



**HAL**  
open science

# Etude structurale du complexe de remodelage de la chromatine NuRD et sa sous-unité MBD3 liée à l'ADN

Rachel Tabaroni

► **To cite this version:**

Rachel Tabaroni. Etude structurale du complexe de remodelage de la chromatine NuRD et sa sous-unité MBD3 liée à l'ADN. Biophysique. Université de Strasbourg, 2018. Français. NNT : 2018STRAJ094 . tel-02918031

**HAL Id: tel-02918031**

**<https://theses.hal.science/tel-02918031>**

Submitted on 20 Aug 2020

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

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

*ÉCOLE DOCTORALE des sciences de la vie et de la santé*  
Centre de Biologie Intégrative  
Institut de Génétique et de Biologie Moléculaire et Cellulaire  
IGBMC - CNRS UMR 7104 – INSERM U 1258

**THÈSE** présentée par :

**Rachel TABARONI**

Soutenue le : **12 décembre 2018**

Pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : Sciences de la Vie

Spécialité : Biophysique et Biologie Structurale

**Etude structurale du complexe de remodelage de  
la chromatine NuRD et sa sous-unité MBD3 liée  
à l'ADN**

**THÈSE dirigée par :**

**Dr. Bruno KLAHOLZ**

Directeur de recherche, IGBMC, Université de  
Strasbourg

**RAPPORTEURS :**

**Dr. VENIEN-BRYAN Catherine**

Professeur, Sorbonne université (Paris)

**Dr. FRIBOURG Sébastien**

Directeur de recherche, IECB (Bordeaux)

**EXAMINATEUR :**

**Dr. LAMOUR Valérie**

Maître de conférence-praticien hospitalier, HDR,  
université de Strasbourg



# Remerciements

J'aimerais tout d'abord remercier mon directeur de thèse Bruno, pour m'avoir permis de travailler sur ce passionnant projet durant ces quatre dernières années. Merci de m'avoir fait confiance et permis d'en apprendre tellement sur les remodelleur de la chromatine et la microscopie électronique.

Je tiens à remercier chaleureusement Catherine Venien-Bryan, Valérie Lamour et Sébastien Fribourg, pour avoir accepté de faire partie de mon jury et d'évaluer mon travail.

Merci à Ali Hamiche pour cette collaboration enrichissante et à Khalid et Hatem pour leurs productions de cellules HeLa.

Mes sincères remerciements à tous les membres de l'équipe passés et présents qui ont su m'accueillir et me faire sentir chez moi ici mais également qui ont toujours été présent pour discuter de mon projet. Merci à Morgan, à l'origine de ce projet passionnant, pour m'avoir transmis ses connaissances et pour sa disponibilité quotidienne lors de mes débuts. Thank you Kareem for all your advices on DNA binding proteins crystallization. Merci à Isabelle Billas, même si nous n'avons jamais travaillé ensemble tu as toujours été là avec un petit mot gentil pour me motiver. Merci à Vanessa et Isabelle Hazemann pour leur bonne humeur communicative au sein de la « partie biochimie » de l'équipe. Merci à Otilie pour ses précieux conseils en microscopie électronique et pour les longues discussions sur mon avenir. Merci à Jean-François qui a été mon premier contact en microscopie. Merci à Sasha et Igor pour leur désir de partage lors des nombreux groupmeeting. Ma thèse n'aurait pas été la même sans Brice, merci d'avoir été là pour moi dans les bons comme dans les mauvais jours et pour toutes ces journées passées au Polara ensemble. Je suis ravie de te compter parmi mes amis, et je sais que dans quelques années on se retrouvera pour travailler ensemble! Merci à notre bande d'IT Mathieu et Jonathan, pour leur aide en informatique et pour m'avoir accueilli dans leur bureau-caféteria. Kundhavai, thank you for your constant happiness, I'm sure that wherever I go I will still here you laughing! Et merci à toi aussi Léo, pour ces agréables moments passés en ta compagnie.

Je voudrais également en profiter pour remercier Simon, mon technicien particulier, merci pour ton aide et ta présence quotidienne. Je ne pense pas que ma thèse aurait été ce qu'elle est sans toi et tes précieux conseils. Je n'oublierai pas nos goûters qui vont tellement me manquer. Et j'espère continuer à te voir encore longtemps!

Merci à Nils, pour toutes nos pauses café intensives pour résoudre nos problèmes de purification et de cristallisation, pour ta bonne humeur et tous nos bons moments.

Je tiens évidemment à remercier tous les services et plateformes de l'IGBMC et du CBI qui rendent notre vie plus simple! Merci à Pierre, Alastair, Catherine et Florence pour partager avec passion leurs connaissances.

Merci également à toutes les personnes de l'institut qui ont participé de près ou de loin à ma thèse au travers de souvent brefs moments dans les couloirs. Un merci spécial à Adam pour

son rotor SW60 dont j'ai pris soin comme la prune de mes yeux. Merci aux membres de la spb et aux membres du cineclub Lyuba, Rocio et Dana. Thank you Moamen for making me smile! Un grand merci à tous les stagiaires qui m'ont soutenu et qui sont entrés dans ma vie, Marine, Fina, Tatiana, Hanna, Clément et Thomas.

Mes remerciements vont également à ma famille pour avoir toujours respecté mes choix et m'avoir poussé jusqu'à aujourd'hui. Merci également à Margaux, Lorène et Phivu ma seconde famille à Bordeaux. Et merci ma chère Olivia, plus de 6000 km nous séparent mais ta constante présence fut un réel soutien.

Finally, a very special thanks to Muminjon. Thank you to be in my life, for your support, your love. These last four years were the most precious of my life. I will never find the right words to thank you as you deserve.





# Table des matières

<b>Table des figures et des tableaux</b>	10
<b>Abréviations</b>	14
<b>Résumé du projet</b>	16
<b>Introduction</b>	
<b>I. ADN: le support de l'information génétique</b>	<b>22</b>
<b>II. L'Epigénétique</b>	<b>24</b>
A. La méthylation de l'ADN	24
1. Les méthyltransférases	26
2. Readers de la méthylation	28
a) Les protéines à domaine MBD	28
(1) MeCP2	30
(2) MBD1	32
(3) MBD2	34
(4) MBD3	36
(5) MBD4	40
(6) MBD5 et 6	42
b) Le domaine SET- et RING-Associé : SRA	42
c) La famille Kaiso	42
3. La déméthylation	44
B. Les modifications post-traductionnelles des histones	46
1. L'acétylation des lysines	46
a) HAT/KAT: acetyltransferase enzyme	48
a) HDAC: deacetylase activity	50
(1) Super-famille des Arginase/Desacetylase	50
(1) Deoxyhypusine synthase like NAD/FAD-binding domain superfamily	54
2. Méthylation des arginine et lysine	54
<b>III. Les re-modeleurs de la chromatine ATP-dépendant</b>	<b>54</b>
A. SWI/SNF	55
B. ISWI	55

C.	INO80/SwRI.....	56
<b>IV.</b>	<b>CHD/Mi2: NuRD .....</b>	<b>57</b>
A.	Structure.....	58
1.	CHD3 et 4.....	58
1.	HDAC1 et 2.....	60
2.	MTA1 et 2.....	66
3.	MBD2 et 3.....	70
4.	RbAp46 et 48.....	70
5.	GATAD2A et B.....	74
6.	DOC-1, LSD1 ; sous-unités et autres protéines associées.....	76
B.	Fonctions, régulation et cancer.....	78

## **Le complexe NuRD**

<b>I.</b>	<b>Méthodes utilisées .....</b>	<b>82</b>
A.	Biochimiques.....	82
1.	Séparation des extraits cellulaires.....	82
2.	Affinité.....	83
3.	Les gradients.....	84
a)	Gradient de densité.....	84
b)	Gradient de fixation.....	85
c)	GraDeR.....	86
4.	Caractérisation.....	88
a)	Séparation des protéines : électrophorèse en condition dénaturante SDS-PAGE .	88
(1)	Révélation au nitrate d'argent.....	89
(2)	Révélation au bleu sensible G-250.....	89
b)	Dosage des protéines : La spectroscopie UV/visible.....	90
c)	Analyse : La spectrométrie de masse.....	90
5.	Préparation des échantillons.....	90
a)	Les dialyses.....	91
b)	Le dessalement.....	91
c)	La concentration.....	92
B.	Biologie structurale.....	94
1.	Cryo-microscopie électronique.....	94
a)	La coloration négative: qualité de l'échantillon.....	96

b) La cryo-microscopie électronique.....	98
<b>II. Résultats.....</b>	<b>104</b>
A. Stabilité du complexe NuRD et la sous-unité CHD4 .....	104
B. Dissociation et instabilité du complexe NuRD.....	111
C. Le gradient de type GraFix .....	115
D. Caractérisation par spectrométrie de masse.....	118
E. Le gradient de type GraDeR.....	120
F. Les grilles de carbone continu.....	127
G. Les agents de réticulation .....	133
H. Traitement de données.....	134
<b>III. Conclusion .....</b>	<b>140</b>

## La sous-unité MBD3

<b>I. Méthodes utilisées .....</b>	<b>144</b>
A. Biologie moléculaire.....	144
1. Les vecteurs .....	144
2. Les techniques de clonage .....	145
a) SLIC .....	145
b) RF-free .....	145
B. Expression .....	146
C. Etudes biochimiques de MBD3 2-72 et 2-133 .....	147
1. Lyse cellulaire.....	147
2. Chromatographie .....	148
a) Affinité.....	148
b) Echangeuse d'ions .....	150
c) Chromatographie d'exclusion stérique .....	150
3. Hybridation des oligonucléotides et formation des complexes .....	151
D. Etudes biophysique.....	152
1. Concentration.....	152
2. Expérience de retard sur gel .....	152
3. DLS.....	153
4. Prometheus .....	154
5. SEC-MALS .....	155
E. Biologie structurale: cristallographie aux rayons X.....	155

1. Cristallisation.....	155
2. Méthodes utilisées .....	157
a) Gouttes assises et gouttes suspendues : la diffusion de vapeur.....	157
b) Cristallisation dans l'huile .....	158
c) Cristallisation en gel .....	158
d) Ensemencement .....	158
e) La diffraction aux rayons-X.....	159
<b>II. Résultats.....</b>	<b>160</b>
A. MBD3 2-72.....	161
1. Purification et caractérisation de MBD3 2-72-histag .....	161
2. Caractérisation de l'interaction MBD-ADN.....	163
B. MBD3 2-72 $\Delta$ 6histag.....	173
1. Purification .....	173
2. Caractérisation du complexe MBD3 2-72 $\Delta$ 6histag.....	178
3. Cristallisation des complexes oligoprotéiques MBD3 2-72 .....	180
C. MBD3 2-133.....	184
1. La construction 2-133 .....	184
2. Purification de MBD3 2-133 .....	187
3. Complexes oligoprotéiques MBD3 2-133:ADN-RPL37-21bp et caractérisations biophysiques .....	189
4. Cristallisation.....	191
<b>III. Conclusion .....</b>	<b>193</b>
<b>Discussion et perspectives</b>	<b>194</b>
<b>Annexes</b>	<b>204</b>
<b>Références</b>	<b>210</b>

# Table des figures et des tableaux

Figure 1	Compaction de l'ADN	21
Figure 2	Le nucléosome	21
Figure 3	Méthylation de la cytosine	25
Figure 4	Méthylation de l'ADN	25
Figure 5	Les protéines de liaison à l'ADN méthylé	27
Figure 6	Détails structuraux du complexe MeCP2-ADN méthylé	29
Figure 7	Le crochet AT et sa reconnaissance des régions AT riches	29
Figure 8	MBD1 et la reconnaissance de l'ADN méthylé	31
Figure 9	Le domaine doigt de zinc de MBD1 et la reconnaissance de l'ADN non méthylé	31
Figure 10	MBD2 et la reconnaissance de l'ADN méthylé	33
Figure 11	L'exclusion de l'arginine dans la triade de reconnaissance	33
Figure 12	MBD3 et la reconnaissance de l'ADN méthylé	37
Figure 13	Comparaison de la triade entre MBD3 et MBD2	37
Figure 14	La triade MBD4	39
Figure 15	Le domaine glycosylase de MBD4	39
Figure 16	Les protéines à domaine SRA	41
Figure 17	Les protéines KAISO	41
Figure 18	Le cycle de déméthylation de l'ADN	43
Figure 19	Modifications post-traductionnelles des histones	45
Figure 20	Les HATs	47
Figure 21	Les HDACs	49
Figure 22	Régulation de la compaction de la chromatine ; équilibre entre HAT et HDAC	51
Figure 23	Les remodeleurs et leurs mécanismes d'action	53
Figure 24	CHD3 et CHD4	59
Figure 25	HDAC1 et 2	61
Figure 26	MTA1(ELM2-SANT) en complexe avec HDAC1	63
Figure 27	La poche basique de liaison à l'IP4 de HDAC	65
Figure 28	Architecture du coeur du complexe NuRD par microscopie électronique	67
Figure 29	RbAp46/48	69
Figure 30	RbAp48 et la poche hydrophile latérale	71
Figure 31	RbAP48 et l'interaction avec l'histone H3	73
Figure 32	L'interface super-hélice entre MBD2 et GATAD2A	75
Figure 33	Gradient, GraFix et GraDeR	85
Figure 34	microscopie électronique à transmission (TEM)	93
Figure 35	Coloration négative	95
Figure 36	Cryo-EM	99
Figure 37	Purification du complexe NuRD, SDS-PAGE	105
Figure 38	Image des contaminations et type de glace en EM	106
Figure 39	Coloration negative et cryo-EM images	108
Figure 40	NuRD analysis and stability	109
Figure 41	glace épaisse vs glace fine	112

Figure 42 dessalement et dialyse	113
Figure 43 Caractérisation des fractions ENI et ENS	115
Figure 44 Analyse par cryo-EM de l'extrait ENI	116
Figure 45 Grandissement de particules observées en figure 45	117
Figure 46 Analyse de spectrométrie de masse	118
Figure 47 SDS gel des fractions du gradient	121
Figure 48 Image en coloration négative	122
Figure 49 Grandissement de particule pour ENI-09	123
Figure 50 Grandissement de particule pour ENI-14	124
Figure 51 Grandissement de particule pour ENI-16	124
Figure 52 Influence du gradient de type GraDeR	126
Figure 53 Analyse par électrophorèse SDS-PAGE de NuRD	127
Figure 54 micrographes de la fraction insoluble ENI	128
Figure 55 Analyse par électrophorèse SDS-PAGE de NuRD	130
Figure 56 micrographes de la fraction insoluble ENS	131
Figure 57 Comparaison des formes observées en coloration négative en cryo-EM.	132
Figure 58 BS3 et grilles en carbone continu	134
Figure 59 Micrographe de la fraction soluble	135
Figure 60 Classes 2D	136
Figure 61 Correction de la CTF par le logiciel cisTEM	137
Figure 62 Sélection des particules sous cisTEM	137
Figure 63 Classes 2D réalisées avec cisTEM	138
Figure 64 Carte et description des plasmides utilisés	143
Figure 65 Diagramme de phase	156
Figure 66 Prédiction des régions désordonnées de MBD3	160
Figure 67 Purification du domaine MBD de MBD3 par chromatographie d'affinité puis d'exclusion stérique	162
Figure 68 Mesures de DLS réalisées sur le MBD purifié	162
Figure 69 Migration sur gel des complexes MBD3-ADN 20bp en conditions natives	164
Figure 70 Migration sur gel en condition natives d'un complexe MBD:DNA en fonction de concentration de NaCl	165
Figure 71 Migration sur gel en condition natives d'un complexe MBD:DNA à différents ratios	166
Figure 72 Cristal monté et cliché de diffraction à 4 Å de résolution	169
Figure 73 Densité électronique	170
Figure 74 Reproduction des cristaux obtenus	172
Figure 75 Tests de d'expression de MBD	173
Figure 76 Courbe de dénaturation du MBD	174
Figure 77 Purification du domaine MBD de MBD3 par chromatographie d'affinité puis d'exclusion stérique.	175
Figure 78 Nouveaux oligonucléotides	177
Figure 79 Expérience de NanoDSF du MBD seul et en complexe avec un ADN de 11 ou 20 bp	178
Figure 80 Comparaison par chromatographie d'exclusion par la taille couplée la détection de la diffusion de lumière laser multi-angles (SEC-MALLS).	179
Figure 81. Cristallisation en goutte assise et goutte suspendue et impact sur la localisation de la croissance cristalline.	182

Figure 82. Nouveaux cristaux obtenus par la méthode de cross-seeding et le screen commercial JCSG+.	183
Figure 83. La région IDR et le crochet AT chez les MBD protéines	185
Figure 84. La région IDR conservée chez MBD3	186
Figure 85. Les étapes de purification de la construction MBD3 2-133	188
Figure 86. gels PAGE en condition natives	189
Figure 87. NanoDSF	190
Figure 88. Organisation du complexe NuRD en domaines	198
Figure 89. La région IDR et le crochet AT chez les MBD protéine	207
Figure 90. La région IDR conservée chez MBD3	208

Tableau 1 Les détergents communément utilisés en Cryo-EM et leurs propriétés .....	87
Tableau 2 Composition des gels de séparation et concentration.....	88
Tableau 3 Composition pour la révélation au bleu G-250 .....	89
Tableau 4 comparaison générale .....	103
Tableau 5 Résumé des conditions testées sur grilles en carbone continu. ....	129
Tableau 6 composition des gels natifs.....	153
Tableau 7 Oligonucléotides.....	163
Tableau 8 Cristaux du complexe MBD3:ADN .....	167
Tableau 9 Statistiques de collecte des données de diffraction .....	169
Tableau 10 Criblage sur la base des conditions de cristallisation mises en évidence .....	171
Tableau 11. Cristaux de complexes MBD 2-72 $\Delta$ 6histag:ADN-RPL37-11bp obtenus avec le screen PACT à 4°C. ....	181
Tableau 12. Cristaux de complexes MBD 2-133 $\Delta$ 6histag:ADN-RPL37-11bp obtenus avec le screen PACT à 4°C. ....	192



# Abréviations

## Acides aminés

Alanine, Ala, A  
Arginine, Arg, R  
Asparagine, Asn, N  
Acide aspartique, Asp, D  
Cystéine, Cys, C  
Acide glutamique, Glu, E  
Glutamine, Gln, Q  
Glycine, Gly, G  
Histidine, His, H  
Isoleucine, Ile, I  
Leucine, Leu, L  
Lysine, Lys, K  
Méthionine, Met, M  
Phénylalanine, Phe, F  
Proline, Pro, P

Sérine, Ser, S  
Thréonine, Thr, T  
Tryptophane, Trp, W  
Tyrosine, Tyr, Y  
Valine, Val, V

## Acides nucléiques

Adénine, A  
Cytosine, C  
Guanine, G  
Thymine, T  
Uracile, U

ADN: Acide désoxyribonucléique

CHAPS: 3-[(3-Cholamidopropyl)diméthylammonio]-1-propanesulfonate

MBD: Méthyl-cytosine binding domain

DLS: Dynamic-light scattering

EDTA: Ethylenediaminetetraacetic acid

HEPES: 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid

IPTG: Isopropyl  $\beta$ -D-1-thiogalactopyranoside

PAGE: Polyacrylamide gel electrophoresis

PEG: Polyéthylène glycol

pI: pH isoélectrique

SDS: Sodium dodecyl sulfate

TCEP: Tris(2-carboxyethyl)phosphine TRIS : Tris(hydroxyméthyl)aminométhane



# Résumé du projet

---

La chromatine est une structure extrêmement dynamique. Toutes les modifications présentes sur la chromatine (ADN et/ou histones) vont pouvoir servir de signalisation pour initier la régulation de l'expression des gènes. En effet ils vont permettre de recruter des complexes de remodelage de la chromatine. Ces complexes de remodelage de la chromatine jouent un rôle clé dans le contrôle de l'expression génique. Ils vont en effet réguler l'accessibilité de l'ADN à la machinerie transcriptionnelle. Cette régulation se réalise au travers de différents processus tels que la modification des histones et le repositionnement des nucléosome sur l'ADN eucaryote. L'action de ces complexes nécessite de l'énergie utilisée sous forme d'ATP ; d'où leur nom de complexe de remodelage ATP-dépendant. Ils sont organisés en différentes familles mais tous possèdent une fonction ATPase translocase. Dans certains cas de cancers ces mécanismes sont dérégulés et sont également connus pour être responsable de phénotypes précancéreux. L'un de ces principaux complexes, NuRD (Nucleosome Remodeling and histone Deacetylation), a fait l'objet de mon projet d'étude. C'est le seul complexe connu couplant deux activités apparemment opposées, une activité de remodelage de la chromatine ATP dépendante et une activité de déacétylation des histones qui va induire la répression de la transcription. Ce complexe est largement exprimé et conservé chez tous les eucaryotes. Le complexe joue un rôle majeur donc dans la régulation de l'expression des gènes, l'organisation de la chromatine, les réparations des dommages causés à l'ADN, et le maintien de la stabilité du génome.

Le complexe NuRD contient 8 sous-unités de base parmi lesquelles on trouve HDAC1/2, CHD3/4, RbAp46/48, MBD2 ou 3, GATAD2a et/ou GATAD2b, MTA1/2/3 et DOC-1. Bien que ce complexe soit à l'origine de nombreux projets biochimiques et structuraux, de nombreuses questions restent en suspens. Des analyses récentes par spectrométrie de masse sur la stœchiométrie du complexe suggèrent que NuRD serait composé d'une protéine CHD3 ou CHD4, une HDAC1 et HDAC2, trois MTA1/2/3, une MBD3, six RbAp46/48, deux GATAD2b ou GATAD2a et deux protéines DOC-1. Néanmoins ces données sont en contradiction avec l'analyse structurale du cœur du complexe par microscopie électronique (à une résolution de 19Å, publication apparu au cours de ma

thèse) soulignant la présence d'un dimère central HDAC1:MTA1 et 4 sous-unités RbAp46/48.

L'objectif de mon projet était l'étude structurale du complexe NuRD par cryomicroscopie électronique. Cette étude structurale du complexe visait à permettre d'obtenir des informations inédites sur la stœchiométrie du complexe mais également des détails moléculaires sur les spécificités d'interaction au sein du complexe, notamment sur la sous-unité MBD3 et ses interactions avec l'ADN. Ce projet faisait l'objet d'une collaboration entre l'équipe de biologie structurale du Dr. Bruno Klaholz et l'équipe « génomique fonctionnelle et cancer » du Dr. Ali Hamiche de l'IGBMC.

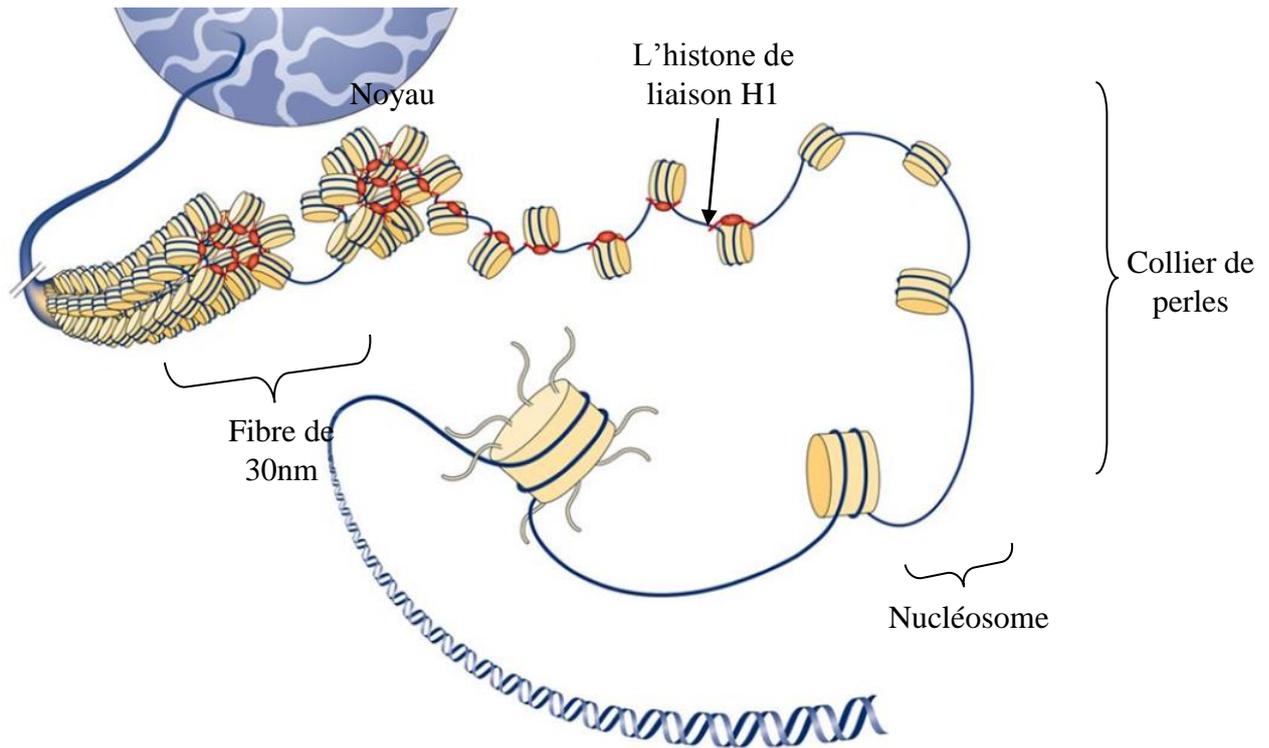
Parmi les sous unités du complexe NuRD, MBD3 pour Methyl-CpG Binding Domain protein 3 présente une importance particulière. Cette sous-unité fait partie de la famille des protéines MBD, qui ont la capacité de se fixer aux cytosines méthylées au niveau d'îlots riches en dinucléotides mCpG et de jouer le rôle de « traducteurs » entre le niveau de méthylation de l'ADN et le remodelage de la chromatine par recrutement du complexe NuRD. Néanmoins, les bases moléculaires de la reconnaissance du nucléosome et de l'ADN nucléosomal par le complexe NuRD et ses sous-unités reste inconnu. Il est important ici de souligner que MBD3, bien que partageant une identité de séquence de 77% avec MBD2, est la seule protéine de la famille des MBD à ne pas présenter de spécificité pour l'ADN méthylé, dû à deux mutations (H30 et F34). La reconnaissance du niveau de méthylation de l'ADN par MBD3 reste néanmoins une question ouverte aujourd'hui ; en effet, bien que certaines données suggèrent que MBD3 interagissait avec l'ADN hydroxyméthylé, un nombre important d'autres expériences ont montré le caractère controversé de ces résultats; et d'autres encore que l'affinité de MBD3 pour l'ADN serait indépendant du niveau de méthylation, comme le prouve la récente structure cristallographique du domaine de MBD3 en complexe avec une séquence d'ADN méthylé. Il reste cependant certain que MBD3 participe au recrutement du complexe sur l'ADN cible, de ce fait il est particulièrement intéressant d'étudier en parallèle du complexe NuRD cette sous-unité pour obtenir des détails moléculaires plus précis en terme de reconnaissance.

Au cours de ma thèse j'ai pu mettre au point la purification de différentes constructions de la protéine MBD3 et caractériser leurs interactions avec divers ADN cibles. Les essais de cristallisation ont permis l'obtention de cristaux pour les différents complexes, et des jeux de données de diffraction aux rayons X ont pu être enregistrés au SLS. Le traitement de ces données est actuellement en cours afin de déterminer la structure du

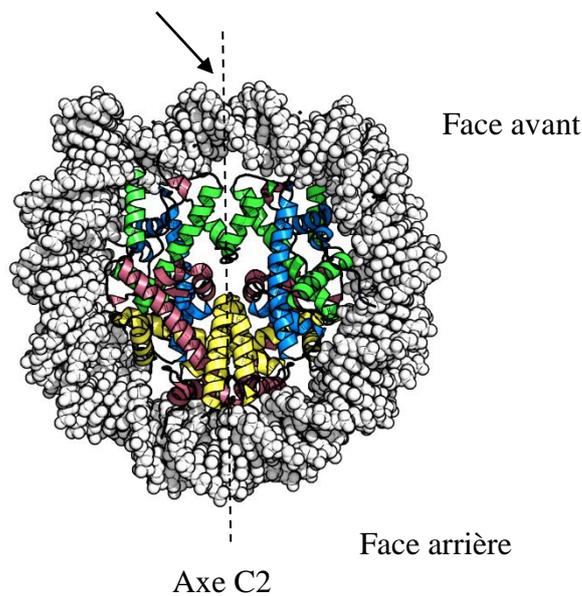
complexe MBD3 avec l'ADN par cristallographie. J'ai également pu optimiser la purification du complexe NuRD, afin d'adapter le protocole déjà développé pour une étude par cryo-microscopie électronique. Le complexe a ainsi pu être purifié et caractérisé par spectrométrie de masse, mettant en évidence la présence des différentes sous-unités mais également de deux autres protéines que sont l'actine cytoplasmique et l'annexin 2. Le complexe a pu être par ailleurs caractérisé par microscopie électronique. Des images en conditions cryo ont pu être obtenues et l'analyse de ces dernières est en cours afin d'étudier la structure du complexe entier







**Figure 1 Compaction de l'ADN** – Représentation schématique des différents états de compaction de l'ADN. Adaptée de Dmitriy V. Fedorov et al. *Emerging roles of linker histones in regulating chromatin structure and function*. *Diade reviews Bol. Cell. Bio* 2018.



**Figure 2 Le nucléosome** – Représentation avec le logiciel Pymol, du nucléosome. En vert sont représentées les histones H3, en bleues H4, en rouge H2A et en jaune H2B. On observe également un axe de symétrie de type C2 représenté en noir. Adaptée de Wakamori, M. et al. *Intra- and inter-nucleosomal interactions of the histone H4 tail revealed with a human nucleosome core particle with genetically-incorporated H4 tetra-acetylation*, *Sci Rep* 2015. 5av6

# INTRODUCTION

---

## I. ADN: le support de l'information génétique

Chez les eucaryotes, l'ADN est le support de l'information génétique. Il consiste en une longue séquence de paires de bases compactée sous forme de chromosomes et codant pour l'ensemble des protéines de l'organisme. L'ADN de l'Homme se compose d'au moins 3 milliard de paires de bases répartis dans 23 paires de chromosomes. Cette structure représente le plus haut degré de compaction de l'ADN. Il est en effet fondamental de compacter cette information, l'ADN mesurant jusqu'à 2 mètre de long doit s'adapter pour être conservé dans le noyau eucaryote environ 200 000 fois plus petit (Figure 1). Cette compaction va également permettre la protection et la régulation de l'information génétique en restreignant son accès. Ainsi la chromatine est en constant remodelage entre un état de compaction, l'hétérochromatine, et un état décondensé transitoirement et localement, l'euchromatine transcriptionnellement active (Li *et al.*, 2007).

Le nucléosome constitue l'unité de base de la chromatine et contient 147 paires de base d'ADN enroulées autour d'un octamère d'histones. Comme on peut l'observer sur la figure 2 chaque octamère se compose d'un tétramère central d'histones H3-H4 encadré par deux dimères H2A-H2B de chaque côté. L'octamère d'histone est maintenu par l'ADN lui-même maintenu par une histone H1 aux extrémités 5' et 3' de l'ADN (Bednar *et al.*, 2017).

Un nombre important de mécanismes vont jouer un rôle dans la transition dynamique de la chromatine incluant l'assemblage/désassemblage des nucléosomes, le remplacement des histones, les modifications post-traductionnelles des histones ou encore la méthylation de l'ADN. Les protéines à l'origine de ces régulations ne possèdent pas de domaine direct de liaison à l'ADN, de ce fait elles sont recruté de façon indirect par des protéines médiatrices de l'activation ou de la répression (Young, 2011).



## II. L'Épigénétique

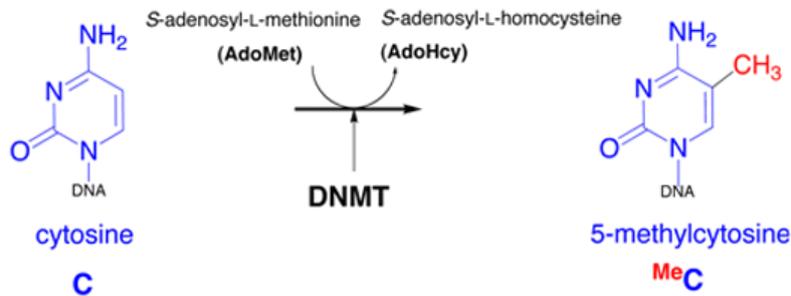
Le terme d'épigénétique fut pour la première fois introduit au début des années 1940 par Conrad Waddington qui la définit comme la branche de la biologie étudiant la causalité entre les gènes et leur produit responsable du phénotype. L'évolution dans la découverte de bon nombre de processus a nécessité par la suite une évolution de cette définition. Aujourd'hui l'épigénétique est l'étude des modifications génétiques altérant l'activité des gènes sans pour autant résulter d'un quelconque changement de la séquence d'ADN. Le dernier caractère important joint à cette définition est l'héritage de ces modifications pouvant être transmise aux cellules filles (Deans and Maggert, 2015). Ces variations sont nécessaires pour le développement normal de tout individu, mais tout dérèglement altérera directement l'activation ou la répression de gènes. C'est pour ces raisons que la plupart sont associées au développement de cancer, retard mental ou encore de syndrome lié à l'instabilité chromosomale.

Ces modifications peuvent être classées en deux groupes majeurs : La méthylation de l'ADN et les Modifications Post-Traductionnelles (MPTs) de la chromatine. Elles seront la cible de protéines et/ou de complexes protéiques à l'origine de la régulation de l'expression des gènes, le maintien de la stabilité du génome ou encore les réparations de l'ADN. Ces modifications vont donc servir d'ancrage aux complexes tels que le complexe nommé NuRD, pour Nucleosome Remodeling and histone Deacetylation complex, qui sera présenté dans le chapitre suivant.

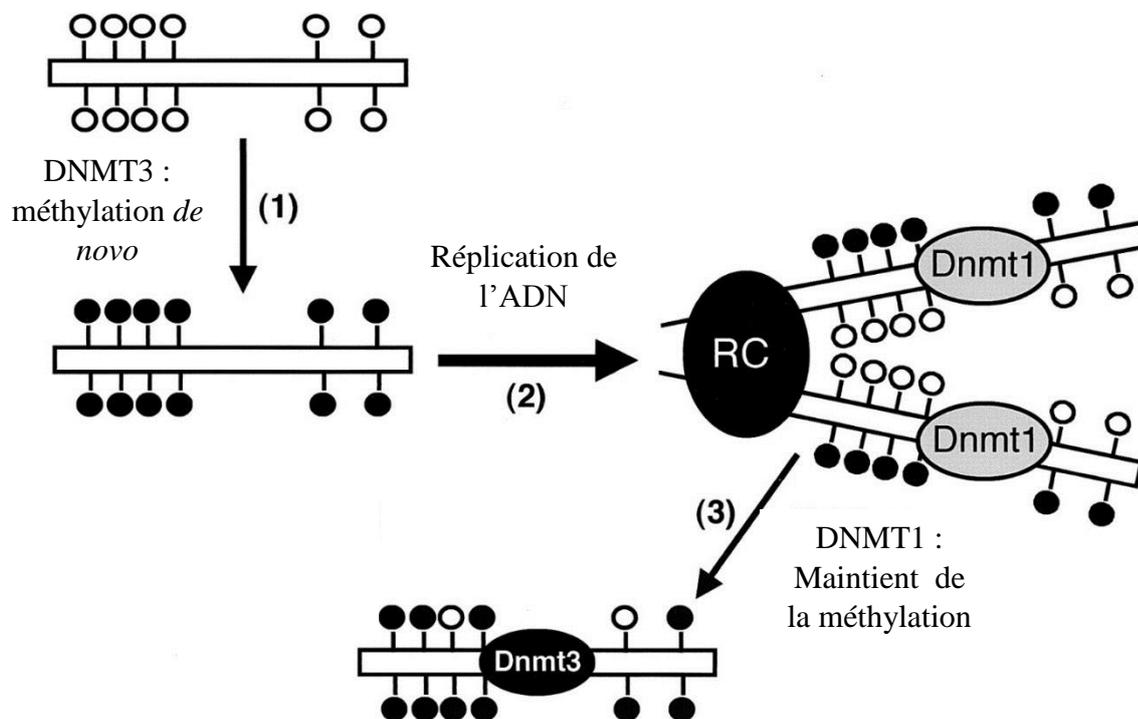
### A. La méthylation de l'ADN

En 1948, Hotchkiss, mis en évidence la présence de la 5-méthyl-cytosine dans l'ADN de thymus de veau par chromatographie sur papier (Hotchkiss, 1948). Cette découverte, qui fut la première preuve de modification de l'ADN, en amont de celle de sa structure par Watson et Crick a permis de mettre en place la première brique à ce géant édifice qu'est l'épigénétique.

Les aspects généraux de cette modification sont aujourd'hui encore à l'étude, néanmoins, il est connu que la méthylation résulte en l'addition d'un groupement méthyle sur une cytosine de l'ADN en position 5' pour former une 5'-méthylcytosine. Elle est retrouvée chez la plupart des êtres vivants eucaryotes comme procaryotes. Cependant, chez certains organismes tels *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* et nombre d'invertébrés cette capacité de modification de l'ADN est absente ou très faible (Wenzel *et al.*, 2011).



**Figure 3 Méthylation de la cytosine** – la méthylation de la cytosine va permettre l’addition d’un groupement méthyle en 5<sup>o</sup> par les DNMS méthyltransférases (DNMTs).



**Figure 4 Méthylation de l’ADN.** Ce schéma représente les deux mécanisme de méthylation principaux retrouvés chez le mammifères. (1) Caractérise la méthylation de l’ADN au stade embryonnaire. (2) Lors de la réplication de l’ADN, celui-ci se retrouve sous forme hémiméthylé et la restauration de la méthylation ce fera par l’intervention des DNMT1 (3). *Adaptée de Chen T., et al. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b, Molecular and cellular biology 2003.*

Il est important de noter l'importance de cette modification chez les eucaryotes, aujourd'hui considérée parfois comme la cinquième base de l'ADN, en addition à l'adénine, la thymine, la cytosine et la guanine. Bien que les méthylations de l'ADN soient principalement trouvées au niveau de cytosine suivi par un résidu guanine (CpG) ; depuis 2006, de nombreuses autres modifications similaires ont pu être découvertes telles que la 5-hydroxyméthylcytosine (5hmC), 5-formylcytosine (5fC), 5-carboxycytosine (5caC), et la *N*<sup>6</sup>-méthyladénine (m6A) qui semblent jouer des rôles cruciaux dans nombres de processus biologiques (Wion and Casadesús, 2006) (He et al., 2011a).

Ainsi au sein du génome humain on retrouve environ 28 millions de sites CpG parmi lesquels 60 à 80% des cytosines sont méthylées (Smith and Meissner, 2013). Ces dinucléotides ont tendance à former des clusters dans des régions promotrices nommés îlots CpG ou ils sont majoritairement non méthylés et associés à une fonction activatrice de l'expression des gènes. Ainsi 60% des promoteurs de gènes sont associé à des îlots CpG.

Différentes protéines vont pouvoir interagir avec cette modification et ainsi moduler l'information contenue dans l'ADN. Trois groupes de protéines ont pu être mis en évidence, classés selon leur rôle:

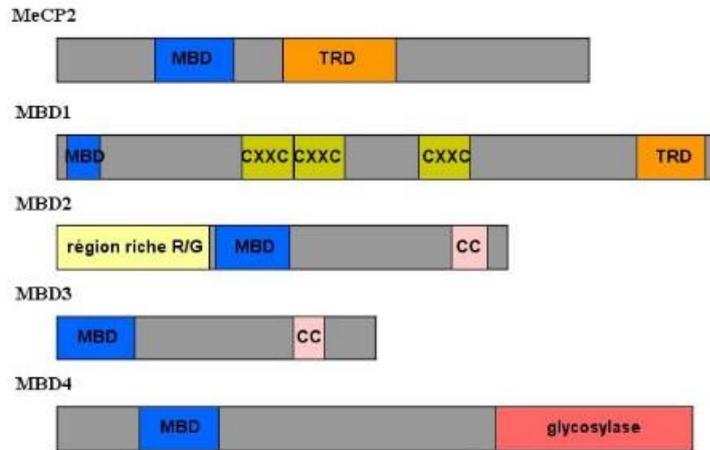
- Les « writer » : composé des DNMT qui catalysent la méthylation.
- Les « reader » : composé des familles MBD, SRA et KAI1, qui vont permettre le transfert de l'information contenu par cette méthylation pour réguler l'expression génique.
- Les « eraser » : composé des ten-eleven translocation dioxygénase via une cascade d'oxydations qui vont oxyder le composé méthyle en vue de son élimination.

Il existe un équilibre entre ces différentes enzymes, impliqué dans bon nombre de processus tels que l'embryogenèse, la différenciation de cellules souches, l'inactivation du chromosome X et l'expression génique. Ainsi, il est facile de comprendre que tout déséquilibre peut être la cause de nombreuses pathologies comme certaines maladies génétiques et le cancer.

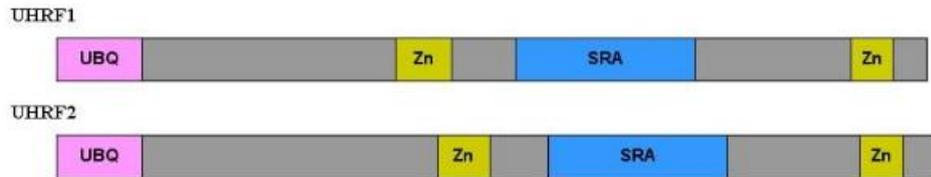
### 1. Les méthyltransférases

Chez les mammifères on compte trois méthyltransférases responsables de la méthylation de la cytosine en position 5'. Ce mécanisme décrit précédemment requiert l'intervention d'un co-facteur, le SAM (S-adénosyl-L-méthionine) (Figure 3).

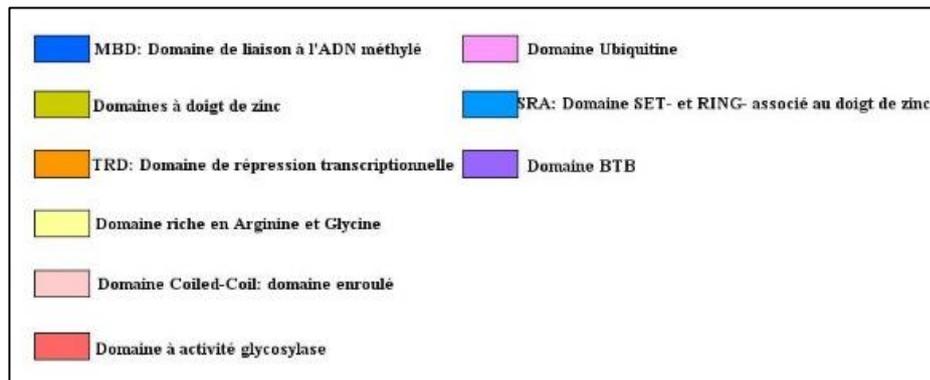
## Methyl-Binding Protéines



## Famille SRA



## Famille Kaiso



**Figure 5 Les protéines de liaison à l'ADN méthylé** – Cette figure illustre les représentants majeurs des différentes familles de MBPs. On retrouve les MBD et leur domaine homonyme de fixation à l'extrémité N-amino terminale. Les SRA et leur domaine de fixation encadré de domaines à doigt de zinc. La famille Kaiso présentant trois domaines à doigt de zinc dans la partie amino-C terminale. Réalisé avec le logiciel IBS (Liu et al., 2015)

- DNMT1, cette enzyme est responsable du maintien de la méthylation et permet de restaurer l'état de méthylation suite à la réplication de l'ADN. De ce fait, il a été prouvé qu'elle possédait une affinité accrue pour les sites hémyméthylés.
- DNMT3A et DNMT3B sont connues pour catalyser *de novo* l'ADN non méthylé et ainsi définir le méthylôme d'où leur expression principalement lors des stades précoces du développement embryonnaire (Guo *et al.*, 2014). (Figure 4)

Ainsi la présence de cytosines méthylées peut prévenir ou au contraire favoriser la reconnaissance par des facteurs de transcription des régions promotrices. Les 5'-mC peuvent également être reconnues directement par des protéines liant l'ADN méthylé présentées plus en détails ci-dessous.

## 2. Readers de la méthylation

Les protéines dites « readers » vont interagir de façon spécifique avec les CpG méthylés et jouer un rôle de traducteur entre le niveau de méthylation de l'ADN et la modification des histones afin de réguler l'expression génétique. De ce fait les MBPs vont servir de point d'ancrage aux autres protéines effectrices au niveau de loci particuliers en vue de réprimer l'expression des gènes.

Différentes familles ont pu ainsi être identifiées et classées selon leur domaine de reconnaissance pour l'ADN méthylé. La première fut celle des Methyl-CpG-binding domain (MBD). Par la suite deux autres familles furent découvertes, les familles Kaiso (domaine C2H2 doigt de zinc) et SET-Ring finger associated (domaine SRA).

### a) Les protéines à domaine MBD

Les protéines à domaine MBD représentent la famille la plus nombreuse de « reader » de la méthylation. Comme on peut le remarquer dans la figure 5, sept protéines ont aujourd'hui été décrites et arborent différents domaines. Dans cette partie je décrirai plus en détail cette famille de protéine, MBD3 faisant l'objet de mon projet de thèse.



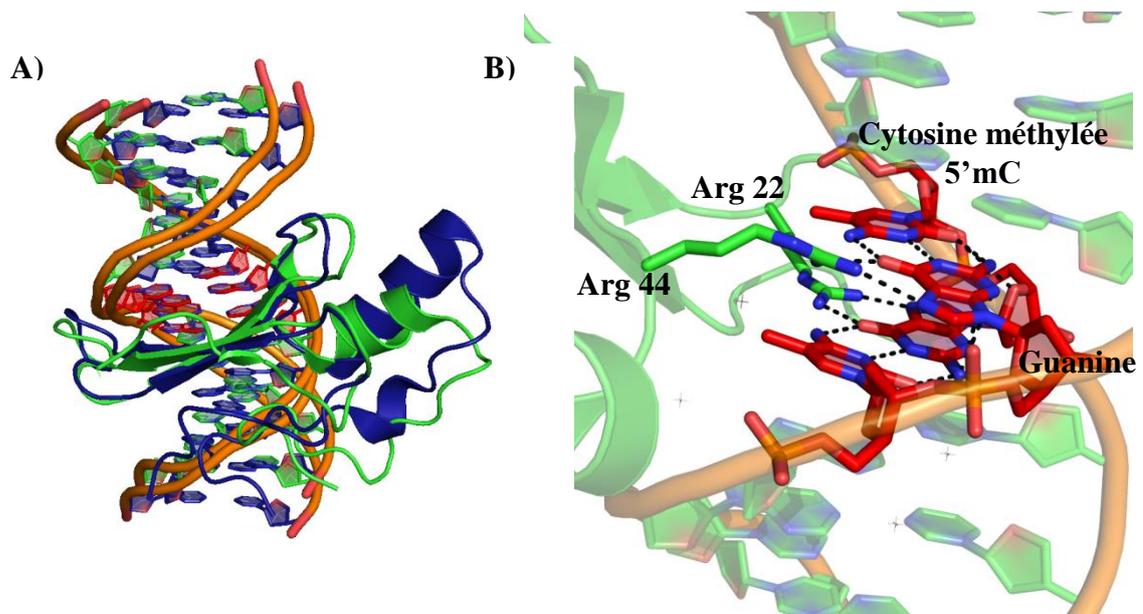
### (1)MeCP2

MeCP2 fut la première protéine MBP à être caractérisée et purifiée par Bird et ses collègues en 1989 (Meehan *et al.*, 1989). De façon surprenante, l'équipe alors intéressée par un mécanisme de protection des îlots CpG non méthylés contre les DNMTs, mis en évidence par des expériences de retard sur gel l'existence de protéine liant l'ADN méthylé. Cette protéine fut nommée MeCP2 (Methylated CpG-binding Protein) et la région responsable de l'interaction fut identifiée par la suite et qualifiée de MBD (Methyl-CpG binding domain). L'identification de ce domaine a permis par des recherches de similarité de séquence d'identifier les autres membres de cette famille (Nan *et al.*, 1993) (Hendrich and Bird, 1998a).

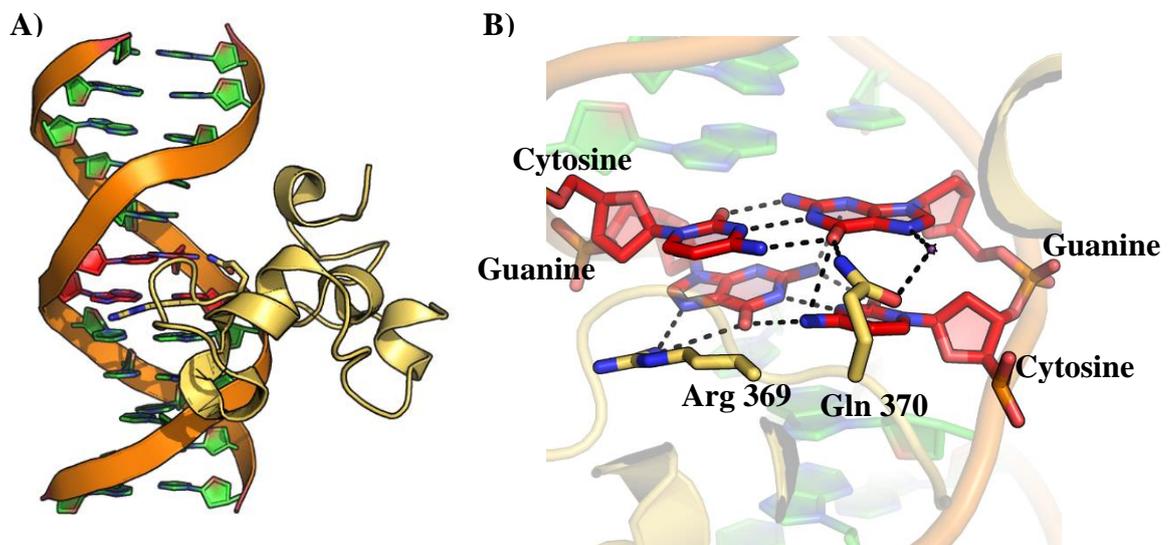
Il a été montré très tôt que l'expression de MeCP2 est dominante dans les tissus cérébraux, soulignant son rôle crucial dans le développement neuronal (Shahbazian *et al.*, 2002). Ainsi, des mutations de MeCP2 sont responsable d'un désordre neuro-développemental rare le syndrome de Rett. Ce syndrome, létal chez l'homme est à l'origine de graves désordres neurobiologiques et psychiatriques chez la femme (Amir *et al.*, 1999).

Comme la majeure partie des "reader" de la méthylation, MeCP2 est connu pour interagir avec des complexes corépresseurs de la transcription comme NCoR ou Sin3. De façon plus surprenante il a été démontré en 2008 que MeCP2 pouvait également être impliqué dans l'activation de la transcription par son association avec le complexe CREB (C-AMP Response Element-binding protein) (Chahrour *et al.*, 2008).

Trois structures cristallographiques de MeCP2 en complexe avec un ADN méthylé ont été publiées en 2008, 2016 et plus récemment en 2018. Ces structures (présenté en figure 6) montrent l'implication de deux triades dans l'interaction avec l'ADN de type : 5mC-Arginine-Guanine (Ho *et al.*, 2008) (Ho *et al.*, 2016). Ainsi deux Arginines, 111 et 133, sont au centre de l'interaction avec l'ADN et vont, de manière symétrique, réaliser des liaisons hydrogènes avec les guanines du dinucléotide mCpG et des contacts de type van der Waals avec les cytosines méthylées. Il est important ici de souligner l'importance des molécules d'eau, en effet cinq molécules d'eau sont impliquées par liaisons hydrogènes dans cette reconnaissance protéine-ADN.



**Figure 8 MBD1 et la reconnaissance de l'ADN méthylé** – A) superposition des domaines MBD de MBD1(en vert, pdb 6D1T) et MeCP2 (en bleu). On observe un léger changement d'orientation entre les deux domaines MBDs. B) La triade d'interaction du domaine MBD de MBD1 est caractérisée par l'absence de molécule d'eau impliquée chez MeCP2 dans des interactions indirectes avec la Tyr 123 (34 pour MBD1).



**Figure 9 Le domaine doigt de zinc de MBD1 et la reconnaissance de l'ADN non méthylé** – A) Organisation générale du domaine doigt de zinc CxxxC, en rouge est représenté le dinucléotide CpG non méthylé (pdb 5WQ9). B) Détail de l'interaction : on observe ici les deux résidus impliqués dans les liaisons hydrogènes. L'Arg 369 et la Gln 370 réalisent des liaisons hydrogène avec les bases guanines des dinucléotides CpG non-méthylés.

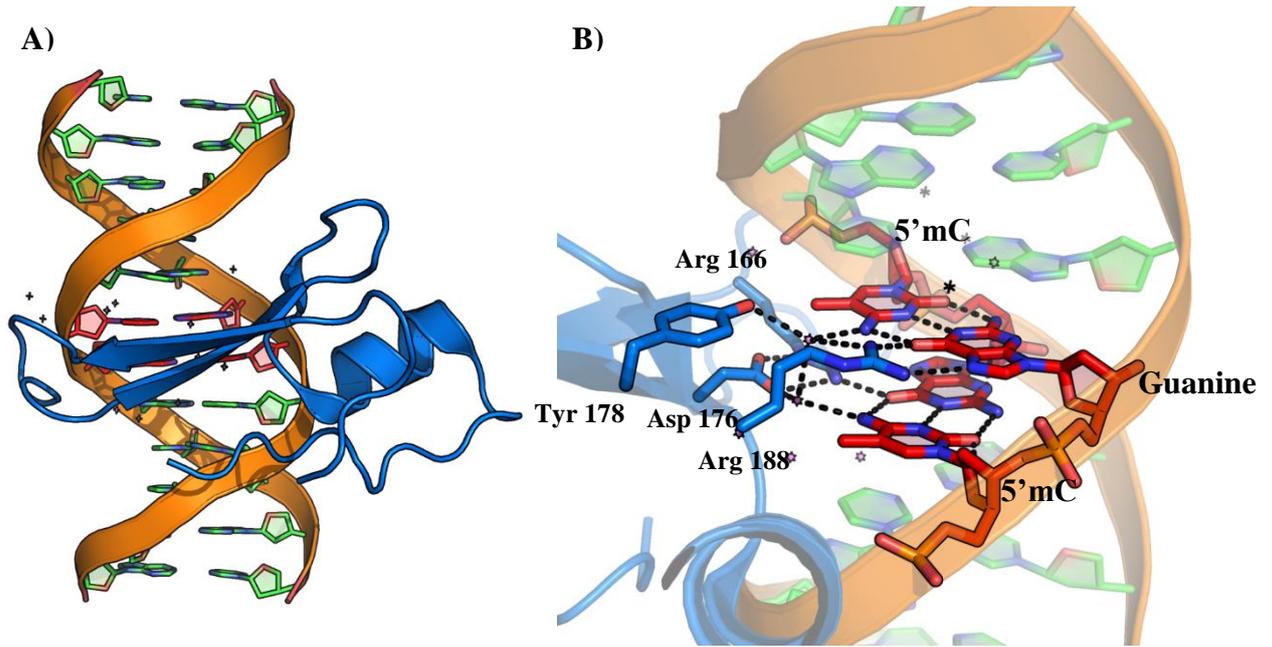
Néanmoins, la spécificité de MeCP2 pour l'ADN ne s'arrête pas au dinucléotide mCpG. En effet en 2013, Baker *et al.* ont mis en avant la présence de deux motifs conservés de type crochet AT (AT-hook) (Baker *et al.*, 2013). Ces deux séquences localisées dans une région désordonnée en région C-terminale aux positions 185-191 et 265-272 présentent le motif RGR (figure 7 B: GRGRGP et PKKRGRKP). Ces motifs vont interagir avec des régions riches en A/T de l'ADN. La présence de région A/T riche entraîne du fait d'un empilement de bases purines-purines le resserrement du petit sillon de l'ADN stabilisé par des molécules d'eau pour permettre le maintien de la forme B. Lors de l'interaction entre MeCP2 et l'ADN le crochet A/T va remplacer ces molécules d'eau, élargir le petit sillon et entraîner une courbure de l'ADN pour conserver la stabilité (Chia *et al.*, 2016) (Lyst *et al.*, 2016). Bien que d'un point de vu biophysique ces interactions sont relativement faibles, les deux motifs RGR sont extrêmement conservés au sein des espèces laissant penser qu'il s'agirait plus que d'un simple domaine secondaire de liaison. (Figure 7 A et B)

## (2) MBD1

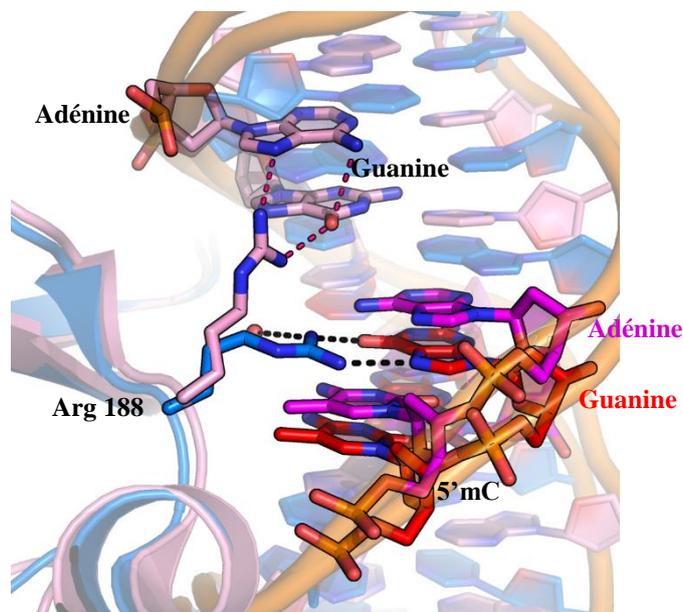
MBD1 est le plus grand représentant de cette famille. En complément du domaine commun MBD il possède également comme MeCP2 un domaine de répression TRD et deux à trois domaines à doigt de zinc CXXC dépendant de l'épissage (Figure 5).

