes.....</b>	<b>39</b>
1. Techniques de la biologie moléculaire .....	40
1.1. Le clonage Gateway®.....	40
1.1.1. Principe du clonage Gateway® .....	40
1.2. Protocole de clonage .....	42
1.2.1. Dessin des oligonucléotides.....	42
1.2.2. Amplification des domaines par la technique de réaction en chaîne par polymérase:.....	42
1.2.3. Purification des produits PCR .....	43
1.2.4. La réaction BP .....	44
1.2.5. Transformation.....	44
1.2.6. Vérification de la recombinaison BP par PCR .....	44
1.2.7. Minipréparation d'ADN .....	45
1.2.8. Séquençage des pENTR.....	46
1.2.9. La réaction LR et la transformation.....	46
1.2.10. Vérification de la recombinaison LR et minipréparation de l'ADN .....	47
2. Production de la protéine dans <i>E.coli</i> .....	48
2.1. Préparation des bactéries E. coli BL21 DE3 Rosetta 2 électro-compétentes.....	49
2.2. Transformation par électroporation.....	50
2.3. Préculture et culture .....	50
3. Lyse cellulaire et purification .....	51
3.1. La lyse cellulaire .....	51
3.2. La purification.....	51
3.2.1. La chromatographie d'affinité .....	52
3.2.2. La chromatographie d'exclusion .....	52

3.2.3. La chromatographie d'échange d'ions.....	53
3.3. Les protocoles utilisés .....	54
3.3.1. Protocole de purification de la protéine PHD-SRA <sub>296-643</sub> .....	54
3.3.2. Protocole de purification de la protéine PHD <sub>296-367</sub> .....	55
3.3.3. Protocole de purification de la protéine PHD <sub>314-367</sub> .....	57
3.4. Contrôle quantitatif et qualitatif des protéines .....	57
3.4.1. Electrophorèse monodimensionnelle en conditions dénaturantes .....	58
3.4.1.1. Composition des gels dénaturants et des solutions utilisés .....	58
3.4.2. Electrophorèse monodimensionnelle en conditions non dénaturantes .....	59
3.4.2.1. Composition du gel non dénaturant.....	60
3.4.2.2. Préparation des échantillons et migration.....	60
4. Caractérisation des protéines et des interactions en solution .....	60
4.1. Thermofluor .....	61
4.2. Diffusion quasi-élastique de la lumière (DLS) .....	63
4.3. La spectrométrie de masse .....	63
4.3.1. La théorie .....	63
4.3.2. Le protocole expérimental .....	64
4.4. Ultracentrifugation analytique .....	65
4.4.1. La théorie .....	65
4.4.2. Le protocole expérimental .....	67
4.5. La diffusion des rayons X aux petits angles.....	67
4.5.1. La théorie .....	68
4.5.2. Le protocole expérimental .....	70
4.6. Le dichroïsme circulaire.....	72
4.6.1. La théorie .....	72
4.6.2. Le protocole expérimental .....	74
4.7. Calorimétrie à titration isotherme .....	75
4.7.1. La théorie .....	77
4.7.2. Protocole expérimentale .....	79
4.7.2.1. VP-ITC .....	79
4.7.2.2. ITC200.....	81
5. Cristallogenèse et cristallographie.....	82
5.1. Principe de la cristallisation .....	82
5.2. Les essais de cristallisation .....	84

5.3. Enregistrement des données de diffraction .....	85
5.4. Détermination des phases expérimentales par la méthode SAD.....	86
<b>Chapitre III: Caractérisation de la protéine PHD-SRA<sub>296-643</sub> .....</b>	<b>87</b>
1. Clonage, production et purification de PHD-SRA .....	88
1.1. Les différentes constructions de PHD-SRA.....	88
1.2. Expression et purification de PHD-SRA <sub>296-643</sub> .....	89
2. Caractérisation en solution de PHD-SRA <sub>296-643</sub> .....	92
2.1. Caractérisation de PHD-SRA <sub>296-643</sub> par spectrométrie de masse.....	92
2.2. Caractérisation de PHD-SRA <sub>296-643</sub> par SAXS.....	92
2.3. Caractérisation de PHD-SRA <sub>296-643</sub> par dichroïsme circulaire .....	96
2.4. Optimisation du conditionnement de PHD-SRA <sub>296-643</sub> par Thermofluor .....	97
3. Caractérisation des interactions de PHD-SRA <sub>296-643</sub> avec ses partenaires .....	98
3.1. Etude de l'interaction de PHD-SRA avec l'ADN.....	98
3.1.1. Caractérisation de l'interaction par spectrométrie de masse en conditions non dénaturantes .....	99
3.1.2. Caractérisation de l'interaction par chromatographie d'exclusion.....	100
3.1.3. Caractérisation de l'interaction par ultracentrifugation analytique .....	101
3.2. Etude de l'interaction de PHD-SRA avec l'histone H3 par calorimétrie à titration isotherme (VP-ITC).....	104
4. Cristallisation.....	106
5. Conclusion.....	106
<b>Chapitre IV: Analyse bioinformatique et clonage du domaine PHD de hUHRF1 .....</b>	<b>108</b>
1. Comparaison de séquences entre le domaine PHD de hUHRF1 et d'autres domaines PHD fixant l'histone H3.....	109
2. Comparaison structurale entre le PHD de hUHRF1 et des PHD fixant l'histone H3 .....	109
2.1. Comparaison structurale avec les PHD fixant l'histone H3K4me0.....	110
2.2. Comparaison avec les PHD fixant H3K4me3.....	112
3. Clonage du domaine PHD .....	113
<b>Chapitre V: Caractérisation de la protéine PHD<sub>314-367</sub>.....</b>	<b>115</b>
1. Expression, purification et caractérisation de PHD <sub>314-367</sub> .....	116
2. Caractérisation de l'interaction de PHD <sub>314-367</sub> avec l'histone H3.....	119
2.1. Caractérisation des complexes PHD <sub>314-367</sub> -histone H3 sur gels natifs .....	119
2.2. Caractérisation de l'interaction par ITC.....	120

3. Cristallisation de la protéine PHD <sub>314-367</sub> .....	120
4. Etude cristallographique et analyse de structures.....	121
5. Conclusion.....	125
<b>Chapitre VI: Caractérisation de la protéine PHD<sub>296-367</sub> .....</b>	<b>127</b>
1. Expression et purification de PHD <sub>296-367</sub> .....	128
2. Caractérisation de l'interaction de PHD <sub>296-367</sub> avec l'histone H3.....	131
2.1. Caractérisation de l'interaction sur gels natifs .....	131
2.2. Caractérisation de l'interaction par ITC.....	133
2.2.1. Caractérisation de la protéine PHD <sub>296-367</sub> issue du protocole initial .....	133
2.2.2. Caractérisation de la protéine PHD <sub>296-367</sub> issue du protocole optimisé.....	136
3. Cristallisation de la protéine PHD <sub>296-367</sub> .....	137
4. Etudes cristallographiques et analyse de la structure .....	138
5. Conclusion.....	140
<b>Chapitre VII: Conclusions et perspectives .....</b>	<b>142</b>
<b>Chapitre VIII: Références bibliographiques.....</b>	<b>145</b>
<b>Chapitre IX: Liste des Figures et des Tableaux.....</b>	<b>155</b>



# Chapitre I

## Introduction

---

- 1. La régulation épigénétique**
- 2. La protéine humaine hUHRF1 et la régulation épigénétique**
- 3. Domaines PHD et régulation épigénétique**

Si toutes les cellules de notre organisme ont le même nombre de gènes, seul un petit nombre d'entre eux est exprimé dans une cellule donnée, selon le stade de développement de l'organisme ou le type cellulaire considéré. C'est la régulation épigénétique qui va permettre de moduler l'activité transcriptionnelle des gènes en modifiant l'accessibilité de la chromatine à la machinerie transcriptionnelle. La régulation épigénétique comprend « l'écriture » des marques épigénétiques par différentes enzymes, « la lecture » de ces marques par différents domaines protéiques et « l'interprétation » de ces marques par le recrutement de complexes activateurs ou répresseurs de la transcription. Toute perturbation dans la régulation épigénétique va conduire à une expression inappropriée de certains gènes et pourra contribuer à l'apparition de maladies. Ainsi, les acteurs de la régulation épigénétique sont devenus des éléments essentiels de la compréhension des mécanismes de régulation transcriptionnelle et des cibles thérapeutiques potentielles d'intérêt croissant.

La protéine humaine UHRF1 est une protéine nucléaire constituée de plusieurs domaines de reconnaissance de marques épigénétiques et également capable d'interagir avec différents acteurs de la régulation épigénétique, ce qui en fait une protéine d'intérêt croissant pour la compréhension des mécanismes épigénétiques. Les deux domaines centraux de la protéine hUHRF1 sont le domaine SRA (SET and RING associated domain) qui est capable d'interagir avec de l'ADN hémiméthylé et le domaine à doigts de zinc de type PHD (Plant HomeoDomain) potentiellement impliqué dans la reconnaissance de l'histone H3.

En effet, différents domaines PHD ont récemment été identifiés comme des modules de reconnaissance de modifications spécifiques sur l'histone H3. Ces domaines PHD reconnaissant l'histone H3 sont classés en fonction de la modification reconnue: la lysine 4 non modifiée, la lysine 4 méthylée ou la lysine 9 méthylée.

Dans ce travail de thèse, nous avons tout d'abord étudié les deux domaines associés PHD et SRA pour aider à la compréhension de leur rôle dans la régulation épigénétique. Ensuite une caractérisation fonctionnelle et structurale du domaine PHD isolé a été initiée, ce qui a conduit à la résolution de la structure tridimensionnelle du domaine PHD de hUHRF1 en complexe avec un peptide de l'histone H3.

# **1. La régulation épigénétique**

L'épigénétique est l'étude des modifications de l'expression des gènes qui sont transmissibles lors de la mitose et/ou la méiose, mais qui ne découlent pas de modifications dans la séquence de l'ADN (Wolffe and Matzke, 1999). Toutes les cellules d'un organisme ont les mêmes informations génétiques, mais la façon dont elles sont lues diffère selon le type de cellules. L'expression génétique est ainsi modulée par des changements d'ordre épigénétique permettant le verrouillage ou le déverrouillage du génome. La cellule transmet à sa descendance non seulement la séquence de ses gènes mais aussi le niveau d'activité génique. Contrairement aux modifications génétiques, les modifications épigénétiques sont réversibles et dynamiques, c'est-à-dire qu'elles peuvent être écrites et effacées. Ces modifications épigénétiques ciblent la chromatine, le support du matériel génétique, et influencent différents processus cellulaires, de la régulation de la transcription à la ségrégation des chromosomes (Deltour et al., 2005; Fuks, 2003; Kepes, 2005).

## **1.1. Composition et structure de la chromatine**

Dans les cellules eucaryotes, l'ADN est présent dans le noyau sous forme d'une structure nucléoprotéique appelée la chromatine. Pendant longtemps, la chromatine n'a été considérée que comme un simple système d'empaquetage de l'ADN génomique qui mesure environ 2 m, dans le noyau cellulaire d'environ 10 µm de diamètre en moyenne. Or les nombreuses études menées ces dernières années démontrent que la chromatine n'est pas inerte mais dynamique et joue un rôle important dans la régulation de nombreux processus cellulaires tels la réplication, la réparation et l'expression de gènes.

### **1.1.1. Composition de la chromatine**

La chromatine est un assemblage composé d'ADN et de protéines. Deux classes de protéines existent dans la chromatine: les histones qui sont très abondantes et jouent un rôle dans la structure de la chromatine et les protéines non histones comprenant une grande variété de protéines, peu abondantes dans la cellule mais jouant un rôle dans la régulation génétique tels les facteurs de transcription. Il y a cinq classes majeures de protéines histones: H1, H2A, H2B, H3 et H4. Ces histones sont de petites protéines de 11 à 24 kDa. Elles sont chargées positivement car riches en arginines et en lysines et elles vont interagir électrostatiquement avec les groupes phosphates chargés négativement de l'ADN.

### 1.1.2. Structure de la chromatine

Deux paires d'histones H2A-H2B et H3-H4 s'associent pour former un octamère d'histones, autour duquel 147 paires de bases d'ADN s'enroulent en 1.7 tour pour former le nucléosome, unité fondamentale de la chromatine. Chacune de ces quatre histones est constituée d'un domaine C-terminal de structure globulaire, formé de trois hélices  $\alpha$ , « histone fold », qui sont impliquées dans la dimérisation des histones, et des queues aminoterminales et carboxy-terminales flexibles qui émergent à la surface du nucléosome. La structure cristallographique d'un nucléosome a été résolue en 1997 à 2.8 Å de résolution (Luger et al., 1997) (Figure 1).



*Figure 1: Structure cristallographique d'un mononucléosome (PDB: 1aoi). La double hélice d'ADN est enroulée autour d'un octamère d'histones formé de deux copies de chacune des histones H2A (en jaune), H2B (en rouge), H3 (en cyan) et H4 (en vert). La queue aminotermine (en cyan) d'une des histones H3 émerge de la surface du nucléosome.*

Dans le noyau, les nucléosomes s'enchainent comme les perles d'un collier, séparés entre eux par des segments d'ADN, dit internucléosomique. Une molécule de l'histone H1 s'associe à l'ADN internucléosomique et permet d'attacher l'ADN nucléosomal sur la surface de l'octamère d'histones, cet ensemble constitue le chromatosome. L'histone H1 permet aussi le pontage internucléosomal et joue un rôle dans la condensation de degré supérieur de l'ADN (Happel and Doenecke, 2009). Le collier de perles se replie ensuite sur lui-même pour former une fibre de 30 nm, la chromatine. D'après des études faites en microscopie électronique et en diffusion des rayons X aux petits angles (SAXS), deux

structures de fibres, solénoïdale ou en zigzag, ont été proposés. Dans le modèle en solénoïde, les nucléosomes forment une hélice autour d'une cavité centrale avec 6 à 8 nucléosomes par tour d'hélice. Les nucléosomes consécutifs sont l'un à côté de l'autre dans cette fibre. Dans le modèle en zigzag, les nucléosomes forment deux rangées et l'ADN internucléosomal zigzague entre eux. Les deux rangées de nucléosomes peuvent ensuite s'enrouler pour former la fibre en double hélice (Figure 2) (Dorigo et al., 2004; Tremethick, 2007; Woodcock, 2005).

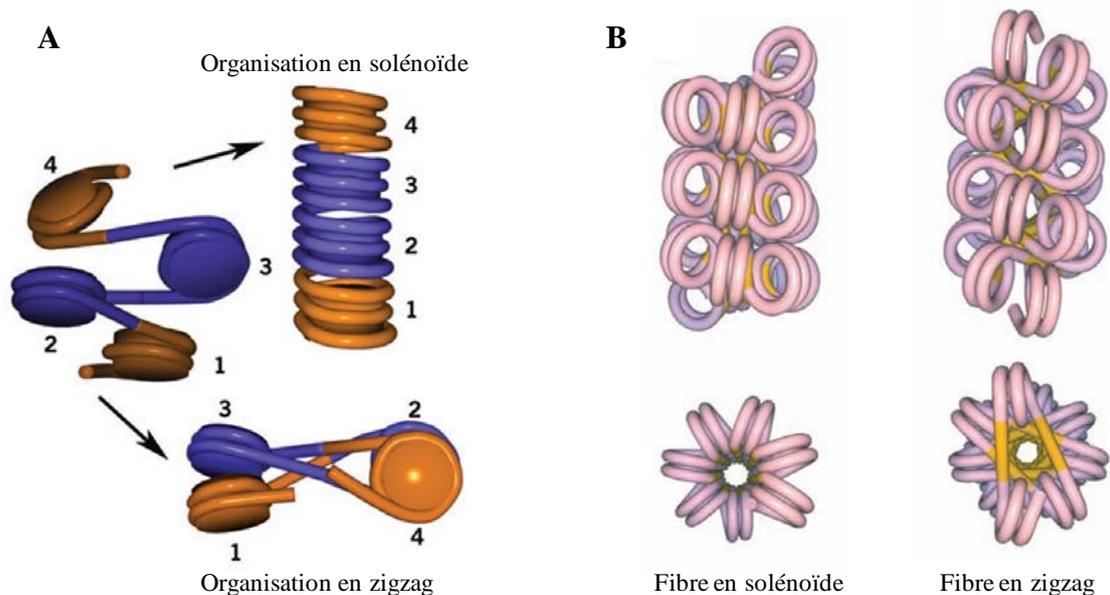


Figure 2: Les deux modèles proposés pour la structure de la fibre chromatinienne de 30 nm. A- Présentation schématique d'une organisation en solénoïde ou en zigzag d'un tétranucléosome (Woodcock, 2005). B- Modèles de l'organisation de l'ADN dans une fibre chromatinienne en solénoïde et en zigzag. L'ADN nucléosomal est présenté en rose et l'ADN internucléosomal en jaune (Dorigo et al., 2004).

La détermination de la structure cristallographique d'un tétranucléosome à 9 Å de résolution a permis de positionner les nucléosomes et l'ADN internucléosomique. Cette structure montre une organisation en zigzag de la chromatine. Les axes traversant les deux rangées de nucléosomes ne sont pas parallèles mais présentent une rotation d'environ 71° (Schalch et al., 2005) (Figure 3). La fibre de 30 nm permet de compacter l'ADN d'environ 50 fois. Des niveaux d'organisation plus compacts sont adoptés pour atteindre le niveau de condensation le plus élevé au sein du chromosome métaphasique. Jusqu'à présent, les niveaux de compaction au delà de la fibre de 30 nm sont très peu caractérisés. Le degré de compaction de l'ADN module son accessibilité et en conséquence l'expression des gènes qu'il porte (Figure 4).



Figure 3: Structure cristallographique en zigzag d'un tétranucléosome (PDB: 1zbb). Pour simplifier, seul l'ADN est présenté dans cette image. Les nucléosomes (N1 à N4) sont positionnés en deux rangées. L'ADN internucléosomal zigzague entre les nucléosomes.

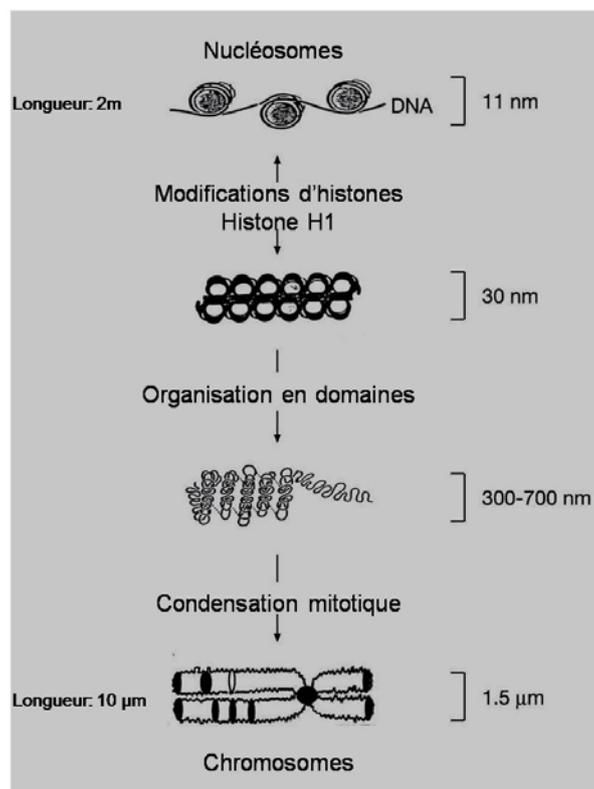


Figure 4: La condensation de la chromatine en chromosomes. La fibre de 11 nm représente l'ADN enroulé autour des nucléosomes, le collier de perles. En impliquant l'histone H1, le collier de perles est compacté en une fibre chromatinienne de 30 nm. La fibre de 300-700 nm présente un niveau élevé de compaction qui correspond à la chromatine de l'interphase et de la métaphase. Le chromosome condensé de 1.5 μm est présent seulement durant les divisions cellulaires (mitose ou méiose) (Allis et al., 2009).

Les modèles et les structures cristallographiques tendent à présenter le nucléosome comme une unité rigide et stable. Mais ce nucléosome peut subir des modifications statiques (remplacement des histones par les variants, variation du nombre de paires de base de l'ADN nucléosomal et internucléosomal) et dynamiques (oligomérisation des histones et changement dans les contacts ADN-histones). Le nucléosome doit être alors considéré comme une particule pouvant différer dans ses propriétés structurales et dynamiques afin d'assurer les différentes fonctions cellulaires (Zlatanova et al., 2009).

### **1.1.3. Les différents types de chromatine**

Historiquement la chromatine a été divisée en euchromatine et hétérochromatine. L'euchromatine est la partie du génome qui est décondensée durant l'interphase et qui peut être transcriptionnellement active ou inactive. Par contre, l'hétérochromatine correspond à la chromatine qui ne change pas d'état de condensation au cours du cycle cellulaire. L'hétérochromatine peut être réprimée de façon permanente, il s'agit de l'hétérochromatine constitutive. Cette dernière comprend peu de gènes et est formée principalement de séquences répétées. Dans certaines cellules, l'hétérochromatine peut être réprimée durant un cycle cellulaire spécifique ou à un certain stade de développement, il s'agit alors de l'hétérochromatine facultative (Allis et al., 2009).

## **1.2. Les modifications de la chromatine**

La chromatine, et non l'ADN seul, est impliquée dans les grandes fonctions nucléaires qui sont la transcription, la réplication, la réparation et la recombinaison de l'ADN. Pour moduler son activité, la cellule a développé différents mécanismes permettant de modifier de façon spatiotemporelle l'organisation de la chromatine et de maintenir une telle organisation au cours de la division cellulaire mitotique ou méiotique:

- Le recrutement des complexes de remodelage de la chromatine: ces complexes requièrent l'hydrolyse de l'ATP pour modifier la structure chromatinienne, exposant ou cachant des régions de l'ADN afin de faciliter ou d'empêcher leur interaction avec des facteurs de réplication, de réparation et de transcription. Cinq familles de complexes existent classés selon la nature de leur ATPase: SWI/SNF, ISWI, Mi-2/NuRD, INO80 et SWR1.
- Les modifications post-traductionnelles des histones qui facilitent ou empêchent l'association des facteurs de transcription et des protéines de réparation de l'ADN à la chromatine.

- Le remplacement des histones canoniques du nucléosome par des variants d'histones: la substitution d'une histone conventionnelle par un variant peut créer un nouveau complexe nucléoprotéique de structure et de stabilité modifiées. Ce remplacement contribue à la spécialisation fonctionnelle de la chromatine et permet la génération de domaines fonctionnels indépendants au sein de la chromatine (euchromatine et hétérochromatine facultative ou constitutive).
- La méthylation des cytosines présentes dans des dinucléotides CpG par des ADN méthyltransférases (DNMTs) facilitant la répression à long terme de certains gènes et assurant la stabilité du génome via la répression des transposons et des éléments répétitifs.
- La présence d'ARN non codants qui sont transcrits à partir de l'ADN mais ne sont pas traduits en protéines. Ces ARN contrôlent différents aspects de l'expression des gènes tels la mémoire épigénétique, la transcription, l'épissage des ARN et la traduction.

Tous ces mécanismes fonctionnent de façon coopérative et forment les bases de la régulation épigénétique qui contrôle l'expression des gènes chez les eucaryotes supérieurs (Mattick and Makunin, 2006; Perche et al., 2003; Reid et al., 2009; Santos-Rosa and Caldas, 2005). Dans ce manuscrit, seuls les mécanismes de méthylation de l'ADN et de modifications des histones seront détaillés.

### **1.2.1. La méthylation de l'ADN**

La méthylation de l'ADN est un élément clé dans la formation et la maintenance de l'hétérochromatine.

#### **1.2.1.1. Les ADN méthyltransférases**

La méthylation de l'ADN est assurée par les DNMTs qui ajoutent un groupe méthyle sur le carbone 5 d'une cytosine d'un dinucléotide palindromique CpG (Figure 5). Les profils de méthylation sont mis en place par l'action d'au moins 3 enzymes ADN méthyltransférases: DNMT1, DNMT3a et DNMT3b qui utilisent la S-adénosyl-méthionine comme donneur du groupe méthyle. DNMT1 est une méthylase de maintenance capable de maintenir le profil de méthylation au cours des divisions cellulaires. DNMT1 agit préférentiellement sur l'ADN hémiméthylé et est localisée au niveau de la fourche de réplication. Par contre DNMT3a et 3b sont des de novo méthylases impliquées respectivement dans la méthylation des séquences

régulatrices de l'expression des gènes et des séquences satellites<sup>1</sup> des centromères (Reid et al., 2009).

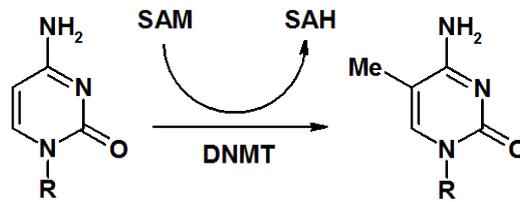


Figure 5: La méthylation d'une cytosine par une DNMT.

#### 1.2.1.2. Les îlots CpG

Les dinucléotides CpG ne sont pas très abondants dans les génomes de mammifères du fait de la désamination spontanée de la 5-méthyle cytosine en thymidine. Les dinucléotides CpG sont répartis de façon non uniforme dans le génome, avec une sous-représentation générale (< 1/100 pb), mais un enrichissement (~ 1/100 pb) dans de courtes régions de 1000 pb, appelées îlots CpG, positionnées au niveau du promoteur et/ou du premier exon d'environ 40% des gènes humains.

#### 1.2.1.3. Les modules de reconnaissance des CpG méthylés

Trois domaines protéiques ont été identifiés, jusqu'à présent, comme des modules de reconnaissance des CpG méthylés: le domaine MBD (methyl-CpG binding domain), les doigts de zinc de type C2H2 et le domaine SRA (SET and RING associated domain) de la famille UHRF. Le domaine MBD, d'environ 70 acides aminés, est présent dans une famille protéique dont les principaux membres sont: MeCP2, MBD1, MBD2, MBD3 et MBD4. Toutes ces protéines à l'exception de MBD3 sont capables de fixer le 5meCpG. Trois protéines humaines contenant des doigts de zinc de type C2H2 sont capables de reconnaître les 5meCpG. La première est un facteur de transcription appelé Kaiso qui reconnaît une séquence consensus composée d'au moins deux dinucléotides CpG méthylés consécutifs. Les deux autres protéines sont ZBTB4 et ZBTB38 qui peuvent interagir avec un ADN contenant un seul dinucléotide méthylé (Lopez-Serra and Esteller, 2008; Reid et al., 2009). Le domaine SRA de la famille UHRF reconnaît spécifiquement les dinucléotides CpG hémiméthylés (Bostick et al., 2007; Unoki et al., 2004).

Les protéines reconnaissant les CpG méthylées sont capables de recruter des complexes de remodelage de la chromatine vers le site de méthylation. Ces complexes contiennent des

<sup>1</sup> Les ADN satellites consistent en de courtes séquences d'ADN répétées un grand nombre de fois.

enzymes de modifications post-traductionnelles des histones, des enzymes responsables de la méthylation de l'ADN et des corépresseurs transcriptionnels. Cette zone devient alors inaccessible aux complexes de transcription, empêchant ainsi l'expression du gène (Deltour et al., 2005).

#### 1.2.1.4. La dérégulation de la méthylation de l'ADN

Des profils aberrants de méthylation de l'ADN ainsi que des mutations touchant les protéines reconnaissant cette marque épigénétique sont impliqués dans différentes maladies humaines.

L'hypométhylation des séquences d'ADN répétitives peut être impliquée dans des instabilités chromosomiques, des translocations et d'autres anomalies nucléaires. Le syndrome ICF (Immunodéficience-instabilité Centromérique-dysmorphie Faciale) par exemple, est directement lié à par des mutations touchant le gène codant pour l'ADN méthyltransférase DNMT3b, conduisant à l'hypométhylation des séquences répétitives péri-centromériques, observée au niveau des chromosomes des malades. Dans certaines cellules cancéreuses, l'hypométhylation de certains promoteurs permet d'exprimer des gènes normalement réprimés, par exemple certains gènes codant pour des facteurs de transcription impliqués dans la prolifération comme *PAX2*.

Le mécanisme le mieux étudié par lequel la méthylation de l'ADN contribue aux cancers est l'hyperméthylation des promoteurs des gènes suppresseurs de tumeur. Dans le cas d'un cancer, les profils de méthylation sont altérés. Bien qu'il y ait une hypométhylation globale des dinucléotides CpG, les promoteurs des gènes suppresseurs de tumeur présentent une hyperméthylation de leurs îlots CpG entraînant leur extinction transcriptionnelle. Ces gènes sont impliqués dans certaines fonctions essentielles pour la cellule, comme le contrôle du cycle cellulaire, la réparation des dommages d'ADN et l'apoptose. Parmi ces gènes réprimés dans différents types de cancer par hyperméthylation des îlots CpG, il y a p16<sup>INK4a</sup>, RB, BRCA1 et p53. Chaque tumeur possède son profil spécifique d'hyperméthylation des îlots CpG dans les gènes suppresseurs de tumeur, donc un hyperméthylome spécifique. Par exemple, l'hyperméthylation des îlots CpG de BRCA1 a lieu essentiellement dans le cancer du sein et le cancer ovarien. L'hyperméthylation aberrante de certains promoteurs est aussi impliquée dans la répression des microARN (miARN), des petits ARN non codants qui inhibent l'expression de certains gènes cibles (Allis et al., 2009; Deltour et al., 2005; Esteller, 2008; Lopez-Serra and Esteller, 2008).

Les mutations qui affectent les protéines reconnaissant l'ADN méthylé sont aussi associées à certaines maladies humaines. Des mutations dans le gène codant pour MeCP2 sont responsables d'un désordre postnatal neurodéveloppemental lié au chromosome X, le syndrome de Rett. Cette maladie touche essentiellement les filles et est caractérisée, entre autre, par des anomalies motrices, une ataxie et une régression du langage (Allis et al., 2009; Deltour et al., 2005).

### 1.2.2. Les modifications des histones

Les modifications d'histones les plus étudiées sont l'acétylation des lysines, la méthylation des lysines (mono-, di- ou tri-) et des arginines (mono-, di- symétrique ou asymétrique), la phosphorylation des sérines et des thréonines et l'ubiquitinylation des lysines. D'autres modifications existent aussi telles que l'ADP-ribosylation, la sumoylation, la glycosylation, la désimination des arginines (ou citrullination), l'isomérisation des prolines et la biotinylation. Les extrémités amino-terminales des histones, qui émergent à la surface du nucléosome sont les cibles privilégiées de ces modifications post-traductionnelles avec quelques exceptions telle l'ubiquitination des queues carboxy-terminales des histones H2A et H2B et l'acétylation et la méthylation du domaine globulaire de H3 au niveau des lysines K56 et K79 respectivement (Figure 6).

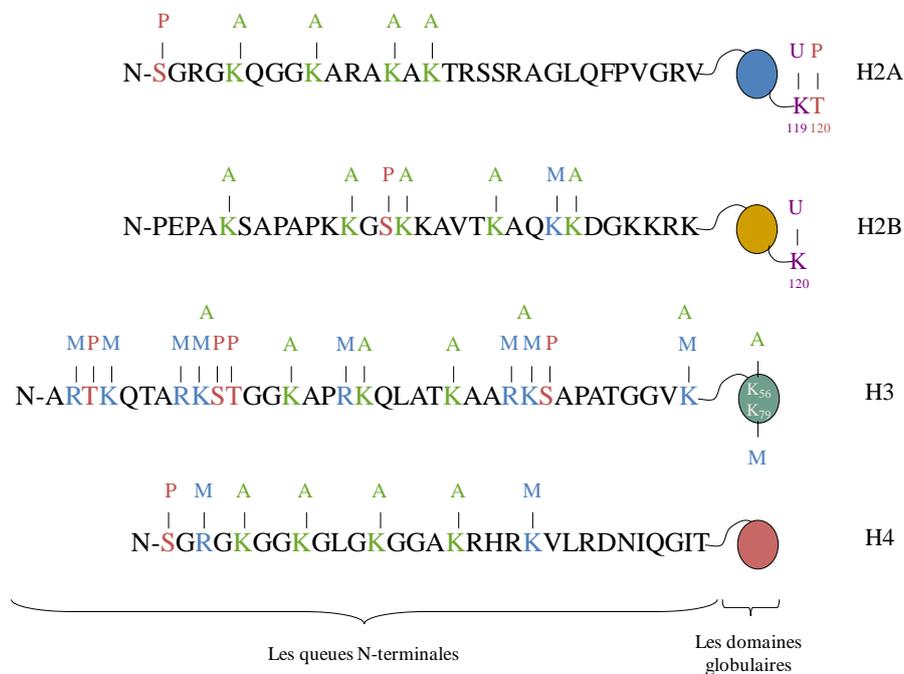


Figure 6: Les modifications post-traductionnelles des histones nucléosomales humaines. Les modifications présentées sont l'acétylation A, la méthylation M, la phosphorylation P et l'ubiquitinylation U (Bhaumik et al., 2007).

### 1.2.2.1. Les enzymes modifiant les histones

Les modifications des histones sont dynamiques, réversibles, ciblées sur des résidus spécifiques pour chaque histone et sont catalysées par deux types d'enzymes antagonistes et spécifiques. Le premier type d'enzymes est capable d'ajouter des modifications sur les histones et le deuxième enlève des modifications déjà existantes. Par exemple, les histones acétyltransférases HATs (Gcn5, HBO1, PCAF, etc.) acétylent des lysines, tandis que les histones déacetylases HDACs (HDAC8, Rpd3, SIRT1, etc.) enlèvent les groupements acétyles des lysines. De même, les histones kinases phosphorylent des sérines ou des thréonines spécifiques tandis que les phosphatases enlèvent les groupes phosphates. Au sein de ce dernier couple, les enzymes les mieux connues sont des kinases mitotiques telle la kinase aurora B qui catalyse la phosphorylation de la sérine S10 de l'histone H3 et la phosphatase de type 1 (PP1) qui effectue la réaction de déphosphorylation quand la cellule sort de la mitose. Deux classes d'histones méthyltransférases ont été décrites: les PRMTs qui sont des arginines méthyltransférases (PRMT1, CARM1, etc.) et les HKMTs qui sont des lysines méthyltransférases (G9a, MLL1, etc.). La réaction de déméthylation des arginines est catalysée par les déiminases qui transforment les méthyle-arginines en citrullines. Plusieurs histone déméthylases (LSD1, JMJD2D, JARID1A, JMJD6, etc.) ont été récemment identifiées, la première déméthylase, LSD1, ayant été décrite en 2004. Cette déméthylase est une amine oxydase capable de déméthyliser H3K4 mono ou di-méthylée et non pas triméthylée (Allis et al., 2009; Kouzarides, 2007; Lall, 2007; Ray-Gallet et al., 2005; Santos-Rosa and Caldas, 2005).

### 1.2.2.2. Effets des modifications sur la chromatine

L'effet des modifications d'histones sur la chromatine peut être en « cis » ou en « trans ». L'effet en « cis » est dû à des changements dans les propriétés physico-chimiques des histones telles les modifications affectant la charge électrostatique ou l'arrangement des queues. Ces modifications affectent directement les contacts inter-nucléosomiaux et donc la compaction de la chromatine. Par exemple, l'acétylation des histones neutralise leurs charges positives, générant une expansion locale de la chromatine et rendant l'ADN plus accessible à la machinerie transcriptionnelle. La phosphorylation permet d'ajouter des charges négatives capables d'affecter l'assemblage du nucléosome. De même l'ajout de gros groupements comme l'ubiquitine ou l'ADP-ribose peut induire différents arrangements des queues d'histones et affecter la compaction de la chromatine. Par contre, l'effet en « trans » consiste à reconnaître les modifications d'histones par des domaines protéiques particuliers présents

dans plusieurs co-régulateurs transcriptionnels et qui à leur tour facilitent le recrutement et/ou la stabilisation au niveau de la chromatine, des complexes de remodelage et/ou de modification de la chromatine. Par exemple, une lysine méthylée peut être reconnue par un chromodomaine qui va faciliter le remodelage de la chromatine en faisant partie ou en recrutant le complexe de remodelage approprié (Allis et al., 2009).

### 1.2.2.3. Le code histone

Les modifications post-traductionnelles des histones constituent un «code histone» permettant d'associer à chaque combinaison de modifications un état particulier de la chromatine. Le «code histone» est très complexe car il implique une conjonction de modifications dont chaque combinaison définit un état transcriptionnel donné. En effet, une modification sur un site donné peut influencer la capacité d'un autre site à être modifié. Cette influence peut être en «cis» c'est-à-dire une modification sur une histone donnée entraîne la modification d'un autre résidu sur la même histone, ou bien en «trans» quand elle entraîne la modification d'un résidu sur une autre histone. Un exemple de la modification en cis est la phosphorylation de la sérine 10 de l'histone H3 qui inhibe la méthylation de la lysine 9 mais favorise l'acétylation de la lysine 14. Par contre l'ubiquitinylation de la lysine 123 de l'histone H2B qui favorise la méthylation de la lysine 4 de l'histone H3 est un exemple d'une modification en trans (Figure 7) (Fingerman et al., 2008; Kouzarides, 2007; Taverna et al., 2007).

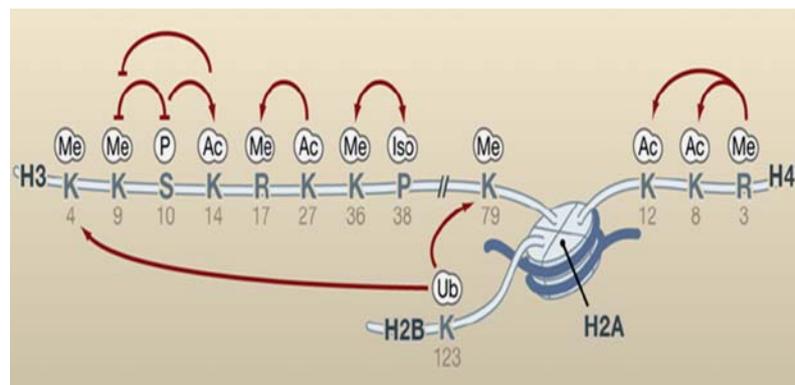


Figure 7: Les modifications des histones en «cis» et en «trans» (Kouzarides, 2007).

### 1.2.2.4. Les lysines méthylées et leur mode de reconnaissance

La méthylation des lysines sur le groupe amine de la chaîne latérale a été mise en évidence en 1964 mais l'identification de la première histone lysine méthyltransférase a eu lieu en 2000 (Murray, 1964; Rea et al., 2000). Depuis, plusieurs lysine méthyltransférases ont été identifiées ainsi que leurs sites de méthylation sur les histones. Toutes ces enzymes, à

l'exception de Dot1, ont un domaine SET qui est le domaine catalytique contenant le site de fixation du cofacteur S-adenosyl-L-méthionine (SAM), donneur du groupe méthyle. Les sites de méthylation les mieux caractérisés sont les cinq lysines K4, K9, K27, K36 et K79 de H3 et la lysine K20 de H4. La méthylation de H3K4, H3K36 et H3K79 est en général liée à une activation de la transcription alors que la méthylation de H3K9, H3K27 et H4K20 est liée à une répression. Deux classes de domaines protéiques sont impliquées dans la reconnaissance des lysines méthylées via des cages aromatiques: les membres de la famille Tudor qui comprend entre autres les domaines Tudor, chromodomaine et MBT (Malignant Brain Tumor), et les doigts de zinc de type PHD (Taverna et al., 2007).

#### 1.2.2.5. La dérégulation des modifications d'histones

Les modifications aberrantes des histones peuvent être impliquées dans l'altération de l'expression des gènes et conduire à l'apparition de différentes maladies humaines.

La perte globale de l'acétylation de H4K16 est associée aux cancers du sein et du foie. Une perte de l'acétylation de l'histone H3 au niveau des gènes suppresseurs des tumeurs ou une altération de certains gènes codant pour des HATs ont été associées à différents types de cancers et d'autres anomalies. Des mutations touchant le gène codant pour HAT p300, par exemple, sont liées à un certain nombre de tumeurs gastro-intestinales et des mutations de HAT CBP (Creb binding protein) sont associées au syndrome Rubinstein-Taybi, caractérisé par un retard mental, des anomalies faciales et une prédisposition au cancer chez l'enfant. Des mutations modifiant le degré de méthylation de l'histone H3 sont aussi impliquées dans le développement d'une variété de cancers et d'autres maladies humaines. La perte de la méthylation de H3K4 et l'augmentation de H3K9 méthylée sont associées à différents cancers. Environ 80% des leucémies infantiles sont causées par des translocations chromosomiques du gène *MLL* (mixed-lineage leukemia) qui est essentiel à la méthylation de H3K4. Certaines maladies humaines sont dues à des mutations touchant les complexes de remodelage de chromatine. BRG1, par exemple, l'ATPase de l'un des complexes SWI/SNF humains est mutée dans de nombreuses tumeurs (Allis et al., 2009; Bhaumik et al., 2007; Deltour et al., 2005).

Les mutations qui affectent les modules de reconnaissance de ces modifications sont aussi associées à différentes maladies humaines. Récemment, différentes maladies comme le syndrome d'immunodéficience, certains types de cancers et certaines maladies d'ordre neurologique ont été associées à des mutations ponctuelles, des délétions ou des translocations chromosomiques touchant des gènes codant pour des domaines PHD (Baker et

al., 2008). Par exemple, des mutations dans le domaine PHD1 du régulateur autoimmun AIRE sont associées à la maladie autoimmune APECED (Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy) (Finnish-German APECED Consortium, 1997).

### 1.3. Thérapie épigénétique

Les acteurs de la régulation épigénétique sont impliqués dans différentes pathologies humaines et peuvent donc être des cibles thérapeutiques potentielles. Des inhibiteurs spécifiques des méthyltransférases (DNMTs) et des histone-déacetylases (HDACs) sont déjà disponibles.

Dans les cellules cancéreuses, les gènes suppresseurs de tumeur sont hyperméthylés et donc réprimés. L'utilisation des médicaments qui bloquent la fonction des DNMTs, permet de ré-exprimer ces gènes suppresseurs de tumeur. Des analogues des bases nucléotidiques tels le 5-azacytidine et le 5-aza-2'-désoxycytidine qui s'incorporent dans l'ADN à la place de résidus cytosines, ont été autorisés comme traitements pour le syndrome myélodysplasique et la leucémie. L'absence de cytosines rend impossible la méthylation de l'ADN par les DNMT permettant ainsi la réexpression des gènes éteints par hyperméthylation. L'inconvénient de ces agents est leur manque de spécificité car ils empêchent la méthylation des îlots CpGs des gènes suppresseurs de tumeur mais provoquent aussi une diminution globale de la méthylation qui contribue à l'instabilité génomique. En plus, ces agents n'ont pas d'effet sur les tumeurs solides.

Les inhibiteurs des HDACs sont capables d'induire la différenciation, l'arrêt du cycle cellulaire et l'apoptose *in vitro*. Parmi les plus connus, il y a les acides gras à courte chaîne, comme le butyrate, les acides hydroxamiques telle la trichostatine A (TSA) et les tétrapeptides cycliques comme la trapoxine. Ces inhibiteurs ont de faibles effets secondaires. L'acide suberoylanilide hydroxamique (SAHA) a été approuvé comme traitement pour le lymphome T cutané. L'efficacité des inhibiteurs des HDACs dans les autres tumeurs reste limitée.

La non spécificité des inhibiteurs des DNMTs et des HDACs peut conduire à des effets indésirables. Mais l'utilisation d'une thérapie épigénétique spécifique et dirigée vers des facteurs de transcription ciblant des promoteurs de gènes particuliers semble prometteuse. Cibler par exemple les protéines à doigts de zinc qui reconnaissent des séquences spécifiques dans certains promoteurs est à l'étude (Allis et al., 2009; Esteller, 2008).

L'importance des mécanismes de régulation épigénétiques, illustrés ci-dessus par la gravité des maladies consécutives à leurs dysfonctionnements, justifie l'intérêt porté à tout

nouveau facteur impliqué dans ces mécanismes. La protéine hUHRF1, dont le rôle dans la régulation du cycle cellulaire a été révélé en 2003, est un de ces nouveaux facteurs, capable via ses domaines fonctionnels, de reconnaître des marques épigénétiques particulières et de recruter des complexes de modifications et de remodelage de la chromatine. Elle pourrait être une cible thérapeutique potentielle.

## **2. La protéine humaine hUHRF1 et la régulation épigénétique**

La protéine humaine hUHRF1 (Ubiquitin like containing PHD and RING finger domains), est une protéine nucléaire de 793 acides aminés et de masse moléculaire théorique de 89758 Da. Elle a été identifiée en 2000, par la technique du simple hybride, comme interagissant *in vitro* avec la deuxième boîte CCAAT inversée (ICB2) du promoteur du gène de la Topoisomérase II $\alpha$  humaine, d'où son nom initial ICBP90 (Inverted CCAT box binding protein of 90 KDa) (Hopfner et al., 2000). Cette protéine est impliquée dans différents processus cellulaires tels que la réplication de l'ADN, la progression du cycle cellulaire et la régulation de la transcription.

### **2.1. La famille UHRF**

La famille UHRF comprend quatre membres d'origine humaine: hUHRF1 (ou ICBP90), hUHRF2 (ou NIRF), hUHRF3 (ou ICBP55), et hUHRF4 (ou ICBP87). Trois membres d'origine murine appartiennent aussi à cette famille: mUHRF1 (ou Np95), mUHRF2 (ou Np97) et mUHRF3 (ou Np55). hUHRF1 présente 73% d'identité en acides aminés avec mUHRF1 et 53% avec hUHRF2. Les connaissances actuelles sur la famille UHRF ont été obtenues grâce à des travaux effectués sur la protéine hUHRF1 et son orthologue mUHRF1 et sur hUHRF2 (Bronner et al., 2007).

### **2.2. Expression cellulaire de hUHRF1**

L'expression du gène *UHRF1* présente une corrélation positive avec la prolifération cellulaire. Les tissus humains actifs en terme de prolifération cellulaire comme le thymus, le foie fœtal et la moelle osseuse sont les plus riches en ARNm d'UHRF1, alors que les tissus hautement différenciés comme le système nerveux central, les hépatocytes et les leucocytes n'en contiennent pas (Hopfner et al., 2000).

Dans les cellules normales, l'expression de la protéine hUHRF1 est dépendante du cycle cellulaire et présente deux pics, le premier à la fin de la phase G1 et le deuxième durant la transition G2/M. Par contre l'expression de cette protéine est constante et élevée tout au

long du cycle cellulaire dans les cellules cancéreuses. Une surexpression de hUHRF1 a été observée dans différentes cellules et tissus cancéreux comme le cancer du sein, le cancer de la prostate, les lésions cervicales, l'ostéosarcome, le rhabdomyosarcome, l'adénocarcinome pancréatique, le cancer du col de l'utérus, le cancer des poumons et les gliomes (Bronner et al., 2007; Hopfner et al., 2000; Jeanblanc et al., 2005; Jenkins et al., 2005; Lorenzato et al., 2005; Mousli et al., 2003; Unoki et al., 2004).

### **2.3. Les différents rôles de hUHRF1**

Deux fonctions ont été attribuées à la protéine hUHRF1: la première dans la progression du cycle cellulaire et la deuxième dans la régulation épigénétique de certains gènes tels les gènes suppresseurs de tumeur.

#### **2.3.1. Rôle de hUHRF1 dans la régulation du cycle cellulaire**

Les principales protéines qui jouent un rôle dans la progression du cycle cellulaire sont les sérines/thréonines kinases dépendantes des cyclines (CDKs), les cyclines, les facteurs de transcription de la famille E2F et la protéine du rétinoblastome (pRB). La protéine pRB est une protéine nucléaire qui est sous forme hypophosphorylée dans la phase de quiescence G0 et au début de la phase G1. Elle est phosphorylée au fur et à mesure de la progression dans la phase G1 et reste sous sa forme hyperphosphorylée jusqu'à la fin de la mitose. Sous sa forme hypophosphorylée, pRB est liée au facteur de transcription E2F-1 et inhibe sa fonction. Dans les cellules en prolifération, la cycline D est synthétisée au début de la phase G1 et se lie aux kinases CDK4 et CDK6 pour former des complexes actifs qui vont phosphoryler un premier site de la protéine pRB. La cycline E est ensuite synthétisée et se lie à CDK2 pour former un nouveau complexe actif qui va phosphoryler un autre site de la protéine pRB durant la phase G1. La protéine pRB hyperphosphorylée se dissocie alors d'E2F-1, lui permettant de transcrire les gènes nécessaires pour la transition G1/S et la progression du cycle cellulaire. Dans les cellules tumorales, plusieurs facteurs peuvent empêcher l'interaction pRB-E2F tels qu'une surexpression des gènes *CDK4* et *CDK6*, une perte de p16<sup>INK4A</sup> (l'inhibiteur de CDK4/CDK6), des mutations touchant le gène *RB*, une phosphorylation inappropriée de pRB due à une surexpression de la cycline D ou une interaction de pRB avec des protéines oncogènes d'un virus tumoral tel le papillomavirus humain E7 (Buolamwini, 2000; Cobrinik, 2005).

L'expression de hUHRF1 dépend de l'expression du facteur de transcription E2F-1 et une surexpression d'E2F-1 régule positivement l'expression de hUHRF1. La protéine

hUHRF1, à son tour, régule négativement la protéine pRB en se fixant sur le promoteur du gène *RB1* sous sa forme méthylée, et en interagissant avec la forme hypophosphorylée de pRB durant la phase G1. En régulant négativement pRB, hUHRF1 permet à E2F-1 de transcrire les gènes nécessaires à l'entrée dans la phase S (Jeanblanc et al., 2005). Quand l'ADN est endommagé, il y a activation de la voie de signalisation de p53 qui induit l'expression de l'inhibiteur des CDK, p21<sup>Cip/WAF1</sup>. Celui-ci inhibe le complexe cycline E/CDK2 maintenant pRB sous sa forme hypophosphorylée et liée à E2F-1. E2F-1 est alors incapable d'activer la transcription d'*UHRF1* ce qui provoque l'arrêt du cycle cellulaire en G1 et empêche la réplication de l'ADN endommagé (Arima et al., 2004; Mousli et al., 2003; Unoki et al., 2004).

Différents travaux effectués sur la protéine hUHRF1 ont montré son implication dans la régulation du cycle cellulaire surtout au niveau des transitions G1/S et G2/M. Dans des cellules normales, l'expression de hUHRF1 diminue quand les cellules entrent dans la phase S du cycle cellulaire et augmente quand elles quittent cette phase. Une surexpression de hUHRF1 dans les fibroblastes pulmonaires augmente la fraction des cellules dans les phases S et G2/M (Arima et al., 2004; Jeanblanc et al., 2005; Jenkins et al., 2005; Mousli et al., 2003; Unoki et al., 2004). Par contre, une diminution des ARNm d'hUHRF1 par des siARN dans les cellules H1299 (cellules de carcinome pulmonaire) et les cellules HeLa (cellules cancéreuses du col de l'utérus) conduit à une augmentation du nombre de cellules en phase G1 et G2 et une diminution du nombre de cellules en phase S. L'inhibition de l'expression de hUHRF1 aussi bien par des siARN, ou par l'activation de la voie de signalisation TCR (récepteurs des cellules T) ou encore par l'inhibition de celle de ERK1/2 provoque l'arrêt du cycle cellulaire au niveau de la transition G1/S (Abbadly et al., 2005; Fang et al., 2009; Unoki et al., 2004).

La Topoisomérase II $\alpha$  joue un rôle essentiel durant la réplication de l'ADN dans la phase S du cycle cellulaire. Le promoteur du gène de la Topoisomérase II $\alpha$  ne possède pas de site de liaison pour le facteur de transcription E2F, dont la présence est courante dans les gènes activés au cours du cycle cellulaire. Ce promoteur possède par contre un site de fixation pour la protéine hUHRF1 qui régule positivement l'expression du gène de la Topoisomérase II $\alpha$ . La phosphorylation de la sérine S298 de hUHRF1 par la protéine kinase A (pKA) via la voie de signalisation de l'adénosine monophosphate cyclique (cAMP) augmente l'affinité de hUHRF1 pour le promoteur du gène de la Topoisomérase II $\alpha$  et stimule la transition des cellules de la phase G0/G1 à la phase S puis aux phases G2/M (Trotzler et al., 2004). La phosphorylation de hUHRF1 par la protéine kinase 2 (CK2)

probablement au niveau de la sérine S354 a été aussi proposée dans la régulation du cycle cellulaire et la transition G1/S (Bronner et al., 2004).

Un arrêt du cycle cellulaire au niveau de la phase G2 est nécessaire pour que le parasite *Toxoplasma gondii* puisse proliférer dans les cellules hôtes après infection. Dans les cellules infectées, une augmentation de l'expression de hUHRF1 est observée. Une suppression de l'expression de hUHRF1, dans ces cellules infectées, induit un arrêt du cycle cellulaire au niveau de la phase G1. Le parasite induit donc la transition G1/S de manière dépendante de hUHRF1 et régule ensuite négativement le complexe cycline B/CDK1, induisant un arrêt du cycle cellulaire au niveau de la phase G2 (Brunet et al., 2008).

Ces observations montrent que la régulation de l'expression de hUHRF1 est essentielle pour la régulation des transitions G1/S et G2/M suggérant que l'inhibition de la protéine aurait un effet antiprolifératif (Figure 8).

### **2.3.2. Rôle de hUHRF1 dans la cancérogenèse**

Certaines propriétés du gène *UHRF1* suggère qu'il est un oncogène alors que d'autres propriétés sont celles d'un gène suppresseur de tumeur. Selon sa localisation intracellulaire, *UHRF1* pourrait avoir l'une ou l'autre des deux fonctions.

Un oncogène cible les régulateurs du cycle cellulaire surtout au niveau de la progression des phases G1 et G2. hUHRF1 jouant un rôle important dans la progression du cycle cellulaire, sa dérégulation a été associée à différents types de cancer. La surexpression de hUHRF1 permet de réprimer l'inhibition par contact des fibroblastes pulmonaires humains (Hopfner et al., 2002), de même que l'expression permanente de hUHRF1 dans les cellules cancéreuses participe à l'inhibition de l'apoptose durant la carcinogénèse (Abbadly et al., 2003). mUHRF1 présente également des propriétés oncogènes car la réexpression de cette protéine permet la réentrée dans la phase S des myotubes totalement différenciés (Bonapace et al., 2002). En reconnaissant les promoteurs méthylés des gènes suppresseurs de tumeur (*RB1*, *p16*, *p14*) et en recrutant l'histone déacétylase HDAC1, hUHRF1 joue un rôle supplémentaire dans la répression de ces gènes dans les cellules cancéreuses (Jeanblanc et al., 2005; Unoki et al., 2004). Une coopération entre hUHRF1, pRB2 et DNMT1 est nécessaire pour maintenir un profil de méthylation spécifique du gène *ER- $\alpha$*  en vue de sa répression (Macaluso et al., 2007). Il a été montré que hUHRF1 interagit avec G9a dans les cellules HeLa pour réprimer l'expression du gène *p21*. En accord avec ce résultat, des cellules souches embryonnaires mUHRF1<sup>-/-</sup> présentent un taux plus élevé de la protéine p21 indiquant un rôle de mUHRF1 dans la répression de p21 (Kim et al., 2009).

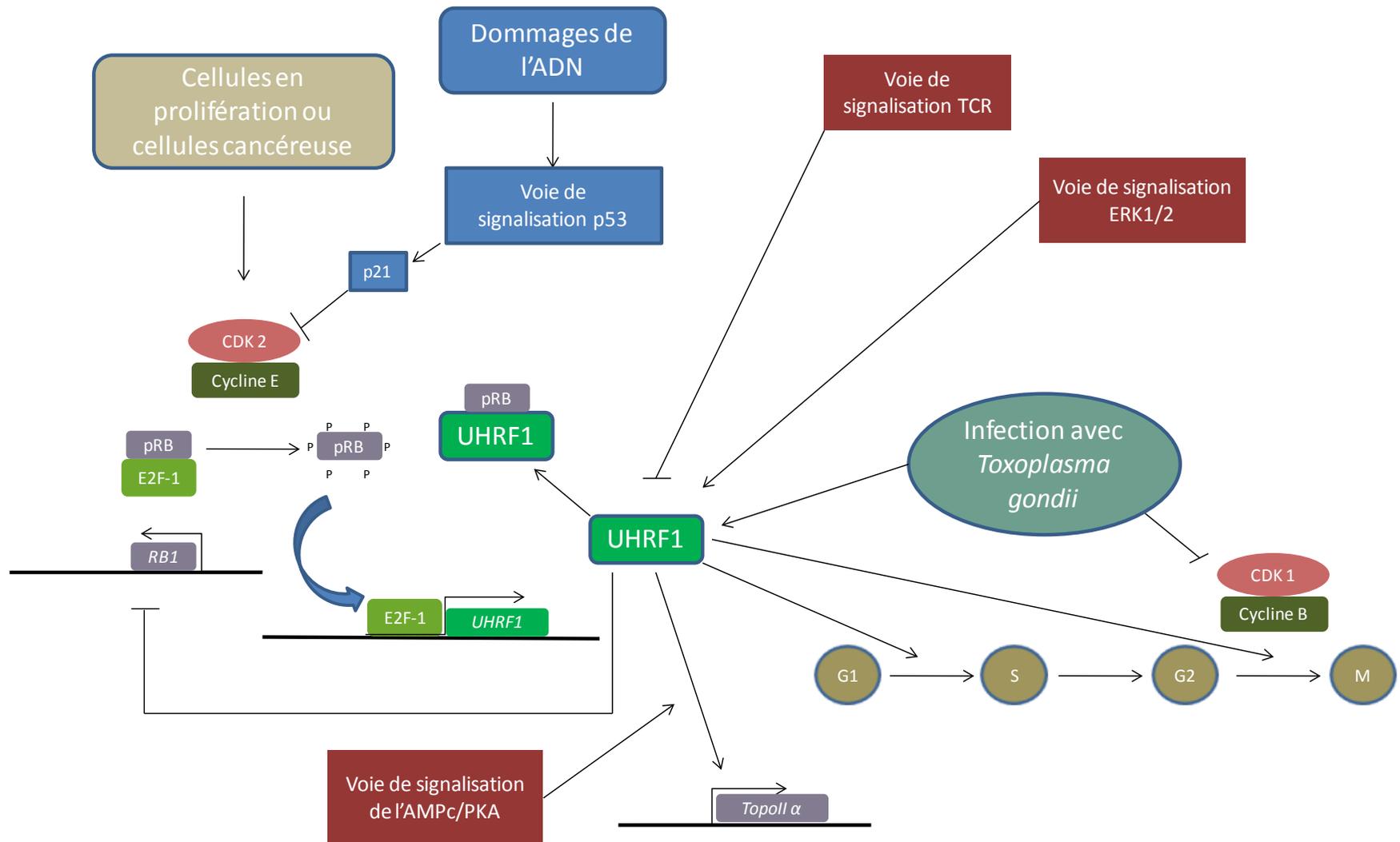


Figure 8: Implication de la protéine UHRF1 dans la régulation du cycle cellulaire au cours de la prolifération cellulaire, d'un endommagement de l'ADN ou d'une infection par le parasite *Toxoplasma gondii*.

*UHRF1* peut aussi se comporter comme un gène suppresseur de tumeur en conférant aux cellules une résistance aux agents cytotoxiques et oncogènes. Dans des cellules cancéreuses résistantes à l'hydroxyurée (HU), l'inhibition de hUHRF1 augmente la sensibilité des cellules à cet agent (Un et al., 2006). En absence de mUHRF1, les cellules murines deviennent également plus sensibles à l'endommagement de l'ADN induit par les radiations ionisantes, la lumière UV, le N-méthyl-N''-nitrosoguanidine (MNNG) et l'hydroxyurée (HU) et la réexpression de cette protéine rétablit la résistance de ces cellules vis-à-vis des agents endommageants (Muto et al., 2002). De plus, mUHRF1 contribue à préserver l'intégrité du génome, jouant un rôle dans la formation et la réplication de l'hétérochromatine péricentromérique via le recrutement de HDAC1 (Papait et al., 2007).

### **2.3.3. Rôle de hUHRF1 dans la régulation épigénétique**

Des travaux effectués sur les protéines humaine et murine (hUHRF1 et mUHRF1) ont révélé leurs rôles importants dans la régulation et le maintien du code épigénétique. mUHRF1 est associée à la chromatine *in vivo* et interagit avec les histones *in vivo* et *in vitro* avec une préférence pour l'extrémité N-terminale de l'histone H3. mUHRF1 présente une activité E3 ubiquitine ligase vis-à-vis des histones, particulièrement l'histone H3 (Citterio et al., 2004). La monoubiquitination est une modification post-traductionnelle qui sert à moduler la fonction et/ou les interactions des histones. Elle va permettre de relaxer la chromatine pour la rendre accessible aux facteurs de transcription et donc favoriser ou inhiber la transcription des gènes. Après la réplication de l'ADN, UHRF1 reconnaît l'ADN hémiméthylé néosynthétisé et recrute DNMT1 pour rétablir le profil de méthylation qui sera transmis aux cellules filles (Achour et al., 2008; Arita et al., 2008; Avvakumov et al., 2008; Bostick et al., 2007; Hashimoto et al., 2008; Qian et al., 2008; Sharif et al., 2007). hUHRF1 est capable de réprimer l'expression de certains gènes via le recrutement d'enzymes modifiant les histones telles que HDAC1 et l'histone H3K9 méthyltransférase G9a. Deux équipes ont montré que hUHRF1 interagit spécifiquement avec l'histone H3 triméthylée à la lysine 9 mais les résultats diffèrent en ce qui concerne le domaine de hUHRF1 impliqué dans cette reconnaissance (Karagianni et al., 2008; Walker et al., 2008a).

Bien que l'ordre des différentes étapes conduisant hUHRF1 à favoriser la formation de l'hétérochromatine soit encore à démontrer, un mécanisme a été proposé par Unoki et ses collaborateurs en 2009 (Unoki et al., 2009) (Figure 9).

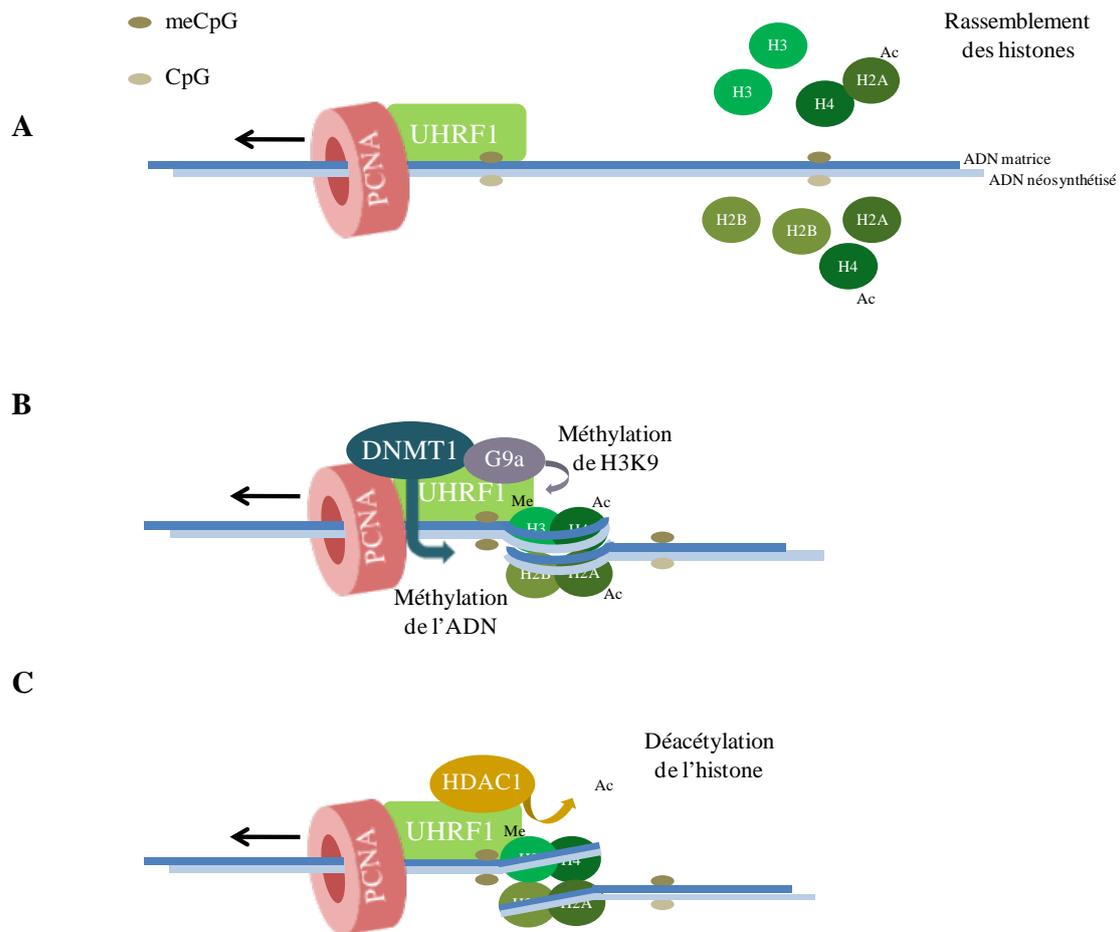


Figure 9: Mécanisme proposé par Unoki et ses collaborateurs en 2009 pour la formation de l'hétérochromatine par UHRF1 au niveau de la fourche de réplication ou d'un site de réparation de l'ADN. A- Interaction entre UHRF1 et PCNA (protéine qui améliore la processivité des ADN polymérases chez les eucaryotes lors de la réplication de l'ADN) et reconnaissance des dinucléotides CpG hémiméthylés via hUHRF1. B- Recrutement de DNMT1 pour méthyle le brin néosynthétisé et rétablir le profil de méthylation de la cellule mère. UHRF1 recrute G9a pour méthyle l'histone. C- UHRF1 recrute HDAC1 pour déacétyler les histones permettant leur interaction avec l'ADN et la formation de l'hétérochromatine (Unoki et al., 2009).

## 2.4. Architecture de la protéine hUHRF1

L'analyse de la séquence en acides aminés a permis d'identifier les domaines protéiques (Figure 10). mUHRF1, mUHRF2 et hUHRF2 présentent la même architecture que hUHRF1. Par contre mUHRF3 et hUHRF3 ne possèdent pas de domaine NIRF-N et hUHRF4 ne possède pas de domaine RING (Mousli et al., 2003).

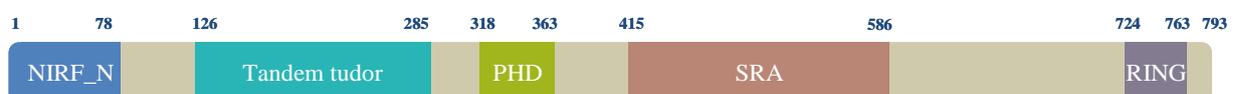


Figure 10: Les cinq domaines de hUHRF1.

### 2.4.1. Les domaines NIRF-N, Tandem tudor et RING de hUHRF1

Dans la partie amino-terminale, un domaine de type ubiquitine NIRF-N (ou UBL pour ubiquitine like) est défini par les acides aminés 1 à 78. Ce domaine, dont la fonction reste mal connue, présente 35% d'identité de séquence avec l'ubiquitine, une protéine impliquée dans la dégradation protéique et la transcription de gènes. La structure de ce domaine, résolue en 2005 par diffraction des rayons X (pdb: 2faz) (Walker et al., 2005), montre un repliement typique d'ubiquitine avec une surface riche en lysines conduisant Bronner à proposer un rôle potentiel dans des interactions protéine-protéine (Bronner et al., 2007) (Figure 11A).

En aval, entre les acides aminés 126 et 285 un domaine tandem tudor a été identifié en 2008. La structure du tandem tudor en complexe avec un peptide de l'histone H3 triméthylé à la lysine 9 (H3K9me3) a été résolue en 2008 (pdb: 3db3) (Walker et al., 2008a). La structure montre une reconnaissance de la lysine méthylée via une cage aromatique formée de deux tyrosines Y188 et Y191 et une phénylalanine F152 (Figure 11B). Ce domaine est généralement associé à la régulation de la transcription, la réparation de l'ADN et le remodelage de la chromatine. Le rôle exact du domaine tudor au sein de hUHRF1 et dans la reconnaissance de la chromatine n'est pas encore documenté.

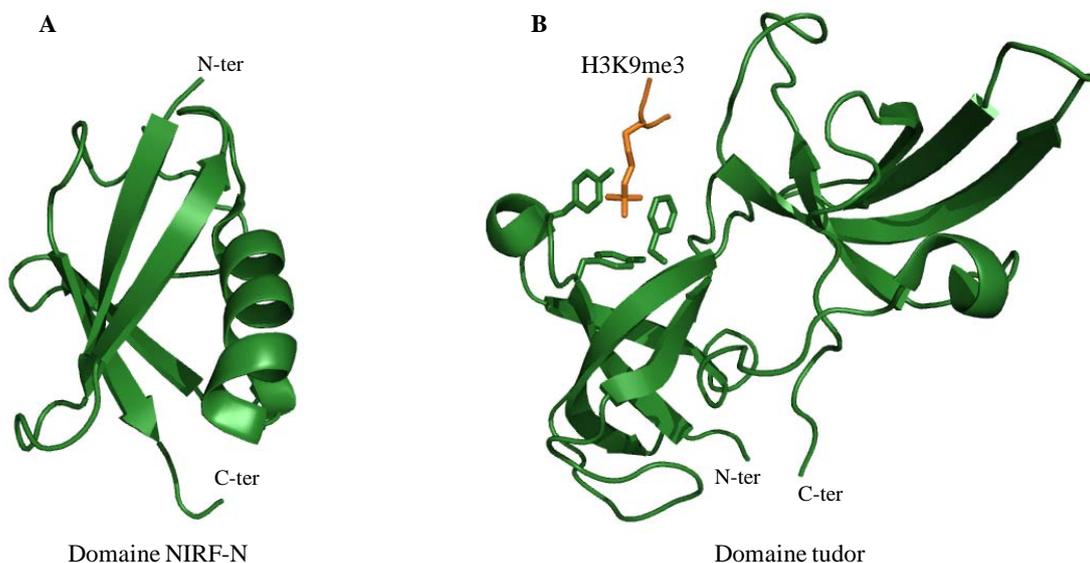


Figure 11: Structures des domaines NIRF-N et tudor de hUHRF1. A- Structure cristallographique du domaine amino-terminal NIRF-N (PDB: 2faz). B- Structure cristallographique du domaine tudor (en vert) en complexe avec un peptide de l'histone H3K9me3 (en orange). La lysine triméthylée est reconnue par une cage aromatique (PDB : 3db3).

La partie carboxy-terminale de hUHRF1 (résidus 724-763) présente une séquence consensus pour un domaine en doigt de zinc mais de type RING (Really interesting New

Gene). Ce domaine présente la signature C3HC4 (3 cystéines, 1 histidine puis 4 cystéines) qui permet la coordination de deux atomes zinc. Le domaine RING possède une activité ubiquitine E3 ligase qui catalyse la dernière étape d'une cascade enzymatique permettant l'ajout d'ubiquitine sur des lysines d'une protéine cible, la marquant ainsi en vue de sa dégradation par le protéasome (polyubiquitination) ou de la régulation de sa fonction (monoubiquitination). Par son domaine RING, mUHRF1 est capable de s'auto-ubiquitiner *in vitro* et d'ubiquitiner les queues des histones *in vitro* et *in vivo* avec une préférence pour l'histone H3 (Citterio et al., 2004). Ce domaine présente plus d'affinité pour G9a que le domaine SRA (résidus 415-586) de hUHRF1 (Kim et al., 2009). La structure cristallographique de ce domaine a été résolue en 2008 (pdb 3fl2) (Walker et al., 2008b) (Figure 12).

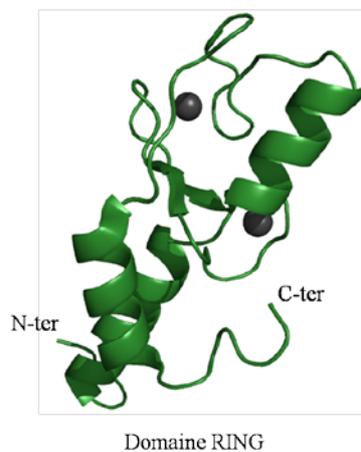


Figure 12: Structure cristallographique du domaine RING de hUHRF1 (PDB: 3fl2).

#### 2.4.2. Le domaine SRA de hUHRF1

Le domaine SRA se trouve entre les acides aminés 415 et 586. Contenant un motif YDG (Tyr-Asp-Gly), SRA est aussi connu sous le nom de «domaine YDG ». Ce domaine est essentiellement trouvé dans des protéines de plantes et uniquement dans la famille UHRF chez les vertébrés. Différentes fonctions ont été attribuées à ce domaine dans UHRF1. Il a été montré que le domaine SRA reconnaît les dinucléotides CpG méthylés des régions promotrices de différents gènes suppresseurs de tumeurs et recrute l'histone déacétylase HDAC1 (Unoki et al., 2004). Le domaine SRA participe aussi à l'interaction avec l'histone H3K9 méthyltransférase G9a (Kim et al., 2009). Les domaines SRA de hUHRF1 et de mUHRF1 jouent un rôle important dans la maintenance du profil de méthylation après la réplication de l'ADN en reconnaissant l'ADN hémiméthylé et en recrutant l'ADN méthyltransférase DNMT1 pour reconstituer le profil de méthylation (Achour et al., 2008;

Bostick et al., 2007; Sharif et al., 2007). La structure cristallographique du domaine SRA d'UHRF1 seul et en complexe avec un ADN hémiméthylé, résolue par plusieurs équipes en 2008, a permis de proposer le mécanisme moléculaire de maintenance du profil de méthylation de l'ADN (Arita et al., 2008; Avvakumov et al., 2008; Delagoutte et al., 2008; Hashimoto et al., 2008; Qian et al., 2008). Le domaine SRA de UHRF1 adopte une conformation en forme de main, dont le « doigt NKR: Asn489-Lys490-Arg491 » et le « pouce: His445-Val446 » font basculer une cytosine méthylée en dehors de la double hélice de l'ADN. Cette cytosine méthylée est alors reconnue dans une cage hydrophobe formée par deux tyrosines (Y466 et Y478) (Figure 13). Après la reconnaissance de la cytosine méthylée, le domaine SRA recrute DNMT1 qui reconnaît alors la cytosine non méthylée dans le site CpG et la méthyle. Dans le travail publié par Karagianni et ses collaborateurs en 2008, il a été montré que le domaine SRA joue le rôle prépondérant dans l'affinité de l'interaction avec l'histone H3K9me3 (Karagianni et al., 2008).

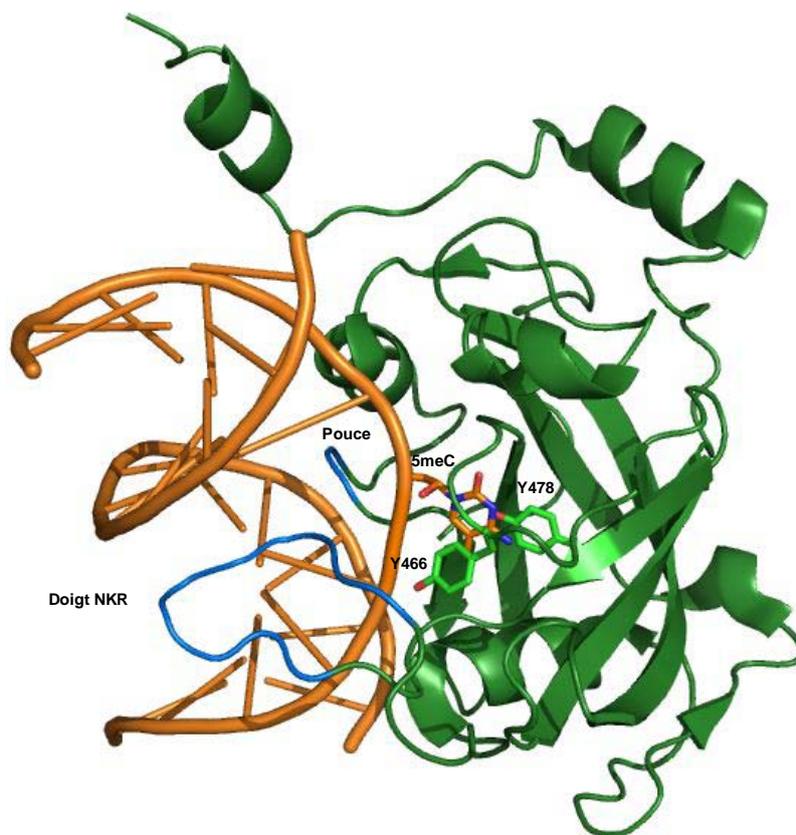


Figure 13: La reconnaissance de la cytosine méthylée d'un dinucléotide CpG hémiméthylé par le domaine SRA de UHRF1 (PDB: 3clz). Le doigt NKR et le pouce du domaine SRA (en bleu) font sortir la cytosine méthylée de son emplacement dans la double hélice d'ADN (en orange), lui permettant ainsi d'être fixée par les deux tyrosines Y466 et Y478 de SRA (en vert).

### 2.4.3. Le domaine PHD de hUHRF1

Le domaine PHD correspond aux résidus 318-363. Des expériences de « pull-down » sur des extraits cellulaires ont montré que le domaine PHD de hUHRF1 reconnaît spécifiquement l'histone H3 di ou triméthylée à la lysine 9 (H3K9me2 et H3K9me3) (Karagianni et al., 2008). Mais aucune évidence d'une interaction directe entre le domaine PHD de hUHRF1 et un peptide H3K9me3 n'a été publiée. Ce domaine semble également être le domaine d'interaction avec DNMT1. Des essais de « GST pull-down » montrent que le domaine PHD de mUHRF1 est capable de retenir DNMT1 (Bostick et al., 2007). Le domaine PHD de mUHRF1 est aussi impliqué dans la réorganisation de l'hétérochromatine périphérique. Les régions hétérochromatiques s'associent progressivement pour former les chromocentres. Au cours de la phase S, ces chromocentres très compacts doivent être ouverts pour permettre à la machinerie de réplication d'accéder à l'ADN et pour reconstituer les marques épigénétiques nécessaires pour la répression de cette région. C'est le domaine PHD de mUHRF1 qui porte la capacité de désagréger ces chromocentres (Papait et al., 2008). Aucune structure du domaine PHD de hUHRF1 ou de mUHRF1 n'est disponible. Par contre, la structure RMN du domaine PHD de hUHRF2 a été résolue en 2006 (PDB: 2e6s) (Kadirvel et al., 2006) (Figure 14).

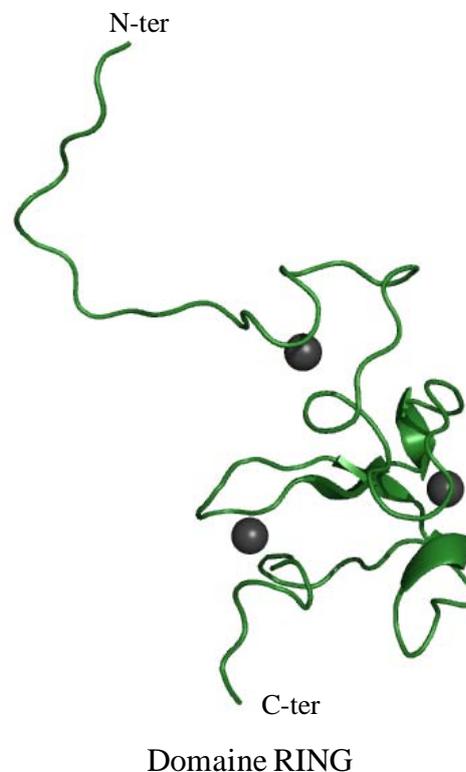


Figure 14: Structure RMN du domaine PHD de hUHRF2 (PDB: 2e6s)

### 3. Domaines PHD et régulation épigénétique

Les domaines à doigts de zinc sont très abondants dans les protéines d'eucaryotes supérieurs et ont d'abord été caractérisés comme reconnaissant des séquences spécifiques d'ADN. Le premier domaine à doigt de zinc a été identifié en 1985 (Miller et al., 1985) et depuis lors plusieurs classes de doigts de zinc ont été définies comme des modules de reconnaissance des ARN et des protéines (Gamsjaeger et al., 2007). La classification de ces domaines dépend de la topologie et de la signature de fixation des zinc et du nombre d'ions zinc fixés. Les doigts de zinc « classiques » fixent un seul zinc avec une signature C2H2 ou C2HC. Certains doigts de zinc sont capables de fixer 2 zinc de manière séquentielle tels les domaines de type LIM et MYND, ou de manière entrecroisée tels les domaines de type RING ou PHD. Les doigts de zinc de type TAZ fixent séquentiellement 3 zincs (Figure 15).

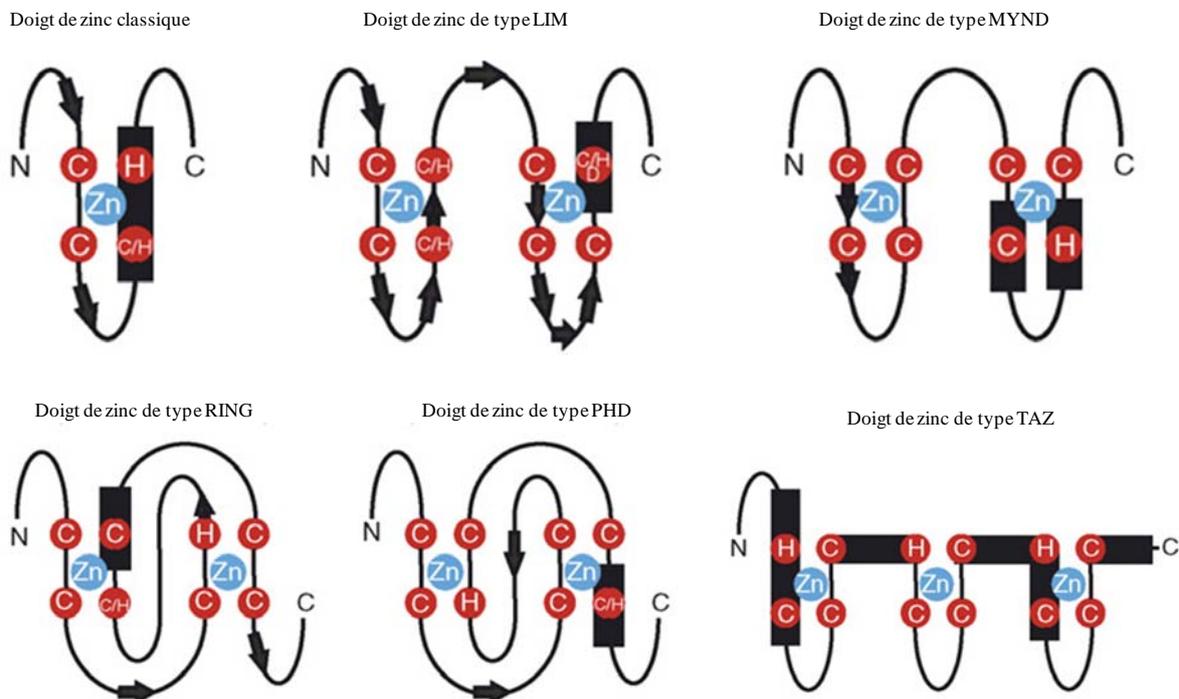


Figure 15: Différentes classes des domaines à doigts de zinc (Gamsjaeger et al., 2007).

#### 3.1. Caractéristiques structurales et fonctionnelles des domaines PHD

##### 3.1.1. Les caractéristiques structurales

Le premier domaine PHD a été identifié en 1993 dans *Arabidopsis* comme un nouveau motif protéique très conservé entre deux protéines homéodomaines des plantes. Les protéines homéodomaines sont des facteurs de transcription contenant un motif hélice-

tournant-hélice capable de fixer l'ADN. Les domaines PHD contiennent huit résidus cystéine/histidine espacés de façon régulière et capable de fixer deux atomes zinc (Schindler et al., 1993). Ils sont formés d'environ 60 résidus avec une signature typique C4HC3 (4 cystéines, 1 histidine, 3 cystéines) mais il existe aussi des PHD atypiques tels le PHD de RAG2 ayant une signature C3H2C2H et le PHD2 de JADE1 ayant la signature C4HC2H. D'autres résidus sont conservés entre les différents PHD tel un tryptophane, ou un autre résidu aromatique, avant la dernière paire de cystéines (Figure 16). Les structures des PHD montrent une topologie entrecroisée pour la fixation des zinc, ce qui signifie que des paires de cystéines consécutives dans la séquence primaire ne fixent pas le même zinc (Figures 17). Les PHD présentent des éléments de structures secondaires comprenant deux courts brins  $\beta$  antiparallèles et parfois une courte hélice. Plusieurs centaines de protéines contenant des PHD ont été identifiées dans le génome humain. Les domaines PHD présentent deux boucles très variables en nombre de résidus et en séquences (boucles 1 et 2).

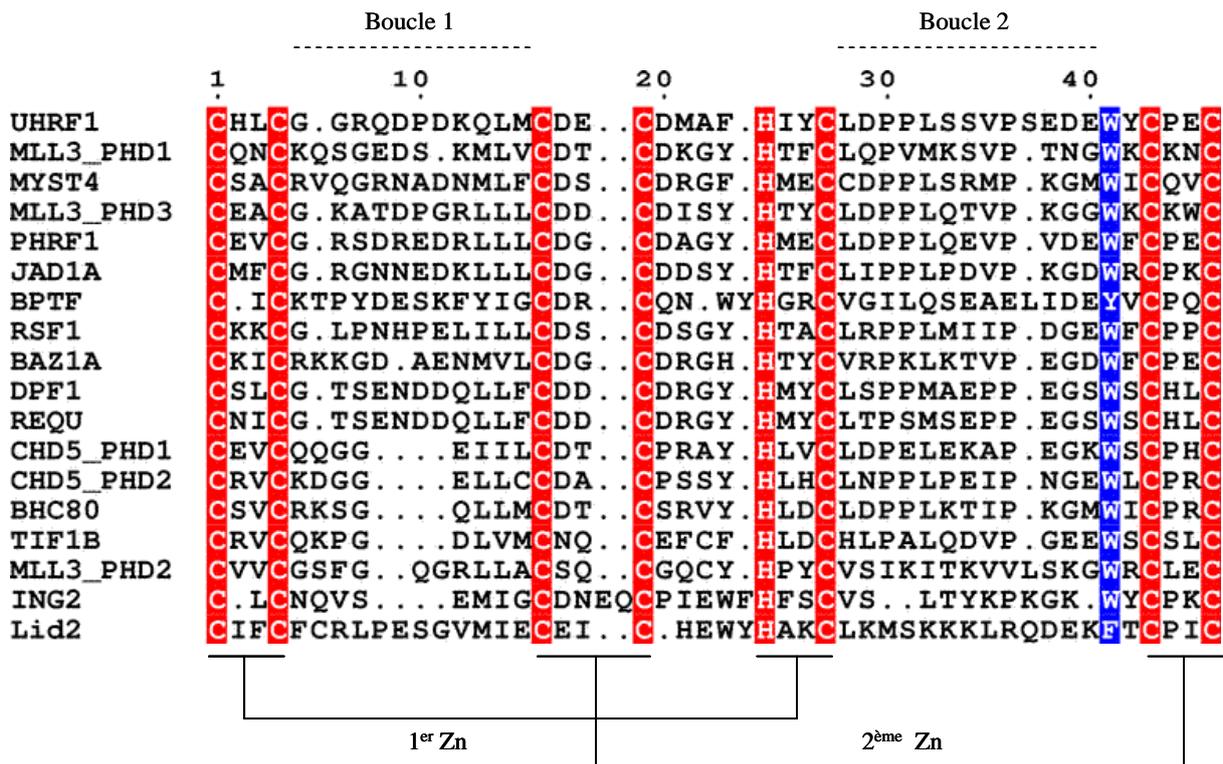


Figure 16: Alignement de séquences de différents PHD humains. Les résidus impliqués dans la fixation des zincs sont encadrés en rouge et le résidu aromatique conservé est encadré en bleu.

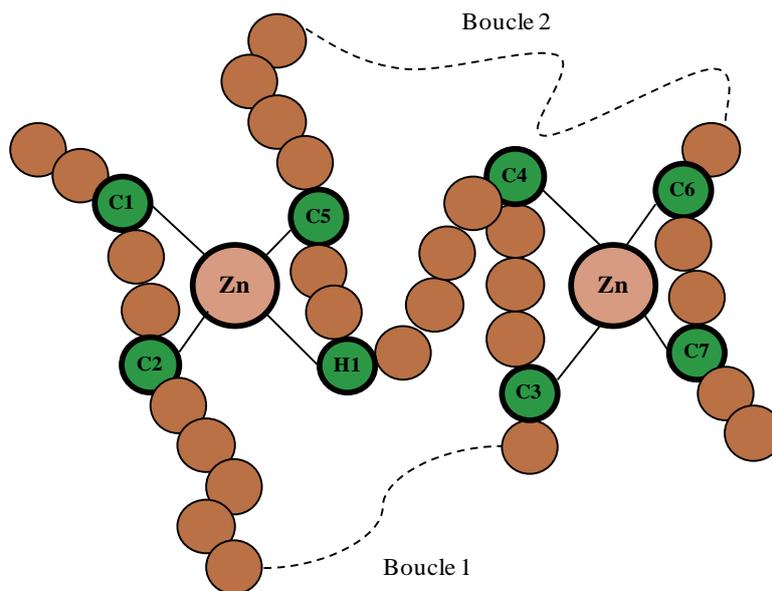


Figure 17: la topologie entrecroisée pour la fixation des zinc dans les domaines PHD (Bienz, 2006).

### 3.1.2. Les caractéristiques fonctionnelles des PHD

La plupart des domaines PHD sont trouvés dans des protéines nucléaires qui jouent un rôle dans la régulation de la chromatine (Bienz, 2006). Certains PHD ont été associés à la fixation des phosphoinositides. C'est le cas du PHD de la protéine ING2. Cette interaction nécessite deux régions riches en résidus basiques: la première correspond à la boucle 2 et la deuxième à la région en aval du domaine PHD. D'autres PHD sont aussi capables de fixer des phosphoinositides tels le PHD de WSTF, d'AIRE1 et d'ING1 (Gozani et al., 2003; Jones et al., 2006). Le PHD du corépresseur KAP1 est le seul domaine PHD, jusqu'à présent, ayant une fonction E3 SUMO ligase. Ce PHD interagit avec l'enzyme SUMO E2, Ubc9, et permet la sumoylation d'un bromodomaine adjacent (Ivanov et al., 2007). Aucune signature propre à cette fonction n'est encore identifiée. Récemment des domaines PHD ont été identifiés comme des modules de « lecture » de certaines marques épigénétiques capables de recruter ensuite des complexes de remodelage de la chromatine selon la marque d'histone reconnue (Eberharter et al., 2004; Ragvin et al., 2004; Taverna et al., 2007). Certaines protéines peuvent contenir plusieurs domaines PHD sans qu'ils aient tous nécessairement la même fonction. La protéine Lid2 par exemple contient 3 domaines PHD dont seul le deuxième est capable de reconnaître des peptides de l'histone H3 (Li et al., 2008).

## **3.2. Les PHD reconnaissant des marques spécifiques de l'histone H3**

Plusieurs domaines PHD ont été identifiés comme des modules de reconnaissance de l'état de méthylation de la lysine 4 de l'histone H3, sous ses formes non modifiée (H3K4me0) ou méthylée (H3K4me). Les modes de reconnaissance de ces deux marques épigénétiques par les domaines PHD ont été caractérisées structurellement depuis 2006 conduisant à l'identification des résidus essentiels pour la reconnaissance de ces deux marques sur K4. Des études récentes ont montrées que certains domaines PHD reconnaissent eux la lysine 9 méthylée (H3K9me) et une autre classe de PHD reconnaissant la lysine 36 méthylée (H3K36me) a été proposée par une étude protéomique. Cette classe comprend jusqu'à présent deux protéines, Ecm5 et Nto1 (Shi et al., 2007). La plupart de ces domaines PHD appartiennent à des protéines faisant partie des complexes de remodelage de la chromatine, ou permettant la régulation de la transcription par activation ou répression de l'expression de certains gènes selon la marque épigénétique reconnue. La protéine ING2, par exemple, reconnaît H3K4me3 via son PHD et fait partie du complexe histone déacétylase mSin3a-HDAC1. En réponse à l'endommagement de l'ADN, la reconnaissance de H3K4me3 par le PHD d'ING2 stabilise le complexe mSin3a-HDAC1 au niveau des promoteurs de certains gènes pour les réprimer (Shi et al., 2006).

### **3.2.1. Les PHD reconnaissant la lysine 4 non modifiée**

Jusqu'à présent deux domaines PHD appartiennent à cette classe: celui de BHC80 (ou PHF21A) et le premier PHD (PHD1) de la protéine AIRE. La protéine BHC80 fait partie d'un complexe contenant LSD1, une déméthylase de l'histone H3K4me3 qui joue un rôle dans la répression de certains gènes. Après la déméthylation par LSD1, BHC80 reconnaît H3K4me0 pour empêcher sa reméthylation. La liaison de BHC80 à l'histone H3 est inhibée par la méthylation de la lysine K4 mais reste insensible aux modifications touchant la lysine K9 ou K14. La structure cristallographique du complexe entre le PHD de BHC80 et H3K4me0 (1-10) a été résolue à 1.43 Å de résolution (PDB: 2puy) (Figure 18A et B). Le peptide se fixe en formant un brin  $\beta$  supplémentaire à la surface du PHD. Sur l'histone H3, l'alanine en position 1 (H3A1) est fixée par des liaisons hydrogènes entre sa chaîne principale et les carbonyles des résidus P523, K524 et G525 de la protéine. L'acide aspartique D489 est le résidu essentiel qui reconnaît H3K4 via sa chaîne latérale. L'histidine H487 est positionnée de manière à empêcher la fixation d'une lysine méthylée (Lan et al., 2007).

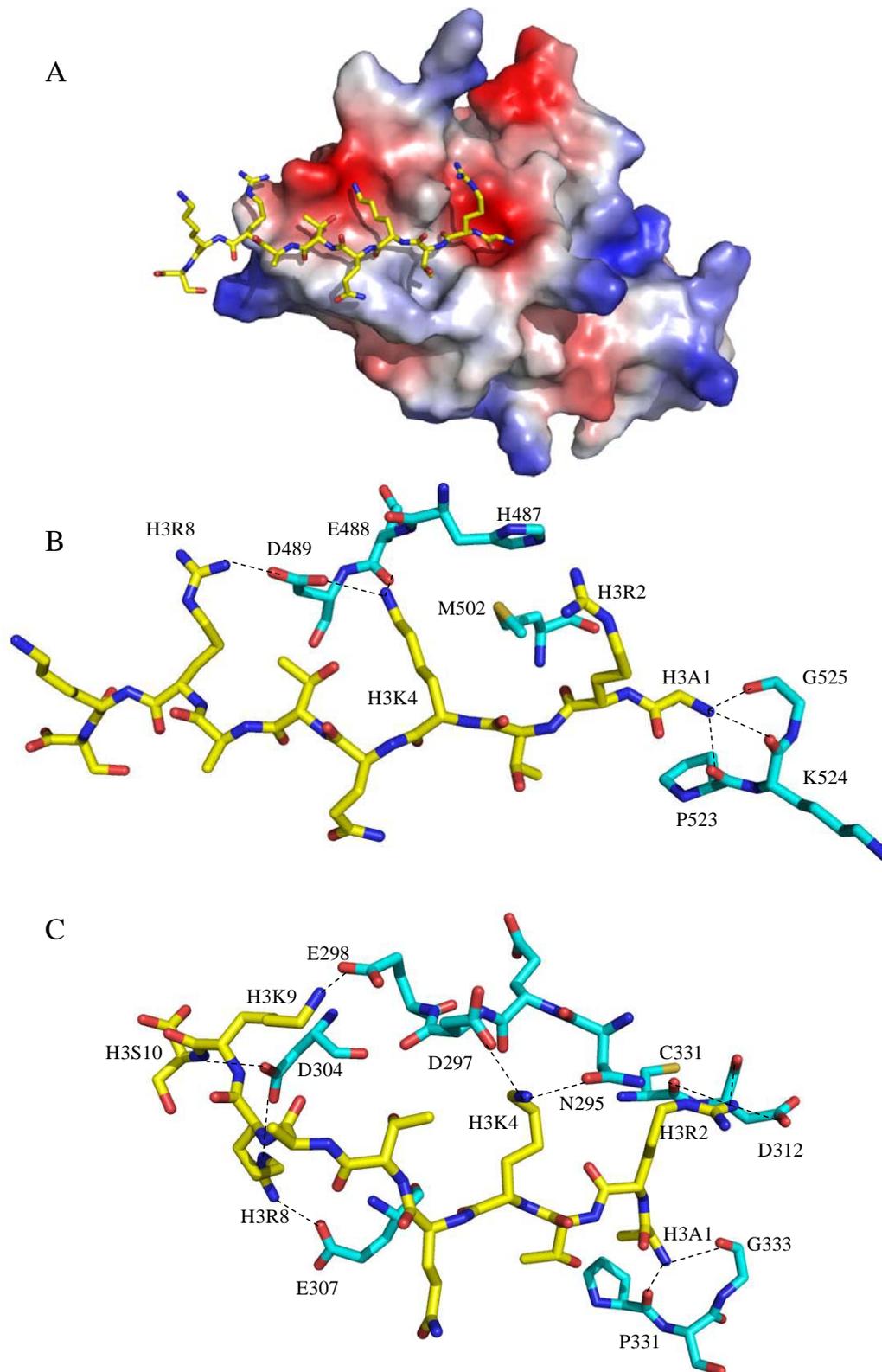


Figure 18: Mode de reconnaissance de l'histone H3K4me0 par les domaines PHD de BHC80 (PDB: 2puy) et AIRE (PDB: 2ke1). A- Représentation de la surface électrostatique du PHD de BHC80 en complexe avec le peptide de H3: la lysine K4 non modifiée est fixée dans une poche chargée négativement. B- Les liaisons hydrogènes et les interactions électrostatiques stabilisant le complexe BHC80-H3K4me0. C- Les liaisons hydrogènes et les interactions électrostatiques stabilisant le complexe AIRE1-H3K4me0.

La protéine AIRE est un régulateur autoimmunitaire humain qui joue un rôle dans la prévention de l'autoimmunité. Cette protéine contient deux domaines PHD. Des mutations touchant cette protéine sont associées à la maladie auto-immune APECED. Seul le domaine PHD1 est capable de fixer l'histone H3 avec la lysine K4 non modifiée (Org et al., 2008). En fixant l'histone H3K4me0, AIRE est capable d'activer la transcription en recrutant des coactivateurs transcriptionnels tel CBP (Creb Binding Protein) au niveau de cette marque répressive. Deux structures RMN du complexe PHD1-H3K4me0 ont été résolues en 2009 (PDB: 2ke1 et 2kft) (Chakravarty et al., 2009; Chignola et al., 2009). Le peptide H3 forme également un brin  $\beta$  supplémentaire à la surface du PHD et la lysine K4 non modifiée est fixée par des liaisons hydrogènes avec l'acide aspartique D297 et l'asparagine N295. Les 9 premiers résidus de l'histone établissent différentes interactions électrostatiques avec des résidus de la surface du PHD stabilisant ainsi le complexe (Figure 18C).

Un alignement de séquences de cette classe de PHD montre une conservation d'un acide aspartique situé deux résidus avant la première cystéine impliquée dans la coordination du premier zinc. Il s'agit du D489, dans le cas de BHC80 et du D297 dans le cas d'AIRE, qui au niveau structural, sont impliqués dans la liaison à la lysine K4 de H3. En janvier 2009, Saksouk et ses collaborateurs ont montré par des expériences de « pull down » que le premier PHD de la protéine JADE1 interagit avec l'histone H3 et que cette interaction est inhibée par la méthylation de la lysine 4 (Saksouk et al., 2009). La séquence de ce PHD montre la présence d'un acide aspartique en position -3 de la première cystéine. Ce PHD pourrait donc appartenir à cette classe de PHD (Figure 19).

	1	10	20	30	40	50
BHC80	I H E D F . C S V C R K S G Q L L M . . . . . C D T C S R V Y H L D C L D P P L K T I P K G M W I C P R C Q D Q M L K K					
AIRE_PHD1	K N E D E . C A V C R D G G E L I C . . . . . C D G C P R A F H L A C L S P P L R E I P S G T W R C S S C L Q A T V Q E					
JADE1_PHD1	. D E D V V C D V C Q S P D G E D G N E M V F C D K C N I C V H Q A C Y G I L K V P E . . G S W L C R T C A L G V Q P K					

Figure 19: Alignement de séquences des PHD de BHC80, AIRE1 et JADE1. Les résidus impliqués dans la coordination des zinc sont en rouge et l'acide aspartique impliqué dans la reconnaissance de H3K4me0 est en vert.

### 3.2.2. Les PHD reconnaissant la lysine 4 méthylée

Une deuxième classe de PHD reconnaît spécifiquement l'histone H3K4me3 qui est une marque épigénétique associée à l'activation de la transcription. Plusieurs structures de domaines PHD appartenant à cette classe sont connues comme celles des PHD de la famille ING, de Yng1, de hPygo1, de RAG2, de BPTF (PHD2), de TAF3 et de JARID1A (PHD3) (Champagne et al., 2008; Fiedler et al., 2008; Li et al., 2006; Liu et al., 2007; Matthews et al.,

2007; Palacios et al., 2008; Pena et al., 2006; Pena et al., 2008; Ramon-Maiques et al., 2007; Taverna et al., 2006; van Ingen et al., 2008; Wang et al., 2009). Cette classe de PHD reconnaît simultanément H3K4me3 et H3R2 dans deux poches adjacentes (Figure 20A). Les deux poches sont séparées par un tryptophane qui est conservé dans cette classe de PHD (Figure 21). Des acides aminés hydrophobes (pour la plupart des aromatiques) forment une cage autour du groupement triméthylammonium et le lient par des interactions cation- $\pi$  et de van der Waals. Dans le cas de BPTF, le groupement méthyle entre dans une cage formée par 4 résidus: trois tyrosines Y10, Y17 et Y23 et le tryptophane W32. Dans le PHD d'ING2, deux résidus aromatiques Y215, W238 et une méthionine M226 forment l'environnement du groupement méthyle de la lysine. Pour RAG2, seuls deux résidus aromatiques Y415 et W453 sont impliqués dans la liaison du groupement méthyle formant dans ce cas un canal et non une cage. La chaîne latérale de l'arginine H3R2 interagit par des interactions électrostatiques avec au moins un résidu acide. Il s'agit de l'acide aspartique D27 et de la glutamine Q30 dans BPTF et de l'acide aspartique D230 et de l'acide glutamique E237 dans ING2. Dans cette classe de PHD, une méthylation de l'arginine 2 diminue l'affinité de l'interaction avec le peptide H3K4me. Une exception est le PHD de RAG2 qui est capable de reconnaître la double méthylation de la lysine 4 et de l'arginine 2. Le PHD de RAG2 n'a pas de poche ni de résidus acides capables de fixer l'arginine R2. De plus la position de la tyrosine Y445 empêche la chaîne latérale de l'arginine d'interagir avec la protéine et celle-ci pointe alors vers le solvant (Figure 20B).

L'alignement de séquences entre différents PHD fixant l'histone H3K4me3 montre une conservation des résidus aromatiques ou hydrophobes dans cette classe de PHD. Ces résidus se trouvent en position -2 ou -3 de la première cystéine, -3 de la troisième cystéine et -2 de l'histidine impliquées dans la coordination des zinc. Chaque PHD possède au moins deux résidus aromatiques conservés et dans BPTF, une tyrosine supplémentaire en position +4 de la deuxième cystéine est également présente. Le résidu aromatique qui est conservé dans toutes les classes de PHD n'est pas impliqué dans la reconnaissance de la lysine triméthylée. De même il y a conservation d'un résidu acide en position +1 de la troisième cystéine. Cet acide est celui impliqué dans la fixation de l'arginine 2 (Figure 21). Le PHD de la protéine Orc1 a été identifié par des expériences de « pull down » comme interagissant avec l'histone H3K4me3 mais ce PHD ne contient pas dans sa séquence les résidus aromatiques conservés (de la Paz Sanchez and Gutierrez, 2009).

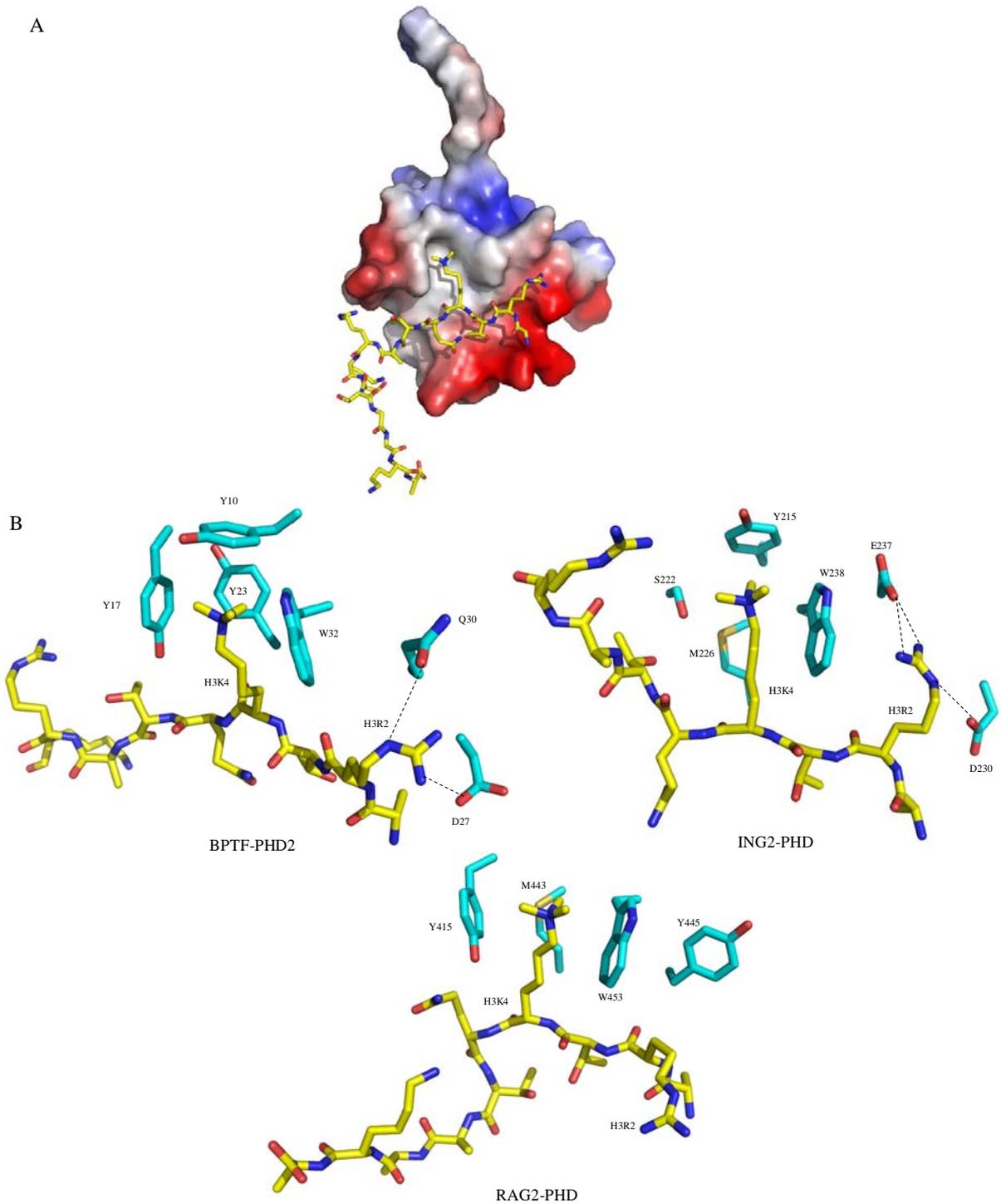


Figure 20: Mode de reconnaissance de l'histone H3K4me3 par les domaines PHD de BPTF (PDB: 2fuu), d'ING2 (PDB: 2g6q) et de RAG2 (PDB: 2v89). A- Représentation de la surface électrostatique du PHD de BPTF en complexe avec le peptide de H3: H3K4me3 est fixée dans une poche hydrophobe et H3R2 dans une poche adjacente électronégative. B- Comparaison entre la fixation de la lysine K4me3 et de l'arginine R2 par les PHD de BPTF, d'ING2 et de RAG2.

	1	10	20	30	40	50																																																																			
Taf3	E	W	G	N	Q	I	W	I	C	P	G	C	N	K	P	D	D	G	S	P	M	I	G	C	D	D	C	D	D	W	Y	H	W	F	C	V	G	I	M	A	A	P	P	E	E	M	Q	W	F	C	P	K	C	A																			
BPTF	G	S	D	T	K	L	Y	C	I	C	K	T	P	Y	D	E	S	K	F	Y	I	G	C	D	R	C	Q	N	W	Y	H	G	R	C	V	G	I	L	Q	S	E	A	E	L	I	D	E	Y	V	C	P	Q	Q																				
ING1	D	P	N	E	P	T	Y	C	L	C	N	Q	V	S	Y	G	E	M	I	G	C	D	N	E	Q	C	P	I	E	W	F	H	F	S	C	V	S	L	T	Y	K	P	K	G	K	W	Y	C	P	K	C	R																					
ING2	D	P	N	E	P	T	Y	C	L	C	N	Q	V	S	Y	G	E	M	I	G	C	D	N	E	Q	C	P	I	E	W	F	H	F	S	C	V	S	L	T	Y	K	P	K	G	K	W	Y	C	P	K	C	R																					
ING4	G	S	N	E	P	T	Y	C	L	C	H	Q	V	S	Y	G	E	M	I	G	C	D	N	P	D	C	S	I	E	W	F	H	F	A	C	V	G	L	T	T	K	P	R	G	K	W	F	C	P	R	C	S																					
ING5	G	S	N	E	P	T	Y	C	L	C	H	Q	V	S	Y	G	E	M	I	G	C	D	N	P	D	C	P	I	E	W	F	H	F	A	C	V	D	L	T	T	K	P	K	G	K	W	F	C	P	R	C	V																					
ING3	Y	S	V	C	A	A	Q	N	C	I	C	N	Q	V	S	Y	G	E	M	V	G	C	D	N	Q	D	C	P	I	E	W	F	H	Y	C	V	G	L	T	E	A	P	K	G	K	W	Y	C	P	Q	C	T																					
JARID1A	S	V	C	A	A	Q	N	C	I	C	N	Q	R	P	C	K	D	K	V	D	W	V	Q	C	D	G	G	C	D	E	W	F	H	Q	V	C	V	G	V	S	P	E	M	A	E	N	E	D	Y	I	C	I	N	C	A																		
Pygo1	S	S	S	D	P	V	Y	P	C	G	I	C	T	N	E	V	N	D	D	Q	D	A	I	L	C	E	A	S	C	Q	K	W	F	H	R	I	C	T	G	M	T	E	T	A	Y	G	L	L	T	A	E	A	S	A	V	W	G	D	T	C	M												
RAG2	G	S	P	E	F	G	Y	W	I	T	C	P	T	C	D	V	D	I	N	T	W	V	P	F	Y	S	T	E	L	N	K	P	A	M	I	Y	C	S	H	G	D	G	H	W	V	H	A	Q	C	M	D	L	E	E	R	T	L	I	H	L	S	E	G	S	N	K	Y	Y	C	N	E	H	V

Figure 21: Alignement de séquences de différents PHD fixant H3K4me3. Les résidus impliqués dans la coordination des zinc sont surlignés en rouges et le résidu aromatique conservé dans tous les PHD est en bleu. Les résidus hydrophobes impliqués dans la fixation de H3K4me3 sont surlignés en jaune. L'acide aspartique ou glutamique impliqué dans la fixation de H3R2 est surligné en bleu cyan.

### 3.2.3. Les PHD reconnaissant la lysine 9 méthylée

Certains domaines PHD sont capables de reconnaître l'histone H3 triméthylée à la lysine 9. Cette marque épigénétique est associée à la répression des gènes. Jusqu'à présent, cette classe de PHD n'est pas caractérisée structuralement et le mode d'interaction entre le PHD et la lysine K9 triméthylée n'est pas connu. La protéine Lid2 joue un rôle dans la régulation croisée de la méthylation de H3K4 et H3K9, étant capable de recruter l'histone H3K4 méthyltransférase Set1 et l'histone H3K9 déméthylase LSD1. Lid2 contient trois domaines PHD. Par des expériences de « pull down » il a été montré que Lid2 reconnaît l'histone H3K9me3 via son domaine PHD2, recrute Set1 pour méthyler la lysine K4 et recrute LSD1 pour déméthyler la lysine K9, permettant ainsi l'activation de la transcription (Li et al., 2008). Un deuxième exemple concerne le domaine PHD de la protéine SMCX. Cette protéine est capable de déméthyler H3K4me3 et H3K4me2 via son domaine jmjC. Cette protéine comprend deux domaines PHD mais seul PHD1 est capable de fixer l'histone H3K9me3 par des expériences de « pull down ». La reconnaissance de K9me3 via PHD1 pourrait favoriser la déméthylation de K4me3 via le domaine jmjC et donc induire la répression de la transcription (Iwase et al., 2007). Concernant le domaine PHD2 de la protéine CDH4, des mesures de fluorescence ont montré qu'il est capable d'interagir avec l'histone H3 non modifiée ( $K_d = 18 \mu\text{M}$ ) mais son affinité augmente lorsque la lysine 9 est méthylée ( $K_d = 0.9 \mu\text{M}$ ) ou acétylée ( $K_d = 0.6 \mu\text{M}$ ). Des expériences de titration par RMN ont montré que ce sont les mêmes résidus de la protéine qui sont impliqués dans la fixation des trois peptides. En comparant les résultats obtenus avec les peptides de H3 non modifiée et méthylé sur K9, trois résidus présentent une amplitude de déplacement chimique plus élevée avec le peptide méthylé sur K9, ce sont D454, G455 et G456 (Figure 22). Sur un modèle du complexe proposé à partir de la structure RMN de PHD2 seul (PDB: 1mm2) et des résultats de titration, ces résidus ne sont pas directement impliqués dans la fixation de la lysine K9me3

(Musselman et al., 2009). Une hypothèse est que le peptide H3K9me3 pourrait adopter une conformation plus favorable à l'interaction avec le domaine PHD ce qui expliquerait l'augmentation de l'affinité mais sans que la lysine K9me3 ne soit spécifiquement reconnue.



Figure 22: Séquence du PHD2 de CDH4. Les résidus surlignés en rouge sont ceux impliqués dans la coordination des zinc. Les résidus indiqués par des étoiles sont ceux qui présentent des déplacements chimiques en présence des peptides H3, H3K9me et H3K9ac. Les résidus surlignés en bleu sont ceux qui présentent une amplitude de déplacement chimique plus élevée en présence d'un peptide H3K9me3 qu'en présence d'un peptide H3 non modifié.