La structure cristallographique du domaine MBD en complexe avec un ADN méthylé fut résolue en 2018 (Liu *et al.*, 2018). Il est intéressant de noter que cette interaction se fait de manière symétrique et identique. En effet deux arginines R22 et R44 vont chacune interagir avec un mCpG d'un des brins d'ADN (Figure 8). De façon générale cette interaction prend la forme d'une pince impliquant majoritairement deux triades comme chez MeCP2.

L'interaction de MBD1<sup>MBD</sup> avec l'ADN se fait de façon aspécifique et ne nécessite que la présence du dinucléotide mCpG. Il est néanmoins intéressant de noter que la présence en position +2 d'une adénine et thymine en position -2 vont augmenter cette affinité de 3 à 5 fois (Clouaire *et al.*, 2010). Ces acides aminés dans les régions adjacentes exercent une influence aussi sur la reconnaissance par MeCP2 pour lequel de façon opposé ils vont réduire l'affinité ; suggérant dès lors que les régions adjacentes joueraient un rôle discriminant dans la fixation des MBPs. La présence de domaines doigt de zinc va influencer grandement l'interaction avec l'ADN méthylé. Alors que le domaine MBD intervient dans la liaison aux CpG méthylés, leurs homologues non méthylés peuvent être spécifiquement reconnus par la famille de protéines à domaine CXXC (Voo *et al.*, 2000).



**Figure 10 MBD2 et la reconnaissance de l'ADN méthylé** – A) Organisation générale du domaine MDB de MBD2 (pdb 6CNP). B) La triade



**Figure 11 L'exclusion de l'arginine dans la triade de reconnaissance** - en rose est représenté l'arginine de MBD en interaction avec un ADN présentant le dinucléotide mCpA (pdb 6C1T) et en bleu avec un dinucléotide mCpG. On voit facilement que l'orientation de l'arginine en rose entraîne une perte de liaison hydrogène expliquant le caractère plus labile de l'interaction avec mCpA. L'arginine en rose va alors réaliser deux autres liaisons hydrogènes avec l'adénine et la guanine en amont du dinucléotide mCpA. La liaison avec la molécule d'eau à l'origine de liaisons hydrogènes indirectes avec les autres résidus impliqués dans la stabilisation du complexe est aussi perdue.

Ainsi la présence du troisième CXXC va entraîner une perte de la sélectivité pour l'ADN méthylé et permettre une répression de la transcription méthylation-indépendante (Jorgensen *et al.*, 2004). De façon générale, la reconnaissance de CpG par le domaine CXXC se fait majoritairement au travers de deux résidus (Q et R) par liaisons hydrogènes (Figure 9) (Xu *et al.*, 2018).

Un nombre croissant d'études ont montré que les protéines à domaine CXXC pourraient agir comme un module ciblant les îlots CpG. Comme pour les autres MBPs, MBD1 joue le rôle de plateforme à la fixation d'autres protéines (MBD1/HDAC3) ou complexes protéiques (MCAF1/MBD1/SETDB1) en vue de réguler la transcription. Ainsi MBD1 va indirectement être impliqué dans la méthylation de l'histone 3 (H3K9). Cette réaction relie deux modifications épigénétiques que sont la méthylation de l'ADN et celle des histones. Cette méthylation va se réaliser par le recrutement de MCAF1 (Activating transcription factor 7-interacting protein 1) et SETB1 (Histone-lysine N-methyltransferase) et ainsi permettre la condensation de la chromatine (Uchimura *et al.*, 2006) (Sarraf and Stancheva, 2004).

### (3) MBD2

Ce troisième représentant des MBPs est connu pour recruter le complexe de remodelage de la chromatine NuRD, une partie décrira ultérieurement MBD2 et MBD3 en contexte non-isolé au sein de NuRD.

Parmi toutes les MBPs, MBD2 est le représentant qui discrimine le plus fortement les dinucléotides mCpG des autres séquences génétiques (Zou *et al.*, 2012). C'est pour cette raison qu'il a été implémenté dans des kits commerciaux de détection de régions d'ADN méthylées (De Meyer *et al.*, 2013). Les spécificités de séquences de MBD2 pour différents ADN méthylés ou non sont le sujet d'un nombre croissant d'étude et les résultats sont souvent controversés. Ainsi, bien qu'il fût initialement décrit que ce domaine MBD possédait une affinité seulement pour les dinucléotides mCpG, Ke Liu et al en 2018 ont modifié entièrement la compréhension de cette reconnaissance. Par des expériences d'ITC ils ont tout d'abord mise en avant que la liaison se réalisait indépendamment des nucléotides encadrant directement le mCpG (Liu *et al.*, 2018a). De plus, ils ont découvert de nouvelles interactions entre les différents domaines MBD et les dinucléotides CpA et mCpA. En effet, bien que les méthylations de l'ADN se produisent principalement au niveau du dinucléotide CpG, 25% des cytosines méthylées des ESC et neurones sont impliquées dans des dinucléotides de type CpA, CpT ou CpC.

Ainsi, la structure cristallographique de hMBD2 en complexe avec mCpG a pu mettre en évidence l'implication des mêmes arginines conservées sous forme de deux triades. On peut reconnaître dans la figure 10 les interactions communes aux MBD que sont : les liaisons hydrogènes entre les arginines et les guanines, les interactions de type  $\pi$  et l'empilement de base CG.

La reconnaissance de MBD2<sup>MBD</sup> pour les dinucléotides mCpA et CpA présente de majeures différences structurales expliquant une interaction de plus faible intensité entre ces partenaires. Ainsi l'arginine 188 (correspondant à l'Arg 133 de MeCP2) impliquée dans une des triades ne peut établir de liaison hydrogène avec l'adénine et sera ainsi éloignée de l'interface de liaison. Il en résulte également une perte des interactions  $\pi$  entre cette arginine et la 5mC. (Figure 11) La reconnaissance ce fait non pas pour le dinucléotide CpA mais pour le dinucléotide complémentaire TpG. Bien que de faible intensité, ces interactions de type non mCpG pourraient avoir un rôle important dans la répression de l'expression de gène au court du développement embryonnaire et neuronal.

Le domaine MBD est suivi d'une région décrite comme intrinsèquement désordonnée. L'équivalent de cette zone chez MeCP2 correspond à la présence d'une séquence crochet AT qui n'est pas retrouvée ici. Néanmoins, cette région très conservée présente une large proportion d'acides aminés chargés positivement qui contribuerait probablement à une interaction dite « fuzzy » avec le squelette phosphate de l'ADN chargé négativement. Cette région sert aussi d'ancrage au cœur du complexe NuRD composé de MTA2, HDAC2 et RbAp48. Enfin le dernier domaine de MBD2 nommé CC pour coiled-coil, que l'on pourrait traduire par domaine enroulé est aussi impliqué dans le recrutement d'une sous-unité du complexe NuRD. En effet ce domaine va interagir avec un domaine homologue présent chez p66 $\alpha$ . Cette liaison va permettre in fine de recruter indirectement la sous-unité CH3/4 (Desai *et al.*, 2015).

Des études récentes ont montré que la protéine MBD2 était exprimée de façon ubiquitaire, avec des taux relativement plus élevé dans les poumons, le foie et le colon (Wood and Zhou, 2016). Des expériences de knockout des différentes protéines MBD chez la souris ont montré que la délétion de MBD2 n'impactait pas sur leur viabilité et leur fertilité à l'inverse de la délétion des autres membres de la famille (Hendrich et al., 2001). Ces informations sont pour le moment interprétées comme l'existence d'une redondance fonctionnelle entre les protéines MBD.

#### (4) MBD3

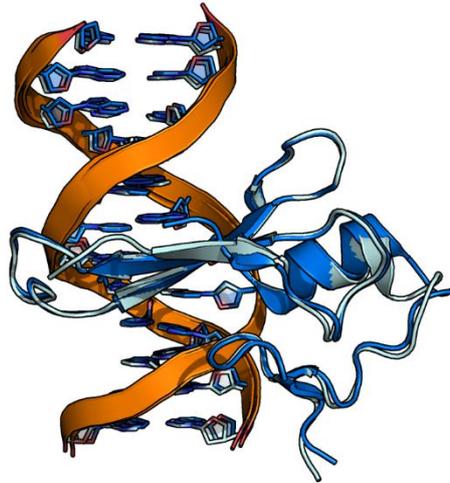
MBD3, présente une similarité de séquence importante pour MBD2, il n'est donc pas surprenant que cette protéine soit membre du complexe NuRD. Les deux représentants de la famille MBD au sein du complexe NuRD sont toutefois mutuellement exclusifs (Le Guezennec *et al.*, 2006).

Cette protéine bien que extrêmement étudiée reste le centre de nombreuses controverses notamment sur ses capacités de reconnaissances pour les dinucléotides CpG. En effet, il a été démontré très tôt que la protéine MBD3, chez les mammifères, a perdu sa capacité à créer des interactions de haute affinité avec l'ADN méthylé (Hendrich and Bird, 1998b) (Fraga *et al.*, 2003).

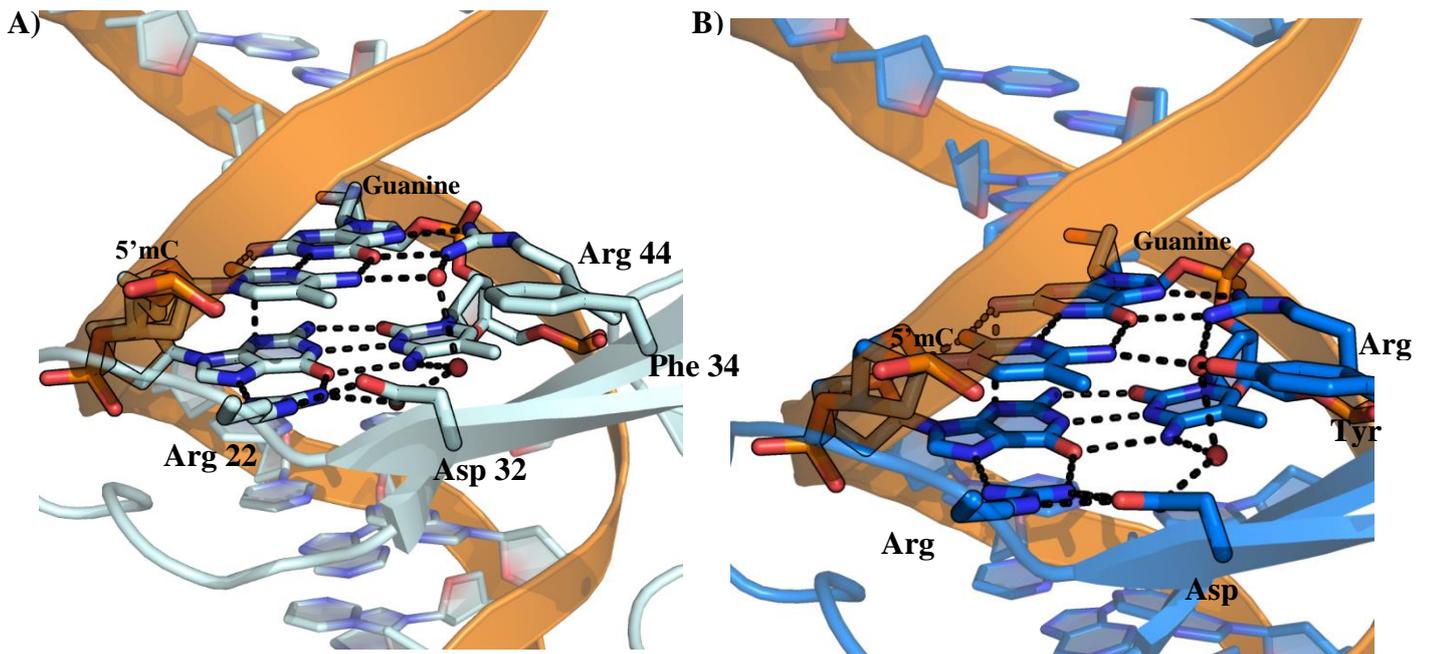
Cramer *et al.* en 2014 ont montré par des expériences de RMN que MBD3 présentait une faible préférence pour l'ADN méthylé, de façon néanmoins moins marqué que MBD2 (Cramer *et al.*, 2014). Leur expérience a montré également une absence totale de reconnaissance pour les 5mCpG oxydé, contrairement à un article aujourd'hui assez controversé sur la capacité de MBD3 à interagir de façon préférentielle avec les dinucléotides 5hmCpG (Yildirim *et al.*, 2011).

Des études de localisation génomique par CHIP sequence ont mise en évidence la localisation de MBD3 au niveau de région riches en CpG non méthylées et notamment les îlots CpG, alors que MBD2 est distribué le long du génome de façon méthyle-dépendante (Shimbo *et al.*, 2013) (Günther *et al.*, 2013). L'enrichissement de MBD3 au niveau d'îlots CpG présente un niveau extrêmement faible de modifications au niveau des cytosines. Ces données supportent le modèle selon lequel MBD3 chez les mammifères ne reconnaît pas la méthylation ou l'hydroxyméthylation de l'ADN (Baubec *et al.*, 2013).

La famille MBPs est intimement liée à la modification de la cytosine et MBD3 est un membre très intéressant de cette famille, étant un facteur de liaison au dinucléotides mCpG chez certains taxons, mais pas tous. Les différentes études de Chromatin ImmunoPrecipitation Sequencing (ChIP seq), ont soulevé une autre particularité imprévue spécifique de MBD3. En effet, cette dernière se retrouve au niveau du génome localisée dans différentes régions au niveau de promoteurs de gènes actifs (Günther *et al.*, 2013) caractérisés par la présence de l'histone H3K4me<sub>3</sub>, mais également dans des régions promotrice de gènes réprimés.



**Figure 12 MBD3 et la reconnaissance de l'ADN méthylé** – A) Organisation générale du domaine MDB de MBD3 en bleu clair superposé à celui de MBD2 en bleu cyan.



**Figure 13 Comparaison de la triade entre MBD3 et MBD2**– A) Le domaine de MBD3 en complexe avec mCpG (pdb 6CCG). B) Le domaine de MBD2 en complexe avec mCpG. On constate la même organisation générale de la triade avec chacune des arginines formant des liaisons hydrogènes avec chacune des guanines. Une Asp stabilisant l'Arg 22 ou 166. La majeure différence étant la présence d'une Phe à la place d'un résidu Tyr chez MBD3. Le résidu Tyr chez MBD2 est impliqué dans une liaison hydrogène indirecte avec la 5'mC par l'intermédiaire d'une molécule d'eau.

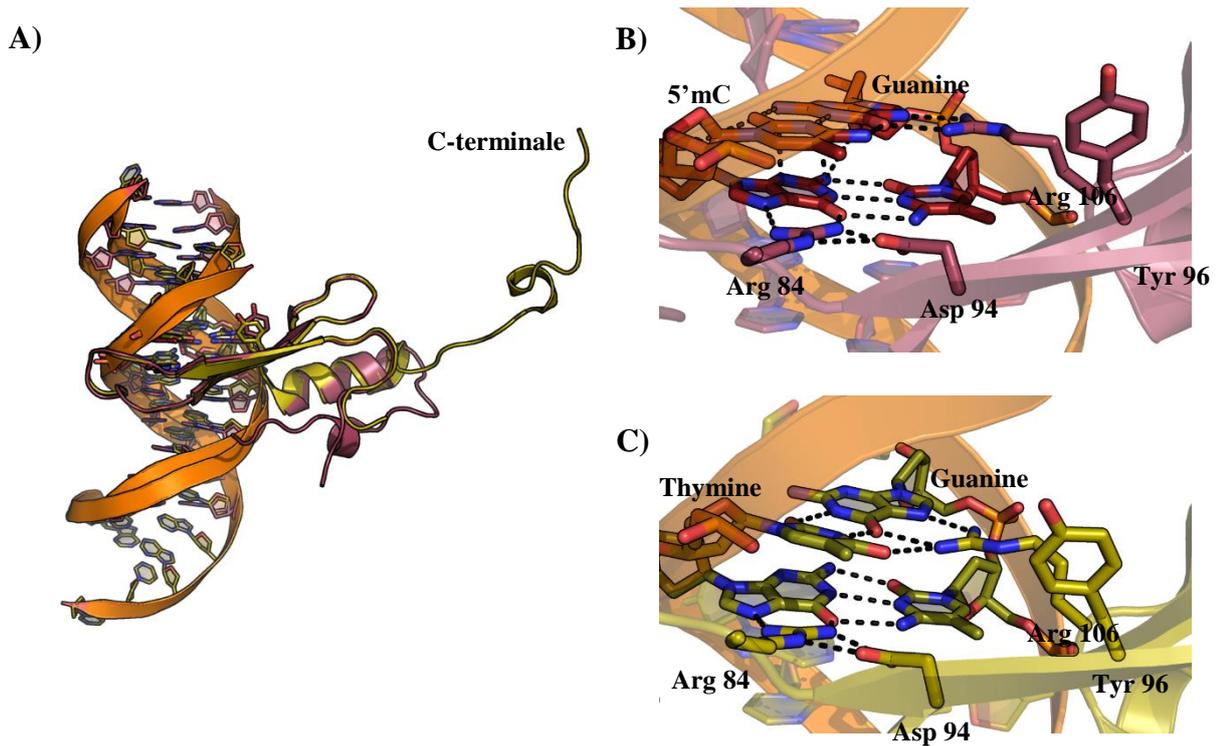
Cette année, Ke Liu *et al.*; ont également publié une étude sur les bases structurales des différents domaines MBD couplée à des études d'interactions par ITC. Leurs résultats pour le domaine MBD3 est en contradiction avec les travaux cités précédemment. Ainsi, ils ont démontrés que le domaine d'affinité de MBD3 pour l'ADN méthylé était 4 à 7 fois inférieur aux autres MBDs mais surtout ils n'ont identifié aucune capacité de fixation à l'ADN non modifié (Liu *et al.*, 2018).

Même après des décennies de recherche, les lecteurs de la méthylation de l'ADN cachent toujours leur véritable nature, d'où la nécessité d'étudier MBD3 et obtenir des informations supplémentaires sur le détail de la reconnaissance à l'ADN.

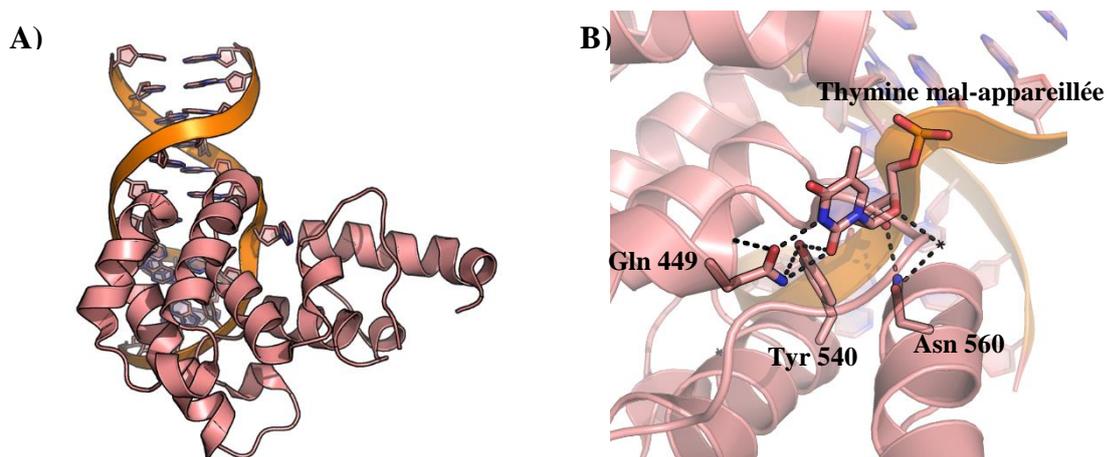
Deux mutations critiques pour la reconnaissance de l'ADN méthylé ont été décrites chez MBD3. Ainsi la tyrosine 178 chez MBD2, impliquée dans l'interaction avec la 5mC est absente chez MBD3 et remplacée par une phénylalanine en position 34. La seconde est une mutation de la lysine qui contacte la guanine +1 en histidine. Ces changements seraient en accord avec les résultats des essais biochimiques mettant en avant une perte de la spécificité pour les dinucléotides mCpG.

D'un point de vue structurale, la structure cristallographie obtenue par Ke Liu *et al.* a permis de poser les bases de la compréhension de ces deux mutations. Le domaine MBD de MBD3 comme ses partenaires présente une structure en sandwich  $\alpha/\beta$  composé en région N-terminale de 4 feuillets  $\beta$  en antiparallèles et une hélice  $\alpha$  en région C-terminale. Cette hélice est maintenue selon une orientation antiparallèle contre les feuillets  $\beta$  par des contacts hydrophobes. (Figure 12)

Ainsi comme on peut le voir sur la figure 13 l'arginine 22 (équivalent de l'arginine 166 chez MBD2) va réaliser deux liaisons hydrogènes avec un résidu guanine du dinucléotide mCpG. Comme pour MBD2, un résidu acide aspartique (Asp32) va stabiliser par liaison hydrogène la position favorable à cette interaction avec l'ADN. Mais il peut également réaliser une liaison hydrogène indirecte avec la cytosine méthylée faisant intervenir une molécule d'eau. La région de reconnaissance implique une dernière arginine caractéristique des liaisons des MBDs. Cette arginine (Arg44) équivalent de l'Arg188 chez MBD2 va en plus de pouvoir réaliser des liaisons hydrogènes avec la guanine, être capable d'interagir indirectement avec la 5mC avec l'intervention encore une fois d'une molécule d'eau. Les liaisons directes à la cytosine méthylée de MBD3 ne peuvent se réaliser du fait de l'absence de tyrosine.



**Figure 14 La triade MBD4** – A) superposition du domaine MBD de MBD4 avec un ADN mCpG en rose (pdb 4LG7) et un ADN avec un substrat désaminé TpG en jaune (pdb 3VXV). On observe une orientation différente de la partie C terminale. B) et C) comparaison des deux triades. Dans les deux structures, la tyrosine n'intervient pas dans la stabilisation de l'interaction ; Elle est expulsée du site de liaison. C) en présence du résidu thymine, le domaine MBD réalise une liaison hydrogène supplémentaire.



**Figure 15 Le domaine glycosylase de MBD4** – A) structure du domaine sous formes d'un paquet d'hélices. B) Détail des acides aminés stabilisant la rotation du dinucléotide mal-appareillé. (pdb 4OFA)

Pour stabiliser le complexe le niveau d'hydratation de la molécule semble jouer dès lors un rôle primordial. La comparaison des interactions MBD3-5mCpG et MBD2-5mCpG permet de comprendre du fait de la formation de liaisons supplémentaire le côté plus stable du complexe avec MBD2. Néanmoins, elle ne permet pas d'expliquer le mécanisme de reconnaissance de l'ADN méthylé ou non dans le cas de MBD3.

MBD3 est principalement exprimé dans les cellules du système nerveux lors du développement. Il a été montré qu'il possédait un rôle central dans la reprogrammation des cellules somatiques et la différenciation cellulaire (Kaji et al., 2007). Des expériences de knockout du gène *mbd 3* chez la souris, ont mis en évidence une létalité précoce au stade embryonnaire (Hendrich et al., 2001).

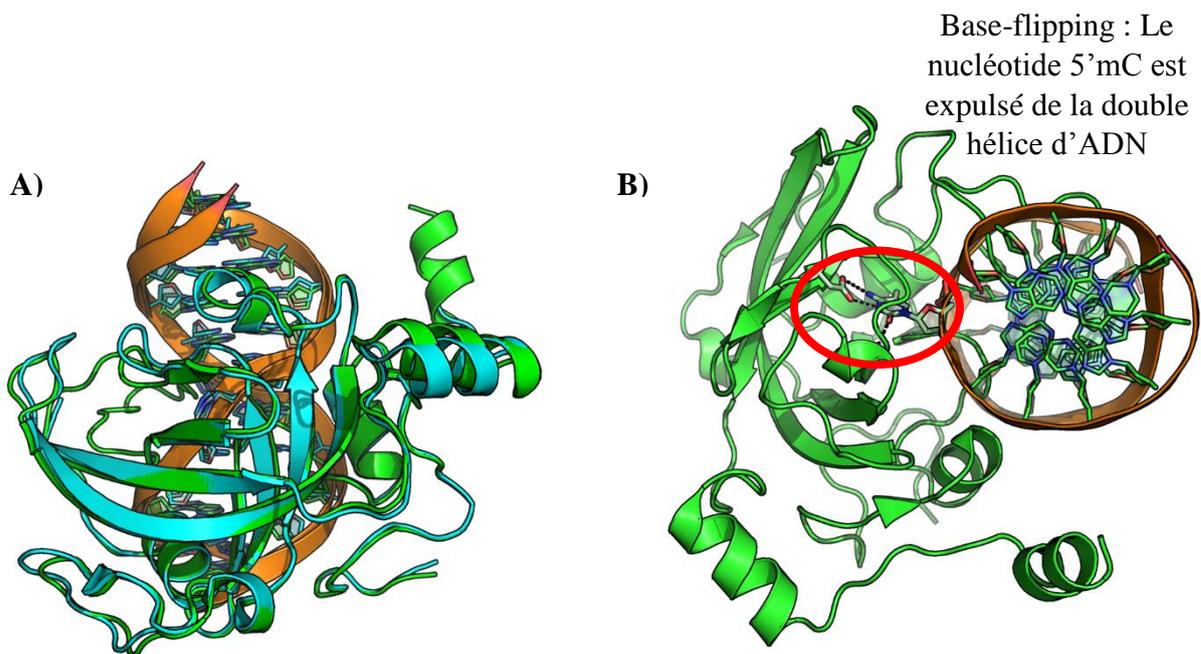
MBD2 et MBD3 sont impliqués dans le même complexe NuRD et possèdent une homologie de séquence importante mais ne posséderaient pas une activité redondante.

#### (5) MBD4

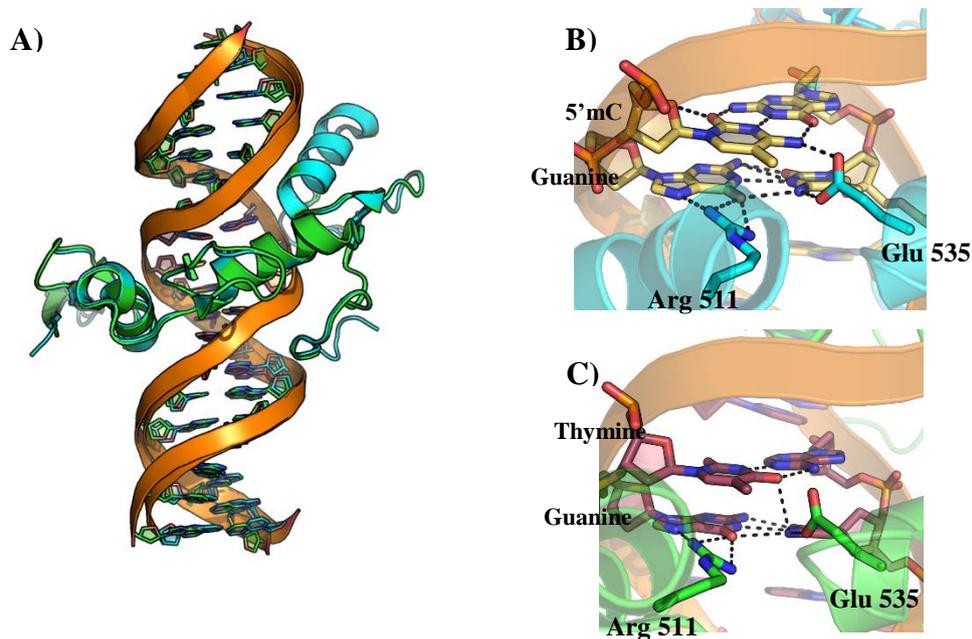
La protéine MBD4 est la seule MBPs non associée à une activité désacétylase. Elle contient de plus deux domaines fonctionnels ; le premier est le MBD domaine classique. Ce dernier en plus de présenter une haute affinité pour les dinucléotides mCpG possède la même affinité pour le substrat désaminé des mCpG le dinucléotide TpG des îlots CpG. La structure du domaine MBD en présence d'ADN méthylé a permis de mettre en évidence l'implication de la triade 5mC-Arg-G similaire aux autre MBPs (Figure 14) (Otani *et al.*, 2013).

Indépendamment de ce domaine MBD, MBD4 possède également un domaine glycosylase impliqué dans le processus de déméthylation similaire au TDG (Thymine DNA Glycosylase). Ce domaine va reconnaître les mésappariements de type mCpG/TpC ou mCpG/hmUpG et va pouvoir les réparer par rotation/excision du nucléotide (Hendrich *et al.*, 1999) (Hashimoto *et al.*, 2012) (Figure 15).

Le lien fonctionnel unissant ces deux domaines reste aujourd'hui non résolu et fait l'étude de nombreuses études. Mais du fait de son impact sur les réparations de mésappariements il est facilement compréhensible que toutes mutations au sein de MBD4 soit à l'origine de carcinomes (Figure 14 et 15) (Tricarico et al., 2015).



**Figure 16 Les protéines à domaine SRA** – (A) superposition des structures de UHRF1 (vert – 3F8I) et 2 (bleu clair – 4PW6) UHRF2-SRA se replie en une structure en forme de selle composée principalement de sept brins  $\beta$  et de quatre hélices  $\alpha$ . (B) Le base flipping



**Figure 17 Les protéines KAISO** – (A) superposition des structures de KAISO en interaction avec de l'ADN mCpG en bleu et TpG en vert. (B) et (C) comparaison du détail de l'interaction. On peut remarquer du fait de la présence de thymine la perte d'une liaison hydrogène avec l'arginine, expliquant la différence de spécificité.

## (6) MBD5 et 6

Ces deux dernières MBPs ne sont que partiellement caractérisées, et bien que leurs associations avec des désordres neuro-développementaux soient connues aucune spécificité pour l'ADN méthylé ne fut trouvée. Laissant penser à une fonction différente des autres membres de la famille (Hendrich and Tweedie, 2003) (Cukier *et al.*, 2012).

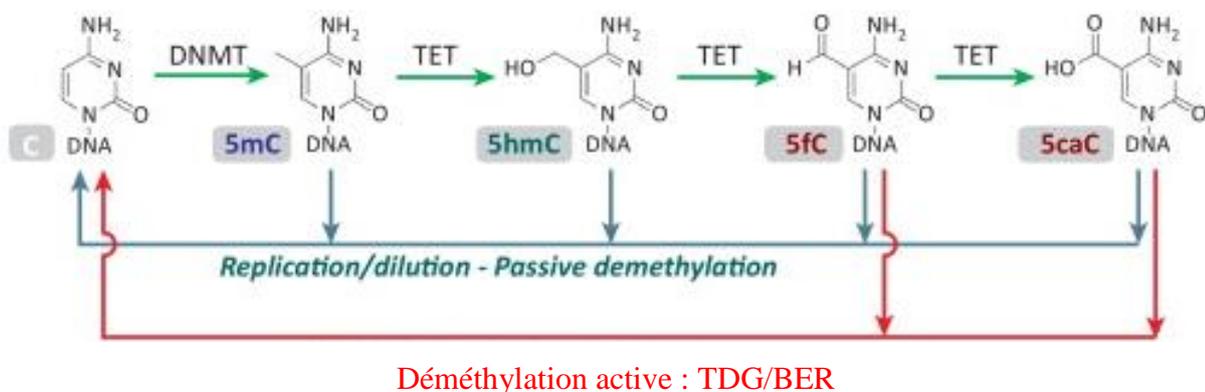
### b) *Le domaine SET- et RING-Associé : SRA*

Chez les mammifères on ne trouve que deux représentant de cette famille UHRF1 et UHRF2. Ces deux paralogues, comme on peut le voir dans la figure 16, présentent une architecture générale similaire et ce sont les seules protéines à posséder un domaine SRA chez l'Homme (Mori *et al.*, 2002).

UHRF1 possède un rôle fondamental dans le maintien du niveau de méthylation de l'ADN. En effet, son domaine SRA lui permet de lier l'ADN hémi-méthylé par un mécanisme nommé le base-flipping, similaire à celui retrouvé chez les DNMTs (Arita *et al.*, 2008) (Hashimoto *et al.*, 2008). Ainsi, l'interaction d'UHRF1 avec l'ADN hémi-méthylé, va permettre au résidu 5mC de pivoter hors de la double hélice et suite au recrutement de DNMT1 permettre la méthylation du brin complémentaire. SRA est le seul domaine identifié aujourd'hui capable de réaliser cette rotation sans activité enzymatique associée (Song *et al.*, 2012). Ainsi, le niveau d'expression d'UHRF1 est en lien direct avec le degré de méthylation et on peut facilement comprendre dès lors son implication dans le développement tumoral. En plus de sa capacité d'interaction avec l'ADN, UHRF1 est capable d'interagir avec l'histone 3 au niveau de la Lys 9 méthylée ou non et de l'Arg 2. UHRF2 présente le même mécanisme d'action que son paralogue, mais en 2014 Zhou *et al.* ont aussi identifié un mécanisme de « base-flipping » pour l'ADN hydroxyméthylé (Zhou *et al.*, 2014). Les deux protéines sont également exprimés de manière opposée, UHRF1 est principalement surexprimée dans les cellules souches embryonnaires puis réduite durant la différenciation (Bostick *et al.*, 2007).

### c) *La famille Kaiso*

La dernière famille de MBP a été découverte en 1999, par Egor Prokhortchouc *et al.* (Prokhortchouc *et al.*, 2001) L'équipe en effet démontra que cette protéine présentant trois domaines à doigt de zinc était capable de discriminer l'ADN méthylé du non-méthylé. Cette nouvelle famille fut par la suite agrandie par la découverte de deux paralogues Zbtb4 et Zbtb38, tous deux possédant 3 motifs en doigt de zinc, dont deux essentiels à la liaison de l'ADN méthylé (Sasai and Defossez, 2009).



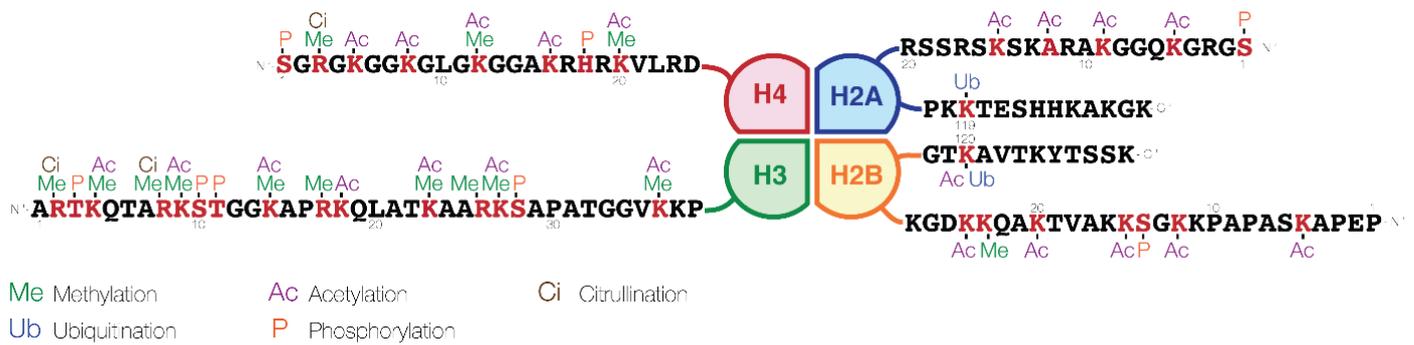
**Figure 18 Le cycle de déméthylation de l'ADN.** La méthylation et la déméthylation de l'ADN sont des processus dynamiques finement régulés. La méthylcytosine (5'mC) est générée par les ADN méthyltransférases (DNMTs), alors que la déméthylation peut être obtenue par de multiples voies. Les 5'mC peuvent être reconvertis passivement en cytosine par dilution dépendant de la réplication en raison d'un manque d'activité de la DNMT. Mais la 5'mC peut être également oxydée par les protéines de type TET en 5-hydroxyméthylcysotine (5'hmC), 5-formylcysotine (5'fC) et 5-carboxylcysotine (5'caC), pouvant être suivi d'une déméthylation passive de type réplication / dilution. Alternativement, la 5'fC et la 5'caC peuvent être éliminés par l'intervention de la thymine ADN glycosylase (TDG) et la réparation par excision de base (BER), résultant en une déméthylation active de 5'mC. *Adaptée de Potential functional roles of DNA demethylation intermediates Chun-Xiao Song and Chuan He, Trends in Biochemical Sciences October 2013.*

La protéine Kaiso est capable de recruter le complexe N-CoR et inhiber l'expression de la protéine MTA2, sous-unité du complexe NuRD (Yoon *et al.*, 2003). Néanmoins, la protéine est aussi connue pour s'associer avec p53 et stimuler l'expression de gènes impliqués dans l'apoptose (Koh *et al.*, 2014). Ainsi on peut comprendre toute la subtilité dans le rôle de ces protéines « readers » dans le sens où elles peuvent être impliquées dans l'activation ou la suppression de l'expression de gènes, de façon partenaire-dépendante. La protéine Kaiso présente une affinité pour la présence de deux CpG méthylés consécutifs mais également pour une séquence d'ADN non-méthylée nommée KBS (Kaiso Binding Site) (Prokhortchouk *et al.*, 2001). Cette séquence CTGCNA présente un TpG, structuralement semblable à un CpG. En effet, les structures cristallographiques de KAISO en complexe avec un ADN méthylé et la séquence KBS présentent une similarité générale manifestant un mode de reconnaissance de l'ADN quasi-similaire (figure 17) (Buck-Koehntop *et al.*, 2012).

### 3. La déméthylation

La méthylation de la cytosine se faisant par formation d'une liaison covalent C-C, le résidu modifié est chimiquement stable. Malgré cette évidente stabilité, l'équilibre méthylation-déméthylation est hautement dynamique et dépendant du contexte cellulaire.

Chez les plantes, la déméthylation implique l'association de différentes DNA glycosylases, ROS1, DME, qui vont agir par un mécanisme direct d'excision et de restauration du résidu cytosine. Les mammifères ne possédant pas d'ADN glycosylases vont pouvoir réguler le niveau de méthylation de deux façons distinctes. D'une part, la modification passive ; au cours de la réplication la méthylation du brin néo-synthétisé est sous la gouvernance de DNMT1 responsable de copier le modèle de méthylation du brin parental. En cas de non-méthylation du brin néo-synthétisé, la méthylation sera perdue au cours des réplifications successives (Ambrosi *et al.*, 2017). D'autre part, la modification active, se traduit par l'oxydation itérative du 5mC comme présenté en (Figure 18) par la famille de protéines Ten-Eleven Translocation (TET) (Tahiliani *et al.*, 2009) (Wu and Zhang, 2017). Par la suite deux éliminations de ces états d'oxydation peuvent avoir lieu. La dilution passive qui comme pour la modification passive sera réplication-dépendante. Ainsi les modifications apportées ne permettront pas la transmission aux cellules filles des méthylations, DNMT1 étant moins efficaces au niveau des cytosines hémi-oxydées.



**Figure 19 Modifications post-traductionnelles des histones.** Les nucléosomes sont représentés par des sphères. Les positions des MPT sont principalement retrouvées sur les queues des histones à l'extrémité N-terminale, au niveau des résidus Lys, Arg et Ser. Ces MPT ont un impact sur l'expression des gènes en modifiant la structure de la chromatine et en servant d'ancrage à d'autres protéines. Elles sont associées à diverses fonctions biologiques telles que l'activation et l'inactivation de la transcription, l'empaquetage des chromosomes et les processus de réparation et de dégradation de l'ADN.

Enfin, la suppression des cytosines modifiées peut se faire par un procédé proche de celui trouvé chez les plantes. Celle-ci qualifiée de suppression active, implique l'intervention d'une TDG qui va reconnaître et exciser les états oxydés de type 5fC et 5caC, créant ainsi un site abasique avant la restauration de la cytosine non-modifiée par le BER. (He *et al.*, 2011b)

## B. Les modifications post-traductionnelles des histones

Alors que la méthylation est comme on a pu le voir précédemment une marque de répression stable pouvant s'étendre sur de longues régions chromosomiques, les MPT possède un caractère plus labile et permettent une régulation de l'expression génique à court terme.

Les histones sont la cible de modifications épigénétiques variées tels :

- La ribosylation,
- L'ubiquitination et la sumoylation,
- Les phosphorylations des serines et des thréonines,
- L'acétylation et la méthylation des lysines et des arginines.

Aujourd'hui près de 160 sites de modifications ont été décrits, au niveau des extrémités amino-terminales des histones.

Dans cette partie introductive, je ne présenterai que les MPT des lysines.

Au niveau des queues d'histones, les résidus lysines sont chargés positivement et sont la cible de nombreuses réactions (Figure 19). Ces charges positives vont permettre l'interaction entre les histones et le squelette phosphate de l'ADN chargé négativement. Ainsi toute modification covalente comme l'acétylation ou la méthylation va altérer ces contacts et *in fine* la structure de la chromatine, régulant l'accès aux gènes.

### 1. L'acétylation des lysines

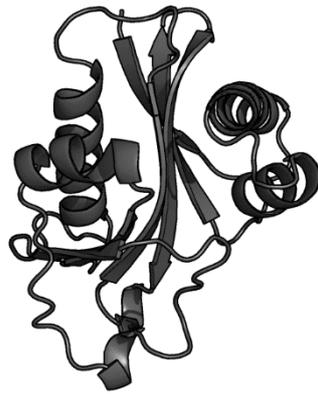
Parmi les modifications de la chromatine, l'acétylation est catalysée par une famille d'enzyme nommées les acetyl-transferases. Celles-ci ont la capacité de transférer un groupement acétyle provenant du coenzyme A au groupe  $\epsilon$ -amine à l'extrémité de la chaîne latérale de la lysine des queues des histones. Ces modifications sont généralement associées à l'activation de la transcription des gènes. En effet, elles vont fournir un site de fixation aux protéines possédant un bromo-domaine, hautement présent chez les protéines nucléaires interagissant avec la chromatine. L'acétylation est une modification facilement réversible, à l'aide d'histones



**HAT1 (1bob)**



**Esa (1fv7)**



**Gcn5 (1vgh)**

**Figure 20 Les HATs.** Morphologie générale de différents membres de chaque famille.

desacétylases. Ainsi la cellule possède du fait de l'action concomitante de ces deux enzymes un système de régulation de la compaction et donc de l'expression génique.

#### a) HAT/KAT: acetyltransferase enzyme

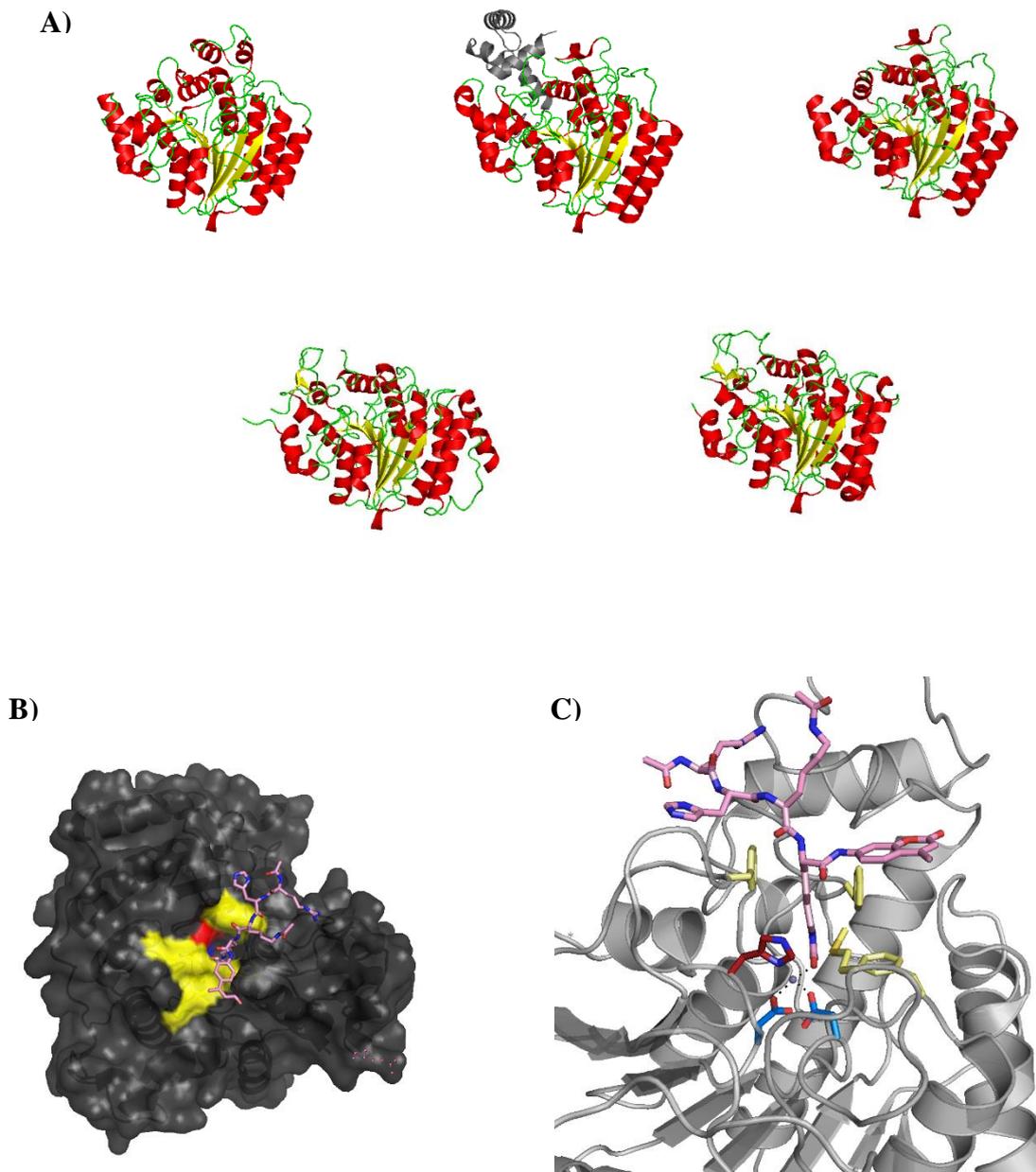
Bien que l'impact de l'acétylation des histones sur la transcription fût découvert dans les années 60, l'aspect enzymatique et l'importance des histones acétyl transférase (HAT) dans ces modifications ne furent décrits que plus de 30 ans plus tard (Allfrey and Mirsky, 1964) (Brownell *et al.*, 1996).

A ce jour, 19 acétyltransférases ont été identifiées chez les mammifères et classées en deux familles selon leur localisation cellulaire. Ainsi furent caractérisées les HAT-B cytoplasmiques, catalysant l'acétylation d'histones néo-synthétisées qui seront par la suite importées dans le noyau et les HAT-A nucléaires. Cette première classification fut supplantée par une autre plus précise se basant sur les similarités de séquences. Chacune de ces trois nouvelles familles a montré avoir des préférences pour des substrats différents, et être ainsi impliqué dans des contextes fonctionnels précis (Figure 20) (Kuo and Allis, 1998) (Yuan and Marmorstein, 2013) (Javaid and Choi, 2017).

La famille GNAT (pour N-acetyltransferases related to GCN5) – Ces HATs, vont acétyler préférentiellement les lysines de l'histone H3 en position 14 (K14) et dans une moindre mesure les lysines K8 et K16 de l'histone 4. Les GNATs sont connues pour interagir avec des complexes protéiques activateurs de la transcription. Parmi elle, on peut citer GCN5 qui fait partie de nombreux complexes tels que SAGA. De plus, GCN5 et PCAF ont la capacité de catalyser la propionylation de l'histone H3, mais ne peuvent pas, à l'inverse de la famille P300, hydrolyser des acyl-coA plus long.

La famille P300/CBP – les éléments de cette famille sont considérés comme des régulateurs globaux de la transcription. Ils sont également impliqués dans des réactions variées comme la propionylation, la crotonylation ou encore la butyrylation de la chromatine (Kaczmarska *et al.*, 2017). Les représentants de cette famille sont connus pour leur interaction avec les récepteurs nucléaires (Leo and Chen, 2000).

La dernière famille d'HATs est nommée MYST. Elle tient son nom des différents représentants de cette famille MOZ, Ybf2/Sas3, Sas2 and Tip60. (Marmorstein and Trievel, 2009) le substrat préférentiel de cette famille est l'histone H4. Parmi les enzymes de cette famille, Esa1 possède en plus de la capacité d'acétylation, celle de propionylation l'histone H4 et est également trouvée au sein du complexe NuA4. (Allard *et al.*, 1999)



**Figure 21 Les HDACs.** A) similarités de structures des protéines HDAC de classe I, II et IV : présence d'un repliement  $\alpha/\beta$  (rouge) ainsi que d'un feuillet  $\beta$  de 8 brins parallèles (jaune). B) le tube lipophile, site actif où la lysine acétylée va interagir (pdb 2V5W). On observe la présence d'acides aminés non-polaires (jaune) formant le tube ; au fond duquel 3 résidus chargés (bleu=négatif et rouge=positif) vont permettre la stabilisation du complexe par liaison hydrogènes faisant intervenir un atome de zinc. C) détail des interactions impliquant l'ion de zinc au sein du tube. Le substrat ici est un peptide de type Arg-His-Lys( $\epsilon$ -acétyle)-Lys( $\epsilon$ -acétyle) caractéristique du suppresseur de tumeur p53.

Malgré leurs divergences de séquences, toutes les protéines présentent une grande homologie structurale pour leur domaine HAT. Comme on peut le voir en figure 20, ce domaine est composé de trois feuillets  $\beta$  antiparallèles suivis d'une hélice  $\alpha$  sous les brins  $\beta$  pour Esa1, Gcn5, HAT1. On retrouve également au cœur de ce domaine le cofacteur acétyl-CoA nécessaire à la réaction d'acétylation. Néanmoins, les régions adjacentes de ce domaine sont structurellement divergentes donnant lieu à la divergence des fonctions régulatrices de ces protéines (Yan et al., 2000) (Kaczmarska *et al.*, 2017).

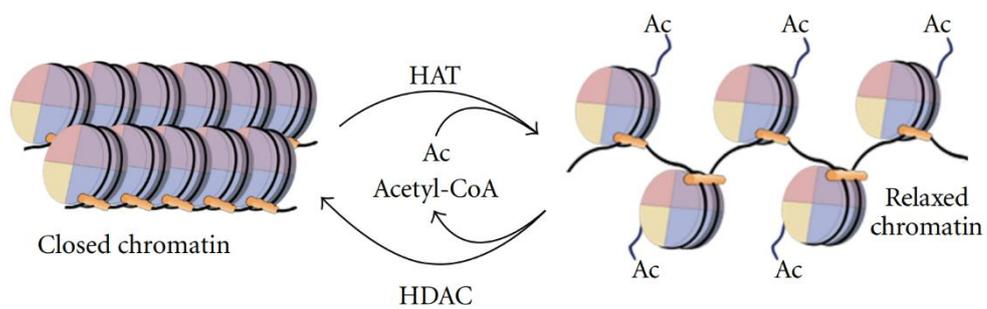
### a) HDAC: deacetylase activity

L'activité de désacétylation a été découverte en 1969 par Inoue et Fujimoto, par la suite de nombreuses études biochimiques ont permis de caractériser précisément cette famille de protéines. Aujourd'hui, 18 différentes HDAC ont été décrites et classées en deux superfamilles et quatre classes selon leur similarités de séquence (Inoue and Fujimoto, 1969) (Seto and Yoshida, 2014).

#### (1) Superfamille des Arginase/Desacetylase

Cette première superfamille est composée des HDAC de classe I, II et IV. En effet, en 1997, Leipe et Landman ont proposé que les HDACs eucaryotes aient pour origine un ancêtre commun procaryote similaire à l'acétyle polyamine amidohydrolase capable de cibler et désacétyler un groupement aminoalkyl d'une composé lié à l'ADN permettant *in fine* la régulation de l'expression de gène (Leipe and Landsman, 1997). Ces informations laissent penser que le substrat primaire de ces HDACs pourrait ne pas être des histones. Les protéines de cette superfamille proviennent non seulement d'un ancêtre commun mais présentent également une similarité de structure et de fonction (Gregoretto et al., 2004).

La classe I de cette superfamille est composée des HDAC1, 2, 3 et 8. Elle a été nommée « Rpd3-like » du fait de la similarité de séquence avec la protéine Rpd3 trouvée chez *Saccharomyces cerevisiae*, et possèdent entre 400 et 500 résidus (Yang et al., 1996). Ces protéines sont exprimées de façon ubiquitaire mais exclusivement dans le noyau cellulaire. Ces quatre enzymes présentent toutes une activité caractéristique de désacétylation néanmoins, cette activité est faible en l'absence d'association avec d'autres protéines. Ces enzymes sont donc majoritairement retrouvée au sein de complexes protéiques permettant de réguler leur activité mais également de les localiser au niveau du site cible. Ainsi HDAC 1 et 2 sont retrouvées dans



**Figure 22 Régulation de la compaction de la chromatine ; équilibre entre HAT et HDAC.** *Adapté de Current and Emerging Treatment Strategies for Cutaneous T-cell Lymphoma, Frederick Lansigan et. al. Drugs, February 2010.*

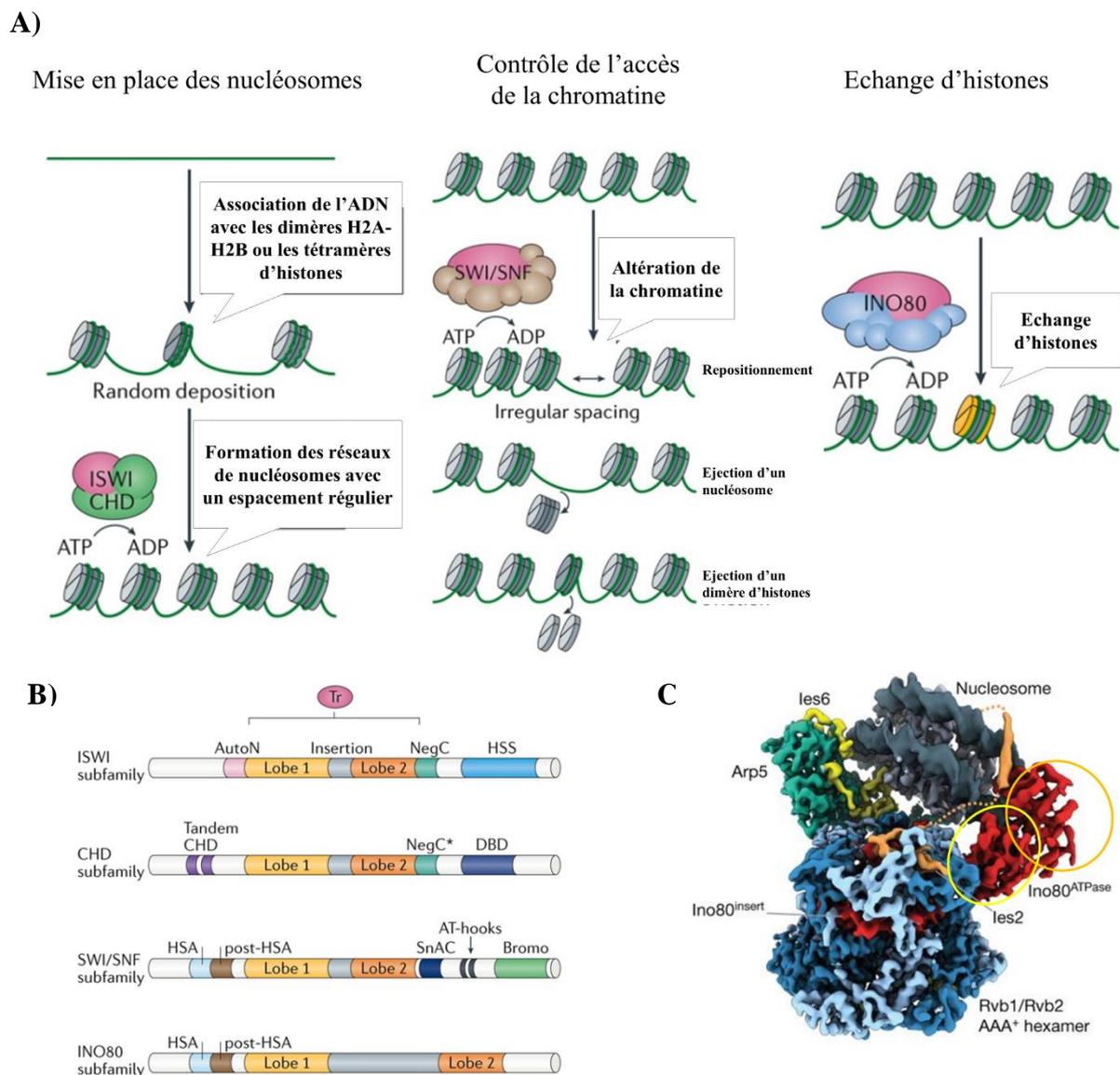
les complexes NuRD, Sin3, CoREST alors que HDAC 3 est spécifique des complexes SMRT et NCoR (Ayer, 1999) (Wen et al., 2000).

Les HDAC 4-7 et 9 appartiennent quant à eux à la classe II. Les enzymes de cette classe présente une similarité avec la protéine Hda1 découverte chez *Saccharomyces cerevisiae*. (Fischer et al., 2002) (Grozinger et al., 1999) Ainsi, bien que le cœur catalytique soit structuralement proche de celui des enzymes de classe I, il est important de noter la présence d'un domaine supplémentaire unique. Au-delà de cette dissimilarité de séquence, cette classe est exprimée de manière tissu-spécifique dans le cerveau, le cœur et le muscle squelettique sous forme de protéines cytoplasmiques puis recrutés dans le noyau ultérieurement. Parmi cette classe on discerne également deux sous-classes dont la sous-classe Iib comprenant HDAC6 et 10 qui possèdent un second site catalytique qui semble fonctionner de façon indépendante. (Kao et al., 2002)

La dernière classe nommée classe IV ne comprend qu'un seul représentant HDAC11 qui partage des similarités avec les deux premières classes présentées. C'est également l'une des dernières HDAC identifiée en 2002 and elle n'a été que faiblement caractérisée (Gao et al., 2002).

La première structure par cristallographie aux rayons X d'une protéine de type HDAC (HDLP) obtenue en 1999 fournit le premier aperçu du mécanisme moléculaire des désacétylases (Finnin et al., 1999). Comme on peut le voir sur la figure 21, la poche du site actif est extrêmement conservée au sein de la classe I, II et IV. De plus, toutes les enzymes de ces classes présentent également une topologie similaire composée d'un repliement  $\alpha/\beta$  ainsi que d'un feuillet  $\beta$  de 8 brins parallèles (Lauffer et al., 2013) (Vannini et al., 2007) (Dowling et al., 2008) (Schuetz et al., 2008).

Du fait de leur similarité structurale il est facilement compréhensible que ces trois classes partagent le même mécanisme enzymatique dépendant de la présence de zinc. (Figure 22) Elles vont en effet, catalyser l'hydrolyse du groupement acétyle de l'acétyle-lysine de protéines histones ou non libérant ainsi un composé acétate et lysine. Au sein du nucléosome, cette modification va altérer la structure de la chromatine et réprimer la transcription. De plus, l'action des HDAC va rendre les résidus lysine disponible à d'autres modifications telles que la méthylation, l'ubiquitination, la sumoylation etc... Le rôle de ces enzymes est donc également de réguler indirectement ces modifications (Tan et al., 2011).



**Figure 23 Les remodeleurs et leurs mécanismes d'action.** (A) Représentation des différents mécanismes de régulation de la chromatine par les remodeleurs ATP-dépendant. Le premier mécanisme caractéristique des sous familles ISWI et CHD, permet la mise en place des histones de manière à positionner les nucléosomes régulièrement. SWI/SNF sont connus pour posséder de nombreuses fonctions permettant de réguler l'accès à la chromatine. Enfin INO80 est connu pour son rôle dans l'échange d'histones au sein de nucléosomes déjà formés. Cette classification des remodeleurs est simplifiée ; en effet, certains membres des familles INO80, ISWI et CHD possèdent également un rôle dans l'accès à la chromatine. (B) Représentation de chaque sous-unité représentative de chaque sous-famille avec un cœur catalytique composé de deux lobes séparés par une région d'insertion, ce domaine est nommé Tr pour domaine ATPase-translocase. *Adapté de Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes Clapier C. et al. Nature Reviews Molecular Cell Biology (2017).* (C) Structure du cœur du complexe INO80 en jaune et orange sont entourés les deux lobes de la sous-unité ATPase interagissant avec le nucléosome. *Adaptée de Structural basis for ATP-dependent chromatin remodelling by the INO80 complex, Eustermann et al. Nature (2018).*

### (1) Deoxyhypusine synthase like NAD/FAD-binding domain superfamily

Cette superfamille fut elle aussi créée en raison de la similarité de ses représentants avec une protéine découverte chez *Saccharomyces cerevisiae* Sir2 impliquée dans la répression de la transcription et conservée de la bactérie à l'Homme (Brachmann et al., 1995). Les enzymes de cette famille sont une classe d'enzyme donc l'action de désacétylation dépend d'un cofacteur NAD. Ces enzymes restent encore peu connues par rapport aux autres HDAC, mais présentent une caractéristique remarquable, à savoir la présence d'une seconde activité enzymatique mono-ADP-ribosyltransférase (Frye, 1999).

#### 2. Méthylation des arginine et lysine

Les PRMT forment un groupe de protéines responsable de l'une des modifications post-traductionnelles majeure des queues d'histones avec la phosphorylation ou encore l'ubiquitination. Il existe aujourd'hui 9 PRMTs identifiées capable de méthyler l'arginine par une réaction proche des méthyltransférases avec l'intervention d'un cofacteur SAM. Les queues des histones H3 et H4 possèdent plusieurs arginines, en fonction du résidu substrat des PRMT, l'impact sur la régulation des gènes sera différent. Ainsi, on peut noter que la méthylation de l'arginine 3 de H3 (H3R3me2a) conduira à la décompaction de la chromatine, celle de l'arginine 2 (H3R2me2a) est associée avec sa répression. Cette famille d'enzyme possède aussi la capacité de méthyler des protéines non-histones impliquée dans la transcription, la signalisation cellulaire, la stabilité des protéines etc...

### III. Les re-modeleurs de la chromatine ATP-dépendant

Comme nous venons de le voir précédemment, la chromatine est une structure extrêmement dynamique. Toutes les modifications présentes sur la chromatine (ADN et/ou histones) vont pouvoir servir de signalisation pour initier la régulation de l'expression des gènes. En effet ils vont permettre de recruter des complexes de remodelage de la chromatine. L'action de ces derniers nécessite de l'énergie utilisée sous forme d'ATP ; d'où leur nom de complexe de remodelage ATP-dépendant. Ils sont organisés en différentes familles mais tous possèdent une fonction ATPase translocase à l'origine du déplacement des histones, leur éjection ou encore l'incorporation de variant d'histones (Figure 23 A).

Cette superfamille est subdivisée en 4 groupes en fonction des similarités et différences de leur domaine ATPase. On distingue ainsi :

- SWI/SNF switch/sucrose non-fermentable
- ISWI imitation switch
- INO80
- CHD chromodomain helicase DNA-binding

Pour tous ses complexes macromoléculaires, une sous-unité est responsable de l'activité catalytique de remodelage, et l'organisation de ce domaine translocase est similairement composé de deux lobes RecA-likes séparés par une région d'insertion plus ou moins longue (Figure 23 B).

#### A. SWI/SNF

Le complexe SWI/SNF est le principal remodeleur à l'origine de la régulation de l'accès à la chromatine. Il va ainsi modifier la chromatine en repositionnant les nucléosomes, en éjectant les octamères, tétramères ou dimères d'histones. Ce complexe va donc favoriser l'expression des gènes en ouvrant la chromatine. Il va permettre ainsi d'exposer les sites de liaison des activateurs de la transcription (Lorch et al., 1999) (Whitehouse et al., 1999).

Le mécanisme d'action de ce remodeleur semble être régulé par les modifications des queues des histones qui pourraient être reconnues par une des nombreuses sous-unités du complexe (Corona et al., 2002).

#### B. ISWI

La sous-famille ISWI pour Imitation Switch, rassemble les complexes possédant une protéine ISWI comme sous-unité ATPase. Cette protéine est caractérisé par deux domaines de liaisons ; le domaine SANT spécifique des queues des histones modifiées et un SLIDE pour les interactions avec l'ADN nucléosomique. Chez les mammifères, deux homologues ont pu être identifié ; SNF2L et SNF2H. A eux deux ils sont impliqués dans pas moins de 8 différents complexes de remodelage (Strohner *et al.*, 2001) (Hakimi *et al.*, 2002) (Banting *et al.*, 2005).

Les complexes ISWI, comme ceux de la famille CHD, sont appelés assembleurs de nucléosomes. En effet, ils vont participer à la déposition des histones, la maturation des nucléosomes et le contrôle de leur espacement. Ce mécanisme de remodelage des nucléosome

va permettre la régulation de l'expression des gènes et des réparations de l'ADN par le biais de NER. Deux fonctions ont ainsi été décrites :

- La maturation des pré-nucléosomes, permettant l'organisation des complexes ADN-histones en nucléosome octamériques.
- La formation des réseaux de nucléosomes espacés à distance relativement fixe (Torigoe *et al.*, 2011).

Ce processus d'assemblage et d'espacement a également lieu pendant la transcription aux endroits où les nucléosomes ont été éjectés de manière dynamique.

Sur le plan fonctionnel, la plupart des complexes des sous-familles de l'ISWI assemblent et espacent régulièrement les nucléosomes pour limiter l'accessibilité de la chromatine et l'expression des gènes; cependant, un des représentants de cette famille (NURF), possède des sous-unités accessoires rendant accessible l'ADN et favorisant la transcription. (Xiao *et al.*, 2001)

### C. INO80/SwRI

INO80 est un complexe macromoléculaire de plusieurs mégadaltons, composé de 15 sous-unités et impliqué dans la régulation de l'expression génique, la réparation de l'ADN et enfin dans la réplication. (Shen *et al.*, 2000) Cette sous-famille est aussi connue sous le nom de « éditeur de nucléosome », il permet en effet le déplacement des nucléosomes et catalyse l'échange du variant d'histones H2A.Z avec H2A. Bien qu'extrêmement étudié biochimiquement et structurellement parlant, ce mécanisme n'est pas encore précisément décrit (Clapier *et al.*, 2017) (Tosi *et al.*, 2013) (Brahma *et al.*, 2017).

La première structure par Cryo-EM du complexe de *Chaetomium thermophilum* en interaction avec le nucléosome fut publiée cette année par Sebastian Eustermann *et al.* (Eustermann *et al.*, 2018) Elle a permis de poser les bases de la compréhension structurale de la reconnaissance de la chromatine par le cœur d'un complexe remodelleur (Figure 23 C).

#### IV. CHD/Mi2: NuRD

Le complexe NuRD (pour Nucleosome Remodeling and histone Deacetylation complex), est un complexe protéique associé au remodelage de la chromatine, hautement conservé. Il fut purifié et caractérisé pour la première fois vers la fin des années 90 à partir de cellules de *Xenopus laevis* et humaines par différents groupes (Zhang et al., 1998) (Wade et al., 1998) (Tong et al., 1998).

Le complexe est composé de 6 sous-unités principales, chacune d'entre elles possédant plusieurs isoformes : HDAC1/2, MTA1/2/3, RBBP4/7, p66alpha/beta, MBD2/3, CHD3/4 (Allen et al., 2013). Il possède la capacité unique de présenter deux fonctions enzymatiques différentes grâce aux sous-unités HDAC (desacétylation) et CHD (remodelage de la chromatine). En raison de la présence de HDAC, associée à la répression de l'expression de gènes, le complexe NuRD a longtemps été considéré comme un complexe répresseur de la transcription. De plus, la protéine MBD2 sert d'ancrage au complexe au niveau des îlots CpG méthylés, induisant la répression (Baubec et al., 2013). Cependant, des études récentes sur le profil d'interaction génomique de NuRD ont révélé que certains complexes NuRD, tels que MBD3-NuRD, sont également présents au niveau de régions promotrices actives, ce qui indique que la régulation de la transcription par NuRD est plus diversifiée que l'on pensait initialement (Whyte et al., 2012). En outre, des expériences fonctionnelles ont révélé un rôle important pour le complexe NuRD dans la régulation des décisions relatives au destin cellulaire (Rais et al., 2013) (dos Santos et al., 2014). Ces différentes fonctions biologiques pourraient être provoquées par des modifications subtiles de la composition en sous-unité du complexe. Le complexe joue un rôle majeur dans la régulation de l'expression des gènes, l'organisation de la chromatine, les réparations des dommages causés à l'ADN, et le maintien de la stabilité du génome (Denslow and Wade, 2007) (Lai and Wade, 2011).

La stœchiométrie de ce complexe est aujourd'hui une question toujours ouverte. En effet, la composition du complexe est dépendante du contexte cellulaire et dynamique et par ailleurs ce complexe interagit avec de très nombreux partenaires comme la Lysine-specific histone demethylase 1A (LSD1), Ikaros, Aiolos, Helios etc.... (Wang et al., 2009) (Kim et al., 1999) (Sridharan and Smale, 2007)

L'équipe de Michiel Vermeulen, s'est penchée sur le sujet en utilisant une technique de spectrométrie de masse pour l'analyse quantitative de la stœchiométrie de type « label-free » (Kloet et al., 2015). Ils ont pu mettre en évidence l'importance de la protéine DOC1 dans la formation du cœur du complexe MBD3-NuRD. Les résultats obtenus sont en accord avec la

plupart des estimations stœchiométriques de ce complexe qui serait : 1 protéine MBD pour 1 HDAC, 3 MTA et 6 RBBP4. Néanmoins, les auteurs précisent que du fait de l'hétérogénéité du complexe, les nombreuses associations avec d'autres facteurs font que ces données ne sont probablement pas strictement exactes.

## A. Structure

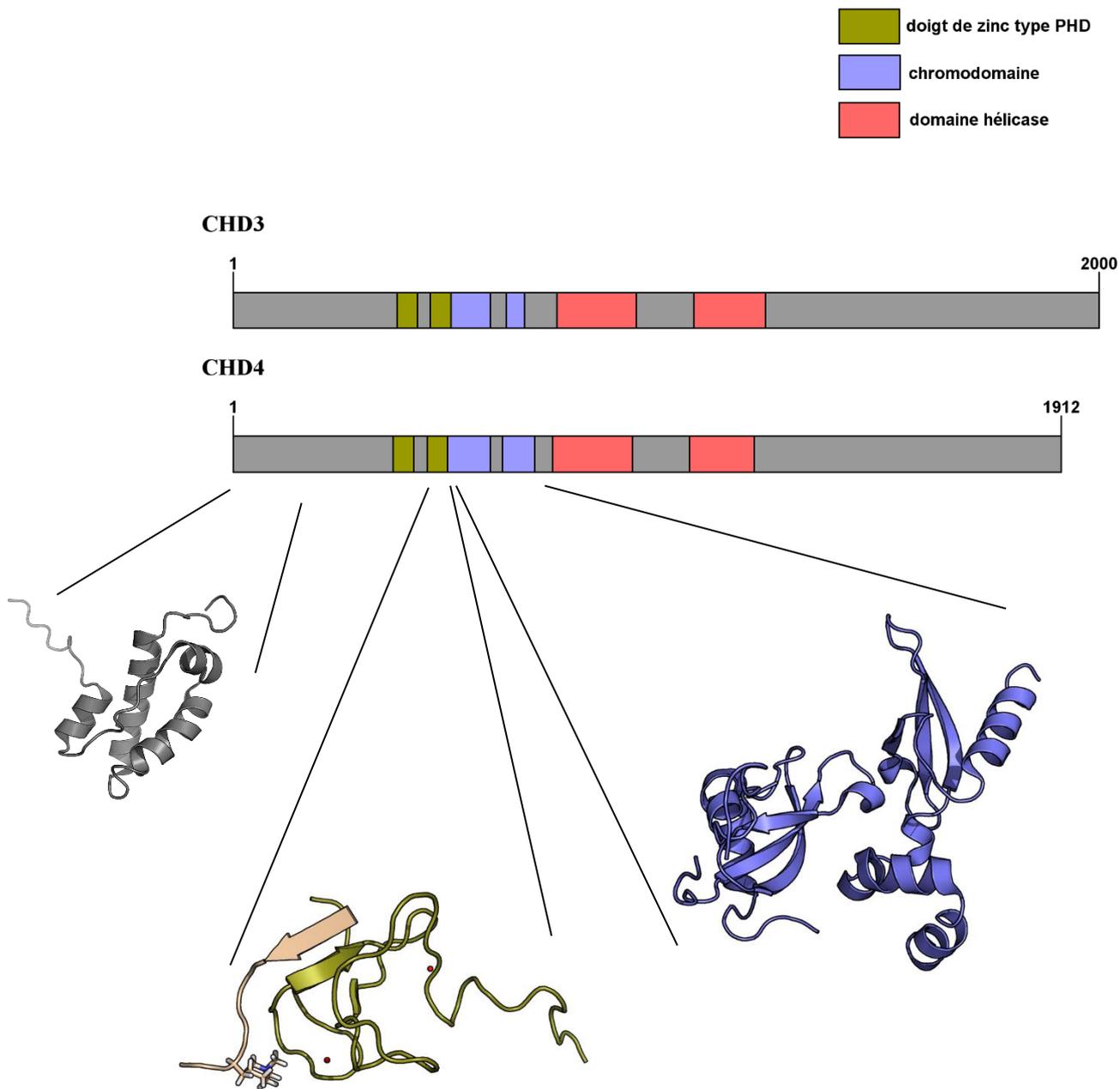
### 1. CHD3 et 4

Les cellules possèdent des enzymes à l'origine du remodelage de la chromatine permettant comme on a pu le voir précédemment, le mouvement, l'éjection et réarrangement des nucléosomes de façon ATP-dépendante. Ces protéines agissent au sein majoritairement de complexes protéiques permettant la régulation de ces enzymes mais également leur localisation aux régions d'intérêt (Clapier and Cairns, 2009).

Le complexe NuRD est caractérisé par la présence d'ATPases nommées CHD pour chromodomain helicase DNA binding protein 3. Parmi les 9 représentants de cette famille, deux homologues CHD3 et CHD4, présentant une similarité de séquence de plus de 71%, sont impliqués dans le complexe NuRD.

D'un point de vue structurale, ces deux homologues sont caractérisés par la présence de deux doigts PHD conservés, deux chromodomains en tandem et un domaine hélicase SW2/SNF-like. Les domaines de doigt de zinc PHD forment une famille de domaines de liaison aux histones et particulièrement sensibles aux MPT des queues d'histones. Ainsi CHD3 présente une certaine sensibilité pour les histones H3 modifiées sur la lysine 9 par méthylation (H3K9me) et/ou acétylation (H3K9ac) permettant de cibler le complexe NuRD à ces endroits en vue du remodelage de la chromatine (Tencer et al., 2017).

Elles sont de plus très conservées dans l'ensemble du règne animal et végétal, bien qu'absente chez la levure. Aucune structure cristallographie de ces protéines complètes n'a été résolue à ce jour. Seules des données en solution par RMN du chromodomaine, des domaines PHD et des données cristallographiques des deux chromodomains en tandem sont disponibles (Figure 24) (Mansfield et al., 2011) (Kwan et al., 2003) (Wiggs et al., to be published).



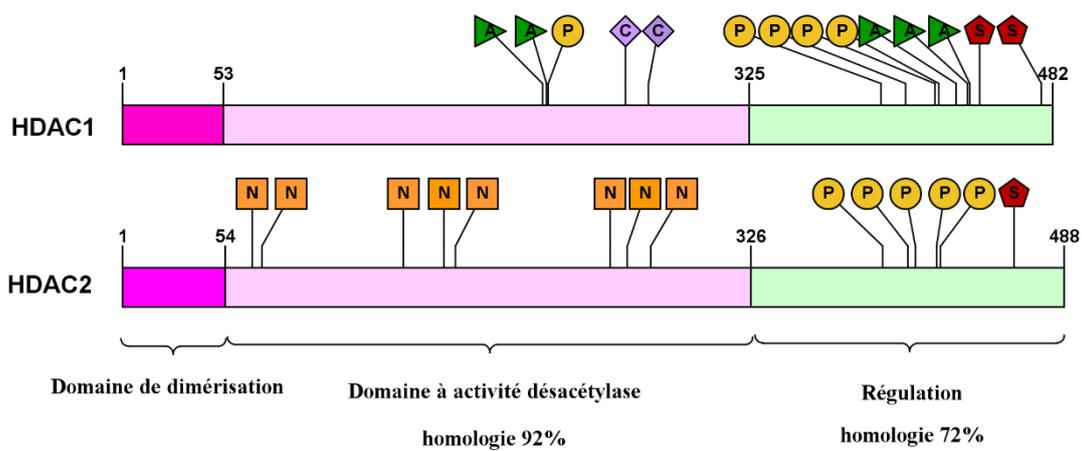
**Figure 24 CHD3 et CHD4.** Eléments structuraux de la sous-unité CHD du complexe NuRD. En jaune sont représentés les domaines PHD, la structure correspondante est une structure en solution par RMN du domaine PHD2 en complexe avec l’histone 3 modifiée H3K9me3 (pdb 2L75). En bleu sont représentés deux chromodomains en tandem, structure cristallographique (pdb 4O9I). Enfin en gris est représentée une région amino-terminale de CHD4 obtenue par RMN, essentielle à l’activité de l’enzyme (pdb 2N5N).

Ces protéines ont été identifiées comme auto-antigènes chez les patients atteints de dermatopolymyosite (Seelig et al., 1996). Par la suite, ces protéines ont été décrites comme agissant seulement au sein du complexe NuRD et sont exprimées de façon ubiquitaires, CHD4 étant plus abondante que son homologue (Su et al., 2004).

Bien que CHD3 et 4 soient hautement similaires, ils forment des complexes NuRD distincts (Hoffmeister et al., 2017). La présence de l'une ou l'autre des sous-unités ATPase va fournir une spécificité fonctionnelle au complexe NuRD dans un contexte cellulaire précis. Cela a été récemment démontré dans le développement du cortex de souris, où il a été exposé que l'incorporation de chaque CHD permet de contrôler différents ensembles de gènes qui interviennent dans les aspects non redondants du développement (Nitaraska et al., 2016). L'expression différentielle de CHD3 et 4 est également observée dans de multiples tissus et maladies (Bergs et al., 2014). Par exemple, CHD4-NuRD est impliqué dans la répression de nombreux gènes lors du développement cardiaque. En absence de cette répression médiée par CHD4, on peut observer une mauvaise expression des chaînes lourdes de la myosine du muscle lisse, l'actine et la troponine entraînant une désorganisation des sarcomères et une altération de la fonction cardiaque (Wilczewski et al., 2018).

## 1. HDAC1 et 2

Comme présenté précédemment dans le chapitre *HAT/KAT: acetyltransferase enzyme*, les protéines HDACs sont impliquées dans la catalyse de la désacétylation des queues des histones. Deux HDACs de type I sont impliquées dans le complexe NuRD, HDAC I et II. Ces protéines d'environ 55 kDa sont très conservées et présentes de façon ubiquitaire chez tous les eucaryotes, elles sont également impliqués comme sous-unité catalytique dans deux autres complexes répresseurs, Sin3a et CoRest. HDAC1 et 2 présentent une similarité de séquence d'environ 86%, suggérant une redondance de la fonction de désacétylation. En effet, des études d'inactivation conditionnelles des gènes dans certains tissus chez la souris ont fourni des informations importantes sur les fonctions compensatoires de HDAC1 et HDAC2 dans la régulation de la prolifération cellulaire, l'apoptose et la différenciation cellulaire (Kelly and Cowley, 2013) (Brunmeir et al., 2009).

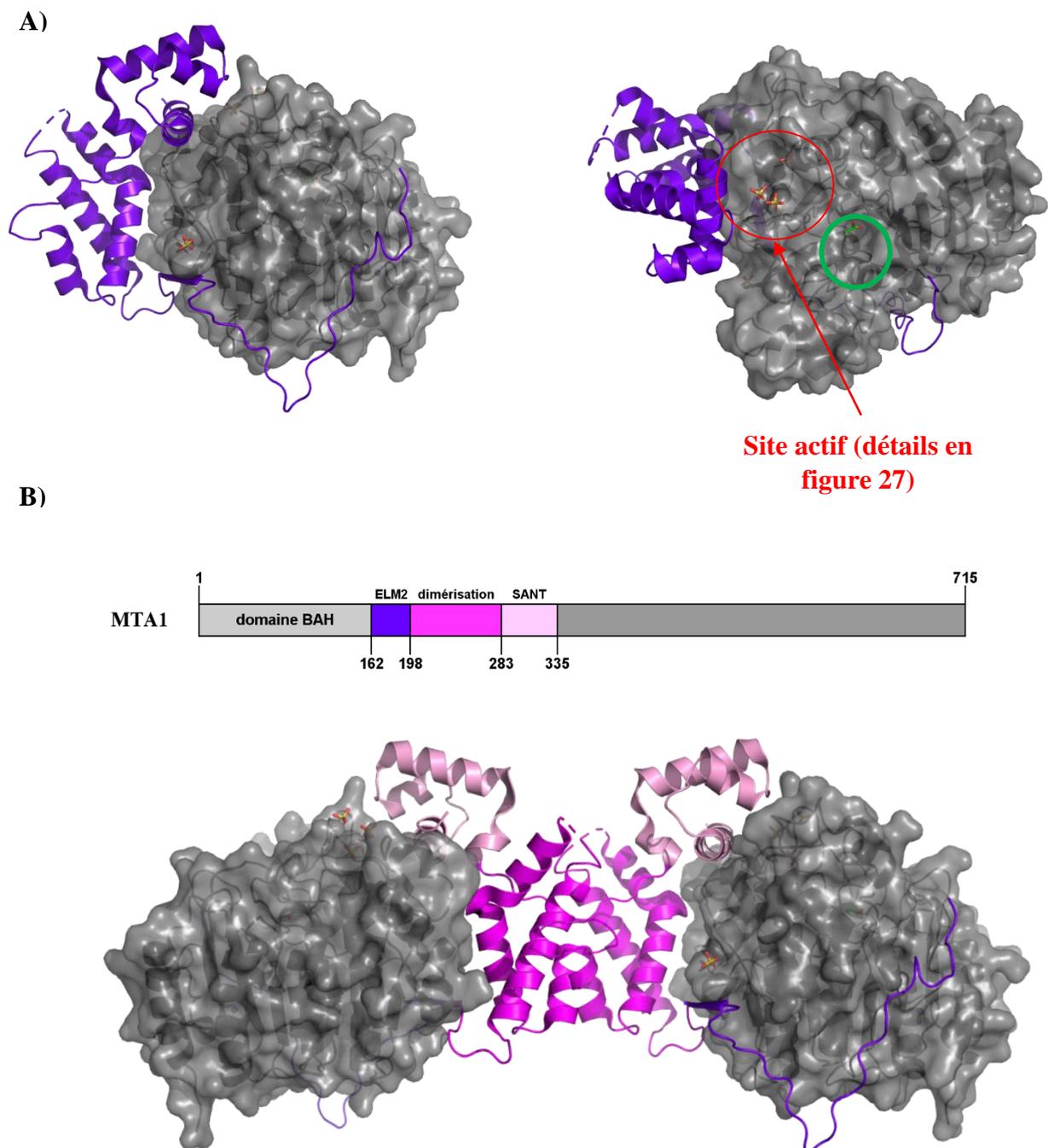


**Figure 25 HDAC1 et 2.** Organisation des domaines fonctionnels de HDAC1 et HDAC2. Pour chaque domaine est indiqué le pourcentage d'identité de séquence calculé à partir de l'alignement de séquence. Les modifications post-traductionnelles de HDAC1 et 2 localisées aux extrémités terminales sont identifiées. P (phosphorylation), A (acétylation), U (ubiquitination), S (SUMOylation), N (nitrosylation) et C (carboxylation).

Bien que certaines fonctions biologiques soient redondantes, HDAC1 et HDAC2 ont également des fonctions biologiques distinctes et même opposées ne pouvant être compensées. Par exemple, la délétion de HDAC1 seule est mortelle chez les embryons de souris (Ma and Schultz, 2016). Il existe également des preuves de fonctions distinctes des HDAC dans les cellules cancéreuses. Ainsi HDAC2, mais pas HDAC1, peut inhiber la prolifération et induire la sénescence des cellules cancéreuses du sein MCF7 (Harms and Chen, 2007), alors que HDAC1 est connu pour être surexprimé dans certains cancers comme le cancer des poumons (Cao et al., 2017). Les protéines HDACs sont, du fait de leur implication dans le cancer, des cibles importantes pour la conception et le développement d'inhibiteurs anti-cancéreux (Witt et al., 2009).

Enfin, bien que ces protéines soient associées à des complexes de répression, il est surprenant que des expériences d'inactivation des gènes *hdac* chez la souris aient révélé qu'une partie significative de gènes se trouvaient réprimés. Cette information suggère que HDAC 1 et 2 puissent avoir un rôle dans l'activation de la transcription de certains gènes (Yamaguchi et al., 2010) (Zupkovitz et al., 2006).

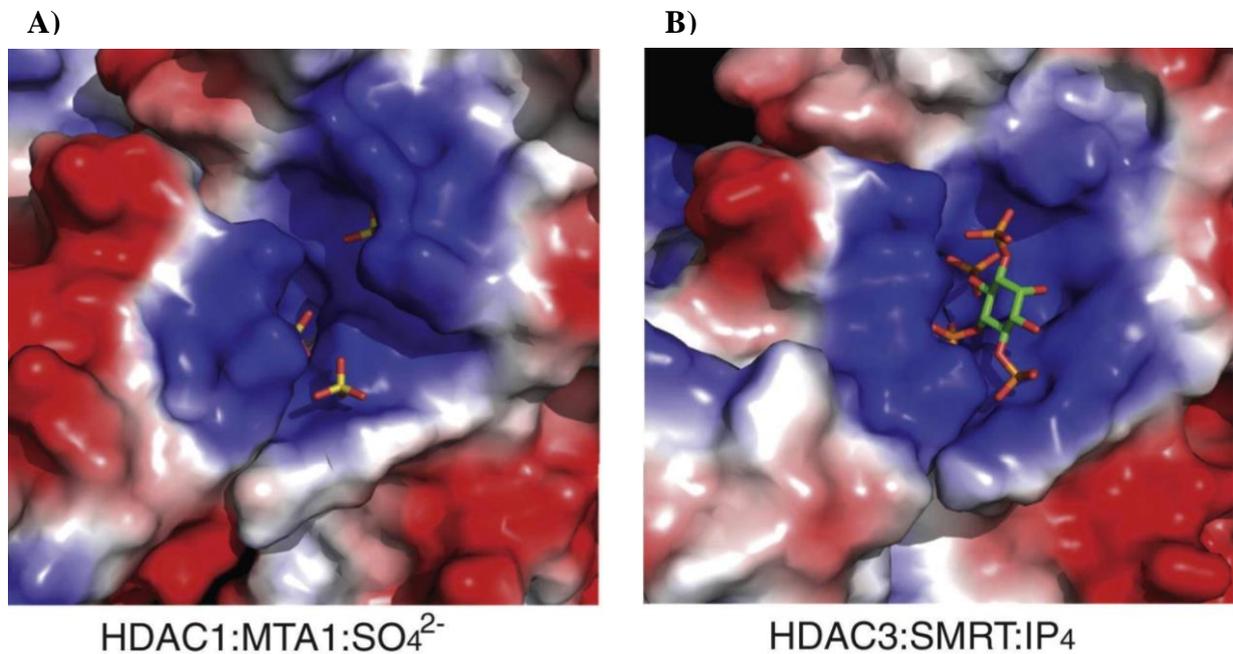
D'un point de vue structurale, les alignements de séquence des HDACs montrent des différences dans leur domaine C-terminal. On retrouve en N-terminal le domaine catalytique commun au HDAC de type I constituant une large part de la protéine. En amont de ce domaine se trouve un domaine d'homodimérisation et hétéro-dimérisation. (Taplick et al., 2001) Les principales différences se situent en C-terminal où l'on retrouve chez HDAC1 un domaine d'interaction à la protéine Chfr régulant la dégradation des protéines et également un domaine de signalisation nucléaire (Oh et al., 2009). HDAC2 quant à elle présente un domaine coiled-coil, c'est le seul représentant de sa famille présentant ce domaine (Figure 26 A). A la différence des autres HDAC de type I, HDAC1 et 2 ne possèdent pas de domaine de liaison à l'ADN et vont exécuter leur fonction en interagissant avec les facteurs de transcription ou des complexes répresseurs de l'expression génique, sous forme dimérique (Reichert et al., 2012).



**Figure 26 MTA1(ELM2-SANT) en complexe avec HDAC1.** A) le motif ELM2 de MTA1 non structuré est enroulé autour de HDAC1 représenté en gris. Après rotation on observe la présence du site actif dans lequel se trouve un composé acétate entouré en vert. B) Dimérisation des domaines de HDAC1. (pdb 4BKX). Sur la structure cristallographique du complexe, on observe également des composés phosphates (en jaune) localisés au niveau de la poche d'interaction de l'IP4. La présence de ces composés est un artéfact de cristallisation, les conditions étant fortement concentrées en phosphate, les composés se sont fixé de façon similaire à l'IP4 dans la poche et permettent probablement de stabiliser la structure.

HDAC1 et 2 ne sont pas seulement des protéines régulant les modifications post-traductionnelles, elles en sont également la cible. Ainsi elles seront la cible de nombreuses modifications tels que l'acétylation, la phosphorylation, la méthylation, l'ubiquitination, la sumoylation etc.... (Segré and Chiocca, 2011) Bon nombre de ces sites sont conservés entre les deux protéines. (Figure 25)

En 2017, l'équipe de John Schwabe publia la première structure cristallographique de HDAC1 en complexe avec MTA1 et plus précisément les domaines ELM2-SANT de cette dernière. Cette structure a apporté de nouvelles informations sur l'organisation du coeur du complexe NuRD mais également sur sa régulation. On peut remarquer (Figure 26) que MTA1 est complètement enroulé autour du cœur catalytique de HDAC1. La surface de contact entre les deux sous-unités de NuRD est extrêmement large et ce fait au travers de nombreux contacts entre les deux extrémités de ELM2-SANT et HDAC1. La partie centrale du double domaine de MTA1 quant à elle semblerait permettre la dimérisation des HDACs. De façon similaire à la structure du complexe HDAC3:SMRT déjà publiée par cette équipe, on retrouve à l'interface HDAC1:SANT une poche extrêmement basique capable d'accommoder un dérivé inositol de type Ins(1,4,5,6)P4 (Figure 27). La présence IP4 augmente drastiquement l'activité de désacétylase du complexe et est considéré dès lors comme un cofacteur du complexe NuRD. Néanmoins, le domaine ELM2 de MTA1 intervient dans l'assemblage du complexe HDAC1 indépendamment de l'inositol phosphate. Il semblerait que HDAC1 soit assemblée de façon constitutive au complexe NuRD et son activité serait régulée par la présence d'IP4. On parle dès lors de MTA1 « attaché » à HDAC1 en opposition à MTA1 « engagé » en présence d'IP4 (Millard et al., 2013a) (Watson et al., 2016).



**Figure 27. La poche basique de liaison à l'IP4 de HDAC.** (A) La poche, chargée positivement, de liaison à l'inositol phosphate de HDAC1: MTA1 est représentée ici. Elle est occupée par des ions sulfate ordonnés (pdb 4BKX). (B) La poche de liaison de HDAC3 en complexe avec SMRT. Ce site est occupé par un Ins(1,4,5,6)P<sub>4</sub>. (pdb 4A69). *Adaptée de Class I HDACs Share a Common Mechanism of Regulation by Inositol Phosphates, Millard C. et al. Mol Cell. (2013).*

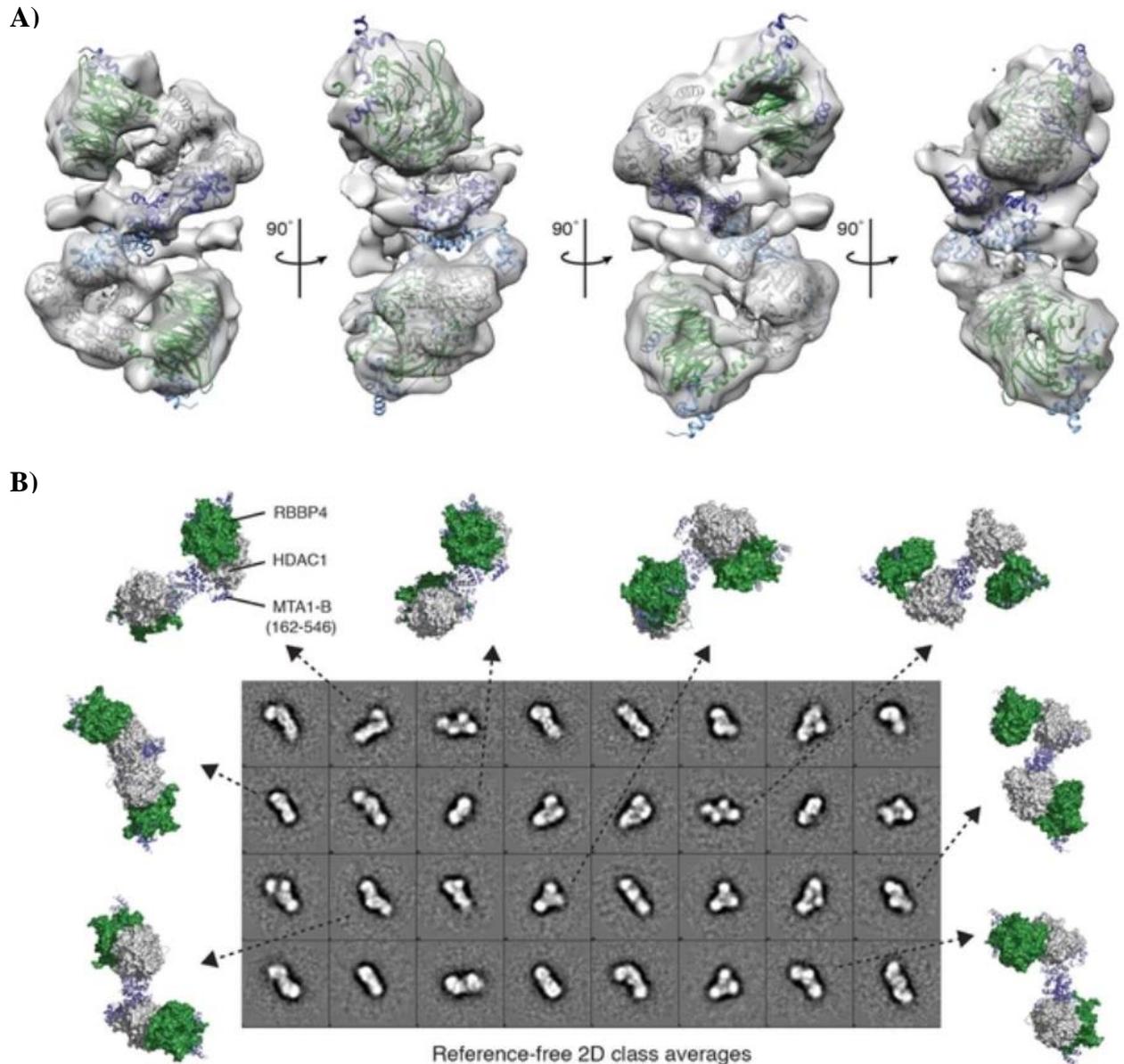
## 2. MTA1 et 2

MTA pour Metastasis-associated protéine, est un élément essentiel du complexe NuRD bien que n'ayant aucune activité enzymatique intrinsèque. Elle joue un rôle dans la stabilisation et l'assemblage du complexe (Millard et al., 2014).

Le premier représentant de cette famille à être identifié fut MTA1, isolé à partir de cellules métastatiques de cancer du sein, d'où son nom. Les protéines de cette famille sont en effet extrêmement impliquées dans la croissance tumorale. On a pu observer qu'une surexpression de MTA1 est fortement associée dans la croissance agressive de cancer de l'endomètre, du sein et des ovaires et a pu dès lors être utilisé comme marqueur de la progression de cancer chez l'Homme. Les représentants de la famille MTA sont des protéines d'environ 80 kDa possédant quatre domaines conservés, tous présent dans leur partie N terminale.

SANT (pour Swi3, Ada2, NCoR and TFIIB domain) : ce domaine est relativement petit, environ 50 acides aminés et est retrouvé chez de nombreuses protéines impliquée dans la régulation de la transcription (Boyer et al., 2004). Comme présenté dans la figure 26, ce domaine est organisé sous forme de 3 hélices, l'hélice H3 formant une poche basique à l'interface MTA1/HDAC1 impliquée dans la régulation de HDAC1 par fixation d'un composé l'IP4. Les résidus impliqués dans cette poche ne sont pas uniquement présent pour SMRT-SANT et MTA-SANT mais également chez d'autres sous-unités de complexe de la répression (RCOR1-3, MIER1-3, RER) recrutant également HDAC1. Ils pourraient dès lors tous former cette poche et être sensible à l'activation à l'IP4. Cette hypothèse est également fortement appuyée par la forte similarité de séquence des représentants de la classe I des HDACs (Millard et al., 2013b).

ELM2 (pour egl-27 and MTA1 homology domain 2): ce domaine est positionné en N-terminal du domaine SANT et est prédit comme majoritairement désordonné. En effet, des expériences de dichroïsme circulaire ont montré que ce domaine ne présenterait aucune structure secondaire quand il était exprimé de façon isolée (Millard et al., 2013b). Néanmoins, comme présenté en figure 26, le domaine va s'organiser sous forme de structure hélicoïdale lors de sa liaison avec HDAC1. On peut noter que ce domaine présente deux régions, une partie N-terminale qui s'enroule autour de HDAC1 et une C-terminale formant un paquet de quatre hélices servant d'interface d'homodimérisation. Ce domaine ELM2 est suffisant à l'interaction avec MTA1 (Lee et al., 2006).



**Figure 28 Architecture du cœur du complexe NuRD par microscopie électronique.** (A) Model par microscopie électronique à une résolution de 19Å du cœur du complexe NuRD dans lequel sont placés les structures cristallographiques des complexes HDAC1-MTA1 et RbAp46. MTA1 est en bleu, RbAp46 en vert et en gris est représenté HDAC1. (B) Classes deux dimensions, représentatives de la structure obtenue. *Adaptée de The structure of the core NuRD repression complex provides insights into its interaction with chromatin, Millard C. et al. eLife (2016).*

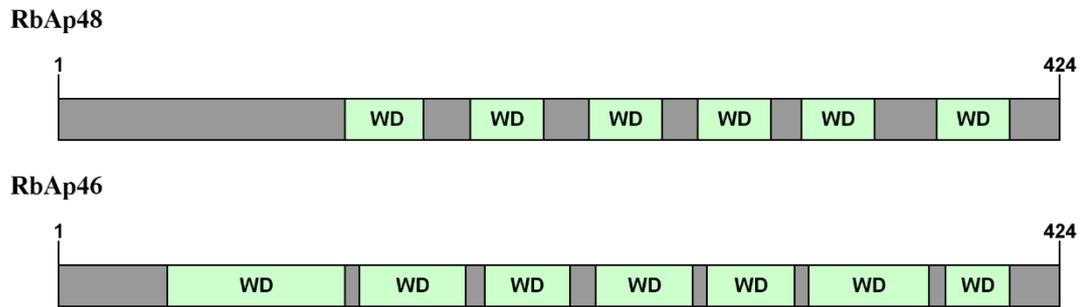
La partie amino-terminale enroulant HDAC1 est extrêmement conservée au sein de la famille, et son positionnement va permettre au domaine BAH adjacent de présenter le nucléosome au site catalytique de HDAC.

Concernant BAH (pour bromo-adjacent domain) et GATA (pour GATA-zing finger domain), peu d'information sur la fonction et aucune sur la structure de ces deux domaines sont présent dans la littérature. Ces domaines ont néanmoins été identifiés par analyse de séquences avec d'autres protéines déjà caractérisées. Grâce à la comparaison et à la modélisation des séquences, nous pouvons aujourd'hui spéculer sur l'aspect de ces domaines et leur fonctionnement. Ainsi, le domaine BAH présent chez d'autres sous-unités de complexes répresseurs, s'est avéré être un médiateur de l'interaction avec le nucléosome et plus précisément les queues des histones H4, permettant le recrutement des complexes aux régions d'intérêts (Callebaut et al., 1999) (Armache et al., 2011).

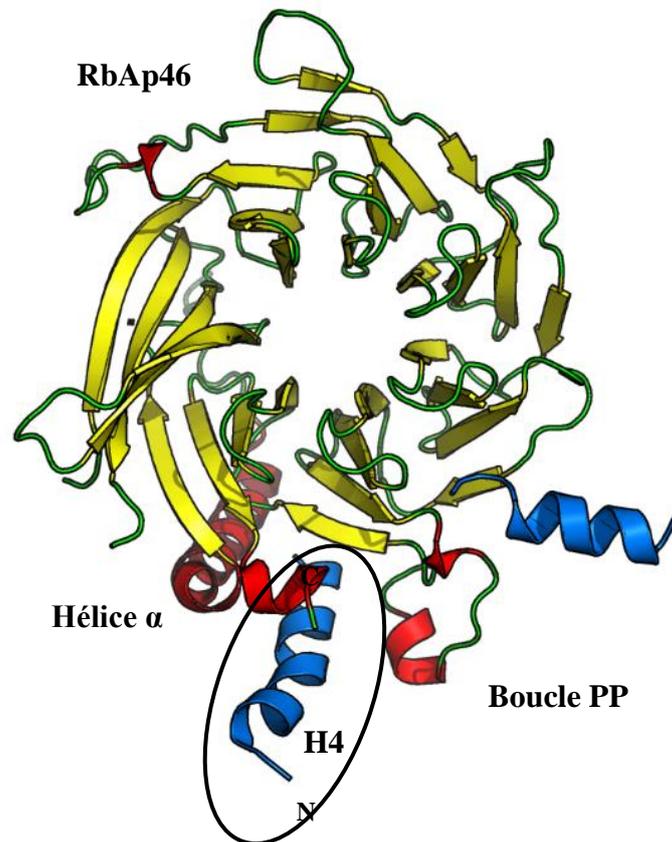
MTA1 et MTA2 partagent une similarité de séquence de 63% et MTA1 et MTA3 72%. Ces trois homologues divergent principalement au niveau de l'extrémité C-terminale désordonnée, pouvant expliquer leur exclusivité mutuelle et leur implication dans différents mécanismes. Cette région est considérée comme principalement désordonnées à l'exception pour MTA1 et MTA2 d'une partie à l'extrémité formant une hélice et interagissant avec RbAp46/48 (Alqarni et al., 2014). De nombreuses études de co-immunoprécipitation ont permis la mise en place d'une hypothèse selon laquelle il n'existerait pas un mais deux sites de fixation à RbAp48/48 au sein de l'extrémité C-terminale de MTA1 (Roche et al., 2008) (Fu et al., 2011). Il faut attendre 2016, pour confirmer ceci. Millard *et al.* ont ainsi identifié et caractérisé par MALLS puis cristallographie aux rayons X ce second domaine. Ces deux sites nommés R1 et R2 interagiraient chacun avec une sous unité RbAp, et comme MTA1 est lui-même un dimère au sein du complexe NuRD, il serait probable que le complexe comporte quatre copies de RbAP. Ces deux régions R1 et R2 sont également extrêmement conservées au sein des espèces, suggérant dès lors une stœchiométrie du complexe NuRD conservée lors de l'évolution. Ce travail a également permis d'obtenir la première structure par coloration négative du cœur du complexe NuRD à une résolution de 19Å. Bien que de faible résolution, l'enveloppe obtenue, après intégration des structures cristallographiques, montre clairement un dimère HDAC1:MTA1 avec de part et d'autres les lobes RbAp (Figure 28) (Millard et al., 2016).

Au sein de cette région désordonnée, un site de liaison à CHD4 a aussi été identifié, mais pas encore décrit de façon structurale (Nair et al., 2013).

A)



B)



**Figure 29 RbAp46/48.** (A) Séquences des sous-unités RbAp caractérisées par un enchaînement de domaine WD40. (B) Structure générale de RbAp48 avec le tonneau de feuillet  $\beta$  en jaune, et la longue hélice  $\alpha$  en rouge qui se place contre le tonneau et permet la formation d'un sillon (entouré en noir) ou vient se positionner l'extrémité N-terminale de l'histone H4 en bleu. (pdb 3CFV)

### 3. MBD2 et 3

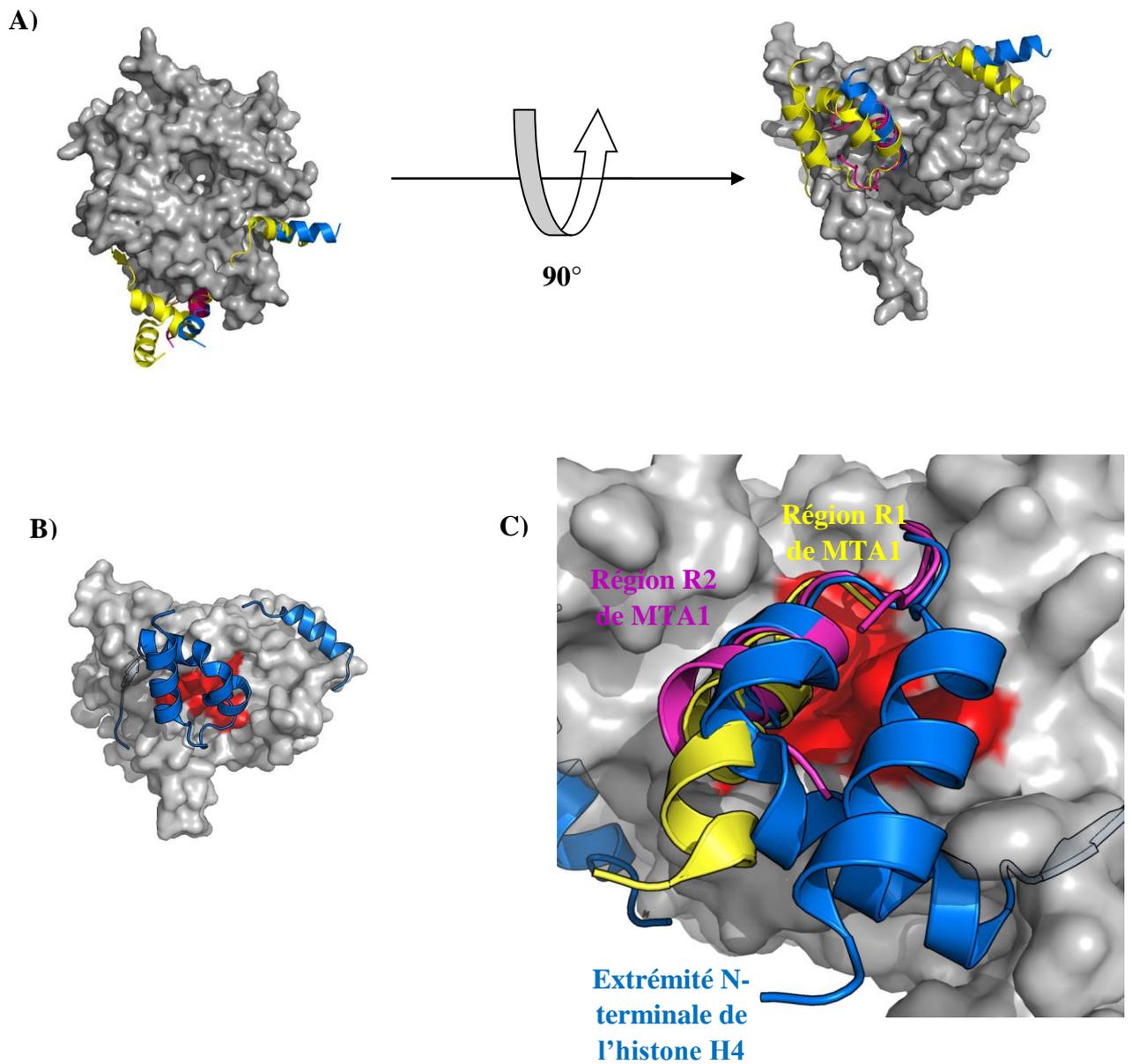
La sous-unité MBD est la plus petite sous-unité du complexe NuRD, déjà présenté dans le chapitre consacré à la méthylation de l'ADN et sa reconnaissance.