### 3.2.4. Comparaison de séquences entre les différentes classes de PHD reconnaissant l'histone H3

Un alignement de séquences des domaines PHD appartenant aux trois classes précédemment détaillées est présenté dans la Figure 23. Comparée aux deux autres classes, la classe des domaines PHD interagissant avec l'histone H3K9me3 ne présente pas de résidus conservés spécifiques. Une caractérisation structurale pour déterminer ce mode de reconnaissance s'avère indispensable. Un tableau récapitulatif caractérisant les différents PHD appartenant à ces trois classes est présenté (Tableau 1).

	1	10	20	30	40	50
1 Taf3	GNQIW	ICPGCNKP	DDGSP	MIGCDD	CD.DWYHWP	CVGIMAAPPEEM.QWF
BPTF_PHD2	DTKLY	ICKTPY	DESKF	YIGCDR	CQ.NWYHGR	CVGILQSEAEIIDEYV
ING2	NEPTY	LCNQ	VSYGE	MIGCDNEQ	CPIEWFHFS	CVSLTYKPKGK
2 BHC80	IHEDF	CSVCRK	SGQ	LLMCDT	CS.RVYHLD	CLDPPLKTIPKG.MWICPRCQ
AIRE_PHD1	KNEDE	CAVCRD	GGE	LICCDG	CP.RAFHLA	CLSPPLREIPSG.TWRCSSCL
JADE1_PHD1	DEDVV	CDVQ	SPDGEDGNEMV	FCDK	CN.ICVHQAC	YGILKVPEG.SWLCRTCA
3 CHD4_PHD2	HMEF	CRVCKD	GGE	LLC	CDT	CP.SSYHIH
Lid2_PHD2	RKKKG	CIFC	FRLPESG	VMIE	CEI	CD.EWYHAK
SMCX_PHD1	FIESY	VRM	SRGDEDDK	LLL	CDG	CD.DNYHIF

Figure 23: Comparaison de séquences entre les trois classes de domaines PHD reconnaissant l'histone H3. La première classe (1) reconnaît H3K4me3 via les résidus aromatiques ou hydrophobes surlignés en jaunes. La deuxième classe (2) interagit avec H3K4me0 via les résidus acides surlignés en vert. Les alignements de séquences de la troisième classe (3) ne permettent pas de faire émerger de façon convaincante les résidus conservés qui seraient impliqués dans la fixation de H3K9me3. Le résidu aromatique conservé dans tous les domaines PHD et l'acide aspartique (ou glutamique) impliqué dans la reconnaissance de H3R2 sont surlignés respectivement en bleu et cyan.

L'équipe dirigée par JP. Samama puis par C. Birck avait commencé le projet hUHRF1 par l'étude du domaine SRA. Sa structure tridimensionnelle, sous forme d'apo-protéine, a été résolue en 2008 par B. Delagoutte (Delagoutte et al., 2008) et mon projet de thèse, démarré en octobre 2006, avait pour objectif de caractériser la structure et la fonction du domaine PHD isolé et fusionné avec le domaine SRA.

Domaine PHD	Peptide	Kd	Type d'interaction	Méthode de détermination de structure	PDB	Rôle	Références
AIRE (PHD1)	H3K4me0 (1-10)	4.7 ± 0.8 µM (FT) 6.5 ± 0.2 (ITC)	Interactions électrostatiques avec D297	RMN	1xwh, 2kft, 2ke1	Activation de la transcription des antigènes tissulaires spécifiques	(Bottomley et al., 2005; Chakravarty et al., 2009; Chignola et al., 2009; Org et al., 2008)
BHC80	H3K4me0 (1-10)	33 ± 6.4 µM (ITC)	Interactions électrostatiques avec D489	X-Ray	2puy	Inhibition de la reméthylation de H3K4 après déméthylation par LSD1	(Lan et al., 2007)
JADE1 (PHD1)	H3K4me0 (1-12)	25.7 ± 1.2 µM (FT)	-	-	-	Acétylation des histones par HAT/HBO1 et élongation de la transcription	(Saksouk et al., 2009)
RAG2	H3K4me3 (1-12)	4.15 ± 0.21 µM (FT)	Cage: Y415, M443, et W453	X-Ray	2v83, 2v85, 2v86, 2v87, 2v88, 2v89, 2jwo	Recombinaison de l'ADN	(Liu et al., 2007; Matthews et al., 2007; Ramon-Maiques et al., 2007)
hPygo1	H3K4me2/3 (1-15)	2.5 ± 0.2 µM (ITC)	Cage: W366, Y341, D352 V350 et A356	X-Ray	2vpb, 2vp7, 2vpd, 2vpe, 2vpg	Activation de la transcription via la voie de signalisation Wnt	(Fiedler et al., 2008; Nakamura et al., 2007)
TAF3	H3K4me3 (1-13)	0.31 ± 0.09 µM (FT)	Cage: W868, W891, M882 et D877	RMN	2k16, 2k17	Fixation du facteur de transcription TFIID au niveau des promoteurs actifs	(van Ingen et al., 2008; Vermeulen et al., 2007)
JARID1A (PHD3)	H3K4me3 (1-15)	0.75 ± 0.04 µM (ITC)	Cage: W1625 et W1635	X-Ray et RMN	2kgg, 2kgi, 3gl6	Inhibition de la déméthylation de H3K4me3 par jarid1	(Wang et al., 2009)
BPTF (PHD2)	H3K4me3 (1-15)	2.7 µM (ITC)	Cage: Y23, Y10, Y17 et W32	RMN	2fuu	Remodelage de la chromatine par le complexe NURF	(Li et al., 2006; Wysocka et al., 2006)
Yng1	H3K4me3 (1-9)	9.1 ± 1.6 µM (FT)	Cage: W180, Y157, M168 et S164	RMN	2jmi, 2jmj	Stabilisation du complexe HAT/ NuA3 au niveau de la chromatine	(Martin et al., 2006; Taverna et al., 2006)
ING1	H3K4me3 (1-12)	3.3 ± 1.6 µM (FT)	Cage: W235, Y212, S219 et M223	X-Ray	2qic	Réparation des endommagements d'ADN et apoptose	(Pena et al., 2008)

ING2	H3K4me3 (1-12)	$1.5 \pm 1 \mu\text{M}$ (FT)	Cage: W238, Y215, S222 et M226	X-Ray	2g6q	Répression de la transcription des gènes en prolifération suite aux endommagements d'ADN	(Martin et al., 2006; Pena et al., 2006)
ING4	H3K4me3 (1-10)	$6.4 \pm 0.8 \mu\text{M}$ (ITC)	Cage: Y198, W221, M209 et S205	X-Ray	2pnx, 2vnf	Activation de la transcription via le recrutement du complexe HAT/HBO1	(Palacios et al., 2006; Palacios et al., 2008)
ING5	H3K4me3 (1-12)	$2.4 \pm 1 \mu\text{M}$ (FT)	Cage: W211, S195, Y188 et M199	X-Ray	3c6w	Activation de la transcription via le recrutement des complexes HAT MOZ/MORF et HBO1	(Champagne et al., 2008)
ORC1	H3K4me3	Pull down	-	-	-	Initiation de la réplication et régulation de la transcription	(de la Paz Sanchez and Gutierrez, 2009)
CHD4 (PHD2)	H3K9me3 H3K9ac (1-12)	$0.9 \pm 0.4 \mu\text{M}$ $0.6 \pm 0.1 \mu\text{M}$ (FT)	-	RMN	PHD2 seul 1mm2	Remodelage de la chromatine par le complexe NuRD	(Musselman et al., 2009)
Lid2 (PHD2)	H3K9me3	Pull down	-	-	-	Régulation de la méthylation et de la déméthylation de H3K4 et H3K9	(Li et al., 2008)
SMCX/ JARID1C (PHD1)	H3K9me3	Pull down	-	-	-	Déméthylation de H3K4me3	(Iwase et al., 2007)
UHRF1	H3K9me3/2	Pull down	-	-	-	Régulation du cycle cellulaire et de l'expression de certains gènes suppresseurs de tumeur	(Karagianni et al., 2008)

*Tableau 1: Différents domaines PHD reconnaissant l'histone H3. Pour les différents PHD, la spécificité de reconnaissance, le mode d'interaction, les constantes de dissociation et les résidus impliqués dans la reconnaissance sont mentionnés. La détermination des Kd par calorimétrie (ITC) ou par fluorescence de tryptophane (FT) est précisée. De même, la taille des peptides utilisés est donnée entre parenthèses.*

## Chapitre II

### Matériels et Méthodes

---

- 1. Techniques de la biologie moléculaire**
- 2. Production de la protéine dans *E.coli***
- 3. Lyse cellulaire et purification**
- 4. Caractérisation des protéines et des interactions en solution**
- 5. Cristallogénèse et cristallographie**

# 1. Techniques de la biologie moléculaire

Toute étude structurale d'une protéine débute par une étape de clonage pour introduire l'ADNc codant pour la protéine d'intérêt dans un vecteur d'expression. Ce vecteur doit être compatible avec le système eucaryote ou procaryote qui va être utilisé pour l'expression de cette protéine.

## 1.1. Le clonage Gateway®

### 1.1.1. Principe du clonage Gateway®

La technologie Gateway®, est une technique de clonage basée sur les propriétés de recombinaison du bactériophage  $\lambda$  (Landy, 1989). Cette technique permet un transfert rapide et efficace de séquences d'ADN dans différents vecteurs, permettant l'expression de la protéine d'intérêt en vue de son étude fonctionnelle.

La recombinaison, site spécifique, du bactériophage  $\lambda$  lui permet d'intégrer son génome dans le chromosome d'*E. coli* et permet le passage du cycle lytique<sup>2</sup> au cycle lysogénique<sup>3</sup>. Cette recombinaison conservative se fait grâce à des sites d'attachement spécifiques, *att*, qui se trouvent sur le chromosome d'*E. coli* (*attB*) et sur le bactériophage (*attP*). Dès l'intégration du bactériophage, la recombinaison se fait entre les sites *attB* et *attP* pour donner les sites *attL* et *attR*. L'enjambement implique deux régions homologues de 15 pb au niveau des deux sites d'attachement. Les séquences qui entourent ces régions sont aussi nécessaires car elles contiennent les sites de fixation des protéines de recombinaison.

La recombinaison  $\lambda$  est catalysée par un mélange d'enzymes qui reconnaissent des séquences spécifiques dans les sites d'attachement. Les protéines impliquées dans la recombinaison varient selon le cycle utilisé par le bactériophage, lytique ou lysogénique. Le cycle lysogénique est catalysé par l'intégrase (Int) du bactériophage  $\lambda$  et les protéines du facteur d'intégration d'*E. coli* (IHF). Le cycle lytique est catalysé par l'intégrase et

---

<sup>2</sup> Cycle lytique: Le virus, incapable de se reproduire par ses propres moyens, injecte son matériel génétique dans la bactérie. Grâce aux enzymes et aux ribosomes de l'hôte, le virus peut être répliqué à plus de cent exemplaires avant que l'hôte n'éclate (cycle infectieux)

<sup>3</sup> Cycle lysogénique: Le matériel génétique du bactériophage s'intègre au chromosome de la bactérie qui le transmet à ses descendants (lysogénie). Dans un cas sur cent mille, l'ADN viral est activé et entame un cycle lytique

l'excisionase (Xis) du bactériophage  $\lambda$  et les protéines du facteur d'intégration d' *E.coli*. Ces enzymes constituent respectivement les mélanges BP clonase<sup>®</sup> et LR clonase<sup>®</sup> (Figure 24).

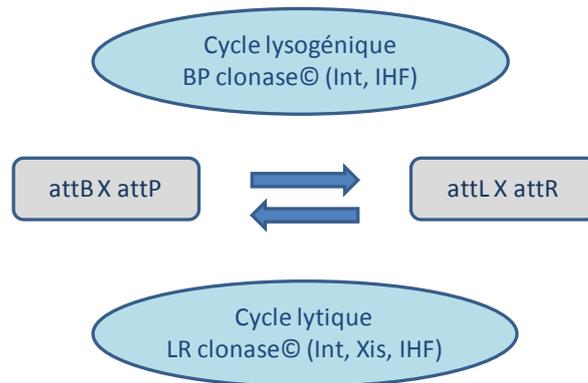


Figure 24: Les réactions de recombinaison et les enzymes nécessaires pour chacun des deux cycles du bactériophage  $\lambda$ .

La technologie Gateway<sup>®</sup> comporte deux réactions:

**La réaction BP:** permet la recombinaison entre un substrat ayant les sites attB (produit PCR ou un clone d'expression linéarisé) et un vecteur donneur ayant les sites attP pour donner un vecteur d'entrée contenant les sites attL. Cette réaction est catalysée par les enzymes BP clonase<sup>®</sup> (Figure 25).

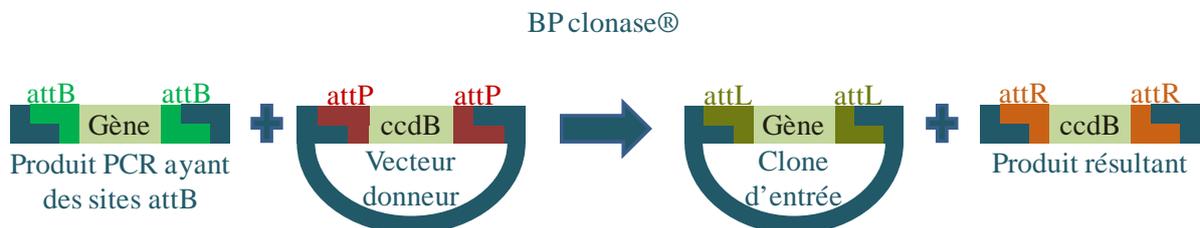


Figure 25: La réaction de recombinaison BP (adapté d'Invitrogen).

**La réaction LR:** permet la recombinaison entre un clone d'entrée ayant les sites attL et un vecteur de destination pour donner un vecteur d'expression contenant des sites attB. Cette réaction est catalysée par les enzymes LR clonase<sup>®</sup> (Figure 26).

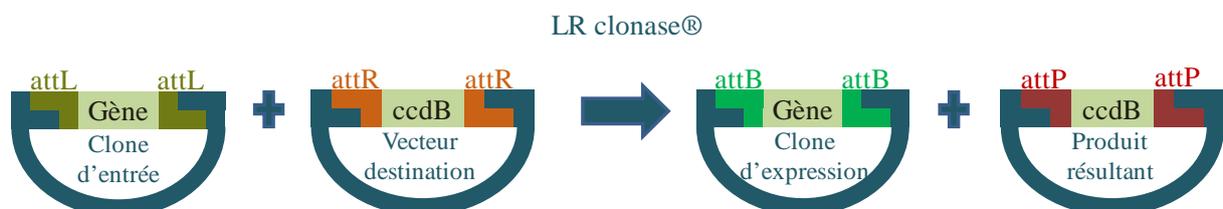


Figure 26: La réaction de recombinaison LR (adapté d'Invitrogen).

La présence du gène *ccdB* permet une sélection négative des vecteurs donneur et destination dans *E.coli* après la recombinaison et la transformation. Les cellules qui reçoivent

un vecteur ayant le gène *ccdB* sont incapables de se multiplier, ce qui permet d'obtenir le clone d'intérêt avec une grande efficacité.

## 1.2. Protocole de clonage

### 1.2.1. Dessin des oligonucléotides

Le clonage Gateway requiert le dessin d'un oligonucléotide « sens » comportant le site attB1 et d'un oligonucléotide « reverse » comportant le site attB2. Deux types d'oligonucléotide sont nécessaires:

- Pour une première amplification: un oligonucléotide court « sens » comportant de 5' vers 3' une séquence « Helper » facilitant la fixation de la protéase sur sa cible, un site de clivage à une protéase, la TEV (Tobacco Ech Viruses) dans notre cas, éventuellement un site de restriction et 18 à 25 nucléotides spécifiques de la cible. L'oligonucléotide « anti-sens » comportera de 5' vers 3' la séquence attB2, un éventuel site de restriction et 18 à 25 nucléotides spécifiques de la cible.
- Pour une deuxième amplification: un oligonucléotide générique « sens » comportant la séquence attB1 suivi de la séquence Helper et du site de clivage TEV. Pour l'oligonucléotide «anti-sens », nous utilisons le même que pour la 1<sup>ère</sup> amplification.

### 1.2.2. Amplification des domaines par la technique de réaction en chaîne par polymérase:

La réaction en chaîne par polymérase ou PCR est une technique de biologie moléculaire permettant d'amplifier *in vitro* une séquence spécifique d'ADN double brin. La PCR est basée sur une répétition de cycles de transition de température, comportant chacun 3 étapes:

- Dénaturation de l'ADN double brin par chauffage à 98°C
- Hybridation des amorces sens et anti-sens sur la matrice d'ADN par diminution de la température à 55°C
- Synthèse du brin complémentaire de l'ADN matriciel à 72°C par les polymérases de l'extrémité 5' vers l'extrémité 3' en utilisant les dNTP libres dans le milieu réactionnel.

L'amplification se fait principalement en utilisant la Phusion ADN polymérase thermostable (2 U/μl) de Finnzymes (Tableau 2):

	Première amplification PCR1	Deuxième amplification PCR2
Matrice	10 ng d'un vecteur d'expression d'une construction plus longue	1 µl du produit de la 1 <sup>ère</sup> PCR
Tampon HF 5x (7.5 mM MgCl <sub>2</sub> )	4 µl	4 µl
Amorce 5' (10 µM)	Amorce spécifique « sens » 0.25 µl	Amorce générique « sens » 0.25 µl
Amorce 3' (10 µM)	Amorce spécifique « anti-sens » 0.25 µl	Amorce spécifique « anti-sens » 0.25 µl
dNTP (2 mM)	2 µl	2 µl
Eau	q.s.p. 19.8 µl	q.s.p. 19.8 µl
Phusion ADN polymérase	0.2 µl	0.2 µl
Volume final	20 µl	20 µl
Protocole de PCR	30 s à 98 °C 20 cycles (7 s à 98 °C ; 20 s à 55 °C ; 15 s/kbp à 72 °C) 7 min à 72 °C	30 s à 98 °C 35 cycles (7 s à 98 °C ; 20 s à 55 °C ; 15 s/kbp. à 72 °C) 7 min à 72 °C

Tableau 2: Protocoles de la première et de la deuxième amplification de l'ADN par PCR.

La réaction PCR est vérifiée en déposant 2 µl de la réaction et 1 µl de bleu de charge sur gel d'agarose 1.3% (masse/volume) contenant 1 µl de Bromure d'ethidium BET (10 mg/ml) pour 100 ml de tampon TAE 1x (Tris-Acétate de sodium-EDTA). Les bandes sont révélées grâce à l'agent intercalant BET, qui devient fluorescent lorsqu'il est exposé à des rayons UV avec une couleur rouge-orangée, 20 fois plus intense lorsqu'il est lié à l'ADN.

Dans les cas où l'amplification avec la Phusion ADN polymérase n'a pas fonctionné, d'autres polymérases ont été testées: Long Expand ADN polymérase (3,5 µl) de Roche Diagnostics et La Herculase II ADN polymérase (0.02 µl) de Stratagene. L'ajout de 2% final de DMSO (Diméthyle Sulfoxyde) était bénéfique dans certains cas pour inhiber les hybridations non spécifiques et déstabiliser les structures secondaires de la matrice ou des oligonucléotides.

### 1.2.3. Purification des produits PCR

Après la deuxième amplification et avant de réaliser la réaction BP, une étape de purification des produits PCR2 s'avère indispensable pour éliminer les oligonucléotides attB et les éventuels dimères de ces oligonucléotides. Les volumes des réactions PCR2 sont complétés à 100 µl avec l'eau MQ et la purification se fait en utilisant le kit NucleoFast® 96 PCR Clean-up de Macherey-Nagel. Les produits PCR2 sont déposés sur des membranes d'ultrafiltration dans une plaque 96 puits. Un vide est appliqué pendant 10 à 15 min. L'ADN

est retenu par la membrane alors que les contaminants passent à travers et sont éliminés. L'ADN est ensuite repris avec 50 µl d'eau MQ. Les produits PCR purifiés sont quantifiés en utilisant le spectrophotomètre ND-1000 NanoDrop® (NanoDrop) en sachant qu'à 260 nm, une unité de densité optique (DO) correspond à une concentration de 50 ng/µl d'ADN double brin.

#### **1.2.4. La réaction BP**

Cette réaction se fait entre le produit PCR-attB purifié et un vecteur donneur pDONR207 d'Invitrogen résistant à la gentamycine, pour donner un clone d'entrée pENTR. La réaction se fait dans un volume final de 5 µl:

30 fmol du produit PCR

100 ng du vecteur donneur pDONR207

1 µl du tampon BP 5x

q.s.p. 4 µl avec l'eau MQ

1 µl du mélange BP clonase

Le mélange est incubé toute la nuit à température ambiante pour favoriser la recombinaison. Le lendemain la réaction est arrêtée par l'ajout de 1,2 µg de la Protéinase K qui va dégrader la BP clonase après incubation à 37 °C pendant 10 min.

#### **1.2.5. Transformation**

A chacune des réactions BP, 50 µl de bactéries DH5α Inoue chimio -compétentes sont ajoutés stérilement. La souche DH5 d' *Escherichia coli* est utilisée durant les étapes de clonages car elle est mutée de manière à ne plus exprimer certaines endonucléases intracellulaires qui dégradent les plasmides d'ADN. Le mélange est incubé sur glace pendant 30 min et ensuite les bactéries sont transformées par choc thermique à 42°C pendant 1 min. 150 µl de milieu LB sont ajoutés et les bactéries sont incubées pendant 1 h à 37°C. Ensuite toute la réaction de transformation est étalée sur boîte de Pétri standard contenant du LB agar additionné de Gentamycine (7 µg/ml final). La boîte est incubée toute la nuit à 37°C.

#### **1.2.6. Vérification de la recombinaison BP par PCR**

Des colonies isolées des différentes transformations sont reprises chacune dans 100 µl LB 2x additionné de gentamycine (7 µg/ml final) et sont incubées 1h30 à 37°C. Ensuite ces précultures sont utilisées comme matrice pour vérifier la présence du gène d'intérêt dans les pENTR par PCR en utilisant la Taq ADN polymérase NEB et 5 µl/ deux

oligonucléotides: sens GWS (5'GTACAAAAAAGCAGGCTTC3') et anti-sens GWR (5'GTACAAGAAAGCTGGGTC 3') (Tableau 3).

Préculture	Tampon Thermo Pol 10x	GWS (10 µM)	GWR (10 µM)	dNTP (2 mM)	Eau	Taq ADN polymérase	Volume final	Protocole
1.5 µl	2 µl	Amorce sens 1 µl	Amorce anti-sens 1 µl	2 µl	q.s.p. 19.5 µl	0.5 µl	20 µl	2 min à 94 °C 25 cycles (30 s à 94°C ; 30 s à 55 °C ; 1 min/kbp à 72 °C) 5 min à 72 °C

Tableau 3: Protocole de l'amplification de l'ADN par PCR pour vérifier la réaction BP.

10 µl de chaque réaction PCR sont déposés en présence de bleu de charge sur gel d'agarose. Si l'insert est présent à la bonne taille, la miniculture est utilisée pour une minipréparation d'ADN.

### 1.2.7. Minipréparation d'ADN

20 µl de chacune des cultures précédentes sont utilisés pour inoculer 2 ml LB 2x additionnés de Gentamycine qui seront incubés pendant la nuit à 37°C.

Pour les minipréparations d'ADN, le kit NucleoSpin® Plasmid de Macherey-Nagel ou bien le kit NucleoSpin® Robot 96 Plasmid peuvent être utilisés. Les précultures sont centrifugées à 2500 rpm pendant 10 min à 4°C pour récolter les cellules. Le surnageant est éliminé et le culot bactérien est resuspendu dans 250 µl de tampon A1 (50 mM Glucose ; 25 mM Tris-HCl pH 8 ; 10 mM EDTA ; 100 µg/ml RNase A). Les cellules sont ensuite lysées par l'addition de 250 µl de tampon A2 (200 mM NaOH ; 1 % SDS). Cette suspension est homogénéisée en retournant le tube plusieurs fois, puis incubée à température ambiante durant 5 min. 300 µl du tampon A3 (3M acétate de potassium ; 3% (v/v) acide formique) sont ensuite ajoutés. Après l'homogénéisation, le mélange est centrifugé 10 min à 12000 rpm et le surnageant est transféré sur une colonne NucleoSpin® Plasmid, où l'ADN se fixe sur la membrane en silice. La colonne est centrifugée 1 min à 11000 rpm. Le filtrat est éliminé. La membrane est lavée une première fois par 500 µl de tampon AW (guanidine-HCl ; isopropanol). Après centrifugation à 11000 rpm pendant 1 min, un deuxième lavage se fait avec 600 µl de tampon A4 (composition non fournie) contenant de l'éthanol. Deux centrifugations permettent ensuite de sécher correctement la membrane. Enfin, l'ADN plasmidique est incubé 2 min à température ambiante avec 50 µl du tampon d'éluion AE (5 mM Tris-HCl pH 8.5) et est élué par centrifugation 2 min à 11000 rpm. La concentration

du plasmide est mesurée par spectrométrie UV à 260 nm. La qualité du plasmide est également vérifiée par PCR avec la Taq Polymérase et les oligonucléotides GWS et GWR et ensuite par dépôt sur gel d'agarose.

#### **1.2.8. Séquençage des pENTR**

Les bornes des produits PCR sous-clonés, ainsi que la séquence du gène d'intérêt, sont vérifiées par séquençage en utilisant les oligonucléotides A1 (sens: 5'-TCGCGTTAACGCTAGCATGGATCTC 3') et A2 (anti-sens: 5'-GTAACATCAGAGATTTTGAGACAC-3'). 30 µl de l'ADN plasmidique à 75 ng/µl et 10 µl de chacun des 2 oligonucléotides à 2 pmol/µl sont nécessaires pour chaque séquençage. Cette étape est assurée par le service commun de séquençage de l'IGBMC.

#### **1.2.9. La réaction LR et la transformation**

Après la vérification des séquences des pENTR, la deuxième réaction de recombinaison peut être faite. Cette réaction se fait entre un pENTR et un vecteur pDEST pour donner un vecteur d'expression. Les deux vecteurs destination utilisés dans notre cas pour l'expression dans *E.coli* sont le pHGWA (fusion amino-terminale (His)<sub>6</sub> et résistance à l'ampicilline) et pHGGWA (fusion amino-terminale (His)<sub>6</sub>-GST et résistance à l'ampicilline). Ces vecteurs ont été développés par la plateforme de biologie et génomique structurales (IGBMC: <http://lbgs.u-strasbg.fr/sbgp>) en se basant sur le vecteur pET22b (Merck-Novagen). La réaction se fait dans un volume final de 5 µl:

- 1 µl pDEST (100 ng/µl)
- 1 µl pENTR (100 ng/µl)
- 1 µl du tampon LR 5x
- 1 µl TE pH8 (10 mM Tris-HCl, 1 mM EDTA)
- 1 µl du mélange LR clonase

Le mélange est incubé 1h à température ambiante et ensuite la réaction est arrêtée par l'ajout de 1.2 µg de la Protéinase K qui va dégrader la LR clonase après incubation à 37 °C pendant 10 min. 50 µl de bactéries DH5α Inoue sont ensuite transformées avec 1µl de cette réaction et étalées sur une boîte de Pétri contenant du LB agar additionné de l'ampicilline (100 µg/ml final). La boîte est incubée toute la nuit à 37°C.

### 1.2.10. Vérification de la recombinaison LR et minipréparation de l'ADN

Des colonies isolées des différentes transformations sont reprises chacune dans 100µl LB 2x supplémentés d'ampicilline (100µg/ml final) et sont incubés 1h30 à 37°C. Ensuite ces précultures sont utilisées comme matrice pour vérifier, par PCR, la présence du gène d'intérêt et de l'étiquette dans les vecteurs d'expression. L'oligonucléotide sens UP1 (5'-GATGCGTCCGGCGTAG-3') est utilisé dans ce cas (Tableau 4).

Préculture	Tampon Thermo Pol 10x	UP1 (10 µM)	GWR (10 µM)	dNTP (2 mM)	Eau	Taq ADN polymérase	Volume final	Protocole
2 µl	2 µl	Amorce sens 1 µl	Amorce anti-sens 1 µl	2 µl	q.s.p. 19.5 µl	0.5 µl	20 µl	2 min à 94 °C 25 cycles (30 s à 94°C ; 30 s à 55 °C ; 1 min/kbp à 72 °C) 5 min à 72 °C

Tableau 4: Protocole de l'amplification de l'ADN par PCR pour vérifier la réaction LR.

10 µl de chaque réaction PCR sont déposés sur gel d'agarose. Si la taille obtenue sur gel correspond au gène d'intérêt avec son étiquette, la préculture est utilisée pour ensemercer des minicultures pour les minipréparations d'ADN. Une vérification des vecteurs d'expression peut aussi être faite par digestion avec des enzymes de restriction spécifiques.

Les 3 vecteurs d'expression correspondant aux protéines utilisées dans ce travail sont présentés dans la Figure 27.

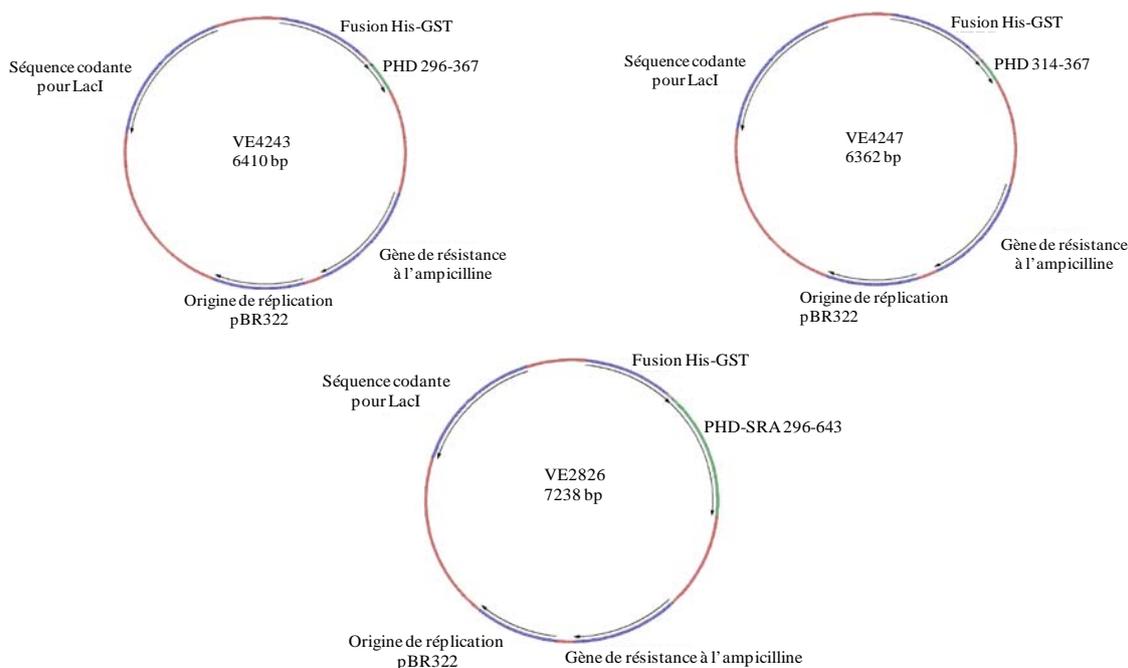


Figure 27: Les vecteurs d'expression correspondant aux protéines étudiées pendant ce travail.

## 2. Production de la protéine dans *E.coli*

*Escherichia coli* est l'organisme de choix pour l'expression des protéines exogènes pour plusieurs raisons: elle est facile à manipuler, elle pousse vite dans des milieux relativement peu cher et permet d'obtenir une grande quantité de protéines. Les différentes protéines utilisées dans ce travail présentent des codons rares: 6 pour PHD<sub>314-367</sub>, 8 pour PHD<sub>296-367</sub> et 35 pour PHD-SRA<sub>296-643</sub>. Pour bien exprimer ces protéines, la souche *E. coli* BL21 Rosetta 2 a été utilisée. Cette souche contient le prophage  $\lambda$ DE3 dont l'un des gènes code l'ARN polymérase du bactériophage T7 sous le contrôle du promoteur lacUV5 qui est inductible par l'IPTG (IsoPropyl  $\beta$ -D-1-ThioGalactopyranoside). Les vecteurs d'expression utilisés sont des dérivés du vecteur pET (Novagen), dans lesquels le gène cible est sous le contrôle du promoteur du bactériophage T7. Le gène codant pour l'ARN polymérase T7 et le gène cible possèdent un site Lac opérateur pour contrôler la transcription. En absence de l'inducteur, le gène *lacI* produit le répresseur lac qui se fixe sur le site lac opérateur et empêche la transcription du gène codant pour l'ARN polymérase T7. En présence de l'IPTG, le répresseur subit un changement conformationnel et ne sera plus capable de se fixer sur l'opérateur. Dans ce cas, l'ARN polymérase est capable de se fixer sur le site promoteur Lac et donc induire la transcription puis la traduction de l'ARN polymérase T7. Cette dernière se fixe alors sur le promoteur T7 et active la transcription du gène cible. Une autre caractéristique de cette souche bactérienne est l'absence des gènes *ompT* et *lon* qui codent pour des protéases capables de dégrader la protéine surexprimée (Figure 28).

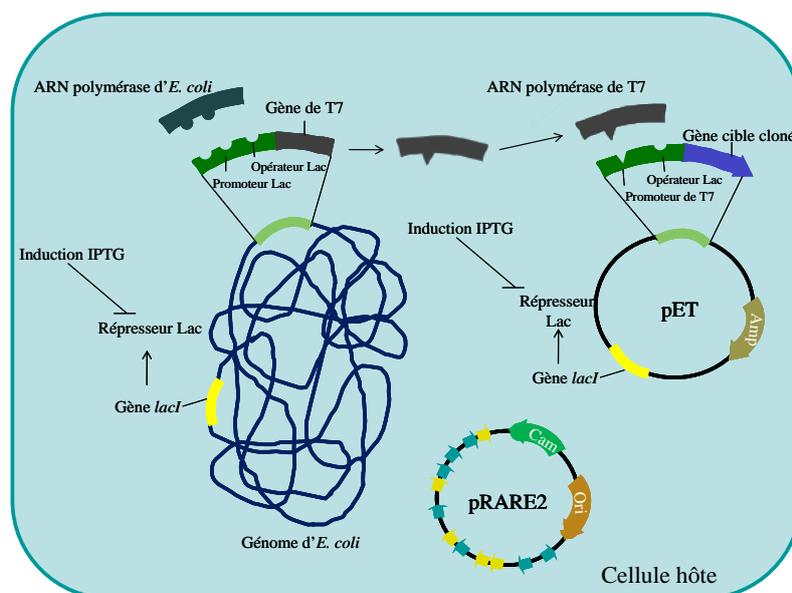


Figure 28: Induction de l'expression du gène cible par l'IPTG dans *E.coli* Rosetta 2.

## 2.1. Préparation des bactéries *E. coli* BL21 DE3 Rosetta 2 électro-compétentes

Le code génétique est dégénéré, plusieurs codons peuvent exister pour un même acide aminé. Il existe, par exemple, six codons différents pour l'acide aminé leucine. La reconnaissance d'un codon donné sur l'ARN messager (ARNm) et l'ajout de l'acide aminé correspondant à la chaîne polypeptidique sont médiés par les ARN de transfert (ARNt). Les concentrations des différents ARNt ne sont pas identiques chez toutes les espèces. Certains ARNt très utilisés chez les mammifères sont rares chez *E.coli*. La présence de ces codons rares lors de l'expression d'une protéine eucaryote dans *E.coli* peut ralentir la cinétique de la traduction, affecter la stabilité de l'ARN messager, conduire à un remplacement d'un acide aminé par un autre, décaler le cadre de lecture ou provoquer une terminaison précoce de la traduction. La souche BL21-DE3-Rosetta 2 d'*E.coli* possède un plasmide pRARE2, résistant au chloramphénicol et exprimant les ARNt rares (AUA, AGG, AGA, CUA, CCC, GGA et CGG), et permet donc d'améliorer l'expression des protéines ayant ces codons. Durant ce travail, cette souche bactérienne a été utilisée pour exprimer les différentes protéines et donc un stock de ces bactéries électro-compétentes a été préparé.

Toutes les étapes de cette préparation sont réalisées dans des conditions stériles et les solutions et les matériels utilisés ont été préalablement autoclavés. 50 µl des bactéries *E. coli* BL21 DE3 Rosetta 2 ont été étalés sur une boîte de pétri LB agar supplémenté en chloramphénicol (35 µg/ml final) et incubés pendant la nuit à 37°C. Le tapis bactérien obtenu est repris dans 3 ml de LB. Cette préculture est utilisée pour inoculer une fiole de 2 l contenant 500 ml de LB de manière à avoir une DO initiale entre 0.050 et 0.100 à 600 nm. Après inoculation, la culture est incubée à 37°C sous agitation jusqu'à une DO<sub>600nm</sub> finale entre 0.5 et 0.8. La culture est ensuite refroidie sur la glace pour arrêter la croissance, répartie dans deux pots froids et stériles de 250 ml et centrifugée à 3000 rpm à 4°C. Le surnageant est alors jeté et les bactéries sont ensuite lavées plusieurs fois, afin d'éliminer toute trace de sel et les rendre plus efficaces à une transformation par électroporation. Chacun des deux culots est repris avec 10 ml d'eau MQ stérile et froide. Les deux suspensions sont mises dans le même pot et le volume est complété à 250 ml avec l'eau MQ stérile et froide. Après centrifugation pendant 15 min à 3000 rpm à 4°C, le culot est repris avec 15 ml du glycérol 10% stérile et froid. Après une dernière centrifugation dans les mêmes conditions qu'avant, le culot est repris avec 2.5 ml de glycérol 10%. Après homogénéisation, les bactéries sont fractionnées par

50 µl dans des tubes eppendorfs stériles, congelées dans de l'azote liquide et conservées à - 80°C.

## 2.2. Transformation par électroporation

Les impulsions électriques entraînent une perméabilité réversible de la membrane bactérienne, permettant ainsi la pénétration du vecteur d'expression portant le gène d'intérêt à l'intérieur de la bactérie. 20 ng du vecteur d'expression résistant à l'ampicilline est mélangé avec 50 µl de bactéries *E. coli* BL21 DE3 Rosetta 2 résistantes au chloramphénicol. Ce mélange est ensuite placé dans une cuve d'électroporation de 0.1 cm de largeur. L'électroporation se fait à 2500 V dans l'Electroporator 2510 (Eppendorf). Les bactéries sont ensuite reprises de la cuve par 500 µl LB et incubées 1h à 37°C. Elles sont ensuite étalées sur des boîtes de pétri LB agar additionnées de chloramphénicol (35 µg/ml) et d'ampicilline (100 µg/ml). Les boîtes sont incubées à 37°C toute la nuit.

## 2.3. Préculture et culture

Chaque préculture de 100 ml supplémentée d'ampicilline et de chloramphénicol est directement inoculée avec une colonie isolée provenant de la transformation et est incubée sous agitation toute la nuit à 37°C.

Les précultures sont utilisées pour ensemercer des milieux LB supplémentés d'antibiotiques. Les cultures sont réalisées dans des erlens de 5 l contenant 1 l de milieu de culture. Le volume de préculture à ajouter à 1l de LB est calculé de façon à avoir une DO initiale de 0.1 à 600 nm. Les cultures sont incubées à 37°C sous agitation jusqu'à une DO<sub>600nm</sub> comprise entre 0.3 et 0.4. La température est ensuite baissée à 20°C. A une DO d'environ 0.5, l'expression des protéines recombinantes est induite par addition de 0.5 mM d'IPTG. Au moment de l'induction, 100 µM de ZnCl<sub>2</sub> sont également ajoutés aux cultures pour stabiliser le domaine PHD présent dans les protéines à exprimer.

Après 5h d'induction à 20°C, les cultures sont centrifugées pendant 20 min à 4000 rpm pour éliminer le LB. Les culots sont ensuite resuspendus dans 20 ml du tampon (30 mM Tris pH 7, 100 mM NaCl) permettant de laver les cellules et éliminer toute trace restante de LB. Une deuxième centrifugation de 20 min à 4000 rpm permet de récupérer les bactéries qui seront ensuite conservées à - 80°C.

### **3. Lyse cellulaire et purification**

Après l'expression en grande quantité, les protéines doivent être récupérées des cellules bactériennes par lyse cellulaire, et ensuite purifiées pour réaliser l'étude structurale.

#### **3.1. La lyse cellulaire**

Les conditions de lyse varient en fonction des protéines étudiées. Chaque protéine nécessite une optimisation des conditions de lyse afin d'obtenir la plus grande quantité possible de protéine soluble. Les conditions de lyse comprennent la technique de lyse et la composition du tampon utilisé (pH, concentration saline, glycérol, détergent, additifs). Dans notre protocole de lyse, chaque culot bactérien, correspondant à 1l de culture, est repris avec 20 ml de tampon de lyse auquel sont ajoutés le cocktail d'inhibiteur de protéases PIC (1 comprimé par 50 ml de tampon de lyse ; Roche) et la benzonase (10 U/ml de tampon de lyse ; Novagen), une nucléase qui dégrade les ADN et les ARN bactériens. Pour la sonication, nous utilisons le sonicateur à ultrasons Labsonic® P équipé d'une sonde 14 mm (Sartorius). Le lysat est maintenu dans la glace lors de la sonication. Sur le sonicateur, l'amplitude est réglée à 30% et le cycle à 0.5, ce qui correspond à une impulsion d'ultrasons de 0.5 s et une latence de 0.5 s par seconde. Pour chaque 20 ml de lysat, le temps de sonication était de 45 s. Après sonication, un aliquot de l'extrait brut (EB) permettant d'estimer la quantité totale de protéine exprimée dans les cellules est prélevé. Le lysat est centrifugé à 14000 rpm pendant 1h à 4°C pour séparer les protéines solubles de la fraction insoluble contenant les débris bactériens et les protéines insolubles. Un aliquot de l'extrait soluble (ES) est prélevé à cette étape. Plusieurs protéines d'E.coli se retrouvent dans l'extrait soluble avec la protéine d'intérêt, d'où la nécessité de mettre au point un protocole de purification de la protéine cible.

#### **3.2. La purification**

La purification de protéines est une succession d'étapes permettant d'isoler une protéine d'intérêt à partir d'un mélange complexe, l'extrait cellulaire. C'est un moyen d'obtenir une protéine pure afin de pouvoir par exemple caractériser sa structure, sa fonction et ses interactions *in vitro*. Pour isoler une protéine donnée, un protocole de purification doit être établi en fonction de la présence d'une étiquette d'affinité et des caractéristiques physico-chimiques de la protéine (masse, forme, charge à un pH donné, etc.). Différentes techniques chromatographiques ont été utilisées afin d'obtenir une protéine pure et homogène pour l'étude structurale: la chromatographie d'affinité, la chromatographie d'exclusion et la

chromatographie d'échange d'ions. Les étapes de purification par chromatographie ont été réalisées sur deux systèmes chromatographiques: le Biocad 700E (Perseptive Biosystems) et l'Äkta™ purifier (GE Healthcare).

### **3.2.1. La chromatographie d'affinité**

Cette technique permet de séparer différentes protéines se trouvant dans le même mélange en fonction de leur différence d'affinité vis-à-vis d'un ligand fixé sur une résine donnée. L'adsorption de la protéine sur le ligand se fait dans des conditions physico-chimiques favorables à l'interaction et hautement spécifiques. Dans ces conditions, les composants du mélange n'ayant pas d'affinité pour le ligand ne sont pas retenus. La protéine spécifiquement liée est récupérée en modifiant les conditions physico-chimiques de sorte à rompre la liaison protéine-ligand. Le plus souvent une molécule qui a plus d'affinité pour le ligand et qui entre en compétition avec la protéine d'intérêt est ajoutée pour éluer cette dernière.

Toutes les protéines utilisées dans ce travail ont été produites avec une étiquette His-GST. Une chromatographie d'affinité sur une résine de nickel a été utilisée comme première étape de purification. Le principe de cette chromatographie est basé sur l'emploi des groupements de type imido-acétates greffés sur une matrice d'agarose, pour chélater des cations métalliques divalents ( $\text{Ni}^{2+}$ ), afin de séparer des protéines portant une étiquette polyhistidines. Les ions  $\text{Ni}^{2+}$  immobilisés interagissent avec les histidines de l'étiquette et permettent de retenir la protéine sur la colonne. L'imidazole qui va entrer en compétition avec les histidines pour lier les ions  $\text{Ni}^{2+}$  est utilisé pour l'élution. Un gradient d'imidazole permet de séparer les protéines fixées faiblement de la protéine d'intérêt fixée plus spécifiquement et donc élue plus tardivement.

Une colonne HiTrap™ Chelating HP de 5 ml a été utilisée. Cette colonne est fournie dépourvue d'ions métalliques et doit être chargée en ions  $\text{Ni}^{2+}$  avant l'utilisation. Deux étapes de chromatographie d'affinité ont été utilisées dans notre protocole de purification: la première après la lyse pour séparer la protéine d'intérêt ayant son étiquette des autres protéines bactériennes et la deuxième après la digestion par la protéase TEV pour séparer la protéine d'intérêt de son étiquette et de la TEV, portant également une étiquette polyhistidine.

### **3.2.2. La chromatographie d'exclusion**

Cette technique de purification, plus couramment connue sous le nom de gel filtration, permet de séparer les protéines en fonction de leur taille et leur forme. La résine utilisée pour

la séparation correspond à un gel formé de billes poreuses. La taille des pores est choisie en fonction de la masse moléculaire des protéines à séparer. Les protéines dont le diamètre est supérieur à celui des pores, ne peuvent y entrer et sont éluées rapidement au niveau du volume mort. Les protéines ayant un diamètre inférieur à celui des pores progressent à travers les billes et sont donc retardées et éluées plus tardivement. Les protéines s'éluent alors dans l'ordre inverse de leur taille. Certaines protéines peuvent avoir un comportement anormal sur la colonne de gel filtration surtout celles qui ne sont pas globulaires. En général, une protéine allongée s'élué plus tôt qu'une protéine globulaire de la même taille. Dans certains cas, surtout quand la salinité du tampon utilisé est faible, certaines protéines peuvent interagir avec la résine de la colonne et s'élueront donc plus tardivement.

La chromatographie d'exclusion peut être utilisée comme étape de purification pour séparer une protéine d'intérêt de ses contaminants mais aussi pour séparer les différents états oligomériques d'une protéine.

Différentes colonnes d'exclusion ont été utilisées durant ce travail en fonction de la quantité de protéine à injecter: Superdex™ 75 10/300 GL, Superdex™ 75 HR 16/50 et HiLoad™ Superdex™ 75 26/60 (GE Healthcare). Ces colonnes permettent de séparer des protéines de masses moléculaires comprises entre 3000 et 70000 Da.

### **3.2.3. La chromatographie d'échange d'ions**

Cette technique consiste à séparer des protéines en fonction de leur différence de charge à un pH donné. En chromatographie d'échange d'ions, une phase stationnaire peut contenir soit des groupements fonctionnels anioniques (pour les échanges de cations), soit cationiques (pour les échanges d'anions). Les molécules en solution ayant une charge opposée que la phase stationnaire seront retenues alors que les molécules ayant la même charge sortent dans les fractions non retenues.

Une protéine donnée est chargée négativement à un pH supérieur à son point isoélectrique (pI) et peut donc s'accrocher sur une résine chargée positivement comme l'UNO™ Q. Cette résine comporte des groupements basiques d'ammonium quaternaires ( $-N^+(CH_3)_3$ ). Par contre, à un pH inférieur à son pI, la protéine est chargée positivement et peut s'accrocher sur une résine chargée négativement telle l'UNO™ S. Cette dernière comporte des groupes acides sulfoniques ( $-SO_3^-$ ).

Ainsi en fonction de son pI et du pH utilisé, la protéine d'intérêt sera retenue sur l'une ou l'autre des différentes résines échangeuses d'ions. L'éluion se fait soit par variation du pH du tampon d'éluion soit en augmentant sa force ionique.

Dans notre cas, la chromatographie échangeuse d'ions a été utilisée pour séparer entre deux conformères de la protéine différant par le nombre des ions zinc fixés par chacun. Une colonne UNO™ Q a été utilisée pour la séparation. Un gradient de NaCl a été utilisé pour l'élution des protéines fixées sur la résine. La capacité maximale recommandée pour la colonne utilisée est de 20 mg de protéine, mais afin d'obtenir une bonne séparation, 10 mg au maximum ont été injectés à chaque run.

Une combinaison de plusieurs étapes de purification et une optimisation des protocoles de purification pour chaque protéine ont été nécessaires pour obtenir des protéines pures, stables et homogènes pour pouvoir faire les essais de cristallisation.

### 3.3. Les protocoles utilisés

Toutes les constructions testées présentent une fusion (His)<sub>6</sub>-GST séparée de la séquence protéique par un site de coupure à la TEV (Tobacco Etch Virus). Le protocole de purification de base consiste en deux chromatographies d'affinité séparées par une digestion à la TEV. La TEV est ajoutée au 1/40 de la quantité totale de protéine (masse/masse). Une étape de chromatographie d'exclusion est ensuite réalisée. Des étapes de dialyses ou l'ajout d'une étape d'échangeuse d'ions ont été indispensables dans certains cas pour obtenir une protéine homogène.

#### 3.3.1. Protocole de purification de la protéine PHD-SRA<sub>296-643</sub>

Le nombre des acides aminés, le point isoélectrique (pI), la masse moléculaire et le coefficient d'extinction molaire théorique de la protéine, avec et sans étiquette, sont présentés dans le tableau suivant (Tableau 5):

	Nombre d'acides aminés	Masse moléculaire (Da)	pI	$\epsilon_{280\text{nm}}$ (M <sup>-1</sup> cm <sup>-1</sup> )
Avec étiquette	608	69008.1	7.25	102270
Sans étiquette	349	38937.6	8.42	56470

Tableau 5: Caractéristiques de la protéine PHD-SRA<sub>296-643</sub>.

Un premier protocole a été mis au point puis a été modifié au fur et à mesure de la caractérisation de la protéine. Le dernier protocole optimisé a permis d'obtenir une protéine plus homogène. La comparaison entre ces deux protocoles est présentée dans le tableau suivant (Tableau 6):

		Protocole initial	Protocole optimisé
Tampon de lyse		50 mM Tris pH 8 ; 500 mM NaCl ; 5% glycérol ; 5 mM $\beta$ ME ; 1 mM PMSF ; PIC; Benzonase	50 mM Tris pH 8 ; 500 mM NaCl ; 5% glycérol ; 10 mM imidazole ; 5 mM $\beta$ ME ; 0.5 mM PMSF, PIC ; Benzonase
Première colonne d'affinité HiTrap	Tampon A	50 mM Tris pH 8 ; 500 mM NaCl ; 5% glycérol ; 1 mM PMSF	50 mM Tris pH 8 ; 500 mM NaCl ; 5% glycérol ; 10 mM imidazole ; 0.5 mM PMSF
	Tampon B	50 mM Tris pH 8 ; 500 mM NaCl ; 5% glycérol ; 300 mM Imidazole ; 1 mM PMSF	50 mM Tris pH 8 ; 500 mM NaCl ; 5% glycérol ; 300 mM imidazole ; 0.5 mM PMSF
	Protocole	3 CV Tp A 10 CV gradient de 0 à 100% Tp B 3 CV Tp B	
Dialyse contre		50 mM tris pH 7; 500mM NaCl; 5 mM $\beta$ ME; 1 mM PMSF	50 mM Tris pH 7; 500 mM NaCl; 5 mM $\beta$ ME; 0.5 mM PMSF
Digestion par la TEV		48h à 4°C	
Dialyses		-	50 mM Tris pH 7; 500 mM NaCl; <b>5 mM EDTA</b> ; 5 mM $\beta$ ME; 0.5 mM PMSF
		-	50 mM Tris pH 7; 500 mM NaCl; 5 mM $\beta$ ME; 0.5 mM PMSF
Deuxième colonne d'affinité HiTrap	Tampon A	50 mM Tris pH 7; 500 mM NaCl; 1 mM PMSF	50 mM Tris pH 7; 500 mM NaCl; 0.5 mM PMSF
	Tampon B	50 mM Tris pH 7 ; 500 mM NaCl ; 300 mM imidazole ; 1 mM PMSF	50 mM Tris pH 7; 500 mM NaCl; 300 mM imidazole ; 0.5 mM PMSF
	Protocole	5 CV Tp A 10 CV gradient de 0 à 40% Tp B 3 CV Tp B	
Dialyse		-	50 mM Tris pH 8 ; 500 mM NaCl ; <b>5 mM TCEP</b> ; <b>10 <math>\mu</math>M ZnCl<sub>2</sub></b> ; 0.5 mM PMSF
Colonne de gel filtration		50 mM Tris pH 7 ; 500 mM NaCl ; 1 mM PMSF	50 mM Tris pH 8 ; 500 mM NaCl, 0.2 mM TCEP

Tableau 6: Comparaison ente le protocole initial et le protocole optimisé de la purification de la protéine PHD-SRA<sub>296-643</sub>.

### 3.3.2. Protocole de purification de la protéine PHD<sub>296-367</sub>

Les caractéristiques de la protéine avec et sans son étiquette sont présentées dans le tableau suivant (Tableau 7):

	Nombre d'acides aminés	Masse moléculaire (Da)	pI	$\epsilon_{280\text{nm}}$ ( $\text{M}^{-1} \text{cm}^{-1}$ )
Avec étiquette	332	38322.8	5.92	54050
Sans étiquette	73	8252.3	5.14	8250

Tableau 7: Caractéristiques de la protéine PHD<sub>296-367</sub>.

Une optimisation du protocole de purification a permis d'obtenir une protéine homogène, capable d'interagir avec les partenaires et de cristalliser (Tableau 8).

		Protocole initial	Protocole optimisé
Tampon de lyse		50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol ; 10 mM imidazole ; 5 mM $\beta$ ME ; 0.5 mM PMSF ; PIC ; Benzonase	50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol ; 10 mM imidazole ; 5 mM $\beta$ ME ; 0.5 mM PMSF ; <b>25 <math>\mu</math>M ZnCl<sub>2</sub></b> , PIC ; Benzonase
Première colonne d'affinité HiTrap	Tampon A	50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol ; 10 mM imidazole ; 0.5 mM PMSF	50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol ; 10 mM imidazole ; 0.5 mM PMSF ; <b>25 <math>\mu</math>M ZnCl<sub>2</sub></b>
	Tampon B	50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol ; 300 mM imidazole ; 0.5 mM PMSF	50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol ; 300 mM imidazole ; 0.5 mM PMSF ; <b>25 <math>\mu</math>M ZnCl<sub>2</sub></b>
	Protocole	3 CV Tp A ; 10 CV gradient de 0 à 100% TpB ; 3 CV TpB	
Dialyse		50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol ; 5 mM $\beta$ ME ; 0.5 mM PMSF	50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol ; 5 mM $\beta$ ME ; 0.5 mM PMSF ; <b>25 <math>\mu</math>M ZnCl<sub>2</sub></b>
Digestion par la TEV		24h à 4°C	
Deuxième colonne d'affinité HiTrap	Tampon A	50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol; 0.5 mM PMSF	50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol ; 0.5 mM PMSF ; <b>25 <math>\mu</math>M ZnCl<sub>2</sub></b>
	Tampon B	50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol; 300 mM imidazole ; 0.5 mM PMSF	50 mM Tris pH 7; 150 mM NaCl ; 5% glycérol ; 300 mM imidazole ; 0.5 mM PMSF ; <b>25 <math>\mu</math>M ZnCl<sub>2</sub></b>
	Protocole	5 CV TpA ; 10 CV gradient de 0 à 40% TpB ; 3CV TpB	
Colonne de gel filtration		20 mM Tris pH 7 ; 150 mM NaCl ; 0.5 mM TCEP	20 mM Tris pH 7 ; 150 mM NaCl ; 0.5 mM TCEP; <b>25 <math>\mu</math>M ZnCl<sub>2</sub></b>
Colonne UnoQ	Dilution de la protéine à 30 mM NaCl avec le tampon	-	20 mM Tris pH 7 ; 0.5 mM TCEP; 25 $\mu$ M ZnCl <sub>2</sub>
	Tampon A	-	20 mM Tris pH 7 ; 30 mM NaCl ; 0.5 mM TCEP; 25 $\mu$ M ZnCl <sub>2</sub>
	Tampon B	-	20 mM Tris pH 7 ; 500 mM NaCl ; 0.5 mM TCEP; 25 $\mu$ M ZnCl <sub>2</sub>
	Protocole	-	35 CV gradient de 30 à 70% Tp B 3CV Tp B

Tableau 8: Comparaison ente le protocole initial et le protocole optimisé pour la purification de la protéine PHD<sub>296-367</sub>.

### 3.3.3. Protocole de purification de la protéine PHD<sub>314-367</sub>

La purification de cette protéine s'est avérée plus facile que les deux autres et le premier protocole testé a permis d'obtenir une protéine homogène et capable de cristalliser de façon reproductible. Les caractéristiques de cette protéine avec et sans l'étiquette sont les suivantes (Tableau 9):

	Nombre d'acides aminés	Masse moléculaire (Da)	PI	$\epsilon_{280\text{nm}}$ ( $\text{M}^{-1} \text{cm}^{-1}$ )
Avec étiquette	316	36562.8	5.66	54050
Sans étiquette	57	6492.3	4.45	8250

Tableau 9: Caractéristiques de la protéine PHD<sub>314-367</sub>.