### 4. RbAp46 et 48

Ces protéines de 50kDa environ sont appelées RbAp46/48 du fait de leur capacité de liaison au suppresseur de tumeur Rb (Qian et al., 1993). Ces protéines sont aujourd'hui caractérisées par la présence d'un motif de séquence répétée WD40 et possèdent une similarité de séquence de plus de 90%. Elles sont principalement retrouvées dans le cœur de NuRD en interaction avec HDAC1 ou 2 mais également comme sous-unités d'autres complexes régulateur de l'expression génétique comme Sin3A et NURF (Barak et al., 2003) et PRC2 (Kuzmichev et al., 2002). Un de ses rôles les plus étudiés est son implication dans l'assemblage des nucléosomes. Ainsi, RBBP4 va former avec sNASP et Asf1 un complexe de chaperonne permettant l'approvisionnement de protéines H3 et H4 néo-synthétisée lorsque la cellule est soumise à un stress réplicatif. Cette protéine fut également caractérisée comme sous-unité du complexe d'assemblage CAF-1 (chromatin-assembly factor-1), dont la fonction d'assemblage des histones H3 et H4 sur le brin néo-synthétisé de l'ADN a été découverte en 1997 (Parthun et al., 1996) (Verreault et al., 1996). De plus, RBBP7 s'est révélé être une sous-unité essentielle du complexe histone-acétyltransférase HAT1, qui acétyle les protéines histones nouvellement synthétisées (Verreault et al., 1998). Comme le prouve ces études, ces protéines sont impliquées dans des mécanismes initiaux de l'organisation de la chromatine.

Il a été démontré que RBBP4 et RBBP7 interagissent directement avec les queues des histones H3 et H4 via deux sites de liaison distincts et peuvent donc servir de modules de recrutement de la chromatine (Murzina et al., 2008) (Zhang et al., 2013).

D'un point de vue structural, de nombreuses structures de RbAp46 et 48 existent et ont permis de révéler des informations sur l'implication de cette sous-unité dans l'architecture du complexe NuRD. De façon générale les deux protéines sont composées d'un tonneau de feuillets  $\beta$  à sept lames, suivi d'une longue hélice  $\alpha$  amino-terminale qui forme une extension au domaine principale. L'orientation de cette hélice va permettre la formation d'un sillon médiant l'interaction avec l'extrémité amino-terminale de l'histone H4. Cette interaction est unique parmi les protéines à domaine WD40, pour lesquelles l'interaction se situait



**Figure 30 RbAp48 et la poche hydrophile latérale.** (A) Vues supérieure et latérale de RbAp48 en complexe avec l'histone H4 (pdb 3CFV) représentée en bleu, la région R2 de MTA1 (pdb 4PBZ) représentée en rose et la région R1 de MTA1 (pdb 5FXY) en jaune. (B) en rouge est représentée la poche hydrophobe dans laquelle se situe la majorité des interactions observées. (C) Zoom sur l'orientation similaire des différents partenaires de RbAp48.

majoritairement sur une face du tonneau voir au centre de celui-ci. La structure présentée en figure 30, montre que pour RBBP4, l'histone H4 se fixe préférentiellement dans une poche hydrophobe sur le côté du tonneau, formée d'une part par la boucle PP et d'autre part par la longue hélice amino-terminale. On observe également un réseau important de ponts salins et de liaisons hydrogènes par des résidus conservés chez RbAp46. Il est intéressant de noter que l'extrémité carboxy-terminale de MTA1, nommée R2 (résidus 671-690) interagit avec RBBP4 / 7 dans ce même sillon, établissant des contacts très similaires à ceux identifiés avec l'histone H4. Par conséquent ces protéines, au sein du complexe NuRD ne pourront interagir qu'avec l'histone H3 et non l'histone H4 (Alqarni et al., 2014) (Murzina et al., 2008).

En plus de cette séquence R2 a été identifiée une séquence très similaire nommée R1 en amont de R2. Il a été récemment montré qu'elle possédait elle aussi l'habileté de lier une protéine RBBP (Millard et al., 2016). Ces informations sur le cœur du complexe NuRD suggèrent que la stœchiométrie du complexe MTA-RBBP est de 2:4 (Schmidberger et al., 2016) (Brasen et al., 2017).

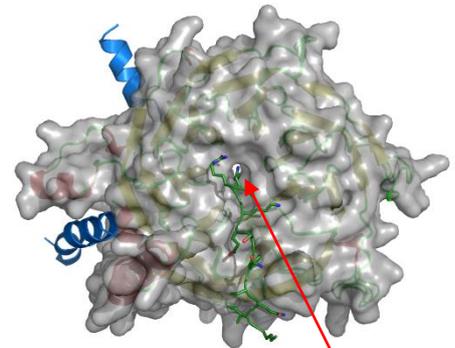
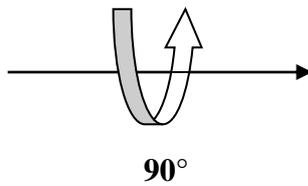
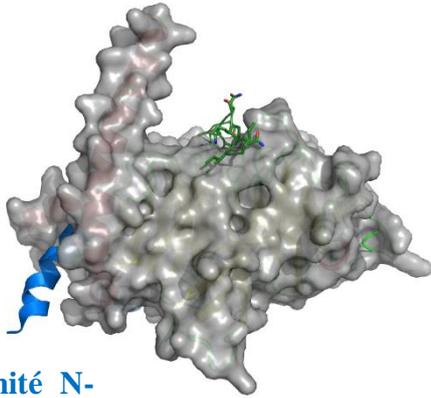
La structure cristalline de RBBP4 liée à un peptide dérivé de la queue de l'histone H3 montre que la surface d'interaction se situe au sommet du domaine WD40 (Schmitges et al., 2011). Ce site n'étant pas bloqué par la liaison à MTA1-RBBP4. Ainsi, RBBP4 resterait théoriquement capable de se lier aux queues de l'histone H3 lors de l'assemblage dans le complexe NuRD. Un tel modèle de recrutement à l'histone H3 est cohérent avec l'architecture du complexe MTA1: HDAC1: RBBP4 décrite en 2016 par Millard et al. (Millard et al., 2016). Comme on peut l'observer en figure 31, l'orientation de RbAp par rapport au dimère HDAC1:MTA1 permet l'exposition de la surface de liaison à l'histone H3. De plus, ce site de recrutement des histones est suffisamment proche du site catalytique de HDAC1 permettant aux autres queues d'histones d'accéder au site actif. Ces observations sont en accord avec le rôle de module de recrutement de cette sous-unité.

En résumé, ces différentes études structurales de la protéine RbAp ont souligné, une fois de plus, le rôle d'échafaudage pour l'assemblage du complexe NuRD de la sous-unité MTA permettant la réunion de différentes sous-unités qui interagissent elles-mêmes avec l'ADN, les histones ou le nucléosome. Ainsi, MTA1 va recruter un partenaire RbAp qui se fixe à la queue d'une histone H3 et en parallèle recrute également deux partenaires HDAC1, par dimérisation permettant leur proximité avec le nucléosome. Cette dernière va faciliter la fixation et la désacétylation des lysines de la queue des histones H3. Ce mécanisme peut être régulé entre autre par la présence de cofacteur de régulation de la transcription tel FOG

A) Vue latérale

Vue supérieur

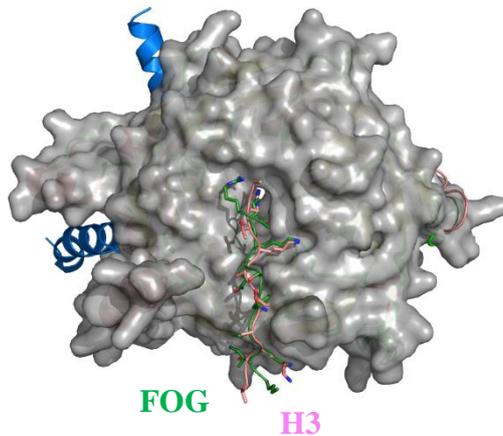
Hélice  $\alpha$



Extrémité N-terminale de l'histone H4

Tunnel d'interaction pour l'histone H3

B)



**Figure 31 RbAP48 et l'interaction avec l'histone H3.** (A) Interaction avec le peptide FOG (pdb 2XU7) représenté en vert et comparaison avec le tunnel d'interaction de l'histone H4. (B) Superposition de la structure de RbAp48 avec FOG et une protéine RbAp-like en interaction avec H3. Le mode d'interaction est très similaire et éloigné du tunnel de H4. (pdb 2YBA)

(Lejon et al., 2011). Ce cofacteur, impliqué dans la différenciation des érythrocytes et des mégacaryocytes présente la capacité de se lier à RbAp au niveau du site d'interaction avec H3. Enfin, MTA1 va permettre le recrutement également de CHD4 au sein du complexe NuRD qui présente des sites de liaison supplémentaires à la chromatine. En effet le domaine PHD présente la capacité de fixer le nucléosome (Mansfield et al., 2011) (Musselman et al., 2012) (Musselman et al., 2009). Comme décrit dans les différentes sous parties de ce chapitre sur le complexe NuRD que l'interaction de NuRD avec le nucléosome est multivalente et ne dépend d'une seule sous-unité. Ainsi la compréhension de l'impact des interactions des différentes sous-unités sur celle plus générale du complexe NuRD avec la chromatine nécessite des études structurales du complexe entier et intact.

## 5. GATAD2A et B

En 2002, des expériences de criblage à double hybride sur la protéine MBD2 ont permis de mettre en évidence une interaction avec deux nouvelles protéines, baptisées p66 $\alpha$  et p66 $\beta$ , et plus tard nommées GATAD2A et GATAD2B (GATA zinc finger domain containing 2A/B) (Brackertz et al., 2002). Ces deux protéines, initialement supposée être deux isoformes d'un même gène, semblent en réalité résulter d'une duplication d'un gène ancestral, survenue lors de l'apparition des mammifères. En effet, un orthologue unique p66 est retrouvé chez *Drosophila melanogaster*, *Caenorhabditis elegans* et *Xenopus laevis*.

La surexpression de chacune des p66 induit une augmentation de l'action répressive de MBD2 tandis qu'une invalidation de gène permet une reprise partielle de l'expression des gènes réprimés par MBD2. Néanmoins, cette reprise n'est que partielle, en effet la répression réalisée par MBD2 est aussi médiée par d'autres protéines que le complexe NuRD.

Les deux protéines p66 présentent deux régions hautement conservées que sont :

CR1 un domaine super-hélice en position amino-terminal. Ce domaine est responsable de l'interaction avec la sous unité MBD2 du complexe NuRD. La structure en solution par RMN du complexe entre la super-hélice CR1 de p66 et celle de MBD2 révèle que chaque protéine adopte une structure régulière en hélice formant une interface en antiparallèle (Figure 32). La constante de dissociation est de l'ordre du nanomolaire. Cette observation est assez surprenante quand on sait que normalement ce type d'interaction n'est pas défini et existe sous une variété de formes antiparallèle, parallèle ou oligomérique, sans préférence d'orientation. (Gnanapragasam et al., 2011)



Cette interaction est médiée par la présence de résidus hautement conservés permettant la réalisation de liaisons ioniques. De façon similaire MBD3 présente un domaine super-hélice capable d'interagir de façon stable avec le domaine CR1 de p66. En complément de ces interactions ioniques, on observe un nombre important de résidus hydrophobes entraînant un réseau d'interaction de type clé-serrure. Ainsi, sur chaque domaine on observe 3 résidus qui vont venir s'intercaler entre deux résidus du domaine opposé. (Figure 32) Ainsi on peut caractériser les résidus servant de « clé » chez p66alpha Ile145, Leu152, Leu159 et MBD2 Ile220, Val227, Leu234 et les résidus dits de « serrure » faisant face à ces résidus hydrophobes. Du fait de leur hydrophobicité et de leur tendance à former des hélices de façon isolé ; ces super-hélices vont former de préférence des homo-oligomères. Or ces domaines de MBD2 et MBD3 restent sous forme monomérique en solution même à haute concentration et le restent jusqu'à la reconnaissance de p66 (Walavalkar et al., 2013) (Brackertz et al., 2006).

CR2 est un domaine C-terminal de type doigt de zinc GATA. Ce domaine est nécessaire à l'association de p66 avec l'extrémité amino-terminale des queues des histones (Feng et al., 2002).

GATAD2A/B peuvent être la cible de modifications post-traductionnelles, telle que la sumoylation. Ainsi, les résidus K30 et K487 de p66 $\alpha$  et K33 de p66 $\beta$  sumoylés, favorisent l'interaction de ces protéines avec d'autres partenaires du complexe NuRD, comme HDAC1 ou RbAp46 (Gong et al., 2006).

#### 6. DOC-1, LSD1 ; sous-unités et autres protéines associées

Du fait de son implications dans de nombreux processus, il n'est pas surprenant que le complexe NuRD puisse exister sous différentes formes en fonction du contexte biologique, environnemental ou encore cellulaire. Ainsi, depuis la découverte du complexe de nouvelles sous-unités ont pu être identifiées. Ces dernières sont qualifiées de sous-unités sous-stœchiométriques dans le sens où elles ne sont pas retrouvées de façon systématique.

Parmi, ces « nouvelles » sous-unités on peut citer LSD1 (pour Lysine Specific Demethylase 1), une amine oxydase qui catalyse la déméthylation des histones via une réaction oxydative dépendante de la flavine adénine dinucléotide (Shi et al., 2004). D'un point de vu biochimique, LSD1 va éliminer les modifications de type mono ou diméthylation de la Lys4 des histones H3 et donc réguler la transcription. À ce jour, LSD1 a été identifié dans un certain nombre de complexes corépresseur, notamment CoREST (Lee et al., 2005) (Shi et al., 2005), CtBP (Wang et al., 2007), et un sous-ensemble de complexes HDAC (You et al., 2001)

(Yang et al., 2018). Il a également été montré que LSD1 interagissait directement avec p53 réprimant la transcription médiée par ce dernier et inhibant l'apoptose (Huang et al., 2007). Son rôle le plus étudié aujourd'hui est son implication dans certaines tumeurs dites à haut risque et particulièrement son intérêt en tant que cible thérapeutique (Forneris et al., 2008) (Kahl et al., 2006) (Shi, 2007).

En 2009, Wang *et al.* ont démontré la présence de LSD1 au sein du complexe NuRD, ajoutant une activité de déméthylation des histones à ce dernier (Wang et al., 2009). Dès lors, il semblerait que NuRD présente une nouvelle activité répressive portée par la sous-unité LSD1. Bien que pouvant paraître surprenant, des études antérieures ont montré que la désacétylation et la déméthylation des histones seraient interdépendantes (Denslow and Wade, 2007) (Lee et al., 2006). L'hypothèse majeure pour la présence de tant d'activité enzymatique au sein d'un même complexe se base sur le fait que ces activités ont toutes pour but la répression de la transcription. Par conséquent, il est concevable que l'évolution ait favorisée cette proximité physique permettant la potentialisation d'activités distinctes de remodelage de la chromatine pour une régulation plus fine (Wang et al., 2009).

Bien qu'initialement décrit en 1995, le lien entre la protéine DOC1 et le complexe NuRD ne fut réalisé que 15 ans plus tard en 2010 par deux équipes en parallèle chez la drosophile et peu après chez l'homme (Reddy et al., 2010) (Spruijt et al., 2010). Le nom DOC1 est pour Deleted Oral Cancer 1; En effet, elle est décrite comme un potentiel suppresseur de tumeur dans les cancers de la bouche. (Todd et al., 1995) De nombreuses études ont ainsi mis en évidence une absence ou une diminution de DOC1 dans environ 70% des cancers oraux. La perte de l'expression étant elle associée à des carcinomes nasopharyngés, gastrique et de l'œsophage (Winter et al., 2011) (Wu et al., 2012). L'absence ou la diminution de la protéine DOC1 peut dès lors servir de marqueur de l'invasion tumorale et de pronostic vital pour les patients. Un des rôles de DOC1 associé au complexe NuRD fut largement décrit en 2017 par Mohd-Sarip *et al.* La perte de DOC1 va contribuer au développement tumoral par un mécanisme de dé-répression de la transcription, DOC1 étant crucial pour le recrutement de NuRD au niveau de certains loci (Mohd-Sarip et al., 2017). Dans quelle mesure ce recrutement intervient ? Comment se déroule la coopération entre DOC1 et les autres sous-unités impliquées dans le recrutement de NuRD au niveau de l'ADN ? Tant de question encore en suspens, qui requierent des investigations plus approfondies.

ZMYND8 est une sous-unité sous-stœchiométrique très intéressante du fait que son recrutement au sein du complexe est réalisé par une GATAD2A. Or comme on a pu le voir précédemment GATAD2A et GATAD2B sont mutuellement exclusive. Ainsi le recrutement de ZMYND8 est médié par la composition du complexe NuRD. On peut parler ici d'une fonction de NuRD paralogue-spécifique. L'identification de cette sous-unité est relativement récente (Smits et al., 2013). ZMYND8 est impliqué dans le recrutement du complexe GATAD2A-NuRD au niveau de site de cassures double brin de l'ADN et favoriser la réparation par recombinaison homologue (Spruijt et al., 2016). De nouvelles études ont mis en avant un mécanisme plus complexe qu'il n'y paraît régulé par la présence de marqueurs épigénétiques au niveau des histones et l'intervention de facteurs tel que KD5MA (Gong et al., 2017).

Comme on a pu le voir jusqu'à présent le complexe NuRD ou plutôt les complexes NuRD présentent à eux seul des variations morphologiques et structurales très importantes. De plus, l'étude structurale de ce complexe pourrait apporter une base dans la compréhension de l'organisation du complexe et dans le type d'interaction réalisé avec le nucléosome et ses partenaires. A la vue des diversités fonctionnelles et biochimiques, NuRD devrait être considéré comme une famille à part entière et non une entité biochimique unique.

## B. Fonctions, régulation et cancer

Bien qu'initialement identifié comme un complexe répresseur de l'expression génétique, NuRD est aujourd'hui connu pour posséder un rôle plus complexe sur la transcription avec la capacité notamment de favoriser l'expression. NuRD possède également d'autres fonctions biologiques, tel que son implication dans les réparations de l'ADN, de par sa capacité de modification de la chromatine et son association à d'autres facteurs de la transcription. Les altérations de la fonction de ce complexe NuRD sont la cause de dérèglements du développement embryonnaire et la cause de nombreuses maladies telles que le cancer.

Les sous-unités HDAC1/2 et RbAp46/48 sont présentes au sein d'autres complexes corépresseur tel que SIN3 et coREST, CHD3/4, MBD3 et MTA sont quant à elles relativement spécifiques du complexe NuRD. Comme on a pu le voir précédemment, les sous-unités du complexe NuRD existent sous la forme de différents paralogues bien souvent mutuellement exclusif. Par conséquent, différentes combinaisons de ces paralogues vont

résulter en différents complexes NuRD tissue et contexte spécifique. De plus le complexe est connu pour s'associer avec une variété de facteurs co-régulateurs entraînant ici aussi des fonctions spécifiques pour le complexe NuRD ou une localisation précise (Kim et al., 1999) (Lauberth and Rauchman, 2006).

Aujourd'hui on ne possède que peu d'informations structurales sur l'assemblage du complexe NuRD et sur la compréhension précise de son recrutement au niveau de sites spécifiques du génome. Néanmoins de nouvelles informations ont permis de mettre en avant notamment un mécanisme de régulation par modifications post-traductionnelles.

Ainsi il a été démontré que la méthylation de la lysine 532 de la sous unité MTA1 était requise à la formation du complexe NuRD présentant de plus une forte affinité pour l'histone H3. Cette association va favoriser la méthylation de l'histone H3 en position K9 permettant le recrutement de CHD4. De façon opposée, quand LSD1 déméthyle MTA1, le complexe NuRD se dissocie, et MTA va s'associer dès lors au complexe NURF et l'HAT p300/CBP pour activer la transcription de gènes. Ainsi, la sous-unité MTA1 ne joue pas uniquement un rôle dans l'assemblage du complexe mais également comme aiguilleur dynamique altérant de façon opposée l'expression des gènes en réponse à des signaux biologiques (Nair et al., 2013).

MTA1 n'est pas la seule sous-unité sensible aux modifications post-traductionnelles, en effet, la famille des HDAC sont régulés par des modifications tels que l'acétylation, la phosphorylation, la méthylation, la sumoylation etc... Il existe en effet un réel code de modification de HDAC1 et HDAC2. Ainsi, la phosphorylation de HDAC2 va permettre son recrutement par la sous unité RbAp48 au sein du complexe Sin3 ou NuRD qui fonctionneront comme répresseur. De façon plus surprenante, l'acétylation de HDAC1 inhibe son activité de déacétylation et HDAC1 et le complexe NuRD sont dès lors retrouvés au niveau de sites de gènes activés lors de la différenciation de cellules érythroïdes associé également à GATA-1. Le complexe NuRD va avoir dès lors une fonction d'activateur et non de répresseur dû à un environnement génétique différent (Segré and Chiocca, 2011) (Yang et al., 2012). Comme on l'a vu précédemment, l'activité déacétylase de HDAC est également régulée par l'inositol phosphate.

NuRD est un complexe présentant un rôle fondamental dans la régulation de l'expression des gènes et le contrôle du devenir des cellules souches embryonnaires. Ce mécanisme est accompli par l'intervention de divers complexes NuRD, leurs associations avec différents cofacteurs et leur habilité à coopérer avec d'autres complexes de modification

de la chromatine. Ainsi, MBD3-NuRD dans ces cellules va restreindre le niveau de l'expression des gènes pluripotents et permettre le renouvellement et le maintien des cellules souches et aussi leur transition en cellules différenciées, par un mécanisme de transcription hétérogène (Reynolds et al., 2012). Un rôle analogue pour NuRD a été décrit dans la peau, où il a été montré qu'il intervient dans les transitions développementales entre les précurseurs épidermiques et les précurseurs des cellules basales, folliculaires et matriciels. (Basta and Rauchman, 2015) (Kashiwagi et al., 2007)

NuRD s'est avéré avoir des effets opposés dans le cancer, à la fois en favorisant et en inhibant la croissance tumorale et les métastases dans différents tissus. Dans une certaine mesure, ces effets paradoxaux pourraient être expliqués par la capacité de NuRD à s'associer ou à moduler à la fois l'activité des suppresseurs de tumeurs (par exemple, p53) et les facteurs oncogènes (par exemple, Bcl6) (Qian et al., 2017) (Shen et al., 2018).



# PARTIE I: Etude du complexe NuRD

---

## I. Méthodes utilisées

### A. Biochimiques

#### 1. Séparation des extraits cellulaires

Le complexe NuRD a été exprimé de façon endogène dans des cellules HeLa par l'équipe du Dr. Ali Hamiche. Ces cellules ont d'abord été conçues pour exprimer la sous-unité MTA2, spécifique du complexe NuRD, avec une étiquette FLAG. De manière à sélectionner les cellules exprimant cette sous-unité modifiée et donc les cellules transfectées, l'équipe a utilisé une technique nommée MACS (Magnetic-Activated Cell Sorting). Les cellules sont en effet co-transfectées avec un gène de marqueur de la surface des cellules, ici l'interleukine IL2, permettant la séparation des cellules qui exprimeront ou non le gène mta-flag. Cette sélection est réalisée en imposant un champ magnétique après l'ajout de billes magnétiques présentant des anticorps ciblant le récepteur IL2.

De manière générale, la purification du complexe NuRD nécessite différentes étapes, et la première présentée ici, va permettre de séparer les différentes fractions cellulaires. NuRD étant un complexe présent dans le noyau, on cherche à isoler les deux fractions nucléaires que sont :

- La fraction nucléaire soluble (nommée par la suite ENS) : elle représente la fraction soluble du noyau qui n'interagit pas avec la chromatine.
- La fraction nucléaire insoluble (nommée ENI) : cette population est caractérisée par son interaction avec la chromatine. Cette fraction fut traitée de façon unique en vue de purifier des complexes NuRD – nucléosome.

#### La séparation des extraits cellulaires

La première étape consiste en la destruction des membranes cytoplasmiques chimiquement et mécaniquement. En effet, le culot cellulaire est resuspendu dans un tampon hypotonique A (10 mM TRIS pH 7.65, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.2 mM PMSF, 5 mM DTT) induisant un gonflement des cellules. En effet, par un processus d'osmose, les cellules vont accumuler de l'eau pour maintenir l'homéostasie ionique. Ensuite, les cellules ont subi

une destruction mécanique à l'aide d'un *dounce*, par 10 allers-retours, puis centrifuger pour permettre l'obtention d'un culot nucléaire. Cela permet la séparation de l'extrait cytoplasmique (surnageant) des noyaux intacts (le culot).

Pour séparer les deux extraits nucléaires, j'ai d'abord remis en suspension dans un tampon B à faible teneur en sel (20 mM TRIS pH 7.65, 60 mM NaCl, 15 mM KCl, 3 mM DTT, 0.34 M sucrose, 0.15 mM spermine and 0.5 mM spermidine) le culot nucléaire puis un second tampon C à haute concentration en sel (20 mM TRIS pH 7.65, 1.5 mM MgCl<sub>2</sub>, 900 mM NaCl, 0.2 mM EDTA, 25% glycérol, 0.2 mM PMSF, 3 mM DTT) est ajouté au goutte à goutte provoquant la libération des protéines solubles des noyaux sans aucun processus de lyse. Après une centrifugation le surnageant nommé ENS est conservé dans la glace et le culot est resolubilisé dans le tampon B, soniqué de façon à rompre les membranes nucléaires et de la DNase Micrococcale est ajouté pour digérer la chromatine et libérer des mono-nucléosomes.

Enfin pour éliminer tous les résidus membranaires, les lipides et les possibles agrégats, les deux extraits nucléaires sont centrifugés 1 heure à 35000 tour par minute. Les deux extraits sont complété d'inhibiteurs de protéase PIC de l'ordre d'une tablette pour 25 mL.

En raison de l'instabilité du complexe, tous les tampons et le matériel étaient maintenus sur la glace ou dans le réfrigérateur à 4 degrés.

Cette première étape de purification permet l'obtention des différents extraits mais n'est absolument pas sélective du complexe NuRD.

## 2. Affinité

L'étiquette FLAG fusionnée à l'extrémité C-terminale de la protéine recombinante permet leur purification par chromatographie d'affinité sur une résine recouverte d'anticorps anti-FLAG (Anti-FLAG® M2 Affinity gel, Sigma Aldrich). L'utilisation de cette étiquette présente de nombreux avantages, notamment dû à la spécificité sélective similaire à celle de l'interaction antigène-anticorps. L'élution est réalisée par compétition avec un peptide FLAG. De plus le degré de pureté est souvent largement satisfaisant avec cette technique et ne produit qu'un faible signal non-spécifique du fait de la sélectivité de la résine anti-FLAG. La reconnaissance ce fera au niveau de l'extrémité de MTA2 modifiée avec un fragment 3xFLAG de 25 acides aminés.

La résine (600 µL pour 30 g de cellules) est équilibrée par trois lavage successifs avec le tampon D composé de 20 mM TRIS pH 7.65, 3 mM MgCl<sub>2</sub>, 150 mM NaCl, 0.1 mM EDTA,

10% glycérol, 0.01% NP40 et 0.5 mM TCEP. Cette étape se fait en batch, et les extraits sont incubés à 4 degrés durant 6 à 7 heures. Après cette incubation, la résine est lavée avec le tampon D et les protéines sont éluées par compétition avec le tampon E contenant 0.5mg.ml de peptide FLAG (DYKDDDD) sur la nuit en chambre froide. Après élution les échantillons sont analysés par SDS-Page.

### 3. Les gradients

#### *a) Gradient de densité*

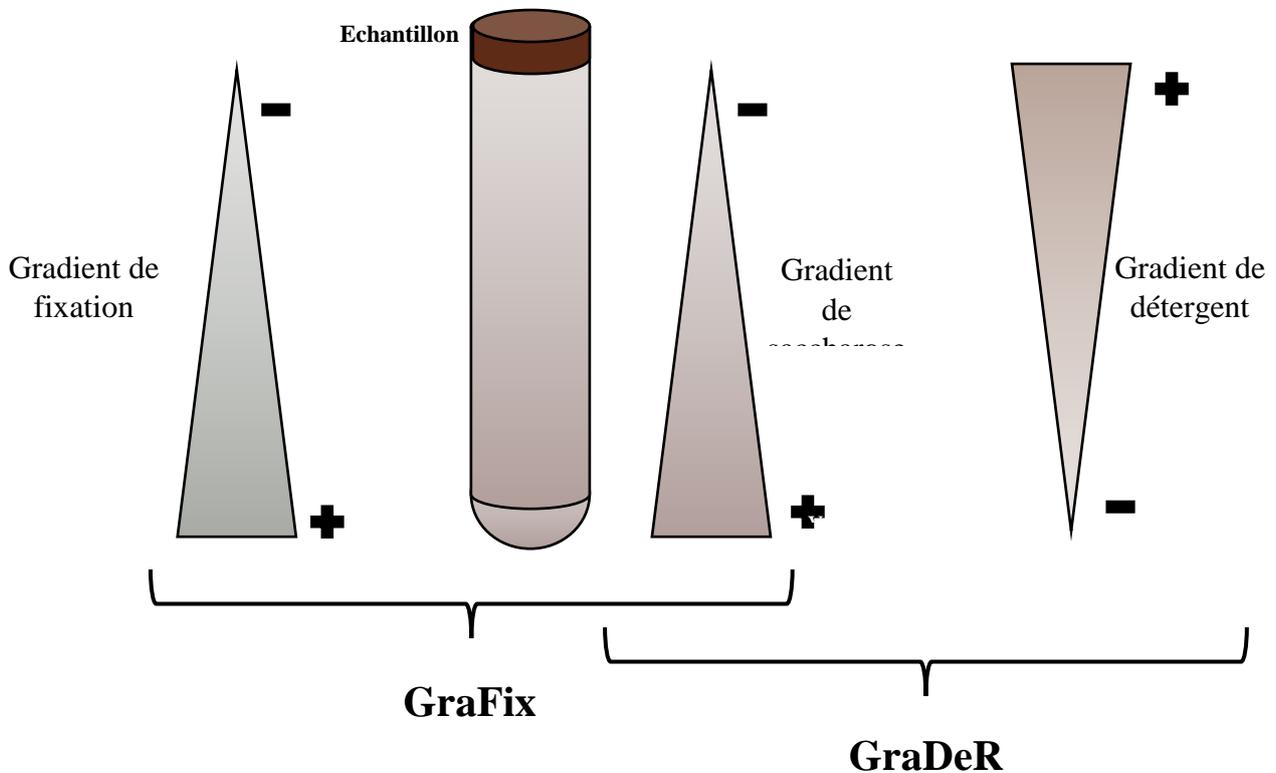
La méthode de séparation par centrifugation en gradient de densité, développée par Brake dans les années 1950, est une des techniques les plus utilisées aujourd'hui pour fractionner et purifier des complexes macromoléculaires, des virus etc... (Brakke, 1951) Elle va permettre ici de réduire l'hétérogénéité de l'échantillon en sub-fractionnant les différents sous-complexes.

La technique est basée sur le principe de sédimentation, selon lequel en solution, les particules plus dense que l'eau ou ici le saccharose vont sédimenter alors que celles moins dense vont flotter. Ainsi, en créant un gradient, on pourra séparer les composés selon leurs densités. De manière à accélérer le processus, les gradients sont soumis à une ultracentrifugation longue.

Pour cette étape un gradient linéaire de saccharose 15-35% est réalisé par l'ajout de deux solutions de saccharose, une de 35% dans la moitié inférieure de tube ultraclear et une de 15% dans la partie supérieure. Le gradient est réalisé 1 heure avant l'ultracentrifugation à l'aide d'un gradient master. Le gradient est par la suite laissé à 4 degrés. Les gradients étant d'environ 5 mL (ultraclear Beckman 5 mL), ils doivent être manipulés avec une grande précaution pour ne pas mélanger les différentes couches.

L'échantillon est appliqué doucement au-dessus du gradient et les tubes sont chargés sur un rotor SW60. Les gradients sont centrifugés à 35 000 rpm pendant 16 heures à 4 degrés avec un profil d'accélération de 9 et un profil de décélération de 9 pour éviter toute diffusion. Les particules se déplaceront ensuite dans le gradient jusqu'à ce qu'elles atteignent l'endroit où leur densité correspond à la densité environnante de saccharose.

Le gradient est fractionné (fractions d'environ 180 µL) directement à la suite de l'ultracentrifugation, à l'aide d'une pompe péristaltique (1 mL par minute), en commençant par la fraction de haute densité.



**Figure 33 Gradient, GraFix et GraDeR**

*b) Gradient de fixation*

Lors de cette méthode, les macromolécules seront fixées chimiquement par un processus de pontage covalent intramoléculaire pendant la purification par ultracentrifugation en gradient de densité. Cela va permettre de stabiliser le complexe durant cette étape mais également de réduire l'hétérogénéité dû à la dissociation. Pour réaliser ce type de gradient, on associe au gradient de sucrose un gradient d'agent réticulant à très faible concentration. En effet, la réticulation est un procédé aspécifique ne s'arrêtant pas spontanément et qui entrainera *in fine*, des réticulations intermoléculaires non désirées aboutissant à l'agrégation de l'échantillon. Les concentrations utilisées sont, de plus, suffisamment faible pour ne causer aucunes contraintes ou artefacts structuraux apparents (Stark, 2010). Pour réaliser ce double gradient, les composés de réticulation sont ajoutés à la solution la plus lourde. Différents agents peuvent être utilisés, mais le plus commun est le L-glutaraldéhyde (**Figure 33**).

Néanmoins, de récentes études ont montré l'importance du type d'agent de réticulation en fonction de la composition du complexe. Par exemple, l'utilisation de formaldéhyde est recommandée pour les complexes chromatinien (Hoffman et al., 2015).

La réticulation va créer dans la majorité des cas des liaisons irréversible entre les groupements amines, il est donc indispensable d'utiliser un tampon différent du TRIS et ses dérivés. De plus l'arrêt de la réticulation se fait de façon extemporanée par l'ajout d'amines à faible concentration telle que le TRIS ou la glycine.

Pour les expériences réalisées durant ma thèse, j'ai pu tester différentes conditions de réticulation en ajoutant entre 0.1 et 0.3% (V/V) de L-glutaraldehyde à la solution de saccharose 35%. Après fractionnement des échantillons, 80 mM de glycine était ajouté pour stopper le processus.

Les pontages covalents ont été également testés après le fractionnement de complexe en utilisant le glutaraldéhyde, un agent de pontage aspécifique, dont les deux groupements aldéhydes vont réagir avec toutes les fonctions amines des protéines.

La BS3, Bis(SulfoSuccinimidyl Suberate, est caractérisée par la présence à ses deux extrémités de fonctions ester de N-hydroxysulfosuccinimide (NHS) réagissant avec les amines primaires des chaînes latérales des lysines. Cette réaction entraîne la formation de liaison amide stable. A la différence de la glutaraldéhyde, la BS3 les deux fonctions réalisant le pontage sont espacées par un bras de 11 Å permettant le pontage de sous-unités plus éloignées. La réaction de réticulation se fait par l'ajout de BS3 à température ambiante pendant 30 min puis la réaction est arrêtée par l'ajout de glycine à une concentration finale de 60 à 80 mM.

### *c) GraDeR.*

Le terme GraDeR définit un nouveau type de gradient développé par l'équipe de Holger Stark en 2015, et signifie gradient basé sur l'élimination de détergents. Cette méthode a en effet largement fait ses preuves dans la purification de complexes membranaires adaptée aux études de microscopie électronique (Hauer et al., 2015). En effet, la purification de ce genre de complexe requiert l'utilisation de détergents à haute concentration créant en microscopie un échantillon très hétérogène et à faible contraste dû à la présence de micelles et monomères de détergents notamment à l'interface eau-air. Du fait de la fragilité de ces complexes, les détergents doivent être retirés de façon progressive et en douceur. Pour cela on peut ajouter au

gradient de saccharose un gradient inverse de détergent permettant l'élimination lente des détergents libres (Figure 33).

La purification de complexes protéiques instables nécessite également l'utilisation de détergents (Tableau 1) pour éviter toute agrégation. Néanmoins, l'élimination de ces détergents ou du moins leur diminution favorable aux études de cryo-EM reste un véritable challenge en raison notamment de leur faible CMC. L'utilisation de cette technique semble donc être une bonne alternative pour diminuer les concentrations en détergents et éviter la présence de micelles. (Singh and Sigworth, 2015)

Name	Type	MW	CMC	Micelle size (kDa)
<b>DDM</b>	Non ionique	510.6	0.17mM 0.009%	65-76
<b>Tween 20</b>	Non ionique	1228	0.06mM 0.0072%	85
<b>NP40</b>	Non ionique	603	0.05 to 0.3mM	60-95
<b>LMNG</b>	Non ionique	1005.19	0.01mM 0.001%	91-393
<b>DMNG</b>	Non ionique	949.08	0.036mM 0.0034%	
<b>Digitonine</b>	Non ionique	1229.3	0.5mM 0.002%	70
<b>Amphipol a8-35</b>	Polymère amphipathique	4300	0.002 g.L <sup>-1</sup>	40
<b>CHAPSO</b>	Zwitterionique	614.9	8-10mM 0.49%	7
<b>CHAPS</b>	Zwitterionique	614.9	8mM 0.49%	6

**Tableau 1 Les détergents communément utilisés en Cryo-EM et leurs propriétés.** Adapté de Stetsenko and Guskov, An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization.

Pour cette méthode du DDM a été ajouté à la solution de 15% de sucrose avant de former le gradient. La quantité choisie était calculée de manière à obtenir une concentration trois à dix fois inférieure de la CMC au niveau de la zone ou NuRD sédimente.

Enfin, le GraDeR présente l'avantage de pouvoir être associé au GraFix, permettant au complexe d'être réticulé en parallèle de la diminution du détergent et ainsi stabiliser le complexe durant cette étape.

#### 4. Caractérisation

##### a) Séparation des protéines : électrophorèse en condition dénaturante SDS-PAGE

L'électrophorèse en conditions dénaturante va permettre l'identification des protéines d'un échantillon en fonction de leurs poids moléculaire. Cette analyse se base sur la migration différentielle au travers d'un gel de polyacrylamide sous l'impulsion d'un champ électrique. La séparation se fera uniquement par la taille du fait des conditions dénaturante dû à la présence de Sodium DodécylSulfate (SDS).

Le gel est composé de deux parties, un gel de concentration et un gel de séparation différant par leur pH et leur pourcentage en acrylamide.

Les échantillons sont chauffés pendant 10 minutes dans un bain sec à 95°C en présence du tampon de charge (20 mM Tris-HCl pH 6.8, 40% glycerol, 8% SDS, 2 M  $\beta$ -mercaptoéthanol (bME) et du bleu de bromophenol). La migration est réalisées dans un tampon composé de 25mM Tris-Base, 250m M glycine et 0.5% SDS pendant environ 1 heure à 170 V. Le complexe étant produit de façon endogène en faible quantité des plaques avec un espacement de 1 mm ont été utilisées au lieu des habituel 0.75 mm. De plus la solution de coloration usuelle n'étant pas assez sensible, deux techniques de coloration ont été utilisées et sont décrites ci-après.

	GEL DE SEPARATION				GEL DE CONCENTRATION	
	10%	12%	15%	17%		6%
<i>Bis-acrylamide 40%</i>	2 mL	2.4 mL	2.8 mL	3.4 mL	<i>Bis-acrylamide 40%</i>	0.75 mL
<i>TRIS 1.5M pH8.8</i>	2 mL	2 mL	2 mL	2 mL	<i>TRIS 0.5M pH6.8</i>	1.25 mL
<i>H<sub>2</sub>O</i>	3.8 mL	3.4 mL	3 mL	2.4 mL	<i>H<sub>2</sub>O</i>	2.9 mL
<i>SDS 10%</i>	80 $\mu$ L	80 $\mu$ L	80 $\mu$ L	80 $\mu$ L	<i>SDS 10%</i>	50 $\mu$ L
<i>APS 10%</i>	80 $\mu$ L	80 $\mu$ L	80 $\mu$ L	80 $\mu$ L	<i>APS 10%</i>	50 $\mu$ L
<i>TEMED</i>	8 $\mu$ L	8 $\mu$ L	8 $\mu$ L	8 $\mu$ L	<i>TEMED</i>	5 $\mu$ L

**Tableau 2** composition des gels de séparation et concentration

### (1) Révélation au nitrate d'argent

La révélation au nitrate d'argent est l'une des méthodes les plus sensibles avec un seuil de détection de l'ordre de 5 à 10 ng. Pour cette méthode je me suis référée aux recommandations du fabricant des kits développés par GE Healthcare. Cette méthode bien que très sensible, est aussi longue et non compatible avec des études de spectrométrie de masse. J'ai donc décidé d'utiliser un autre type de méthode de coloration aussi sensible, utilisant le bleu de coomassie colloïdale G-250, selon la méthode développée par Kang en 2002.

### (2) Révélation au bleu sensible G-250

Protocole de révélation au bleu de coomassie G-250 :

- Le gel est d'abord rincé trois fois dans de l'eau distillée. Lors de chaque rinçage, le gel est laissé 5 minutes dans l'eau distillée pour éliminer le SDS limitant la coloration.
- Le gel est ensuite placé dans la solution de bleu colloïdale (Tableau 3). Après deux heures d'incubation minimum on peut observer l'apparition des premières bandes.
- Après la coloration sur la nuit, le gel est rincé de nouveau deux fois dans de l'eau.
- Enfin une étape de révélation par incubation pendant 15 minutes dans la solution de fixation.

Cette coloration au bleu sensible est basée sur la conversion des molécules de coomassie en particules colloïdales. Pour la préparation des solutions j'ai suivi le protocole publié par Nadine Dyballa et Sabine Metzger (Dyballa and Metzger, 2009).

<b>Coomassie colloïdale G-250</b>		<b>fixation</b>	
<b>G-250</b>	0.2% (W/V)	-	-
<b>Aluminium sulfate-(14-18)-hydrate</b>	5% (W/V)	-	-
<b>Ethanol (96%)</b>	10% (V/V)	<b>Ethanol (96%)</b>	10% (V/V)
<b>Acide orthophosphorique (85%)</b>	2% (V/V)	<b>Acide orthophosphorique (85%)</b>	2% (V/V)

**Tableau 3** Composition pour la révélation au bleu G-250

### *b) Dosage des protéines : La spectroscopie UV/visible*

Cette technique non colorimétrique est basée sur le fort pouvoir absorbant des résidus aromatiques, en particulier le tryptophane, à 280 nm. L'absorbance mesurée par spectrophotomètre est corrélée à la concentration protéique en appliquant la relation de Beer-Lambert :

$$A_{280\text{nm}} = -\log_{10} \frac{I}{I_0} = \epsilon_{280\text{nm}} \cdot l \cdot c$$

Avec A, l'absorbance à 280nm,  $\frac{I}{I_0}$  la transmittance de la solution,  $\epsilon_{280\text{nm}}$  le coefficient d'absorption molaire calculable *in silico* à partir de la séquence primaire de la protéine, en utilisant des outils comme ProtParam, l la longueur de la solution à traverser en cm et c la concentration en (mol/L).

### *c) Analyse : La spectrométrie de masse*

La spectrométrie de masse est une technique physique permettant de détecter et d'identifier des molécules d'intérêts. Elle réside dans la séparation en phase gazeuse des composants d'un échantillon par déplacement dans un champ électromagnétique. L'échantillon est au préalable ionisé, et les molécules chargées sont analysées en fonction de leur rapport masse/charge. Les analyses ont été réalisées par le service commun de protéomique de l'IGBMC, à l'aide d'un spectromètre de masse de type MALDI-TOF. Cet appareil permet de caractériser les protéines extraites d'un gel SDS-PAGE. Les bandes à analyser sont découpées dans le gel, déshydratées et enfin clivées spécifiquement à la trypsine. L'analyse MALDI-TOF permet de déduire la séquence primaire des peptides ionisés, dont les valeurs masse/charge correspondent à des enchaînements d'acides aminés définis. Par recouvrement des fragments déterminés, il est ainsi possible de remonter à la séquence de protéine entière. L'identification se fait par comparaison dans les banques de données, par exemple le NCBI.

## **5. Préparation des échantillons**

Pour permettre l'étude des complexes en microscopie électronique, l'échantillon doit se trouver dans des conditions optimales permettant à la fois sa stabilité mais également l'obtention d'un contraste adéquat à l'analyse des données et l'obtention d'une structure à haute résolution. Dans cette optique, différentes techniques ont dû être testées afin de concentrer l'échantillon, éliminer des composés ou simplement changer le tampon.

### *a) Les dialyses*

La dialyse est la première méthode utilisée pour le changement de tampon d'échantillon. Elle permet de traiter l'échantillon sans contrainte mécanique. Elle se base sur un mécanisme osmotique qui va permettre l'échange passif de composés par diffusion au travers d'une membrane semi-perméable.

Pour tester de nombreuses solutions de pH, sels, additifs, de nombreuses expériences de dialyse sur de petits échantillons de NuRD d'environ 50 à 100  $\mu$ L ont été réalisées. Un système de dialyseur réalisé à partir tubes eppendorfs 1.5 mL et de membrane « peau de serpent » avec des pores inférieurs à 7 kDa a été optimisé pour cela.

Les échantillons étaient déposés au contact de la membrane à l'interface avec le tampon de dialyse (1000 à 5000 fois le volume d'échantillon). La dialyse était laissée sous rotation à 4°C pour une durée de 6 à 16 heures.

### *b) Le dessalement*

Le dessalement est une technique se basant sur le principe de la chromatographie d'exclusion stérique. Elle va ainsi permettre l'élimination de petites molécules de masse moléculaire inférieure à 2 kDa, les sels ou encore simplement le changement de tampon. Le passage de l'échantillon dans ces petites colonnes ne nécessite pas l'utilisation de FPLC (Fast protein liquid chromatography) mais par gravité accéléré par l'utilisation d'une centrifugeuse. Cette technique permet de plus de pouvoir réaliser des tests sur de petits échantillons inférieurs à 100  $\mu$ L de façon extrêmement rapide, la méthode est donc adaptée au criblage de conditions en cryo-EM. Néanmoins, cela provoquera la perte d'une partie du complexe dû au passage au travers d'une colonne sous l'effet d'une force centrifuge du complexe.

Le dessalement d'échantillons de NuRD a été utilisé juste après le gradient de saccharose pour éliminer ce dernier, modifier le tampon ou encore éliminer les composés de pontage.

J'ai travaillé avec de très petits échantillons, afin de maximiser le nombre de tests en parallèle et de pouvoir tester plusieurs conditions. Le protocole reste le même, environ 70  $\mu$ L d'échantillon est appliqué sur la résine, préalablement équilibrée, lentement jusqu'à ce qu'elle pénètre dans cette dernière. Puis la colonne est centrifugée 2 minutes à 1500 g. L'échantillon est utilisé directement après pour préparer les grilles de microscopie. Il est très important de laisser l'échantillon pénétrer dans la résine avant de centrifuger pour optimiser l'élimination

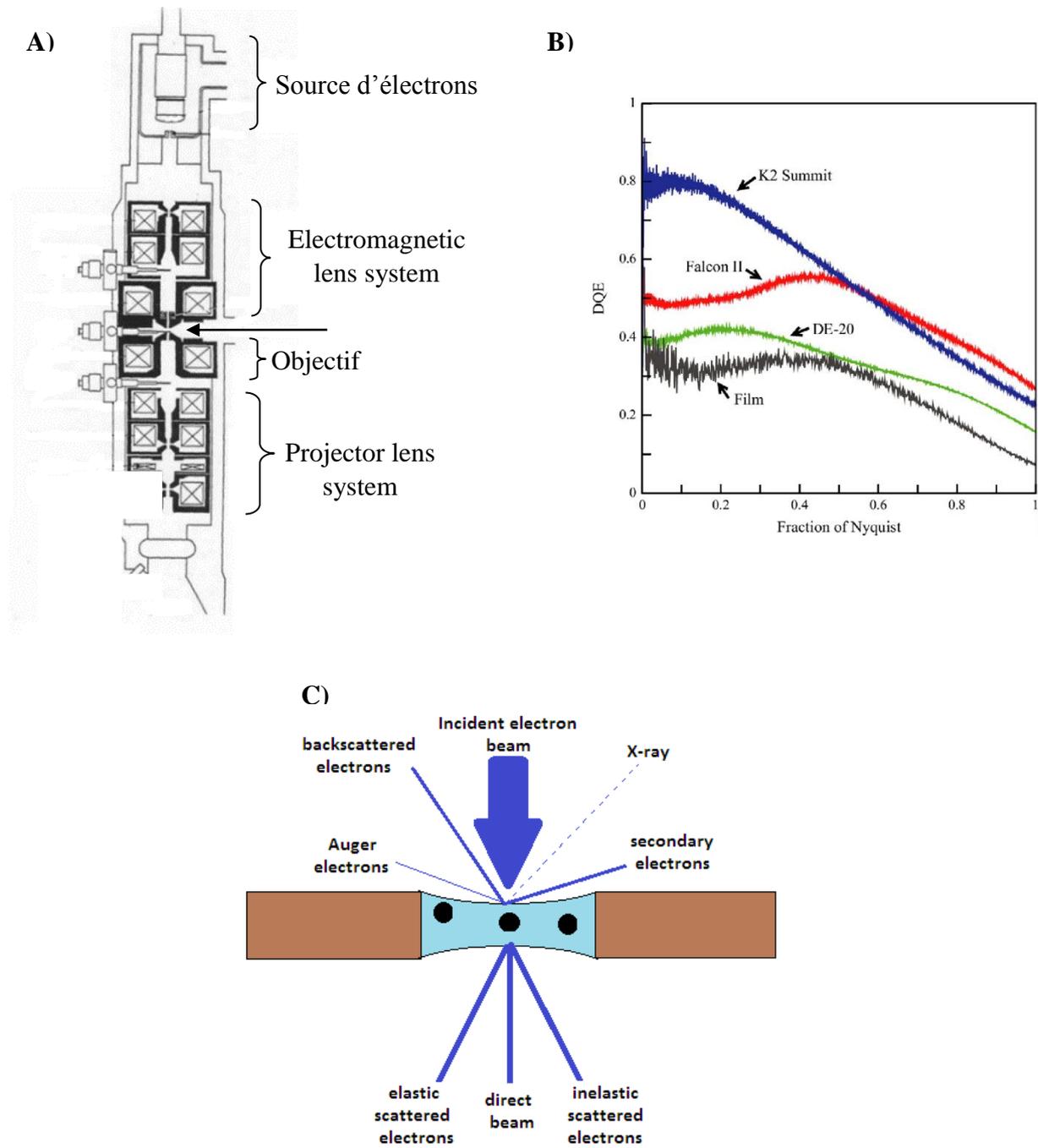
du saccharose d'une part et améliorer la récupération du complexe. En effet, du fait de la viscosité du sucre, le dessalement peut être ralenti.

### *c) La concentration*

Le complexe NuRD est produit en cellules HeLa, de façon endogène et non surexprimé. Cette méthode permet la purification d'un complexe proche de son état physiologique, néanmoins les rendements de ce genre de technique sont faible et l'échantillon nécessite d'être concentré pour pouvoir l'étudier. Différentes techniques sont connues pour concentrer un échantillon, toutes présentant des avantages et des inconvénients.

Les techniques testées pour le complexe NuRD étaient :

- L'utilisation de concentrateur Amicon ultra-0.5 avec une taille de pores de l'ordre de 100 kDa. La concentration est ici une simple ultrafiltration dépendante de la force centrifuge. Différents paramètres de centrifugation peuvent être testé en fonction de la sensibilité de l'échantillon. Généralement l'échantillon est centrifugé en étape itératives de 5 minutes à 2000-4000g.
- Une seconde technique se base sur la précipitation des protéines permettant par exclusion de la solution de précipiter sans dénaturer la ou les protéines d'intérêts. Pour cela, différents composés chimiques peuvent être utilisé tel que le PEG 20K ou le sulfate d'ammonium. La précipitation au PEG 20K est une étape clé dans la purification du ribosome, utilisée en routine dans l'équipe. Cette précipitation est adéquate pour les complexes protéiques macromoléculaires. Les tests de précipitation de NuRD se sont donc inspirés du protocole de précipitation du ribosome. Ainsi, 5 à 15% de PEG 20K ont été additionné aux échantillons de NuRD et incubé sur glace pendant 10 min. Le tout est centrifugé 10 min à 174000 g.



**Figure 34 microscope électronique à transmission (TEM)** A) Anatomie d'un microscope électronique. B) DQE des différents type de détecteurs. C) Interaction électron-matière.

## B. Biologie structurale

### 1. Cryo-microscopie électronique

L'un des objectifs de ma thèse était l'étude structurale du complexe NuRD. Comme on a pu le voir dans l'introduction, NuRD est un grand complexe protéique, hétérogène et dynamique. De ce fait, la méthode de choix pour cette étude semble être la cryo-EM, permettant l'étude de protéines ou complexes protéiques dans des conditions proches de l'état natif au sein de la cellule (Orlov et al., 2017).

Le premier prototype de microscope électronique fut développé en 1933 par Ruska et Knoll, et depuis cette technique n'a cessé d'évoluer et permis de se rapprocher de structure à résolution quasi atomique comme en atteste les récentes publications (Natchiar et al., 2017) (Bartesaghi et al., 2018) (Tan et al., 2018). L'un des points clé dans l'évolution de la microscopie électronique est le développement dans les années 80 par Jacques Dubochet d'une nouvelle technique de préparation cryogénique des échantillons permettant leur conservation et leur protection. Cette évolution a permis de poser un nouveau regard sur la microscopie électronique et d'ouvrir la porte à de nombreuses autres avancées sur la préparation des échantillons mais également pour les microscopes, les détecteurs etc. Tous ces développements sont à l'origine de ce qu'on qualifie aujourd'hui de « révolution en résolution ». (Kuhlbrandt, 2014) (von Loeffelholz et al., 2018)

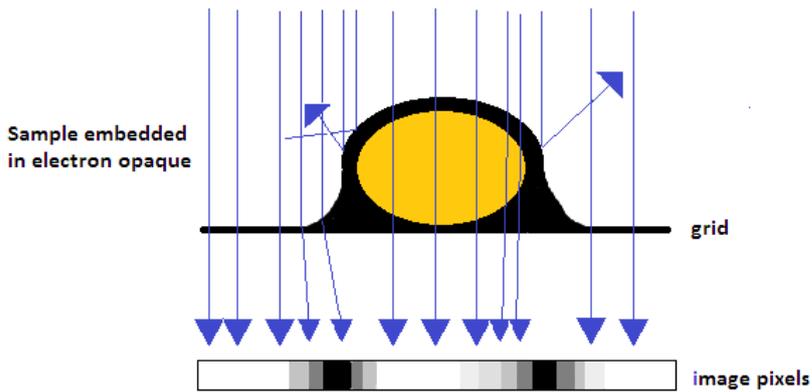
Néanmoins, deux majeures améliorations ont permis d'amener la microscopie électronique dans la zone des structures à résolution quasi-atomique :

- Les détecteurs : DDE et DQE (Figure 34 B)

La première avancée fut le développement de détecteurs directs d'électrons capable de détecter les électrons directement à une fréquence élevée d'images par second, sans nécessité le moindre arrêt durant l'acquisition. Ces détecteurs permettent dès lors, une amélioration de l'efficacité quantique (DQE pour Detective Quantum Efficiency). Ce terme DQE est utilisé pour décrire l'efficacité des détecteurs. Il correspond à la capacité des photons à être absorbés par le détecteur, qui les convertit en charge électrique, transférée et mesurée au niveau de chaque pixel. La valeur DQE est toujours inférieure à 1, 1 représentant un détecteur parfait pour lequel tout le signal sera transféré immédiatement permettant l'obtention d'images sans ajouter de bruit. Néanmoins, le temps d'acquisition d'un micrographe élevé (15/min) de ces détecteurs est à l'origine de la formation d'un flou dû au mouvement en x,y de l'échantillon par rapport au faisceau.