Le protocole de purification de cette protéine est présenté dans le tableau suivant (Tableau 10):

Tampon de lyse		50 mM Tris pH 7; 150 mM NaCl; 5% glycérol; 10 mM imidazole; 5 mM $\beta$ ME; 0.5 mM PMSF; PIC; Benzonase
Première colonne d'affinité HiTrap	Tampon A	50 mM Tris pH 7; 150 mM NaCl; 5% glycérol; 10 mM imidazole; 0.5 mM PMSF
	Tampon B	50 mM Tris pH 7; 150 mM NaCl; 5% glycérol; 300 mM imidazole; 0.5 mM PMSF
	Protocole	3 CV Tp A; 10 CV gradient de 0 à 100% Tp B; 3 CV TpB
Dialyse		50 mM Tris pH 7; 150 mM NaCl; 5% glycérol; 5 mM $\beta$ ME; 0.5 mM PMSF
Digestion		24h à 4°C
Deuxième colonne d'affinité HiTrap	Tampon A	50 mM Tris pH 7; 150 mM NaCl; 5% glycérol; 0.5 mM PMSF
	Tampon B	50 mM Tris pH 7; 150 mM NaCl; 5% glycérol; 300 mM imidazole; 0.5 mM PMSF
	Protocole	5 CV Tp A; 10 CV gradient de 0 à 40% Tp B; 3 CV Tp B
Colonne de gel filtration		20 mM Tris pH 7; 150 mM NaCl; 0.5 mM TCEP

Tableau 10: Protocole de purification de la protéine PHD<sub>314-367</sub>.

### 3.4. Contrôle quantitatif et qualitatif des protéines

Le contrôle de la pureté de la protéine après chaque étape de purification se fait par électrophorèse sur gel SDS-polyacrylamide. Dans certains cas, des électrophorèses en conditions natives ont aussi été faites pour étudier la qualité de l'échantillon ou détecter des interactions avec les partenaires. La quantification des protéines est réalisée par dosage de

Bradford après la première étape de purification et ensuite par spectrophotométrie NanoDrop® quand la protéine est plus pure.

### 3.4.1. Electrophorèse monodimensionnelle en conditions dénaturantes

Cette technique permet de vérifier la pureté de la protéine et de rassembler les fractions contenant la protéine d'intérêt après chaque étape de purification. Une séparation électrophorétique dépend normalement de la masse moléculaire, de la charge et de la forme des protéines. En conditions dénaturantes, la séparation se fait en fonction de la masse moléculaire seulement. En fait, les échantillons sont dénaturés par chauffage à 95°C en présence d'une forte concentration d'un agent réducteur, l $\beta$ ME. Le gel d'électrophorèse et le tampon de migration contiennent du dodécylsulfate de sodium (SDS), un détergent anionique qui se fixe sur les protéines, masquant ainsi leurs charges propres. Les protéines chargées négativement migrent en fonction de leur masse moléculaire de la cathode (-) vers l'anode (+).

#### 3.4.1.1. Composition des gels dénaturants et des solutions utilisés

Les gels d'électrophorèse utilisés sont discontinus: l'échantillon passe tout d'abord à travers d'une première couche de gel, dite de concentration. Ensuite la séparation a lieu dans la deuxième couche de gel ayant une composition différente. Quelque soit le pourcentage du gel utilisé pour la séparation, la composition du gel de concentration est toujours la même mais c'est le pourcentage en acrylamide/bisacrylamide dans le gel de séparation qui varie. Des gels ayant différents pourcentages en acrylamide/bisacrylamide ont été utilisés durant ce travail (Tableau 11):

	Gel de séparation			Gel de concentration
	10%	12%	15%	5%
Acrylamide/bisacrylamide (40%)	3 ml	3.5 ml	4.5 ml	0.66 ml
Eau MQ	6.4 ml	5.8 ml	4.9 ml	3.8 ml
Tampon (1.5 M Tris-HCl pH 8.8)	2.5 ml	2.5 ml	2.5 ml	-
Tampon (0.5 M Tris-HCl pH 6.8)	-	-	-	0.5 ml
SDS 20%	12.5 $\mu$ l	12.5 $\mu$ l	12.5 $\mu$ l	25 $\mu$ l
Persulfate d'ammonium 10%	60 $\mu$ l	60 $\mu$ l	60 $\mu$ l	34 $\mu$ l
TEMED (N,N,N',N'-Tetramethylethylenediamine)	9 $\mu$ l	9 $\mu$ l	9 $\mu$ l	10 $\mu$ l

Tableau 11: Composition des différents gels SDS utilisés durant ce travail.

La migration se fait dans une cuve d'électrophorèse contenant du tampon de migration 1x, pendant 35 min à 250 V et 100 mA.

***Solution de charge (5x):***

312.5 mM Tris-HCl pH6.8  
50 % Glycérol (v/v)  
10 % SDS (v/v)  
0.2 % Bleu de bromophénol (m/v)  
20 %  $\beta$ ME (v/v)

***Tampon de migration 5x (II):***

15 g Tris  
94 g Glycine  
25 ml SDS (20 %)  
Eau MQ

***Solution de coloration:***

0.1 % Bleu de Coomassie R250 (m/v)  
30% éthanol (v/v)  
20 % acide acétique (v/v)  
Eau MQ

***Solution de décoloration:***

30% éthanol (v/v)  
10 % acide acétique (v/v)  
Eau MQ

### **3.4.2. Electrophorèse monodimensionnelle en conditions non dénaturantes**

Dans l'électrophorèse en conditions non dénaturantes, les molécules sont séparées dans leur état natif. La vitesse de migration des protéines dépend dans ce cas de leur charge, leur masse et leur forme. La migration peut ainsi être différente selon la conformation adoptée. Les gels non dénaturants sont alors sensibles à tout processus qui va modifier la charge, la masse ou la conformation d'une protéine comme l'oligomérisation, l'agrégation, les interactions protéine-protéine et protéine-ligand. Durant ce travail, les gels natifs ont été surtout utilisés pour étudier l'interaction entre le domaine PHD et différents peptides d'histones. Les gels utilisés sont à pH 9.4. A ce pH, les peptides d'histones sont chargés positivement et ne peuvent pas entrer dans le gel et les protéines sont chargées négativement et migrent alors de la cathode (-) vers l'anode (+). Si un peptide interagit avec la protéine étudiée, la migration de cette dernière se trouve retardée et une bande correspondante au complexe est observée.

### 3.4.2.1. Composition du gel non dénaturant

Un seul type de gel non dénaturant a été utilisé durant ce travail, c'est un gel pH 9.4 (6%) dont la composition est présentée ci-dessous:

#### ***Gel non dénaturant pH 9.4:***

3050  $\mu\text{l}$  eau MQ  
1250  $\mu\text{l}$  tampon natif pH 9.4 (5x)  
750  $\mu\text{l}$  acrylamide/bisacrylamide (40%)  
25  $\mu\text{l}$  persulfate d'ammonium (10%)  
5  $\mu\text{l}$  TEMED

#### ***Tampon natif pH 9.4 (5x):***

18.15 g Tris-HCl ( $121.1 \text{ g}\cdot\text{mol}^{-1}$ )  
22.13 g CAPS ( $221.3 \text{ g}\cdot\text{mol}^{-1}$ )  
Eau MQ qsp 500 ml

### 3.4.2.2. Préparation des échantillons et migration

Le matériel utilisé ne doit contenir aucune trace de SDS et toutes les solutions doivent être froides pour éviter la dénaturation de la protéine. Juste avant le dépôt sur gel, un volume 1/1 de bleu natif pH 9.4 est ajouté à chaque échantillon. Pour l'étude des interactions, la protéine a été incubée avec les peptides à tester au moins 30 min à 4°C avant le dépôt sur gel. La migration se fait à 4°C dans le tampon de migration natif pH 9.4 (1x), pendant 30 à 40 min à 150 V. La coloration et la décoloration du gel sont les mêmes que pour les gels SDS.

#### ***Solution de charge (2x):***

50% glycérol (v/v)  
25% tampon natif pH 9.4 (1x)  
0.01% bleu de bromophénol  
Eau MQ

## **4. Caractérisation des protéines et des interactions en solution**

Différentes méthodes biophysiques ont été utilisées durant ce travail pour caractériser en solution l'état oligomérique d'une protéine, sa stabilité, son homogénéité, sa masse moléculaire, son enveloppe moléculaire, sa composition en éléments de structures secondaires et ses interactions avec ses partenaires. Le Thermofluor permet de déterminer les conditions optimales pour stabiliser la protéine. La diffusion quasi-élastique de la lumière est utilisée pour vérifier la monodispersité de la protéine en solution. La spectrométrie de masse

permet de vérifier les masses des protéines purifiées et d'étudier leur interaction avec les partenaires. L'ultracentrifugation analytique et la diffusion des rayons X aux petits angles sont deux techniques complémentaires permettant d'avoir des informations sur l'état oligomérique et l'état conformationnel de la protéine en solution. Quant au dichroïsme circulaire, il permet de déterminer la composition en éléments de structures secondaires de la protéine. Les constantes d'association entre la protéine d'intérêt et ses partenaires peuvent être déterminées par la calorimétrie à titration isotherme.

#### 4.1. Thermofluor

La stabilité, l'homogénéité et la solubilité d'une protéine sont fortement liées à sa capacité de cristalliser. Le Thermofluor est une technique de criblage qui permet d'identifier pour une protéine donnée, le tampon de conditionnement le plus favorable à sa stabilité. Elle permet aussi d'identifier les partenaires potentiels de cette protéine. Il s'agit de suivre la dénaturation de la protéine en fonction de la température et de déterminer sa température de fusion  $T_m$ . Une augmentation de la température de fusion indique une augmentation de la stabilité de la protéine étudiée. Les mesures se font en présence d'un fluorophore, le Sypro Orange, qui devient fluorescent quand il se fixe sur les parties hydrophobes de la protéine, qui deviennent accessibles lors de la dénaturation. Au début de l'expérience, la protéine est bien repliée et la fluorescence est faible. En augmentant petit à petit la température, la protéine se déplie, le fluorophore va se fixer et la fluorescence augmente. Le maximum de fluorescence est atteint lorsque la protéine est complètement dénaturée (Figure 29). Ensuite l'intensité de fluorescence diminue et ceci est probablement dû à la précipitation du complexe protéine-fluorophore (Ericsson et al., 2006).

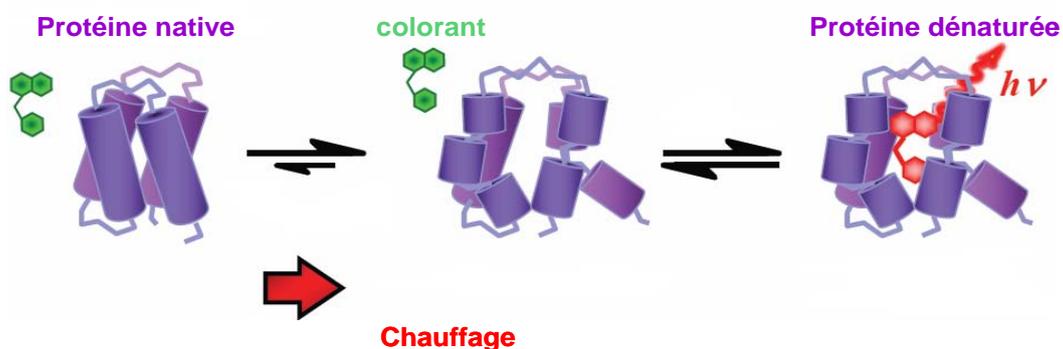


Figure 29: Le principe du Thermofluor consiste à suivre la fixation d'un fluorophore sur les régions hydrophobes de la protéine sous l'effet d'une dénaturation par chauffage.

L'appareil MiniOpticon (Bio-Rad) permet de mesurer la fluorescence de 48 puits en même temps. La plaque de 48 puits a été divisée en deux parties: 24 puits contenant la protéine dans les différentes conditions à tester et les 24 puits restants contenant les blancs correspondants. Les mesures ont été réalisées dans un volume final de 30  $\mu$ l et la composition des puits échantillons et blancs sont les suivants:

Chaque puit échantillon contient:

- 15  $\mu$ l du tampon à tester
- 5  $\mu$ l de la protéine à 1mg/ml, de protéine-ADN, ou de protéine-peptide d'histone
- 3  $\mu$ l du tampon de la protéine
- 7  $\mu$ l du colorant Sypro Orange (1: 200) ajouté à la dernière minute

Chaque puit blanc contient:

- 15  $\mu$ l du tampon à tester
- 8  $\mu$ l du tampon de la protéine
- 7  $\mu$ l de colorant Sypro Orange (1:200)

Le premier protocole de mesure utilisé consistait à lire la fluorescence après chaque augmentation de température de 0.2°C. Comme une différence de  $T_m$  n'est significative qu'au-delà de 2°C, ce protocole a été modifié et la lecture a été réalisée après une augmentation de 1°C. Les deux protocoles utilisés sont présentés ci-dessous:

Premier protocole:

1. Incubation à 20°C pendant 5 min
2. Courbe de fusion de 20°C à 90°C ; lecture tous les 0.2°C, délai de 1 s après la lecture
3. Incubation à 20°C pendant 5 min
4. Fin

Deuxième protocole:

1. Incubation à 20°C pendant 5 min
2. Courbe de fusion de 20°C à 90°C ; lecture tous les 1°C, délai de 2 s après la lecture
3. Incubation à 20°C pendant 5 min
4. Fin

Le programme Opticon Monitor 3 a été utilisé pour définir les paramètres de mesures, démarrer les mesures et analyser les résultats obtenus.

## 4.2. Diffusion quasi-élastique de la lumière (DLS)

Cette technique permet un contrôle qualitatif de la protéine (rayon hydrodynamique, polydispersité, agrégation). Un faisceau de lumière monochromatique produit par un laser est envoyé sur la solution. Une très faible partie de la lumière est diffusée par les particules du liquide. Cette diffusion provient des fluctuations locales de densité et de concentrations de soluté sous l'effet de l'agitation thermique (mouvement Brownien). Un photomultiplicateur permet de mesurer la variation au cours du temps de l'intensité de la lumière diffusée. Le paramètre calculé est le coefficient de diffusion translationnel qui à son tour permet de calculer le rayon hydrodynamique de la particule par la relation de Stokes-Einstein:

$$R_H = \frac{kT}{6\pi\eta D_t}$$

$R_H$ : rayon de Stokes,  $T$ : la température absolue,  $\eta$  : la viscosité du fluide,  $D_t$ : le coefficient de diffusion translationnel et  $k$ : la constante de Boltzmann. La DLS fournit deux types d'analyse: monomodale qui permet de déterminer le pourcentage de polydispersité, et bimodale qui, dans le cas d'un échantillon polydispersé, fournit la distribution qualitative et quantitative des entités présentes dans la solution. Pour qu'une solution de macromolécules soit considérée monodisperse, c'est-à-dire composée d'une seule population, le pourcentage de polydispersité doit être inférieur à 20%. Ce pourcentage décrit la distribution de la taille des molécules présentes dans la solution étudiée.

15  $\mu$ l d'échantillon sont centrifugés à 13000 rpm pendant 15 minutes à 4°C et mis dans une cuve en quartz. Les mesures et l'analyse sont réalisées en utilisant l'appareil Dynapro-MS/X (Protein Solutions) et le logiciel Dynamics v.5.0.1.

## 4.3. La spectrométrie de masse

La spectrométrie de masse a été utilisée pour plusieurs raisons durant ce travail: en conditions dénaturantes, pour vérifier l'identité de la protéine purifiée en comparant les masses moléculaires théorique et expérimentale et en conditions natives pour déterminer le nombre d'ions zincs fixés, l'état d'oligomérisation de la protéine et aussi pour étudier l'interaction avec les partenaires.

### 4.3.1. La théorie

L'ensemble des résultats ont été acquis à l'aide d'un spectromètre de masse muni d'une source d'électronébulisation (Electrospray: ES) et d'un analyseur à temps de vol (Time Of Flight: TOF). Les protéines à analyser sont conditionnées dans un tampon volatile (acétate

d'ammonium). A pression atmosphérique, la solution protéique est transformée en gouttelettes à l'extrémité d'un fin capillaire porté à un potentiel élevé (entre  $\pm 3$  et  $\pm 5$  kV). L'électronébulisation est basée sur un processus électrolytique:



Si le voltage de la source est positif, l'ionisation de l'échantillon se fait par protonation des résidus basiques et si le voltage est négatif, elle se fait par déprotonation des acides carboxyliques. Dans notre cas le voltage était positif et donc les gouttelettes étaient chargées positivement. Sous l'effet d'un courant d'air chauffé, les gouttelettes s'évaporent progressivement. Leur densité de charge devenant trop importante, les gouttelettes explosent en libérant des microgouttelettes constituées de molécules protonées, porteuses d'un nombre de charges variable. Une interface située entre la source et l'analyseur, permet de transmettre, désolvater et focaliser les ions formés afin qu'ils puissent passer dans l'enceinte sous vide de l'analyseur. Les ions vont évoluer en fonction de leur rapport masse/charge ( $m/z$ ) au sein de l'analyseur à temps de vol. Le TOF consiste à mesurer le temps que met un ion, accéléré préalablement par une tension, à parcourir une distance donnée. Le rapport masse sur charge est directement mesurable à partir du temps de vol. Les ions de rapport  $m/z$  le plus petit parviendront au détecteur avant ceux de rapport élevé ce qui permet leur séparation. Un traitement informatique des mesures du rapport masse/charge permet de calculer la masse de la macromolécule qui a généré ces ions (Figure 30).

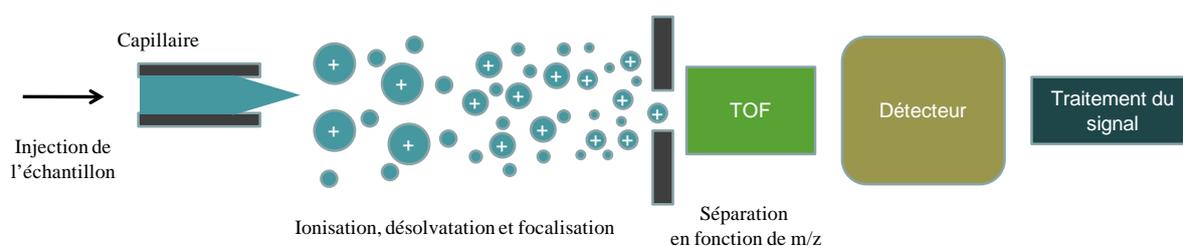


Figure 30: Principe du spectromètre de masse ayant une source d'électronébulisation et un analyseur à temps de vol

#### 4.3.2. Le protocole expérimental

L'analyse en conditions dénaturantes est réalisée en diluant la protéine conditionnée dans 50 mM acétate d'ammonium pH 6.8 dans un solvant eau/Acétonitrile 50/50 (v/v) + 1 % d'acide formique de sorte à obtenir une concentration en protéine de 2 à 5  $\mu\text{M}$ . Dans ces conditions, l'ensemble des interactions non covalentes est dénaturé. Les expériences en conditions non dénaturantes ont été réalisées en acétate d'ammonium. Les solutions

protéiques ont été diluées à l'aide de ce tampon de sorte à obtenir une concentration en protéine entre 5 et 20  $\mu\text{M}$ . En conditions natives, les liaisons ioniques spécifiques sont conservées ce qui permet d'étudier le nombre de zinc fixés par la protéine, l'état d'oligomérisation des macromolécules ainsi que leur interaction potentielle avec des partenaires, tels que l'ADN ou des peptides d'histones. Les mesures de spectrométrie de masse ont été réalisées au Laboratoire de Spectrométrie de Masse Bio-Organique (LSMBO) par Cédric Atmanene et Sarah Sanglier, à l'IGBMC par Hélène Nierengarten, et au CEBGS par Adeline Page et Frank Ruffenach.

#### 4.4. Ultracentrifugation analytique

L'ultracentrifugation analytique permet de déterminer les propriétés des macromolécules biologiques en solution, soumises à une force centrifuge. Elle permet d'étudier en temps réel la sédimentation des macromolécules et de déterminer leur coefficient de sédimentation ( $s$ ), leur masse moléculaire ( $M$ ) et leur coefficient de friction ( $f$ ). Des informations peuvent également être obtenues sur la conformation des macromolécules et la constante d'association des complexes, ainsi que sur l'homogénéité des échantillons. Deux types d'expériences peuvent être réalisés avec cet instrument: des expériences de vitesse de sédimentation et des expériences d'équilibre de sédimentation. Durant ce travail, des expériences de vitesse de sédimentation ont été réalisées sur la protéine PHD-SRA afin de déterminer ses caractéristiques hydrodynamiques et d'étudier son interaction avec l'ADN.

##### 4.4.1. La théorie

Une molécule de masse  $m$ , de volume  $V_p$  et de densité  $\rho_p$ , se déplace dans un milieu de densité  $\rho_m$  sous l'effet d'une force  $F$  résultante de la force centrifuge ( $F_1$ ), de la poussée d'Archimède ( $F_2$ ) et de la force de frottement ( $F_3$ ):

$$F_1 = m \omega^2 r = V_p \rho_p \omega^2 r$$

$$F_2 = \rho_m V_p \omega^2 r$$

$$F_3 = f \, dr/dt$$

$r$  est la distance de l'axe de rotation,  $\omega$  est la vitesse angulaire de rotation,  $f$  est le coefficient de friction qui dépend de la taille et de la forme de la molécule et de la viscosité du milieu et  $dr/dt$  est la vitesse de la molécule. La force est résultante  $F$  est donc égale à:

$$F = F_1 - F_2 - F_3 = V_p (\rho_p - \rho_m) \omega^2 r - f \, dr/dt$$

Sous l'effet de la force  $F$ , la molécule se déplace en s'écartant de l'axe de rotation et va se déposer peu à peu sur le fond de la cellule, c'est la sédimentation. Durant un temps très

court, inférieur à  $10^{-6}$  s, la molécule est considérée en équilibre et la force résultante F est alors nulle:

$$F = V_p (\rho_p - \rho_m) \omega^2 r - f \, dr/dt = 0$$

$$\text{Donc } V_p (\rho_p - \rho_m) \omega^2 r = f \, dr/dt \quad (1)$$

La masse m de la molécule qui est égale à  $V_p \rho_p$  peut être remplacée par  $M/N_A$ , où M est la masse moléculaire de la molécule et  $N_A$  est le nombre d'Avogadro. Et en introduisant le volume partiel spécifique de la molécule  $v_p$  à la place de sa densité ( $v_p = 1/\rho_p$ ), l'équation (1) devient:

$$M (1 - v_p \rho_m) \omega^2 r = N_A f \, dr/dt \quad (2)$$

Le coefficient de friction f est relié au coefficient de diffusion D par la relation de Stokes-Einstein:

$$f = RT/D \quad (3)$$

R est la constante molaire des gaz et T est la température absolue. Le coefficient de sédimentation s, défini par Svedberg, correspond à la vitesse de la particule dans un champ centrifuge égal à l'unité. Ce coefficient est exprimé en unité Svedberg S ( $1 \text{ S} = 10^{-13}$  secondes):

$$s = (dr/dt) / \omega^2 r$$

En considérant les équations (3) et (4), l'équation (2) donne la relation de Svedberg qui permet de calculer la masse moléculaire de la molécule étudiée à partir des grandeurs mesurables:

$$M = \frac{N_A R T}{(1 - v_p \rho_m) D} s \quad (4)$$

Durant une expérience de vitesse de sédimentation, il s'agit de mesurer la vitesse à laquelle une molécule sédimente lorsqu'elle est soumise à une forte vitesse de rotation. Avant la centrifugation, les molécules sont réparties uniformément dans la cellule. Quand la force centrifuge est appliquée, le matériel protéique sédimente en formant un front. Au cours de la centrifugation, le front se déplace vers le fond de la cellule. La mesure de l'avancée de ce front en fonction du temps permet de déterminer le coefficient de sédimentation s. La masse moléculaire de la macromolécule et le coefficient de friction f peuvent alors être déterminés en utilisant la relation de Svedberg. Experimentalement, nous obtenons le coefficient de friction sous forme de  $f/f_0$  correspondant au rapport entre le coefficient de friction expérimentale et le coefficient d'une sphère équivalente. Pour une entité globulaire et symétrique, il doit être entre 1.2 et 1.3.

#### 4.4.2. Le protocole expérimental

Les expériences de vitesse de sédimentation ont été réalisées dans une ultracentrifugation analytique modèle XL-A (Beckman), équipée d'un rotor An 55 class R pour étudier la protéine PHD-SRA<sub>296-643</sub>. Les solutions de protéines ou d'ADN sont placées dans des cellules de centrifugation à compartiments doubles (un compartiment pour la solution de macromolécule et l'autre pour le tampon correspondant qui va servir de référence pour les mesures). Trois cellules différentes peuvent être utilisées dans une même expérience. Pour étudier l'interaction avec de l'ADN, une première cellule doit contenir la protéine seule, une deuxième l'ADN seul et une 3<sup>ème</sup> le mélange protéine-ADN. Le suivi du déplacement des protéines ou de l'ADN au cours de la centrifugation se fait par des mesures d'absorption grâce à un système optique intégré dans la centrifugeuse à 280nm et 260 nm respectivement. Le bras optique va ainsi balayer la cellule de centrifugation à des temps définis et va permettre d'étudier la répartition radiale des particules en solution lors de la centrifugation.

Pour les expériences de vitesse de sédimentation, les cellules sont remplies avec 400  $\mu$ l de l'échantillon ayant une absorbance comprise entre 0.5 et 1, et 425  $\mu$ l du tampon correspondant. Pour étudier l'interaction avec l'ADN, des mesures ont été réalisées sur la protéine seule à 17.7  $\mu$ M, l'ADN méthylé seul (20 pb) à 2.9  $\mu$ M et un mélange protéine (8.8  $\mu$ M) + ADN (1.4  $\mu$ M). La protéine a été conditionnée dans le tampon (25 mM Tris pH8, 50 mM NaCl, 0.2 mM TCEP) et ce même tampon a été utilisé pour diluer l'ADN à partir de la solution stock à 272  $\mu$ M. 200 scans espacés de 3 min ont été enregistrés à 50000 rpm et à 4°C.

Les programmes Sedfit et Sedphat ont été utilisés pour analyser les résultats obtenus. Ces programmes utilisent des solutions numériques pour l'équation de Lamm pour traiter les données expérimentales. L'équation de Lamm décrit la distribution de la concentration des molécules soumises à une ultracentrifugation, en fonction du temps et de la position radiale  $r$ . Le programme sedfit est utilisé pour analyser des systèmes simples de données de vitesse de sédimentation. Par contre, le programme sedphat, permet d'analyser des systèmes plus complexes, par exemple des complexes ternaires et de réaliser une analyse globale de données provenant de différentes expériences biophysiques (vitesse de sédimentation, équilibre de sédimentation, ITC, DLS, SPR ...).

#### 4.5. La diffusion des rayons X aux petits angles

La diffusion des rayons X aux petits angles ou SAXS (pour Small-Angle X-ray Scattering) est une technique qui permet une caractérisation structurale, à basse résolution,

des macromolécules biologiques en solution. Elle permet de déterminer l'enveloppe moléculaire de la macromolécule étudiée et différents paramètres moléculaires tels le rayon de giration ( $R_g$ ), la distance intramoléculaire maximale ( $D_{max}$ ) et la masse moléculaire ( $M$ ). L'utilisation du rayonnement synchrotron, l'amélioration des détecteurs et le développement des logiciels pour l'analyse des données ont fait du SAXS une technique de choix en biologie structurale.

#### 4.5.1. La théorie

Les rayons X sont des rayonnements électromagnétiques caractérisés par un champ électromagnétique sinusoïdal dont la longueur d'onde varie entre 0.01 et 10 nm. L'énergie du rayonnement est quantifiée et son transport se fait sous forme de photons. Lorsque le rayonnement X de vecteur d'onde  $s_0$  ( $|s_0| = 2\pi/\lambda$ ) rencontre un atome de la matière les électrons de ce dernier entrent en vibration et l'atome émet à son tour un rayonnement électromagnétique dans toutes les directions et de même longueur d'onde que le rayonnement incident (il s'agit d'une diffusion élastique). L'onde mesurée selon un angle  $\theta$  possède un vecteur d'onde  $s_1$  de même amplitude que  $s_0$  ( $|s_1| = |s_0| = 2\pi/\lambda$ ). L'amplitude de l'onde diffusée par chaque atome est définie par le facteur de diffusion atomique  $f = N_e r_0$  où  $N_e$  est le nombre d'électrons et  $r_0$  est le rayon de Thomson ( $r_0 = 2.82 \times 10^{-13}$  cm). Le processus de diffusion est décrit par la transformation de Fourier qui lie l'espace réel contenant l'échantillon de coordonnées  $r$  à l'espace réciproque contenant les vecteurs de diffusion  $s = s_1 - s_0$  avec  $|s| = 4\pi \sin(\theta)/\lambda$  (Figure 31).

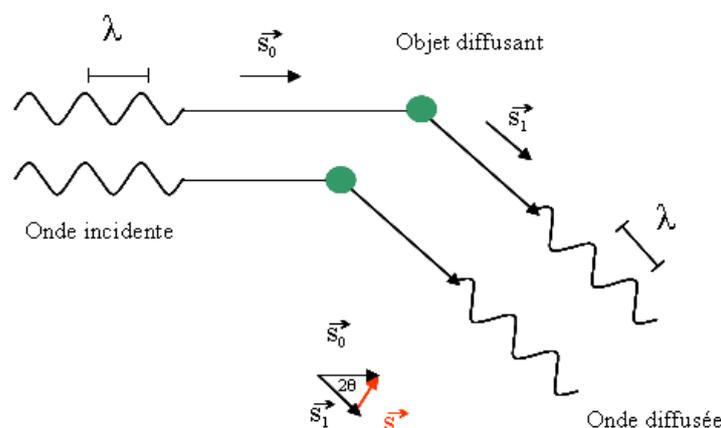


Figure 31: Principe de base de la diffusion élastique des rayons X.  $s_0$  et  $s_1$  sont les vecteurs d'onde incident et diffusé respectivement.  $s$  est le vecteur de diffusion correspondant à un angle de diffusion  $2\theta$ .  $\lambda$  est la longueur d'onde des ondes incidente et diffusée.

En solution, les particules ont des positions et des orientations aléatoires. Pour une particule donnée et à un instant donné, l'intensité diffusée dans la direction définie par le vecteur de diffusion  $s$  est égale à:

$$I(s) = |F(s)|^2$$

Le facteur de structure  $F(s)$  de la particule est lié à sa densité électronique  $\rho(r)$  située à une position  $(r)$  dans une unité de volume  $V_r$ .

L'orientation de la particule varie très rapidement en solution, l'intensité diffusée est alors la moyenne de  $I(s)$  pour toutes les orientations du vecteur  $s$ , ou moyenne sphérique. Dans une solution diluée (solution idéale), les interférences entre les différentes particules sont négligeables et les contributions moyennes de chaque particule à l'intensité diffusée s'ajoutent:

$$I(s) = N \langle |F(s)|^2 \rangle$$

$N$  est le nombre de particules et  $\langle \rangle$  indique la moyenne sphérique. Cette intensité  $I(s)$ , aussi appelée facteur de forme, dépend de la forme et de la structure interne de la particule.

Dans une solution concentrée où les particules interagissent entre elles, l'intensité totale de diffusion des particules  $I_s(s)$  correspond à  $I(s)$ , facteur de forme, modulé d'un terme  $S(s)$  qui correspond à l'interaction entre les particules:

$$I_s(s) = I(s) \times S(s)$$

Dans une expérience de SAXS, il s'agit de placer une solution de macromolécules de concentration  $c$  dans un faisceau fin de rayons X monochromatiques et d'enregistrer le flux des photons diffusés à un angle  $\theta$  du faisceau incident à l'aide d'un détecteur. Le profil de diffusion enregistré et exprimé en fonction du vecteur de diffusion  $s$ , donne des informations sur la structure globale des molécules étudiées en fonction des conditions physico-chimiques. Pour déterminer le rayon de giration, les mesures doivent être réalisées sur une solution monodispersée. Dans ce cas, l'effet d'interaction entre les particules est négligeable et le facteur d'interaction  $S(s)$  est égal à 1. Pour les petites valeurs de l'angle de diffusion  $2\theta$ , l'intensité diffusée  $I(s)$  suit la loi de Guinier qui est valable quelle que soit la forme de la particule pour des valeurs de  $s$  telles que  $1/s$  soit nettement inférieur à ses dimensions:

$$I(s) = I(0) e^{-4/3 \pi^2 R_g^2 s^2}$$

La mesure des intensités diffusées aux petits angles permet de déterminer le rayon de giration  $R_g$  avec une précision satisfaisante. Le rayon de giration est la moyenne quadratique des distances de tous les atomes de la particule à son centre de gravité. Ce rayon dépend de la forme des particules, il est minimum pour une sphère homogène, et il augmente pour des

formes moins symétriques. En portant  $\text{Ln } I(s)$  en fonction de  $s^2$ , la pente de la droite obtenue correspond à  $-4/3 \pi^2 R_g^2 s^2$ , d'où l'on déduit  $R_g$ . Par extrapolation à  $s = 0$ , on obtient  $I(0)$  qui est lié au nombre d'électrons dans la particule, donc à sa masse  $M$ , pourvu que les intensités soient calibrées à une échelle absolue et que la concentration des particules  $c$  soit connue (Figure 32A):

$$I(0) = cM/K$$

$K$  est une constante mesurée par calibration avec une protéine de concentration et de masse moléculaire connues. Après la caractérisation du  $R_g$  et de la masse  $M$  de la macromolécule dans une solution diluée, le comportement de l'ensemble des particules peut être étudié dans une solution concentrée pour obtenir plus d'informations structurales et une meilleure résolution en augmentant le ratio signal/bruit aux grands angles. Mais un inconvénient, dans ce cas, est que le facteur d'interaction  $S(s)$  est différent de 1 et les interactions entre les particules peuvent être attractives ou répulsives (Figure 32B).

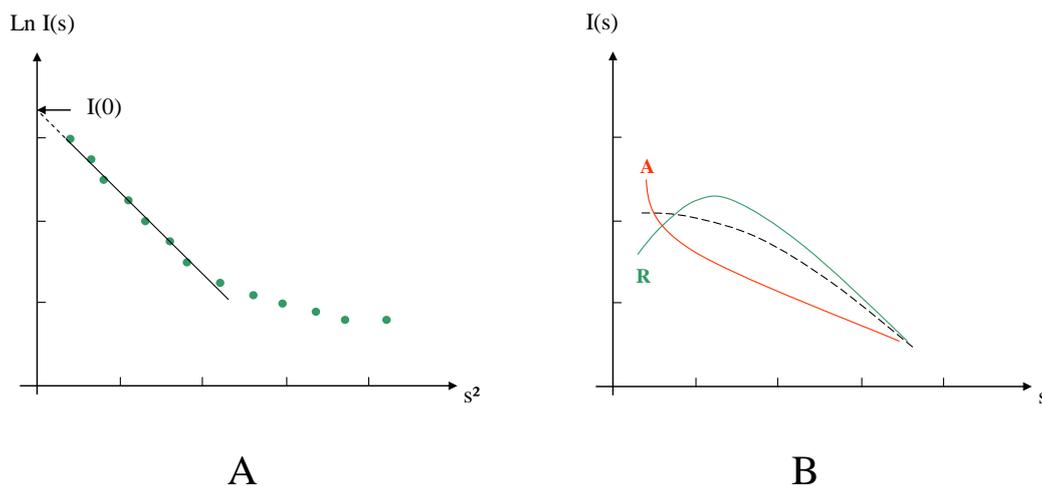


Figure 32: **A-** diagramme de Guinier. Pour les faibles valeurs de  $s$ , le logarithme de l'intensité diffusée est une fonction linéaire de  $s^2$  (Delepierre, 1994). **B-** spectres caractéristiques d'interactions attractives (A), répulsives (R), et d'une solution idéale (en pointillés).

#### 4.5.2. Le protocole expérimental

Les expériences de SAXS ont été réalisées au synchrotron DESY (EMBL, Hambourg) sur la ligne X33. Le dispositif expérimental comprend:

- Une source de rayons X, l'anneau magnétique DORIS
- Un système de collimation et de monochromatisation du faisceau
- Un système de détection et d'acquisition des données, le détecteur 2D Mar345 relié à un ordinateur (Figure 33).

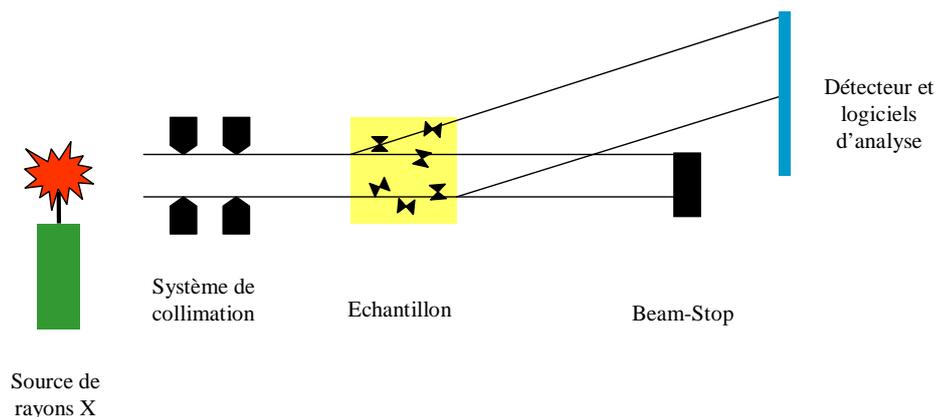


Figure 33: Les composants d'un instrument SAXS.

La longueur d'onde est fixée à  $1.5 \text{ \AA}$ . L'absorption et la diffusion des rayons X par l'air sont limités par un vide primaire sur leur trajet. Quatre paires de fentes en tungstène servent à éliminer les rayonnements parasites provenant principalement du monochromateur et, accessoirement aussi, de la diffusion de l'air résiduel. L'échantillon est placé dans une cellule sous vide en polycarbonate de  $50 \text{ }\mu\text{l}$  de volume. Le remplissage et le rinçage de la cellule est automatisé et sa température peut être contrôlée entre  $5$  et  $50^\circ\text{C}$ . Un problème de l'utilisation du rayonnement synchrotron à flux élevé est l'endommagement des macromolécules causé par les radiations. Ce problème a été réglé par l'ajout de  $2 \text{ mM}$  dithiothréitol (DTT), un agent réducteur qui permet d'éliminer les radicaux libres formés en solution lors de l'irradiation par les rayons X. Les photons diffusés sont mesurés par le détecteur 2D Mar345. La distance standard entre l'échantillon et le détecteur est de  $2.7 \text{ m}$ .

La protéine BSA (Bovine serum albumin) de concentration  $5.3 \text{ mg/ml}$  et de masse moléculaire  $M = 67000 \text{ Da}$  a été utilisée pour déterminer la constante  $K$ . Pour chaque échantillon, la diffusion de la solution de macromolécules et la diffusion du solvant (tampon) sont mesurées successivement et dans les mêmes conditions. La différence entre les deux mesures correspond à la diffusion des macromolécules sans le bruit dû à l'air résiduel et au tampon lui-même. Différents programmes informatiques sont disponibles pour l'analyse des données, la détermination des différents paramètres, la construction ab initio d'un modèle en solution et la comparaison de ce modèle avec des structures tridimensionnelles disponibles. Le programme PRIMUS permet la manipulation des données expérimentales et la détermination, à partir du Guinier, du  $R_g$  et de la masse moléculaire  $M$  (Konarev et al., 2003). Ensuite le programme GNOM permet de déterminer la dimension moléculaire maximale  $D_{\text{max}}$  à partir de la distribution de distances interatomiques dans la particule (Svergun, 1992).

Ce  $D_{\max}$  est utilisé par le programme DAMMIN pour construire un modèle ab initio d'enveloppe moléculaire à basse résolution (Svergun, 1999). Le programme BUNCH permet de proposer un modèle tridimensionnel en accord avec les données de diffusion en utilisant les structures tridimensionnelles connues et en construisant ab initio les régions manquantes (Petoukhov and Svergun, 2005). Deux programmes sont utiles pour la comparaison entre les données à haute et basse résolution: CRY SOL qui est utilisé pour déterminer des courbes de diffusion à partir de fichiers de structures atomiques et les comparer avec les courbes de diffusion expérimentales et SUPCOMB qui permet la superposition entre des enveloppes moléculaires déterminées par DAMMIN et les modèles tridimensionnels proposés par BUNCH (Kozin and Svergun, 2001).

## 4.6. Le dichroïsme circulaire

Le dichroïsme circulaire (CD) est une technique spectroscopique utilisée pour étudier la conformation des acides nucléiques et des protéines. Il s'agit d'enregistrer des spectres dont l'allure permet de déterminer la composition en hélices, brins  $\beta$ , tournants et structures désordonnées. Le traitement des données par des logiciels appropriés permet une estimation de la proportion des différents éléments de structures secondaires.

### 4.6.1. La théorie

La polarisation de toute onde lumineuse peut se décomposer en deux parties, l'une circulaire droite (PCD) et l'autre circulaire gauche (PCG). Si les deux composantes, après passage à travers un échantillon, ne sont pas absorbées ou bien sont absorbées de façon égale, la radiation résultante sera polarisée dans le plan original. Mais si les deux composantes sont absorbées différemment, la résultante possède une polarisation elliptique et la matière présente un dichroïsme circulaire. Seules les molécules chirales, c'est-à-dire celles qui n'ont ni un centre ni un plan de symétrie sont optiquement actives et possèdent un dichroïsme circulaire. Les molécules biologiques sont chirales: les acides aminés des protéines (sauf la glycine), les riboses et les désoxyriboses des acides nucléiques, les lipides, etc., et possèdent des carbones asymétriques. Si on mesure l'absorbance d'un milieu dichroïque pour de la lumière polarisée circulairement successivement droite ( $A_D$ ) et gauche ( $A_G$ ) à une longueur d'onde déterminée, une différence d'absorbance est observée (Figure 34):

$$\Delta A = A_G - A_D$$

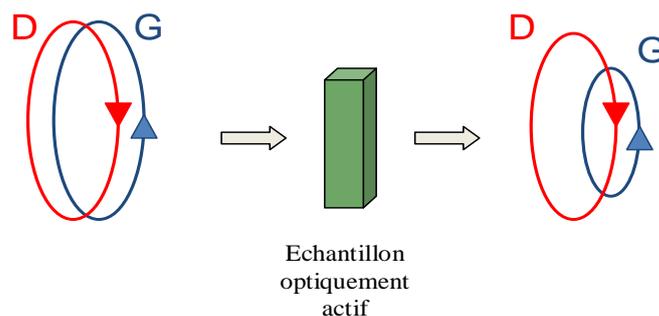


Figure 34: Un échantillon absorbant différemment les lumières polarisées droite et gauche.

Si le trajet optique est  $l$  et la concentration du chromophore est  $C$ , l'égalité devient d'après la loi de Beer-Lambert:

$$\Delta A = (\epsilon_G - \epsilon_D) l C$$

$\epsilon_G$  et  $\epsilon_D$  sont les coefficients d'extinction molaires de la lumière PCG et PCD respectivement, en  $L \cdot mol^{-1} \cdot cm^{-1}$ . Le pouvoir dichroïque molaire est alors défini par:

$$\Delta \epsilon = \epsilon_G - \epsilon_D = \Delta A / l C$$

Cependant les résultats sont souvent reportés en termes de degré d'ellipticité (en degrés) qui est un angle correspondant à la forme de la polarisation de la lumière (Figure 35).

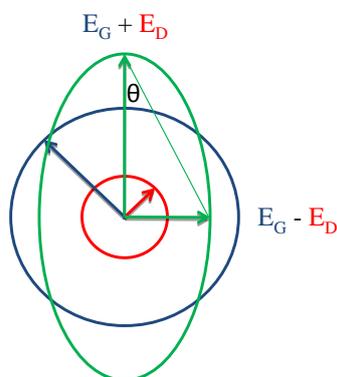


Figure 35: L'onde polarisée elliptiquement (en vert) est due à une contribution inégale de la polarisation circulaire gauche (en bleu) et droite (en rouge).

Il faut noter que:

$$\tan \theta = \frac{E_G - E_D}{E_G + E_D} = \frac{(I_G)^{1/2} - (I_D)^{1/2}}{(I_G)^{1/2} + (I_D)^{1/2}}$$

$E_G$ ,  $E_D$ ,  $I_G$  et  $I_D$  sont les amplitudes des vecteurs de champs électriques et les intensités des polarisations circulaires gauches et droites respectivement. L'intensité  $I$  est proportionnelle au carré de l'amplitude du vecteur de champ électrique  $E$ . L'effet du dichroïsme circulaire est très petit,  $\Delta A \ll 1$  donc  $\tan \theta = \theta$  (radians):

$$\theta \text{ (radians)} = \tan \theta = \frac{1}{4} \Delta A \ln 10$$

Une relation numérique simple existe alors entre la différence d'absorbance et l'ellipticité  $\theta$ :

$$\theta \text{ (degré)} = 180 (\ln 10) \Delta A / 4\pi = 32.98 \Delta A$$

L'ellipticité molaire (degrés. cm<sup>2</sup>. dmol<sup>-1</sup>) est déterminée à partir de l'ellipticité par la formule suivante:

$$[\theta] = \frac{100 \theta}{c l} = 3298 \Delta \epsilon$$

#### 4.6.2. Le protocole expérimental

Les spectres ont été enregistrés à température constante (17°C) en utilisant l'appareil de dichroïsme circulaire Jobin-Yvon CD6. Il existe différents logiciels pour analyser les données de dichroïsme circulaire enregistrées à des longueurs d'onde correspondantes à l'UV lointain (180-260 nm) et estimer la composition en éléments de structures secondaires des protéines. La plupart de ces logiciels utilisent des bases de données contenant des spectres de dichroïsme circulaire provenant de différentes protéines avec différents repliements et dont les structures ont été résolues par cristallographie. Pour éliminer l'effet du tampon, des spectres correspondants aux tampons utilisés ont été enregistrés et soustraits des spectres protéiques. Trois logiciels, inclus dans CDPro ont été utilisés pour l'analyse des spectres: SELCON (Sreerama and Woody, 1993), CDSSTR (Johnson, 1999) et CONTIN (Provencher and Glockner, 1981) et quatre bases de données présentées dans le Tableau 12 (<http://lamar.colostate.edu/~sreeram/CDPro/main.html>).

Numéro de la base	Gamme de $\lambda$ (nm)	Nombre de protéines solubles	Structures secondaires
1	178-260	29	$\alpha$ R, $\alpha$ D, $\beta$ R, $\beta$ D, T, U
3	185-240	37	$\alpha$ R, $\alpha$ D, $\beta$ R, $\beta$ D, T, U
4	190-240	43	$\alpha$ R, $\alpha$ D, $\beta$ R, $\beta$ D, T, U
5	185-240	37	$\alpha$ , $\beta$ , T, P2, U

Tableau 12: Bases de données utilisées pour l'interprétation des spectres.  $\alpha$ R: hélice  $\alpha$  régulière ;  $\alpha$ D: hélice  $\alpha$  déformée ;  $\beta$ R: brin  $\beta$  régulier ;  $\beta$ D: brin  $\beta$  déformé ; T: tournant (turns) ; U: région non ordonnée (unordered), P2: Structure poly(Pro)II.

## 4.7. Calorimétrie à titration isotherme

La calorimétrie par titration isotherme ou ITC (pour Isothermal Titration Calorimetry) permet la caractérisation thermodynamique des interactions entre molécules en solution, et des réactions biochimiques en général. En mesurant la variation de la température due à une interaction, l'ITC permet une détermination directe de la constante d'association ( $K_a$ ), la stoechiométrie ( $n$ ) et l'enthalpie ( $\Delta H$ ). L'énergie libre ( $\Delta G$ ) et l'entropie ( $\Delta S$ ) peuvent ensuite être déterminées à partir de  $K_a$ . L'ITC permet d'étudier les interactions en solution de façon directe et dans des conditions non destructives. Elle ne nécessite pas de marquage des protéines, ni d'immobilisation et il n'y a aucune limite pour les masses moléculaires des macromolécules à analyser. Elle permet de déterminer des constantes d'association allant de  $10^3$  à  $10^8 \text{ M}^{-1}$ . En ayant recours à une méthode de compétition, des constantes d'association plus élevées ( $K_a > 10^{10} \text{ M}^{-1}$ ) peuvent être déterminées.

L'appareil consiste en une enceinte adiabatique contenant deux cellules: la cellule d'échantillon contenant par exemple la macromolécule (A) et la cellule référence contenant le tampon correspondant. Le ligand (B) est ajouté dans la cellule échantillon par injection de petits volumes à l'aide d'une seringue avec une aiguille à ailettes mobile pour mieux homogénéiser le mélange (Figure 36). A chaque ajout de B correspond un échange thermique signant l'interaction A-B, signal proportionnel à la quantité de complexe formé. La puissance nécessaire pour neutraliser cet échange thermique est mesurée par le calorimètre. La chaleur absorbée ou dégagée durant la réaction est déterminée à partir de l'intégrale de la puissance en fonction du temps. Un développement récent de l'ITC consiste à injecter le ligand B en une seule fois dans la cellule échantillon et permet de réduire le temps de l'expérience. Les composants A et B doivent être conditionnés dans le même tampon et leurs concentrations doivent être déterminées avec précision. La concentration de B dans la seringue doit être calculée de manière à obtenir à la fin de la titration une concentration finale du ligand dans la cellule qui est égale à 2-3 fois la concentration de A multipliée par le nombre de ses sites de fixation. La température de l'expérience peut être réglée entre 2 et 80°C. Toutes les solutions utilisées doivent être préalablement dégazées pour éviter les fluctuations de la ligne de base dues à des bulles d'air.

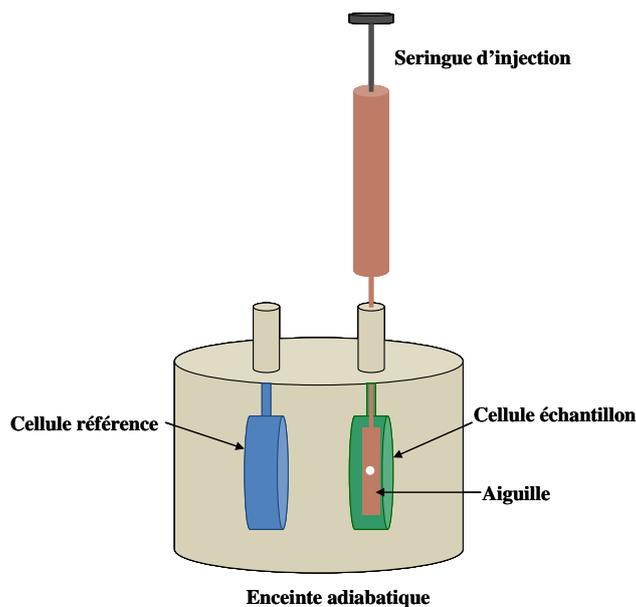


Figure 36: Représentation schématique de l'ITC

La courbe expérimentale obtenue en ITC correspond à la puissance ( $\mu\text{cal/s}$ ), nécessaire pour maintenir une différence de température constante entre les deux cellules, en fonction du temps (Figure 37a). Chaque pic correspond à une injection unique et son intégrale, après soustraction de la ligne de base correspond à la chaleur associée à la fixation ( $q_i$ ). Les pics deviennent progressivement plus petits car il y a moins de composant (A) libre après chaque nouvelle injection. Quand le composant (A) est saturé, le pic obtenu correspond à la ligne de base. Par la suite, la courbe de  $q_i$  en fonction du ratio molaire  $[B]/[A]$  dans la cellule échantillon peut être obtenue (Figure 37b).

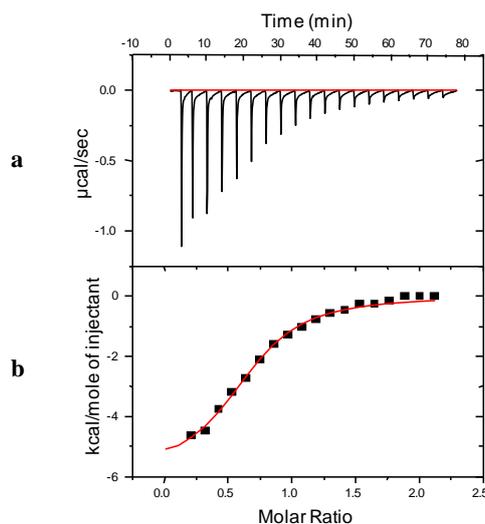


Figure 37: Exemple de courbes obtenues en ITC. a) Courbe expérimentale correspondant à la puissance nécessaire pour neutraliser la chaleur dégagée durant une injection, en fonction du temps. b) La chaleur dégagée suite à une injection en fonction du ratio molaire  $[B]/[A]$ .

### 4.7.1. La théorie

Durant une expérience d'ITC, la quantité de chaleur  $q_j$ , absorbée ou émise pendant une injection  $j$  est égale à:

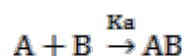
$$q_j = v \Delta H [AB]_f \quad (1)$$

$v$  est le volume de la cellule,  $[AB]_f$  correspond à la concentration du complexe formé pendant l'injection  $j$ . Le volume  $v$  est constant et la variation de l'enthalpie par mole de B fixé,  $\Delta H$ , peut être calculée facilement à partir de la courbe expérimentale. Quand la quantité de B injectée est faible, B interagit complètement avec A et la variation d'enthalpie peut être déterminée à partir de l'équation précédente. Si la constante  $K_a$  est connue, l'énergie libre  $\Delta G$  et l'entropie de la réaction peuvent être déterminées en utilisant l'équation:

$$\Delta G = - R T \ln(K_a) = \Delta H - T \Delta S$$

$R$ , la constante des gaz parfaits, est égale à  $8,31 \text{ J mol}^{-1} \text{ K}^{-1}$  et  $T$  est la température absolue en Kelvin.

Dans le cas le plus simple où la macromolécule possède un seul site de fixation du ligand, la réaction de l'interaction est alors:



$$K_a = \frac{[AB]}{[A][B]}$$

$$[A]_{\text{tot}} = [A] + [AB]$$

$$[B]_{\text{tot}} = [B] + [AB]$$

$[AB]$  est la concentration du complexe dans la cellule à un moment donné.  $[A]_{\text{tot}}$  et  $[B]_{\text{tot}}$  sont les concentrations de la macromolécule dans la cellule et du ligand dans la seringue avant les injections.  $K_a$  est la constante d'association du complexe. D'après ces équations la concentration  $[AB]$  peut être écrite sous les formes suivantes:

$$[AB] = K_a [A] [B]$$

$$[AB] = K_a ([A]_{\text{tot}} - [AB]) [B]$$

$$[AB] = [A]_{\text{tot}} \frac{K_a [B]}{(1 + K_a [B])}$$

La concentration de  $[AB]_f$  formée pendant une injection  $j$  correspond à la différence de concentration entre les injections  $j$  et  $j-1$ :

$$[AB]_f = [AB]_j - [AB]_{j-1}$$

$$[AB]_f = [A]_{\text{tot}} \left( \frac{K_a [B]_j}{1 + K_a [B]_j} - \frac{K_a [B]_{j-1}}{1 + K_a [B]_{j-1}} \right) \quad (2)$$

La combinaison des deux équations (1) et (2) correspond à l'intégrale de la surface du pic obtenu pendant l'injection  $j$ :

$$q_j = v \Delta H [A]_{\text{tot}} \left( \frac{K_a [B]_j}{1 + K_a [B]_j} - \frac{K_a [B]_{j-1}}{1 + K_a [B]_{j-1}} \right)$$

Cette dernière équation est utilisée pour modéliser les données expérimentales et déterminer, si le fit est correcte, les valeurs de  $K_a$  et  $\Delta H$  par la suite tous les autres paramètres thermodynamiques. En utilisant les équations d'interaction appropriées, une étude de réactions à deux sites de fixation est également possible.

Le paramètre critique qui détermine la forme de la courbe isotherme d'interaction est la constante  $c$  qui dépend de la constante d'association  $K_a$ , de la concentration de la macromolécule  $[A]_{\text{tot}}$  au début de l'expérience et de la stœchiométrie:  $c = n K_a [A]_{\text{tot}}$ . La valeur de  $\Delta H$  peut toujours être déterminée de façon précise quel que soit l'aspect de la courbe isotherme mais la détermination expérimentale de  $K_a$  dépend de l'aspect de la courbe et donc de  $c$ . Pour des valeurs de  $c$  très élevées ( $c > 5000$ ), la courbe est « rectangulaire » et ne varie pas en fonction de  $K_a$ . Pour des valeurs de  $c$  très faible ( $c = 0.1$ ), la courbe isotherme devient presque une ligne horizontale, ce qui indique une très faible interaction. La forme d'une courbe isotherme varie en fonction de  $K_a$  pour  $1 < c < 1000$ , ce qui correspond à des affinités allant jusqu'à  $10^8$ - $10^9 \text{ M}^{-1}$ . Les valeurs de  $c$  allant de 5 à 500 sont idéales pour la mesure de  $K_a$  (Figure 38). Dans le cas d'une première expérience sans une idée préalable de la valeur de  $K_a$ , un bon point de départ serait de commencer avec une concentration de  $5 \mu\text{M}$  dans la cellule et  $50 \mu\text{M}$  dans la seringue, pour une stœchiométrie de 1. Avec ces concentrations, des valeurs de  $K_a$  entre  $2 \cdot 10^5$  et  $2 \cdot 10^8 \text{ M}^{-1}$  peuvent être déterminées (Bjelic and Jelesarov, 2008; Pierce et al., 1999; Serdyuk et al., 2007).

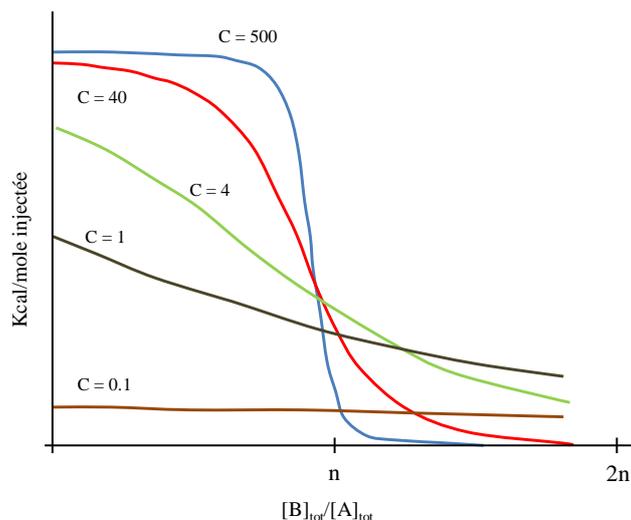


Figure 38: La variation de la courbe isotherme en fonction de la constante  $c$ . Pour déterminer expérimentalement la constante d'association  $K_a$ , il faut idéalement être entre  $5 < c < 500$  (Microcal VP-ITC user's manual).

En fonction de l'appareil utilisé, le volume et la quantité de matériel nécessaire varient. Pour un VP-ITC, le volume actif dans la cellule est de 1.4 ml et la quantité minimale de macromolécule nécessaire est de 50  $\mu\text{g}$ . L'ITC<sub>200</sub> nécessite beaucoup moins de matériel, 5 à 10  $\mu\text{g}$  dans un volume actif de 200  $\mu\text{l}$  sont suffisants.

Durant mon travail de thèse, j'ai utilisé deux appareils d'ITC: le VP-ITC et l'ITC<sub>200</sub> de MicroCal. Par rapport au VP-ITC, l'ITC<sub>200</sub> présente plusieurs avantages: les quantités et les volumes utilisés sont plus faibles, le rinçage des cellules est automatisé et le temps de l'expérience est réduit.

## 4.7.2. Protocole expérimentale

### 4.7.2.1. VP-ITC

Le calorimètre se compose de deux parties essentielles: l'unité calorifique formée de l'enceinte adiabatique, les cellules et la seringue d'injection et de l'ordinateur contenant le programme VPViewer 2000 qui permet de contrôler le calorimètre. La température de l'enceinte doit être réglée avant le début de l'expérience, dans l'onglet **thermostat/calibration**, à une température 5°C en dessous de la température de l'expérience pour éviter une longue durée d'équilibration du calorimètre une fois la méthode lancée.

Les solutions à utiliser doivent être préalablement dégazées 15 à 20 min au Thermovac (860 mBars) à une température 5°C en dessous de la température prévue pour l'expérience. Au début de l'expérience, les deux cellules doivent être rincées avec de l'eau MQ puis du

tampon dégazé. Le volume de chaque cellule est de 1.4 ml mais il vaut mieux prévoir un minimum de 1.8 ml pour la remplir afin d'éviter la formation des bulles d'air. Après le rinçage, les cellules référence et échantillon sont remplies, très lentement, à l'aide d'une seringue avec le tampon et la protéine respectivement. Pendant toutes ces étapes il faut éviter la formation des bulles d'air qui provoquent des fluctuations et empêchent le retour à la ligne de base ce qui rend les données non interprétables. Pour remplir la seringue d'injection, il faut la plonger dans un tube contenant 500 µl du peptide. Dans l'onglet **ITC controls**, il faut cliquer **open fill port** pour pouvoir monter la solution de peptide dans la seringue d'injection par aspiration. Une fois la seringue remplie, il faut cliquer sur **close fill port**. La seringue est ensuite remplie et vidée trois fois en cliquant sur le bouton **purge/refill** pour être sûr de l'absence de bulles d'air. Après avoir enlevé les traces du peptide se trouvant sur les parois de l'aiguille, cette dernière est plongée délicatement dans la cellule échantillon.

Ensuite les paramètres expérimentaux doivent être définis dans l'onglet **ITC controls** \ **Experimental parameters**. A ce niveau il faut définir un certain nombre de paramètres:

- Le nombre d'injections: 30 injections dans notre cas
- La température de la cellule: nous avons travaillé à 20 et 25°C
- La puissance de référence: 17 µcal/s
- Le délai initial avant la 1<sup>ère</sup> injection: 120 s
- Les concentrations dans la cellule et la seringue en mM: 0.6 mM pour les peptides et 0.028 et 0.052 mM pour PHD-SRA<sub>296-643</sub>, 0.065 mM pour PHD<sub>296-367</sub> et 0.072 mM pour PHD<sub>314-367</sub>
- La vitesse de rotation de l'aiguille pour homogénéiser la solution: 310 rpm
- Le nom du fichier des données
- Le mode de feedback: élevé si l'échange thermique dû à la réaction est rapide
- Les options de l'équilibration: rapide et automatique

Concernant les injections, l'onglet **ITC controls** \ **Injection parameters** permet de choisir le volume et la durée de chaque injection et aussi la durée entre les injections. Dans notre cas, les injections sont de 10 µl sauf la 1<sup>ère</sup> qui est de 2 µl. En fait, dans certains cas, la 1<sup>ère</sup> injection d'une série peut montrer un effet calorifique plus faible que prévu et ceci peut être dû à une petite fuite dans la seringue ou à un heurt de l'aiguille contre la paroi lors de son placement dans la cellule. Une 1<sup>ère</sup> injection d'un petit volume permet de surmonter ce problème et le point correspondant ne sera pas utilisé dans l'analyse des données. La période

de filtrage (**filter period**), le temps durant lequel les données sont moyennées pour produire un point sur la courbe de données, doit être définie. Dans notre cas cette période est de 2s.

Après avoir défini tous les paramètres de l'expérience, la méthode peut être démarrée en cliquant sur l'icône **Start** dans le menu principal du programme VPViewer. Un certain temps d'équilibration permet au calorimètre d'atteindre la puissance et la température demandées. Une fois stable, le calorimètre commence les injections. L'analyse des données obtenues a été faite à l'aide du programme Microcal Origin®.

Pour éliminer l'effet de la dilution du peptide et l'effet de la rotation de l'aiguille, des expériences à blanc doivent être réalisées. Il s'agit de remplacer la protéine par son tampon et d'y injecter le peptide en suivant le même protocole. Il faut répéter deux à trois fois chaque expérience pour s'assurer de la reproductibilité des résultats.

#### 4.7.2.2. ITC200

L'utilisation de l'ITC<sub>200</sub> ne diffère pas de celle du VP-ITC. Il s'agit de rincer les cellules tout d'abord avec le tampon et ensuite de les remplir délicatement sans faire de bulles. Le rinçage des cellules et de seringue peut être fait automatiquement à l'aide du programme ITC<sub>200</sub> dans l'onglet **Instrument controls\ Accessory station**. Le volume actif de la cellule est 203 µl mais il vaut mieux prévoir un volume de 300 µl d'échantillon pour la remplir. Le volume de la seringue d'injection est de 38.9 µl mais il vaut mieux préparer 100 à 150 µl pour faciliter le remplissage. De même il faut régler la température de l'enceinte et rentrer les paramètres de l'expérience dans **Instrument controls** et **Advanced experimental design** respectivement. Les paramètres de nos expériences sont les suivants:

- Le nombre d'injections: 20 injections
- La température de la cellule: 20°C
- La puissance de référence: 5 µcal/s
- Le délai initial avant la 1<sup>ère</sup> injection: 60 s
- Les concentrations dans la cellule et la seringue en mM: les peptides 1.5 et 3 mM et la protéine PHD<sub>296-367</sub> 0.119 et 0.154 mM
- La vitesse de rotation de l'aiguille pour homogénéiser la solution: 1000 rpm
- Le mode de feedback: élevé

Concernant les paramètres d'injection, le volume des injections est de 2 µl sauf pour la 1<sup>ère</sup> injection qui est de 0.2 µl. La durée de chaque injection est de 0.4 s et la période de filtrage est de 5 s. Après avoir défini tous les paramètres, la méthode peut être réalisée en

cliquant sur **start**. De même l'analyse des données peut être faite à l'aide du programme Microcal Origin®.

## **5. Cristallogenèse et cristallographie**

La détermination de la structure cristallographique d'une protéine nécessite l'obtention des cristaux bien ordonnés qui diffractent fortement les rayons X. La cristallisation est souvent l'étape limitante dans ce processus. En fait, la formation des cristaux dépend de différents paramètres tels le pH, la température, la concentration protéique, la force ionique, la nature et la concentration du tampon, la nature et la concentration de l'agent précipitant, l'ajout de certains additifs et la présence d'éventuels partenaires. Il s'agit donc de rechercher, en tenant compte de tous ces paramètres, les quelques combinaisons qui peuvent donner des cristaux. Les propriétés de diffraction des cristaux obtenus sont ensuite testées et des données de diffraction sont collectées. Après la détermination des phases expérimentales, un modèle peut être construit par interprétation des cartes de densité électronique. Plusieurs cycles d'affinement sont nécessaires pour obtenir la structure finale de la macromolécule étudiée.

### **5.1. Principe de la cristallisation**

Un cristal est un arrangement périodique ordonné dans l'espace tridimensionnel et la maille cristalline est l'unité de base qui se répète par translations à travers tout le cristal. La cristallisation est un changement d'état qui fait passer la protéine de l'état soluble désordonné à un état solide ordonné. Dans le cristal les protéines sont empilées de façon périodique et régulière dans les trois directions et la cohésion entre les molécules protéiques voisines est assurée par des interactions non covalentes (interactions hydrophobes, liaisons hydrogènes, liaisons de Van der Waals, etc.). Étant volumineuses et globulaires et ayant des surfaces irrégulières, les protéines dans le cristal ne peuvent pas s'empiler très étroitement et des espaces occupés par le solvant séparent alors entre les molécules. Le solvant occupe en moyenne 50 % du volume du cristal donc la protéine dans le cristal possède les mêmes propriétés qu'en solution. Le haut pourcentage du solvant rend les cristaux protéiques très fragiles et leur manipulation délicate.

L'obtention de cristaux nécessite une protéine pure et homogène et peut se décomposer en quatre étapes: la formation d'un état de sursaturation, un état où la concentration de la protéine est supérieure à sa solubilité, la nucléation ou formation des germes cristallins, la croissance cristalline et l'arrêt de la croissance. La technique de cristallisation qui a été utilisée durant ce travail est la diffusion en phase vapeur. Il s'agit de préparer un réservoir à

une concentration donnée en agent précipitant (des sels, des polymères, des solvants organiques ou des composés organiques non volatils). Ensuite la goutte est préparée en ajoutant un faible volume de la solution protéique à un faible volume du réservoir. La goutte et le réservoir sont placés sans être en contact dans une enceinte fermée. L'équilibre s'établit entre la goutte et le réservoir par diffusion des espèces volatiles (l'eau par exemple) jusqu'à ce que la tension de vapeur soit la même dans la goutte et dans le réservoir. La goutte est plus diluée que le réservoir ce qui entraîne une perte d'eau de la goutte vers le réservoir et une augmentation de la concentration de la protéine et de l'agent précipitant dans la goutte. Une fois la saturation est atteinte, une précipitation est observée. Si en précipitant les molécules s'organisent de manière ordonnée, un crystal apparait sinon un précipité amorphe est obtenu. Le comportement de la protéine en fonction de sa concentration et celle de l'agent précipitant peut être décrit sous forme d'un diagramme de phases (Figure 39). La stratégie idéale pour obtenir des grands cristaux serait de commencer dans des conditions correspondant à la zone de nucléation (la zone orange) pour obtenir quelques germes cristallins, et ensuite passer à la zone métastable (en jaune) où la croissance peut avoir lieu. Mais ces conditions optimales sont difficiles à prédire en raison du trop grand nombre de variables dont elles dépendent. Le criblage est donc nécessaire pour déterminer des conditions initiales de cristallisation.

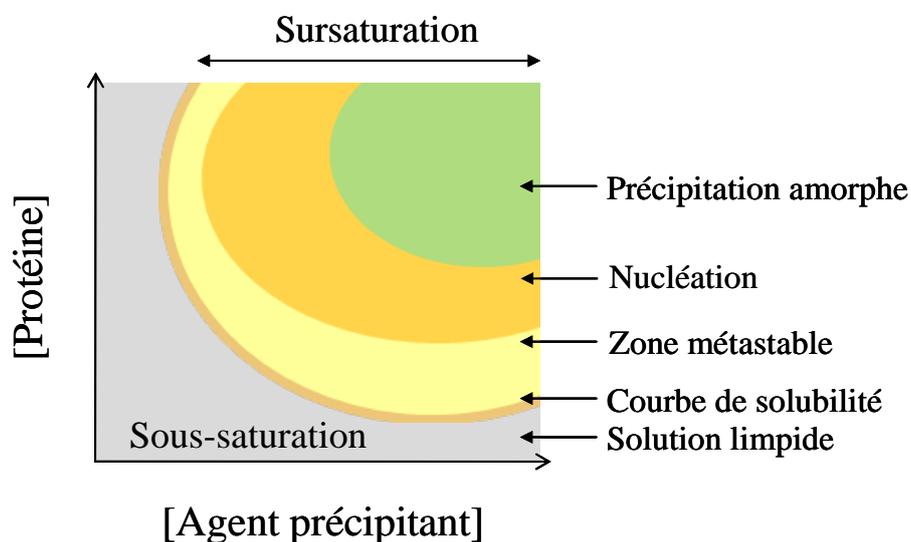
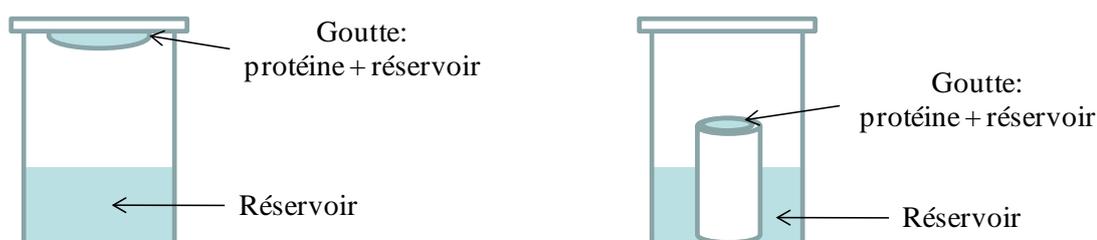


Figure 39: Le diagramme des phases. Dans la zone de sous-saturation, la protéine ne pourra jamais cristalliser ni précipiter (zone grise). La courbe de solubilité délimite la frontière zone sous-saturée/zone sursaturée. En fonction des cinétiques et des taux de sursaturation, la zone de sursaturation peut être divisée en trois parties: la zone de précipitation où il y a formation d'un précipité amorphe (en vert), la zone de nucléation où des germes cristallins peuvent se former (en orange) et une zone métastable qui est une zone favorable pour la croissance cristalline des germes déjà formés et pour l'ensemencement (en jaune).

## 5.2. Les essais de cristallisation

Deux dispositifs expérimentaux ont été utilisés: la goutte suspendue et la goutte assise (Figure 40). Des essais de cristallisation en gouttes assises ont été réalisés pour le criblage des conditions de cristallisation. L'optimisation des cristaux a été réalisée en gouttes suspendues en utilisant des boîtes Nextal.



Cristallisation en goutte suspendue

Cristallisation en goutte assise

Figure 40: Les deux dispositifs utilisés pour la cristallisation par diffusion de vapeur.

De nombreux essais de cristallisation sont nécessaires pour trouver des conditions de cristallisation. L'utilisation des robots de cristallisation, disponibles sur la plateforme de cristallisation du département de biologie et génomique structurales (DBGS), permet de tester de nombreuses conditions en un temps réduit. Le robot Tecan Workstation 150, muni d'une tête 96 supports de cônes, permet le remplissage des réservoirs d'une boîte de cristallisation en moins d'une minute. Ensuite le robot de cristallisation, Cartesian Honeybee 8+1, permet de mélanger la protéine, seule ou en mélange avec un ou plusieurs de ses partenaires, avec la solution du réservoir. Dans notre cas, les gouttes ont été réalisées en mélangeant 100 nl de la solution protéique avec 100 nl du réservoir. Plusieurs kits de criblage commerciaux peuvent être utilisés en premier lieu pour tester très rapidement un grand nombre de conditions de cristallisation. Lors des criblages des cristaux peuvent apparaître, et il reste alors à optimiser les conditions de croissance cristalline. Par contre, en l'absence de cristaux, des informations sur la solubilité de la protéine en fonction des différents paramètres testés peuvent être obtenues et des essais de cristallisation plus spécifiques peuvent être élaborés. Différents types de boîtes à 96 puits peuvent être utilisés selon le nombre de gouttes à faire par condition: une, deux ou trois gouttes. Les boîtes à plusieurs gouttes par condition permettent de tester plusieurs tampons de conditionnement ou plusieurs concentrations de la protéine ou de comparer le comportement de la protéine seule et en présence de différents ratios de ses partenaires. Le volume de réservoir utilisé est de 500  $\mu$ l. Les essais de cristallisation ont été

réalisés à deux températures, 4 et 17°C. Pour l'optimisation des cristaux en boîtes Nextal, le volume du réservoir était de 1 ml et les gouttes, le plus souvent, contenaient 1 µl protéine et 1 µl réservoir.

### **5.3. Enregistrement des données de diffraction**

La cristallographie aux rayons X est la méthode la plus utilisée pour déterminer la structure atomique des macromolécules biologiques. Les rayons X interagissent avec les électrons qui constituent l'enveloppe des atomes. L'interaction entre ces rayons et une molécule génère une diffusion de faible intensité, c'est pour cela que nous utilisons un cristal dans lequel les molécules sont ordonnées dans les trois directions de l'espace, pour amplifier le signal. Les rayons X interagissent alors avec le cristal, ils sont alors diffractés et le détecteur enregistre le cliché de diffraction du cristal.

Pour augmenter la durée de vie du cristal et réduire la dégradation due aux rayons X, le cristal doit être congelé avant l'enregistrement des données. Un agent cryoprotectant doit être utilisé préalablement à la congélation pour empêcher la formation de glace à partir du solvant contenu dans le cristal qui détériorerait ses propriétés de diffraction. Le cristal est manipulé à l'aide d'une boucle, transféré dans une solution cryoprotectante puis congelé dans de l'azote liquide.

Les sources de rayons X utilisées sont le diffractomètre du laboratoire (Micromax 007 HF) à anode tournante et le rayonnement synchrotron (la ligne PROXIMA 1 du synchrotron SOLEIL). La longueur d'onde du rayonnement utilisé dans le laboratoire correspond à la raie du cuivre ( $\lambda = 1.54 \text{ \AA}$ ). Le rayonnement synchrotron est un rayonnement blanc très intense qui permet entre autres de réduire le temps d'exposition du cristal aux rayons X, d'améliorer le rapport signal/bruit, d'augmenter la résolution et de choisir la longueur d'onde. La possibilité de choisir la longueur d'onde permet l'utilisation de la diffusion anormale pour la détermination des phases expérimentales par la méthode SAD ou MAD.

Une première image de diffraction permet d'estimer la qualité de diffraction du cristal. La répartition des tâches de diffraction sur le cliché permet d'extraire les informations concernant le réseau cristallin: c'est l'indexation. Cette étape permet de déterminer le groupe d'espace et les paramètres de maille du cristal. L'indexation d'une ou plusieurs images permet de définir une stratégie d'enregistrement: la distance cristal-détecteur, le domaine angulaire minimal à mesurer, l'angle d'oscillation et le temps d'exposition. Ensuite un jeu complet de diffraction est enregistré. Les clichés de diffraction sont traités en trois étapes: indexation, intégration et mise à l'échelle. Lors de l'étape d'intégration, l'intensité des

réflexions est déterminée. L'ensemble des tâches de diffraction est ensuite mise à l'échelle afin de constituer un jeu de données complet et homogène. En effet, les réflexions équivalentes par symétrie ou enregistrées plusieurs fois sont moyennées. Cependant, le pouvoir de diffraction du cristal diminue au cours de l'enregistrement suite aux dommages dus aux radiations, le rapport signal sur bruit,  $I/\sigma I$  permet de donner une estimation de l'intensité moyenne des réflexions mesurées en fonction de la résolution, où  $I$  est l'intensité des réflexions mesurées et  $\sigma I$  l'écart type des intensités de réflexions. Les programmes HKL-2000 et XDS ont été utilisés pour le traitement des données de diffraction.

#### **5.4. Détermination des phases expérimentales par la méthode SAD**

Pour résoudre la structure d'une macromolécule en cristallographie il faut trouver la phase et l'amplitude du facteur de structure  $F_{hkl}$  de chaque réflexion  $hkl$ . Ce facteur de structure est un nombre complexe liée à la densité électronique par une transformée de Fourier. La mesure de l'intensité diffractée  $I(hkl)$  de chacune des tâches de diffraction d'indice  $(hkl)$  permet de déterminer l'amplitude du facteur de structure car l'intensité est proportionnelle au carré de l'amplitude du facteur de structure. Mais la phase de ce dernier ne peut pas être déterminée à partir des clichés de diffraction. Pour obtenir les phases expérimentales, nous avons utilisé la méthode SAD (Single wavelength Anomalous Diffraction) avec les atomes de zinc fixés par le domaine PHD de hUHRF1, comme diffuseurs anomaux. Avant l'enregistrement des données cristallographiques, un spectre de fluorescence est enregistré sur le cristal pour obtenir la longueur d'onde correspondant au maximum d'absorption des atomes de zinc et déterminer les valeurs des termes de la diffusion anormale  $f'$  et  $f''$ . Ensuite un jeu de données doit être enregistré à la longueur d'onde correspondant au pic d'absorption qui est le maximum de  $f''$ . Les programmes Solve/Resolve ont été utilisés pour déterminer les phases expérimentales et obtenir une première carte de densité électronique permettant de construire un premier modèle de la macromolécule étudiée. Ensuite les programmes PHENIX et REFMAC ont été utilisés pour affiner le modèle issu de l'interprétation de la carte de densité électronique expérimentale, par un procédé mathématique qui minimise l'écart entre les modules de facteurs de structure observés et les modules de facteur de structure calculés à partir du modèle. Plusieurs cycles d'affinement et d'optimisation du modèle structural sont nécessaires pour aboutir au modèle final. Le programme COOT a été utilisé pour visualiser les cartes de densité et modifier les modèles cristallographiques.

## **Chapitre III**

# **Caractérisation de la protéine PHD-SRA<sub>296-643</sub>**

---

- 1. Clonage, production et purification de PHD-SRA**
- 2. Caractérisation en solution de PHD-SRA<sub>296-643</sub>**
- 3. Caractérisation des interactions de PHD-SRA<sub>296-643</sub>  
avec ses partenaires**
- 4. Cristallisation**
- 5. Conclusion**

Ce travail de thèse a débuté par la caractérisation des deux domaines associés PHD et SRA dans le but de déterminer la structure des deux domaines centraux de hUHRF1 impliqués dans la régulation épigénétique et d'appréhender les interactions interdomaines en relation avec les propriétés fonctionnelles de hUHRF1.

## **1. Clonage, production et purification de PHD-SRA**

Différentes constructions correspondant à différentes bornes des domaines PHD et SRA ont été clonées par la technologie Gateway®. La protéine d'intérêt est exprimée avec une étiquette (His)<sub>6</sub> ou (His)<sub>6</sub>-GST en amino-terminal et un site de coupure à la protéase TEV a été introduit permettant d'obtenir la protéine d'intérêt sans étiquette.

### **1.1. Les différentes constructions de PHD-SRA**

Le domaine PHD canonique correspond à la région 318-363 et le domaine SRA à la région 415-586. En amino-terminal du domaine PHD, se trouve un motif contenant 4 cystéines conservées dans la famille UHRF, entre les résidus 302 et 316. Trois résidus N-terminaux ont été choisis comme bornes pour les différentes constructions: R296, D308 ou A317. Les constructions qui commencent au résidu R296 comprennent le domaine PHD canonique ainsi que le motif amino-terminal riche en cystéines. Seulement deux cystéines du motif amino-terminal sont présentes en amont du PHD pour la construction D308. Les constructions qui commencent à l'alanine 317 comprennent uniquement le domaine PHD canonique. La structure du domaine SRA a été résolue dans notre équipe et la construction qui a permis de résoudre cette structure correspond à la région 408-643. Dans cette structure, la densité électronique correspondant à la région 616-643 n'était pas définie. En se basant sur la structure de SRA, trois résidus de terminaison ont été choisis: E593, Q606 et G643. Les constructions qui se terminent à l'acide glutamique 593 contiennent le domaine SRA canonique, celles qui se terminent à la glutamine 606 comprennent une hélice supplémentaire et celles se terminant à la glycine 643 correspondent au domaine SRA qui a cristallisé. Les différentes constructions qui ont été clonées sont présentées dans le Tableau 13 et la Figure 41. Les différentes protéines sont bien exprimées dans les bactéries *E.coli* Rosetta 2 et la construction la plus longue, correspondant à la région 296-643, a été choisie pour commencer l'étude structurale car le motif conservé riche en cystéines en amino-terminal pourrait conférer des propriétés particulières à la protéine.

Domaine protéique	Fusion amino-terminale	Vecteur d'expression	Concentration (ng/μl)
296-593	(His) <sub>6</sub>	3652	58
	(His) <sub>6</sub> -GST	3653	56
296-606	(His) <sub>6</sub>	3654	41
	(His) <sub>6</sub> -GST	3655	40
296-643	(His) <sub>6</sub>	2825	60
	(His) <sub>6</sub> -GST	2826	66
308-593	(His) <sub>6</sub>	3656	57
	(His) <sub>6</sub> -GST	3657	59
308-606	(His) <sub>6</sub>	3658	64
	(His) <sub>6</sub> -GST	3659	125
317-593	(His) <sub>6</sub>	3697	110
	(His) <sub>6</sub> -GST	3698	83
317-606	(His) <sub>6</sub>	3699	48
	(His) <sub>6</sub> -GST	3700	42
317-643	(His) <sub>6</sub>	3701	45
	(His) <sub>6</sub> -GST	3702	102

Tableau 13: Les différentes constructions clonées correspondant à la protéine PHD-SRA.

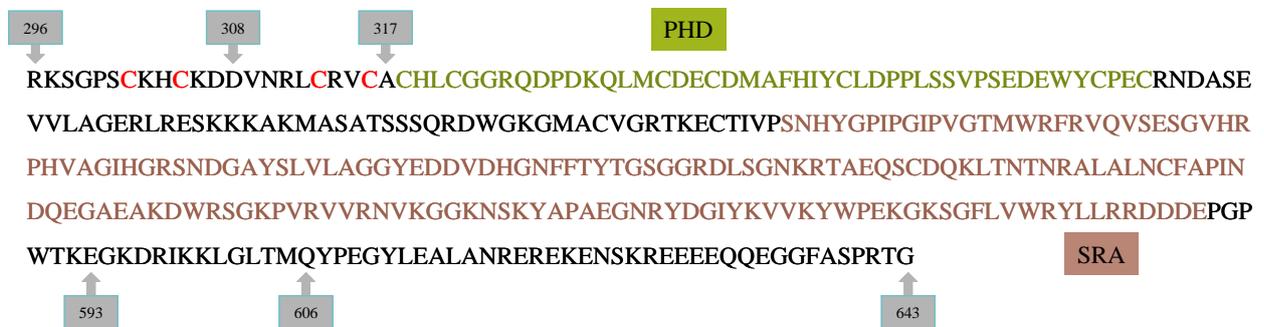


Figure 41: La séquence de la protéine PHD-SRA avec les différentes bornes choisies pour le clonage.

## 1.2. Expression et purification de PHD-SRA<sub>296-643</sub>

La protéine est surexprimée dans la souche *E. coli* Rosetta 2 par induction avec 0.5 mM d'IPTG pendant 5 h à 20°C en présence de 100 μM ZnCl<sub>2</sub>. L'addition de ZnCl<sub>2</sub> permet de stabiliser le domaine PHD.

Les protocoles de purification, initial et optimisé, de la protéine PHD-SRA<sub>296-643</sub> sont présentés dans le chapitre II partie 3.3.1. Après la lyse des bactéries par sonication, deux pics ont été séparés sur la première chromatographie d'affinité: le premier pic contient les

protéines fixées de façon faiblement spécifique et le deuxième qui présente son maximum à ~ 220 mM imidazole contient la protéine d'intérêt. Ensuite une étape de dialyse à pH 7 permet d'éliminer l'imidazole et de diminuer le pH pour être compatible avec le pI de la protéine sans son étiquette. Après digestion par la protéase TEV, une deuxième colonne d'affinité est réalisée. Deux pics ont été obtenus dans les lavages et un pic dans le gradient d'imidazole à 40 mM. Des analyses de spectrométrie de masse en conditions dénaturantes ont montré la présence de la protéine d'intérêt dans les trois pics. Pour les mesures en conditions non dénaturantes, les spectres n'étaient pas assez résolutifs pour déterminer le nombre exact de zinc fixés par ces protéines, du fait de la présence d'adduits de NaCl sur la protéine et de la précipitation de la protéine dans l'ammonium acétate. Partant de l'hypothèse d'une variation du nombre de zinc entre les différents pics de protéine et pour homogénéiser notre échantillon, deux dialyses successives en présence de 5 mM EDTA ont été réalisées pour enlever les atomes zinc fixés aspécifiquement à la protéine suivies d'une dialyse sans EDTA avant la deuxième colonne d'affinité. Après ces dialyses, le pic à 40 mM imidazole est obtenu. Les fractions correspondant à ce pic ont été injectées sur une colonne de gel filtration et trois pics ont été obtenus (Figure 42A). Le premier pic ayant le volume d'élution le plus faible correspond à la protéine agrégée ( $V_e = 33$  ml). D'après la calibration de la colonne, le deuxième pic qui est élué à un volume de 35 ml contient essentiellement de dimères et le troisième pic élué à un volume de 40 ml contient essentiellement la forme monomérique. Une hypothèse a été faite que le dimère se forme suite à la perte du zinc du motif N-terminal. Les cystéines ainsi libérées formeraient des ponts disulfures avec les cystéines d'un autre monomère. Pour vérifier cette hypothèse, une série de dialyse en présence d'agent réducteur, d'EDTA et de  $ZnCl_2$  a été réalisée. La dialyse contre un tampon contenant de l'EDTA conduit à la l'augmentation de la proportion de dimères et d'agrégats. Par contre une dialyse contre 1 mM TCEP (agent réducteur) permet de déplacer le pic de dimères vers le pic de monomères. L'ajout d'EDTA chélate le zinc du motif amino-terminal et conduit à la dimérisation et à l'agrégation de la protéine via la formation de ponts disulfures. L'ajout de TCEP réduit les ponts disulfures et favorise les monomères. Pour empêcher la reformation du dimère après la réduction des cystéines, l'ajout du  $ZnCl_2$  est indispensable. D'après ces expériences une étape de dialyse en présence de TCEP et de  $ZnCl_2$  avant l'étape de gel filtration était nécessaire pour obtenir une majorité de protéine monomérique (Figure 42B). Le protocole de purification de PHD-SRA optimisé permet d'obtenir environ 2.25 mg de la protéine monomérique à partir d'un litre de culture.

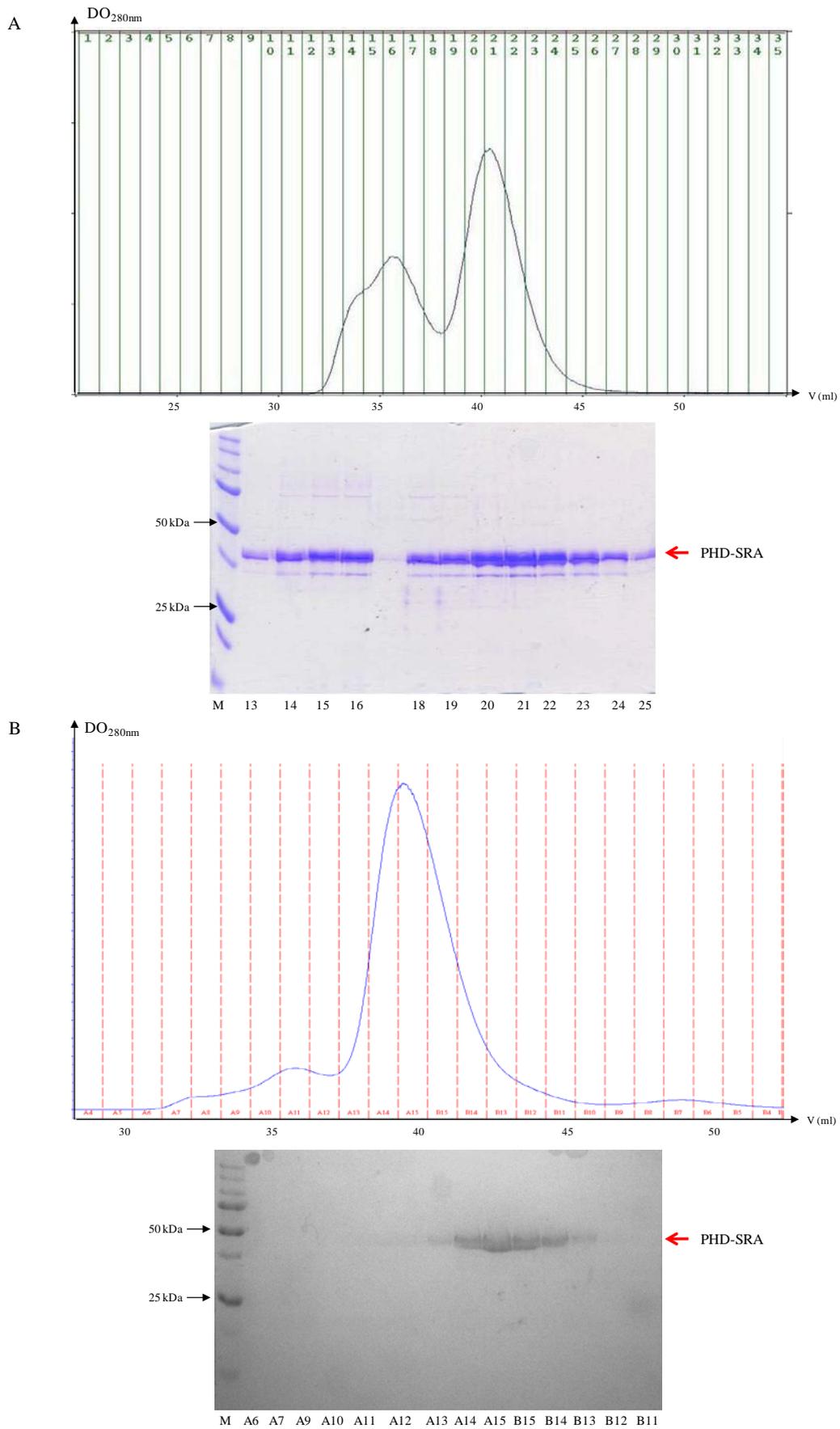


Figure 42: Le chromatogramme et le gel SDS correspondant de l'étape de chromatographie d'exclusion dans le protocole initial (A) et optimisé (B).

## 2. Caractérisation en solution de PHD-SRA<sub>296-643</sub>

Différentes techniques ont été utilisées pour caractériser en solution la protéine PHD-SRA<sub>296-643</sub> purifiée.

### 2.1. Caractérisation de PHD-SRA<sub>296-643</sub> par spectrométrie de masse

Pour déterminer sa stœchiométrie en zinc, des mesures de spectrométrie de masse ont été réalisées sur la protéine monomérique. Les premières mesures ont été réalisées sur le spectromètre de masse LCT (Waters Micromass, Manchester, UK) en conditions dénaturantes à 5  $\mu$ M et en conditions natives à 15  $\mu$ M. La masse moléculaire déterminée en conditions dénaturantes est 38 938.65 Da qui est très proche de la masse théorique (38 937.63 Da). En conditions natives, le pic majoritaire a une masse de 39 123.82 Da (Figure 43A). L'augmentation théorique de masse pour trois atomes de zinc liés à la protéine est:  $(3 \times 65.39) - (3 \times 2) = 190.17$  Da, car pour chaque atome de zinc fixé, il y a une perte de deux atomes d'hydrogène. La différence de masse entre les conditions natives et dénaturantes est de 185.17 Da. Cette masse ne correspond pas exactement à celle attendue pour la fixation de 3 trois atomes zinc mais cela peut être dû à la mauvaise résolution des pics obtenus. Dans le but d'affiner les pics, d'autres analyses ont été réalisées sur le micrOTOF-Q (Bruker Daltonics), une machine plus sensible et plus résolutive, qui requiert des quantités plus faibles d'échantillon (2  $\mu$ M pour les mesures en conditions dénaturantes et 5  $\mu$ M en conditions natives). La résolution des pics a été améliorée mais est restée insuffisante, et les masses trouvées sont de 38 937.18 Da en conditions dénaturantes et de 39 129.84 Da en conditions natives. Dans ce cas, la différence de masse entre les deux mesures est de 192.66 Da, ce qui correspond à 3 trois atomes zinc fixés (Figure 43B).

### 2.2. Caractérisation de PHD-SRA<sub>296-643</sub> par SAXS

En l'absence de structure tridimensionnelle, des mesures de SAXS ont été réalisées pour déterminer une enveloppe moléculaire de la protéine PHD-SRA<sub>296-643</sub> en solution, afin de voir si des éventuelles interactions existent entre les deux domaines. La protéine, à deux concentrations, a été conditionnée en 50 mM Tris pH 8, 500 mM NaCl et 0.1 mM TCEP. La polydispersité des échantillons a été étudiée par diffusion dynamique de la lumière (DLS) avant les mesures SAXS (Tableau 14).

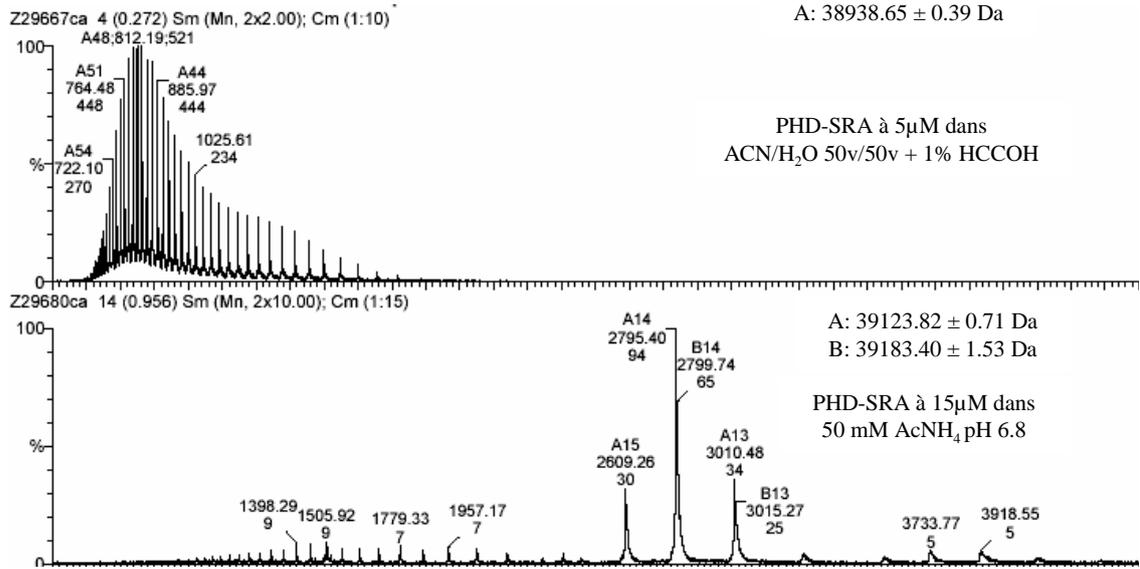
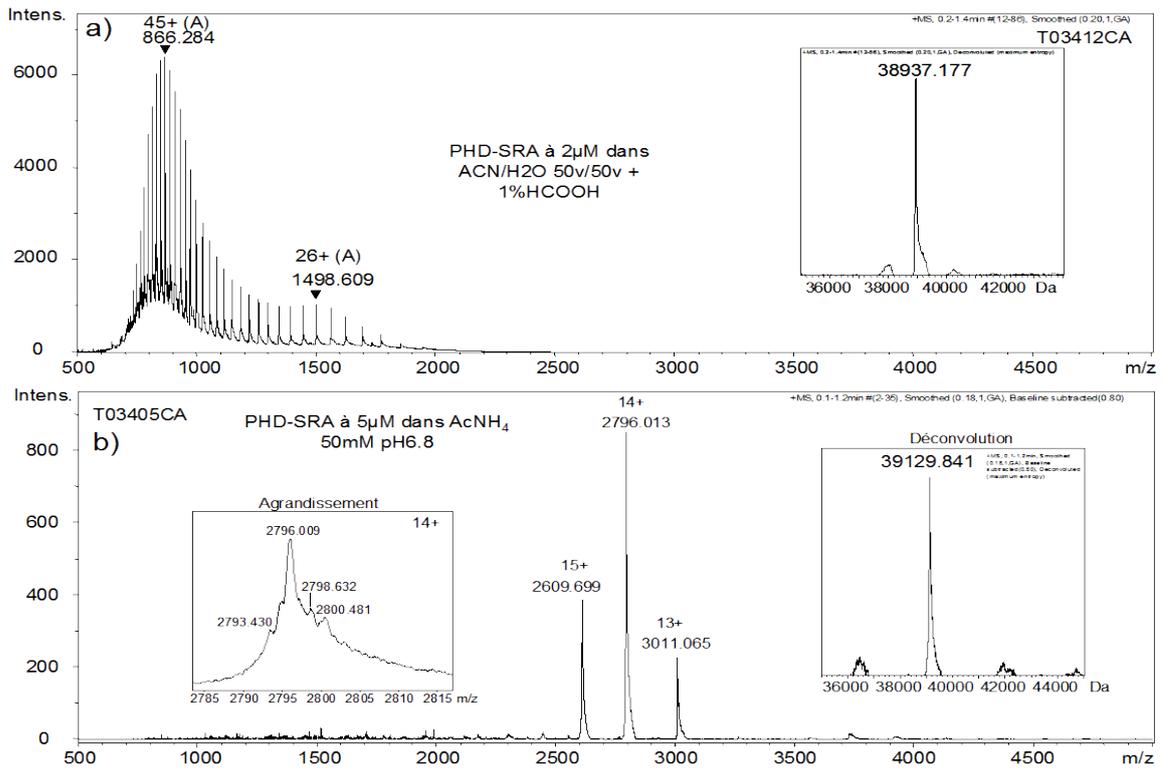
**A****B**

Figure 43: Analyse par spectrométrie de masse de la protéine PHD-SRA<sub>296-643</sub>. A- Résultats obtenus en conditions dénaturantes et natives en utilisant le spectromètre de masse LCT. B- Résultats obtenus en conditions dénaturantes et natives en utilisant le spectromètre de masse microTOF-Q.

Protéine	Concentration	Polydispersité	Rayon hydrodynamique
PHD-SRA <sub>296-643</sub>	2.36 mg/ml	Cp/Rh = 17.7 %	3.51 nm
	4.99 mg/ml	Cp/Rh = 20.5 %	Deux espèces Rh = 3.38 nm ; 98.3 % en intensité Rh = 135.16 nm ; 1.7 % en intensité

Tableau 14: Caractéristiques de la protéine PHD-SRA<sub>296-643</sub> utilisée en SAXS

Les spectres de diffusion correspondant à la zone angulaire  $0.0064 \text{ \AA}^{-1} < s < 0.507 \text{ \AA}^{-1}$  de la protéine et du tampon seul ont été enregistrés à 15°C. L'intensité  $I_0$  normalisée à 1 mg/ml obtenue avec la BSA est  $I_0 = 223.53$  et la valeur de la constante K calculée est alors 299.7. Seules les données obtenues avec la protéine à 2.36 mg/ml ont été exploitées du fait de la présence importante d'agrégats dans la solution concentrée à 4.99 mg/ml. Le tracé de Guinier montre toutefois la présence de traces d'agrégats avec la protéine à 2.36 mg/ml (Figure 44A). Le rayon de giration déterminé est de  $30.1 \pm 0.158 \text{ \AA}$ , pour  $s$  compris entre 0.0164 et  $0.0443 \text{ \AA}^{-1}$  et  $sRg$  compris entre 0.501 et 1.33. La valeur de  $I_0 = 127.56 \pm 0.497$  donne  $M = 38.29 \text{ kDa}$  ce qui correspond à la masse moléculaire du monomère. Les données sont bruitées aux grands angles, seules les données de diffusion  $0.0173 \text{ \AA}^{-1} \leq s \leq 0.4002 \text{ \AA}^{-1}$  ont été utilisées. La distribution de distances  $P(R) = f(R)$ , obtenue avec le programme GNOM permet de déterminer un  $D_{\max}$  de  $90 \text{ \AA}$  (Figure 44B). Le modèle d'enveloppe ab initio construit par DAMMIN est comparé en utilisant SUPCOMB à un modèle construit par BUNCH à partir de la structure cristallographique de SRA (PDB: 2PB7) et d'un modèle par homologie du domaine PHD (Figure 44C et D). Le modèle BUNCH se superpose bien avec le modèle ab initio. La comparaison des courbes de diffusion expérimentales et celles calculées à partir du modèle obtenu par BUNCH a été réalisée en utilisant le programme CRY SOL et la valeur de Chi obtenue est de 1.263 (Figure 44E).

Nous avons ainsi déterminé un modèle à basse résolution de la protéine PHD-SRA. Ce modèle montre un arrangement linéaire entre les deux domaines qui suggère que chacun des deux domaines peut librement interagir avec ses partenaires.

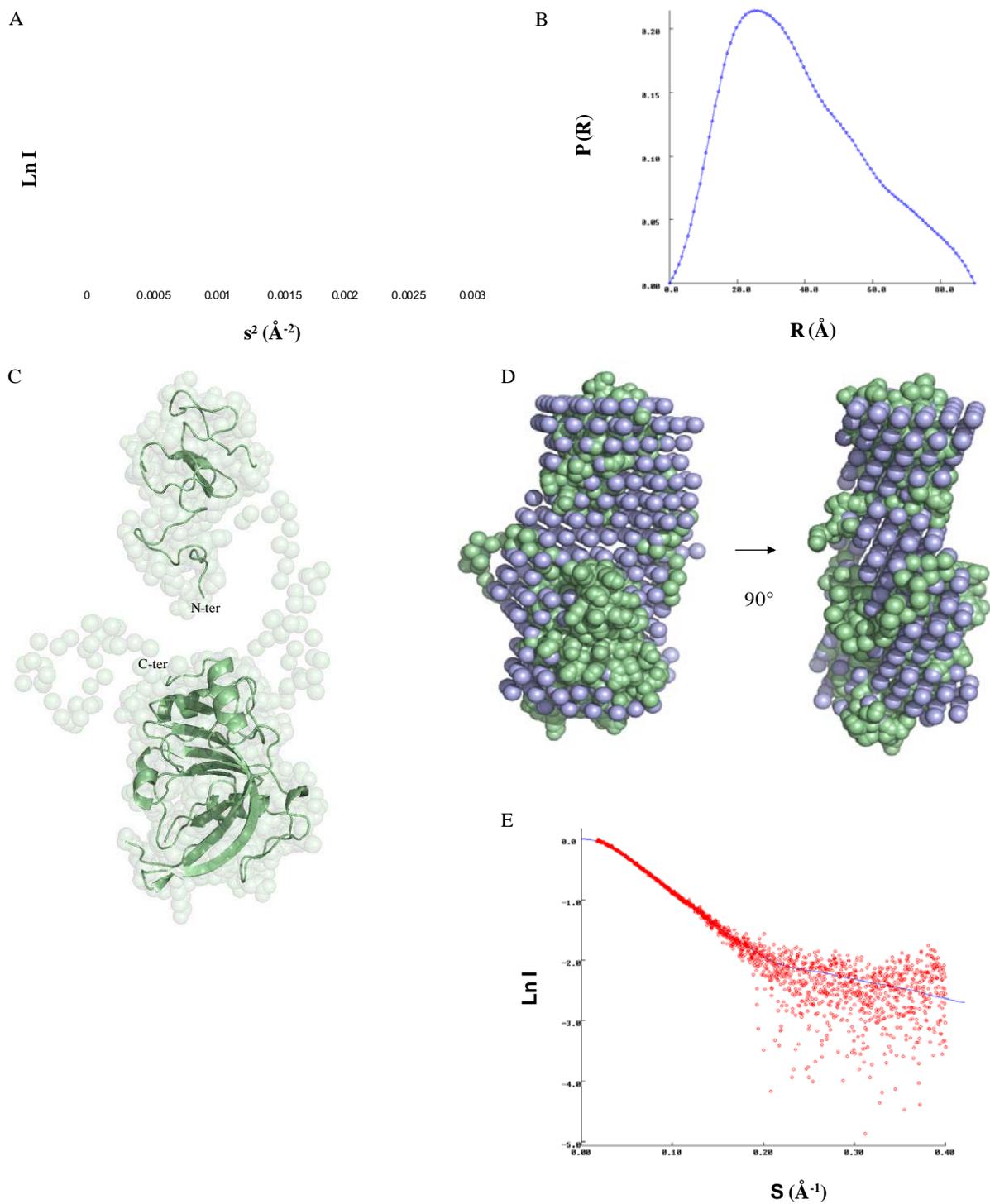


Figure 44: Données de diffusion obtenues avec la protéine PHD-SRA<sub>296-643</sub>. A- Tracé de Guinier. B- La courbe de distribution de distances obtenue avec GNOM. C- Le modèle obtenu avec BUNCH à partir des structures atomiques. D- Comparaison entre le modèle ab initio construit avec DAMMIN (en bleu) et le modèle obtenu avec BUNCH (en vert). E- Comparaison des courbes de diffusion expérimentales (en rouge) et calculées à partir du modèle BUNCH (en bleu) en utilisant le programme CRY SOL.

### 2.3. Caractérisation de PHD-SRA<sub>296-643</sub> par dichroïsme circulaire

Des mesures de dichroïsme circulaire ont été réalisées sur la protéine PHD-SRA<sub>296-643</sub> pour déterminer la composition en éléments de structures secondaires de la région interdomaine. La protéine a été conditionnée dans 20 mM phosphate pH 7 et 500 mM NaF. Différentes concentrations protéiques et différents trajets optiques ont été testés pour obtenir le meilleur ratio signal/bruit. Le meilleur ratio est obtenu avec une concentration protéique de 25 µM et un trajet optique de 0.2 mm. Le spectre enregistré à 17°C entre 182 et 260 nm est présenté dans la Figure 45 et son analyse dans le Tableau 15.

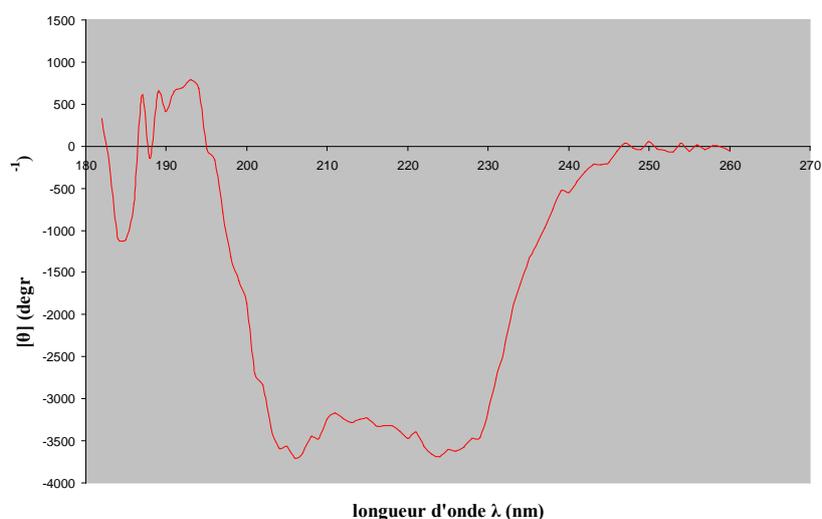


Figure 45: Spectre de dichroïsme circulaire de la protéine PHD-SRA<sub>296-643</sub>.

Tampon	Base de données	Programme utilisé	Hélices $\alpha$	Brins $\beta$	Tournants	Région non ordonnée	RMSD	NRMSD	
Haut sel	1	SELCON3	8,7	38,6	20,8	30,4	0,266	0,396	
		CDSSTR	5,0	36,0	24,0	33,0	0,163	0,243	
		CONTINLL	17,2	30,7	20,1	31,8	0,136	0,203	
	3	SELCON3	9,5	36,5	20,1	32,4	0,231	0,291	
		CDSSTR	6,0	35,0	24,0	34,0	0,120	0,150	
		CONTINLL	7,8	35,8	21,8	34,6	0,055	0,069	
	4	SELCON3	6,8	33,7	20,1	35,1	0,139	0,168	
		CDSSTR	6,0	35,0	24,0	34,0	0,101	0,121	
		CONTINLL	7,3	35,6	21,8	35,2	0,058	0,070	
	5	SELCON3	7,5	24,0	11,4	46,6	0,221	0,278	
		CDSSTR	6,0	25,0	14,0	45,0	0,124	0,156	
		CONTINLL	7,4	25,2	11,6	46,2	0,055	0,070	
		Moyenne		<b>7.9 ± 3.2</b>	<b>32.6 ± 5.1</b>	<b>19.5 ± 4.6</b>	<b>36,5 ± 5.9</b>		

Tableau 15: Analyse du spectre de dichroïsme circulaire de la protéine PHD-SRA<sub>296-643</sub> par 3 algorithmes et en utilisant quatre bases de données.

Les pourcentages en éléments de structures secondaires calculés à partir du spectre sont très hétérogènes en fonction du programme utilisé et de la base de données choisie. La protéine PHD-SRA<sub>296-643</sub> présente en moyenne 8 % d'hélices  $\alpha$ , 33 % de brins  $\beta$  et 20 % de tournants. Nous avons comparé les pourcentages en éléments de structures secondaires obtenus en solution avec les valeurs calculées à partir des structures des deux domaines isolés: le modèle par homologie du domaine PHD et la structure cristallographique du domaine SRA (PDB: 2pb7). Pour valider l'analyse, des mesures de dichroïsme circulaire ont également été réalisées sur le domaine SRA<sub>408-643</sub> et l'analyse des données a été comparée aux valeurs calculées à partir de la structure cristallographique de SRA<sub>408-643</sub> (Tableau 16).

	Analyse de dichroïsme circulaire		Analyse des structures		
	PHD-SRA (296-643) 349 résidus	SRA (408-643) 239 résidus	Modèle PHD <sub>296-363</sub> 68 résidus	SRA <sub>409-615</sub> (PDB: 2pb7) 207 visibles	PHD + SRA 275 résidus
Hélices $\alpha$	8 % 28 résidus	17 % 41 résidus	0 résidu	32 résidus	32 résidus
Brins $\beta$	33 % 115 résidus	28 % 67 résidus	6 résidus	53 résidus	59 résidus
Tournants	20 % 70 résidus	19 % 45 résidus	12 résidus	32 résidus	44 résidus

Tableau 16: Composition en éléments de structures secondaires pour PHD-SRA et SRA obtenues par dichroïsme circulaire et analyse structurale.

La comparaison entre les données expérimentales du domaine SRA isolé et les données calculées à partir de la structure cristallographique suggère que la partie carboxy-terminale, de 28 résidus, présente en solution mais invisible dans la structure doit être très riche en éléments de structures secondaires. Tous les résidus de cette partie carboxy-terminale sont soit des hélices, des brins  $\beta$  ou des tournants. Ce résultat paraît aberrant et en contradiction avec les prédictions de structures secondaires pour cette région par analyse bioinformatique. De même, la comparaison entre les données expérimentales de la protéine PHD-SRA<sub>296-643</sub> et les données calculées à partir des domaines isolés suggère que toute la région interdomaine, soit 44 résidus, doit être constituée de brins et de tournants. Ce résultat est également en contradiction avec les prédictions de structures secondaires pour cette région.

## 2.4. Optimisation du conditionnement de PHD-SRA<sub>296-643</sub> par Thermofluor

La technique de Thermofluor permet de déterminer les meilleures conditions stabilisant la protéine, par criblage de différents tampons, pH, additifs, sels et différentes concentrations salines et détermination des températures de dénaturation (ou de fusion  $T_m$ ) correspondantes. La protéine éluée de la colonne de gel filtration est conditionnée en 50 mM Tris pH 8, 500

mM NaCl et 0.2 mM TCEP. Dans ce tampon, sa température de fusion ( $T_m$ ) est de 39°C. En utilisant un tampon phosphate à pH 5.5, la valeur de  $T_m$  diminue à 31°C mais à des pH variant entre 6.5 et 8.5, la variation de  $T_m$  est négligeable (Figure 46). En diminuant la concentration saline à 50 mM, la valeur de  $T_m$  augmente faiblement à 41°C. L'ajout de 5 % glycérol, 5% pentaerythritol ethoxylate ou de 10% xylitol permet d'augmenter la valeur de  $T_m$  d'environ 3°C. L'ajout de différents peptides de l'histone H3 ou d'ADN n'a pas modifié de façon significative la valeur de  $T_m$ .

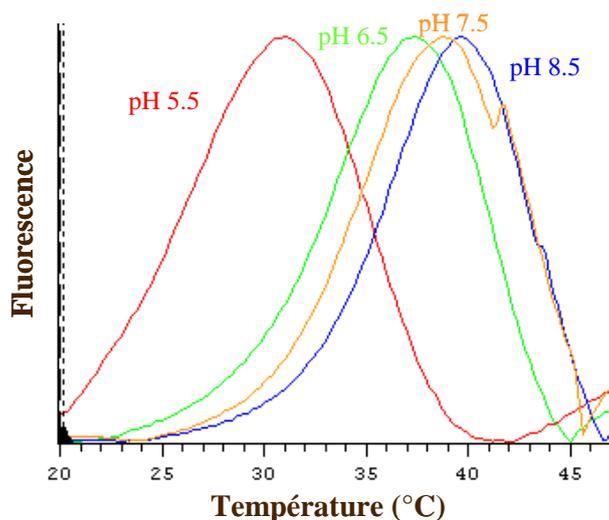


Figure 46: Courbes de dénaturation de PHD-SRA<sub>296-643</sub> obtenues par Thermofluor. Différents pH ont été testés à une concentration saline de 500 mM.