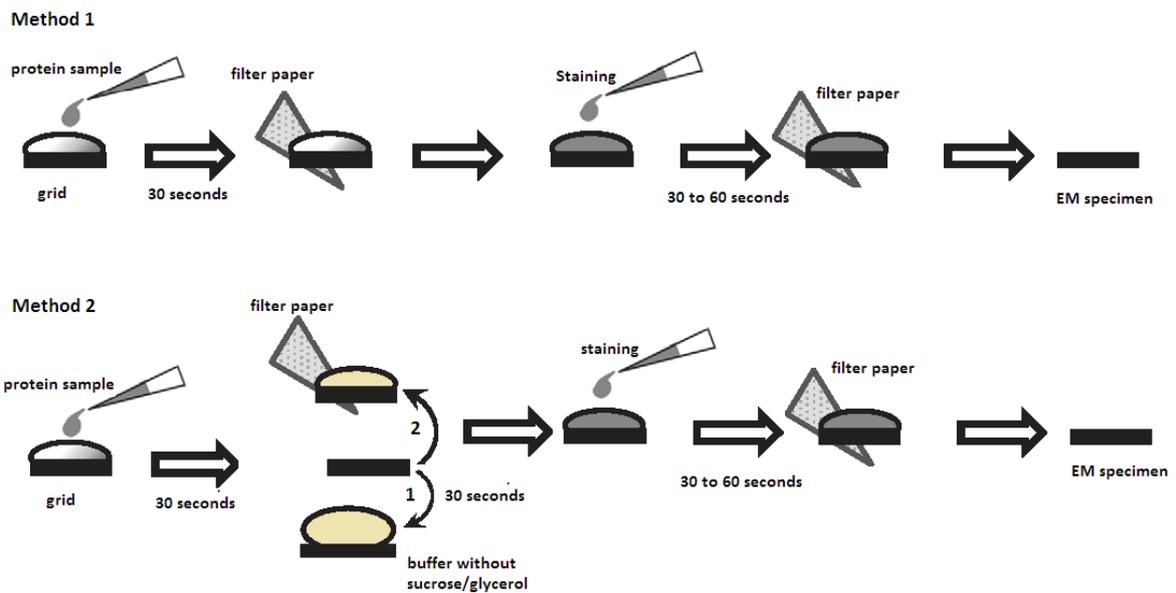
A)

Principe de coloration négative :



B)

Méthodes:



**Figure 35 Coloration négative** A) Coloration négative et contraste. B) Préparation des grilles de coloration négative.

Ce flou de mouvement entraîne une perte de contraste affectant la haute résolution. La correction de ces mouvements est corrigé par le mode film ou « movie ». Ce mode d'acquisition permet d'acquérir différentes images, durant une exposition. Ces images seront ensuite alignées afin de corriger le mouvement des particules dans la glace et formeront les micrographes.

- La seconde évolution est moins en lien avec la préparation des échantillons ou l'acquisition, mais tout aussi importante. Il s'agit du développement du traitement des données qui s'est caractérisée au fil des années par l'amélioration des performances de calcul (CPU, GPU etc...) et par l'évolution des logiciels (algorithmes).

En comparaison avec d'autres techniques de biologie structurale, la cryo-EM présente de nombreux avantages. Par exemple, la quantité de matériel nécessaire est inférieure à celle requise en RMN ou en cristallographie aux rayons-X. Elle nous permet également de nous affranchir des étapes limitantes de cristallisation et de traitement des cristaux. De plus, la microscopie nous permet de travailler en « real space » et donc d'éviter les problèmes de phase. Enfin, la cryo-EM peut dans une certaine mesure être utilisée pour des échantillons présentant une hétérogénéité et permettre l'obtention d'informations sur la dynamique des protéines ou des complexes protéiques. Bien que promettante, cette technique n'est pas parfaite et présente des limites du à l'imperfection des lentilles causant des aberrations sphériques (Cs) et chromatiques (Cc).

#### a) La coloration négative: qualité de l'échantillon.

La coloration négative est une technique de fixation développée en 1959 par Brenner et Horne. Elle permet d'accéder rapidement à la qualité de l'échantillon en termes d'homogénéité, de concentration, et de comportement de l'échantillon sur la grille.

Cette méthode se base sur l'adsorption de l'échantillon sur une grille de carbone, par la suite recouvert en surface par des sels de métaux lourds servant d'agent contrastant. Le principal agent est l'acétate d'uranyle, un composé dense aux électrons, qui va permettre de préserver l'échantillon en le protégeant contre les dégâts dus aux radiations mais surtout d'augmenter drastiquement le contraste des images. (Figure 35 A) Même si l'utilisation de ce type de sel nous permet de travailler à une dose d'électrons plus élevée et également d'augmenter le contraste d'amplitude, la coloration négative n'est pas adéquate pour les études structurales à

haute résolution et ne donnera que quelques informations sur la morphologie, la forme et la taille des particules. Les solutions de sels de métaux lourds, vont en effet enrober l'échantillon et interagir avec les zones hydrophile de ce dernier, ce qui va avoir pour effet le séchage de l'échantillon à l'origine de distorsions physiques. De plus, le pH acide des solutions va avoir un impact direct sur la stabilité du complexe, pouvant entraîner la dissociation ou l'agrégation des particules. Toutes ces limitations vont se répercuter sur la limite de résolution de cette technique, qui n'est que de 15 à 20 Å. Mais bien que peu résolutive elle permet d'obtenir de réelles informations soit par couplage avec d'autres techniques de biologie structurale soit comme technique de caractérisation en amont de la cryo-EM.

Le protocole utilisé pour les observations par coloration négative des différents échantillons est la suivante : les grilles subissent tout d'abord un traitement par décharge lumineuse, pendant 30 secondes à 200 mbar de pression d'air ou d'argon. Cette décharge a été réalisée à l'aide d'un appareil Elmo cordouan et permet de rendre la grille plus hydrophile et ainsi faciliter l'adsorption de l'échantillon. Deux techniques différentes adaptées du protocole classique dit de goutte à goutte ont été utilisées lors des préparations des échantillons.

Pour les échantillons ne présentant pas de composés chimiques tels que le saccharose ou le glycérol. 3 µL d'échantillon est pipeté sur la grille et laissé 30 secondes pour permettre l'adsorption de l'échantillon sur le carbone continu. L'excès de tampon est absorbé à l'aide d'un papier filtre appliqué sur la tranche de la grille. Puis 3 µL d'une solution d'acétate d'uranyle à 3% est ajouté et laissé 30 secondes à 1 minute avant d'être à son tour absorbée. Les échantillons ainsi colorés sont analysés directement par microscopie électronique (microscopes Tecnai TF20 ou CM 120)

La seconde méthode concerne les échantillons pour lesquelles on retrouve du saccharose ou du glycérol dans le tampon. Du fait de l'impact négatif de ces composés sur le contraste de l'image, ils doivent être éliminés. Pour cela j'ai utilisé une seconde méthode, présentée en figure 35 B. La grille avec l'échantillon est équilibré sur une goutte de tampon sans saccharose (ou glycérol) pendant environ 30 secondes. Si nécessaire, cette étape peut être répétée plusieurs fois. Puis comme précédemment, l'échantillon est coloré et séché avant d'être analysé.

## b) La cryo-microscopie électronique

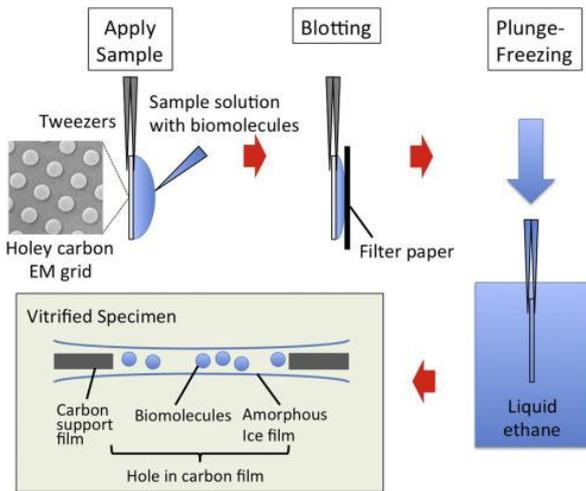
### **Préparation des grilles et congélation**

A la différence de la coloration négative, la préparation des grilles en cryo-EM ne nécessite pas l'utilisation d'un quelconque colorant. Dès lors, pour permettre la protection de l'échantillon des dommages causés par le faisceau d'électrons, l'échantillon est conservé dans une fine couche de glace amorphe, pour préserver au mieux l'état natif du complexe. En effet, la glace amorphe, est un état non-cristallin de la glace qui n'interagit donc pas avec le faisceau d'électrons à l'inverse de la glace hexagonale ou cubique. Pour obtenir cette glace amorphe, la phase de congélation se doit d'être extrêmement rapide et contrôlée pour éviter toute variation de température. Cette étape n'est donc pas manuelle mais assistée par un robot, le vitrobot, qui permet de déposer l'échantillon sur grille dans une enceinte à température et humidité contrôlée suivi de la congélation rapide dans un bain d'éthane liquide. (Figure 36 A) Ce contrôle permet d'éviter toute évaporation ou dégradation de l'échantillon lors de son dépôt sur la grille. L'absorption de l'excès de solution est réalisée ici, par un mécanisme mécanique entre deux papiers filtres puis la grille est congelée rapidement dans l'éthane. Tous ses paramètres sont ajustés par l'utilisateur pour obtenir un spécimen vitrifié d'une finesse parfaite avec une bonne répartition des particules et de la glace amorphe. L'épaisseur de la glace est un aspect limitant pour la qualité des grilles, une glace trop épaisse va réduire le contraste alors qu'une glace trop fine favorisera les orientations préférentielles ou l'exclusion des particules des trous. Il est donc primordial de contrôler ces paramètres d'absorption par les papiers filtres tel que la force et le temps de pression. Cette pression mécanique peut influencer les orientations préférentielles ou la formation d'agrégats mais également peut entraîner de véritables dégâts physiques sur les complexes et leur dissociation. C'est dans cette optique que certains laboratoires se sont lancés dans le développement de nouvelles machines afin d'optimiser la préparation des grilles et limiter les tensions physiques. On peut citer ici le spotiton, qui va permettre la répartition de l'échantillon sans l'intervention de papier filtre (Jain et al., 2012).

Similairement à la coloration négative, les grilles sont prétraitées par une décharge lumineuse pendant 15 à 30 secondes, pour créer une surface négative hydrophile mais aussi éliminer les contaminants hydrophobes. Cette étape nécessite également une optimisation, puisque l'hydrophilicité aura un effet sur l'orientation et la distribution des particules. Environ 3 $\mu$ L d'échantillon est adsorbé sur des grilles de carbone perforé.

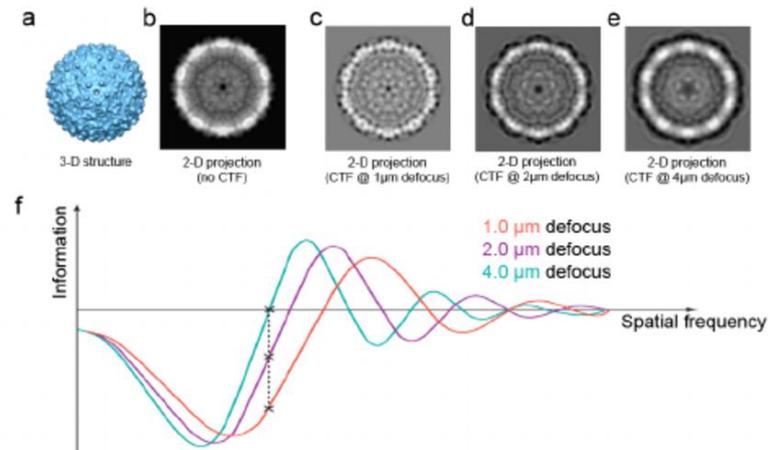
A)

### Préparation des grilles



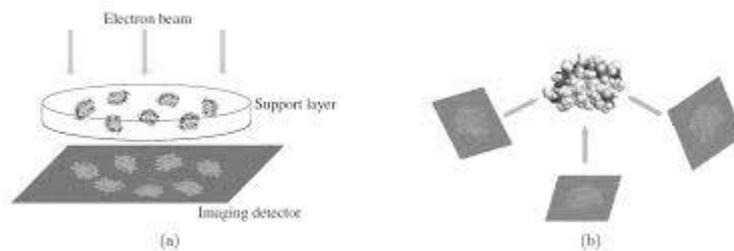
B)

### CTF



C)

### Traitement des données en cryo-EM



**Figure 36 Cryo-EM** A) Préparation et congélation des grilles de microscopie électronique. *Adaptée de Kazuyoshi Murata et al. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules, Biochimica et Biophysica Acta (BBA) - General Subjects 2018.* B) Fonction CTF. Les différentes figures montrent l'impact de la défocalisation sur la projection 2D simulée et sur le CTF. La courbe d'oscillation de la CTF diminue en amplitude à mesure que la fréquence augmente. De plus, nous pouvons voir certaines fréquences sans qu'aucune information ne soit transférée. Seule une combinaison à des défocalisations différentes peut restaurer complètement les informations à toutes les résolutions. *Adaptée de Neil A. Ranson et al. Cryo-Electron Microscopy of Viruses, Emerging Topics in Physical Virology 2010* C) des projections 2D à la structure 3D.

Le choix du type de grille est important et va influencer la qualité des échantillons. Il existe dans le commerce une variété importante de grilles, différant par le matériel utilisé (cuivre, or, molybdène etc...), la taille et la forme des trous ou encore par la présence ou non de support (carbone, or fin...). Chaque type de grille présente des propriétés spécifiques ayant pour effet de contrôler la répartition des particules, ou d'influencer l'épaisseur de la glace par exemple. Après congélation les grilles sont placées dans la colonne du microscope sous vide à basse température pour empêcher toute transformation de la glace et réduire les effets d'irradiations sur les complexes.

### **La formation de l'image**

Toute image, provient de la différence d'intensité entre l'objet d'intérêt et les régions environnantes, et c'est cette différence qui entraîne le contraste. Le contraste nous permet donc de discerner les différents composants d'une image.

La formation de l'image en microscopie électronique à transmission provient de l'interaction entre le faisceau d'électrons et l'échantillon.

Il est important de noter que les échantillons biologiques sont caractérisés de « transparent aux électrons », en effet les atomes composants ces particules (H, C, O, N, P, S) sont majoritairement des atomes légers qui n'interagiront que faiblement avec le faisceau conduisant donc à un faible contraste.

Comme on peut le voir sur la figure 34 C, les électrons vont interagir de différentes manières au contact de l'échantillon. Une partie du faisceau n'interagira pas avec l'échantillon et ne sera ni dévié ni énergétiquement modifié. Deux types de diffusion sont observés, une diffusion élastique où les électrons sont déviés de leur trajectoire sans perte d'énergie. Une diffusion inélastique qui va contribuer à la formation du bruit en MET. En effet, dans ce cas les électrons déviés vont transférer de l'énergie à l'échantillon causant des dégâts au niveau moléculaire. Pour minimiser ces dommages dus à l'irradiation, on travaille à une faible dose d'électrons qui conduit également à un faible ratio signal sur bruit (SNR). L'absorption d'une partie du faisceau est responsable du contraste d'amplitude et est de faible intensité. Un second type de contraste, vient du décalage de phase entre les électrons non déviés et déviés.

Le contraste des images en microscopie électronique à transmission peut être dû à deux effets liés à l'amplitude et à la phase des ondes électroniques impliqués.

Le contraste d'amplitude : L'amplitude va être modifiée lors de l'absorption ou la diffraction des électrons. De ce fait, les parties plus épaisses ou composés d'atomes plus lourds

apparaissent plus sombres sur l'image. Ce contraste est négligeable en cryo-EM mais plus important en coloration négative du fait de l'utilisation de sels de métaux lourds.

Le contraste de phase provient de la superposition d'ondes qui ne sont pas en phase et provoque également une diminution de l'amplitude et de ce fait un contraste. Ce contraste de phase permet la visualisation de détail du spécimen composé d'éléments à faible nombre atomique qui ne donneraient pas lieu à un contraste d'amplitude.

Les lentilles vont permettre la focalisation des électrons sur l'échantillon. Puis les électrons transmis passent une dernière fois à travers des lentilles de projection afin d'obtenir un agrandissement, avant d'être collectés par le système d'acquisition (Figure 35 A). Ces lentilles ne sont néanmoins pas parfaites et vont entraîner une limite physique à la formation de l'image due aux aberrations sphérique et chromatique.

### **Défocalisation et contraste de phase**

Le transfert de l'information par le microscope peut être représentée mathématiquement par une fonction, appelée fonction de transfert de contraste (CTF). Cette fonction représente le contraste dépendant de la fréquence spatiale, décrivant l'impact des aberrations sur l'image (Figure 36 B). Cette fonction a une amplitude d'oscillation comprise entre 1 et -1, caractéristique de l'inversion de contraste. Comme on peut le voir, la fonction rencontre le zéro plusieurs fois sur l'axe X, entraînant une perte de l'information. En faisant varier la défocalisation, la trajectoire de diffusion élastique des électrons varie également, ce qui se répercute sur la CTF par un décalage de la courbe en fonction des fréquences. La collecte de données dans une plage de défocalisation permettra donc de récupérer les informations manquantes.

En conservant une défocalisation faible, on observe moins de transitions entre le contraste positif et le contraste négatif, ce qui se traduit par une perte moindre d'informations à haute résolution. Cependant, une défocalisation plus élevée confère plus de contraste aux résolutions plus faibles, ce qui rend les particules plus facilement visibles sur les micrographes. Il est important de souligner ici la présence du Volta Phase Plate installé sur le Titan Krios de l'institut. Ce dispositif va permettre de collecter des données proches de la focale avec une plage de défocalisation plus réduite et donc va augmenter le contraste.

## Le traitement d'images

On parle communément d'analyse en particules isolées. Les données collectées sont un ensemble de micrographes, présentant une multitude de copies de la même macromolécule. Ces macromolécules, sont figées dans la glace sous différentes orientations, et chaque particule est une représentation bidimensionnelle (2D) du complexe 3D. Elles correspondent en effet à une projection 2D d'un objet tridimensionnel (3D).

La correction de la CTF est la première étape réalisée et elle permet de restaurer, du moins en partie, les détails à haute résolution perdu du fait notamment de l'imperfection des lentilles et la sous-focalisation. Cette étape a été réalisée à l'aide du programme Gctf.

Les particules isolées sur les micrographes vont être sélectionnée dans des boîtes de taille défini et caractéristiques de la taille des particules étudiées de manière à ce que la particule représente  $2/3$  de la boîte. Cette étape consiste à implémenter la position des particules dans chaque micrographe. Pour permettre l'obtention d'une première reconstruction 3D à basse résolution, quelques milliers de particules peuvent être suffisant. Alors que pour une structure à haute résolution, plusieurs centaines de milliers de particules sont requises. Cette étape est principalement réalisée de façon automatique ou semi-automatique par les logiciels de traitement tels que RELION, EMAN2, cisTEM etc... Néanmoins, pour les données collectées, la qualité de l'échantillon ne permettait pas l'utilisation de cet auto-sélection et les particules ont donc dû être sélectionnées manuellement. Les particules sélectionnées sont ensuite classées par la méthode de maximum de vraisemblance. Chaque classe est spécifique d'une orientation et représente l'addition de plusieurs particules augmentant le contraste du fait de l'augmentation du rapport signal sur bruit. (Figure 36 C).

Coloration négative		Cryo-EM	
Avantages	Inconvénients	Avantages	Inconvénients
<ul style="list-style-type: none"> <li>- Contraste important</li> <li>- Résistant aux radiations</li> <li>- Simple et rapide à mettre en œuvre</li> </ul>	<ul style="list-style-type: none"> <li>- Entraîne des distorsions</li> <li>- Les images obtenues sont caractéristiques de la coloration et non de l'échantillon</li> <li>- Limite de pénétration du colorant</li> <li>- Conditions physiologiques modifiées dû au pH du colorant</li> <li>- Limite de résolution</li> </ul>	<ul style="list-style-type: none"> <li>- Préparation : pas de colorant donc on évite tous les inconvénients suscité de leurs utilisations</li> <li>- Image de l'échantillon directement</li> <li>- Les images sont des projections de l'objet 3D entière et non uniquement une surface</li> </ul>	<ul style="list-style-type: none"> <li>- Faible contraste</li> <li>- Sensible aux irradiations</li> </ul>

**Tableau 4 comparaison générale**

## II. Résultats

### A. Stabilité du complexe NuRD et la sous-unité CHD4

La première purification du complexe NuRD fut réalisée en collaboration avec le Dr. Khalid Ouarahni de l'équipe du Dr. Ali Hamiche. Le protocole utilisé suit celui présenté dans la section « matériels et méthodes » page 82. Après resuspension des cellules dans le tampon A (10 mM TRIS pH 7.65, 1.5m M MgCl<sub>2</sub>, 10m M KCl, 0.2m M PMSF), les cellules sont cassées par l'utilisation d'un Dounce B et 0.33 mL par gramme de cellules de tampon B (20m M TRIS pH 7.65, 60 mM NaCl, 15 mM KCl, 0.34 M sucrose, 0.15 mM spermine et 0.5 mM spermidine) est utilisé pour transférer l'échantillon dans un tube falcon. L'échantillon est centrifugé 10 min à 5000 rpm à 4°C après l'ajout d'inhibiteur de protéase.

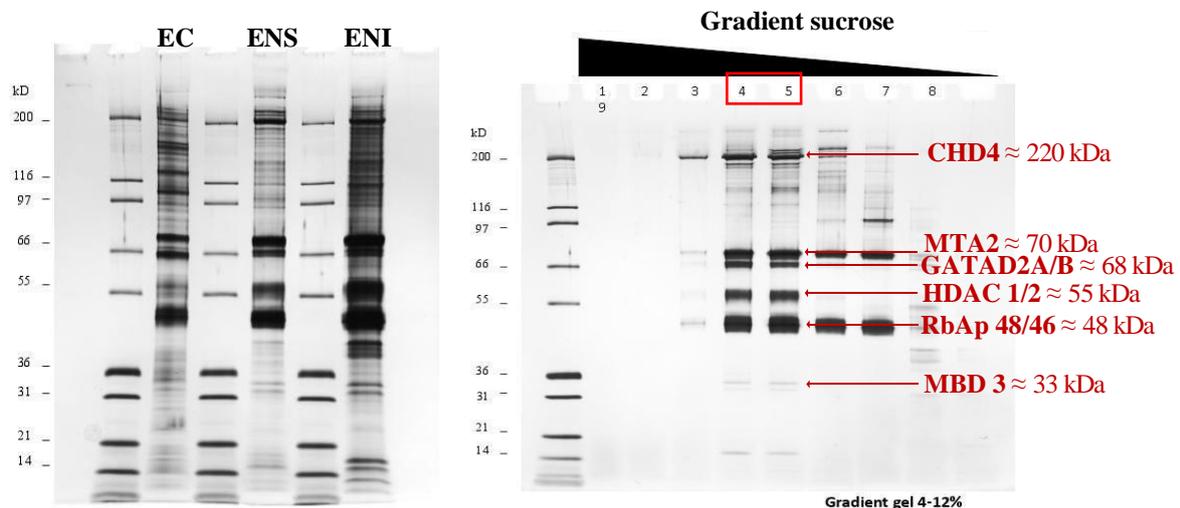
Les noyaux récupérés sont resuspendus dans le même tampon B et le tampon à haut sel (20 mM TRIS pH 7.65, 1.5 mM MgCl<sub>2</sub>, 900 mM NaCl, 0.2 mM EDTA, 25% glycérol, 0.2 mM PMSF) est ajouté aux gouttes à gouttes. Une étape de centrifugation pendant 10min à 5500 rpm permet de séparer le surnageant représentant la fraction nucléaire soluble (ENS) du culot ou fraction nucléaire insoluble.

La fraction nucléaire insoluble (ENI) est solubilisée dans le tampon B auquel est ajouté 1 mM de CaCl<sub>2</sub>. La solution est placée dans un bain marie à 37°C pendant 1 min. Cette étape est primordiale pour la préparation de l'activation de la nucléase permettant de réduire la chromatine en mono-nucléosome. La nucléase, DNase micrococcace, est ajoutée à une concentration de 2.5 U par gramme de cellules et le tout est mis à incubé 10 min à 37°C. La réaction est stoppée par l'ajout d'EDTA (4mM). Pour finir l'échantillon ENI est soniqué 3 fois pendant 1 minute avec une amplitude de 10% sur glace. Comme pour les échantillons précédents du PIC est ajouté avant la centrifugation.

Les trois échantillons sont ultracentrifugés pendant 1 heure à 35 000 rpm pour culoter les débris cellulaires.

La seconde étape de purification a été réalisée avec la résine FLAG. Lors de cette étape le complexe a été incubé sur la nuit avec la résine pré-équilibrée puis éluée le lendemain pendant 6 heures (20 mM TRIS pH 7.65, 3 mM MgCl<sub>2</sub>, 150 mM NaCl, 0.1 mM EDTA, 10% glycérol, 0.01% NP40) avec 0.5 mg/mL de peptide FAG.

La qualité de la purification a d'abord été vérifiée par gel SDS-PAGE, présentés en figure 37. On observe pour les fractions ENS et ENI différentes bandes dont les fronts de migration correspondent aux poids moléculaires des différentes sous-unités du complexe. De façon intéressante on peut observer différentes bandes dans la zone de migration de MBD3 pouvant correspondre aux différentes isoformes de la protéine. Pour toutes les fractions on observe également la présence de bandes supplémentaires pouvant être dû à des contaminants ou encore à la présence de sous-unités sub-stoechiométriques ou spécifiques des cellules HeLa. Lors de cette première purification la fraction cytoplasmique a été étudiée, et on peut observer dans cette fraction une co-purification des sous-unités de NuRD. Néanmoins, la stoechiométrie même du complexe cytoplasmique semble différente des fractions nucléaires.

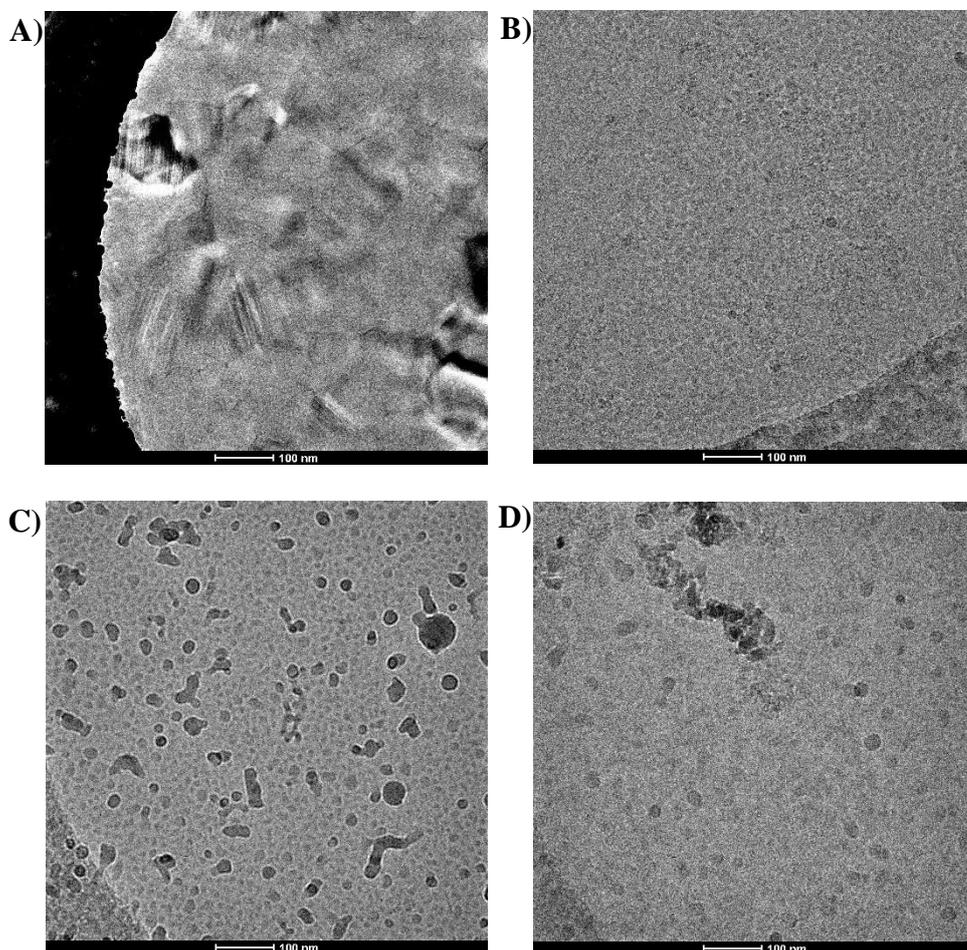


**Figure 37 Purification du complexe NuRD, SDS-PAGE.** Le premier gel présente les trois fractions après l'éluion sur la résine FLAG. La fraction cytoplasmique est annotée EC, la fraction nucléaire soluble ENS, et l'insoluble ENI. A droite on observe un gel des différentes fractions post-gradient, les fractions 4 et 5 sont les fractions sélectionnées pour l'étude structurale.

Les fractions 4 et 5 post-gradient de saccharose semblent être les plus concentrées et possèdent les sous-unités principales du complexe NuRD. Les fractions 6 et 7 migrent quant à elles à un plus faible pourcentage de saccharose, équivalent à une densité du complexe plus faible. On observe en effet sur le gel une diminution d'intensité pour les bandes qui correspondent aux sous-unités CHD4, GATAD2A/B et MBD3. Cette perte d'intensité peut

être due à une perte des sous-unités ou une modification stœchiométrique du complexe NuRD.

Les deux fractions 4 et 5 ont été dialysées sur la nuit contre un tampon E (20 mM HEPES pH7, 150 mM NaCl, 4 mM MgCl<sub>2</sub>). Cette dialyse avait pour but d'éliminer la présence de saccharose ou du moins la réduire à moins de 1%. En effet le saccharose, bien que nécessaire à la purification du complexe, est aussi connu pour ses caractéristiques cryo-protectantes. Le saccharose va donc interférer avec le processus de congélation des grilles et même à faible concentration va diminuer le contraste des images (Figure 38 - A et D).

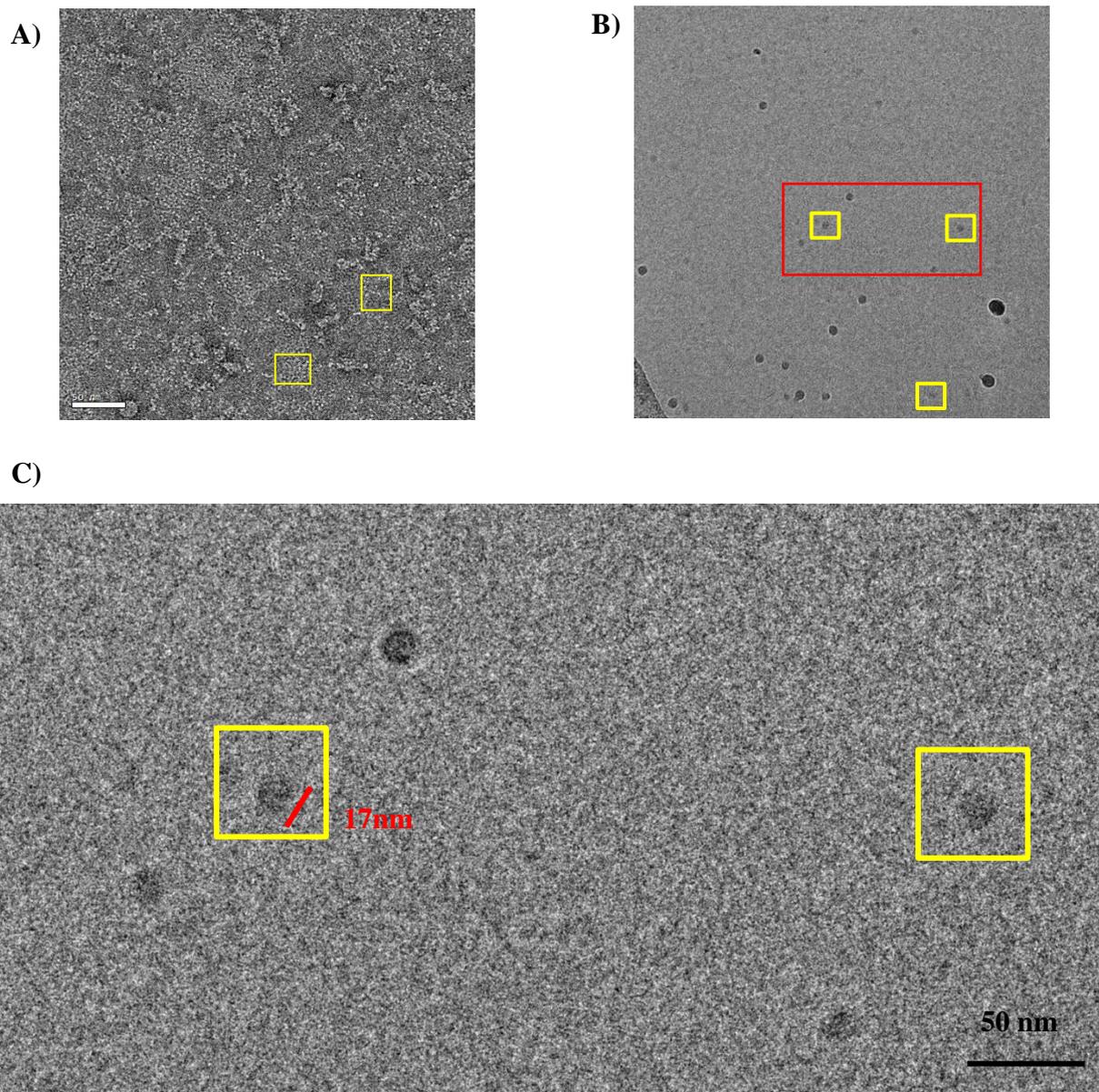


**Figure 38 Image des contaminations et type de glace en EM.** A) Concentration élevée de saccharose dans l'échantillon, présence de glace. B) Présence de glace cristalline due au réchauffement de la grille. Ce réchauffement entraîne une transformation de la glace et peut être dû à la manipulation même de la grille ou un contact avec le clip non adéquat ou simplement un échantillon trop épais ralentissant la congélation. C) Transformation de la grille lors de son réchauffement dans le bras de charge du Polara. On observe également une contamination importante à l'éthane. D) Présence de saccharose dans l'échantillon qui entraîne un flou et une perte de contraste.

L'échantillon a d'abord été étudié par coloration négative pour obtenir des informations sur son comportement sur grille mais également en termes de concentration, agrégation et dissociation. Communément, une bonne concentration de particules correspond à une quantité permettant d'observer une couche de complexes non superposés, bien répartie et également en quantité importante. Lors du passage de l'observation en coloration négative à l'observation en conditions cryogénique, il est nécessaire que cette concentration soit de 5 à 10 fois supérieure pour compenser les artefacts de la congélation.

Pour ces tests, 3  $\mu\text{L}$  d'échantillon ont été appliqués pendant 30 secondes sur une grille traitée au préalable par décharge lumineuse. L'échantillon est ensuite absorbé latéralement par papier filtre avant d'être mis pendant 30 secondes en contact avec l'acétate d'uranyle. Après élimination de l'excès de colorant, les grilles sont imagées directement à l'aide d'un microscope CM120. Mise à part un environnement qui semble contaminé, probablement dû à la mauvaise coloration de l'échantillon, on peut observer, en figure 39 A, la présence de particules d'environ 20 nm. Le bruit de fond observé peut être le résultat également d'une dissociation/agrégation du complexe.

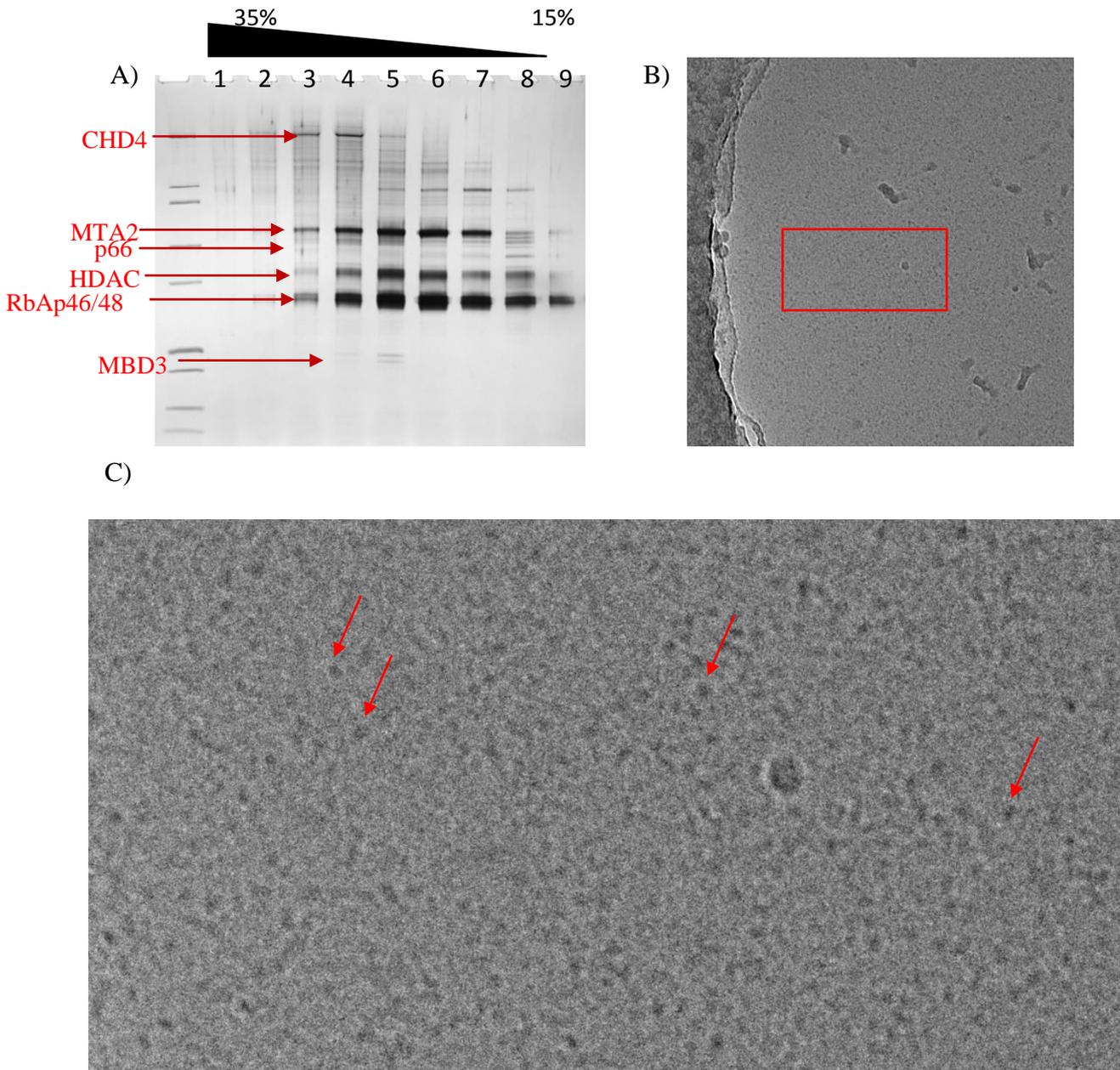
Le même échantillon a été utilisé pour réaliser des grilles de microscopie électronique en condition cryogénique. 3  $\mu\text{L}$  de cette même fraction a été déposée sur une grille Quantifoil R2/2 traitée pendant 30 secondes par décharge lumineuse au préalable. Pour ces expériences les paramètres de pression du papier filtre réglés sur le vitrobot sont les suivants : la force est fixée à 6 et le temps de 1 seconde. Le microscope Tecnai Polara a été utilisé pour l'observation des grilles à un voltage d'accélération de 100 kV afin d'augmenter le contraste en comparaison des 300 kV normalement utilisés lors de la collecte de données. Ce criblage des grilles réalisées est manuel et nécessite l'utilisation du logiciel EPU. La figure 39 A représente une des images réalisées lors de ce criblage et présente un défocus de  $-5 \mu\text{m}$  et un grossissement de 59 000 fois. On observe des particules contrastées mais en très faible quantité.



**Figure 39 Coloration négative et cryo-EM images** A) Image en coloration négative de l'échantillon ENS réalisé sur le CM120. B) Image par cryo-EM du même échantillon en jaune sont encadrés les particules. C) Grandissement de la zone encadrée en rouge de l'image B).

Lors de ces premières observations, j'ai été confronté aux principaux problèmes récurrents de l'étude structurale de ce complexe qui sont la faible concentration due à la production endogène et également l'instabilité du complexe qui comme on peut le remarquer sur l'image en coloration négative présente une forte propension à l'agrégation. On peut se demander quelle est la stabilité du complexe au cours du temps au sein même des cellules HeLa, en vue d'optimiser le temps de préparation des cellules. En effet, en partant de cellules fraîches, le

temps de génération est d'approximativement trois semaines. Dans le but de réduire ce temps de préparation, les cellules ont été laissées dans l'incubateur et divisées pour atteindre la même quantité précédemment utilisée. La purification a été réalisée selon le même protocole et les résultats sont présentés en figure 40.



**Figure 40 NuRD analyses et stabilité.** A) SDS-PAGE post-gradient de saccharose de la fraction ENI. B) Image de cryo-EM de la fraction 5. C) Détail du tapis de particules dissociées (encadré en rouge sur l'image B). On peut observer de nombreuses particules de très petites tailles réparties sur la grille pointées par une flèche.

Le gel de SDS-PAGE des différentes fractions du gradient, permet comme précédemment d'identifier les différentes sous-unités principales de NuRD. Néanmoins, on observe que la fraction la plus représentative de l'échantillon n'est plus stabilisée dans la même zone du gradient. On observe donc une variation de la densité du complexe caractérisée par une absence de la sous-unité CHD4. Les sous-unités GATAD2A et 2B semblent également moins présentes sur le gel. La perte de la sous-unité CHD4 ne semble pas impacter la formation d'un complexe NuRD, laissant supposer une possible position périphérique de cette sous-unité. La même hypothèse peut être émise pour la sous-unité GATAD qui ne se localiserait pas au niveau du cœur du complexe. Au vu de ces résultats des grilles ont été réalisées pour la fraction 5, après une dialyse sur la nuit pour éliminer le saccharose. Les grilles présentaient toute une dissociation totale du complexe. Comme on peut le voir en figure 40 B on observe un tapis de particules de très petites tailles, représentant les différentes sous-unités de NuRD. La perte des sous-unités CHD4 et GATAD2A/B ne semble pas perturber la formation d'un complexe mais impacte le comportement du complexe au contact de la grille et lors de la congélation.

Pour conclure, le complexe NuRD présente une importante instabilité lors de son étude en microscopie, avec une tendance à la dissociation et l'agrégation. L'hypothèse sur la localisation en périphérie de la sous-unité CHD4 a été confirmée par les travaux publiés en 2016 par Low qui ont mis en évidence l'existence d'un complexe NuDe, formé en l'absence de CHD4 et le caractère dynamique de cette sous-unité périphérique (Low et al., 2016).

Le comportement du complexe NuRD sur les grilles et sa dissociation ou agrégation pourrait venir des changements du tampon juste avant la congélation. En effet, en sortie de gradient le complexe se trouve dans un milieu riche en saccharose, un agent stabilisant de type osmolyte. Ainsi, en plus d'augmenter la viscosité de l'échantillon, la présence d'osmolytes permet la stabilisation de complexes ou protéines sous leur forme structurée. L'élimination de ce composé peut donc entraîner des dissociations et l'agrégation des états dénaturés des sous-unités du complexe.

## B. Dissociation et instabilité du complexe NuRD

De manière à optimiser la stabilité des échantillons et maintenir la structure 3D des particules sur les grilles, de nombreux tests de mise au point ont été effectués, que ce soit en termes de pH, solution tampon, sels et quantités de sel utilisées, détergents, ou encore le temps et la force de pression impactant sur la finesse de la glace. Une des optimisations usuelles pour la réduction de l'agrégation est l'utilisation d'agents réducteurs, permettant de limiter l'oxydation des cystéines qui conduisent à la formation de ponts disulfures inappropriés. Le choix de l'agent réducteur se fait selon sa stabilité, son pouvoir réducteur et sa compatibilité avec les techniques employées. Au final, 5 à 3 mM de DTT ont été ajoutés aux tampons A à D, remplacée par la suite par du TCEP à 0.5 ou 1 mM.

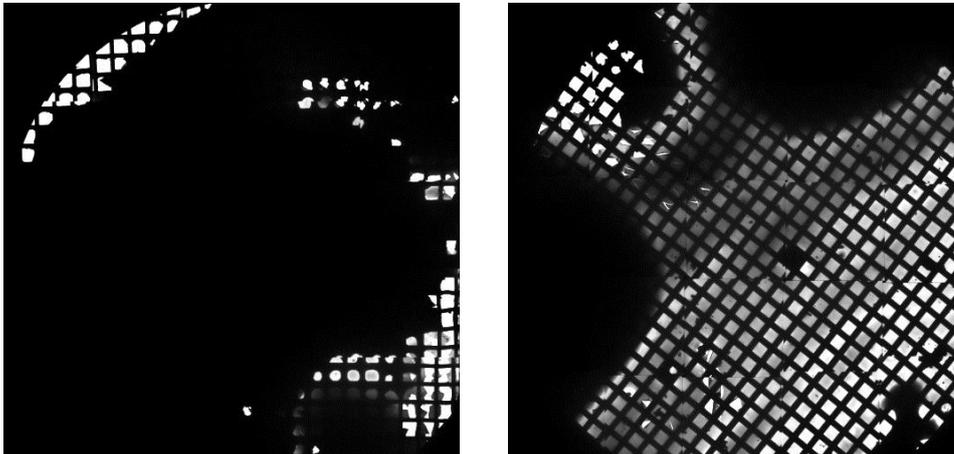
Comme on a pu le voir tout le long de l'introduction mais également lors des premières expériences réalisées, le complexe NuRD est connu pour son hétérogénéité, qui semble due principalement ici à l'aspect dynamique du complexe. Pour augmenter l'homogénéité des fractions, le volume des fractions collectées suite au gradient a été réduit de 500  $\mu\text{L}$  à 165  $\mu\text{L}$ .

Lors de la préparation d'échantillon de complexe macromoléculaire, il est commun de réaliser un gradient de saccharose ou de glycérol. Néanmoins, comme on a pu le voir, ces composés sont inadéquats à l'étude par cryo-EM. Cette étape permet ici d'une part l'élimination du peptide FLAG qui diminue le contraste du fait de la formation d'un tapis protéique mais également l'obtention de fractions homogènes. En effet, les différents sous-complexes vont pouvoir être séparés. Le désavantage étant que l'élimination de ce genre de composés est relativement compliquée et peut endommager le complexe mais également leur élimination entraîne une dilution de l'échantillon. Lors de la dialyse des fractions le saccharose ne peut diffuser au travers de la membrane. Ainsi, les molécules d'eau et les ions vont diffuser pour équilibrer les compartiments par osmolarité. En conséquence, le volume de l'échantillon est multiplié par deux après la dialyse et la concentration sera réduite par deux. La question se pose de savoir s'il est possible dès lors d'éliminer le peptide FLAG sans avoir à réaliser le gradient de saccharose de façon non-agressive. Pour ces tests, différentes techniques avant et après gradient de densité ont été comparées et analysées en cryo-EM. (Figure 42).

Pour l'observation des grilles la préparation est quasi similaire à précédemment; 2.5  $\mu\text{L}$  d'échantillon ont été déposés sur des grilles Quantifoil R2/2 traité par décharge lumineuse durant 20 secondes. La réduction de 3 à 2.5  $\mu\text{L}$  d'échantillon a résulté en la présence de glace épaisse sur l'intégralité des grilles. La quantité n'étant pas adaptée à la répartition adéquate de

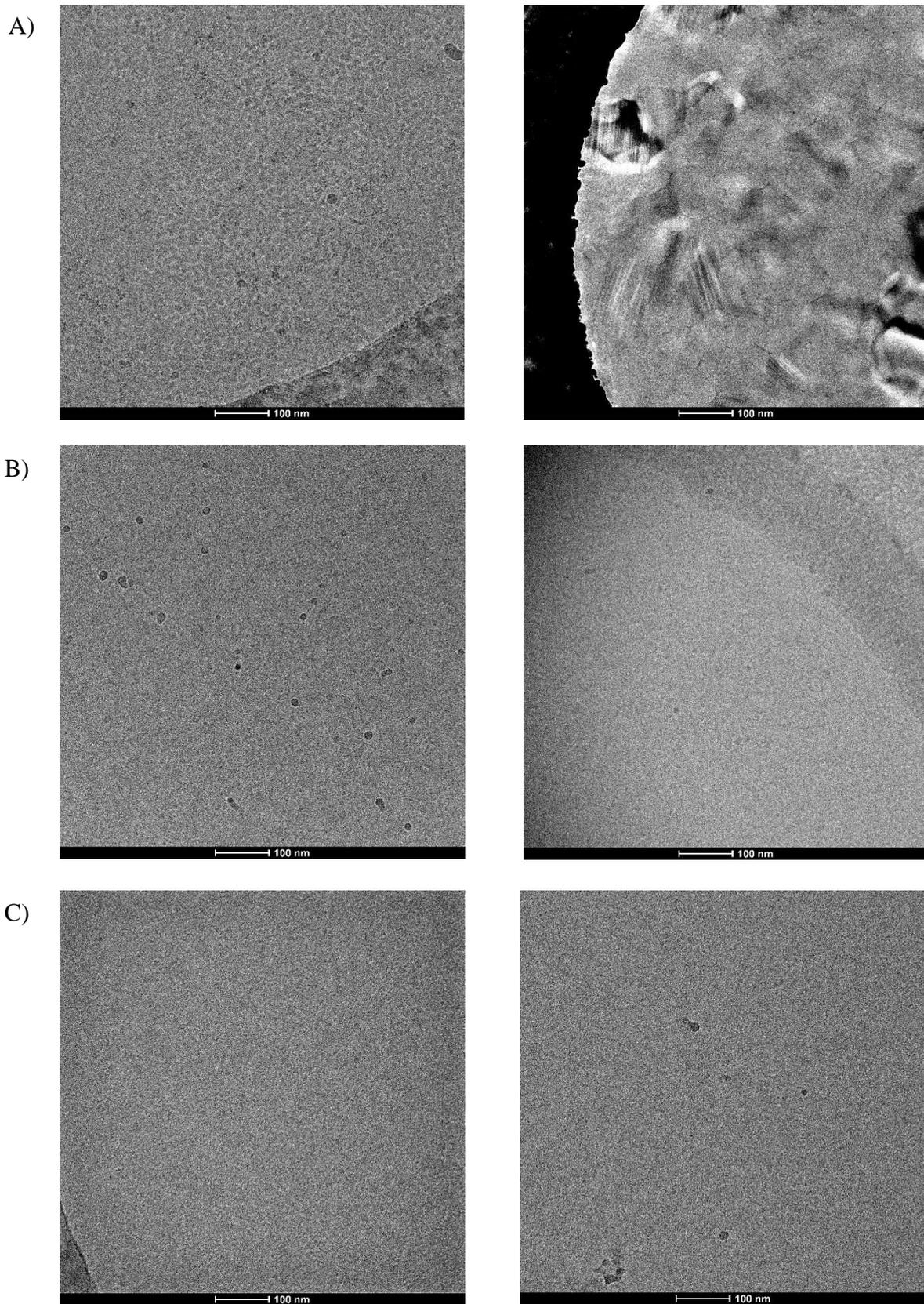
l'échantillon sur la grille. Bien que la différence soit minime la réalisation de nouvelles grilles, utilisant les mêmes paramètres mais avec 3  $\mu\text{L}$  d'échantillon a permis l'obtention de glace fine permettant l'analyse des échantillons figure 41.

Pour cette partie, les analyses de microscopie sont réalisées sur un microscope électronique Tecnai Polara opérant à 300 kV, équipé d'une caméra Falcon I, avec une défocalisation de -5  $\mu\text{m}$  et un grandissement de l'ordre de 59000x. Généralement, le criblage de données est réalisé à 100 kV de manière à augmenter le contraste. Le complexe NuRD étant de l'ordre du MégaDalton, cela nous permet de travailler à 300 kV, la diminution du contraste sera en partie compensée par la taille du complexe.



**Figure 41** glace épaisse vs glace fine. A gauche 2.5  $\mu\text{L}$  d'échantillon sont déposés contre 3  $\mu\text{L}$  à droite.

En amont du gradient, seul le dessalement par l'utilisation des colonnes zeba spin a été utilisé, la dialyse n'étant pas adaptée à l'élimination de composés peptidiques. Les colonnes zeba spin ne permettent en effet pas uniquement le changement de tampon mais vont favoriser l'ultrafiltration de petits composés d'environ 12k Da ou moins. Néanmoins, comme on peut l'observer en figure 42 A, tous les essais ont abouti à la formation de glace relativement épaisse avec une anatomie en forme d'écaille de serpent ou de glace non-amorphe, probablement dû à une élimination non optimale du glycérol.



**Figure 42 dessalement et dialyse** A) Dessalement après élution pour les extraits ENS à gauche et ENI à droite. B) Dessalement pour l'extrait ENI sur grille d'or à gauche et carbone à droite. C) Dialyse pour l'extrait ENS à gauche et ENI à droite.

Au vu de l'inefficacité d'élimination du peptide FLAG directement suite à l'élution, j'ai lancé un gradient de saccharose sur la nuit (16h à 35 000 rpm), suite auquel les échantillons concentrés en complexe NuRD ont été dessalés ou dialysés et les deux techniques ont pu être comparées par microscopie (tampon : 20 mM HEPES pH8, 150 mM NaCl, 3 mM MgCl<sub>2</sub>, 1% glycérol). Les deux techniques présentent leurs avantages et leurs inconvénients, ainsi bien que le dessalement soit rapide il entraîne la perte d'une partie de l'échantillon. La dialyse est beaucoup plus longue, de ce fait le complexe peut se dissocier au cours du temps, néanmoins les forces exercées lors de cette étape sont moins importantes.

Les deux techniques entraînent une perte ou une dilution des fractions, l'effet sur la concentration est donc relativement similaire. Comme lors des premiers tests, on observe que peu de particules présentes sur les grilles (Figure 42 B et C). Les images sont également assez floues. Bien que les étapes de purification semblent fonctionner, le complexe n'est toujours que peu visible dans les trous des grilles. Les raisons peuvent être nombreuses pour expliquer ces observations et ont guidées les nombreux tests effectués par la suite. Je vais décrire ici brièvement les stratégies utilisées.

- Considérant que l'échantillon ne soit pas assez concentré. La première idée serait de reconcentrer les échantillons. Les tests de concentration utilisant les systèmes amicon ont tous confirmé la sensibilité du complexe. Ainsi, lors de l'observation d'échantillons concentrés je n'ai pu observer qu'un tapis protéique caractéristique de la dissociation du complexe. La seule option fut donc ici d'augmenter la quantité de cellules utilisée et réduire les volumes d'élution. Précédemment, les purifications du complexe NuRD étaient réalisées avec 8 g de cellules HeLa, concentration extrêmement faible pour une étude structurale sachant que le complexe n'est pas surexprimé.
- Dissociation du complexe à l'interface air-tampon. Cette hypothèse sera décrite dans les parties suivantes et permettra de décrire l'intérêt de l'utilisation de composé de pontage tel que le glutaraldéhyde ou encore l'utilisation des grilles de carbone continu.
- L'une des « dernières » hypothèses est due à la préférence d'interaction du complexe pour le carbone. Ainsi le complexe ne se retrouvera pas dans les ménisques de glace. Pour éviter ce genre de problème, la composition des grilles peut être changée (ex grilles en or ou recouverte de carbone).

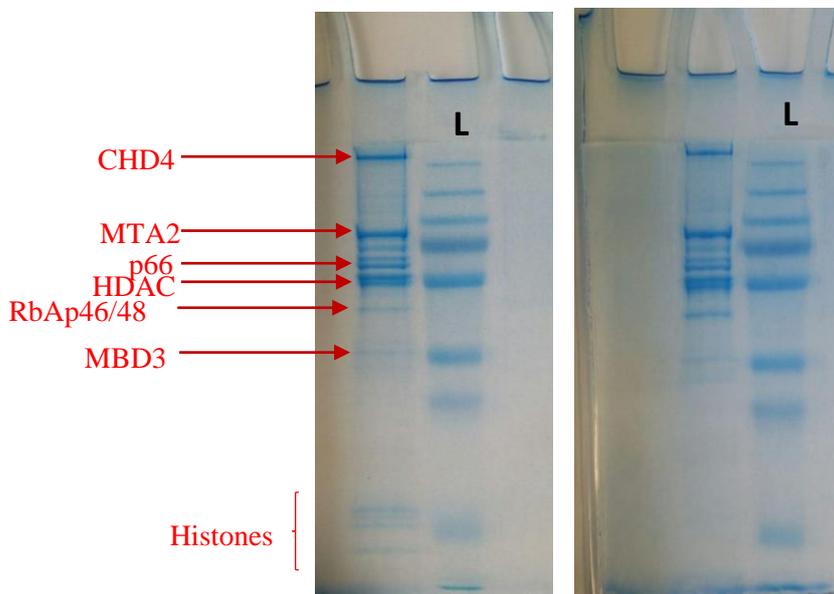
Ainsi, lors de tests supplémentaires, j'ai décidé d'utiliser des grilles d'or au lieu des classique R2/2 en cuivre avec un support carbone. Comme on peut le voir sur la figure 42 B, le contraste est amélioré lors de l'utilisation de ces grilles ce qui permet d'identifier quelques

particules d'environ 20 nm. Dû à une meilleure conductivité, les grilles en or permettent de réduire les mouvements induits par le faisceau et donc de réduire la dérive. On observe également plus nettement la dissociation environnement du complexe qui tapisse la glace.

Au vu de ces informations j'ai décidé donc de doubler la quantité de cellules et réduire le volume d'élution de la résine FLAG. Enfin, l'emploi d'un gradient d'agent de fixation en parallèle du gradient de densité a été utilisé.

### C. Le gradient de type GraFix

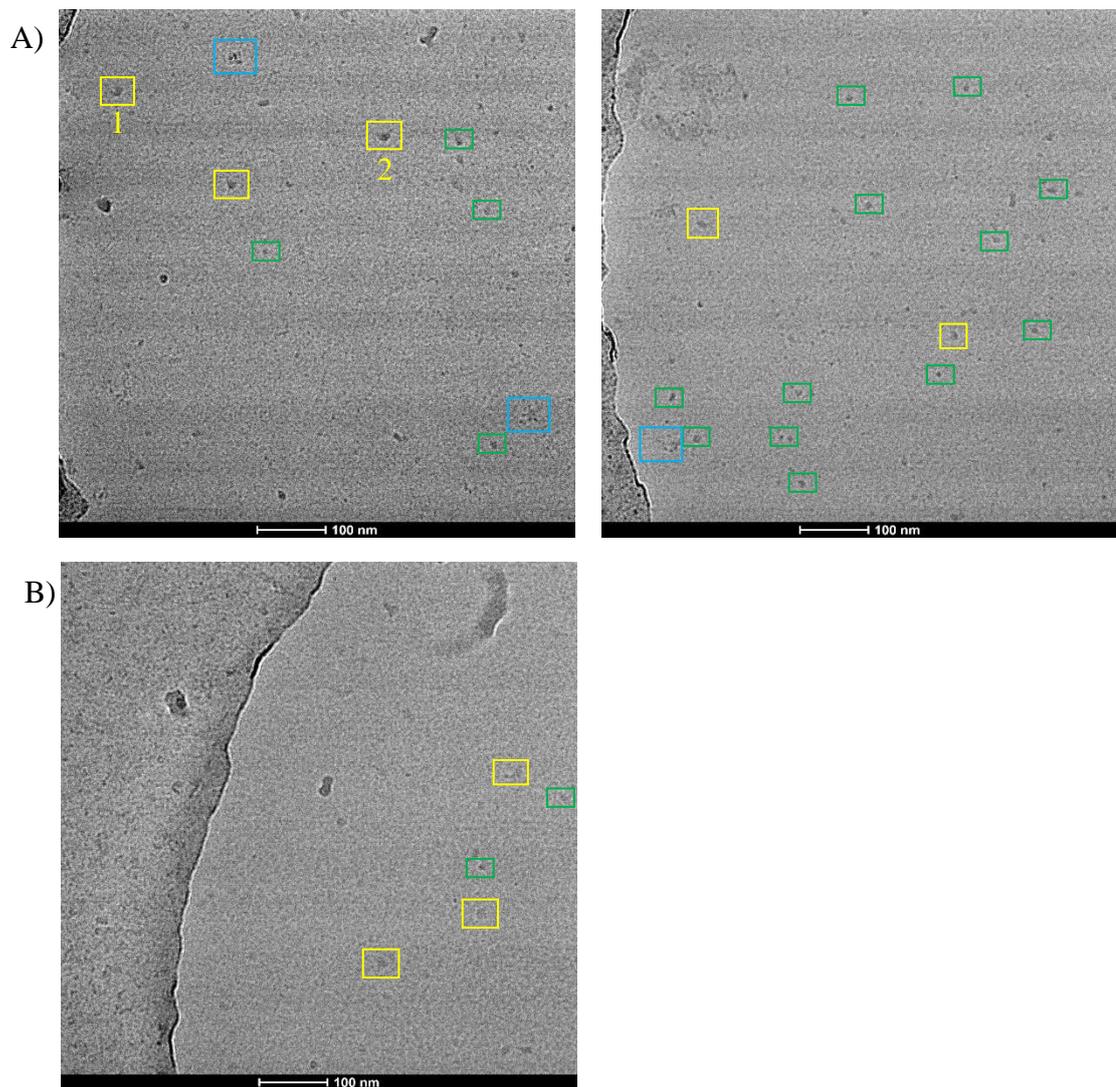
Les nouvelles purifications du complexe NuRD ont été réalisées à partir de 15 grammes de cellules, deux fois plus que précédemment (Figure 43).



**Figure 43** Caractérisation des fractions ENI et ENS. (A) SDS-PAGE suite à l'élution de la fraction insoluble à gauche, on peut noter la présence dans la partie inférieure de bandes caractéristiques des histones et donc de la présence de nucléosomes et soluble à droite.

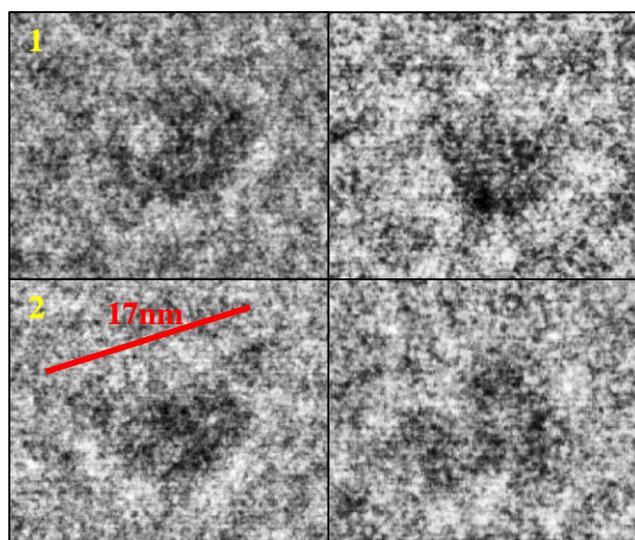
Le protocole est le même que présenté précédemment, excepté pour le gradient de densité qui a été modifié par l'ajout d'un agent de pontage dans la fraction à haut pourcentage de sucrose avant la formation du gradient. Mon premier choix c'est porté sur le L-glutaraldéhyde, agent de pontage employé en cryo-EM. Communément dans la littérature une concentration comprise entre 0.05% et 0.3% (V/V) est suffisante à la stabilisation de complexes. Des expériences de migration sur gel en conditions dénaturantes ont été établies afin de mettre en

évidence la plus faible concentration nécessaire à la stabilisation du complexe. Le but étant d'obtenir une concentration suffisante pour le pontage du complexe tout en évitant la formation de pontages intermoléculaire et l'agrégation du complexe. Quelque soit la concentration ajouté aucune bande correspondante aux sous-unités du complexe n'ont été observées. J'ai donc décidé de réaliser le gradient avec une concentration à 0.1% dans la solution à 35% de saccharose, me permettant d'obtenir entre 0.5-0.75% pour les fractions où le complexe est stabilisé. Après fractionnement du gradient 80 mM (concentration finale) de glycine est ajouté afin de stopper tout pontage supplémentaire.



**Figure 44 Analyse par cryo-EM de l'extrait ENI.** A) GraFix et dessalement. B) GraFix et 12 heures de dialyse. En jaune sont encadrées les particules de 17 nm en vert celles de 7 nm environ.

Comme on peut l'observer sur les grilles (Figure 44), la concentration semble toujours extrêmement faible pour une étude structurale par cryo-EM, néanmoins on observe plus de particules isolées entières que précédemment. Lors de l'étude de ces échantillons, j'ai pu observer une variation morphologique importante des particules mais également en termes de taille. Le diamètre des particules varie de 7 à 17 nm pour les plus grandes et de nombreuses petites particules qui pourraient correspondre aux sous-unités isolées sont observées. De plus, la qualité de la glace ne semble pas encore optimale, sur de nombreuses images un effet de flou peut-être noté. Il semblerait bien que l'ajout d'un agent de réticulation à faible concentration nous permette de stabiliser une partie des particules résultant en moins de dissociation lors de la congélation des grilles. Les particules de plus grande taille semblent présenter une conformation en croissant plus ou moins repliée (Figure 45) et on observe également des formes plus circulaires. Ces différences observées dans la morphologie de l'objet peuvent être différentes vues de l'objet en 3D ou la mise en évidence d'une flexibilité potentielle du complexe, pouvant aboutir à une grande hétérogénéité de conformations présentes sur les grilles et réduire la résolution lors de l'analyse des données. Il est donc important d'optimiser la quantité de particules sur les grilles et leur répartition.



**Figure 45 Grandissement de particules observées en figure 44.** Vue détaillée des particules indiquée 1 et 2, on observe une forme de croissant ouvert en 2 et fermé en 1.

La qualité de la glace reste également un problème et présente un niveau élevé du bruit de fond dû au tampon. Lors de l'analyse des différents composés qui pouvait entraîner ce bruit de fond je me suis rendu compte que lors du gradient de densité la concentration en détergent,

le DDM (dodecyl  $\beta$  d maltopyranoside) était bien supérieur aux limites conseillées par la cryo-EM. En effet, la concentration utilisée était de 0.05% (W/V) ce qui correspond à 10 fois la concentration micellaire critique (CMC=0.005%). La présence d'une concentration élevée en détergent impacte directement sur la répartition des particules et à tendance à repousser les complexe des trous vers le support carbone et favorise la présence de micelles. De plus la formation d'une interface air-eau au niveau de laquelle viennent s'orienter les molécules de détergents, cause un bruit de fond important, diminuant la possible caractérisation des particules. Le DDM, présente une CMC relativement faible, ce qui rend son élimination par dialyse difficile. Au vu de la concentration initiale en DDM des échantillons, il est possible qu'une partie des effets observés sur les grilles, comme le flou, en soit la cause.

#### D. Caractérisation par spectrométrie de masse

Dans le but de caractériser et de confirmer la présence de différentes sous-unités de NuRD, j'ai choisi d'effectuer des analyses par spectrométrie de masse, réalisées par le service de protéomique de l'IGBMC. Le protocole utilisé était une digestion en-gel avec de la trypsine et les résultats sont présentés figure 46.

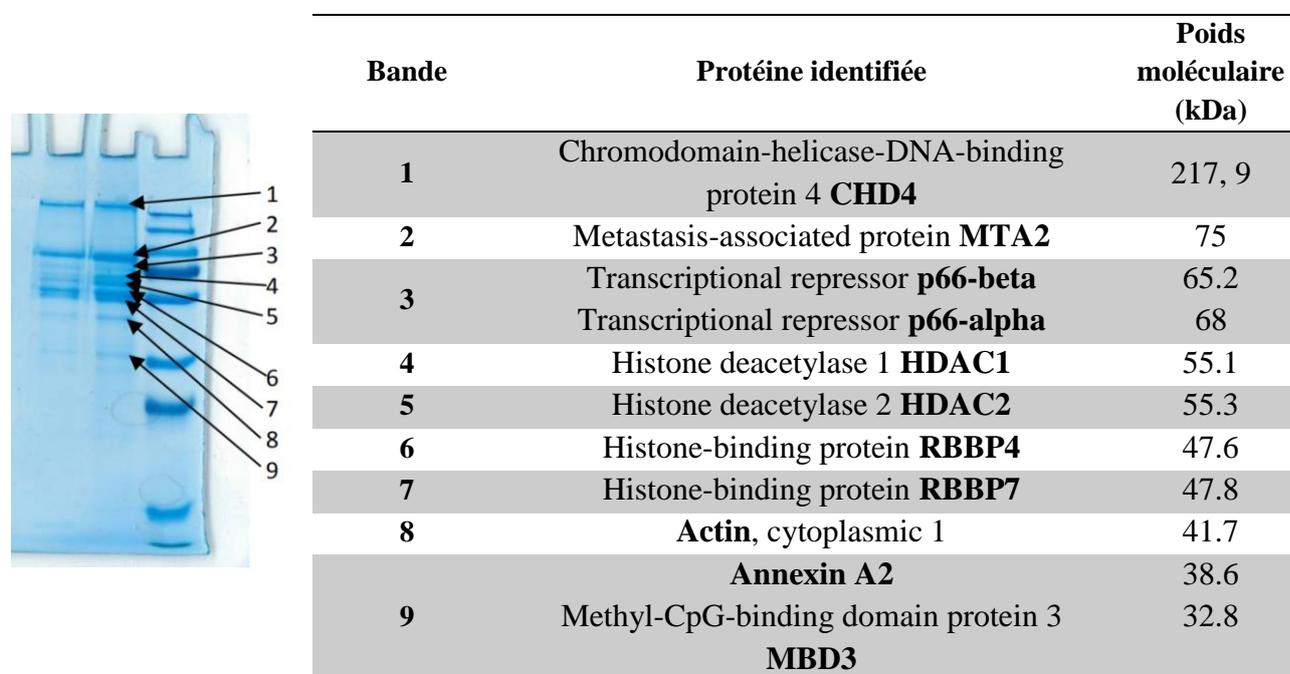


Figure 46 Analyse par spectrométrie de masse

Les résultats de spectrométrie de masse, confirment la présence de toutes les sous-unités majeures de NuRD, mais également la présence uniquement de la sous-unité CHD4 et non de son homologue CHD3. Cette information laissant penser que la présence de MTA2 est sélective pour la sous-unité CHD4. De la même façon, seul MBD3 a été caractérisé par spectrométrie de masse, en accord avec le caractère mutuellement exclusif de MBD2 et 3 au sein du complexe NuRD. Il semblerait qu'en sélectionnant le complexe MTA2-NuRD on sélectionne en réalité un complexe NuRD de type MTA2-CHD4-MBD3.

Deux protéines supplémentaires sont co-purifiées avec le complexe: l'actine et l'annexine 2.

La présence d'actine, co-immunoprécipitée avec le complexe NuRD n'est pas réellement surprenant en soit. Il est en effet bien connu que l'actine est un composant du cytosquelette régulateur les mouvements cellulaires, mais également impliqué dans les processus nucléaires. L'actine facilite la transcription en interagissant avec l'ARN polymérase et divers facteurs de transcription et joue un rôle de co-facteur pour de nombreux complexes remodeleurs de la chromatine tels INO80, SWI/SNF, NCoR, NuA4 etc... (Kapoor et al., 2013) (Szerlong et al., 2008) (Kapoor and Shen, 2014) Récemment, Serebryznnyy *et. al.*, ont mis en avant l'importance de l'actine dans la régulation de l'activité désacétylase des HDACs seules ou impliquées dans les complexes NuRD et CoREST (Serebryannyy et al., 2016). Ainsi, l'actine sous forme monomérique possède la capacité d'interagir avec HDAC 1 ou 2 et ainsi va inhiber la fonction désacétylase. De façon opposée, la polymérisation de l'actine dans le noyau, limite l'inhibition et entraîne donc une augmentation de l'activité des HDACs. Le niveau de polymérisation de l'actine régule donc le mécanisme de transcription en régulant la désacétylation des lysines des histones.

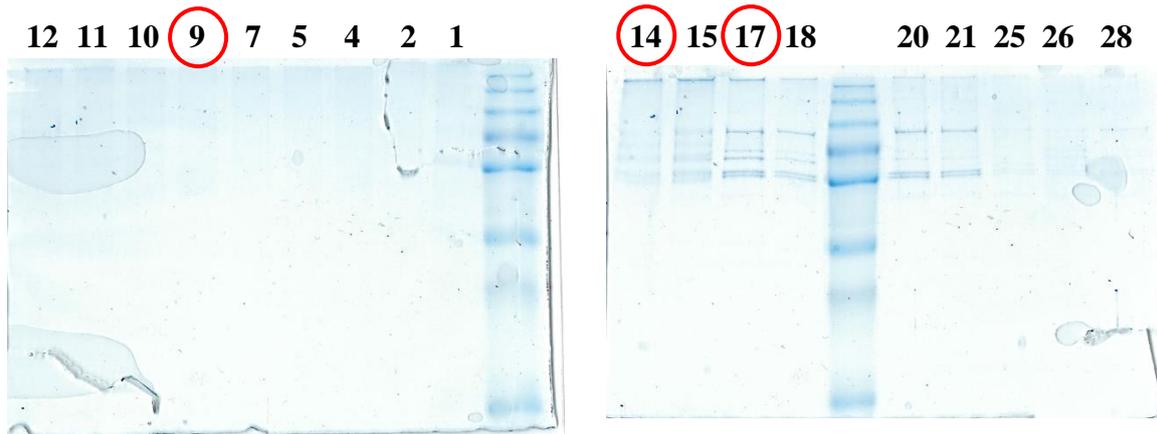
L'annexine 2 est quant à elle une protéine hautement conservée, surexprimée dans certains types de cellules cancéreuses. Elle est capable de se lier à l'actine en vue de réguler son remodelage au voisinage de membranes cellulaires dynamiques. La principale hypothèse serait que cette protéine soit co-purifiée avec l'actine.

Ces deux protéines sont co-purifiées avec le complexe NuRD et sont protéines associées au complexe.

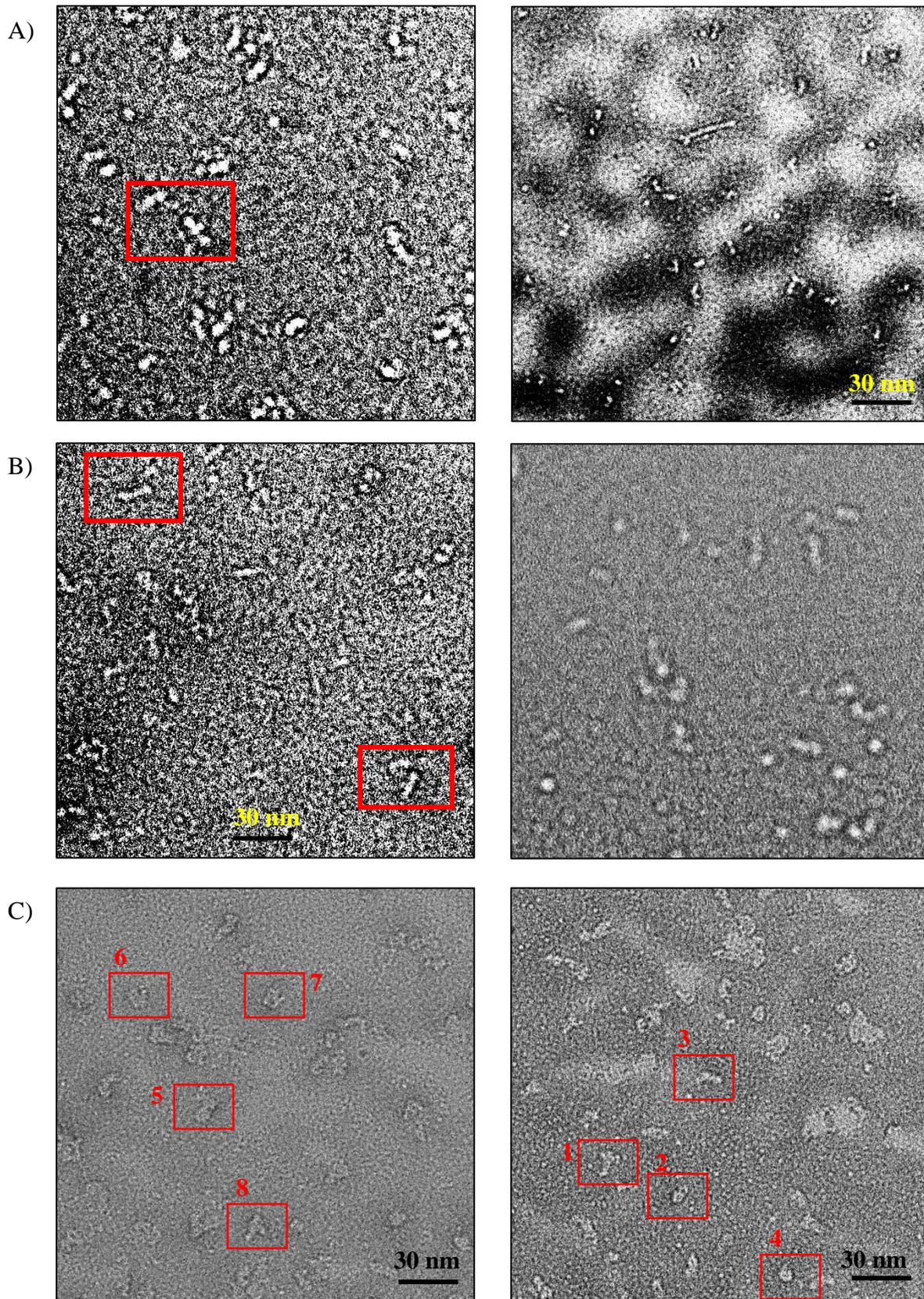
## E. Le gradient de type GraDeR

Lors des expériences de GraFix réalisées la concentration en DDM était d'environ dix fois la CMC. Communément, les détergents aident à la solubilisation des protéines et notamment par la formation de patches hydrophobes réduisant ainsi l'agrégation. La famille des détergents non-ioniques sont des détergents plus doux, moins susceptibles de dénaturer les protéines. La présence de détergents est donc fondamentale lors de la purification spécialement pour les complexes macromoléculaires ou les protéines interagissant avec l'ADN. Elle l'est beaucoup moins en cryo-EM, où la concentration nécessite d'être réduite au minimum et à une concentration bien inférieure à la CMC. Du fait de la difficulté d'éliminer le détergent par les méthodes dites classiques, j'ai décidé de m'inspirer de la technique de GraDeR initialement développée pour les protéines membranaires. L'intérêt de cette technique est la formation d'un gradient inverse de détergent. Ainsi la concentration de détergent est réduite le long du gradient et l'élimination se fait progressivement sans être trop violent pour le complexe. De manière à stabiliser le complexe au cours de cette élimination un gradient de L-glutaraldéhyde a été ajouté. L'élimination de détergent est donc normalement compensée par la réaction de fixation qui stabilise le complexe. La purification suit le même protocole que précédemment à la différence des solutions pour le gradient. 0.01% de DDM est ajouté au tampon à 15% de saccharose et 0.1% de L-glutaraldéhyde est ajouté à celui à 35% de saccharose. En se basant sur les fractions où se stabilise le complexe, on devrait avoir dans notre milieu environ 0.005-0.006% de DDM et 0.05% de L-glutaraldéhyde. Lors de la réalisation des grilles, on a pu observer une glace plus épaisse empêchant toute analyse. La modification de concentration en DDM a eu un impact sur l'absorption de l'échantillon par le vitrobot. Les paramètres du vitrobot ont dû être modifiés et la force de pression a été augmentée à 7.

Les échantillons ont d'abord été vérifiés par coloration négative selon la méthode 2 présentée en figure 35 B. En parallèle du gradient de type GraDeR-GraFix, j'ai réalisé un gradient avec peu de matériel pour sélectionner les fractions d'intérêt. Bien que communément le complexe soit stabilisé aux environs de 25% de sucrose, le gradient supplémentaire me permet de sélectionner de façon plus précise les fractions (Figure 47).



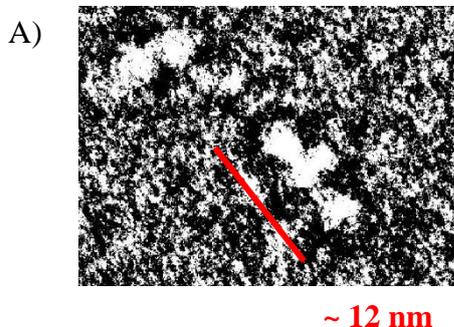
**Figure 47 SDS gel des fractions du gradient.** Le gradient s'étend de la fraction 1 correspondant à 35% de saccharose à 28 à 15% de saccharose. Les fractions encerclées de rouge sont les fractions étudiées.



**Figure 48 Image en coloration négative.** A) Coloration négative de la fraction 9 (33% de saccharose). B) Coloration négative de la fraction 14 (27% de saccharose). C) Coloration négative de la fraction 16 (25% de saccharose). En rouge sont encadrées les particules pour lesquels un agrandissement est présenté dans les figures suivantes.

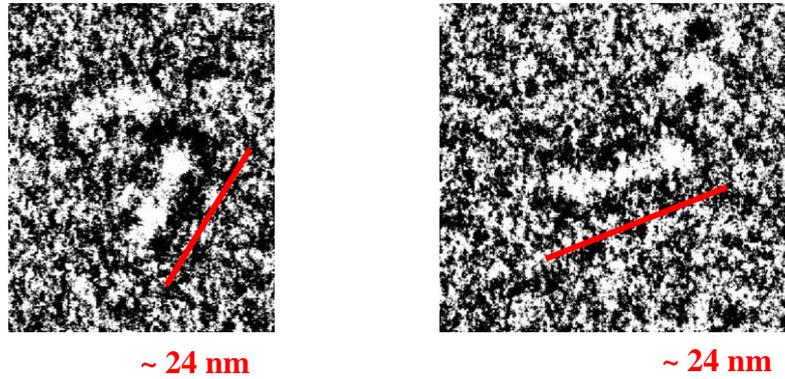
Lors de l'étude par coloration négative, la fraction 9 est également étudiée. En effet, la mesure de l'absorbance des différents échantillons à mise en évidence la présence d'une importante absorbance à 280 nm en absence de bande de migration sur le gel pour la fraction.

Les images réalisées sur le Tecnai TF20 sont présentées figure 48. On observe la présence pour la fraction 9 de particules allongées et d'autres de faibles tailles (Figure 49). De façon surprenante aucune information sur gel SDS ne corrobore la présence de protéines. L'hypothèse la plus probable serait qu'il s'agisse soit de contaminants soit du complexe NuRD ayant subi une réticulation trop importante. Les composés de cette fraction sont stabilisés à 33% de saccharose et donc doivent posséder une densité importante.

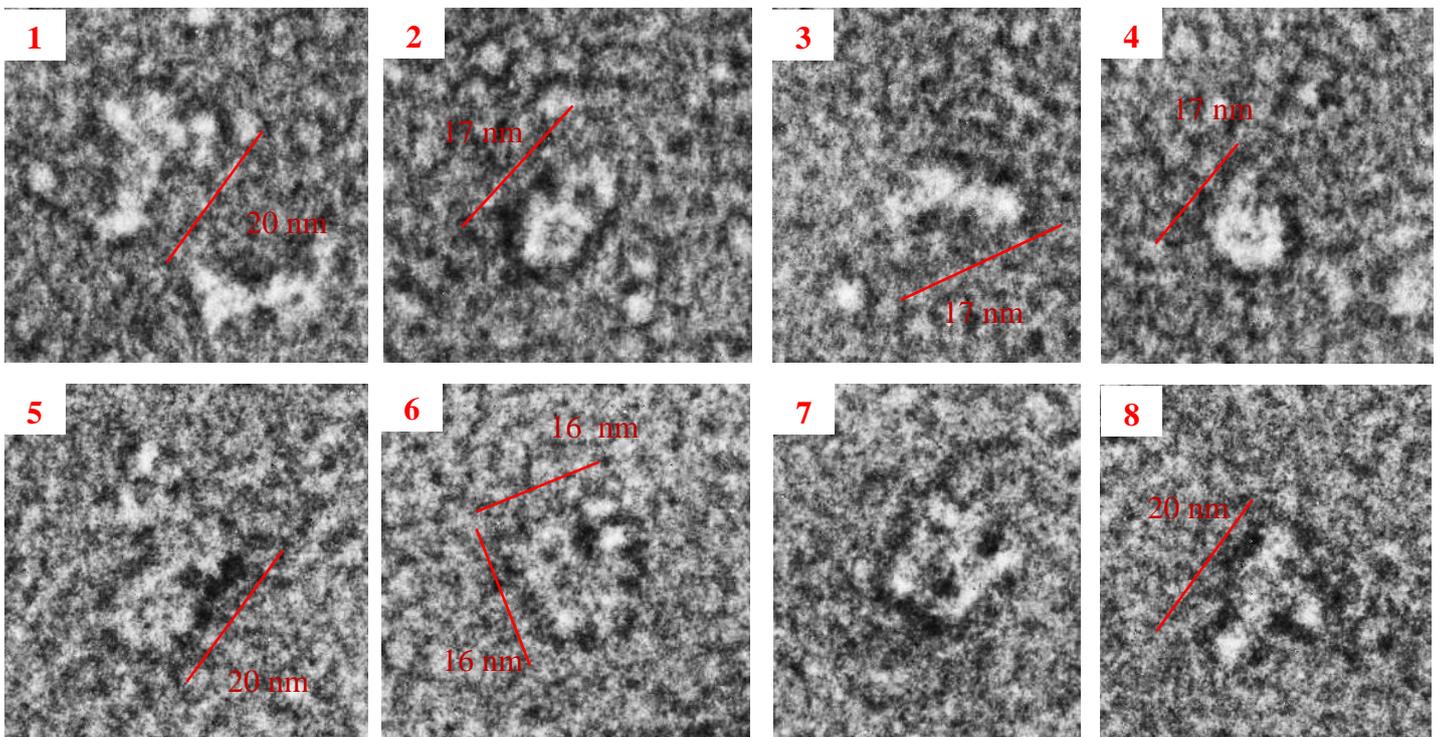


**Figure 49 Grandissement de particule pour ENI-09.** Particule de petite taille d'environ 12 nm. On peut discerner une organisation en domaines. Détail de la figure 48-A.

Les fractions 14 et 16, correspondent sur le gel SDS-PAGE aux fractions enrichies en complexe NuRD à une densité de saccharose de 25%. La fraction 14 semble se caractériser par des particules allongées d'environ 17-25 nm (Figure 50). Ce n'est pas la fraction la plus enrichie et elle présente peut-être un complexe NuRD de stœchiométrie différente du pic de la fraction 17. En effet lors de l'analyse de la fraction 17, on observe clairement la présence de complexes d'environ 20 nm (Figure 51). Cette observation de complexe nous permet de visualiser une organisation en domaines bien visible ici également dans différentes orientations. Néanmoins, on note une certaine hétérogénéité de forme (Figure 48 C). L'instabilité du complexe est la cause également de conformations différentes dans notre échantillon. Enfin, la dernière observation concerne la dissociation du complexe, on peut observer des complexes qui semblent se dissocier même en présence d'agents de réticulation. Ces observations peuvent être dû à l'instabilité du complexe mais également être des artefacts de la coloration négative.



**Figure 50 Grandissement de particule pour ENI-14.** (A) et (B) particules allongées d'environ 24 nm identifiées sur la grille.

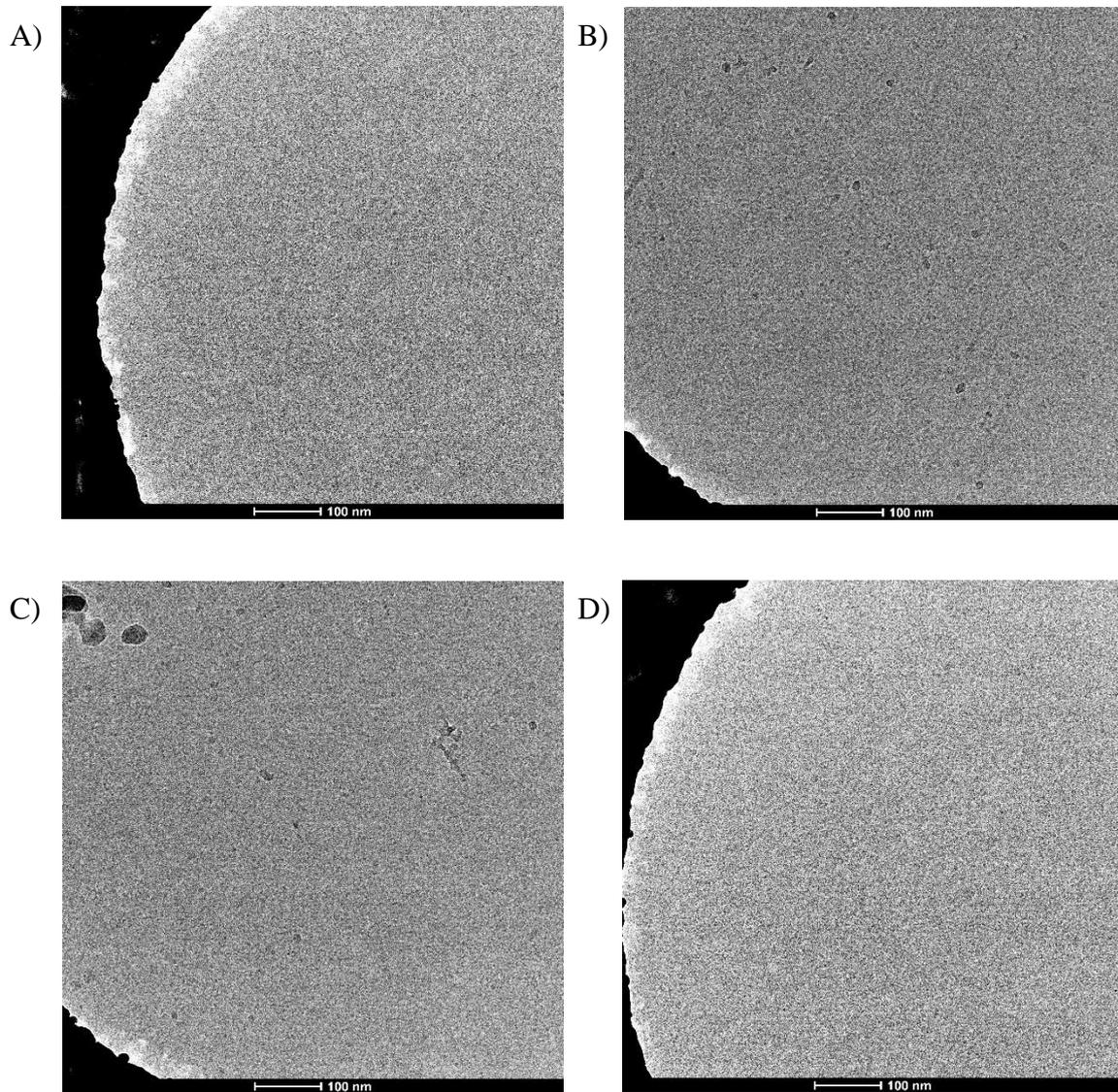


**Figure 51 Grandissement de particule pour ENI-16.** On observe une variété de forme pour ces particules. Les particules 2, 3, 4 et 5 semblent similaires mais dans des orientations différentes. Les particules 6 et 7 sont identiques. Les tailles des différentes particules varient entre 15 et 20 nm.

L'échantillon lors de cette analyse ne fut que peu dilué, la concentration observée est donc caractéristique de la fraction post-gradient. Pour la préparation des grilles de cryo-EM, la concentration devrait être de 5 fois au moins celle observée en coloration négative. Pour remédier à ce problème, après élimination du saccharose par dialyse ou dessalement,

l'échantillon a été appliqué deux fois sur les grilles afin de le reconcentrer sur grille. Pour cela 3  $\mu\text{L}$  de la fraction 16 ont été déposés sur la grille dans le vitrobot, et l'excès de liquide est absorbé latéralement et manuellement à l'aide d'un papier filtre ou re-pipeté directement sur la grille. Suite à cela, 3  $\mu\text{L}$  sont ajoutés rapidement et la grille est préparée normalement avec le vitrobot. Bien que le complexe soit visible en coloration négative, lors du passage en condition cryogénique, quasiment aucune particule du complexe ne peut être identifiée sur les images réalisées que ce soit dans de la glace fine ou de la glace épaisse (Figure 52). De plus, l'association d'un gradient de type GraFix au gradient GraDeR n'entraîne aucune modification et n'améliore pas la visibilité du complexe. Une importante dissociation est néanmoins observée même en présence d'agent de pontage, la concentration utilisée n'est peut-être donc pas suffisante.

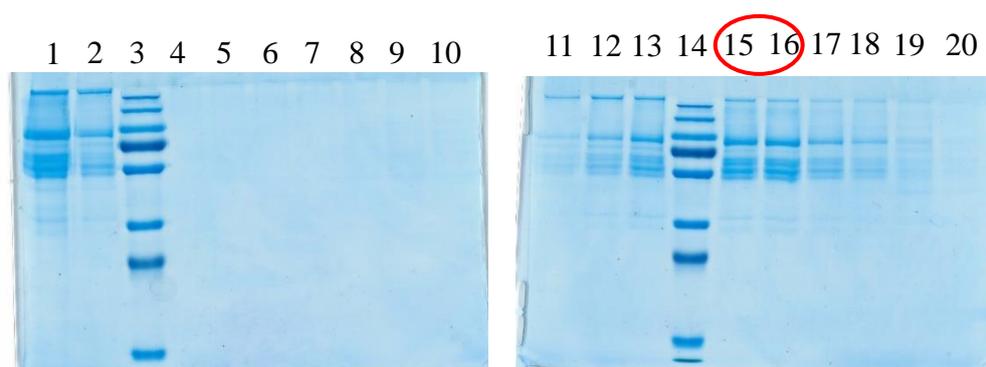
La concentration utilisée étant supérieure, on devrait pouvoir observer une amélioration dans la distribution des particules et en identifier plus dans les trous des grilles. Les expériences présentées ici, m'ont permis d'identifier que les modifications de type gradient, de quantité de complexes, la réticulation, n'avaient pas un impact important sur la distribution des particules. J'ai donc décidé d'utiliser de nouvelles grilles présentant une fine couche de carbone continue. Cette méthode permet d'adsorber le complexe protéique sur le film de carbone. Ainsi, si l'absence de complexe au sein des trous des grilles de type R2/2 vient d'une affinité préférentielle du complexe pour le carbone, l'utilisation d'une couche continue de carbone devrait permettre de répartir uniformément le complexe sur la grille. En effet, les complexes macromoléculaires qui ont la capacité d'interagir avec l'ADN présentent une plus grande affinité pour le support et seront exclus des trous. Bien qu'extrêmement répandue, cette technique présente certaines limites ; ainsi l'utilisation de carbone va induire un bruit de fond limitant la visibilité du complexe, ce qui peut être compensé ici par la taille importante de NuRD. Mais elle peut aussi être la cause d'orientations préférentielles de nos particules.



**Figure 52 Influence du gradient de type GraDeR** A) et B) représentent la purification du complexe NuRD avec l'utilisation d'un gradient de type GraDeR suivi d'une étape d'élimination du saccharose par dessalement en A) et dialyse en B). C) et D) représentent le même type de traitement de fraction mais lors de la purification un gradient de type GraFix associé à un GraDeR est réalisé.

## F. Les grilles de carbone continu

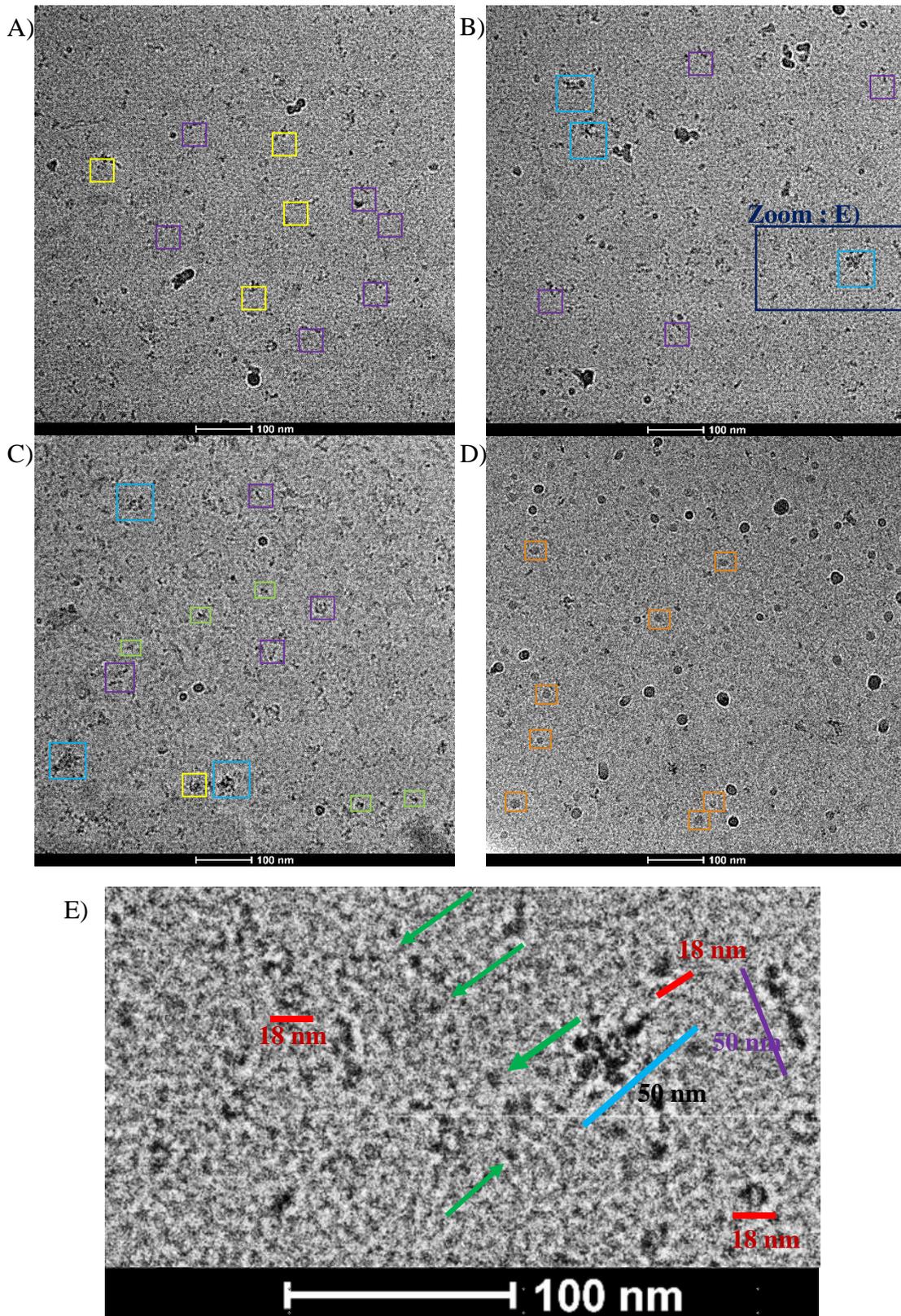
Les grilles de carbone continu ont été réalisées à partir de grilles commerciales R2/2 avec la méthode du carbone flotté. Lors du dépôt de l'échantillon, comme pour les précédentes expériences l'échantillon a été déposé en deux fois (3  $\mu$ L à chaque fois) avec un temps d'incubation d'environ 5-10 minutes lors du dépôt de la première goutte pour laisser le temps au complexe de se déposer sur la grille par gravité. Néanmoins, il est important de ne pas laisser l'échantillon trop longtemps lors de cette incubation. L'incubation de gouttes de 3  $\mu$ L va augmenter le contact à l'interface air-tampon et aboutir à la dissociation du complexe. Au vu de l'importante dissociation du complexe, j'ai décidé d'augmenter la concentration d'agent de réticulation (0.15% L-glutaraldéhyde dans la partie à 35% de saccharose) et de détergent pour le GraDeR (0.03% de DDM dans la partie à 15% de saccharose). Pour ne pas endommager la couche de carbone, la décharge lumineuse a été réduite à 15 secondes.



**Figure 53 Analyse par électrophorèse SDS-PAGE de NuRD.** 1 et 2 : les fractions d'éluion des fractions insolubles et soluble. 4 à 20 les fractions post-gradient de saccharose de la fraction insoluble.

Différentes grilles ont pu être réalisées avec les fractions 15 et 16 observées sur le gel SDS comme étant caractérisées par l'ensemble des sous-unités, puis observé sur le Polara à une tension d'accélération de 300 kV. Pour cette analyse, un défocus de -5  $\mu$ m et un grandissement de 59000x ont été utilisés avec la caméra Falcon I.

L'utilisation des grilles en carbone continu, nous permettent de repérer plus de particules que précédemment.



**Figure 54** micrographes de la fraction insoluble ENI. A), B) et C) On peut observer sur ces trois micrographes la présence de particules encadrées en jaune, en violet sont encadrés les particules de formes allongés pouvant être le complexe débobiné et en orange les particules circulaires. Enfin, en bleu sont encadrés les agrégats. E) En vert sont pointé les sous-unités dissociées avec un diamètre

inférieur à 10 nm. La ligne bleue sert de barre d'échelle pour l'agrégat et en mauve pour la particule allongée dissociée. En rouge sont caractérisées les particules d'environ 20 nm.

On observe principalement des particules longilignes qui attestent de l'instabilité et de la flexibilité importante du complexe (Figure 54). De plus, on constate, que le complexe s'agrège encore énormément sur les grilles. Pour résoudre ce problème d'agrégation de nombreuses modifications de tampon, dans le but de changer la force ionique et ainsi limiter ce problème ont pu être testés. Le tableau 5 illustre les différentes conditions testées. Malgré tous ces efforts, aucune réelle amélioration n'a été observée.

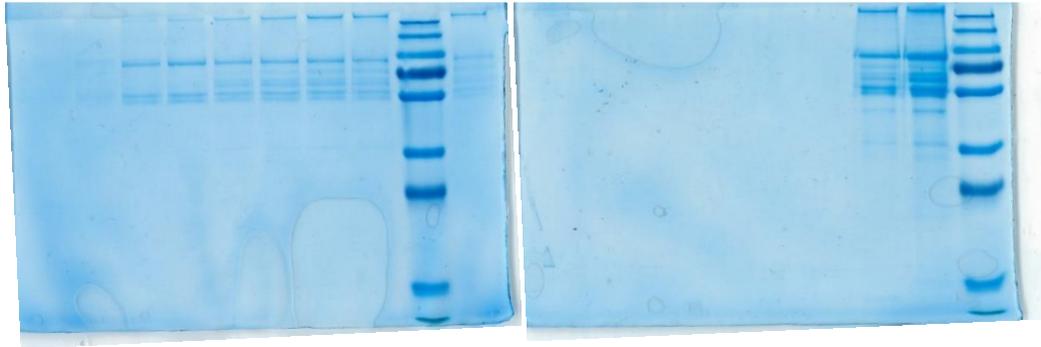
Tampon et pH	Sel	MgCl <sub>2</sub>	EDTA	Agent réducteur	Détergent	Glycérol
HEPES pH 7,65	150 mM NaCl	3 mM	0.1 mM	DTT 1 mM	DDM 0.005 à 0.0004%	0 à 1%
HEPES pH 8,0	125 mM NaCl		0 mM	TCEP 0.5 mM		
HEPES pH 8,5	100 mM KCl					
CHES pH 9,0	80 mM KAcO					
	100 mM KAcO					

**Tableau 5 Résumé des conditions testées sur grilles en carbone continu.**

L'intérêt majeur dans ces changements de sels ou de pH est d'essayer de réduire l'agrégation est également de stabiliser le complexe. Les protéines sont théoriquement moins solubles lorsque leur pH est égal au pI ; dès lors changer le pH va changer la charge nette des protéines et la répartition des charges en surface. En plus d'avoir une influence sur l'agrégation, le changement de charge devrait influencer directement la répartition des particules sur la grille. Aucun effet direct lors des changements de pH ne fut néanmoins observé.

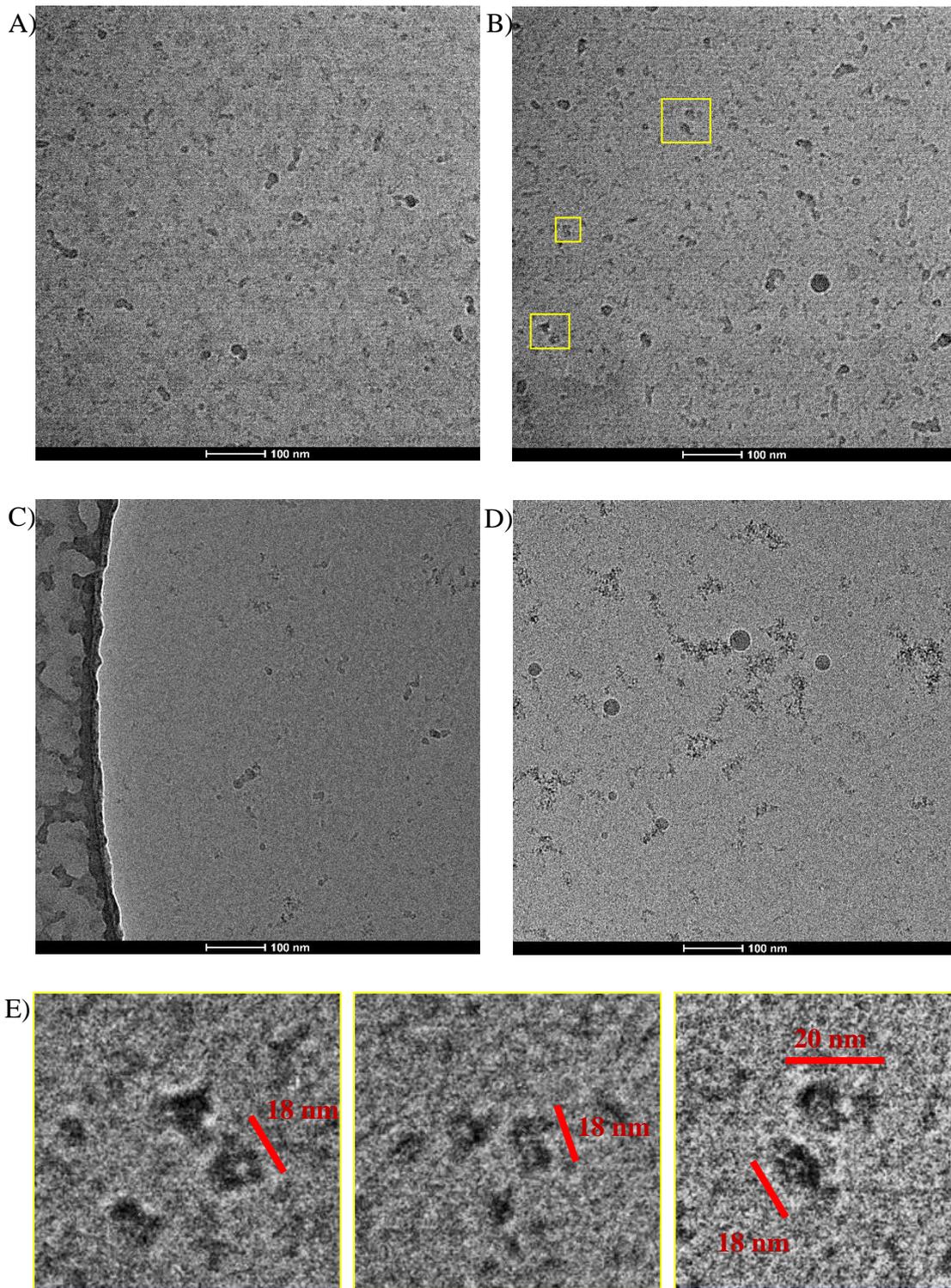
Les changements de sels et de concentration de sel vont avoir un effet quant à eux sur la force ionique. En modifiant la force ionique, on modifie les interactions électrostatiques au sein des protéines et entre les protéines. Ces modifications peuvent stabiliser ou déstabiliser nos échantillons et donc prévenir ou être la cause d'agrégations. Le changement de type de sel peut affecter également la stabilité de l'échantillon, par l'incorporation d'ions qui vont pouvoir interagir avec les protéines. De la même façon qu'avec les variations de pH, aucune véritable amélioration ne fut observée.

Néanmoins, toutes ces modifications ont été réalisées par la technique de dessalement après le gradient saccharose réalisé avec un tampon de type HEPES pH 8,0 et 150 mM NaCl, 3 mM MgCl<sub>2</sub>, 0,1 mM EDTA, 0,5 mM TCEP, 1% glycérol et 0.005% DDM. Il serait intéressant de réaliser ces changements lors de la purification, durant l'étape du gradient de densité.