### 3. Caractérisation des interactions de PHD-SRA<sub>296-643</sub> avec ses partenaires

Comme le domaine SRA est impliqué dans la reconnaissance de l'ADN et que le domaine PHD peut être impliqué dans la reconnaissance de l'histone H3, les interactions de PHD-SRA avec l'ADN, méthylé ou non méthylé, et avec des peptides de l'histone H3 ont été étudiées.

#### 3.1. Etude de l'interaction de PHD-SRA avec l'ADN

Différentes techniques ont été utilisées pour étudier l'interaction de PHD-SRA avec l'ADN: la spectrométrie de masse en conditions natives, la chromatographie d'exclusion et l'ultracentrifugation analytique.

### 3.1.1. Caractérisation de l'interaction par spectrométrie de masse en conditions non dénaturantes

L'interaction entre PHD-SRA et un ADN de 20 pb contenant une boîte CpG méthylée (M = 12264 Da) ou non (M = 12236 Da) sur ses cytosines a été caractérisée (Figure 47).



Figure 47: La séquence de l'ADN utilisé pour étudier l'interaction avec PHD-SRA<sub>296-643</sub> en spectrométrie de masse. La boîte CpG est surlignée en gris et les deux cytosines méthylées ou non sont en rouges.

Les spectres obtenus montrent que la protéine est capable de former un complexe 1:1 avec l'ADN qu'il soit méthylé ou non. Les signaux étant peu intenses et les pics étant larges, la mesure d'une masse précise a été difficile et la détermination des constantes d'affinité n'a pas été possible (Figure 48).

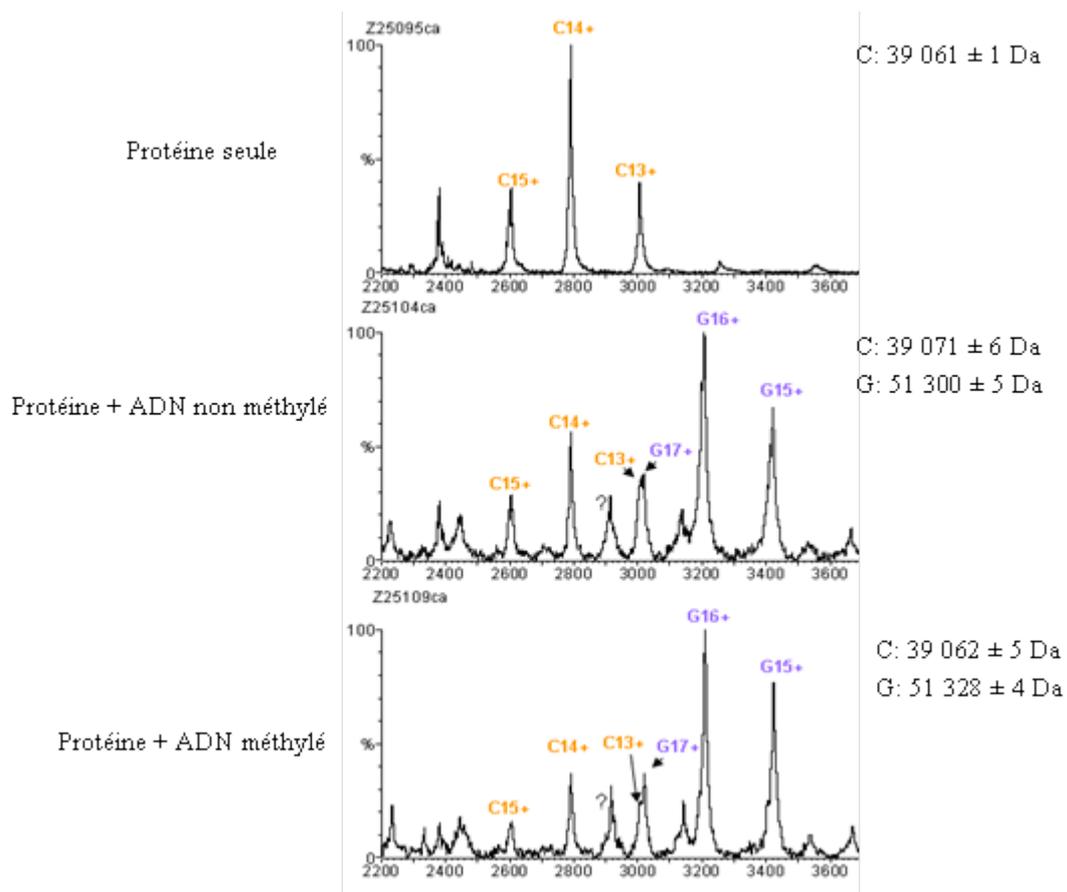


Figure 48: Etude de l'interaction de PHD-SRA<sub>296-643</sub> avec un ADN méthylé et non méthylé par spectrométrie de masse en conditions natives.

### 3.1.2. Caractérisation de l'interaction par chromatographie d'exclusion

L'étude de l'interaction entre PHD-SRA<sub>296-643</sub> et l'ADN a aussi été réalisée par injection de la protéine seule, l'ADN seul et du mélange sur une colonne de gel filtration analytique s75 10/300. Différentes concentrations salines ont été testées pour déterminer la concentration saline optimale pour la formation du complexe. Nous avons choisi d'utiliser, pour des raisons de coût, l'ADN non méthylé pour mener cette étude. Des mélanges ADN-protéine à 70  $\mu$ M, ont été préparés et dialysés contre des tampons (25 mM Tris pH 7 ; 0.1 mM TCEP) contenant différentes concentrations salines: 500 mM, 300 mM, 200 mM, 100 mM ou 50 mM NaCl. 100  $\mu$ l de chaque mélange sont injectés sur la colonne de gel filtration ainsi que l'ADN seul et la protéine seule conditionnés dans le tampon correspondant. A des concentrations salines variant de 500 à 100 mM NaCl, il n'y a pas formation du complexe ADN-protéine. En diminuant la concentration saline à 50 mM, un pic correspondant au complexe ADN-protéine est apparu (Figure 49).

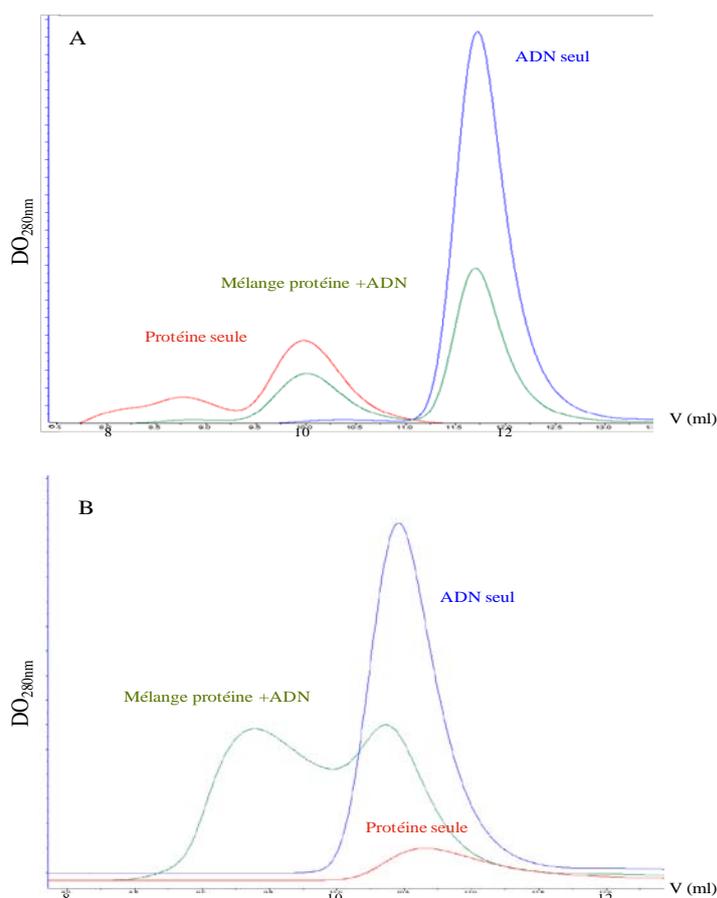


Figure 49: Caractérisation du complexe entre PHD-SRA<sub>296-643</sub> et l'ADN non méthylé sur colonne de gel filtration analytique. A- Superposition des chromatogrammes obtenus en présence de 500 mM NaCl pour l'ADN seul (bleu), la protéine seule (rouge) et le mélange protéine-ADN(vert). Aucun pic de complexe n'est observé. B- Superposition des chromatogrammes obtenus en présence de 50 mM NaCl pour l'ADN seul (bleu), la protéine seule (rouge) et le mélange protéine-ADN (vert). Un pic de complexe est visible à 9.3 ml d'éluion.

### 3.1.3. Caractérisation de l'interaction par ultracentrifugation analytique

Nous avons réalisé des expériences de vitesse de sédimentation pour étudier l'interaction entre la protéine PHD-SRA<sub>296-643</sub> et l'ADN contenant la boîte CpG méthylée (20 pb, 12264 Da). Nous avons également déterminé l'état de monodispersité, le coefficient de sédimentation et le coefficient de friction des entités séparées (Tableau 17). Les mesures d'absorption ont été réalisées à 260 et 280 nm pour suivre l'ADN et la protéine respectivement.

Cellule	Espèce étudiée	Concentration		Coefficient d'extinction molaire théorique	volume partiel spécifique	
		Protéine	ADN		4°C	20°C
1	Protéine	17.7 $\mu$ M	-	$\epsilon_{280 \text{ nm}} = 56470$	0.7103 ml.g <sup>-1</sup>	0.7171 ml.g <sup>-1</sup>
				$\epsilon_{260 \text{ nm}} = 28235$		
2	ADN	-	2.9 $\mu$ M	$\epsilon_{280 \text{ nm}} = 173996$	0.5 ml.g <sup>-1</sup>	
				$\epsilon_{260 \text{ nm}} = 347993$		
3	Protéine + ADN	2.2 $\mu$ M	2.2 $\mu$ M	-	0.7103 ml.g <sup>-1</sup>	0.7171 ml.g <sup>-1</sup>
				-		

Tableau 17: Les concentrations, coefficients d'extinction molaire et les DO des différentes espèces étudiées par ultracentrifugation analytique en vitesse de sédimentation.

Les différentes entités ont été conditionnées dans un tampon 25 mM Tris pH 8, 50 mM NaCl et 0.2 mM TCEP. L'analyse de la distribution des coefficients de sédimentation en utilisant le programme sedfit, montre un pic majoritaire (92 % du signal) pour la protéine seule à une valeur de  $s$  de  $2.03 \pm 0.14$  S (Figure 50A). Le rapport  $f/f_0$  obtenu est de 1.35, indiquant que la protéine est asymétrique. La masse moléculaire calculée à partir de ces valeurs expérimentales est de  $37\,649 \pm 13\,199$  Da correspondant à la forme monomérique de la protéine. Concernant l'ADN seul, un problème de sédimentation anormale a été observé sur les premières courbes de sédimentation. Ce problème pourrait être dû à un mauvais positionnement de la cellule et/ou à un problème de non idéalité. L'analyse de ces données montre un pic majoritaire (79 % du signal) à une valeur de  $s$  de  $1.68 \pm 0.17$  S. Le rapport  $f/f_0$  déterminé est de 1.5 qui est en accord avec la forme allongée de l'ADN (Figure 50B). La masse moléculaire calculée est de  $12\,997 \pm 1014$  Da.

Pour étudier le mélange protéine-ADN, nous avons utilisé le programme sedphat qui permet une analyse globale de données de sédimentation enregistrées à 260 et 280 nm. Les coefficients de sédimentation obtenus sont corrigés à 20 °C dans l'eau ( $s_{20,w}$ ). Les résultats obtenus sont présentés dans le Tableau 18 et la Figure 51.

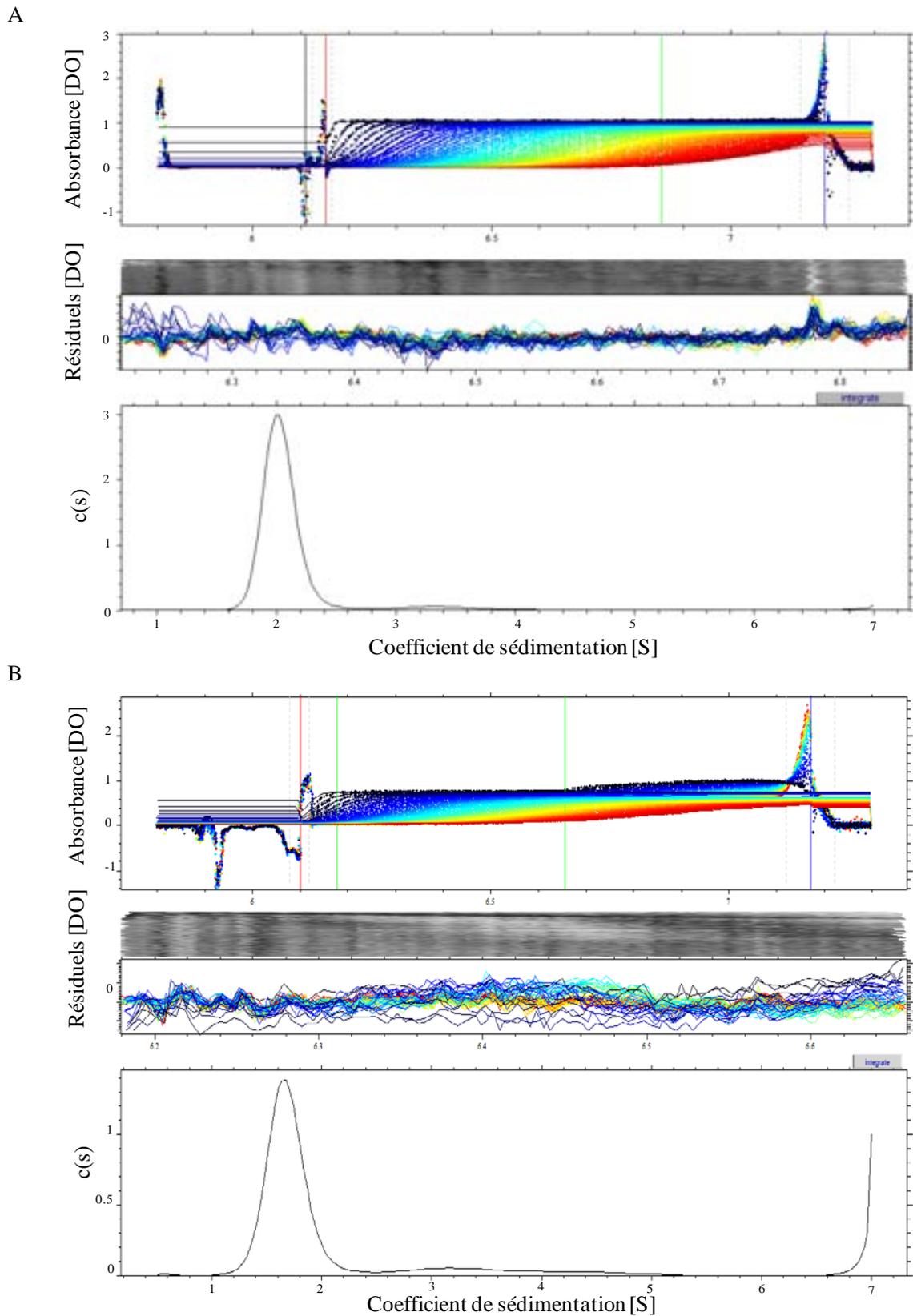


Figure 50: Données de vitesse de sédimentation de la protéine PHD-SRA<sub>296-643</sub> seule à 280 nm (A) et de l'ADN seul à 260 nm (B) analysées avec le programme Sedfit. Pour chaque longueur d'onde, le panneau supérieur représente les courbes de sédimentation enregistrées, le panneau central représente les résiduels associés au fit et le panneau inférieur représente la distribution de coefficients de sédimentation obtenue.

		Pics obtenus	
PHD-SRA	$s_{20,w} = 3.30 \text{ S}$ $C = 14.36 \mu\text{M}$ 81.1 %		
ADN	$s_{20,w} = 2.71 \text{ S}$ $C = 1.35 \mu\text{M}$ 46.6 %		
PHD-SRA + ADN	$s_{20,w} = 2.38 \text{ S}$ $C_{\text{ADN}} = 1.57 \mu\text{M}$ 71 %	$s_{20,w} = 4.37 \text{ S}$ $C_{\text{protéine}} = 0.94 \mu\text{M}$ 43%	$s_{20,w} = 6.61 \text{ S}$ $C_{\text{protéine}} = 0.27 \mu\text{M}$ 12 %
		$s_{20,w} = 4.11 \text{ S}$ $C_{\text{ADN}} = 0.43 \mu\text{M}$ 20 %	

Tableau 18: Les coefficients de sédimentation et les concentrations des espèces majoritaires après analyse des données de vitesse de sédimentation de PHD-SRA<sub>296-643</sub>, de l'ADN et du mélange protéine-ADN.

L'expérience réalisée avec l'ADN seul montre que moins de 50 % de l'ADN total sédimente à une valeur de  $s$  en accord avec un ADN de 20 pb. En présence de la protéine, l'ADN semble plus stable puisque 91 % de l'ADN total est retrouvé dans deux pics de sédimentation correspondant à l'ADN seul et au complexe. La valeur du coefficient de sédimentation de l'ADN obtenu à partir des données de sédimentation de l'ADN seul ( $s_{20,w} = 2.71 \text{ S}$ ) est légèrement différente de celle obtenue à partir des données du mélange ( $s_{20,w} = 2.38 \text{ S}$ ) avec également un pic plus large. Ceci est dû au fait d'avoir utilisé le volume partiel spécifique de la protéine pour traiter les courbes de sédimentation du complexe. Une comparaison des distributions de coefficients de sédimentation entre le mélange protéine-ADN et la protéine seule ou l'ADN seul, montre l'apparition d'un nouveau pic de sédimentation contenant les deux espèces à une valeur de  $s_{20,w} \approx 4.2 \text{ S}$ . Il apparait que seulement 20 % de l'ADN total et 43 % de la protéine totale sont impliqués dans la formation du complexe. Le ratio protéine/ADN dans ce pic est égal à 2 ce qui signifie que le complexe serait formé par deux molécules de la protéine fixant une molécule d'ADN. Ces résultats indiquent que le domaine SRA lié au PHD est capable de reconnaître l'ADN. Par contre, dans la structure cristallographique du complexe SRA-ADN, une seule molécule de SRA fixe un ADN hémiméthylé. Le fait d'avoir utilisé un ADN portant un dinucléotide CpG totalement méthylé peut être à l'origine de la stœchiométrie observée. Dans ce cas, chaque cytosine méthylée serait accessible pour fixer une molécule de SRA. Etant donné qu'environ la moitié de la protéine est impliquée dans la formation du complexe, ceci suggère que la concentration utilisée est du même ordre de grandeur que le  $K_d$  du complexe. Une détermination précise de cette constante nécessiterait de réaliser plusieurs expériences sur le complexe à différentes

concentrations mais la gamme de concentrations pouvant être balayée est limitée par la saturation du signal d'absorption mesuré.

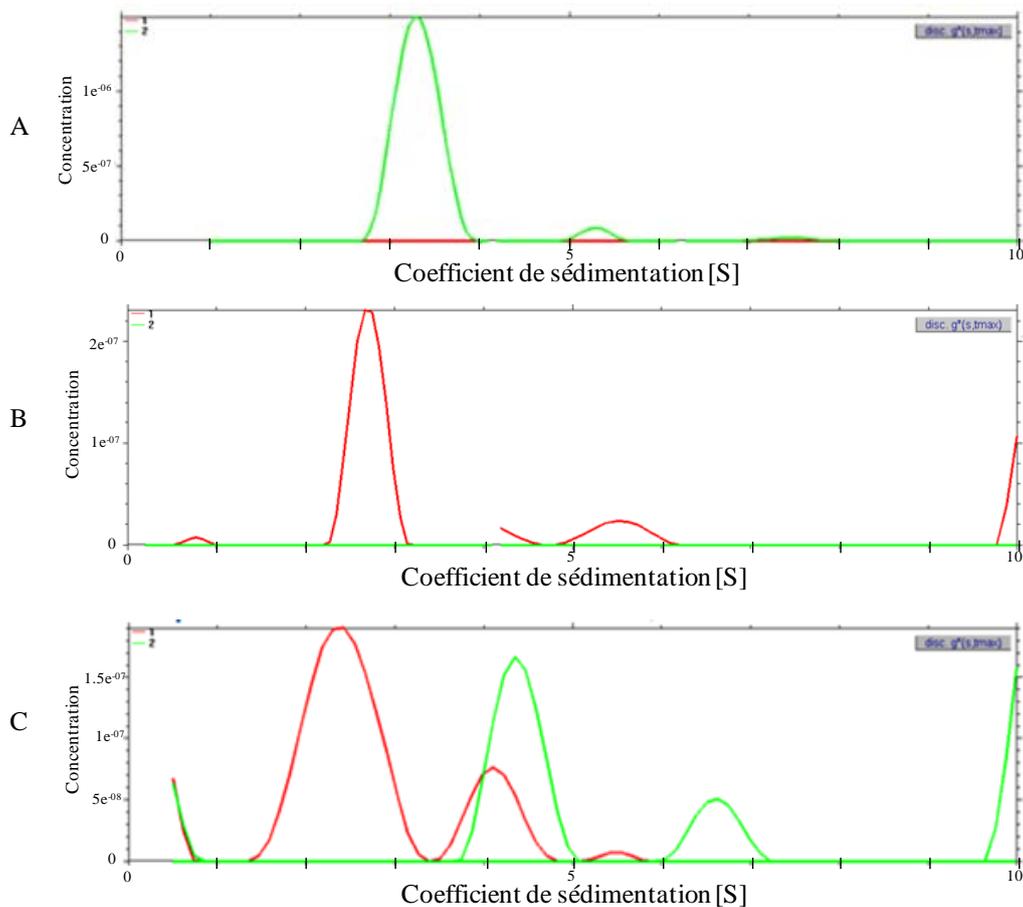


Figure 51: Courbes de distribution des coefficients de sédimentation de PHD-SRA seul (A), de l'ADN seul (B) et du mélange PHD-SRA avec ADN (C). L'analyse globale des données de sédimentation enregistrées à 280 et 260 nm a été effectuée avec le programme Sedphat.

### 3.2. Etude de l'interaction de PHD-SRA avec l'histone H3 par calorimétrie à titration isotherme (VP-ITC)

Karagianni et ses collaborateurs ont montré en 2008, par des expériences de « pull down » que chacun des domaines PHD et SRA contribue à l'interaction avec l'histone H3 triméthylée à la lysine 9. Afin de valider cette interaction et déterminer sa constante de dissociation, la calorimétrie à titration isotherme a été utilisée.

La protéine PHD-SRA<sub>296-643</sub> conditionnée en 25 mM Tris pH 7, 150 mM NaCl et 0.2 mM TCEP et concentrée à 28  $\mu$ M a été utilisée. L'analyse par DLS montre une polydispersité de 11%. Une première expérience a été réalisée avec cette protéine et un peptide de l'histone H3 non méthylé à la lysine 4 et triméthylé à la lysine 9 (H3K9me3) à 600  $\mu$ M (Figure 52).

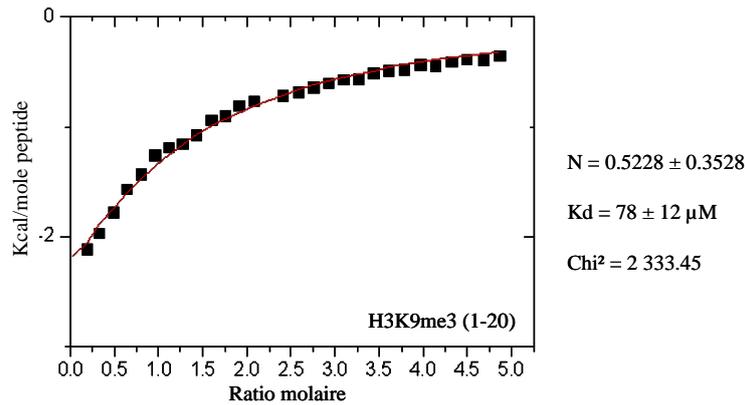


Figure 52: Etude de l'interaction par ITC entre PHD-SRA<sub>296-643</sub> à 28  $\mu\text{M}$  et le peptide d'histone H3K9me3 (1-20).

A partir d'une telle courbe, où il manque le premier plateau, le fit entre les données expérimentales et un modèle théorique d'interaction n'est pas bon, il est donc impossible de déterminer précisément un  $K_d$  ou une stœchiométrie d'interaction. Une deuxième série d'expériences a donc été réalisée en utilisant une protéine plus concentrée (52.4  $\mu\text{M}$ ) et présentant 14% de polydispersité en DLS. La concentration du peptide a également été augmentée à 2 mM (Figure 53). Nous avons observé une précipitation de la protéine dans la cellule en fin d'expériences, ce qui révèle un problème de stabilité de la protéine dans ces conditions expérimentales. La conséquence de cette précipitation est un problème de retour à la ligne de base au niveau des quatre dernières injections qui ont donc été supprimés pour l'analyse des données. Cette deuxième expérience nous a permis de confirmer une interaction entre PHD-SRA de hUHRF1 et le peptide de l'histone H3 triméthylé à la lysine 9. Le fit entre les données expérimentales et un modèle théorique s'est amélioré et une constante de dissociation d'environ 50  $\mu\text{M}$  avec une stœchiométrie de 1 a été obtenue. Ces valeurs restent approximatives en l'absence du premier plateau.

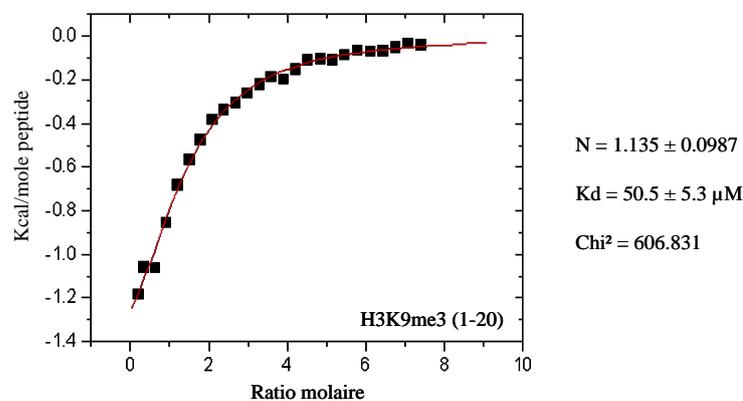


Figure 53: Etude de l'interaction par ITC entre PHD-SRA<sub>296-643</sub> à 52 $\mu\text{M}$  et le peptide d'histone H3K9me3 (1-20).

## 4. Cristallisation

Des essais de cristallisation ont été réalisés sur la protéine PHD-SRA seule ou en complexe avec de l'ADN et/ou des peptides de l'histone H3. Différents screens ont été testés en variant la concentration de la protéine, son conditionnement, le ratio des partenaires, la taille des peptides de l'histone H3, la température de cristallisation et en ajoutant certains additifs. Après plusieurs séries d'optimisation, des petits cristaux (70  $\mu\text{m}$  dans la plus grande dimension) ont été obtenus à 17°C (Figure 54). Ces cristaux sont apparus dans une goutte contenant 1  $\mu\text{l}$  du mélange ternaire: protéine (185  $\mu\text{M}$ ) + ADN non méthylée + H3K9me3 (6-17) à un ratio 1: 1: 5 et 1  $\mu\text{l}$  de réservoir (0.1 M Bis-Tris pH 5.5, 75 mM  $\text{MgCl}_2$ , 25 % PEG 3350 et 4% PEE). Ces cristaux très fragiles n'ont jamais été testés en diffraction.

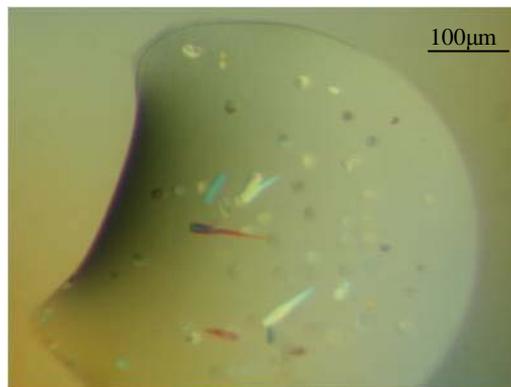


Figure 54: Cristaux obtenus dans une goutte contenant un mélange ternaire PHD-SRA<sub>296-643</sub>, peptide H3 et ADN non méthylé.

## 5. Conclusion

Le travail sur la protéine PHD-SRA<sub>296-643</sub> a permis la mise au point d'un protocole de purification permettant d'obtenir une protéine pure et homogène. Cette protéine fixe trois atomes zinc, deux par le domaine PHD canonique et un par le motif amino-terminal comprenant quatre cystéines. Une structure à basse résolution déterminée par SAXS montre un arrangement linéaire des deux domaines PHD et SRA. Nous avons pu montrer par des expériences de spectrométrie de masse, de chromatographie d'exclusion et de vitesse de sédimentation que cette protéine est capable de se fixer à l'ADN. Le site de liaison à l'ADN du domaine SRA reste donc accessible en présence du domaine PHD. De plus, des expériences préliminaires en ITC ont montrées une interaction entre la protéine PHD-SRA et un peptide de l'histone H3 ce qui suggère la fonctionnalité des deux domaines PHD et SRA. Des essais de cristallisation ont été réalisés avec cette protéine seule ou en complexe binaire et

ternaire avec l'ADN et le peptide de l'histone H3. Des microcristaux fragiles ont été obtenus mais n'ont pas pu être optimisés. Nous avons donc décidé d'entamer la caractérisation structurale et fonctionnelle du domaine PHD isolé. Les résultats obtenus sur le domaine PHD isolé devraient aider à faire aboutir l'étude structurale et fonctionnelle de la protéine PHD-SRA.

## Chapitre IV

# Analyse bioinformatique et clonage du domaine PHD de hUHRF1

---

1. Comparaison de séquences entre le domaine PHD de hUHRF1 et d'autres domaines PHD fixant l'histone H3
2. Comparaison structurale entre le PHD de hUHRF1 et des PHD fixant l'histone H3
3. Clonage du domaine PHD

L'étude structurale et fonctionnelle du domaine PHD de hUHRF1 a débuté par une analyse bioinformatique de ce domaine et des comparaisons structurales entre un modèle du domaine PHD de hUHRF1 et les structures disponibles d'autres domaines PHD. Plusieurs constructions ont été clonées. Deux constructions correspondant au domaine PHD canonique seul ou précédée du motif riche en cystéines ont été caractérisées durant ce travail.

## 1. Comparaison de séquences entre le domaine PHD de hUHRF1 et d'autres domaines PHD fixant l'histone H3

La séquence du PHD de hUHRF1 a été comparée aux séquences des trois classes connues de PHD reconnaissant l'histone H3 (Figure 55). Le PHD de hUHRF1 présente une conservation du résidu acide impliqué dans l'interaction avec l'arginine R2 de l'histone H3 (D334). Mais ce PHD ne possède pas les résidus aromatiques et hydrophobes nécessaires à l'interaction avec l'histone H3K4me3, ni le résidu acide en position -2 ou -4 de la première cystéine, nécessaire à la reconnaissance électrostatique de l'histone H3K4me0. De même, les résidus du PHD de CHD4 qui présentent le maximum de déplacements chimiques en présence d'un peptide H3K9me3 ne sont pas conservés dans le PHD de hUHRF1. A partir de la comparaison de séquences, nous n'avons pas pu déterminer la classe du PHD de hUHRF1.

		320	330	340	350	360
1	hUHRF1	CRVCA..	CHLGGGR.QDPDK.QLMCDE..	CD.MAFHIYCLDPPLSSVPSSEDEWYCPECR		
	Taf3	GNQIW..	ICPGCNKP.DDGSP.MIGCDD..	CD.DWYHWPVGVGIMAAPPEEM.QWFCPKCA		
	BPTF_PHD2	DTKLY..	C.ICKTPYDESKF.YIGCDR..	CQ.NWYHGRCVGILOSEAELIDEVVCPQCQ		
	ING2	NEPTY..	C.ICNQ..VSYGE.MIGCDNEQ	CPIEWFHFSVSLTYKPKGK..WYCPKCR		
2	BHC80	IHEDF..	CSVCRK..SGQ..LLMCDT..	CS.RVYHLDCLDPPLKTIPKG.MWICPRCQ		
	AIRE_PHD1	KNEDE..	CAVCRD..GGE..LICDGD..	CP.RAFHLACLSPPLREIPSG.TWRCSSCL		
	JADE1_PHD1	..DEDVV	CDVQSPDGEDGNEMVFCDK..	CN.ICVHQACYGILKVPEG..SWLCRTCA		
3	CHD4_PHD2	.HMEF..	CRVCKD..GGE..LLCDDT..	CP.SSYHIHCLNPPLPEIPNG.EWLCPRCT		
	Lid2_PHD2	RKKKG..	CIFCFRLPESG.VMIECEI..	CH.EWYHAKCLKMSKKKLRQDEKFTCPICD		
	SMCX_PHD1	FIESY.V	CRMCSRGDDEDK..LLLCDG..	CD.DNYHIFCLLPPLPEIPKG.VWRCPKCV		

Figure 55: Alignement de séquences entre le PHD de hUHRF1 et des PHD interagissant avec H3K4me3 (en jaune), H3 non modifiée (en vert) et H3K9me3 (en gris).

## 2. Comparaison structurale entre le PHD de hUHRF1 et des PHD fixant l'histone H3

Dans la famille UHRF, la structure RMN du PHD de hUHRF2 a été résolue (PDB: 2e6s). En amont du domaine PHD canonique, il existe dans cette structure un motif riche en cystéines capable de fixer un troisième zinc. Ce motif riche en cystéines est également présent dans le domaine PHD de hUHRF1. Ces deux séquences présentent 48% d'identité et



l'acide glutamique E488 de BHC80 qui forment des liaisons hydrogènes avec H3K4 sont absents dans hUHRF1 mais le carbonyle de la cystéine C316 de hUHRF1 qui pointe dans la même direction pourrait jouer un rôle similaire (Figure 58A). Dans la structure du premier domaine PHD de la protéine AIRE en complexe avec un peptide H3K4me0, le peptide de l'histone H3 n'adopte pas la même conformation que dans le complexe avec BHC80 (Figure 58B). Les arginines R2 du peptide de l'histone H3 occupent des positions très différentes dans les deux structures de complexes. Le modèle du PHD de hUHRF1 semble plus proche de la structure du PHD de BHC80 que de celui du PHD1 de AIRE.

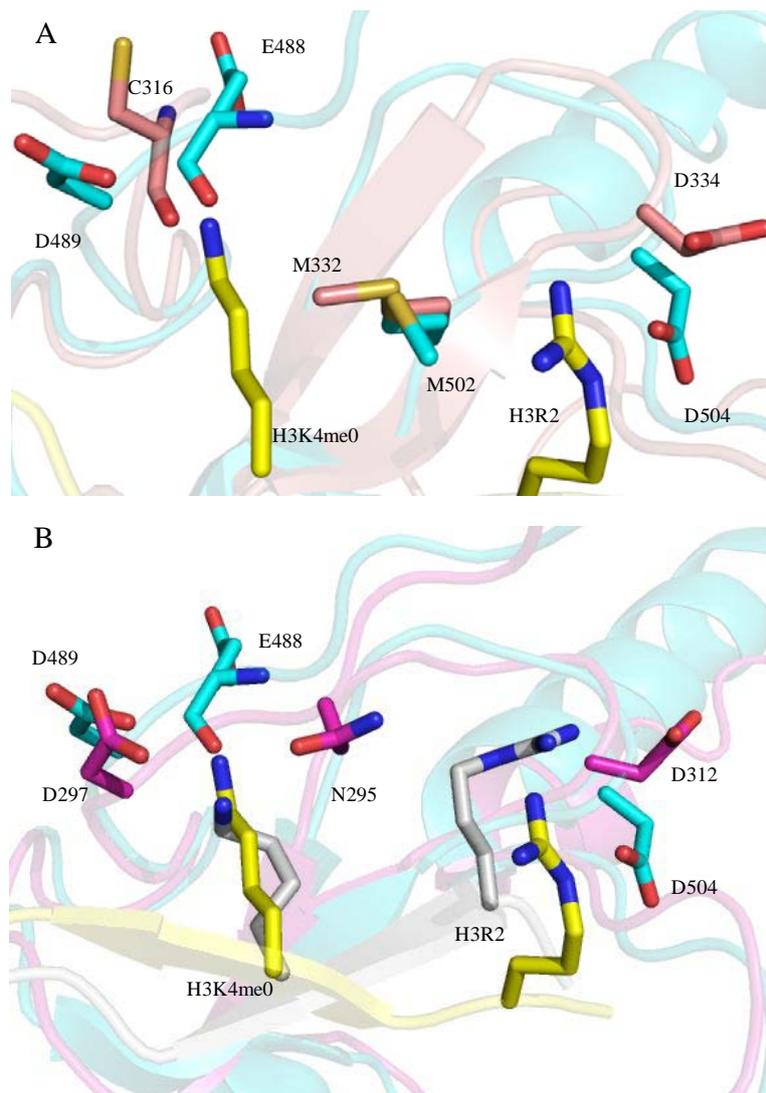


Figure 58: Comparaison structurale entre le modèle par homologie du PHD de hUHRF1 et des PHD fixant un peptide de l'histone H3K4me0. A- Comparaison entre hUHRF1 (en rose) et BHC80 (en cyan) en complexe avec le peptide de l'histone H3 (en jaune); PDB 2puy : l'acide aspartique qui fixe l'arginine R2 et la méthionine qui se positionne entre H3R2 et la H3K4 sont conservés dans hUHRF1. Le carbonyle de la cystéine C316 pointe vers H3K4. B- Comparaison entre le PHD de BHC80 (en cyan) et le PHD1 de AIRE (en magenta ; PDB 2ke1): Le peptide de l'histone H3 fixé par BHC80 (en jaune) adopte une conformation différente de celui fixé par le PHD1 de AIRE (en gris).

## 2.2. Comparaison avec les PHD fixant H3K4me3

Une superposition de structures entre le modèle par homologie du PHD de hUHRF1 et les structures des PHD d'ING2 et de BPTF en complexe avec l'histone H3K4me3 ont également été analysées (Figure 59). Le tryptophane conservé dans les PHD interagissant avec H3K4me3 (W238 dans ING2 et W32 dans BPTF) n'est pas présent dans hUHRF1. Dans le PHD de hUHRF1, nous trouvons une alanine à la place de ce tryptophane, et la méthionine M332 est le seul résidu hydrophobe qui pointe dans la même direction. Le carbonyle de la cystéine C316 de hUHRF1 vient pointer vers H3K4me3 alors qu'une tyrosine (Y215 dans ING2 et Y10 dans BPTF) qui participe à la cage aromatique est normalement présente à cette position. Le résidu hydrophobe (M226 dans ING2 et Y23 dans BPTF) qui participe à la cage hydrophobe est remplacé par un résidu polaire (Q330) dans hUHRF1. Ces superpositions structurales montrent l'absence d'une cage aromatique permettant de fixer H3K4me3 dans le PHD de hUHRF1.

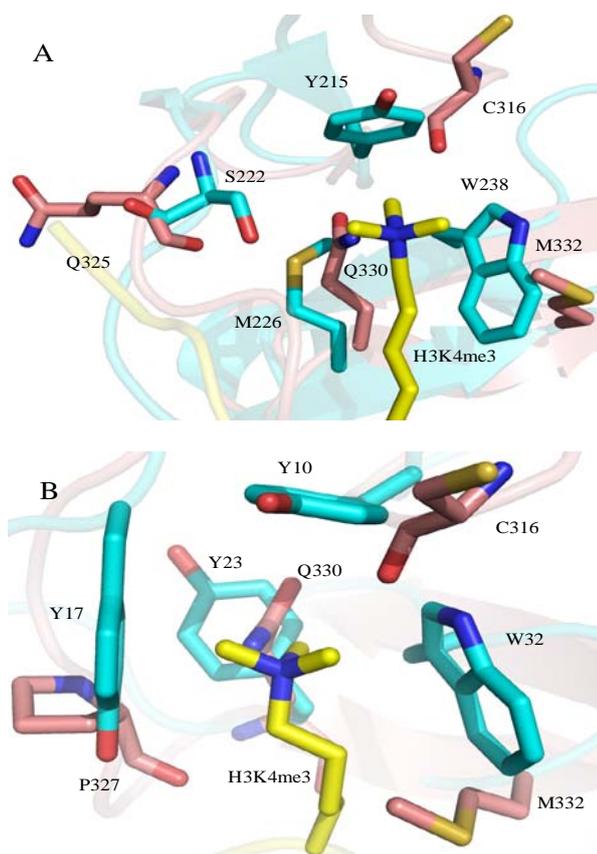


Figure 59: Comparaison structurale entre un modèle par homologie du PHD de hUHRF1 (en rose) et les PHD d'ING2 et de BPTF (en cyan). A- Comparaison avec ING2 (PDB :2g6q): les résidus formant la cage hydrophobe (Y215, M226 et W238) et les résidus en position équivalente (C312, Q330 et M332) dans hUHRF1 sont représentés. B- Comparaison avec BPTF (PDB : 2fuu): les résidus formant la cage aromatique (Y10, Y17, Y23 et W32) et les résidus en position équivalente (C312, P327, Q330 et M332) dans hUHRF1 sont représentés.

En conclusion, ces superpositions de structures suggèrent que le PHD de hUHRF1 appartiendrait plutôt à la classe de PHD reconnaissant H3K4me0 puisqu'aucun des résidus aromatiques conservés formant la cage hydrophobe de reconnaissance de H3K4me3 n'est présent dans hUHRF1. Mais, du fait de l'absence dans hUHRF1 de l'acide aspartique conservé et impliqué dans l'interaction avec H3K4me0 (D489 dans BHC80 et D297 dans AIRE), une étude structurale du PHD de hUHRF1 en complexe avec un peptide de l'histone H3 s'est avérée indispensable pour classifier ce domaine et établir les bases moléculaires de la reconnaissance de l'histone H3.

### 3. Clonage du domaine PHD

Différentes constructions du domaine PHD ont été clonées par la technologie Gateway®. Quatre limites amino-terminales ont été choisies: R296, D308, R314 et A317. Ces limites, à part R314, sont les mêmes que celles choisies pour PHD-SRA. La construction commençant à l'arginine R314 correspond au domaine PHD canonique et quatre résidus supplémentaires pour avoir la cystéine C316 qui serait potentiellement impliquée dans l'interaction avec H3K4me0. Toutes les constructions se terminent au résidu 367 du côté carboxy-terminal (Figure 60). Les différentes constructions sont présentées dans le Tableau 19.

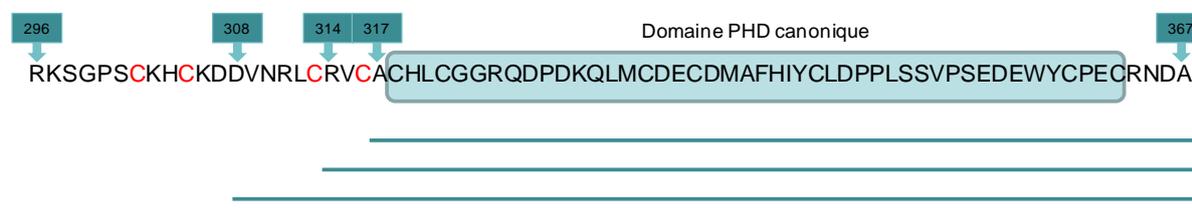


Figure 60: Les différentes constructions du domaine PHD. Le domaine PHD canonique est encadré en bleu et les cystéines formant le motif amino-terminal sont colorées en rouge. Les résidus qui ont été choisis comme limites pour les différentes constructions sont identifiés par des flèches et des numéros.

Domaine protéique	Etiquette amino-terminale	Numéro du vecteur d'expression	Concentration plasmidique (ng/μl)
<b>296-367</b>	(His) <sub>6</sub>	4242	84
	(His) <sub>6</sub> -GST	4243	120
<b>308-367</b>	(His) <sub>6</sub>	4244	166
	(His) <sub>6</sub> -GST	4245	134
<b>314-367</b>	(His) <sub>6</sub>	4246	139
	(His) <sub>6</sub> -GST	4247	280
<b>317-367</b>	(His) <sub>6</sub>	4248	53
	(His) <sub>6</sub> -GST	4249	92

Tableau 19: Les différentes constructions du domaine PHD de hUHRF1.

Durant ce travail de thèse, nous avons caractérisée deux constructions : PHD<sub>296-367</sub> qui contient le PHD canonique et le motif amino-terminal riche en cystéines et PHD<sub>314-367</sub> contenant uniquement le domaine PHD canonique. Les résultats obtenus avec ces deux constructions sont présentés dans les deux chapitres suivants.

## Chapitre V

# Caractérisation de la protéine PHD<sub>314-367</sub>

---

1. Expression, purification et caractérisation de PHD<sub>314-367</sub>
2. Caractérisation de l'interaction de PHD<sub>314-367</sub> avec l'histone H3
3. Cristallisation de la protéine PHD<sub>314-367</sub>
4. Etude cristallographique et analyse de structures
5. Conclusion

La caractérisation du domaine PHD de hUHRF1 a débuté par la purification de la protéine PHD<sub>296-367</sub>. L'absence de reproductibilité des premiers cristaux obtenus nous a conduit à entamer la caractérisation de la protéine correspondant au domaine PHD canonique, PHD<sub>314-367</sub>. Les résultats obtenus avec PHD<sub>314-367</sub> nous ont par la suite permis d'améliorer la qualité de la protéine PHD<sub>296-367</sub> qui a abouti à la résolution de la structure de cette protéine en complexe avec un peptide de l'histone H3.

## 1. Expression, purification et caractérisation de PHD<sub>314-367</sub>

La protéine PHD<sub>314-367</sub> avec l'étiquette (His)<sub>6</sub>-GST est surexprimée dans la souche *E. coli* Rosetta 2 après induction par 0.5 mM d'IPTG pendant 5h à 20°C en présence de ZnCl<sub>2</sub>.

Après une lyse du culot bactérien par sonication dans le tampon 50 mM Tris pH 7, 150 mM NaCl, 5 % glycérol et 10 mM imidazole, la fraction soluble est purifiée sur une première colonne d'affinité. La protéine PHD<sub>314-367</sub> est éluée à 220 mM imidazole (Figure 61A). Les fractions contenant la protéine sont dialysées pour éliminer l'imidazole avant la digestion par la protéase TEV qui permet d'enlever l'étiquette. La protéine digérée est ensuite séparée de son étiquette sur une deuxième colonne d'affinité. La majorité de la protéine sort dans les fractions non retenues et dans les fractions de lavage (fractions A5-A12 ; Figure 61B). Une très faible proportion est éluée à 40 mM imidazole en présence de plusieurs contaminants (fractions C5-C8 ; Figure 61B). Les fractions non retenues sont rassemblées, concentrées et injectées sur une colonne de gel filtration. Le pic majoritaire obtenu contient la protéine PHD<sub>314-367</sub> (Figure 61C). Ce protocole de purification permet d'obtenir environ 1 mg de protéine pure à partir d'un litre de culture.

Des mesures de spectrométrie de masse ont été réalisées sur cette protéine à l'aide du spectromètre MicrOTOF II (Bruker Daltonics) pour vérifier que la masse moléculaire de la protéine purifiée correspond à la masse théorique calculée à partir de la séquence et pour déterminer le nombre d'atomes zinc fixés. En conditions dénaturantes, avec une protéine PHD<sub>314-367</sub> diluée à 5 µM, l'analyse montre un pic majoritaire de masse moléculaire 6 491.7 Da, ce qui correspond à la protéine dépourvue de ses atomes zinc (masse théorique = 6 492.3 Da). La dénaturation étant insuffisante, un deuxième pic de masse moléculaire 6 554.6 Da est détecté correspondant à la protéine ayant conservée un seul zinc (masse théorique = 6 555.7 Da) (Figure 62A). En conditions natives, avec la protéine diluée à 20 nM dans 50 mM ammonium acétate (pH 6.8), l'analyse montre un pic majoritaire de 6 618.4 Da qui correspond à la protéine fixant deux zinc (masse théorique = 6 619.1 Da) (Figure 62B).

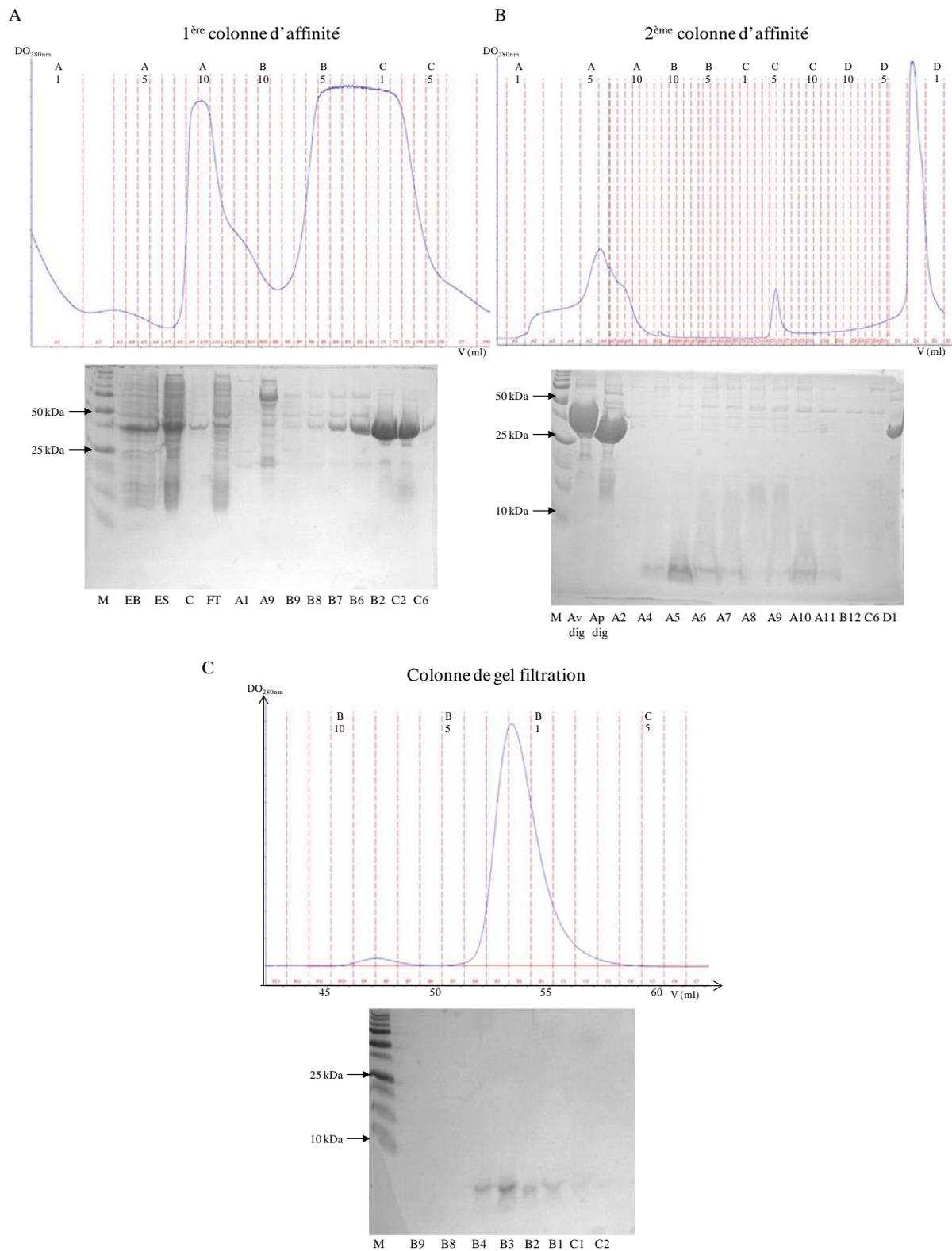
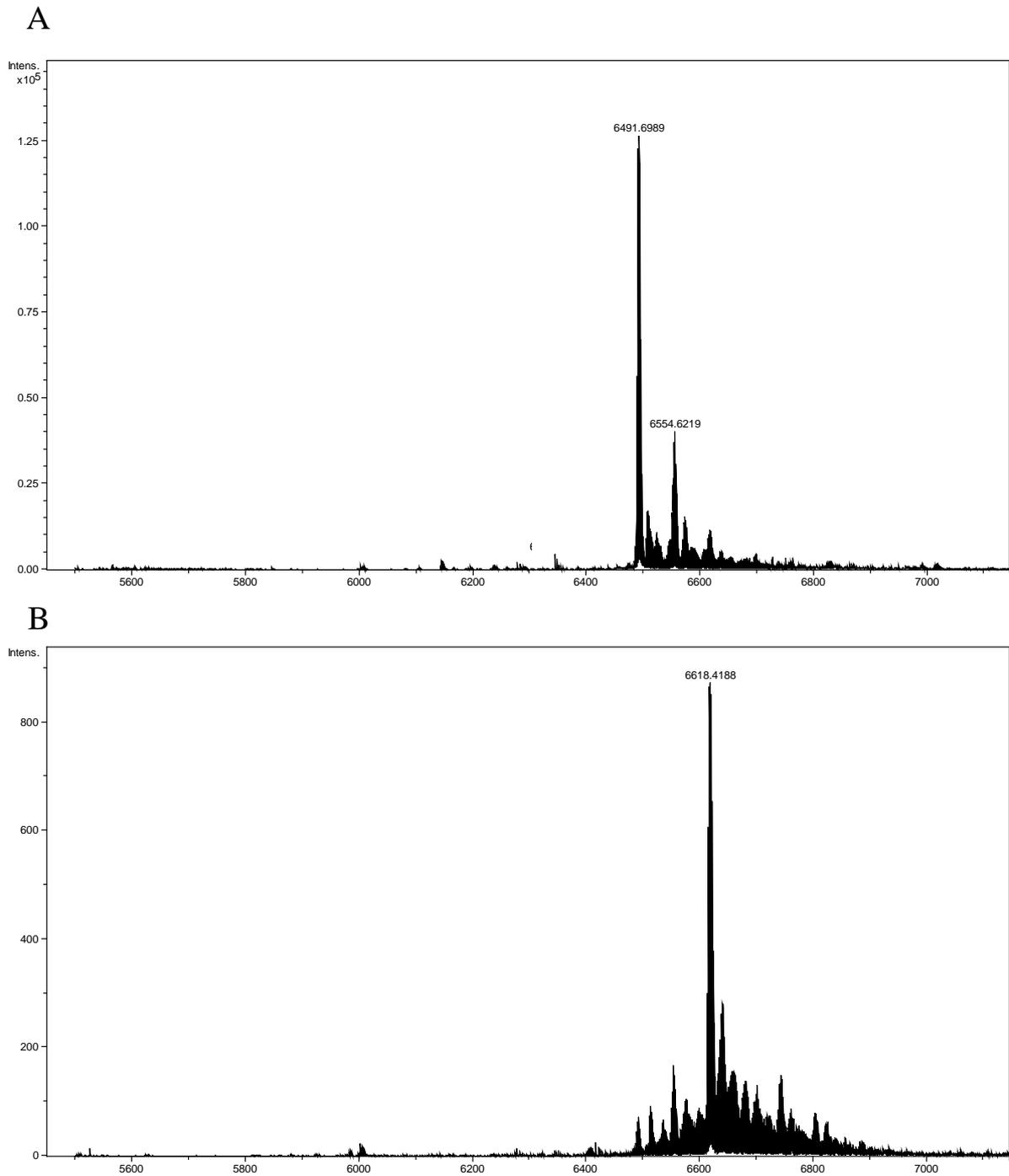


Figure 61: Les différentes étapes de purification de la protéine PHD<sub>314-367</sub>. Les chromatogrammes et les gels SDS correspondant sont présentés. M: marqueur de masse moléculaire, EB: extrait brut, ES: extrait soluble, FT: fraction non retenue, Av dig et Ap dig: aliquots de la protéine avant et après digestion par la TEV.



*Figure 62: Mesures de spectrométrie de masse effectuées sur la protéine PHD<sub>314-367</sub> pure. A- Mesures effectuées en conditions dénaturantes. B- Mesures effectuées en conditions natives.*

## 2. Caractérisation de l'interaction de PHD<sub>314-367</sub> avec l'histone H3

### 2.1. Caractérisation des complexes PHD<sub>314-367</sub>-histone H3 sur gels natifs

Afin de déterminer la spécificité de reconnaissance du domaine PHD de hUHRF1, l'interaction entre la protéine PHD<sub>314-367</sub> et différents peptides de l'histone H3 a été étudiée sur des gels natifs 6% (pH 9.4). Pour chaque réaction, 4µg de protéine ont été utilisées. Les peptides de l'histone H3 étudiés sont: H3 (1-20), H3K9me3 (1-20), H3 (1-8) et H3K9me3 (6-13). Différents ratios protéine-peptide ont été testés. Après incubation à 4°C, les mélanges protéine-peptide ainsi que la protéine seule, utilisée comme témoin, ont été déposés sur le gel natif. Avec les peptides longs (1-20), une bande correspondant au complexe est observée avec l'histone H3 non modifiée et H3K9me3 mais nous ne voyons pas de différences entre les complexes formés avec ces deux peptides. Avec les peptides courts contenant une seule des modifications H3K4me0 (1-8) ou H3K9me3 (6-13), aucune bande correspondant au complexe n'est observée (Figure 63). L'affinité du domaine PHD vis-à-vis de l'histone H3 est donc également dépendante de la taille des peptides utilisés. L'affinité augmente avec la taille des peptides indiquant que des résidus plus loin que la modification reconnue sont impliqués dans l'interaction.