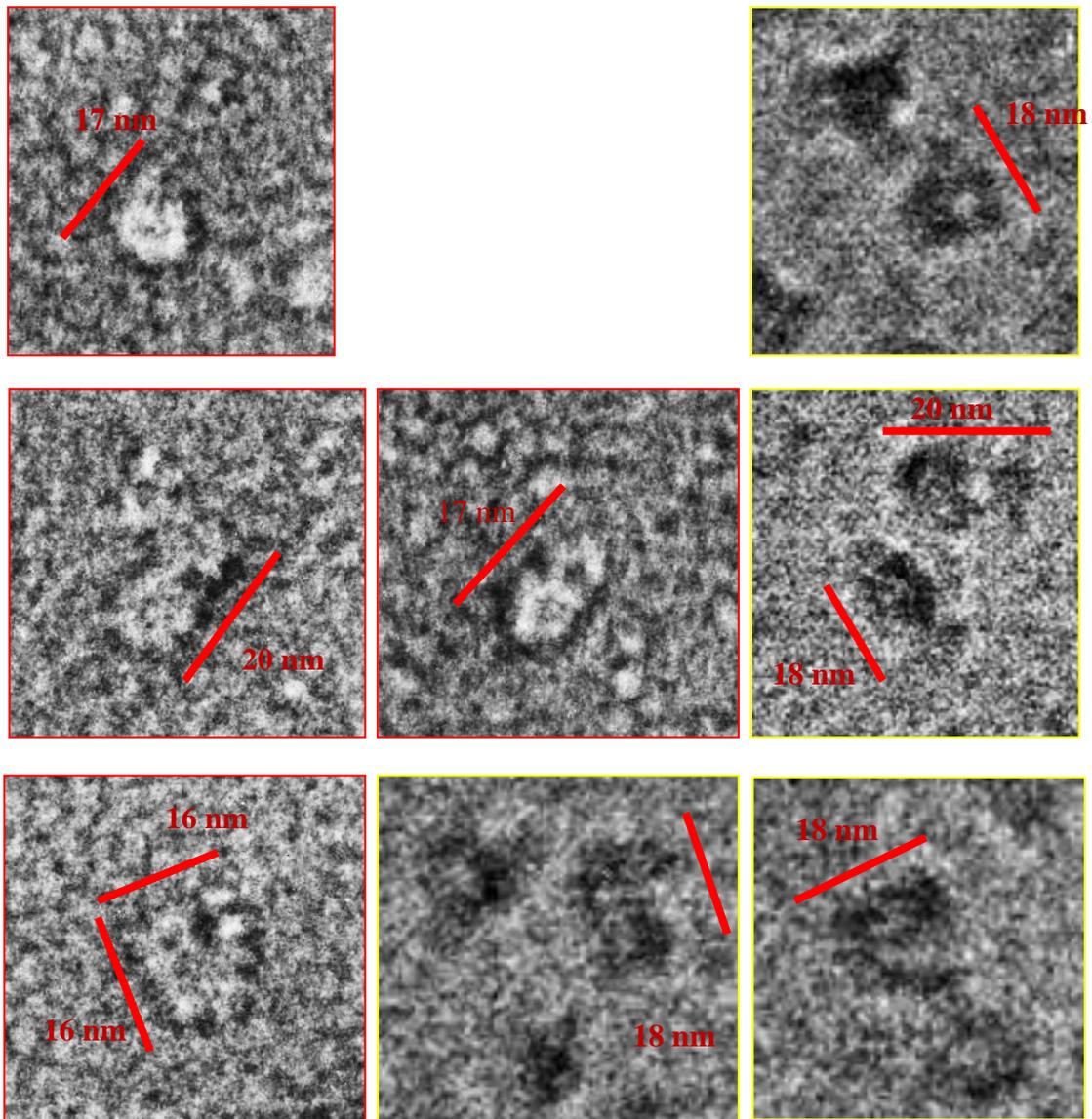


**Figure 55 Analyse par électrophorèse SDS-PAGE de NuRD.** Fraction soluble

Pour la fraction soluble (Figure 56), on observe la présence de complexe en A) et B). De même que pour la fraction insoluble, les particules semblent allongées. La réticulation avec le glutaraldéhyde ne semble pas optimale ; en effet, lors de l'observation de fractions en amont de celles étudiées on observe une importante agrégation des particules. Le composé de réticulation peut-être trop concentré ou la méthode de GraFix entraîne une réticulation trop longue qui pontent toutes les sous-unités de façon intermoléculaire. Les expériences précédentes avec cette fraction ont montrés que la diminution de cet agent entraînait la dissociation du complexe lors de la congélation. J'ai donc décidé ici de tester un autre agent de réticulation plus spécifique que le glutaraldéhyde, le BS3. Le bras liant les deux fonctions réticulantes de ce composé mesure environ 11 Å, bien plus long que la glutaraldéhyde. De ce fait, le BS3 est capable de ponter des composés plus éloignés au sein du complexe et peut-être stabiliser l'intégralité de NuRD. Il est nécessaire de contrôler très finement la concentration de ce composé qui pourra ponter différents complexes à plus grande distance. Le BS3 est de plus non adapté la méthode de GraFiX, tous les tests réalisés avec cet agent ont été réalisés à la suite du gradient de saccharose.



**Figure 56 micrographes de la fraction insoluble ENS.** A), B) et C) On peut observer sur ces trois micrographes la présence de particules encadrées en jaune. En rouge sont encadrés les agrégats. Ces fractions ont été traitées par dessalement pour éliminer le saccharose. D) fraction 9, cette fractions en amont des fractions ou s'équilibre le complexe, est composée de complexe agrégés de façon irréversible. E) grandissement des particules observées. Les formes des particules observées ici sont très similaires à celles observées en coloration négative en **Figure 51** et pourraient représenter différentes orientation du complexe NuRD.



**Figure 57** Comparaison des formes observées en coloration négative (images encadrées en rouge) et celles observées en cryo-EM. On retrouve dans les deux échantillons des particules similaires en formes et en tailles.

Comme on peut l'observer en **Figure 57**, bien que la réticulation fut trop importante on observe des particules de taille et formes similaires que l'on a pu identifier également précédemment au cours des différents tests réalisés. Ces observations sont en accord avec celles déjà observée en coloration négative et mettent en avant une organisation en domaines du complexe NuRD.

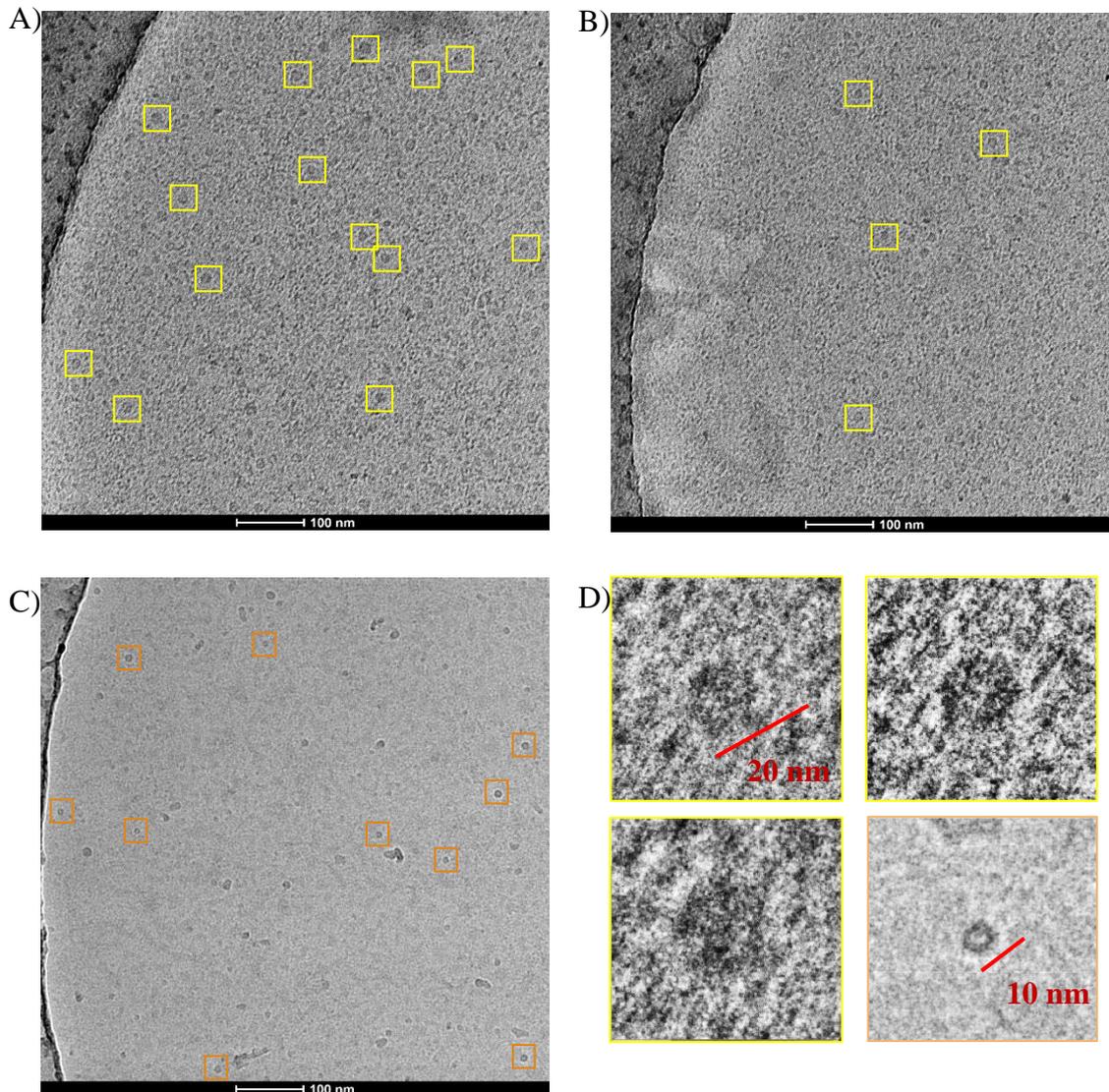
Une acquisition sur le Titan Krios par les Dr. Otilie Loeffelholz Von Colberg et Dr. Brice Beinsteiner a été quand même envisagée, mais la grille a été perdue lors de son chargement dans le microscope.

## G. Les agents de réticulation

Comme je l'ai expliqué précédemment, l'utilisation du BS3 pourrait apporter des modifications pour la stabilisation du complexe, en créant des pontages entre les sous-unités plus éloignées au sein de NuRD. Différents protocoles d'utilisation sont publiés pour l'utilisation de cet agent. Du fait de la sensibilité du complexe, j'ai décidé d'éviter toute étape à plus de 25°C. Lors des tests de réticulation l'échantillon fut placé en présence de concentration croissante de BS3 (0.1 à 3 mM) pendant 30 minutes à température ambiante ou 1 heure sur glace. Le processus de réticulation a été stoppé par l'ajout de 50 mM de carbonate d'ammonium pendant 30 minutes toujours à température ambiante. Comme pour les tests en présence de L-glutaraldéhyde aucune bande n'est pas détectée même à basse concentration en agent réticulant. J'ai donc décidé d'utiliser la concentration la plus utilisée pour la réticulation de complexe qui est de 0.2 mM. Les conditions de purification et les paramètres du vitrobot ont été conservés. Toutefois du fait de l'importante dissociation observée précédemment, le temps d'incubation sur les grilles a été réduit à 5 minutes et 1% de glycérol a été ajouté au tampon final pour stabiliser le complexe.

Lors de l'analyse de ces grilles, un nombre important de particules de forme circulaires a pu être identifié. La forme de ces particules diffère de celle de forme plus allongées observées jusqu'à présent. La concentration en détergent est d'environ 0.007%, en-dessous de la CMC mais pas de façon significative, il se pourrait que ce soit des artefacts dû à la formation de micelles. La taille de ces particules se situe entre 19 et 22 nm. La présence de ces particules est principalement identifiée pour la fraction soluble du complexe (Figure 58 A et B). Pour la fraction insoluble, on peut noter toujours une dissociation du complexe et la présence importante de micelles (Figure 58 C). Les micelles présentes dans la fraction insoluble et les particules de la fraction soluble sont relativement différentes en termes de taille et de forme. En effet, la taille des micelles est d'environ 7-10 nm de diamètre contre environ 20 nm pour les particules encadrées de jaune. De plus, on observe pour les micelles une partie centrale non dense aux électrons caractéristique des micelles, à l'inverse les particules présentes une variation de densité même dans leurs parties centrales. Deux hypothèses sont alors émises en ce qui concerne ces particules ; qui peuvent être le complexe NuRD stabilisé sous une forme arrondie par le BS3 ou il s'agit de micelles qui auraient enfermées des protéines par stabilisation des régions hydrophobes des protéines. Pour vérifier ces hypothèses un jeu de

donnés a été collecté sur le Titan Krios, équipé de la Falcon III et du correcteur Cs et en présence du Volta Phase Plate à 300 kV par le Dr. Otilie Loeffelholz Von Colberg.

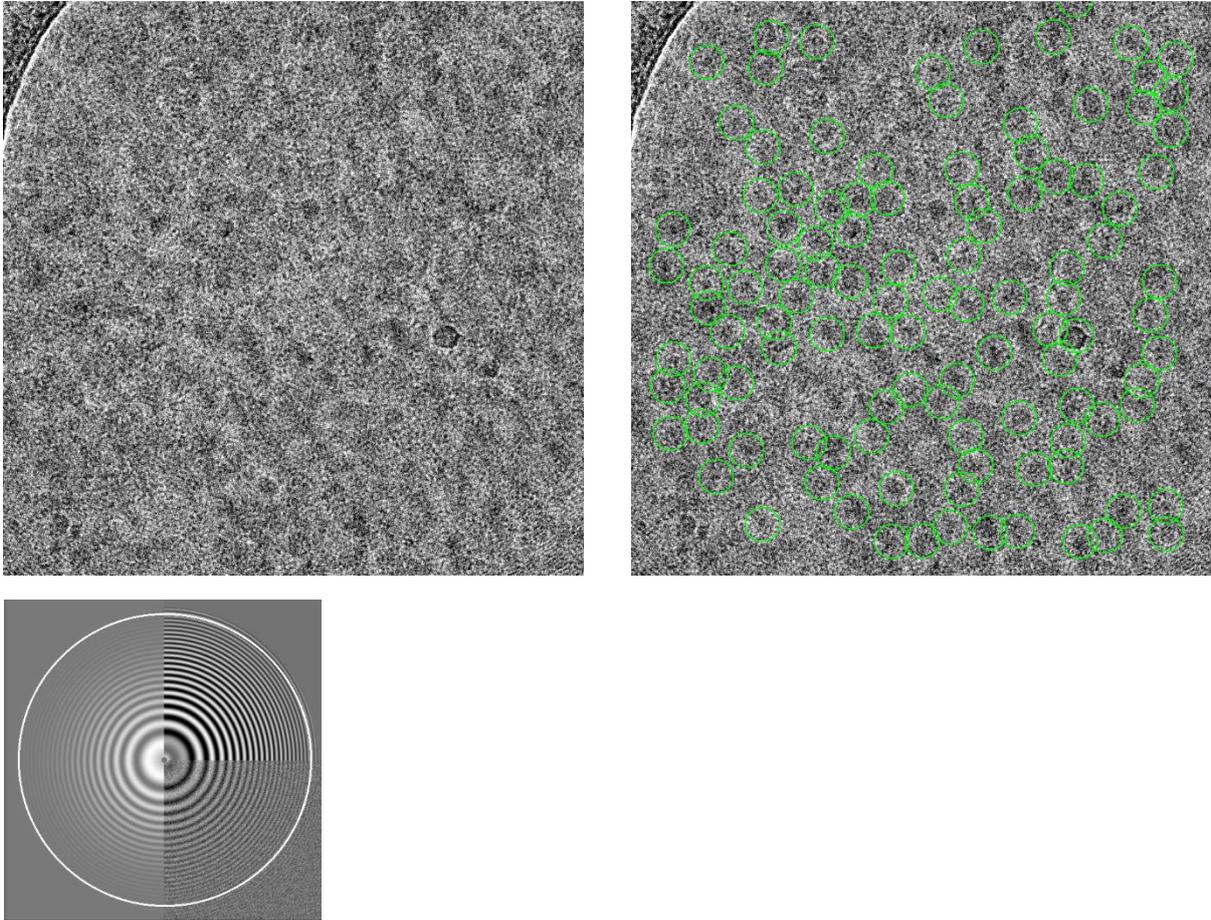


**Figure 58 BS3 et grilles en carbone continu** A et B) fraction soluble avec certaines particules encadrées en jaune. C) fraction insoluble, en orange sont encadrés les micelles présents sur la grille. Enfin en D) on observe un grandissement de particules en A) et d'une micelle provenant de l'image C) on observe une variation de taille mais également de forme. La micelle présente en effet une cavité non retrouvée pour les trois autres particules.

#### H. Traitement de données

Lors de la première acquisition 2454 micrographes ont été enregistré grâce au logiciel SerialEM. Un grandissement de 78 000 fois, un temps d'exposition de 66 secondes pour une dose totale de  $80\text{e}^-/\text{A}^2$  répartie entre 78 frames.

Dans un premier temps différents tests pour sélectionner les particules de façon automatique ont été réalisées avec le logiciel Relion 1.4.

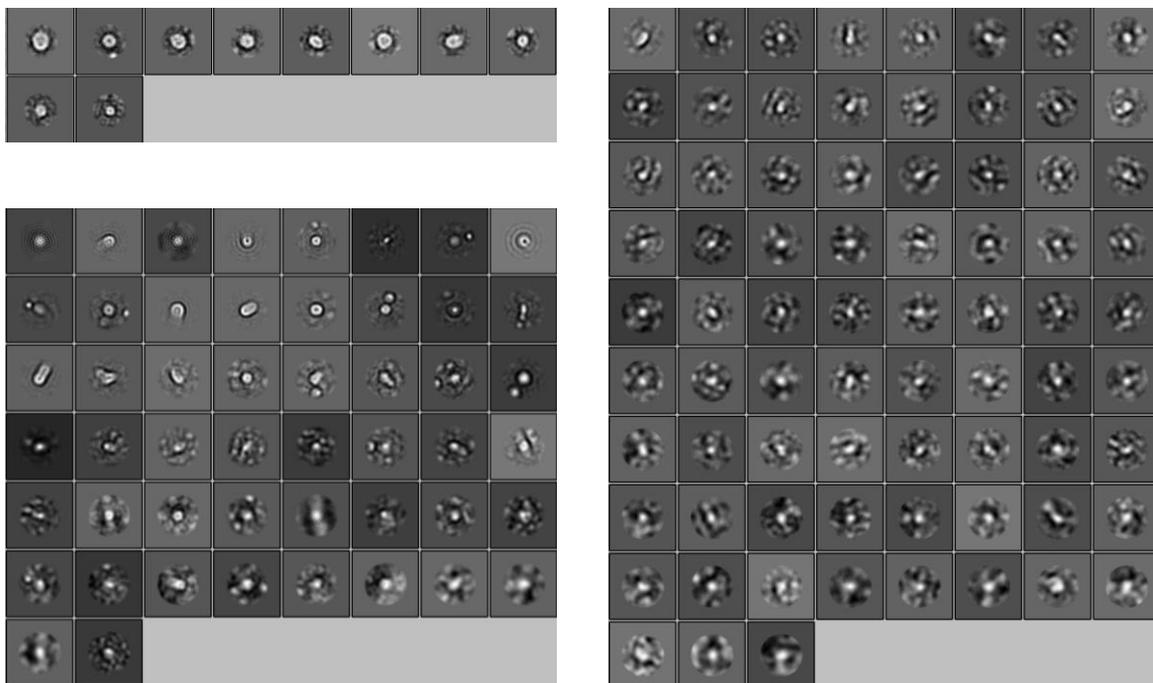


**Figure 59 Micrographe de la fraction soluble.** Image réalisée sur le microscope Titan Krios avec une caméra Falcon III et une Volta phase plate, défocalisation  $-0.5 \mu\text{m}$ , une tension d'accélération de 300 kV. A) image avant la sélection automatique de particules B) Particules sélectionnées automatiquement en vert. C) correction de la CTF par Gctf.

Comme on peut observer sur la figure 59, l'utilisation du Volta phase plate et d'une caméra plus puissante que sur le Polara ont permis d'obtenir une augmentation de contraste même pour le carbone continu. La plupart des micrographes sont caractérisés par une très faible présence de complexes et un contraste important du carbone rendant difficile l'interprétation et la sélection des particules. La sélection automatique avec le logiciel Relion 1.4, fut testé en premier lieu, les différents paramètres ont été modifiés pour tenter d'ajuster la sélection automatique. Néanmoins, aucune réelle amélioration ne fut notée, et la majorité du temps seul le carbone continu était sélectionné.

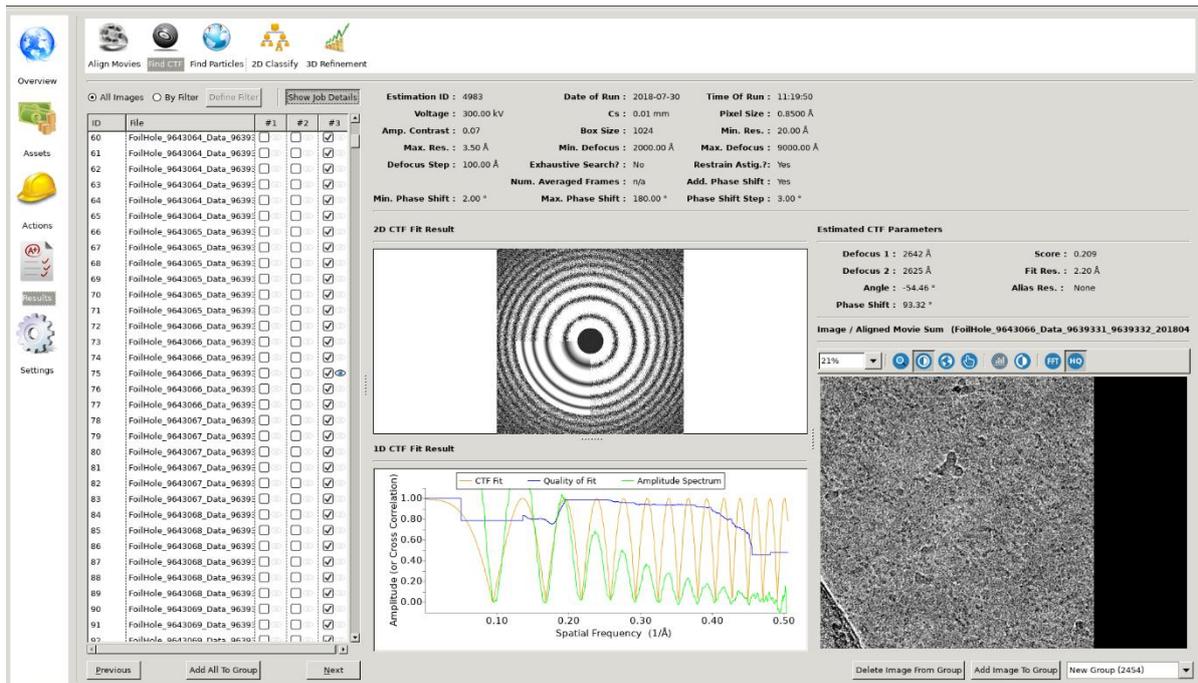
Afin d'obtenir des informations préliminaires sur le jeu de données, j'ai donc sélectionné manuellement 2500 particules. La sélection manuelle des particules assure un meilleur

centrage des particules comparé aux méthodes automatiques de références. En effet, du fait du contraste du carbone continu, de la glace relativement épaisse et de la présence de contaminant, les logiciels sont limités pour la sélection fine de particules. Après extraction de ces particules, les particules ont été classées par le logiciel Relion en différentes classes bidimensionnelles. L'augmentation du nombre de classes 2D nous permet de discerner quelques micelles et des particules protéiques. Au vu de ces résultats, j'ai décidé de sélectionner de façon toujours manuelle un plus grand nombre de particules. Les classes 2D de ces 5973 particules sont présentées en figure 60 C. La sélection de particules supplémentaires semble orienter la classification sur des particules qui ne ressemble pas à des micelles. De plus pour ces classes 2D, on n'observe pas de forme caractéristique, mais plutôt de points, dus à un problème de sélection et de centrage des particules.

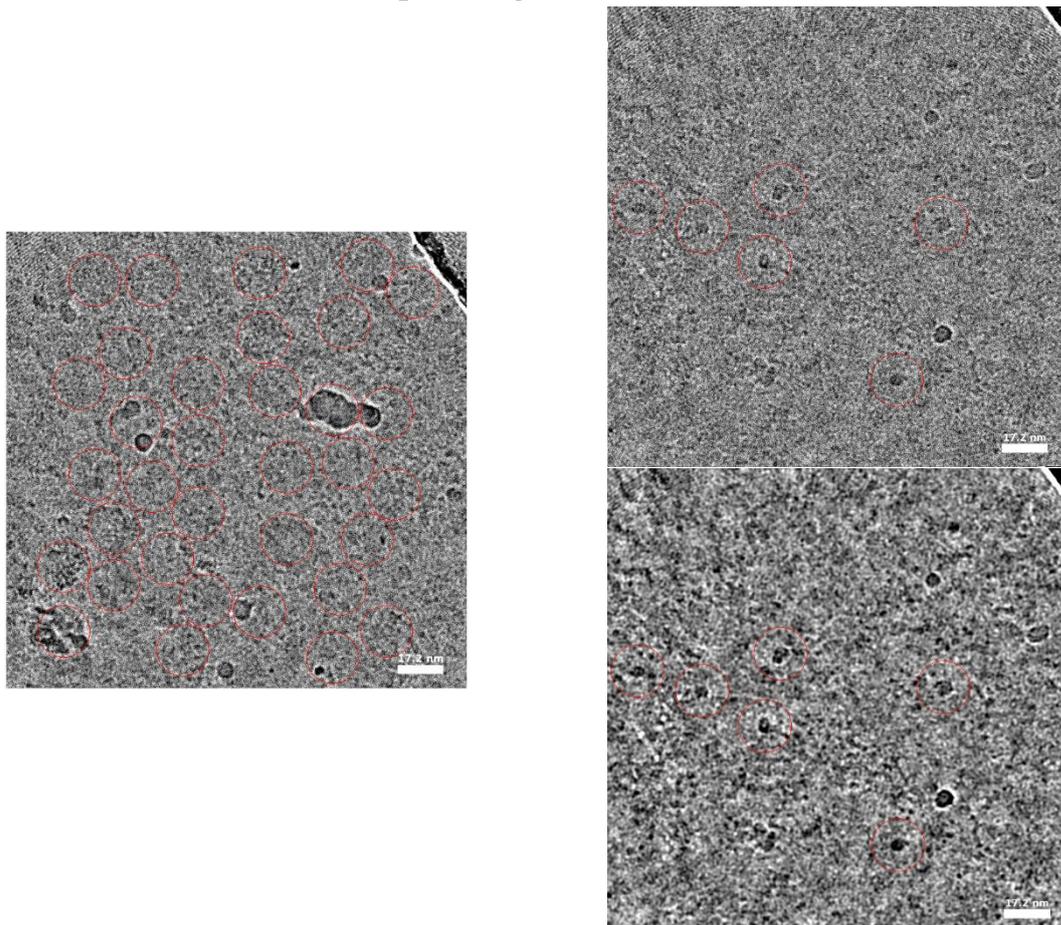


**Figure 60 Classes 2D.** A) 2500 particules sélectionnées manuellement réparties en 10 classes, on observe ici des particules ressemblant à des micelles. B) même particules qu'en A mais réparties en 50 classes. C) 5973 particules sélectionnées et réparties en 75 classes 2D.

Les classes ne semblent pas de bonnes qualité. La question est de savoir si cela est dû aux données elles-mêmes ou au logiciel. J'ai donc décidé de traiter le jeu de données avec le logiciel cisTEM. L'avantage ici, est l'aspect « user-friendly » du logiciel, notamment lors de la correction de la CTF qui nous permet de pouvoir sélectionner les micrographes de bonnes qualités plus rapidement en fonction de la présence de glace, de mouvement etc... (Figure 61)



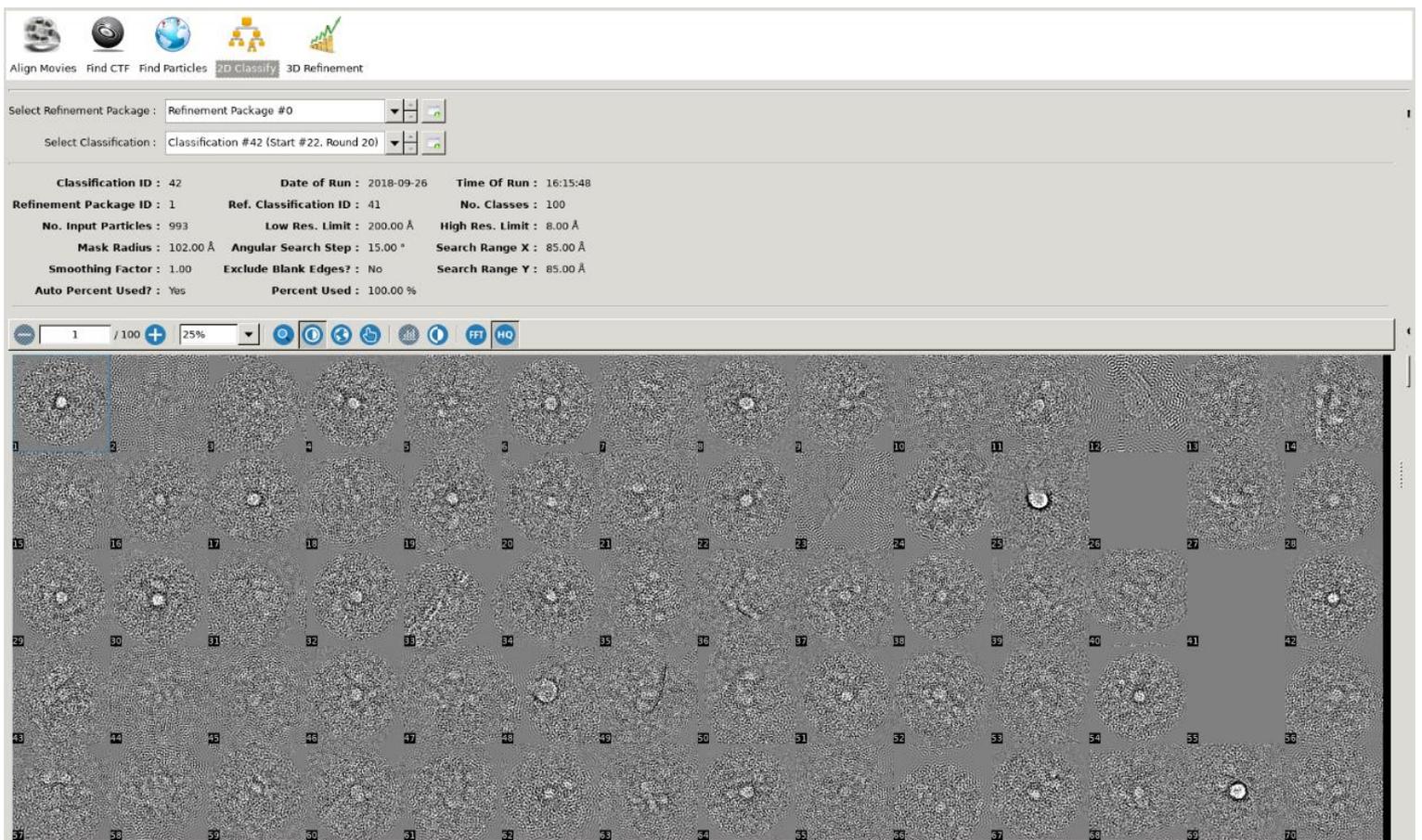
**Figure 61 Correction de la CTF par le logiciel cisTEM**



**Figure 62 Sélection des particules sous cisTEM.** A) sélection automatique, il n'y a pas de distinction entre particules et le carbone continu. B) sélection manuelle. C) Ajout d'un filtre passe-bande permettant de discerner plus finement les particules.

De même que pour le logiciel Relion 1.4, le picking automatique a été préalablement testé, mais comme précédemment le contraste ne permet pas au logiciel de discerner les particules de façon fine. Pour la sélection manuelle, le logiciel possède l'avantage de pouvoir ajouter des filtres après ouverture des images et ainsi sélectionner au mieux les particules. (Figure 62)

La classification 2D réalisée par le logiciel cisTEM semble moins adéquate que celle de Relion. Les classes 2D semblent présenter des incohérences au vu des particules sélectionnées (Figure 63). Certaines ne présentent aucune représentation 2D (classe 2) d'autres présente des particules en forme de grappes (classe 57).



**Figure 63 Classes 2D réalisées avec cisTEM.**

La difficulté majeure rencontrée ici est la faible présence de complexes structurés d'une part et le bruit de fond important. Bien qu'une classification puisse sortir le signal par moyennation, il est nécessaire de sélectionner correctement les particules et les centrer. Il

serait ici intéressant dans l'avenir de reprendre le dataset et essayer d'autres logiciels tels que Warp ou encore cryoSPARC.

Néanmoins, la préparation des grilles de cryo-EM, nécessite encore quelques optimisations en vue d'améliorer encore la qualité des images et la stabilité du complexe.

### III. Conclusion

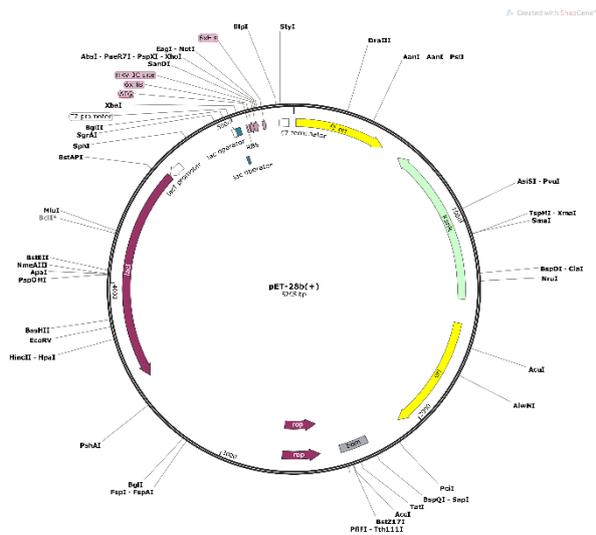
Pour conclure bien qu'aucune structure à haute résolution ne fut obtenue durant mes quatre années de thèse ; les expériences réalisées ont permis l'obtention d'informations préliminaires d'intérêts. En effet, le complexe NuRD a pu être visualisé et des informations morphologiques obtenues. Les conditions de purification et la réalisation des grilles en vue d'obtenir des complexes stables ont été optimisées permettant ici de mettre en avant des perspectives prometteuses dans l'obtention future d'informations structurales sur le complexe. Dans un premier temps il serait intéressant de reproduire les grilles (Figure 48 et 56). Ainsi ces échantillons pourraient être utilisés pour collecter des données en coloration négative de manière à obtenir des informations tridimensionnelles de morphologie du complexe. Puis dans un second temps passer à une étude en conditions cryogénique.

Parmi les centaines de conditions testées en conditions cryogénique, il est évident que deux techniques ont permis la première visualisation de la topologie du complexe : il s'agit d'une part de l'association d'un gradient de type GraDeR à un GraFix mais également à l'utilisation de grilles en carbone continu. L'utilisation de ce dernier pourrait néanmoins être encore améliorée par une variation dans l'épaisseur du carbone ou l'utilisation de graphène oxyde transparent aux électrons.

L'obtention d'informations structurales permettrait ici d'enrichir les connaissances sur le complexe NuRD qui restent aujourd'hui que peu renseigné notamment en termes de stœchiométrie et sur les bases moléculaires de son fonctionnement.

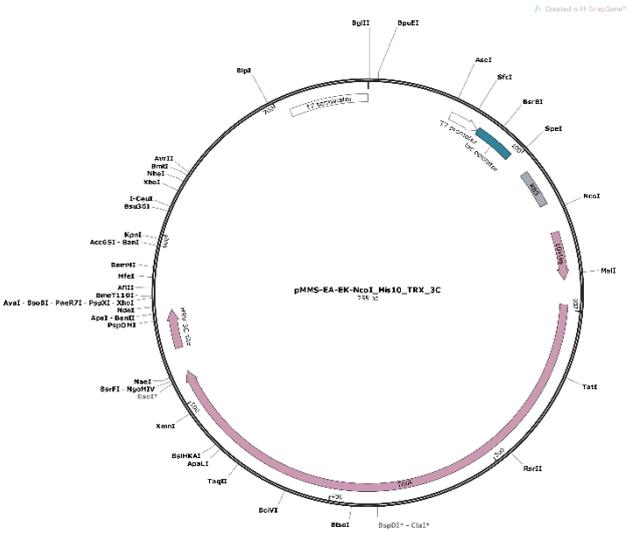
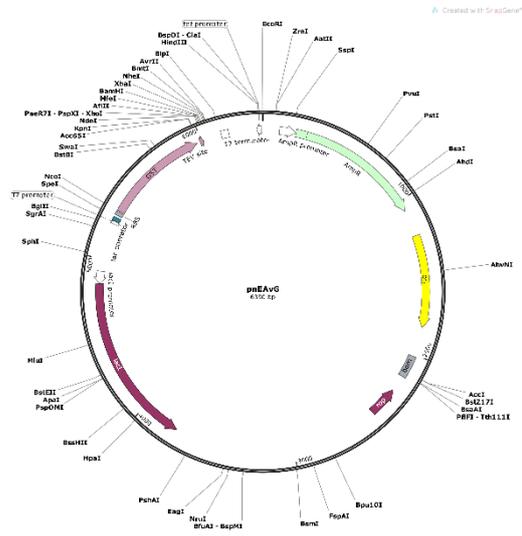






**pET28b** : le site de clonage est localisé entre le site de coupure HRV 3C et le codon stop, permettant la fusion en N-terminale d'une étiquette poly-histidine clivable par la protéase 3C. Présence également d'une étiquette poly-histidine, facultative en position C-terminale.

**pnEAVG** : le site de clonage est localisé entre le site de coupure TEV et le codon stop. L'étiquette fusionnée à la protéine sera ici la GST connu pour sa capacité à améliorer la solubilité et l'expression de protéines recombinantes.



**pMMS** : le site de clonage est localisé entre le site de coupure HRV 3C et le codon stop, permettant la fusion en N-terminale d'une étiquette histidine et de l'étiquette thioredoxine pouvant servir d'étiquette de purification mais va également aider au repliement des protéines.

Figure 64 Carte et description des plasmides utilisés

# PARTIE II: La sous-unité MBD 3 et sa spécificité à l'ADN

---

La seconde partie de mon projet de thèse concernait l'étude de la sous-unité MBD3 de NuRD et sa capacité de fixation et de reconnaissance à l'ADN. Ainsi de la même façon que pour l'étude du complexe NuRD une démarche de biologie structurale a été mise au point, avec de nombreuses différences. En effet, la sous-unité MBD3 présente une taille inférieure à la limite pour une étude en microscopie électronique. La méthode choisie est donc la cristallographie aux rayons-X qui va nécessiter la cristallisation de la protéine et donc sa production et purification en quantité importante.

## I. Méthodes utilisées

### A. Biologie moléculaire

Afin de produire les protéines d'intérêts dans un système d'expression donné, son gène doit tout d'abord être cloné dans un vecteur d'expression adéquat. Pour les études de MBD3 mon choix s'est porté sur un système bactérien de surexpression : *Escherichia coli*.

#### 1. Les vecteurs

Le vecteur de clonage typiquement utilisé en système bactérien est le plasmide. Cette molécule d'ADN circulaire possède la capacité du à son origine de répllication de pouvoir se répliquer indépendamment de l'ADN chromosomique de la bactérie. De nombreux vecteurs commerciaux existent et peuvent être modifiés de manières à ajouter différents sites nécessaires à l'expression, tels que des gènes codant pour des étiquettes, des sites de clivage etc...

Les plasmides du système pET, développé par Novagen, permettent de placer l'expression de la protéine sous la dépendance du promoteur du phage T7 et de la séquence opératrice *lac*. Durant ma thèse, trois plasmides dérivés du système pET ont été utilisés et sont présentés ci-

contre. (Figure 64) Après identification du gène à cloné et le choix du vecteur, différentes techniques de clonage peuvent être utilisées.

## 2. Les techniques de clonage

### a) *SLIC*

La méthode SLIC permet comme son nom l'indique de s'affranchir de l'étape de ligation et ne nécessite pas l'intervention d'enzymes de restrictions. Elle est basée sur le principe de la recombinaison homologue entre l'insert et le vecteur de destination et se déroule en plusieurs étapes :

Etape 1 : Linéarisation du vecteur par PCR en utilisant une polymérase Phusion. Les primers utilisés présentent une partie de la séquence du gène à intégrer de manière à permettre la reconnaissance homologue. Une fois le vecteur linéarisé amplifié, un traitement à DpnI permet de dégrader le vecteur circulaire initial, méthylé, et de conserver le vecteur néo synthétisé, linéaire et non-méthylé. De plus le vecteur est également traité avec la CIP, une phosphatase alcaline permettant la déphosphorylation en 5' et 3' de l'ADN. Parallèlement l'insert est également traité par PCR, pour l'insertion de bornes complémentaires à celles du vecteur.

Etape 2 : Le vecteur et l'insert sont traités ensemble par la T4 DNA polymérase. En absence de dNTP, la polymérase présente une activité 3'-5' exonucléase qui permet la digestion un à un des nucléotides, formant ainsi des extrémités cohésives. Cette étape est rapide et ne nécessite qu'une incubation d'environ 3 minutes à température ambiante.

Etape 3 : Le produit est alors transformé dans des bactéries *E. coli* DH5 $\alpha$ . La ligation du vecteur et de l'insert s'effectue naturellement dans la bactérie au cours du processus de réplication. Les bactéries sont finalement étalées sur une boîte de pétri de LB agar additionné d'antibiotique, et laissé à 37°C sur la nuit.

### b) *RF-free*

Cette méthode de clonage, permet l'insertion de séquence à n'importe quelle position dans le plasmide vecteur sans l'intervention d'enzymes de restrictions. Pour cela une paire d'amorces est conçue contenant une séquence complémentaire à la fois de l'insert désiré et du plasmide cible. Ces deux amorces sont par la suite utilisées pour amplifier le gène d'intérêt par PCR. Le produit résultat est purifié et utilisé comme méga-amorce pour une seconde PCR utilisant le

plasmide vecteur comme matrice. Le plasmide est dès lors amplifié dans les deux directions. Par la suite l'ajout de DpnI permet la dégradation des plasmides parentaux n'ayant pas incorporé le gène, dû à leur méthylation. Le produit final est utilisé pour transformer des souches bactériennes compétentes.

Cette méthode de clonage a été utilisée pour cloner les différents gènes ci-dessous en vue de leur co-expression:

MBD3 2-260 : Cette construction se caractérise par la présence du domaine de fixation MBD et également du domaine de super-hélice responsable de l'interaction avec GATAD2A. De façon plus précise cette construction comme la suivante présentent une délétion de la queue poly-E, à l'origine des difficultés rencontrées pour la purification de la protéine en entière dans le l'équipe précédemment.

MBD3 2-267

GATAD2A : Cette construction est caractéristique de la protéine GATAD2A entière.

GATAD2A domaine CC : cette construction est caractéristique du domaine super-hélice interagissant avec MBD3.

## B. Expression

Les protéines étudiées sont toutes exprimées chez *Escherichia coli*. Pour tous les plasmides synthétisés, l'expression du gène d'intérêt est sous le contrôle du promoteur T7. Pour contrôler l'expression de la protéine par la bactérie, le plasmide recombinant est introduit dans des souches dite DE3 possédant dans leur génome un gène codant pour l'ARN polymérase T7 (provenant d'une infection par le bactériophage lambda DE3) dépendant du promoteur lac UV5 inductible par l'IPTG. Ainsi la présence d'IPTG va permettre par un mécanisme de dé-répression la transcription et l'expression de l'ARN polymérase T7 qui va initier la transcription du gène d'intérêt.

Différentes souches bactérienne d'expression sont disponibles à l'institut et ont pu être testée pour la surexpression des différentes protéines : BL21(DE3), BL21(DE3) pLysS, BL21(DE3) pLysSpRARE, BL21(DE3) pRARE.

Le protocole de transformation est le même pour les différentes souches bactériennes. 100 µl de cellules compétentes sont mises à décongélés sur glace puis 2 µl de plasmide recombinant

(ng/ $\mu$ l) sont ajoutés. Le tout est laissé sur glace pendant 15 minutes. La transformation est réalisée par un choc thermique à 42 °C pendant 45 secondes à l'aide d'un bain sec. Les échantillons sont placés par la suite sur la glace pendant 3 minutes puis suspendus dans 200  $\mu$ l de milieu LB2x et incubés 1 heure à 37 °C. Pour finir les suspensions bactériennes sont étalées sur des boîtes de pétri de géloses de LB supplémentées des antibiotiques adéquats. Cette méthode va permettre de sélectionner les clones recombinants. Pour les bactéries transformées avec le vecteur pET28b la kanamycine (50  $\mu$ g/ml) est utilisée. Pour celles transformée avec les vecteurs pnEAVG et pMMS il s'agira de l'ampicilline (50  $\mu$ g/ml) pour BL21(DE3) et supplémenté de pour les autres souches. Les boîtes sont par la suite laissées sur la nuit à 37 degrés Celsius.

Les conditions optimales d'expression pour chacune des protéines recombinante sont identifiées par variation de nombreux paramètres tels que la souche bactérienne mais aussi la concentration en IPTG, les températures et les temps d'induction ou encore le type de milieu utilisé. Ces optimisations se font sur de petits volumes.

Le protocole d'expression à grande échelle reste similaire, une colonie est sélectionnées et inoculées dans 5 mL de milieu liquide dans une fiole contenant le ou les antibiotiques appropriés. La fiole est placée à 37°C toute la nuit sous agitation horizontal à 180 rpm. Ces mini-cultures vont servir à ensemercer des précultures de 100 mL de milieu supplémenté d'antibiotiques. Après quelques heures d'incubation à 37°C sous agitation, les bactéries se trouvent en phase exponentielle de croissance et sont utilisées pour ensemercer les 6L de culture de LB2x contenant les antibiotiques appropriés. Les cultures sont incubées à 37°C sous agitation (180 rpm) jusqu'à une  $DO_{600nm}$  comprise entre 0,3 et 0,4. Puis la température est progressivement diminuer jusqu'à 15°C, permettant d'atteindre de façon simultanée une  $DO_{600nm}$  comprise entre 0,6 et 0.8. L'expression de la protéine est alors induite par l'ajout de 0.4 mM d'IPTG sur la nuit à 15°C sous agitation.

Le lendemain, les bactéries sont sédimentées par centrifugation (4200 rpm pendant 20 minutes) et conservées à -20°C.

## C. Etudes biochimiques de MBD3 2-72 et 2-133

### 1. Lyse cellulaire

Les bactéries sédimentées sont pesées, le rendement obtenu pour l'expression en cellules BL21(DE3) pRARE2 est d'environ 6 grammes par litre de culture. Les cellules sont

resuspendues dans un tampon de lyse (200 mL pour 6 L de culture) composé de 50 mM Tris-HCl pH7.5, 200 mM NaCl, 4 mM CHAPS, 1 mM PMSF, 15 mg de lysozyme, 2 mM BME. On ajoute également des inhibiteurs de protéase (cOmplete EDTA-free, Roche), une pastille par 25 mL. Suite à la resuspension, l'échantillon est soniqué pendant deux fois 5 minutes avec une pause de 1 minute à une amplitude de 60% et des cycles de 0,6. Toutes ces étapes sont réalisées en chambre froide ou sur glace. C'est conditions sont celles optimisées pour la purification du domaine MBD de MBD3.

L'extrait total est par la suite ultracentrifugé pendant 1 heure à 35000 g de manière à séparer l'extrait soluble composé des protéines solubles, des acides nucléiques de l'extrait insoluble composé des protéines insolubles, des débris cellulaires.

Différentes étapes de purification sont nécessaires pour isoler la protéine d'intérêt à un haut niveau de pureté.

## 2. Chromatographie

### a) Affinité

- Nickel-NTA

La première étape de purification utilisée est une chromatographie d'affinité utilisant un métal ( $\text{Ni}^{2+}$ ) immobilisé sur une résine d'agarose. Elle repose sur la capacité des noyaux imidazoles des histidines à chélater des ions divalents comme le nickel ou le cobalt. Cette technique est possible du fait de l'ajout d'une étiquette à l'extrémité N-terminale de la protéine caractérisée par six résidus histidines. La méthode choisi ici était réalisé en « batch » et ne requiert pas l'utilisation de colonne ou d'un système de type Akta. Cette méthode requiert l'incubation de l'échantillon sur la résine sous agitation. Lors de cette première étape, l'extrait soluble était directement utilisé, il est composé de résidus de membranes et de nombreux contaminant qui peuvent détériorer les systèmes Akta. De plus cette méthode permet également de pouvoir couper l'étiquette histidine lorsque la protéine est fixée à la résine et limiter la quantité de contaminant libérer par une élution.

Pour cette méthode la résine est équilibrée avec cinq volumes de tampon À (20 mM Tris-HCl pH7,5, 200 mM NaCl, 10 mM d'imidazole, 4 mM de CHAPS et 2 mM de BME). L'extrait soluble est incubé sur la résine pendant environ 1 heure en chambre froide à 4 degrés sous rotation constante. La résine est par la suite lavée avec le même tampon jusqu'à ce que

l'absorbance à 280 nm soit proche de 0. Le dernier lavage est réalisé avec une solution similaire au tampon A avec 20 à 40 mM d'imidazole, ce qui va permettre d'éliminer les protéines fixées de façon non spécifiques.

Pour récupérer la protéine, deux méthodes sont utilisées :

La première est réalisée pour conserver l'étiquette 6xHis. Pour cela la protéine est éluée en plusieurs étapes avec un tampon B (20 mM Tris-HCl pH 7,5, 200 mM NaCl, 250-300 mM d'imidazole, 4 mM de CHAPS et 2 mM de BME). L'éluion est fractionnée manuellement soit par centrifugation successive soit par l'utilisation de colonne dans laquelle la résine est coulée est l'éluion se fait par gravité.

La seconde technique est utilisée pour éliminer la queue histidine. Pour cela suite au lavage de la résine environ 15 mL de tampon A est utilisé pour resuspendre la résine implémenté par la protéase P3C (1/100 de la concentration en protéine). L'incubation avec la protéase se fait sur la nuit en chambre froide sous rotation continue. Le lendemain le surnageant est récupéré par centrifugation (700 g pendant 2 minutes). La résine est lavée pour récupérer la protéine clivée. Pour cette méthode une étape supplémentaire dite de Nickel inverse est nécessaire à l'élimination de la protéase HRV 3C présentant une étiquette 6xHis.

- Héparine

Cette étape de purification est réalisée à la suite de la purification sur résine Nickel. L'héparine est un glycosaminoglycane, qui, dû à sa structure et ses charges négatives est capable de mimer la structure de l'ADN. Cette méthode est donc idéale à la purification de protéines liant l'ADN.

Pour cela, une colonne HiTrap Héparine 5 mL (GE Healthcare) a été utilisée. La résine de cette colonne correspond à des billes de Sepharose-6B modifiée par la présence d'héparine. L'échantillon est appliqué sur la colonne et la colonne est nettoyée avec un tampon héparine A à 70% (20 mM Tris-HCl pH7,5, 5 mM EDTA, 2 mM CHAPS, 3 mM DTT) et 30% d'un tampon B (20 mM Tris-HCl pH7,5, 1 M NaCl, 5 mM EDTA, 1 mM CHAPS, 2 mM DTT). La protéine est éluée par un gradient croissant en tampon héparine B qui va augmenter la force ionique et ainsi déstabiliser l'interaction protéine-héparine.

### *b) Echangeuse d'ions*

De façon à éliminer l'ADN contaminant qui serait fixé au domaine MBD de MBD3 un autre type de chromatographie fut testé. Il s'agit d'une échangeuse d'ions et plus précisément ici, d'une échangeuse d'anions. Cette technique se base sur la charge des protéines. En effet, la charge nette des protéines varie en fonction du pH de la solution tampon. A un pH supérieur au pI la protéine est chargée négativement et lorsque le pH est inférieur au pI la charge nette sera positive. Les colonnes échangeuses d'anions de type mono Q-sepharose, les protéines possédant une charge nette négative vont se fixer aux groupements triméthyl ammonium chargés positivement de la résine. Le point isoélectrique du domaine MBD est d'environ 9.1, à pH7,5, la charge nette de la protéine est donc positive et ne va pas permettre l'interaction avec la résine. A l'inverse de l'ADN chargé négativement qui lui sera retenu. L'ADN sera élué en appliquant un gradient de NaCl de 300 mM à 1 M.

Une colonne de gel monoQ-sepharose de 5 mL (GE Healthcare) est équilibrée avec cinq volumes de tampon MonoQ-A (20 mM Tris-HCl pH7,5, 300 mM NaCl, 5 mM EDTA, 2 mM CHAPS, 3 mM DTT). L'échantillon est appliqué sur la colonne à un débit de 2 mL/min. L'échantillon non fixé à la résine est récupéré il s'agit de l'échantillon contenant MBD3 nettoyé de l'ADN. Une fois cette fraction récoltée, la colonne est lavée avec le tampon MonoQ-A et les contaminants sont élués en appliquant un gradient de 300 mM à 1 M de NaCl. La fraction non retenue et les fractions éluées sont caractérisées sur gel d'électrophorèse SDS-PAGE. La présence d'ADN dans les différentes fractions est contrôlée en mesurant le ratio A260/A280 avec le nanodrop.

### *c) Chromatographie d'exclusion stérique*

La méthode de chromatographie d'exclusion stérique ou gel filtration permet de séparer les molécules en fonction de leur rayon hydrodynamique. Ces colonnes sont composées d'une résine de granules poreuses dont la taille des pores est choisie en fonction du diamètre des protéines à séparer. A la différence des méthodes de chromatographies précédente, cette méthode ne repose donc sur aucun type d'interaction, la solution tampon n'affectera donc pas la résolution de cette technique. Cette méthode se base sur un mécanisme d'exclusion, en effet, les molécules dont le diamètre est supérieur à celui des pores, sont exclues des billes de gel et seront éluées en premières. A l'inverse, les petites protéines vont diffuser dans le gel et seront donc éluées plus tardivement. L'élué des protéines est donc inversement

proportionnelle à leur rayon hydrodynamique. Cette méthode permet de mettre en évidence les différentes formes moléculaires d'une protéine comme une possible dimérisation ou agrégation mais également c'est une étape supplémentaire pour séparer la protéine d'intérêt d'autres contaminants. Elle est classiquement utilisée à la fin des étapes de purification pour les changements de tampons.

Pour MBD3 deux colonnes Superdex S200 10/300 branchées en série sur le système akta, sont équilibrées avec 1 volume de tampon GF (20 mM HEPES pH7,5, 150 mM NaCl, 0.5 mM TCEP et 2 mM MgCl<sub>2</sub>). Environ 1mL d'échantillon était injecté et élué sur un volume de tampon. Des fractions de 200uL étaient collectées et analysées sur gel SDS-PAGE.

### 3. Hybridation des oligonucléotides et formation des complexes

Les oligonucléotides utilisés pour la formation des complexes avec le domaine MBD de MBD3 ont été commandé chez eurofin sous forme de simples brins pour lesquels une étape de renaturation était nécessaire appelé annealing.

Pour cela chaque ADN simple brin précipité est resuspendu dans un volume de tampon d'annealing (10 mM TRIS pH7,5, 100 mM NaCl, 1% DMSO, 0.2 mM EDTA) permettant d'obtenir la même concentration pour chaque simple brin. La concentration est vérifiée par spectrophotométrie. Les deux oligonucléotides complémentaires sont mélangés de façon équimolaire et hybridés à l'aide d'un thermocycler selon le protocole suivant :

La température est préchauffée à 95°C, et maintenue pendant 2 min afin de permettre la dissociation des structures secondaires dû à des complémentarités de bases au sein des oligonucléotides. La température est ensuite diminuée jusqu'à 25°C sur 45 minutes environ. Suite à cela 1mM en concentration finale est ajouté aux différents échantillons et la concentration est vérifiée par spectrophotométrie. L'hybridation quant à elle est vérifiée par migration sur gel natif 20% TBE (**Tableau 6**). Les solutions sont conservées sous formes d'aliquots à -20°C.

Pour former les complexes, l'ADN est ajouté à la protéine à différents ratios molaires pour trouver les conditions adéquates de complexes. Principalement ici les ratios étaient de 1:1 ou 1,1:1, un léger excès d'ADN peut être favorable à la formation de complexes. Les complexes sont laissés 20 minutes à 12 heures à 4°C avant d'être utilisé pour des études de caractérisation biophysiques ou des essais de cristallisation.