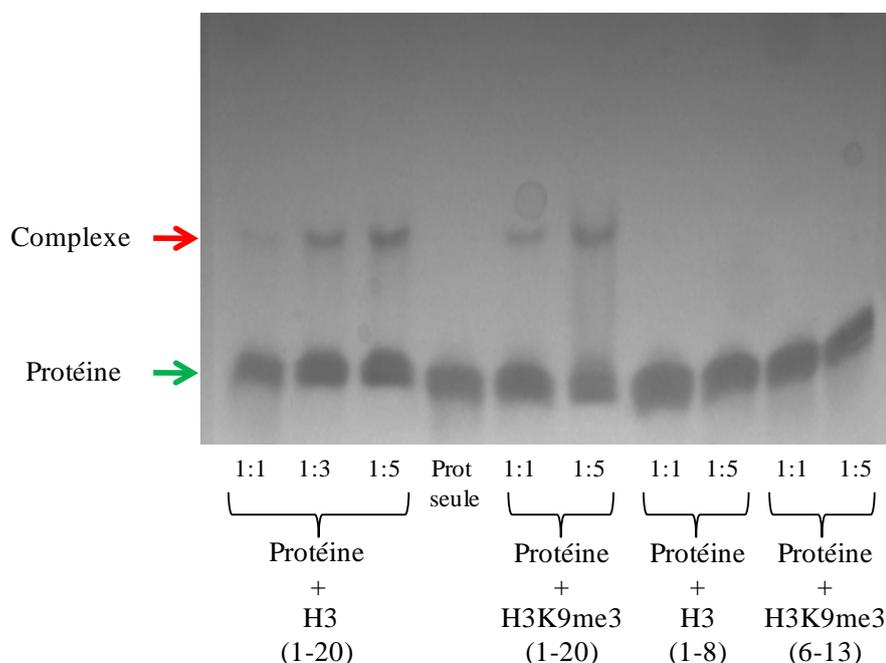


Figure 63: Interaction entre la protéine PHD<sub>314-367</sub> et des peptides de l'histone H3.

## 2.2. Caractérisation de l'interaction par ITC

Afin de valider les observations précédentes, nous avons essayé de déterminer des constantes de dissociation entre le PHD de hUHRF1 et différents peptides de l'histone H3 par ITC. La protéine pure à 72 $\mu$ M et des peptides à 600  $\mu$ M ont été utilisés. Les mesures ont été réalisées à 25°C sur l'appareil VP-ITC avec un délai entre les injections de 300 s. Malheureusement la protéine a précipité dans la cellule. Une deuxième tentative à 20°C a également échoué. Donc l'étude du domaine PHD<sub>314-367</sub> par ITC nécessite au préalable d'améliorer la stabilité de la protéine.

## 3. Cristallisation de la protéine PHD<sub>314-367</sub>

Des essais de cristallisation ont été réalisés sur la protéine PHD<sub>314-367</sub> seule ou en présence des peptides de l'histone H3. Des cristaux ont été obtenus dans différentes conditions de criblages généraux. Les cristaux ont été optimisés en variant la concentration de la protéine, les ratios protéine : peptide, la taille des peptides et la température de cristallisation. Des cristaux de taille suffisante ont été obtenus avec la protéine seule ou en présence de peptides (Figure 64).

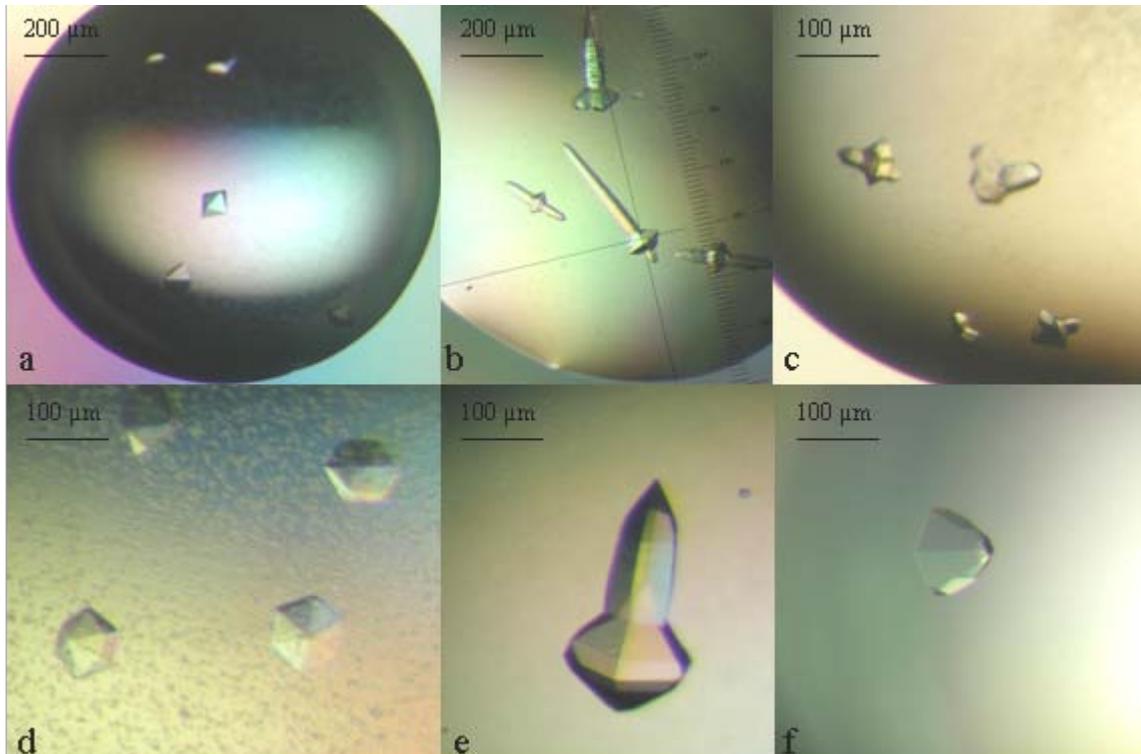


Figure 64: Cristaux obtenus avec la protéine PHD<sub>314-367</sub>. Les conditions de cristallisation correspondantes sont décrites dans le Tableau 20.

	Concentration de la protéine	Peptide H3 et ratio (protéine:peptide)	Composition du réservoir	Température	Méthode de cristallisation
a	770 $\mu$ M	H3K9me3 (1-20) Ratio (1:3)	0.1 M Hepes pH 7.5 1.3 M tri sodium citrate dihydrate 5% éthylène glycol	17 °C	Goutte suspendue
b	847 $\mu$ M	H3K9me3 (1-20) Ratio (1:10)	0.1 M Bis Tris pH 5.5 0.2 M MgCl <sub>2</sub> 25 % PEG 3350	4 °C	Goutte suspendue
c	770 $\mu$ M	H3K9me3 (6-13) Ratio (1:3)	0.1 M Bis Tris pH 5.5 0.2 M MgCl <sub>2</sub> 25 % PEG 3350 0.2 M 1.6 hexanediol	17°C	Goutte suspendue
d	770 $\mu$ M	H3 (1-8) Ratio (1:3)	0.1 M Hepes pH 7.5 1.3 M tri sodium citrate dihydrate 5% éthylène glycol	17°C	Goutte suspendue
e	847 $\mu$ M	H3 (1-8) Ratio (1:20)	0.1 M Bis Tris pH 5.5 0.2 M MgCl <sub>2</sub> 25 % PEG 3350 5% éthylène glycol	4°C	Goutte suspendue
f	847 $\mu$ M	-	2.1 M DL-Malic acid pH7	17 °C	Goutte assise

Tableau 20: Conditions de cristallisation de la protéine PHD<sub>314-367</sub>.

#### 4. Etude cristallographique et analyse de structures

Les propriétés de diffraction des cristaux ont été testées au laboratoire sur un diffractomètre Micromax 007 HF et au synchrotron SOLEIL sur la ligne PROXIMA 1. Certains cristaux ont été testés à température ambiante et les autres ont été préalablement congelés dans l'azote liquide, soit directement, soit en ajoutant des agents cryoprotectants tels l'éthylène glycol, le sucrose, le glycérol, la paraffine, le PEG 200 et le PEG 400. Ces cristaux appartiennent au groupe d'espace hexagonal P<sub>6</sub>22 avec les paramètres de maille:  $a = b = 51.0 \text{ \AA}$ ,  $c = 83.64 \text{ \AA}$ ,  $\alpha = \beta = 90^\circ$  et  $\gamma = 120^\circ$ . Différents jeux de données ont été collectés et traités (Tableau 21). Différentes structures ont été résolues, en utilisant le phasage SAD grâce au signal anomal des zinc, puis en utilisant la méthode de remplacement moléculaire avec la première structure résolue. Dans aucune structure, un peptide de l'histone H3 n'est visible. Les structures résolues sont similaires et présentent un domaine PHD canonique fixant deux atomes de zinc. La structure déterminée à partir du cristal « phd32 » est présentée dans la Figure 65. Les éléments de structures secondaires présents dans cette structure, déterminés par DSSP, sont deux brins  $\beta$  antiparallèles et deux courtes hélices  $\alpha$ . Les statistiques d'affinement de cette structure sont présentées dans le Tableau 23.

Cristal	Composition de la goutte	Condition de cristallisation	Agent cryoprotectant	Enregistrement	Résolution
Phd01	Protéine + H3K9me3 (1-20) ratio (1:3)	2,1 M DL Malic acid pH7 5% éthylène glycol	Direct	Synchrotron	1.47 Å
Phd02	Protéine + H3 (1-8) ratio (1:3)	1,8M NH4 citrate pH7 5% éthylène glycol	Direct	Synchrotron	1.60 Å
Phd11	Protéine + H3 (1-8) ratio (1:3)	0,1 M Bis-Tris pH 5,5 0,2 M MgCl <sub>2</sub> 25 % PEG 3350	10 % PEG 400	Synchrotron	1.75 Å
Phd12	Protéine + H3K9me3 (6-13) ratio (1:3)	0,1 M Bis-Tris pH 5,5 0,2 M MgCl <sub>2</sub> 25 % PEG 3350 0,2 M 1,6 hexanediol	10 % PEG 400	Synchrotron	1.80 Å
Phd21	Protéine + H3 (1-8) ratio (1:3)	0,1M Hepes pH 7,5 1,3 M tri Na citrate 5% éthylène glycol	Paraffine	Synchrotron	1.70 Å
Phd32	Protéine + H3 (1-8) ratio (1:3)	2,2 M Na Malonate pH 7 0,2 M 1,6 hexanediol	5 % glycérol	Synchrotron	1.43 Å
Crys1	Protéine seule	2.1 M DL-Malic acid pH7 0.1 M 1,6 hexanediol	-	CEBGS	3.40 Å
Crys2	Protéine + H3 (1-8) ratio (1:3)	2.1 M DL-Malic acid pH7 0.1 M 1,6 hexanediol	-	CEBGS	3.60 Å
Crys5	Protéine seule	2.1 M DL-Malic acid pH7 5 % éthylène glycol	-	CEBGS	3.60 Å
Crys6	Protéine + H3 (1-8) ratio (1:3)	.1 M Hepes pH 7.5 1.3 M tri sodium citrate dihydrate 5 % éthylène glycol	-	CEBGS	3.50 Å
Crys8	Protéine seule	1.6 M ammonium citrate dibasique pH 7 0.2 M 1,6 hexanediol	-	CEBGS	2.90 Å
B45A2G2	Protéine + H3 (1-8) ratio (1:20)	0.1 M Bis Tris pH 5.5 0.2 M MgCl <sub>2</sub> 25 % PEG 3350 3 % éthylène glycol	-	CEBGS	1.90 Å
B45A1G3	Protéine + H3K9me3 (1-20) ratio (1:10)	0.1 M Bis Tris pH 5.5 0.2 M MgCl <sub>2</sub> 25 % PEG 3350	5 % éthylène glycol	CEBGS	1.86 Å

Tableau 21: Les différents jeux de données enregistrés avec des cristaux de la protéine PHD<sub>314-367</sub> au synchrotron SOLEIL et au laboratoire (CEBGS). Les jeux de données sur les cristaux crys1 à crys8 ont été enregistrés à 17°C.

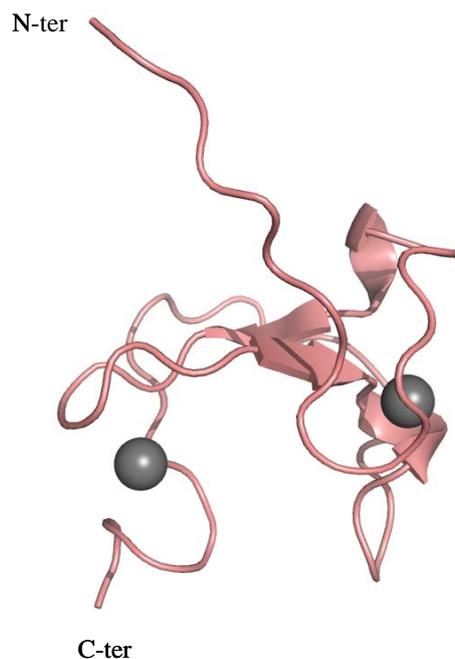
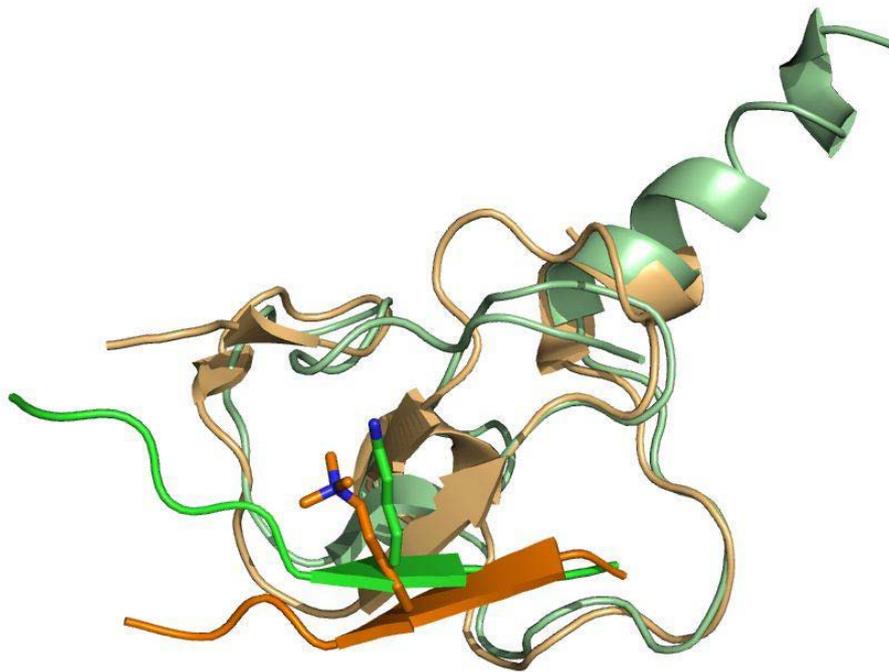


Figure 65: Structure du domaine PHD<sub>314-367</sub> de hUHRF1 (phd32). Les atomes de zinc sont représentés par des sphères grises.

phd32	
Beamline	Proxima-1 SOLEIL
Space group	P6522
Unit cell parameters (Å)	a = b = 51.00 c = 83.64
Resolution (Å)	24.4-1.43 (1.45-1.43)
No. of unique reflections	20 810
Redundancy	3.7 (1.6)
Completeness (%)	92.2 (54.7)
average I/σ(I)	19.3 (2.1)
R <sub>merge</sub>	0.089 (0.482)
mean B, Wilson plot (Å <sup>2</sup> )	24.80
No. of molecules in ASU	1
Refinement statistics	
R(%)	0.208
R <sub>free</sub> (%)	0.215
No. of protein residues/atoms	55/466
No. of nonprotein atoms	50
mean B, protein (Å <sup>2</sup> )	32
mean B, nonprotein (Å <sup>2</sup> )	42
Geometry	
R.m.s.d, bonds (Å)	0.006
R.m.s.d, angles (°)	1.089
Residues in	
Most favoured regions (%)	97.62
Additionally allowed regions (%)	2.38
Disallowed regions (%)	0

Tableau 23: Les statistiques de traitement et d'affinement du jeu de données phd32.

Une comparaison structurale entre le PHD d'ING2 interagissant avec H3K4me3 et celui de BHC80 interagissant avec H3K4me0, montre que la surface d'interaction avec le peptide est la même dans les deux protéines (Figure 66). La superposition structurale du domaine PHD<sub>314-367</sub> de hUHRF1 avec ces PHD nous a permis de positionner le peptide d'histone à la surface du PHD de hUHRF1 (Figure 67A). Dans ce modèle d'un complexe entre PHD<sub>314-367</sub> et un peptide portant H3K4me0, des résidus de la région amino-terminale viennent encombrer le site potentiel d'interaction avec l'histone. La chaîne latérale de la cystéine C316 entre en contact avec la chaîne latérale de la lysine K4 de l'histone. Les chaînes latérales de l'arginine R314 et de la méthionine initiale, issue du clonage (séquence GHM), encombrant également le site potentiel de fixation de l'histone H3. Ceci suggère que dans le cristal, le PHD<sub>314-367</sub> de hUHRF1 adopte une conformation autoinhibée qui empêche l'interaction avec le peptide de l'histone H3 (Figure 67B). Cette conformation explique l'absence de peptides dans toutes les structures de PHD<sub>314-367</sub> que nous avons résolues malgré la variation des conditions de cristallisation, de la taille et des concentrations des peptides utilisés.



*Figure 66: Comparaison entre les structures des domaines PHD de BHC80 et d'ING2 en complexe avec des peptides de l'histone H3. Le domaine PHD de BHC80 et l'histone H3K4me0 sont représentés respectivement en vert clair et vert. Le domaine PHD d'ING2 et l'histone H3K4me3 sont représentés respectivement en orange clair et orange. Dans ces deux structures de complexes, la région d'interaction se situe au même endroit sur le domaine PHD.*

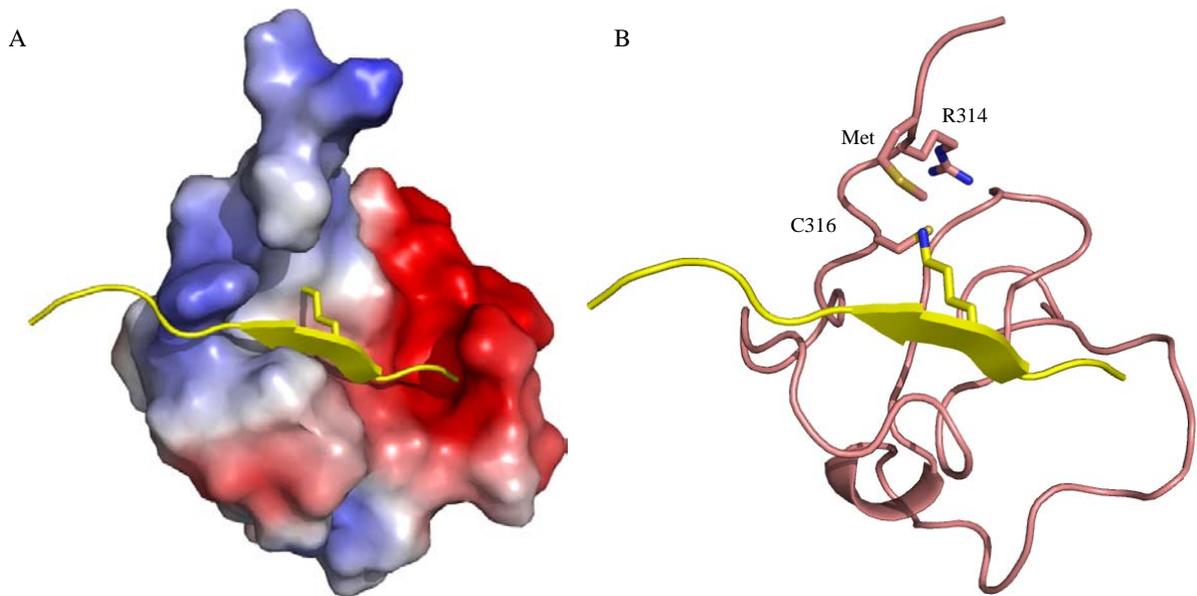
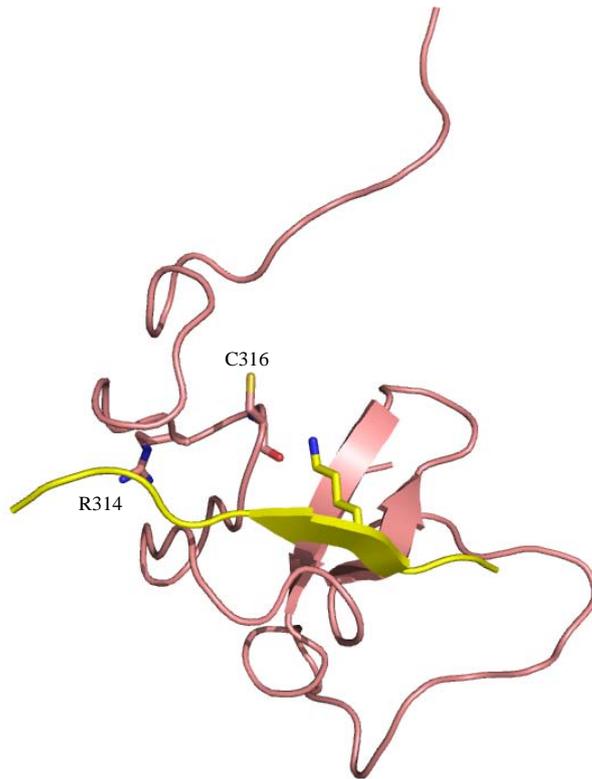


Figure 67: Modèle du PHD<sub>314-367</sub> de hUHRF1 en complexe avec un peptide de l'histone H3. A- Représentation de la surface électrostatique du PHD de hUHRF1 en complexe avec l'histone H3K4me0. B- Représentation du modèle d'interaction entre le PHD (en rose) et l'histone H3K4me0 (en jaune). Les résidus C316, R314 et Met qui encombrent le site de fixation de l'histone sont représentés.

Pour étudier si la présence du motif amino-terminal riche en cystéines pourrait empêcher cette conformation d'autoinhibition et donc favoriser l'interaction avec l'histone, nous avons utilisé le modèle par homologie du PHD commençant au résidu R296 pour faire la superposition avec l'histone H3. En présence du troisième doigt de zinc, la chaîne latérale de la cystéine C316 est impliquée dans la fixation du zinc, et c'est son carbonyle qui pointe vers la lysine non modifiée, ce qui pourrait favoriser une interaction électrostatique avec H3K4me0 (Figure 68).

## 5. Conclusion

Nous avons résolue la structure du domaine PHD canonique de hUHRF1. L'analyse de la structure montre que la construction PHD<sub>314-367</sub> n'est pas propice pour résoudre la structure du complexe PHD-histone H3 du fait d'une conformation « autoinhibée » de la protéine. Par contre la construction PHD<sub>296-367</sub> apparaît plus appropriée pour l'étude du complexe puisque la partie amino-terminale de la protéine adopte une conformation différente en présence du troisième doigt de zinc



*Figure 68: Modélisation de l'interaction entre le modèle par homologie du PHD de hUHRF1 et l'histone H3K4me0. La présence du motif amino-terminal riche en cystéines semble favorable à une interaction électrostatique entre le carbonyle de la cystéine C316 et la lysine K4me0.*

## Chapitre VI

# Caractérisation de la protéine PHD<sub>296-367</sub>

---

1. Expression et purification de PHD<sub>296-367</sub>
2. Caractérisation de l'interaction de PHD<sub>296-367</sub> avec l'histone H3
3. Cristallisation de la protéine PHD<sub>296-367</sub>
4. Etudes cristallographiques et analyse de la structure
5. Conclusion

Le rôle potentiel du motif riche en cystéines, en amont du domaine PHD, pour l'interaction avec l'histone H3, nous a conduit à caractériser la construction PHD<sub>296-367</sub>.

## 1. Expression et purification de PHD<sub>296-367</sub>

La protéine PHD<sub>296-367</sub> avec l'étiquette (His)<sub>6</sub>-GST a été surexprimée dans la souche *E. coli* Rosetta 2 après induction par 0.5 mM d'IPTG pendant 5h à 20°C en présence de ZnCl<sub>2</sub>.

Le premier protocole de purification utilisé est le même que celui de PHD<sub>314-367</sub>. Après lyse du culot bactérien par sonication, la fraction soluble est déposée sur la première colonne d'affinité et la protéine est éluée à 240 mM imidazole (Figure 69A). Après digestion par la TEV et une nouvelle étape de chromatographie d'affinité, la protéine PHD<sub>296-367</sub> est éluée majoritairement dans les fractions non retenues et dans les fractions de lavage (lot 1). Une faible proportion de la protéine est éluée à 50 mM imidazole en présence de protéines contaminantes (Figure 69B). Le lot 1 déposé sur une colonne de gel filtration conduit à un pic unique pour un volume d'élution de 52 ml (Figure 69C). De manière inattendue, la protéine pure migre sous forme de deux bandes diffuses sur gel SDS. La proportion de ces deux bandes est variable d'une purification à l'autre. Une mesure de spectrométrie de masse en conditions dénaturantes donne un pic majoritaire de masse moléculaire 8 248.5 Da, ce qui est très proche de la masse attendue pour une protéine dépourvue de ses atomes de zinc (masse théorique = 8 252.3 Da) mais avec toutefois une différence inexplicée de 4 Da. Deux autres pics, d'intensité plus faible, ont été observés, avec de masses moléculaires de 8 313.5 Da et 8 375.4 Da correspondant à la protéine fixant respectivement un et deux zinc suggérant une dénaturation incomplète (Figure 70A). En conditions natives, le spectre de masse obtenu présente plusieurs pics attribués à des adduits de sels sur la protéine. Deux pics majoritaires sont observés: le premier de masse moléculaire 8 375.3 Da correspond au domaine PHD<sub>296-367</sub> fixant deux atomes de zinc et le deuxième de masse moléculaire 8 442.3 Da correspond à la protéine fixant 3 atomes de zinc (Figure 70B). D'après ces résultats, l'échantillon de protéine PHD<sub>296-367</sub> apparaît hétérogène en terme de nombre de zinc fixés sur la protéine. Il est probable que le zinc fixé sur le motif amino-terminal soit le plus labile. Afin d'homogénéiser notre échantillon, 25 µM de ZnCl<sub>2</sub> ont été ajoutés à tous les tampons de purification. Notre hypothèse étant que les espèces protéines portant deux ou trois zinc devaient avoir des charges différentes, une étape finale de chromatographie échangeuse d'anions a été mise en œuvre. Le pic majoritaire, qui s'élue à 210 mM NaCl, migre sous forme d'une seule bande en gel SDS et en gel natif (Figure 71). Cette protéine pure et homogène a été utilisée pour caractériser les spécificités de reconnaissance de l'histone H3.

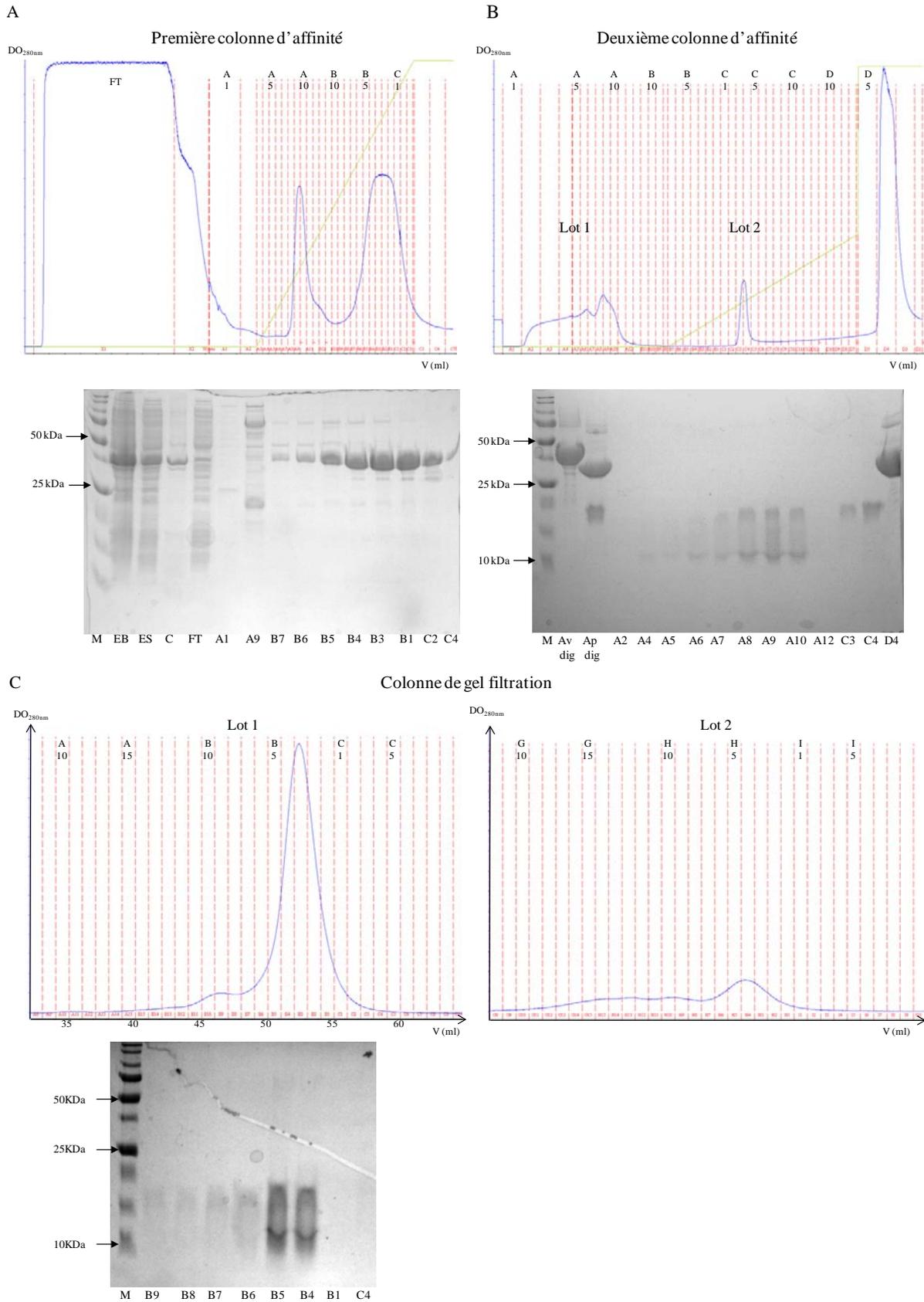


Figure 69: Protocole initial de purification de la protéine PHD<sub>296-367</sub>. Les chromatogrammes et les gels SDS correspondants sont présentés. M: marqueur de masses moléculaires, EB: extrait brut, ES: extrait soluble, C: culot, FT: fraction non retenue, Av dig et Ap dig: aliquots de la protéine avant et après digestion par la TEV.

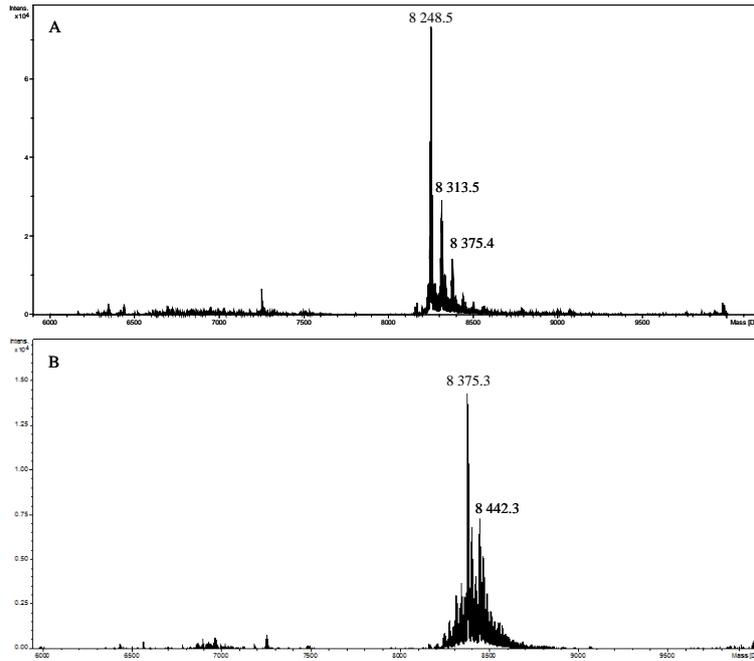


Figure 70: Mesures de spectrométrie de masse effectuées sur la protéine PHD<sub>296-367</sub> pure. A- Mesures effectuées en conditions dénaturantes. B- Mesures effectuées en conditions natives.

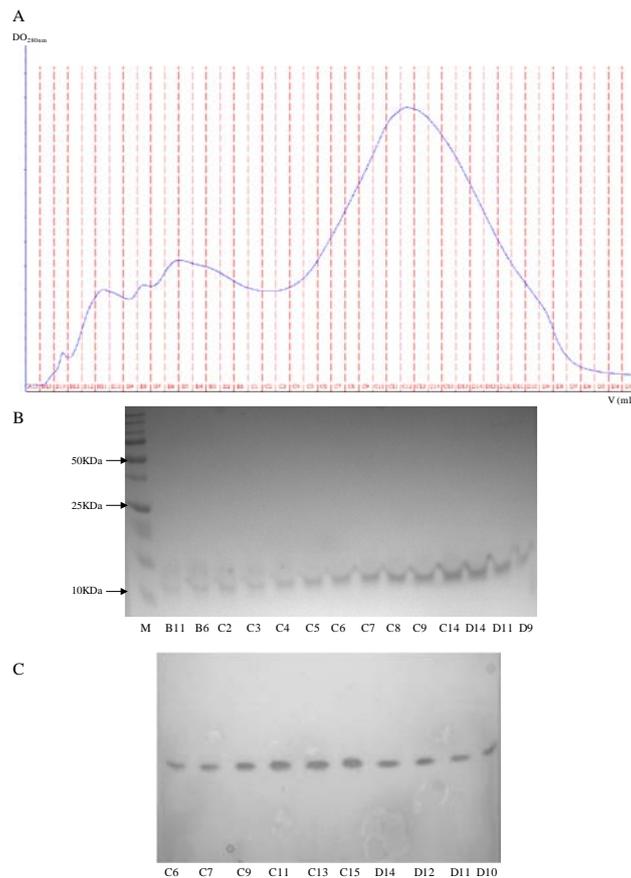


Figure 71: Chromatographie d'échange d'anions de la protéine PHD<sub>296-367</sub>. A- Chromatogramme montrant l'espèce majoritaire éluée dans les fractions C5 à D9. B- Analyse des fractions sur gel SDS 15%. C- Analyse des fractions sur gel natif 6 % (pH 9.4). La protéine PHD<sub>296-367</sub> retrouvée dans le pic majoritaire migre sous d'une seule bande sur gel SDS ou natif.

## 2. Caractérisation de l'interaction de PHD<sub>296-367</sub> avec l'histone H3

La caractérisation de l'interaction entre PHD<sub>296-367</sub> et différents peptides de l'histone H3 a été réalisée par électrophorèse en conditions natives et par calorimétrie à titration isotherme.

### 2.1. Caractérisation de l'interaction sur gels natifs

Les différents peptides qui ont été testés sont présentés dans le Tableau 24. Une première expérience a été réalisée en utilisant 4 µg de protéine et un rapport molaire protéine:peptide de 1:5. Après une incubation de 2 h dans la glace, les mélanges protéine-peptide ainsi que la protéine seule ont été déposés sur gels natifs à pH 9.4 (Figure 72). Une bande correspondant aux complexes protéine-peptide a été observée en présence des peptides portant les résidus 1 à 20 et quelque soit l'état de méthylation de la lysine K4 ou de la lysine K9.

Histone H3	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25																		
	M	A	R	T	K	Q	T	A	R	K	S	T	G	G	K	A	P	R	K	Q	L	A	T	K	A	A																		
H3S10A (0-19)	A																																											
H3K9me3 (6-17C)										Me3																C																		
H3S10ph (7-20C)										ph																								C										
H3 (3-25)																																												
H3 (1-20)																																												
H3K9me3 (1-20)										Me3																																		
H3K9me2 (1-20)										Me2																																		
H3K9me1 (1-20)										Me1																																		
H3K4me3 (1-20)					Me3																																							

Tableau 24: Les différents peptides de l'histone H3 utilisés.

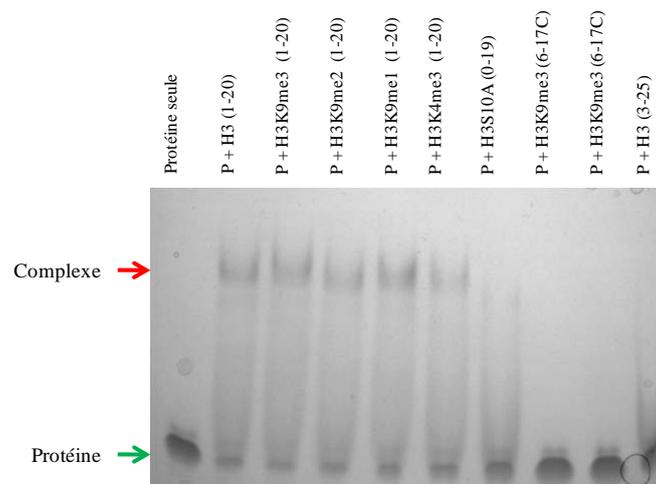
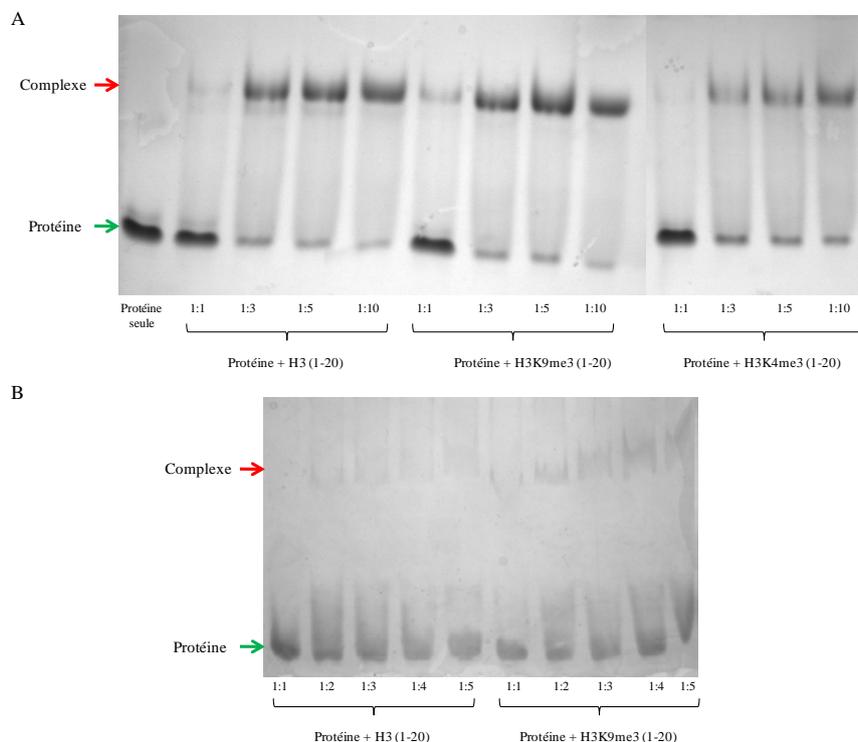


Figure 72: La caractérisation sur gel natif de l'interaction entre PHD<sub>296-367</sub> et différents peptides de l'histone H3. Les bandes correspondant à la protéine seule et au complexe protéine-peptide sont identifiées par des flèches.

Afin de déterminer la spécificité de reconnaissance du domaine PHD, des rapports protéine: peptide de 1:1, 1:3, 1:5 et 1:10 ont été testés avec des peptides (1-20) de H3 non modifiées ou portant les modifications K4me3 ou K9me3. Une bande correspondant aux complexes protéine-peptide a été observée avec les trois peptides testés, même lorsque le rapport molaire était de 1:1. Bien que dans ces conditions expérimentales l'interaction avec le peptide H3K4me3 semble être la plus faible, les observations ne permettent pas d'affirmer l'existence de différences significatives entre les différents peptides utilisés (Figure 73A). La concentration de protéine étant de 88  $\mu$ M et n'ayant aucune idée des Kd des différentes interactions, nous avons diminué la concentration protéique d'un facteur 10 pour essayer de mettre en évidence des différences d'affinité. Dans ces nouvelles conditions, nous avons testé les rapports 1:1, 1:2, 1:3, 1:4 et 1:5 avec les peptides H3 (1-20) et H3K9me3 (1-20) et nous avons utilisé la coloration à l'argent pour visualiser les protéines (figure 73B) (Helmut et al., 1987). Les bandes de complexes formées entre le PHD et le peptide H3K9me3 paraissent plus intenses que celles correspondant au complexe formé avec le peptide non modifié. Pour valider ces résultats et déterminer des constantes de dissociation pour chacun de ces complexes, nous avons choisi de réaliser des expériences de calorimétrie à titration isotherme.



**Figure 73:** Caractérisation de l'interaction entre PHD et différents peptides de l'histone H3 en variant la concentration en protéine et les rapports protéine: peptide. A- Dans les mélanges, la concentration en protéine est de 88  $\mu$ M et les rapports protéine : peptide sont indiqués. Le gel est coloré au bleu de Coomassie. B- Dans les mélanges, la concentration en protéine est de 8.8  $\mu$ M et le gel est coloré au nitrate d'argent.

## 2.2. Caractérisation de l'interaction par ITC

Plusieurs séries de mesures ont été réalisées pour étudier l'interaction entre la protéine PHD<sub>296-367</sub> et différents peptides de l'histone H3. Les premières mesures ont été effectuées sur l'appareil VP-ITC. Afin de réduire la quantité de matériel, et le temps nécessaire pour chaque mesure, l'appareil ITC-200 a été préféré par la suite.

### 2.2.1. Caractérisation de la protéine PHD<sub>296-367</sub> issue du protocole initial

La protéine à une concentration de 65.4  $\mu\text{M}$  et les peptides à une concentration de 600  $\mu\text{M}$  ont été utilisés. Les mesures ont été réalisées sur l'appareil VP-ITC à 25°C avec un délai de 300 s entre chaque injection. Les résultats obtenus sont représentés sur la Figure 74 et dans le Tableau 25. La protéine PHD<sub>296-367</sub> de hUHRF1 interagit de la même façon avec un peptide (1-20) d'histone H3 méthylé ou non méthylé à la lysine 9. Par contre la méthylation de la lysine 4 diminue sensiblement l'affinité du PHD vis-à-vis de l'histone H3. L'utilisation de peptides plus courts incluant soit la lysine K4 (peptide 1-8), soit la lysine K9 (peptide 6-17C) montre que le PHD interagit seulement avec un peptide H3 contenant la lysine 4 non modifiée. Il faut noter que l'affinité est dépendante de la taille des peptides utilisés, l'affinité étant plus faible quand le peptide non modifié est plus court. Cette observation suggère que des résidus éloignés du site de modification reconnu par le PHD contribuent à l'interaction, et ces premiers résultats nous conduisent à émettre l'hypothèse que c'est la lysine 4 non modifiée, et non la lysine 9 triméthylée, qui pourrait déterminer la spécificité de reconnaissance de l'histone H3 avec le domaine PHD de hUHRF1.

Une série de mesures a été réalisée sur l'ITC<sub>200</sub> en utilisant la protéine PHD<sub>296-367</sub> à une concentration de 154  $\mu\text{M}$  et les peptides d'histone H3 à une concentration de 1.5 mM. Les mesures ont été réalisées à 20°C avec un délai de 240 s entre chaque injection. Les résultats obtenus sont présentés dans la Figure 75 et le Tableau 26. Les résultats obtenus confirment l'hypothèse de la lysine 4 non modifiée comme déterminant de la reconnaissance de l'histone H3. En effet ces mesures ne montrent pas une différence d'affinité significative entre un peptide portant une lysine 9 triméthylée ou non méthylée. Par contre, une perte d'affinité significative est liée à la méthylation de la lysine 4. Les mesures montrent une variabilité de la stœchiométrie de 1 à 0.5, ce qui signifie qu'une molécule du peptide interagit avec une ou deux molécules de la protéine ou bien qu'une partie de la protéine (jusqu'à 50 %) perd ses propriétés de reconnaissance dans les conditions expérimentales de l'expérience ou encore qu'une erreur jusqu'à 50 % sur la concentration de la protéine ou du peptide a été faite. La concentration de la protéine a été mesurée plusieurs fois à 280 nm au spectrophotomètre

Nanodrop. Par contre nos peptides n'absorbent pas à 280 nm et leur concentration n'a donc pas pu être déterminée expérimentalement mais seulement calculée en fonction de la dilution à partir du peptide lyophilisé.

L'hétérogénéité de l'échantillon protéique peut également être responsable de la variation des paramètres expérimentaux rapportés dans les Tableaux 25 et 26. Pour obtenir des résultats de meilleure qualité en ITC, de nouvelles séries de mesures ont été réalisées, après que l'homogénéité de la protéine ait été améliorée par l'addition de  $ZnCl_2$  suivi de l'étape de chromatographie échangeuse d'anions.

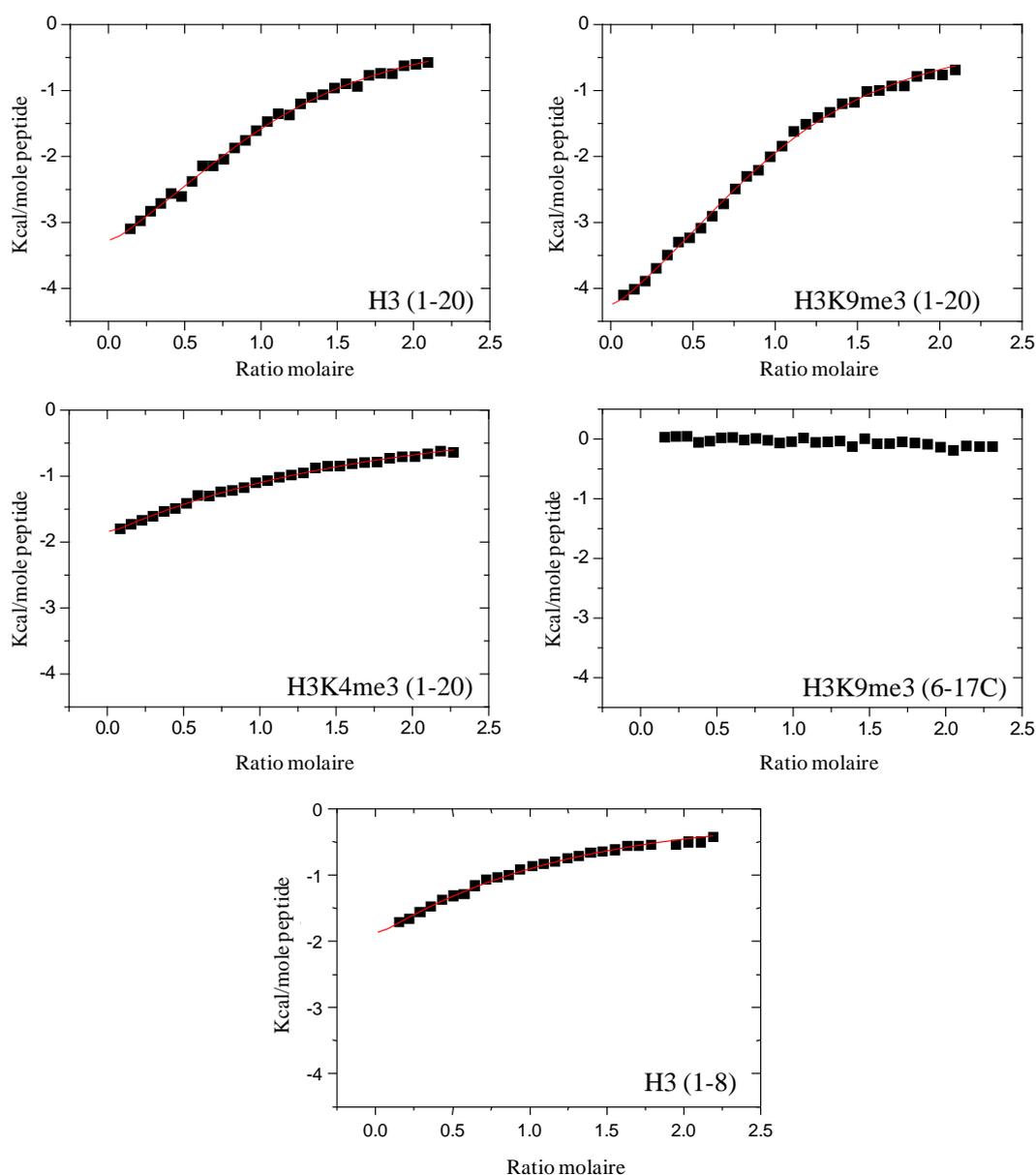


Figure 74: Expériences d'ITC réalisées sur VP-ITC avec la protéine  $PHD_{296-367}$  hétérogène (mélange de protéines fixant deux ou trois zinc) à  $65.4 \mu M$  et différents peptides de l'histone H3. Les peptides de l'histone H3 utilisés pour chaque mesure sont indiqués sur les courbes.

Protéine PHD <sub>296-367</sub> (65.4 μM)					
Peptide H3	H3 (1-20)	H3K9me3 (1-20)	H3K4me3 (1-20)	H3K9me3 (6-17C)	H3 (1-8)
Chi <sup>2</sup>	2252.14	2907.06	488.019	Pas d'interaction	992.552
N	1.025 ± 0.02676	0.952 ± 0.0149	0.963 ± 0.2138		0.8002 ± 0.1167
Kd	29.7 ± 2.4 μM	25.2 ± 1.32 μM	157 ± 24 μM		75.2 ± 10.6 μM

Tableau 25: Valeurs de  $\chi^2$ , stœchiométries (N) et constantes de dissociation (Kd) déterminées pour chacune des interactions entre PHD<sub>296-367</sub> et un peptide de l'histone H3.

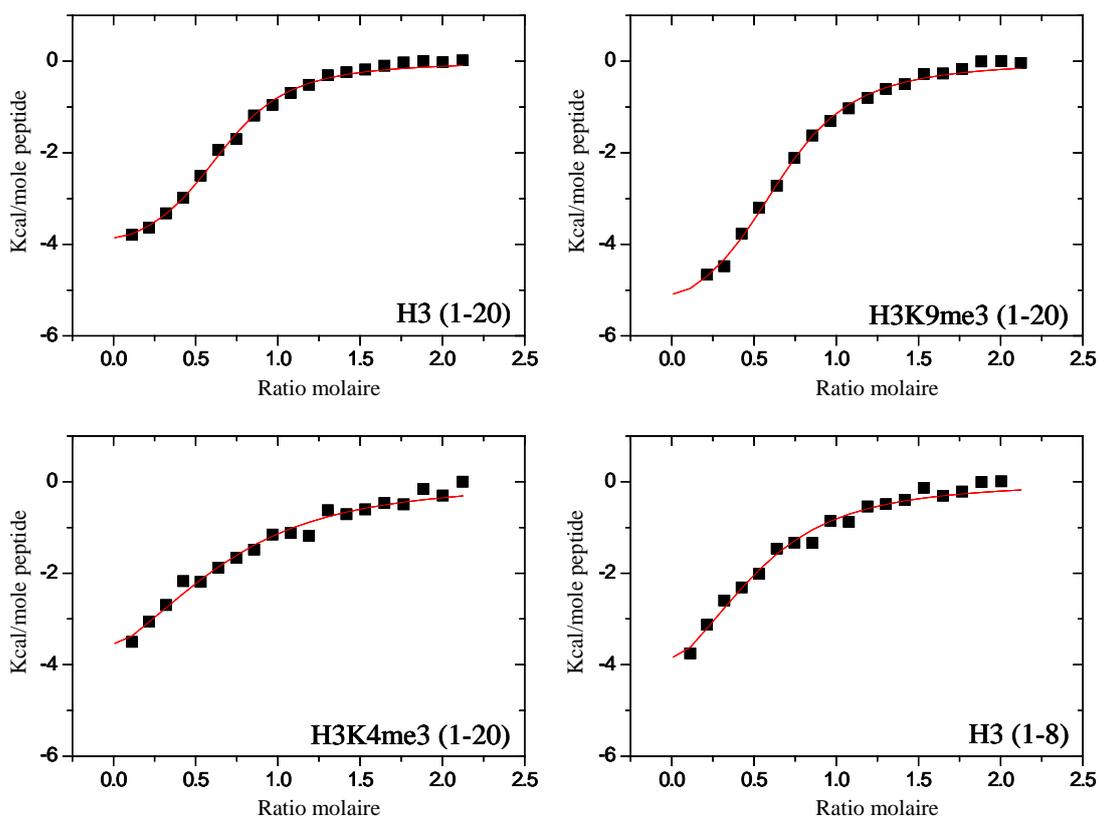


Figure 75: Expériences d'ITC réalisées sur ITC<sub>200</sub> avec la protéine PHD<sub>296-367</sub> hétérogène à 154 μM et différents peptides de l'histone H3.

Protéine PHD <sub>296-367</sub> (154 μM)					
Peptide H3	H3 (1-20)	H3K9me3 (1-20)	H3K4me3 (1-20)	H3K9me3 (6-13)	H3 (1-8)
Chi <sup>2</sup>	7547.97	11342.0	24838.9	Pas d'interaction	22523.5
N	0.6518 ± 0.01394	0.6559 ± 0.01853	0.6473 ± 0.0803		0.4998 ± 0.0533
Kd	11.4 ± 1.2 μM	14.4 ± 1.6 μM	52.4 ± 11.4 μM		34.8 ± 6.9 μM

Tableau 26: Valeurs de  $\chi^2$ , stœchiométries (N) et constantes de dissociation (Kd) déterminées pour chacune des interactions entre PHD<sub>296-367</sub> et un peptide de l'histone H3.

## 2.2.2. Caractérisation de la protéine PHD<sub>296-367</sub> issue du protocole optimisé

Pour cette série de mesure, la protéine était à une concentration de 119  $\mu\text{M}$  et les peptides à une concentration de 3 mM. La température était de 20°C avec un temps de 160 s entre chaque injection. Des courbes de meilleure qualité que celles obtenues lors des expériences précédentes ont été enregistrées (Figure 76 et Tableau 27). Les résultats confirment la spécificité du domaine PHD de hUHRF1 vis-à-vis de la lysine 4 non modifiée. Les constantes de dissociation ont des valeurs comparables avec les mesures précédentes. Par contre, le problème de stœchiométrie persiste. Les autres domaines PHD décrit comme reconnaissant les histones contiennent un seul site de fixation du peptide.

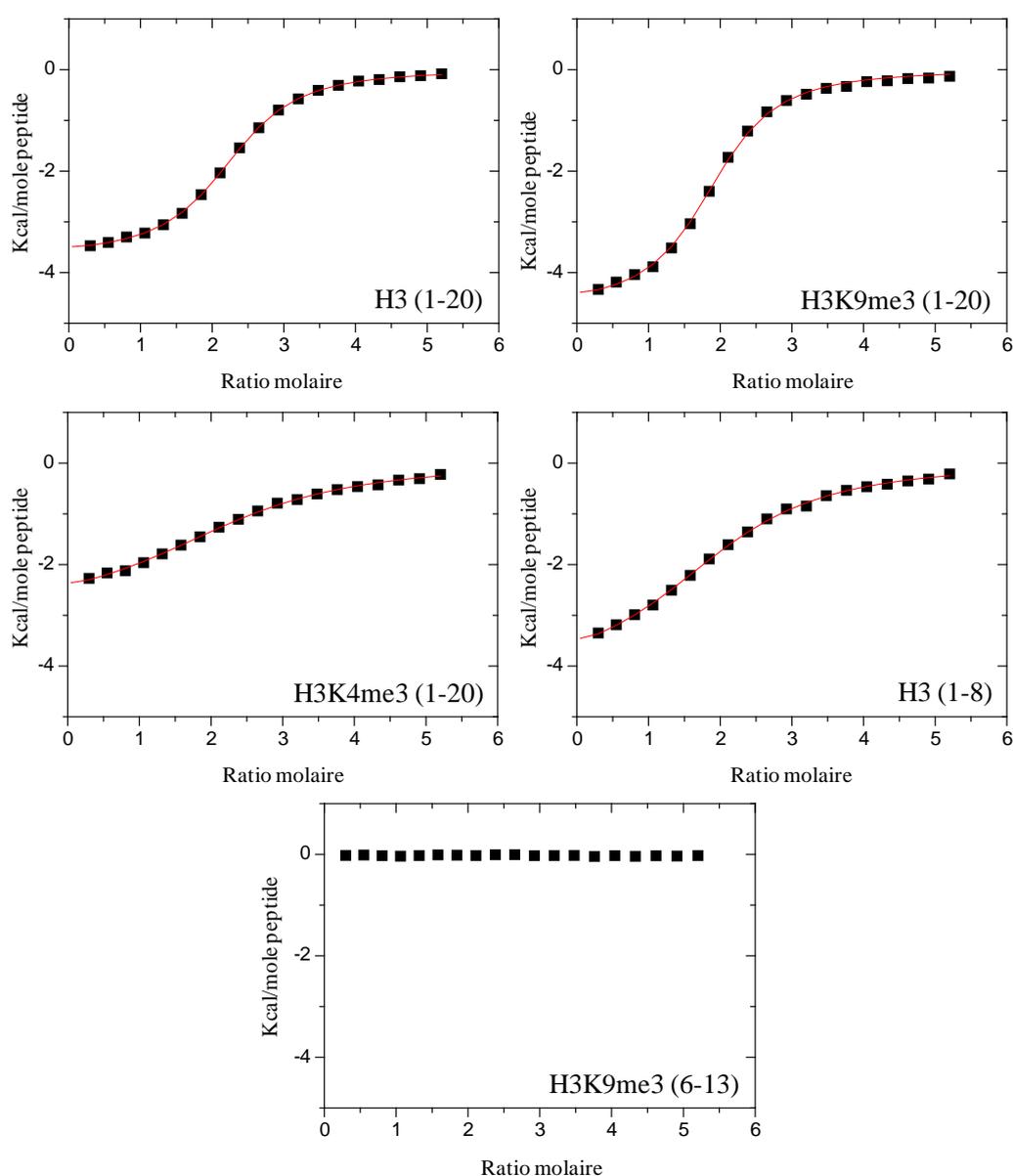


Figure 76: Courbes d'ITC sur l'appareil ITC<sub>200</sub> avec la protéine PHD<sub>296-367</sub> homogène (une seule espèce à 3 zinc) à 119  $\mu\text{M}$  et différents peptides de l'histone H3.

Protéine PHD <sub>296-367</sub> (119 $\mu$ M)					
Peptide H3	H3 (1-20)	H3K9me3 (1-20)	H3K4me3 (1-20)	H3K9me3 (6-13)	H3 (1-8)
Chi <sup>2</sup>	172.466	2074.53	312.094	Pas d'interaction	829.97
N	2.229 $\pm$ 0.00459	1.878 $\pm$ 0.01280	2.368 $\pm$ 0.02010		2.037 $\pm$ 0.01992
Kd	12.8 $\pm$ 0.2 $\mu$ M	14 $\pm$ 0.7 $\mu$ M	60.5 $\pm$ 2.4 $\mu$ M		51.3 $\pm$ 2 $\mu$ M

Tableau 27: Valeurs de  $\chi^2$ , stœchiométries (N) et constantes de dissociation (Kd) déterminées pour chacune des interactions entre PHD<sub>296-367</sub> fixant 3 zinc et un peptide de l'histone H3.

### 3. Cristallisation de la protéine PHD<sub>296-367</sub>

Les premiers essais de cristallisation ont été réalisés sur la protéine PHD<sub>296-367</sub> hétérogène. De cristaux sous forme d'aiguilles ont été obtenus dans différentes conditions. Le meilleur cristal en termes de taille et de diffraction est présenté dans la Figure 77a. La difficulté à reproduire ces cristaux est probablement due à l'hétérogénéité de l'échantillon protéique. Avec la protéine PHD<sub>296-367</sub> homogène fixant 3 atomes de zinc, des cristaux ont été facilement obtenus. Des essais de co-cristallisation en présence de peptides de l'histone H3 ont été menés afin de résoudre la structure du complexe. Des cristaux de morphologies différentes ont été obtenus dans différentes conditions de cristallisation (Figure 77b-f et Tableau 28).

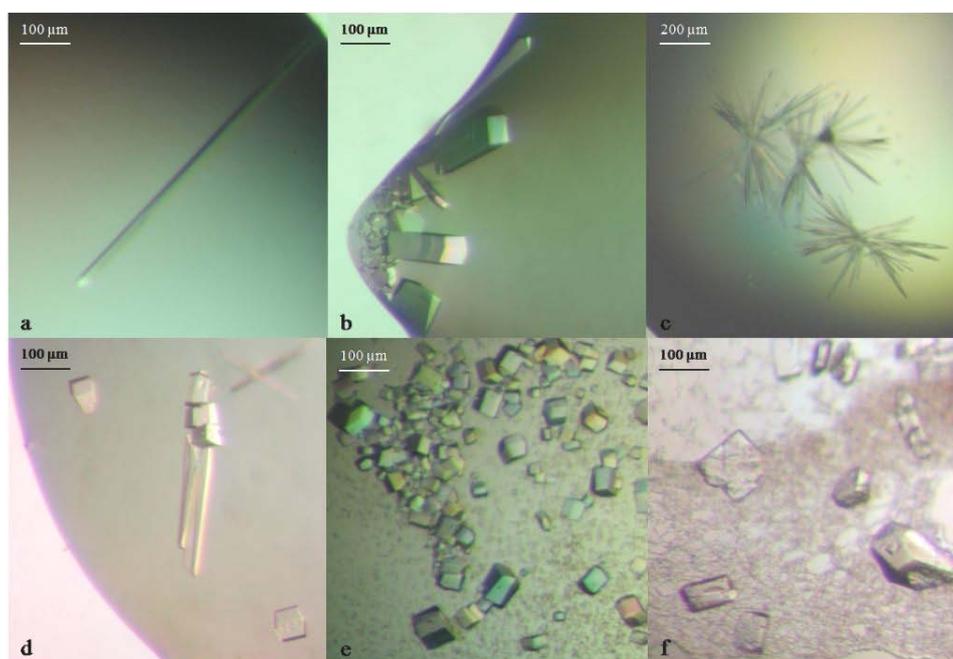


Figure 77: Cristaux obtenus avec la protéine PHD<sub>296-367</sub> seule (a) ou en présence de peptides de l'histone H3 (b-f). les cristaux ont été obtenus à 17°C par la méthode de diffusion en goutte suspendue. Seul le cristal présenté en (a) a été obtenu avec la protéine hétérogène. Les conditions de cristallisation correspondantes sont décrites dans le Tableau 29.

	Concentration de la protéine	Ratio protéine:peptide	Composition du réservoir
a	485 $\mu$ M	-	0.1 M Hepes pH 7 0.1 M MgCl <sub>2</sub> 18 % PEG 6000
b	606 $\mu$ M	1:3	0.1 M MES pH 7 0.2 M CaCl <sub>2</sub> 18% PEG 6K
c	606 $\mu$ M	1:3	0.1 M Tris pH 8.5 0.2 M MgCl <sub>2</sub> 28 % PEG 3350
d	606 $\mu$ M	1:10	0.1 M Tris pH 8.5 20% isopropanol 20 % PEG 6K
e	606 $\mu$ M	1:3	0.1 M MES pH 6.5 0.2 M CaCl <sub>2</sub> 20% PEG 6K
f	606 $\mu$ M	1:3	10% isopropanol 30% PEG 3350

Tableau 28: Conditions de cristallisation de la protéine PHD<sub>296-367</sub>. Pour les essais de cocrystallisation, seul le peptide de H3 non modifié 1-8 a été utilisé.

#### 4. Etudes cristallographiques et analyse de la structure

La diffraction des cristaux a été testée soit au laboratoire, soit sur la ligne PROXIMA1 du synchrotron SOLEIL. Avec la protéine hétérogène, la meilleure diffraction a été obtenue avec le cristal présenté dans la Figure 77a, congelé dans l'azote liquide en présence de 20 % glycérol. Ce cristal appartient au groupe d'espace tétragonal P422 et présente les paramètres de maille suivants:  $a = b = 90 \text{ \AA}$ ,  $c = 34 \text{ \AA}$  et  $\alpha = \beta = \gamma = 90^\circ$ . La limite de diffraction est de  $3.2 \text{ \AA}$  de résolution mais la présence de très nombreuses taches de diffraction allongées et dédoublées n'a pas rendu possible l'exploitation de ce jeu de données.

Les propriétés de diffraction des cristaux obtenus avec la protéine homogène ont été bien meilleures. Différents jeux de données ont pu être enregistrés et traités (Tableau 29). Différentes structures ont été résolues, en utilisant le phasage SAD grâce au signal anomal du zinc, puis en utilisant la méthode de remplacement moléculaire avec la première structure résolue. 4 sites de fixation de zinc ont été identifiés dans la protéine: deux dans le domaine PHD canonique, un dans le motif amino-terminal riche en cystéines et un quatrième zinc qui participe à l'empilement cristallin. Les différentes structures résolues sont en cours d'affinement. La structure résolue à partir du cristal « ph1 » est présentée dans la Figure 78 et les statistiques d'affinement actuels sont présentés dans le Tableau 30.

Nom du cristal	Condition de cristallisation	Agent cryoprotectant	Enregistrement des données	Résolution
RTB52B2G2	0.1 M MES pH6.5 0.2 M CaCl <sub>2</sub> 22% PEG 6K	-	CEBGS	2.4 Å
ph1	0.1 M Tris pH 8.5 0.2 M MgCl <sub>2</sub> 30 % PEG 4K	15 % MPD	SOLEIL	1.8 Å
ph3	0.1 M MES pH6.5 0.2 M CaCl <sub>2</sub> 22% PEG 6K	15 % isopropanol	SOLEIL	2.2 Å
ph4	0.1 M Tris pH 8.5 0.2 M MgCl <sub>2</sub> 30 % PEG 4K	35 % PEG 3350	SOLEIL	2.1 Å
ph9	0.1 M Tris pH 8.5 0.2 M MgCl <sub>2</sub> 30 % PEG 4K	15 % EG	SOLEIL	2 Å

Tableau 29: Les différents jeux de données enregistrés sur les cristaux de la protéine PHD<sub>296-367</sub> homogène à 3 zinc, en présence du peptide de l'histone H3 (1-8).

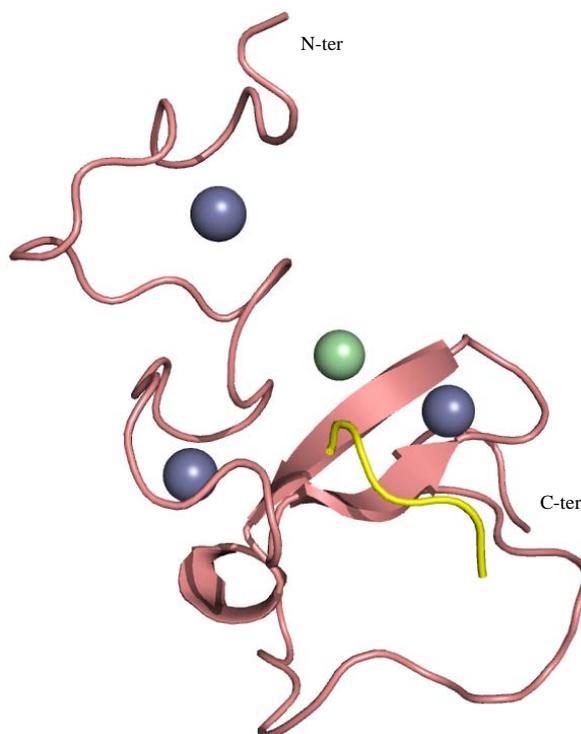


Figure 78: Structure du domaine PHD<sub>296-367</sub> (en rose) de hUHRF1 en complexe avec un peptide de l'histone H3 (en jaune). Les atomes de zinc sont représentés par des sphères grises (l'atome de zinc qui participe à l'empilement cristallin est représenté en vert). Le jeu de données ph1 a été utilisé pour résoudre cette structure.

ph1	
Beamline	proxima-1 SOLEIL
Space group	P43212
Unit cell parameters (Å)	a = b = 42.30 c = 182.70
Resolution (Å)	38.4-1.95 (2.0-1.95)
No. of unique reflections	22160
Redundancy	4.96 (4.7)
Completeness (%)	95.8 (83.9)
average I/σ(I)	21.46 (3.12)
R <sub>merge</sub>	0.048 (0.528)
No. of molecules in ASU	2
mean B, Wilson plot (Å <sup>2</sup> )	40.6
Refinement statistics	
R(%)	0.178
R <sub>free</sub> (%)	0.219
No. of protein residues/atoms	/1181
No. of nonprotein atoms	87
mean B, protein (Å <sup>2</sup> )	44.6
mean B, nonprotein (Å <sup>2</sup> )	47.7
Geometry	
R.m.s.d, bonds (Å)	0.007
R.m.s.d, angles (°)	1.095
Residues in	
Most favoured regions (%)	96.12
Additionally allowed regions (%)	3.88
Disallowed regions (%)	0

Tableau 30: Les statistiques de traitement et d'affinement du jeu de données ph1.

Dans cette structure, le domaine PHD canonique adopte une topologie entrecroisée pour la fixation des zinc et le motif amino-terminal riche en cystéines fixe un zinc supplémentaire. Les résidus 1 à 5 du peptide de l'histone H3K4me0 sont également visibles. Le domaine PHD interagit avec les quatre premiers résidus de l'histone H3. La lysine K4 non modifiée de H3, via son groupement ammonium, établit une liaison hydrogène avec le carbonyle de la cystéine C316 de hUHRF1. L'arginine R2 de H3 interagit via des liaisons hydrogènes avec les deux acides aspartiques D334 et D337 de hUHRF1. L'alanine A1 de H3 est tenue par des liaisons hydrogènes entre son groupe ammonium et les carbonyles de la proline P353 et de l'acide glutamique E355 du domaine PHD (Figure 79).

## 5. Conclusion

Dans ce chapitre, nous avons montré que le domaine PHD<sub>296-367</sub> de hUHRF1 reconnaît spécifiquement la lysine K4 non modifiée de l'histone H3. Des cristaux ont été obtenus avec

cette protéine et différentes structures du complexe PHD<sub>296-367</sub>-H3 ont été résolues et sont en cours d'affinement. L'interaction entre la protéine PHD<sub>296-367</sub> et le peptide de l'histone H3 est médiée par une série de liaisons hydrogènes impliquant l'alanine A1, l'arginine R2 et la lysine K4 de l'histone H3 et les acides aminés C316, D334, D337, P353 et E355 de hUHRF1. La lysine K4 non modifiée de l'histone interagit avec le carbonyle de la cystéine C316 de hUHRF1 ce qui explique la nécessité d'avoir le motif amino-terminal fixant un zinc pour obtenir la structure du complexe. Le domaine PHD de hUHRF1 appartient donc à la même famille que BHC80 et AIRE.

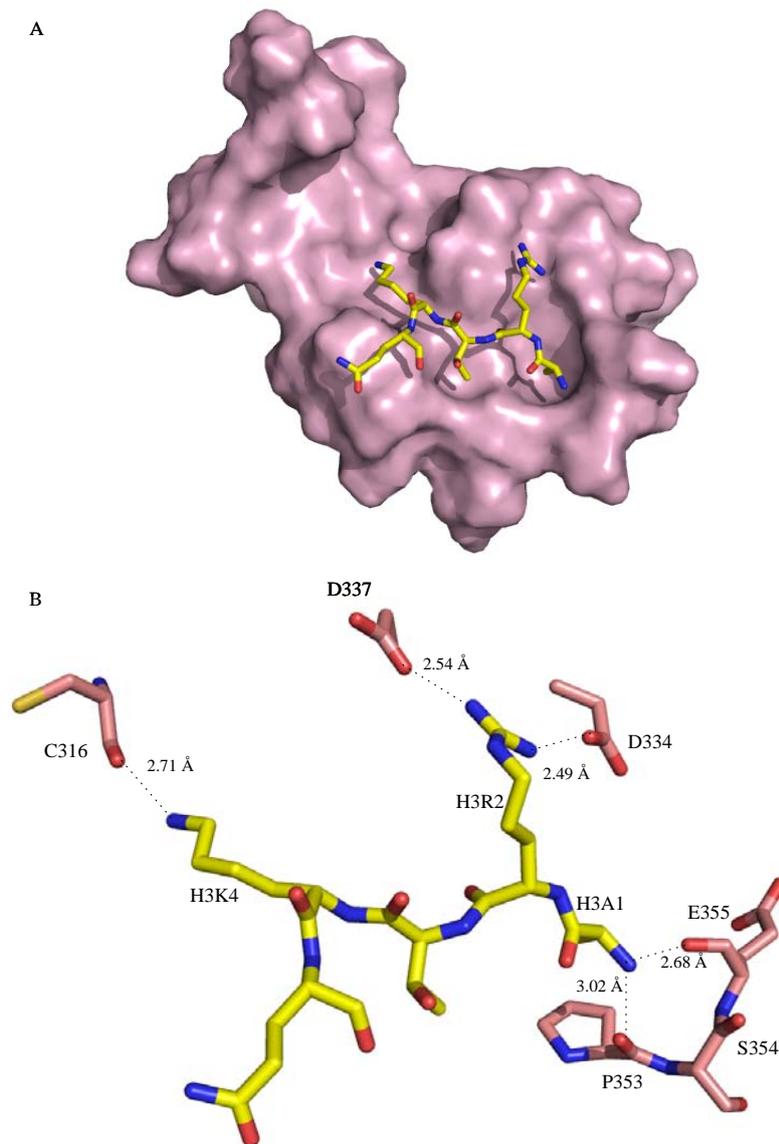


Figure 79: Mode de reconnaissance de la lysine K4 non modifiée de l'histone H3 par le PHD<sub>296-367</sub>. A- Représentation de la surface du PHD de hUHRF1(en rose) en complexe avec l'histone H3K4me0 (en jaune). B- Les bases moléculaires de l'interaction électrostatique entre le PHD de hUHRF1 et l'histone H3K4me0. Les résidus de la protéine hUHRF1 impliqués dans l'interaction avec le peptide H3 sont représentés.

## **Chapitre VII**

# **Conclusions et Perspectives**

---

## Conclusions et Perspectives

Les modifications épigénétiques qui ciblent la chromatine jouent un rôle important dans la régulation de la transcription. Différentes maladies ont été associées aux dysfonctionnements des mécanismes de la régulation épigénétique. Les facteurs impliqués dans ces mécanismes sont devenus des cibles thérapeutiques potentielles.

Nous nous sommes intéressés à la protéine humaine hUHRF1 qui est capable de reconnaître des marques épigénétiques particulières et de recruter des complexes de modifications et de remodelage de la chromatine. Notre travail a consisté à caractériser structurellement les deux domaines centraux de hUHRF1, PHD et SRA, qui sont capables respectivement d'interagir avec l'histone H3 et l'ADN hémiméthylé.

Dans un premier temps, nous avons caractérisé une protéine correspondant aux deux domaines associés PHD et SRA. Nous avons mis au point un protocole d'expression et de purification de cette protéine. Une structure à basse résolution, déterminée par SAXS, a montré un arrangement linéaire des deux domaines étudiés. Cet arrangement permet à chacun des domaines d'interagir librement avec son partenaire. En présence du domaine PHD, le site d'interaction avec l'ADN du domaine SRA reste accessible et en présence de SRA, le domaine PHD est capable d'interagir avec un peptide de l'histone H3. Des microcristaux fragiles ont été obtenus avec cette protéine mais nous n'avons pas pu les optimiser.

Dans un deuxième temps, nous avons caractérisé le domaine PHD isolé pour étudier son interaction avec des peptides de l'histone H3. Une comparaison de séquences et une superposition de structures entre le domaine PHD de hUHRF1 et des domaines PHD reconnaissant des marques spécifiques sur l'histone H3 ont été réalisées. D'après ces comparaisons, le domaine PHD de hUHRF1 ressemble aux domaines PHD reconnaissant la lysine 4 non modifiée de l'histone H3. Une étude structurale a été entamée pour déterminer, au niveau moléculaire, la spécificité d'interaction du domaine PHD avec l'histone H3.

Une première construction correspondant au domaine PHD canonique, PHD<sub>314-367</sub>, a été caractérisée. Différents cristaux ont été obtenus en présence de peptides de l'histone H3 et différentes structures ont été résolues. Dans aucune des structures résolues, un peptide de l'histone H3 n'a été visible. L'analyse des différentes structures a montré que cette construction n'est pas propice pour obtenir la structure du complexe PHD-histone H3 car la région amino-terminale adopte une conformation d'autoinhibition et vient se positionner au niveau du site potentiel d'interaction avec l'histone H3. Un modèle par homologie du

domaine PHD précédé du motif riche en cystéines qui est conservé dans la famille UHRF et qui fixe un zinc supplémentaire a montré que cette partie amino-terminale du PHD adopte une conformation différente en présence de ce motif qui pourrait favoriser l'interaction avec l'histone H3.

Une deuxième construction, PHD<sub>296-367</sub>, correspondant au domaine PHD précédé du motif riche en cystéines a donc été caractérisée. La spécificité d'interaction du domaine PHD de hUHRF1 vis-à-vis de la lysine 4 non modifiée de l'histone H3 a été déterminée à partir des constantes de dissociation obtenues en utilisant la calorimétrie à titration isotherme. Des cristaux ont été obtenus et différentes structures du complexe PHD-H3K4me0 ont été résolues. L'interaction entre la protéine PHD<sub>296-367</sub> et le peptide de l'histone H3 est médiée par une série de liaisons hydrogènes impliquant l'alanine A1, l'arginine R2 et la lysine K4 de l'histone H3 et les acides aminés C316, D334, D337, P353 et E355 de hUHRF1. La lysine K4 non modifiée de l'histone interagit avec le carbonyle de la cystéine C316 de hUHRF1 ce qui explique la nécessité d'avoir le motif amino-terminal fixant un zinc pour obtenir la structure du complexe.

Une perspective à ce travail, en ce qui concerne le domaine PHD, sera l'utilisation de la mutagenèse dirigée pour valider l'importance des acides aminés du domaine PHD, identifiés par l'analyse structurale, dans la reconnaissance de l'histone H3. Les propriétés d'interaction avec différents peptides de l'histone H3 des protéines mutées seront étudiées par ITC. De même, l'effet de ces mutations *in vivo* permettra d'aider à la compréhension de la fonction du domaine PHD au sein de la protéine hUHRF1.

Concernant les domaines associés PHD et SRA de hUHRF1, l'obtention d'une protéine pure et homogène permet d'envisager l'étude de la modulation de la fonction du domaine PHD en présence de SRA, et réciproquement, pour mieux comprendre la fonction biologique de l'association de plusieurs modules de reconnaissance de la chromatine (reconnaissance des histones et de l'ADN) dans une même protéine. Les résultats obtenus sur le domaine PHD isolé devraient aider à augmenter la stabilité de la protéine et à faire aboutir son étude structurale. L'étude structurale du complexe ternaire entre PHD-SRA, l'histone H3K4me0 et l'ADN hémiméthylé serait un des rares exemples de deux modules associés de reconnaissance de la chromatine en complexe avec leurs marques épigénétiques, ce qui pourrait documenter le rôle des interactions interdomaines dans cette reconnaissance.

## **Chapitre VIII**

### **Références Bibliographiques**

---

## Références bibliographiques

Abbady, A. Q., Bronner, C., Bathami, K., Muller, C. D., Jeanblanc, M., Mathieu, E., Klein, J. P., Candolfi, E., and Mousli, M. (2005). TCR pathway involves ICBP90 gene down-regulation via E2F binding sites. *Biochem Pharmacol* 70, 570-579.

Abbady, A. Q., Bronner, C., Trotzier, M. A., Hopfner, R., Bathami, K., Muller, C. D., Jeanblanc, M., and Mousli, M. (2003). ICBP90 expression is downregulated in apoptosis-induced Jurkat cells. *Ann N Y Acad Sci* 1010, 300-303.

Achour, M., Jacq, X., Ronde, P., Alhosin, M., Charlot, C., Chataigneau, T., Jeanblanc, M., Macaluso, M., Giordano, A., Hughes, A. D., *et al.* (2008). The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene* 27, 2187-2197.

Allis, C. D., Jenuwein, T., and Reinberg, D. (2009). *Epigenetics*, (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Arima, Y., Hirota, T., Bronner, C., Mousli, M., Fujiwara, T., Niwa, S., Ishikawa, H., and Saya, H. (2004). Down-regulation of nuclear protein ICBP90 by p53/p21Cip1/WAF1-dependent DNA-damage checkpoint signals contributes to cell cycle arrest at G1/S transition. *Genes Cells* 9, 131-142.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* 455, 818-821.

Avvakumov, G. V., Walker, J. R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C. H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* 455, 822-825.

Baker, L. A., Allis, C. D., and Wang, G. G. (2008). PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. *Mutat Res* 647, 3-12.

Bhaumik, S. R., Smith, E., and Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol* 14, 1008-1016.

Bienz, M. (2006). The PHD finger, a nuclear protein-interaction domain. *Trends Biochem Sci* 31, 35-40.

Bjelic, S., and Jelesarov, I. (2008). A survey of the year 2007 literature on applications of isothermal titration calorimetry. *J Mol Recognit* 21, 289-312.

Bonapace, I. M., Latella, L., Papait, R., Nicassio, F., Sacco, A., Muto, M., Crescenzi, M., and Di Fiore, P. P. (2002). Np95 is regulated by E1A during mitotic reactivation of terminally differentiated cells and is essential for S phase entry. *J Cell Biol* 157, 909-914.

- Bostick, M., Kim, J. K., Esteve, P. O., Clark, A., Pradhan, S., and Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* *317*, 1760-1764.
- Bottomley, M. J., Stier, G., Pennacchini, D., Legube, G., Simon, B., Akhtar, A., Sattler, M., and Musco, G. (2005). NMR structure of the first PHD finger of autoimmune regulator protein (AIRE1). Insights into autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) disease. *J Biol Chem* *280*, 11505-11512.
- Bronner, C., Achour, M., Arima, Y., Chataigneau, T., Saya, H., and Schini-Kerth, V. B. (2007). The UHRF family: oncogenes that are drugable targets for cancer therapy in the near future? *Pharmacol Ther* *115*, 419-434.
- Bronner, C., Trotzier, M. A., Filhol, O., Cochet, C., Rochette-Egly, C., Scholler-Guinard, M., Klein, J. P., and Mousli, M. (2004). The antiapoptotic protein ICBP90 is a target for protein kinase 2. *Ann N Y Acad Sci* *1030*, 355-360.
- Brunet, J., Pfaff, A. W., Abidi, A., Unoki, M., Nakamura, Y., Guinard, M., Klein, J. P., Candolfi, E., and Mousli, M. (2008). *Toxoplasma gondii* exploits UHRF1 and induces host cell cycle arrest at G2 to enable its proliferation. *Cell Microbiol* *10*, 908-920.
- Buolamwini, J. K. (2000). Cell cycle molecular targets in novel anticancer drug discovery. *Curr Pharm Des* *6*, 379-392.
- Chakravarty, S., Zeng, L., and Zhou, M. M. (2009). Structure and site-specific recognition of histone H3 by the PHD finger of human autoimmune regulator. *Structure* *17*, 670-679.
- Champagne, K. S., Saksouk, N., Pena, P. V., Johnson, K., Ullah, M., Yang, X. J., Cote, J., and Kutateladze, T. G. (2008). The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide. *Proteins* *72*, 1371-1376.
- Chignola, F., Gaetani, M., Rebane, A., Org, T., Mollica, L., Zucchelli, C., Spitaleri, A., Mannella, V., Peterson, P., and Musco, G. (2009). The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation. *Nucleic Acids Res* *37*, 2951-2961.
- Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Di Fiore, P. P., and Bonapace, I. M. (2004). Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol Cell Biol* *24*, 2526-2535.
- Cobrinik, D. (2005). Pocket proteins and cell cycle control. *Oncogene* *24*, 2796-2809.
- Consortium, F.-G. A. (1997). An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet* *17*, 399-403.
- De la Paz Sanchez, M., and Gutierrez, C. (2009). Arabidopsis ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. *Proc Natl Acad Sci U S A* *106*, 2065-2070.
- Delagoutte, B., Lallous, N., Birck, C., Oudet, P., and Samama, J. P. (2008). Expression, purification, crystallization and preliminary crystallographic study of the SRA domain of the human UHRF1 protein. *Acta Crystallogr Sect F Struct Biol Cryst Commun* *64*, 922-925.

- Deltour, S., Chopin, V., and Leprince, D. (2005). [Epigenetics and cancer]. *Med Sci (Paris)* 21, 405-411.
- Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R. R., and Richmond, T. J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* 306, 1571-1573.
- Eberharter, A., Vetter, I., Ferreira, R., and Becker, P. B. (2004). ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts. *Embo J* 23, 4029-4039.
- Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N., and Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* 357, 289-298.
- Esteller, M. (2008). Epigenetics in cancer. *N Engl J Med* 358, 1148-1159.
- Fang, Z., Xing, F., Bronner, C., Teng, Z., and Guo, Z. (2009). ICBP90 mediates the ERK1/2 signaling to regulate the proliferation of Jurkat T cells. *Cell Immunol* 257, 80-87.
- Fiedler, M., Sanchez-Barrena, M. J., Nekrasov, M., Mieszczanek, J., Rybin, V., Muller, J., Evans, P., and Bienz, M. (2008). Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex. *Mol Cell* 30, 507-518.
- Fingerman, I. M., Du, H. N., and Briggs, S. D. (2008). Controlling histone methylation via trans-histone pathways. *Epigenetics* 3, 237-242.
- Fuks, F. (2003). [DNA methyltransferases: from chromatin remodeling to cancer]. *Med Sci (Paris)* 19, 477-480.
- Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M., and Mackay, J. P. (2007). Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci* 32, 63-70.
- Gozani, O., Karuman, P., Jones, D. R., Ivanov, D., Cha, J., Lugovskoy, A. A., Baird, C. L., Zhu, H., Field, S. J., Lessnick, S. L., *et al.* (2003). The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. *Cell* 114, 99-111.
- Happel, N., and Doenecke, D. (2009). Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* 431, 1-12.
- Hashimoto, H., Horton, J. R., Zhang, X., Bostick, M., Jacobsen, S. E., and Cheng, X. (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* 455, 826-829.
- Helmut, B., Hildburg, B., and Hans, J. G. (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* 8, 93-99.
- Hopfner, R., Mousli, M., Jeltsch, J. M., Voulgaris, A., Lutz, Y., Marin, C., Bellocq, J. P., Oudet, P., and Bronner, C. (2000). ICBP90, a novel human CCAAT binding protein, involved in the regulation of topoisomerase IIalpha expression. *Cancer Res* 60, 121-128.

- Hopfner, R., Mousli, M., Oudet, P., and Bronner, C. (2002). Overexpression of ICBP90, a novel CCAAT-binding protein, overcomes cell contact inhibition by forcing topoisomerase II alpha expression. *Anticancer Res* 22, 3165-3170.
- Ivanov, A. V., Peng, H., Yurchenko, V., Yap, K. L., Negorev, D. G., Schultz, D. C., Psulkowski, E., Fredericks, W. J., White, D. E., Maul, G. G., *et al.* (2007). PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol Cell* 28, 823-837.
- Iwase, S., Lan, F., Bayliss, P., de la Torre-Ubieta, L., Huarte, M., Qi, H. H., Whetstine, J. R., Bonni, A., Roberts, T. M., and Shi, Y. (2007). The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* 128, 1077-1088.
- Janin, J., Delepierre, M. (1994). *Biologie Structurale, principes et méthodes biophysiques* (Hermann, Editeurs des sciences et des arts).
- Jeanblanc, M., Mousli, M., Hopfner, R., Bathami, K., Martinet, N., Abbady, A. Q., Siffert, J. C., Mathieu, E., Muller, C. D., and Bronner, C. (2005). The retinoblastoma gene and its product are targeted by ICBP90: a key mechanism in the G1/S transition during the cell cycle. *Oncogene* 24, 7337-7345.
- Jenkins, Y., Markovtsov, V., Lang, W., Sharma, P., Pearsall, D., Warner, J., Franci, C., Huang, B., Huang, J., Yam, G. C., *et al.* (2005). Critical role of the ubiquitin ligase activity of UHRF1, a nuclear RING finger protein, in tumor cell growth. *Mol Biol Cell* 16, 5621-5629.
- Johnson, W. C. (1999). Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins* 35, 307-312.
- Jones, D. R., Bultsma, Y., Keune, W. J., Halstead, J. R., Elouarrat, D., Mohammed, S., Heck, A. J., D'Santos, C. S., and Divecha, N. (2006). Nuclear PtdIns5P as a transducer of stress signaling: an in vivo role for PIP4Kbeta. *Mol Cell* 23, 685-695.
- Kadirvel, S., He, F., Muto, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Yokoyama, S. (2006). Solution structure of the PHD domain in RING finger protein 107 (PDB: 2e6s).
- Karagianni, P., Amazit, L., Qin, J., and Wong, J. (2008). ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Mol Cell Biol* 28, 705-717.
- Kepes, F. (2005). [Epigenetics as an aspect of post-genomics]. *Med Sci (Paris)* 21, 371-376.
- Kim, J. K., Esteve, P. O., Jacobsen, S. E., and Pradhan, S. (2009). UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. *Nucleic Acids Res* 37, 493-505.
- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* 36, 1277-1282.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.

- Kozin, M. B., and Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography* *34*, 33-41.
- Lall, S. (2007). Primers on chromatin. *Nat Struct Mol Biol* *14*, 1110-1115.
- Lan, F., Collins, R. E., De Cegli, R., Alpatov, R., Horton, J. R., Shi, X., Gozani, O., Cheng, X., and Shi, Y. (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* *448*, 718-722.
- Landy, A. (1989). Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem* *58*, 913-949.
- Li, F., Huarte, M., Zaratiegui, M., Vaughn, M. W., Shi, Y., Martienssen, R., and Cande, W. Z. (2008). Lid2 is required for coordinating H3K4 and H3K9 methylation of heterochromatin and euchromatin. *Cell* *135*, 272-283.
- Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D., and Patel, D. J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* *442*, 91-95.
- Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R., and Desiderio, S. (2007). A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity* *27*, 561-571.
- Lopez-Serra, L., and Esteller, M. (2008). Proteins that bind methylated DNA and human cancer: reading the wrong words. *Br J Cancer* *98*, 1881-1885.
- Lorenzato, M., Caudroy, S., Bronner, C., Evrard, G., Simon, M., Durlach, A., Birembaut, P., and Clavel, C. (2005). Cell cycle and/or proliferation markers: what is the best method to discriminate cervical high-grade lesions? *Hum Pathol* *36*, 1101-1107.
- Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* *389*, 251-260.
- Macaluso, M., Montanari, M., Noto, P. B., Gregorio, V., Bronner, C., and Giordano, A. (2007). Epigenetic modulation of estrogen receptor-alpha by pRb family proteins: a novel mechanism in breast cancer. *Cancer Res* *67*, 7731-7737.
- Martin, D. G., Baetz, K., Shi, X., Walter, K. L., MacDonald, V. E., Wlodarski, M. J., Gozani, O., Hieter, P., and Howe, L. (2006). The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3. *Mol Cell Biol* *26*, 7871-7879.
- Matthews, A. G., Kuo, A. J., Ramon-Maiques, S., Han, S., Champagne, K. S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D. N., *et al.* (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* *450*, 1106-1110.
- Mattick, J. S., and Makunin, I. V. (2006). Non-coding RNA. *Hum Mol Genet* *15 Spec No 1*, R17-29.

- Miller, J., McLachlan, A. D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *Embo J* 4, 1609-1614.
- Mousli, M., Hopfner, R., Abbady, A. Q., Monte, D., Jeanblanc, M., Oudet, P., Louis, B., and Bronner, C. (2003). ICBP90 belongs to a new family of proteins with an expression that is deregulated in cancer cells. *Br J Cancer* 89, 120-127.
- Murray, K. (1964). The Occurrence of Epsilon-N-Methyl Lysine in Histones. *Biochemistry* 3, 10-15.
- Musselman, C. A., Mansfield, R. E., Garske, A. L., Davrazou, F., Kwan, A. H., Oliver, S. S., Oleary, H., Denu, J. M., Mackay, J. P., and Kutateladze, T. G. (2009). Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem J* 423, 179-187.
- Muto, M., Kanari, Y., Kubo, E., Takabe, T., Kurihara, T., Fujimori, A., and Tatsumi, K. (2002). Targeted disruption of Np95 gene renders murine embryonic stem cells hypersensitive to DNA damaging agents and DNA replication blocks. *J Biol Chem* 277, 34549-34555.
- Nakamura, Y., Umehara, T., Hamana, H., Hayashizaki, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Tanaka, A., Padmanabhan, B., and Yokoyama, S. (2007). Crystal structure analysis of the PHD domain of the transcription co-activator Pygopus. *J Mol Biol* 370, 80-92.
- Org, T., Chignola, F., Hetenyi, C., Gaetani, M., Rebane, A., Liiv, I., Maran, U., Mollica, L., Bottomley, M. J., Musco, G., and Peterson, P. (2008). The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep* 9, 370-376.
- Palacios, A., Garcia, P., Padro, D., Lopez-Hernandez, E., Martin, I., and Blanco, F. J. (2006). Solution structure and NMR characterization of the binding to methylated histone tails of the plant homeodomain finger of the tumour suppressor ING4. *FEBS Lett* 580, 6903-6908.
- Palacios, A., Munoz, I. G., Pantoja-Uceda, D., Marcaida, M. J., Torres, D., Martin-Garcia, J. M., Luque, I., Montoya, G., and Blanco, F. J. (2008). Molecular basis of histone H3K4me3 recognition by ING4. *J Biol Chem* 283, 15956-15964.
- Papait, R., Pistore, C., Grazini, U., Babbio, F., Cogliati, S., Pecoraro, D., Brino, L., Morand, A. L., Dechampsme, A. M., Spada, F., *et al.* (2008). The PHD domain of Np95 (mUHRF1) is involved in large-scale reorganization of pericentromeric heterochromatin. *Mol Biol Cell* 19, 3554-3563.
- Papait, R., Pistore, C., Negri, D., Pecoraro, D., Cantarini, L., and Bonapace, I. M. (2007). Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing. *Mol Biol Cell* 18, 1098-1106.
- Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., Zhao, R., and Kutateladze, T. G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* 442, 100-103.
- Pena, P. V., Hom, R. A., Hung, T., Lin, H., Kuo, A. J., Wong, R. P., Subach, O. M., Champagne, K. S., Zhao, R., Verkhusha, V. V., *et al.* (2008). Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. *J Mol Biol* 380, 303-312.

- Perche, P. Y., Robert-Nicoud, M., Khochbin, S., and Vourc'h, C. (2003). [Nucleosome differentiation: role of histone H2A variants]. *Med Sci (Paris)* 19, 1137-1145.
- Petoukhov, M. V., and Svergun, D. I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 89, 1237-1250.
- Pierce, M. M., Raman, C. S., and Nall, B. T. (1999). Isothermal titration calorimetry of protein-protein interactions. *Methods* 19, 213-221.
- Provencher, S. W., and Glockner, J. (1981). Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 20, 33-37.
- Qian, C., Li, S., Jakoncic, J., Zeng, L., Walsh, M. J., and Zhou, M. M. (2008). Structure and hemimethylated CpG binding of the SRA domain from human UHRF1. *J Biol Chem* 283, 34490-34494.
- Ragvin, A., Valvatne, H., Erdal, S., Arskog, V., Tufteland, K. R., Breen, K., AM, O. Y., Eberharter, A., Gibson, T. J., Becker, P. B., and Aasland, R. (2004). Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J Mol Biol* 337, 773-788.
- Ramon-Maiques, S., Kuo, A. J., Carney, D., Matthews, A. G., Oettinger, M. A., Gozani, O., and Yang, W. (2007). The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proc Natl Acad Sci U S A* 104, 18993-18998.
- Ray-Gallet, D., Gerard, A., Polo, S., and Almouzni, G. (2005). [Variations on the topic of the "histone code"]. *Med Sci (Paris)* 21, 384-389.
- Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406, 593-599.
- Reid, G., Gallais, R., and Metivier, R. (2009). Marking time: the dynamic role of chromatin and covalent modification in transcription. *Int J Biochem Cell Biol* 41, 155-163.
- Saksouk, N., Avvakumov, N., Champagne, K. S., Hung, T., Doyon, Y., Cayrou, C., Paquet, E., Ullah, M., Landry, A. J., Cote, V., *et al.* (2009). HBO1 HAT complexes target chromatin throughout gene coding regions via multiple PHD finger interactions with histone H3 tail. *Mol Cell* 33, 257-265.
- Santos-Rosa, H., and Caldas, C. (2005). Chromatin modifier enzymes, the histone code and cancer. *Eur J Cancer* 41, 2381-2402.
- Schalch, T., Duda, S., Sargent, D. F., and Richmond, T. J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* 436, 138-141.
- Schindler, U., Beckmann, H., and Cashmore, A. R. (1993). HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *Plant J* 4, 137-150.
- Serdyuk, I. N., Zaccai, N. R., and Zaccai, J. (2007). *Methods in molecular biophysics structure, dynamics, function*, (Cambridge: Cambridge University Press).

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., *et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450, 908-912.

Shi, X., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M. R., *et al.* (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* 442, 96-99.

Shi, X., Kachirskaja, I., Walter, K. L., Kuo, J. H., Lake, A., Davrazou, F., Chan, S. M., Martin, D. G., Fingerman, I. M., Briggs, S. D., *et al.* (2007). Proteome-wide analysis in *Saccharomyces cerevisiae* identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J Biol Chem* 282, 2450-2455.

Sreerama, N., and Woody, R. W. (1993). A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem* 209, 32-44.

Svergun, D. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* 25, 495-503.

Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76, 2879-2886.

Taverna, S. D., Ilin, S., Rogers, R. S., Tanny, J. C., Lavender, H., Li, H., Baker, L., Boyle, J., Blair, L. P., Chait, B. T., *et al.* (2006). Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Mol Cell* 24, 785-796.

Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* 14, 1025-1040.

Tremethick, D. J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* 128, 651-654.

Trotzler, M. A., Bronner, C., Bathami, K., Mathieu, E., Abbady, A. Q., Jeanblanc, M., Muller, C. D., Rochette-Egly, C., and Mousli, M. (2004). Phosphorylation of ICBP90 by protein kinase A enhances topoisomerase IIalpha expression. *Biochem Biophys Res Commun* 319, 590-595.

Un, F., Qi, C., Prosser, M., Wang, N., Zhou, B., Bronner, C., and Yen, Y. (2006). Modulating ICBP90 to suppress human ribonucleotide reductase M2 induction restores sensitivity to hydroxyurea cytotoxicity. *Anticancer Res* 26, 2761-2767.

Unoki, M., Brunet, J., and Mousli, M. (2009). Drug discovery targeting epigenetic codes: The great potential of UHRF1, which links DNA methylation and histone modifications, as a drug target in cancers and toxoplasmosis. *Biochem Pharmacol* 78, 1279-1288.

Unoki, M., Nishidate, T., and Nakamura, Y. (2004). ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene* 23, 7601-7610.

- van Ingen, H., van Schaik, F. M., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A. C., Kruijzer, J. A., Liskamp, R. M., Timmers, H. T., and Boelens, R. (2008). Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. *Structure* *16*, 1245-1256.
- Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W., van Schaik, F. M., Varier, R. A., Baltissen, M. P., Stunnenberg, H. G., Mann, M., and Timmers, H. T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* *131*, 58-69.
- Walker, J. R., Avvakumov, G.V., Xue, S., Dong, A., Li, Y., Bountra, C., Weigelt, J., Arrowsmith, C.H., Edwards, A.M., Bochkarev, A., Dhe-Paganon, S. (2008a). Cryptic tandem tudor domains in UHRF1 interact with H3K9ME and are important for pericentric heterochromatin replication (PDB: 3db3).
- Walker, J. R., Avvakumov, G.V., Xue, S., Li, Y., Bountra, C., Weigelt, J., Arrowsmith, C.H., Edwards, A.M., Bochkarev, A., Dhe-Paganon, S. (2008b). Structure of the Ring Domain of the E3 Ubiquitin-Protein Ligase Uhrf1 (PDB: 3fl2).
- Walker, J. R., Wybenga-Groot, L., Doherty, R.S., Finerty Jr., P.J., Newman, E., Mackenzie, F.M., Weigelt, J., Sundstrom, M., Arrowsmith, C., Edwards, A., Bochkarev, A., Dhe-Paganon, S (2005). Ubiquitin-Like Domain of Human Nuclear Zinc Finger Protein NP95 (PDB: 2faz).
- Wang, G. G., Song, J., Wang, Z., Dormann, H. L., Casadio, F., Li, H., Luo, J. L., Patel, D. J., and Allis, C. D. (2009). Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* *459*, 847-851.
- Wolffe, A. P., and Matzke, M. A. (1999). Epigenetics: regulation through repression. *Science* *286*, 481-486.
- Woodcock, C. L. (2005). A milestone in the odyssey of higher-order chromatin structure. *Nat Struct Mol Biol* *12*, 639-640.
- Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P., *et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* *442*, 86-90.
- Zlatanova, J., Bishop, T. C., Victor, J. M., Jackson, V., and van Holde, K. (2009). The nucleosome family: dynamic and growing. *Structure* *17*, 160-171.

## **Chapitre IX**

### **Liste des Figures et des Tableaux**

---

# Liste des Figures

## Chapitre I: Introduction

- Figure 1:** Structure cristallographique d'un mononucléosome (PDB: 1aoi).
- Figure 2:** Les deux modèles proposés pour la structure de la fibre chromatinienne de 30 nm.
- Figure 3:** Structure cristallographique en zigzag d'un tétranucléosome (PDB: 1zbb).
- Figure 4:** La condensation de la chromatine en chromosomes.
- Figure 5:** La méthylation d'une cytosine par une DNMT.
- Figure 6:** Les modifications post-traductionnelles des histones nucléosomales humaines.
- Figure 7:** Les modifications des histones en « cis » et en « trans ».
- Figure 8:** Implication de la protéine UHRF1 dans la régulation du cycle cellulaire.
- Figure 9:** Mécanisme pour la formation de l'hétérochromatine par UHRF1 au niveau de la fourche de réplication ou d'un site de réparation de l'ADN.
- Figure 10:** Les cinq domaines de hUHRF1.
- Figure 11:** Structures des domaines NIRF-N et tudor de hUHRF1.
- Figure 12:** Structure cristallographique du domaine RING de hUHRF1 (PDB: 3fl2).
- Figure 13:** La reconnaissance de la cytosine méthylée d'un dinucléotide CpG hémiméthylé par le domaine SRA de UHRF1 (PDB: 3clz).
- Figure 14:** Structure RMN du domaine PHD de hUHRF2 (PDB: 2e6s).
- Figure 15:** Différentes classes des domaines à doigts de zinc.
- Figure 16:** Alignement de séquences de différents PHD humains.
- Figure 17:** la topologie entrecroisée pour la fixation des zinc dans les domaines PHD.
- Figure 18:** Mode de reconnaissance de l'histone H3K4me0 par les domaines PHD de BHC80 (PDB: 2puy) et AIRE (PDB: 2ke1).
- Figure 19:** Alignement de séquences des PHD de BHC80, AIRE1 et JADE1.
- Figure 20:** Mode de reconnaissance de l'histone H3K4me3 par les domaines PHD de BPTF (PDB: 2fuu), d'ING2 (PDB: 2g6q) et de RAG2 (PDB: 2v89).
- Figure 21:** Alignement de séquences de différents PHD fixant H3K4me3.
- Figure 22:** Séquence du PHD2 de CDH4.
- Figure 23:** Comparaison de séquences entre les trois classes de domaines PHD reconnaissant l'histone H3.

## Chapitre II: Matériels et Méthodes

**Figure 24:** Les réactions de recombinaison et les enzymes nécessaires pour chacun des deux cycles du bactériophage  $\lambda$ .

**Figure 25:** La réaction de recombinaison BP.

**Figure 26:** La réaction de recombinaison LR.

**Figure 27:** Les vecteurs d'expression correspondant aux protéines étudiées pendant ce travail.

**Figure 28:** Induction de l'expression du gène cible par l'IPTG dans E.coli Rosetta 2.

**Figure 29:** Le principe du thermofluor.

**Figure 30:** Principe du spectromètre de masse ayant une source d'électronébulisation et un analyseur à temps de vol

**Figure 31:** Principe de base de la diffusion élastique des rayons X.

**Figure 32:** Diagramme de Guinier (A) et les spectres caractéristiques d'interactions attractives ou répulsives d'une solution idéale (B).

**Figure 33:** Les composants d'un instrument SAXS.

**Figure 34:** Un échantillon absorbant différemment les lumières polarisées droite et gauche.

**Figure 35:** L'onde polarisée elliptiquement est due à une contribution inégale de la polarisation circulaire gauche et droite.

**Figure 36:** Représentation schématique de l'ITC

**Figure 37:** Exemple de courbes obtenues en ITC.

**Figure 38:** La variation de la courbe isotherme en fonction de la constante  $c$ .

**Figure 39:** Le digramme des phases.

**Figure 40:** Les deux dispositifs utilisés pour la cristallisation par diffusion de vapeur.

## Chapitre III: Caractérisation de la protéine PHD-SRA<sub>296-643</sub>

**Figure 41:** La séquence de la protéine PHD-SRA avec les différentes bornes choisies pour le clonage.

**Figure 42:** Le chromatogramme et le gel SDS correspondant de l'étape de chromatographie d'exclusion dans le protocole initial (A) et optimisé (B).

**Figure 43:** : Analyse par spectrométrie de masse de la protéine PHD-SRA<sub>296-643</sub>.

**Figure 44:** : Données de diffusion obtenues avec la protéine PHD-SRA<sub>296-643</sub>.

**Figure 45:** Spectre de dichroïsme circulaire de la protéine PHD-SRA<sub>296-643</sub>.

**Figure 46:** Courbes de dénaturation de PHD-SRA<sub>296-643</sub> obtenues par Thermofluor.

- Figure 47:** La séquence de l'ADN utilisé pour étudier l'interaction avec PHD-SRA<sub>296-643</sub> en spectrométrie de masse.
- Figure 48:** Etude de l'interaction de PHD-SRA<sub>296-643</sub> avec un ADN méthylé et non méthylé par spectrométrie de masse en conditions natives.
- Figure 49:** Caractérisation du complexe entre PHD-SRA<sub>296-643</sub> et l'ADN non méthylé sur colonne de gel filtration analytique.
- Figure 50:** Données de vitesse de sédimentation de la protéine PHD-SRA<sub>296-643</sub> seule à 280 nm (A) et de l'ADN seul à 260 nm (B) analysées avec le programme Sedfit.
- Figure 51:** Courbes de distribution des coefficients de sédimentation de PHD-SRA seul (A), de l'ADN seul (B) et du mélange PHD-SRA avec ADN (C) .
- Figure 52:** Etude de l'interaction par ITC entre PHD-SRA<sub>296-643</sub> à 28  $\mu$ M et le peptide d'histone H3K9me3 (1-20).
- Figure 53:** Etude de l'interaction par ITC entre PHD-SRA<sub>296-643</sub> à 52 $\mu$ M et le peptide d'histone H3K9me3 (1-20).
- Figure 54:** Cristaux obtenus dans une goutte contenant un mélange ternaire PHD-SRA<sub>296-643</sub>, peptide H3 et ADN non méthylé.

## **Chapitre IV: Analyse bioinformatique et clonage du domaine PHD de hUHRF1**

- Figure 55:** Alignement de séquences entre le PHD de hUHRF1 et des PHD interagissant avec H3K4me3, H3 non modifiée et H3K9me3.
- Figure 56:** Alignement de séquences entre les PHD de hUHRF1 et de hUHRF2.
- Figure 57:** Les structures des domaines PHD de hUHRF1 et hUHRF2.
- Figure 58:** Comparaison structurale entre le modèle par homologie du PHD de hUHRF1 et des PHD fixant un peptide de l'histone H3K4me0.
- Figure 59:** Comparaison structurale entre un modèle par homologie du PHD de hUHRF1 et les PHD d'ING2 et de BPTF
- Figure 60:** Les différentes constructions du domaine PHD.

## **Chapitre V: Caractérisation de la protéine PHD<sub>314-367</sub>**

- Figure 61:** Les différentes étapes de purification de la protéine PHD<sub>314-367</sub>.
- Figure 62:** Mesures de spectrométrie de masse effectuées sur la protéine PHD<sub>314-367</sub> pure.
- Figure 63:** Interaction entre la protéine PHD<sub>314-367</sub> et des peptides de l'histone H3.

**Figure 64:** Cristaux obtenus avec la protéine PHD<sub>314-367</sub>.

**Figure 65:** Structure du domaine PHD<sub>314-367</sub> de hUHRF1 (phd32).

**Figure 66:** Comparaison entre les structures des domaines PHD de BHC80 et d'ING2 en complexe avec des peptides de l'histone H3.

**Figure 67:** Modèle du PHD<sub>314-367</sub> de hUHRF1 en complexe avec un peptide de l'histone H3.

**Figure 68:** Modélisation de l'interaction entre le modèle par homologie du PHD de hUHRF1 et l'histone H3K4me0.

## **Chapitre VI: Caractérisation de la protéine PHD<sub>296-367</sub>**

**Figure 69:** Protocole initial de purification de la protéine PHD<sub>296-367</sub>.

**Figure 70:** Mesures de spectrométrie de masse effectuées sur la protéine PHD<sub>296-367</sub> pure.

**Figure 71:** Chromatographie d'échange d'anions de la protéine PHD<sub>296-367</sub>.

**Figure 72:** La caractérisation sur gel natif de l'interaction entre PHD<sub>296-367</sub> et différents peptides de l'histone H3.

**Figure 73:** Caractérisation de l'interaction entre PHD et différents peptides de l'histone H3 en variant la concentration en protéine et les rapports protéine: peptide.

**Figure 74:** Expériences d'ITC réalisées sur VP-ITC avec la protéine PHD<sub>296-367</sub> hétérogène à 65.4  $\mu$ M et différents peptides de l'histone H3.

**Figure 75:** Expériences d'ITC réalisées sur ITC<sub>200</sub> avec la protéine PHD<sub>296-367</sub> hétérogène à 154  $\mu$ M et différents peptides de l'histone H3.

**Figure 76:** Expériences d'ITC réalisées sur ITC<sub>200</sub> avec la protéine PHD<sub>296-367</sub> homogène à 119  $\mu$ M et différents peptides de l'histone H3.

**Figure 77:** Cristaux obtenus avec la protéine PHD<sub>296-367</sub> seule (a) ou en présence de peptides de l'histone H3 (b-f).

**Figure 78:** Structure du domaine PHD<sub>296-367</sub> de hUHRF1 en complexe avec un peptide de l'histone H3.

**Figure 79:** Mode de reconnaissance de la lysine K4 non modifiée de l'histone H3 par le PHD<sub>296-367</sub>.

# Liste des Tableaux

## Chapitre I: Introduction

**Tableau 1:** Différents domaines PHD reconnaissant l'histone H3.

## Chapitre II: Matériels et Méthodes

**Tableau 2:** Protocoles de la première et de la deuxième amplification de l'ADN par PCR.

**Tableau 3:** Protocole de l'amplification de l'ADN par PCR pour vérifier la réaction BP.

**Tableau 4:** Protocole de l'amplification de l'ADN par PCR pour vérifier la réaction LR.

**Tableau 5:** Caractéristiques de la protéine PHD-SRA<sub>296-643</sub>.

**Tableau 6:** Comparaison entre le protocole initial et le protocole optimisé de la purification de la protéine PHD-SRA<sub>296-643</sub>.

**Tableau 7:** Caractéristiques de la protéine PHD<sub>296-367</sub>.

**Tableau 8:** Comparaison entre le protocole initial et le protocole optimisé pour la purification de la protéine PHD<sub>296-367</sub>.

**Tableau 9:** Caractéristiques de la protéine PHD<sub>314-367</sub>.

**Tableau 10:** Protocole de purification de la protéine PHD<sub>314-367</sub>.

**Tableau 11:** Composition des différents gels SDS utilisés durant ce travail.

**Tableau 12:** Bases de données utilisées pour l'interprétation des spectres de dichroïsme circulaire.

## Chapitre III: Caractérisation de la protéine PHD-SRA<sub>296-643</sub>

**Tableau 13:** Les différentes constructions clonées correspondant à la protéine PHD-SRA.

**Tableau 14:** Caractéristiques de la protéine PHD-SRA<sub>296-643</sub> utilisée en SAXS

**Tableau 15:** Analyse du spectre de dichroïsme circulaire de la protéine PHD-SRA<sub>296-643</sub> par 3 algorithmes et en utilisant quatre bases de données.

**Tableau 16:** Composition en éléments de structures secondaires pour PHD-SRA et SRA obtenues par dichroïsme circulaire et analyse structurale.

**Tableau 17:** Les concentrations, coefficients d'extinction molaire et les DO des différentes espèces étudiées par ultracentrifugation analytique en vitesse de sédimentation.

**Tableau 18:** Les coefficients de sédimentation et les concentrations des espèces majoritaires après analyse des données de vitesse de sédimentation de PHD-SRA<sub>296-643</sub>, de l'ADN et du mélange protéine-ADN.

## **Chapitre IV: Analyse bioinformatique et clonage du domaine PHD de hUHRF1**

**Tableau 19:** Les différentes constructions du domaine PHD de hUHRF1

## **Chapitre V: Caractérisation de la protéine PHD<sub>314-367</sub>**

**Tableau 20:** Conditions de cristallisation de la protéine PHD<sub>314-367</sub>.

**Tableau 21:** Les différents jeux de données enregistrés avec des cristaux de la protéine PHD<sub>314-367</sub> au synchrotron SOLEIL et au laboratoire (CEBGS).

**Tableau 23:** Les statistiques de traitement et d'affinement du jeu de données phd32.

## **Chapitre VI: Caractérisation de la protéine PHD<sub>296-367</sub>**

**Tableau 24:** Les différents peptides de l'histone H3 utilisés.

**Tableau 25:** Valeurs de  $\chi^2$ , stœchiométries (N) et constantes de dissociation (Kd) déterminées pour chacune des interactions entre PHD<sub>296-367</sub> et un peptide de l'histone H3.

**Tableau 26:** Valeurs de  $\chi^2$ , stœchiométries (N) et constantes de dissociation (Kd) déterminées pour chacune des interactions entre PHD<sub>296-367</sub> et un peptide de l'histone H3.

**Tableau 27:** Valeurs de  $\chi^2$ , stœchiométries (N) et constantes de dissociation (Kd) déterminées pour chacune des interactions entre PHD<sub>296-367</sub> fixant 3 zinc et un peptide de l'histone H3.

**Tableau 28:** Conditions de cristallisation de la protéine PHD<sub>296-367</sub>.

**Tableau 29:** Les différents jeux de données enregistrés sur les cristaux de la protéine PHD<sub>296-367</sub> homogène à 3 zinc, en présence du peptide de l'histone H3 (1-8).

**Tableau 30:** Les statistiques de traitement et d'affinement du jeu de données ph1.