## D. Etudes biophysique

### 1. Concentration

La méthode de biologie structurale choisie pour l'étude de l'interaction ADN-MBD3 étant la cristallographie aux rayons-X, le complexe doit passer par une étape de cristallisation. Cette étape nécessite des concentrations plus importantes que la microscopie présentée dans le chapitre précédent. Pour cela différentes techniques de concentration ont pu être utilisées durant ma thèse :

- La concentration en amicon ultra
- La concentration par précipitation au sulfate d'ammonium
- La concentration avec des concentrateurs de type Vivaspin

### 2. Expérience de retard sur gel

L'expérience de retard sur gel, également connue sous le nom d'EMSA (pour electromobility shift assay), est une méthode d'électrophorèse en condition native, permettant de mettre en évidence des interactions ADN-protéine. Cette technique se base sur l'idée simple que les acides nucléiques seuls présentent une mobilité électrophorétique différente s'ils sont liés à une protéine, par rapport à une référence sans protéine. Ainsi, un fragment d'ADN complexé à une protéine devrait migrer plus lentement, en raison de sa masse moléculaire plus importante. Il en résulte un décalage sur le gel, ou « shift ». Cette méthode est plus qualitative que quantitative et présente de nombreuses limites. Ainsi, la stabilité du complexe va impacter sur la migration. L'électrophorèse native, bien qu'effectuée à 4°C, génère un dégagement de chaleur qui peut déstabiliser les interactions au sein du complexe et dénaturer ce dernier. Typiquement, la migration s'effectue à un voltage réduit (environ 150 V), intensité faible (15 mA) pour une puissance totale maximale d'environ 2 W.

<b>Gels natifs</b>		
<b>Type de gel</b>	<b>TBE pH8,3</b>	<b>Tris-Glycine pH8,9</b>
<b>gel</b>	8 à 20% Acryl/Bisacrylamide 39 :1 89 mM Tris-HCl pH 8,3 89 mM acide borique 2 mM EDTA 0.15% APS 0.2% TEMED	8 à 20% Acryl/Bisacrylamide 39 :1 300 mM Tris-HCl pH 8.9 0.15% APS 0.2% TEMED
<b>Solution de migration</b>	Solution 0.5X: 47 mM Tris 45 mM acide borique 1 mM EDTA	Solution 1X : 27 mM Tris 192 mM glycine
<b>Solution de charge</b>	1% de sucrose 40% est ajouté pour 10 µL de solution	1% de sucrose 40% est ajouté pour 10 µL de solution
<b>Solution de coloration</b>	Bromure d'éthidium Bleu de coomassie (nécessite une décoloration)	

**Tableau 6 composition des gels natifs**

Pour les gels natifs, il est conseillé d'utiliser des solutions de moindre force ionique pour éviter toute déstabilisation de complexe. De nombreux types de tampons peuvent être utilisés et largement décrits dans la littérature. Il est important de se baser également sur le pI de la protéine. L'absence de l'ajout de SDS, entraîne qu'en plus de migrer selon sa taille, la migration de la protéine va dépendre de sa charge. Le pH de la solution de migration et celui du gel doivent donc être supérieurs au pI de la protéine. Le pI de MBD3 étant supérieur à 9, le pH de mes gels natifs, devraient être au-dessus de cette limite. Néanmoins, l'utilisation d'un tel pH peut entraîner également des dissociations. Pour les différentes études présentées ici, je me suis contenté d'utiliser une solution de tris-glycine à pH8.9 ou TBE à pH8.2 permettant la détection de complexe et d'ADN seul mais pas de la protéine seule. (Tableau 6)

### 3. Dynamic light scattering ou DLS

La diffusion dynamique de la lumière ou DLS, est l'une des techniques les plus répandues pour déterminer la taille et la distribution de petites particules en solution. Elle consiste à

envoyer un faisceau lumineux monochromatique, tel qu'un laser, sur une solution contenant des particules en suspension. De manière générale, cette méthode se base sur une détection de l'intensité de la diffusion par les particules à 90 degrés au cours du temps. Cette mesure va permettre, par application de la relation de Stoke-Einstein, d'interpréter le rayon de stoke ou rayon hydrodynamique des particules. La méthode se base sur deux hypothèses : la première étant que les particules en solution sont soumises à un mouvement brownien, aléatoire. La seconde est que les particules sont sphériques et de petite tailles.

Ces mesures nous permettent de déterminer la polydispersité d'un échantillon protéique, c'est-à-dire l'écart de la distribution de taille d'une population de particule. Pour la cristallisation des protéines, une faible polydispersité est généralement requise. On considère qu'un indice de polydispersité supérieur à 20% réduit la probabilité de nucléation et de croissance cristalline.

A l'institut, nous utilisons un DynaPro DLS system (Wyatt), permettant l'analyse d'un échantillon à une concentration de 1 mg/mL dans un volume de 50  $\mu$ L dans une cuve en plastique ou de 3  $\mu$ L dans une cuve en quartz. Préalablement à toute mesure, l'échantillon est ultracentrifugé pour éliminer les agrégats.

#### 4. Prometheus

Le prometheus se base sur une technique de nanoDSF pour nano Differential Scanning Fluorimetry. Cette méthode permet de déterminer la stabilité d'une protéine en utilisant la fluorescence intrinsèque des résidus tryptophanes principalement, et dans une moindre mesure la tyrosine. La fluorescence des tryptophanes dans une protéine dépend fortement de son environnement. En suivant les changements de fluorescence, la stabilité chimique et thermique peut être évaluée de manière rapide sans avoir à utiliser un quelconque marqueur. Ainsi la fluorescence des résidus va être associée à une dissociation de la protéine d'intérêt.

La stabilité thermique d'une protéine est généralement décrite par la «température de fusion» ou «T<sub>m</sub>», à laquelle 50% de la population de protéines est sous forme dépliée, ce qui correspond au point médian de la transition de l'état structuré à déstructuré. Cette technique permet donc de cribler un large panel de conditions différentes, sans nécessité de marqueurs. Elle est également très sensible et ne requiert que de petits volumes.

Pour toutes les expériences réalisées, la protéine était diluée à hauteur de 1/10 dans les tampons à tester de manière à obtenir la même concentration pour tous les échantillons et pouvoir négliger l'impact du tampon de base (volume final 12  $\mu$ L). La température initiale

était fixée à 20°C puis augmenté progressivement à 95°C. Toute variation de fluorescence est quasi-instantanément détectée de façon sensible et retranscrit sous forme de courbes représentant le rapport F350/F330 en fonction de la température. La température de fusion de l'échantillon est calculée à partir de la courbe et rend compte de l'impact du milieu sur la stabilité thermique de la protéine que l'on peut extrapoler comme l'impact sur la stabilité de la protéine.

## 5. SEC-MALS

SEC-MALS, pour Size Exclusion Chromatography Multi Angle Light Scattering, est une technique permettant une mesure précise de la diffusion de la lumière par les molécules en solution. Cette mesure est faite en parallèle à la séparation des composés dans une colonne d'exclusion stérique. L'analyse informatique de ces informations permet d'obtenir les différentes masses moléculaires des composés en solution (Tarazona and Saiz, 2003).

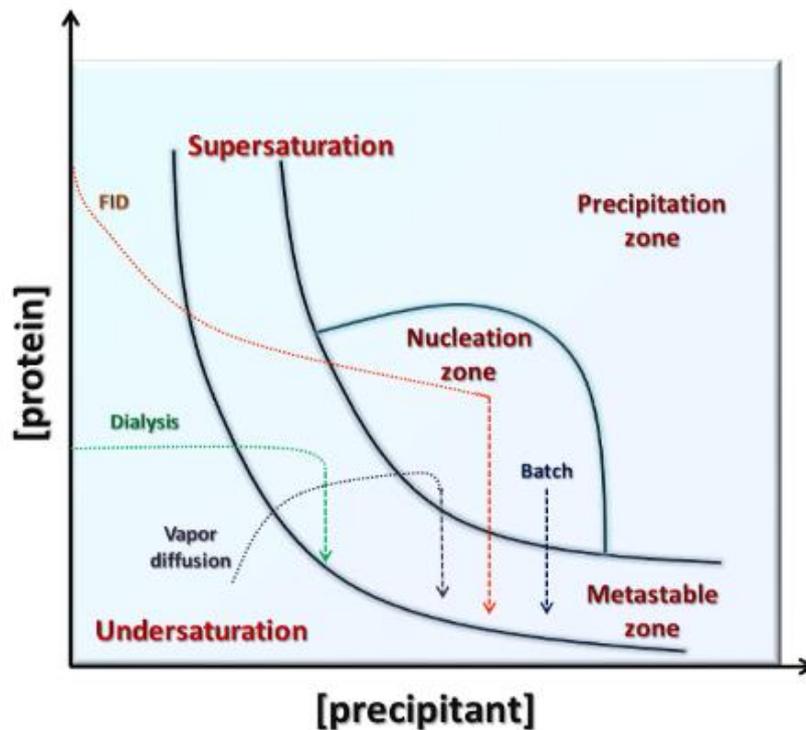
Pour cela, la colonne est équilibrée avec le tampon dans lequel se trouve notre échantillon, et 50uL à 4mg/mL est injecté sur une Superdex 200 10/300. Le spectre MALS a été enregistré par un laser à une longueur d'onde de 690 nm en utilisant un détecteur DAWN-HELEOS. L'indice de réfraction (IR) a été mesuré par un détecteur Optilab T-rEX. La masse a été déterminée en analysant l'indice de réfraction différentiel au pic d'élution de chaque échantillon en utilisant le modèle de Debye pour les protéines (Edelman, 1992), intégré au logiciel ASTRA (version 6.5.0.3).

## E. Biologie structurale: cristallographie aux rayons X

### 1. Cristallisation

Parmi les techniques de biologie structurale, la cristallographie aux rayons X fut choisie pour l'étude des complexes protéines-acides nucléiques entre MBD3 et différents oligonucléotides. Pour permettre cette étude structurale, la première étape est la formation de cristaux du complexe. Le but étant l'obtention de cristaux présentant une qualité de f=diffraction optimale pour le traitement bio-informatique des données. Pour on va chercher à diminuer progressivement la solubilité de notre complexe soluble et pure de manière à atteindre un état de sursaturation. La difficulté ici, étant de trouver les conditions idéales à une sursaturation qui va permettre aux complexes de s'organiser de façon ordonnée sous forme de cristaux et éviter toute agrégation. La cristallisation est réalisée en plusieurs étapes, la première étant la nucléation, est l'étape limitante dans la croissance cristalline. Suite à la nucléation, le cristal

va continuer à se former en trois dimension dû à la sursaturation environnement. Ce processus de cristallisation va être réalisé d'une part par l'ajout de précipitant qui vont favoriser la sursaturation accéléré par l'évaporation. Le diagramme de phase illustre l'impact des concentrations en protéines et en précipitant sur la solubilité de la protéine d'intérêt.



**Figure 65 Diagramme de phase**

De nombreux paramètres physico-chimiques vont permettre de contrôler cette sursaturation en vue d'obtenir des monocristaux.

- La température, elle va permettre de contrôler la vitesse de nucléation et de cristallisation. Néanmoins ici les expériences étaient réalisées à 4 degrés dû à l'instabilité du complexe à température ambiante.
- Le pH, comme on l'a vu précédemment, les variations de pH vont permettre de modifier les charges en surface de nos protéines.
- La concentration en protéine
- La concentration et le type de sel
- La présence de polymères, comme le PEG, permettent d'altérer les contacts entre le solvant et l'environnement de la protéine.

- La présence d'additifs, qui peuvent se fixer, modifier ou encore stabiliser la conformation de protéines.

Un certain nombre de techniques ont été développées pour amener une solution de protéines dans un état sursaturé. En figure 64, les différentes techniques permettant d'une part d'atteindre le point de nucléation puis la zone métastables sont présentées. Dans les parties suivantes je présenterais les principales techniques utilisées pour ce projet, leurs avantages et les différents protocoles.

## 2. Méthodes utilisées

### a) *Gouttes assises et gouttes suspendues : la diffusion de vapeur*

Cette technique est la plus couramment utilisée et se base sur un procédé de diffusion. Une goutte contenant le complexe et la solution de cristallisation communément dans un ratio 1:1 est mise en équilibre avec le réservoir contenant la même solution de cristallisation. Le puit est scellé et la goutte va se déshydrater au cours du temps pour atteindre un équilibre osmotique. De ce fait, les concentrations en précipitant et en protéine au sein de la goutte vont augmenter et amener le complexe dans un état de sursaturation favorable à la cristallisation.

Pour ces tests de cristallisation deux types de plaque ont été utilisés : Des plaques en gouttes assises, ou la goutte repose directement sur un support solide. Mais également des plaques en gouttes suspendues ou la goutte est placée sur la face inférieure d'une lame de protection en verre siliconé. Les tests préliminaires étaient réalisés en goutte assises à l'aide du robot mosquito® puis stockés et imagés par le système Rock Imager® à 4 ou 20°C. Les gouttes étaient de 0.1+0.1 nl équilibrées avec un réservoir contenant 40uL de solution de cristallisation. Lors de ces tests, des screens commerciaux ont été utilisé de manière à pouvoir tester un maximum de conditions avec un minimum de complexe. Une fois des conditions identifiées, l'optimisation par variation de pH, sels, et précipitants étaient réalisés dans des plaques de gouttes suspendues. En effet, les gouttes assises sont faciles à mettre en place mais lors de tous les essais de cristallisation les cristaux nécessitent un support solide pour leur croissance. C'est pour cette raison que les différentes autres techniques telles que la méthode dite en microbatch, en gel ou encore des méthodes d'ensemencement ont été préféré par la suite.

### *b) Cristallisation dans l'huile*

La méthode de gouttes dans l'huile, est une technique s'affranchissant du mécanisme de diffusion. Tous les réactifs impliqués dans la cristallisation sont présents à une concentration spécifique et aucune augmentation significative de concentration en protéine et en précipitant ne peut se produire dans la goutte. La supersaturation doit être réalisée directement en avant de placer la goutte sous huile. Pour ces expériences, les conditions de cristallisation identifiées doivent être de nouveau optimisées en terme de concentration en protéines et en précipitant, mais également le choix du type d'huile doit être pris en compte. Pour cela trois type d'huile ont été testées permettant une évaporation plus ou moins importante de la goutte, l'huile de paraffine qui ne permet aucune évaporation, et des mélange de paraffine et silicone qui permet de contrôler l'évaporation (ratio 1:1 et 0.75:0.25) de la goutte.

### *c) Cristallisation en gel*

Comme il sera présenté dans les résultats, les cristaux nécessitent un support solide lors de la croissance cristalline. C'est pour cette raison que des tests de cristallisation en présence d'agarose furent réalisés. L'avantage ici, étant que l'agarose va créer un gel composé d'un réseau de fibre au sein de la goutte pouvant servir de support solide (Biertümpfel et al., 2002). De plus ces gels d'agarose n'interagissent pas avec les protéines et permettent de pouvoir mettre en place tout type de cristallisation. Pour ces tests, entre 0.1 et 0.4% (m/v) d'agarose est ajouté aux conditions de cristallisation. Ainsi lors de la formation des gouttes en 1+1  $\mu\text{L}$  la concentration finale d'agarose se situera entre 0.05 et 0.2%. La technique en goutte suspendue fut ici utilisée.

### *d) Ensemencement*

Toujours dans l'optique d'optimiser les conditions de cristallisation et permettre la croissance des cristaux au sein de la goutte sans être accolé au support. Une technique d'ensemencement a été utilisée, et plus particulièrement de micro-ensemencement. Pour cela les cristaux obtenus collés au support sont écrasés et réduits en microcristaux par un mécanisme dit de « drop and crash ». Il consiste en l'ajout progressif de buffer de cristallisation (10  $\mu\text{L}$ ) puis d'écraser les cristaux, pipeter l'échantillon, rajouter encore de la solution et écraser. Cette itération de dilution et broyage permet d'obtenir le « seed stock » de 50  $\mu\text{L}$ . Cette solution est légèrement vortexée avec une bille de verre pour homogénéiser la taille des microcristaux. A partir de ce stock 3 dilutions sont réalisées permettant l'obtention de 4 solutions : le seed stock, 1/10, 1/100, 1/1000.

Ces différentes solutions d'ensemencement ont été utilisées pour réaliser des ensemencements croisés à l'aide du mosquito dans des screen commerciaux. Pour cela la concentration en complexe a été réduite à 2.5 mg/mL et supplémenté des différentes solutions seed dans un ratio 0.3:0.7.

Mais également pour ensemer les conditions de cristallisation identifiées en faisant varier la concentration en protéines, en précipitants, en sel, le pH. Pour cette technique une moustache de chat est plongée dans les solutions seed et utilisée pour transporter de petits fragments de cristaux de complexes dans une goutte. Cette forme de semis est très facile et peut être utilisée pour ensemer rapidement une série de gouttes. Une attention particulière a été apportée ici, de manière à ensemer les microcristaux au milieu de la goutte sans toucher le fond et éviter ainsi toute interaction avec le support.

#### *e) La diffraction aux rayons-X*

Du fait de l'organisation spatiale précise et répétée des molécules au sein d'un cristal, lors de l'exposition à des rayons X de longueur d'onde inférieure aux distances interatomiques, les rayonnements réémis par les atomes équivalents interfèrent négativement, sauf à des angles particuliers (définies par la loi de Bragg) où ils interfèrent positivement. Le diffractogramme du cristal est donc composé d'un nombre discret de points de diffraction.

Il est possible d'obtenir l'enveloppe électronique d'une protéine à partir de clichés de diffraction obtenus en exposant différentes orientations d'un même cristal à un faisceau de rayons X de longueur d'onde définie et une distance cristal-détecteur donnée. Le cristal est monté sur un goniomètre qui permet de le faire tourner. Des clichés de diffraction sont collectés pour différentes orientations. L'angle minimum sur lequel il est nécessaire de collecter des données pour résoudre la structure dépend de la symétrie du cristal.

Cependant, les clichés de diffraction correspondent à des images qui n'ont pas été refocalisées (il n'existe pas de lentilles pour rayons X) et ne contiennent plus l'information de phases. Différentes stratégies permettent de résoudre ce problème parmi lesquelles on peut noter le remplacement moléculaire utilisant les phases d'une structure homologue ou la diffusion anormale.

## II. Résultats

Un des objectifs de mon projet de thèse était la caractérisation structurale de la sous-unité MBD3 du complexe NuRD impliquée dans la reconnaissance à l'ADN au niveau de séquence de type CpG. En particulier, je me suis intéressée à l'interaction de cette sous-unité à l'ADN pour comprendre d'un point de vue moléculaire les détails de cette interaction. Mais également comprendre la perte de spécificité de la sous-unité pour l'ADN méthylé et comparer cette interaction avec celle réalisée par la sous-unité MBD2.

La protéine MBD3 est, comme on a pu le présenter dans l'introduction, une protéine d'environ 33kDa. Comme on peut le voir sur la figure 66, la protéine présente notamment à l'extrémité C-terminale une région hautement désordonnée caractérisée par la présence d'une séquence poly-E (acides aminés 272-281) diminuant grandement la stabilité de la protéine lors de sa purification. Ainsi durant sa thèse, le Docteur Morgan Torchy qui a initié ce projet, dans le passé a dû faire face à de nombreux problèmes lors de la purification et la formation de complexes ADN.

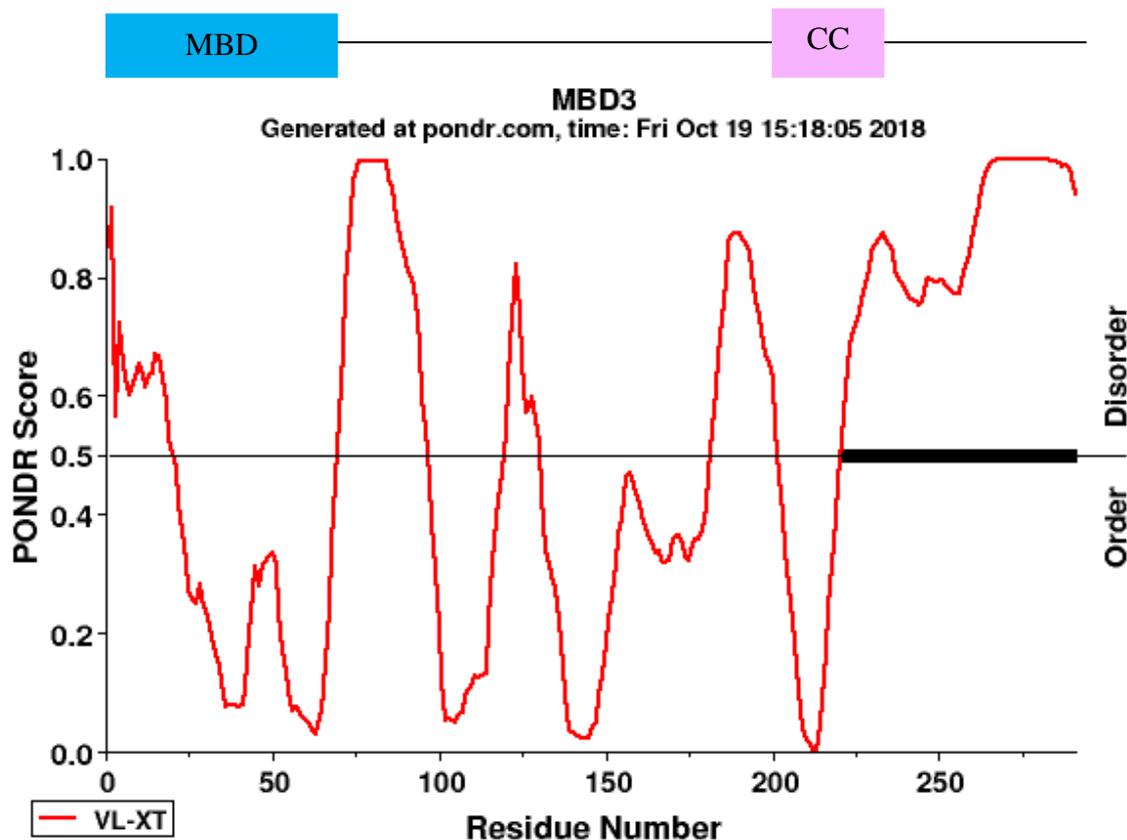


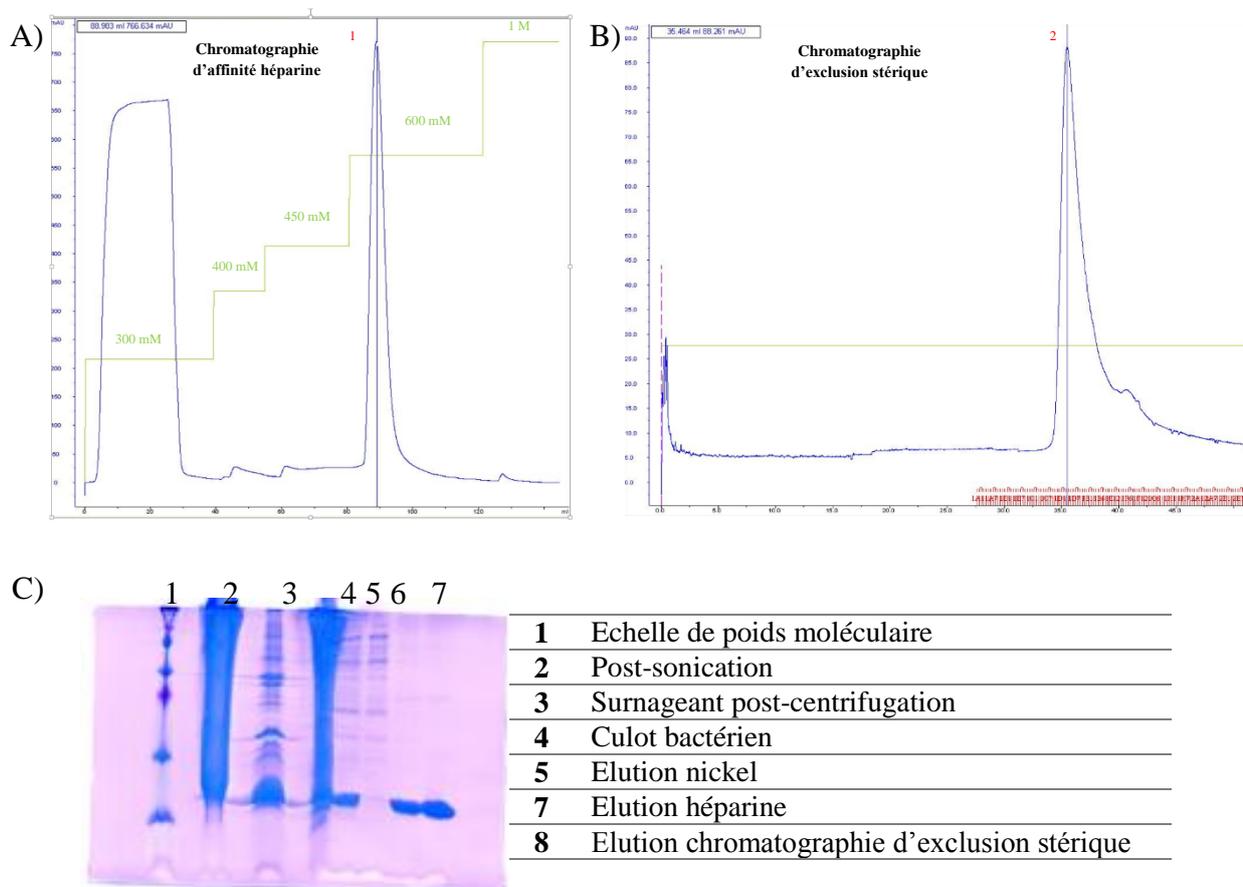
Figure 66 Prédiction des régions désordonnées de MBD3

Les protéines qui se lient à des molécules d'ADN ou d'ARN, comme ici pour MBD3, présentent souvent au sein de leur structure des boucles désordonnées ou flexibles pouvant gêner la formation d'un réseau cristallin bien ordonné (**Figure 66**). Ainsi, pour surmonter ce problème, l'identification du fragment structuré minimal à l'interaction avec l'ADN est souvent préférée. Il a été décidé de s'intéresser au domaine MBD de la protéine à l'origine de la reconnaissance à l'ADN. Ce domaine est caractérisé par les acides aminés 2 à 72 et a été cloné dans un vecteur de type pET28b avec une étiquette 6xHis et un site de clivage à la protéase 3C en N-terminal. Ce travail a été réalisé par le service de biologie moléculaire à l'IGBMC.

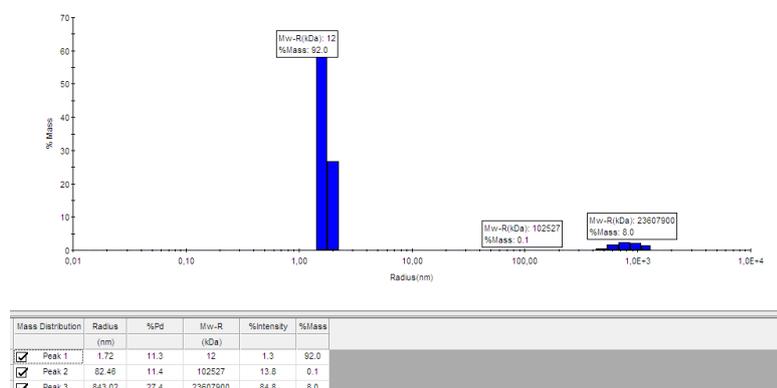
## A. MBD3 2-72

### 1. Purification et caractérisation de MBD3 2-72-histag

Lors de mon arrivée au laboratoire, un premier essai de purification a été réalisé dans des conditions similaires à celles développées pour la protéine MBD3 entière. Pour cela, le culot de 20 litres de culture cellulaire a été lysé dans un tampon de lyse (300 mM de NaCl, 50 mM MES pH6,5, 4 mM de CHAPS, 2 mM BME, PMSF et cOmplete EDTA-free). Puis l'échantillon a été lysé en utilisant du lysozyme et la sonication. Suite à la centrifugation de l'échantillon, le surnageant est incubé avec de la résine Ni-NTA, lavé et élué avec un tampon d'éluion (300 mM NaCl, 50 mM MES pH6,5, 300 mM imidazole, 4 mM de CHAPS et 2 mM de BME). L'éluât obtenu était trouble et une grande partie des protéines été agrégées. Suite à la centrifugation de cette échantillon, il fut injecté sur une colonne HiTrap Héparine 5 mL, et élué par un gradient linéaire en sel de 300 à 1 M NaCl). Les fractions enrichies en MBD3 2-72 se caractérisées par une concentration entre 600 et 700 mM de NaCl. Après concentration en amicon, l'échantillon était de nouveau trouble et suite à sa concentration il fut injecté sur deux colonnes Superdex 200 10/300 GL montées en séries (150 mM NaCl, 50 mM MES pH6.5, 5 mM EDTA, 1mM de CHAPS et 500  $\mu$ M de TCEP). Le profil de GF n'a montré qu'une espèce éluée à 35 mL comme une protéine d'environ 8 à 11 kDa, en accord avec la taille attendue, le domaine MBD étant de 10,3 kDa. La migration sur gel d'un échantillon du pic d'éluion montre clairement que le MBD 3 est très pur (Figure 67 C)



**Figure 67 Purification du domaine MBD de MBD3 par chromatographie d'affinité puis d'exclusion stérique.** (A) Chromatogramme d'élution de la chromatographie d'affinité héparine. Le pic d'élution du domaine MBD est annoté 1. En vert sont représentées les concentrations croissantes en NaCl en mM. Le MBD est élué à 600 mM. (B) Chromatogramme d'élution de la chromatographie d'exclusion stérique. Le pic d'élution du MBD est annoté 2 et se situe à 35 mL. (C) Suivi de la purification du domaine MBD sur gel SDS-PAGE-17%.



**Figure 68 Mesures de DLS réalisées sur le MBD purifié.** Le poids moléculaire relatif est estimé de 12 kDa correspondant à un monomère de MBD avec sa couche d'hydratation. Sur le schéma est représenté le pourcentage de masse en fonction du rayon hydrodynamique exprimé en nm.

La migration sur gel en conditions dénaturantes a permis de vérifier la pureté de l'échantillon après les différentes étapes de chromatographies, l'homogénéité de l'échantillon a été aussi vérifiée par diffusion dynamique de la lumière (DLS). La moyenne des paramètres mesurés sur les dix acquisitions réalisées montre un rayon hydrodynamique de 1,72 nm pour une masse moléculaire relative de 12 kDa, ce qui corrobore l'état monomérique du MBD du MBD3 2-72. De plus le faible pourcentage de polydispersité évalué à 1,3% indique que le MBD purifié est homogène en solution (Figure 68). On observe également un troisième pic correspondant à des agrégats.

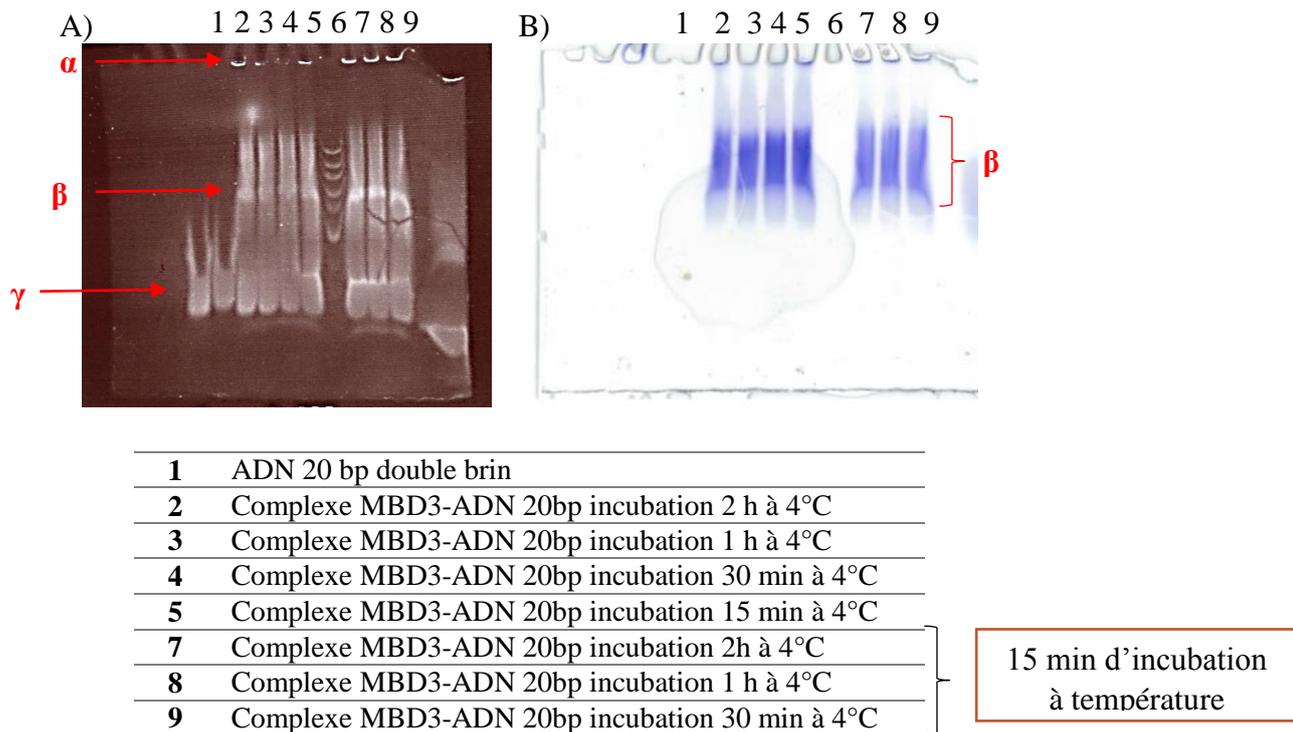
## 2. Caractérisation de l'interaction MBD-ADN

Afin d'étudier les spécificités d'interaction de MBD3 2-72 avec l'ADN, des tests d'ITC ont été réalisés par le Docteur Morgan Torchy avec différents ADN désignés en se basant sur le travail de K. Günther, qui avait identifié en 2013, par des expériences de ChIP-seq les régions de l'ADN où l'on observe des interactions préférentielles de MBD3. Parmi ces régions on peut noter les promoteurs de gène de ZSCAN22, MED23, AURKAIP1, RPL37 et CRABP1 (Tableau 7). Les oligonucléotides de 11 bp désignés ont été utilisés pour réaliser les expériences d'ITC et de retard sur gel. Néanmoins, aucune interaction entre le domaine MBD et ces oligonucléotides ne furent mise en avant.

Nom de l'oligonucléotide	Séquence 5'-3'
ZSCAN22/1	CAGGCGGAGC
ZSCAN22/2	GAGACGAAGT
MED23/1	GACACCGTCTC
MED23/2	GGTGTGGCTG
AURKAIP1/1	GGTCTCGAACT
AURKAIP1/2	ACCTGCGGGTA
RPL37/1	ATCCTCGTCCT
RPL37/2	AGCCCCGCATT
CRABP1/1	CCTCACGCTTT
CRABP1/2	CCTTGGAGCT
20bp	GCCTGATCGATCGATCGCTA

**Tableau 7 Oligonucléotides**

Suite à cette observation de nouveaux oligonucléotides ont été conçus et utilisé présentant une alternance de CG et AT pour maximiser l'interaction du domaine MBD avec l'ADN. Des complexes ont été réalisés avec un ratio 1:1,5, et vérifié par des expériences de retard sur gel, après une migration de 30 min à 2W, un décalage est observé sur le gel caractéristique de la présence de complexe mais également un excès d'ADN (Figure 69).

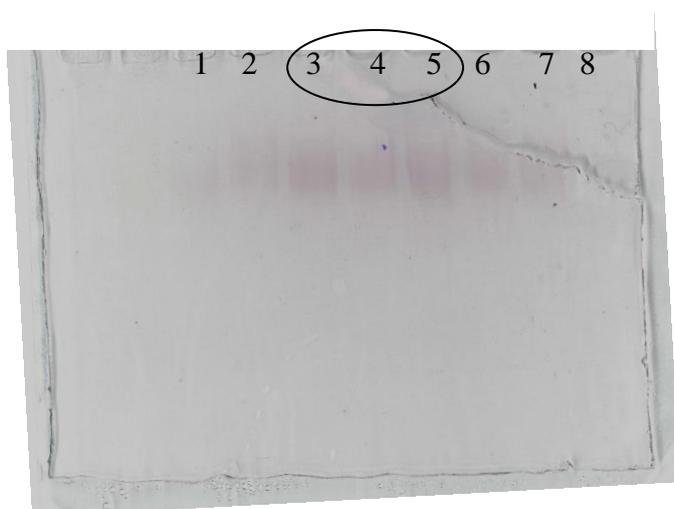


**Figure 69 Migration sur gel des complexes MBD3 2-72 -ADN 20bp en conditions natives.** Profil de migration sur gel natif-PAGE-8%, en (A) le gel est révélé au UV après incubation au BET, on identifie ici l'ADN. (B) le gel est coloré au bleu de coomassie permettant d'identifier les protéines.

Les expériences de gel natif ici ont permis de mettre en évidence les conditions de stabilité et de formation du complexe MBD-ADN optimales. Les gels réalisés sont toujours colorés en deux temps, une première révélation sous UV avec du BET pour identifier la présence d'ADN et une seconde coloration au bleu de coomassie pour mettre en évidence la coprésence d'ADN et protéines. En effet, une simple coloration au bleu de coomassie n'était pas la méthode la plus adaptée. Le MBD seul présente un pI théorique de 9.8 quasiment égal au pH du gel (ici TBE à pH 8.3), le contrôle de la protéine seule ne pénétrera donc pas dans le gel et aucun décalage de taille ne peut être observé. En présence d'ADN en revanche, le complexe migre

dans le gel. En **Figure 69**, différentes conditions d'incubation ont été testées, on remarque que l'incubation du complexe au préalable 15 min à température ambiante n'entraîne aucun changement. De plus 15 min sur glace est suffisant à la formation des complexes. Ce temps d'incubation sera celui utilisé tout au long de mes études d'interaction entre MBD3 2-72 et l'ADN. Le profil de migration sur gel des complexes présente néanmoins une particularité, on n'observe pas comme attendu une bande bien distincte mais une bande étalée communément caractérisée par le terme « smear ». Cet aspect est spécifique des interactions protéine-ADN ayant tendance à se dissocier au cours de la migration dû à une faible interaction ou à des problèmes de temps de migration ou de puissance de migration trop importante.

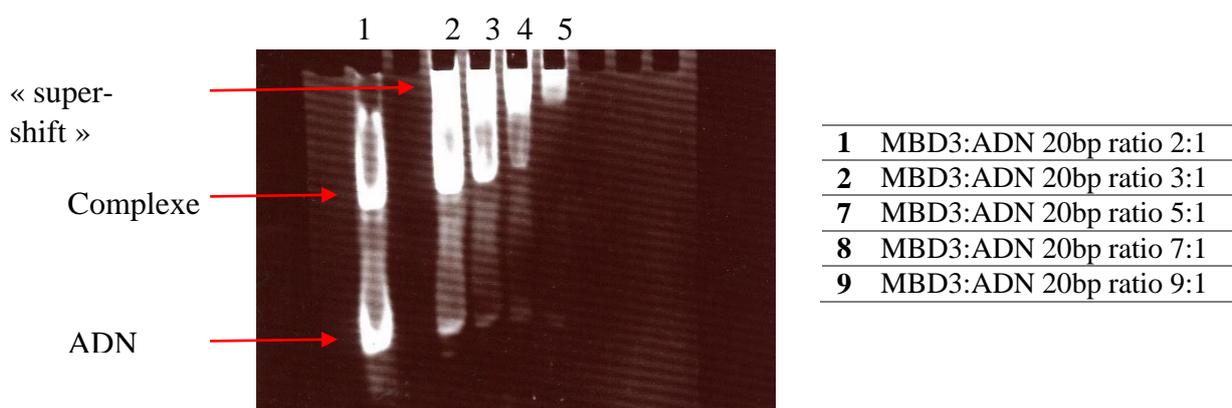
Les études structurales de complexes protéo-nucléotidiques, nécessite la formation de complexes stables en vue de leur cristallisation, et requiert de nombreuses optimisations dans leur préparation. Ainsi, différentes études de retard sur gel pour le complexe MBD3 2-72:DNA 20bp ont été réalisées dans le but d'identifier les conditions de stabilité du complexe. Différentes concentrations en NaCl ont pu être testées. Ainsi, des concentrations élevées en sel vont être néfastes à la formation du complexe, elles peuvent masquer les charges à la surface de la protéine nécessaires à la liaison de l'ADN et conduire à la non-formation du complexe. Néanmoins une concentration trop faible en sel va induire la précipitation du complexe. Ainsi, pour suivre l'évolution de la stabilité du complexe, des complexes MBD3 2-72:ADN ont été réalisés dans des concentrations croissantes en NaCl allant de 50 à 250 mM (Figure 70). Sur le gel on observe la présence de complexes pour les différentes concentrations en NaCl, avec une bande plus importante pour les concentrations comprises entre 100 et 150 mM.



<b>1</b>	50 m M NaCl
<b>2</b>	75 mM NaCl
<b>3</b>	100 mM NaCl
<b>4</b>	125 mM NaCl
<b>5</b>	150 mM NaCl
<b>6</b>	175 mM NaCl
<b>7</b>	200 mM NaCl
<b>8</b>	250 mM NaCl

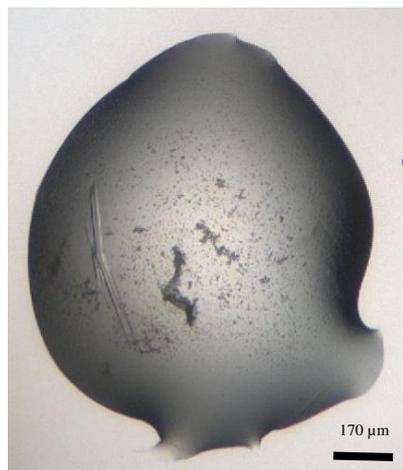
**Figure 70 Migration sur gel en condition natives d'un complexe MBD3 2-72:DNA en fonction de concentration de NaCl.**

Des expériences similaires ont été également réalisées pour identifier les ratios molaires adéquats. De façon générale, on utilise un léger excès d'acide nucléiques, partant de l'hypothèse que le rendement de l'hybridation de l'ADN n'est pas de 100% et donc que la concentration en ADN double brin n'est qu'approximative. De plus un excès d'ADN ne représente pas un problème majeur lors des tests de cristallisations car leur capacité à cristalliser seul reste relativement faible comparé aux protéines. Une gamme de ratios MBD3 2-72:DNA 20bp a été ainsi testées (présentées en figure 71). On observe qu'une augmentation molaire en MBD entraîne un super décalage sur le gel natif. Ces complexes se présentent dès lors sous forme de deux bandes sur le gel, témoignant de la présence de différents complexes. Il s'agit ici d'une saturation de l'ADN en MBD3 2-72. En effet, l'oligonucléotide utilisé possède trois sites de liaison à l'ADN permettant la liaison d'au moins trois MBD par oligonucléotide.



**Figure 71 Migration sur gel en condition natives d'un complexe MBD3 2-72:DNA à différents ratios.**

Cette observation m'a amené à reconsidérer les oligonucléotides utilisés lors de cette étude. En effet, en cristallographie il est important de travailler avec un échantillon pur et le plus homogène possible. La formation de différents complexes dus aux différents sites de liaison pourrait rendre la cristallisation impossible. Néanmoins, d'un point de vu thermodynamique lors de l'utilisation d'un ratio de 1:1, le complexe le plus favorable devrait se former et ne pas entraver la cristallisation. C'est dans cette optique que des tests de cristallisations ont été réalisés avec deux échantillons de ratio MBD3 2-72:DNA 1:1.1 et 9:1.



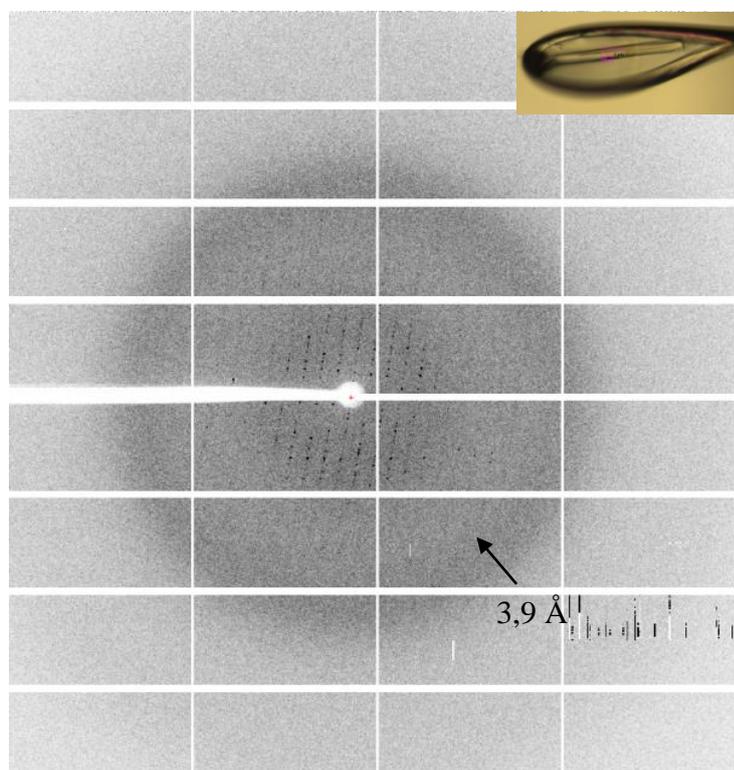
<b>Complexe et ratio</b>	MBD3 2-72:ADN(20bp) 1:1.1 Séquence ADN : GCCTGATCGATCGATCGCTA	MBD3:2-72ADN(20bp) 1:1.1 Séquence ADN : GCCTGATCGATCGATCGCTA
<b>Screen et conditions</b>	<b>Nucleix C3 :</b> 200 mM ammonium acetate 10 mM Mg acétate 50 mM sodium cacodylate pH6,5 30% PEG 8K	<b>Classics F11:</b> 100 mM sodium cacodylate pH 6.5 200 mM sodium acétate 30% PEG 8K
<b>Apparition</b>	J+9	J+19
<b>Cryoprotectant</b>	200 mM ammonium acétate 10 mM Mg acétate 50 mM sodium cacodylate pH6,5 30% PEG 8K 25% PEG 2K	100 mM sodium cacodylate pH 6.5 200 mM sodium acétate 30% PEG 8K 25% PEG 2K
<b>Résolution</b>	10 Å	3,9 Å

**Tableau 8 Cristaux du complexe MBD3 2-72:ADN**

Considérant les conditions optimisées de stabilité du complexe MBD3 2-72:DNA, des tests de cristallisation avec le robot mosquito ont été menés. Le complexe est concentré à 4 mg/mL et des gouttes 100 nL+100 nL sont choisies de manière à pouvoir tester plus de screens commerciaux vu les faibles rendements de la purification. Les screens commerciaux Nucleix, PACT, PEGs et Classics, disponibles sur la plateforme de biologie et génomique structurale de l'IGBMC, ont été choisis.

Des cristaux de complexes MBD3 2-72:DNA 20bp sont apparus après 10 et 20 jours d'incubation dans deux conditions différentes. Aucun cristal n'est par ailleurs identifié pour les gouttes contrôles de ces conditions pour lesquels seule la protéine est présente. Les cristaux sont présentés dans le tableau 8, avec les conditions de cristallisations et de cryoprotectant utilisé. Ces cristaux ont été congelés en utilisant des solutions cryoprotectantes identiques à celles de cristallisation respectives, supplémentées de 25% v/v final de PEG 2000.

L'analyse aux rayons-X du cristal I (Tableau 8), n'a pas permis d'obtenir une diffraction allant au-delà de 10 angströms de résolution mais a confirmé son contenu protéique. Le second cristal testé, dont un cliché de diffraction est présenté en figure 72, a diffracté jusqu'à environ 4 Å.



### Figure 72 Cristal monté et cliché de diffraction à 4 Å de résolution.

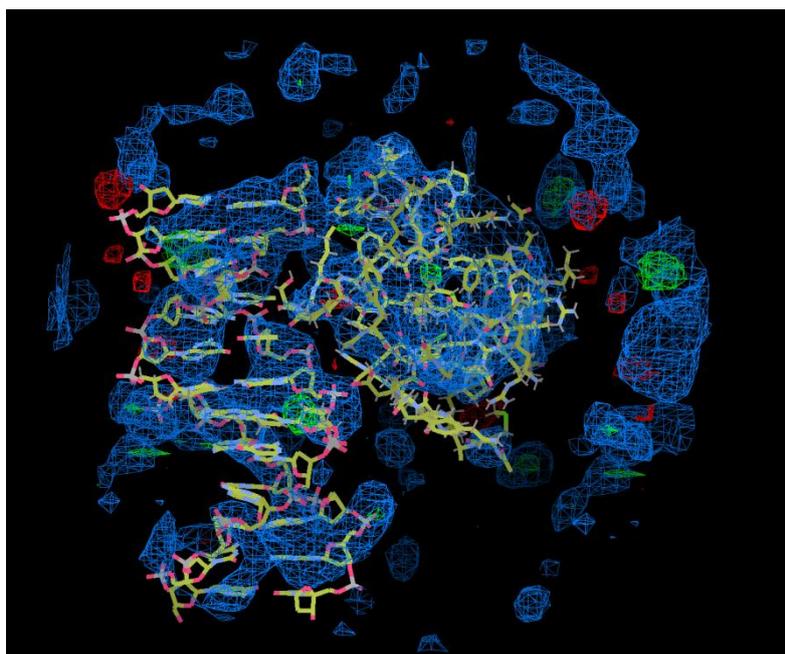
Le jeu de données collecté a été traité avec le logiciel XDS, permettant après indexation l'obtention des paramètres de maille. Les statistiques de collecte de données sont présentées dans le **Tableau 9**, elles ont été obtenues suite à l'intégration des données obtenues avec XDS sur le serveur STARANISO.

#### Collecte de données

<b>Résolution</b>	48,42 Å – 3,9 Å (4 Å – 3,9 Å)
<b>Groupe d'espace</b>	C <sub>1</sub> 2 <sub>1</sub>
<b>Paramètres de maille</b>	a = 212,61 Å ; b = 44,50 Å ; c = 143,30 Å $\alpha = 90^\circ$ ; $\beta = 114,044^\circ$ ; $\gamma = 90^\circ$
<b>R<sub>meas</sub></b>	0.213 (1.47)
<b>I/<math>\sigma</math></b>	4.43 (0.817)
<b>Nombre de réflexions</b>	39307 (3052)
<b>Complétude</b>	98.3% (99.5%)
<b>Redondance</b>	2.9
<b>CC<sub>1/2</sub></b>	

**Tableau 9** Statistiques de collecte des données de diffraction

Les données calculées ont pu être utilisées pour réaliser différents tests de remplacement moléculaire en utilisant le domaine MBD de MBD3 cristallisé seul, la structure d'un ADN seul ou encore par implémentation des deux structures. Ces tests bien qu'initiaux ont permis de mettre en évidence la présence de l'ADN, on peut remarquer en effet en figure 73 la présence de la double hélice d'ADN localisé dans notre densité électronique. On observe également la présence d'une densité qui semblerait correspondre au domaine MBD3 2-72, mais des étapes d'affinement sont encore nécessaires. En effet, la basse résolution rend le remplacement moléculaire compliqué, à cette résolution seul les structures secondaires telles que les hélices  $\alpha$  peuvent être distinguées et ne nous permettra pas d'obtenir des informations précises sur le type d'interaction réalisés et sur la reconnaissance de MBD3 2-72 pour l'ADN non modifié.



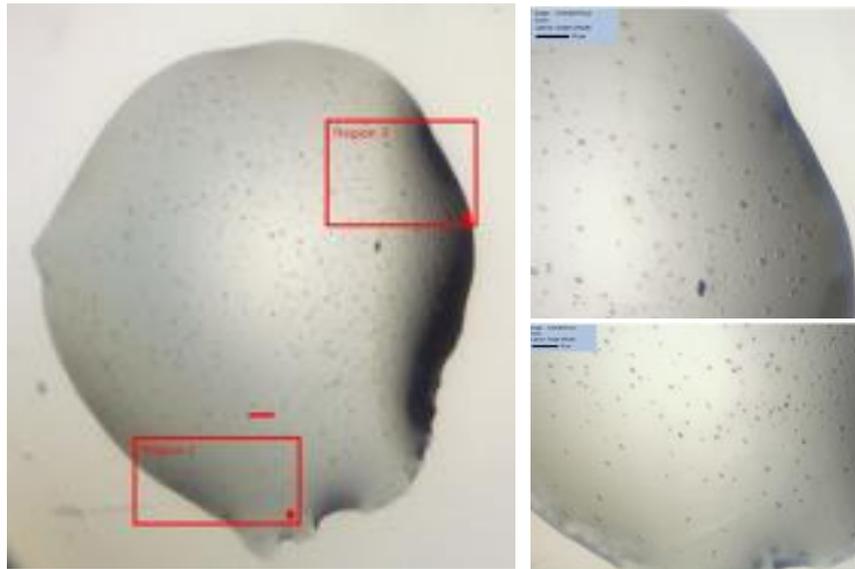
**Figure 73 Densité électronique.** Densité électronique obtenue après remplacement moléculaire avec le logiciel Phenix. On distingue clairement sur la gauche la densité correspondant à l'ADN. Sur la droite la partie qui devrait correspondre à MBD3 2-72 est quant à elle moins bien définie. Des étapes supplémentaires d'affinement seraient nécessaires.

Au vu des difficultés rencontrées lors du traitement des données ainsi que la résolution relativement basse, de nombreux tests de reproduction des cristaux ont été mis en œuvre. Le but étant ici dans un premier temps de reproduire les cristaux mais également d'optimiser les conditions de cristallisation. Pour cela de nombreuses conditions ont été criblées, en faisant varier différents paramètres présentés dans le tableau 10. Ces différents tests ont été de plus réalisés en faisant varier les concentrations en complexes MBD3 2-72:DNA20bp ainsi que la température. Ainsi, le screen présenté dans le tableau fut utilisé pour des concentrations de 2, 4 et 6 mg/mL à 4°C et 20°C.

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	
<b>A</b>	0.075 M	0.15 M	0.175 M	0.2 M	0.225 M	0.25 M	0.1 M NaCaCo pH 6,0
	30% PEG 8000						
<b>B</b>	24%	26%	28%	30%	32%	34%	0.1 M NaCaCo pH 6,0
	0.2 M Na acétate						
<b>C</b>	0.075 M	0.15 M	0.175 M	0.2 M	0.225 M	0.25 M	0.1 M NaCaCo pH 6,5
	30% PEG 8000						
<b>D</b>	24%	26%	28%	30%	32%	34%	0.1 M NaCaCo pH 6,5
	0.2 M Na acétate						
<b>E</b>	0.075 M	0.15 M	0.175 M	0.2 M	0.225 M	0.25 M	0.1 M NaCaCo pH 7
	30% PEG 8000						
<b>F</b>	24%	26%	28%	30%	32%	34%	0.1 M NaCaCo pH 7
	0.2 M Na acétate						

**Tableau 10 Criblage sur la base des conditions de cristallisation mises en évidences**

Il est important ici de noter que la purification présentait à ce stade de nombreux inconvénient du fait du caractère instable de MBD3 2-72 entrainant assez souvent la formation d'un précipité blanc lors de la concentration et la formation du complexe. De ce fait tous les tests réalisés n'étaient que très peu reproductibles et de ce fait aucun cristal de taille suffisante pour leur étude par diffraction aux rayons-X n'a pu être obtenue. En figure 74, on observe des microcristaux obtenus dans la plupart des gouttes.



**Figure 74 Reproduction des cristaux obtenus**

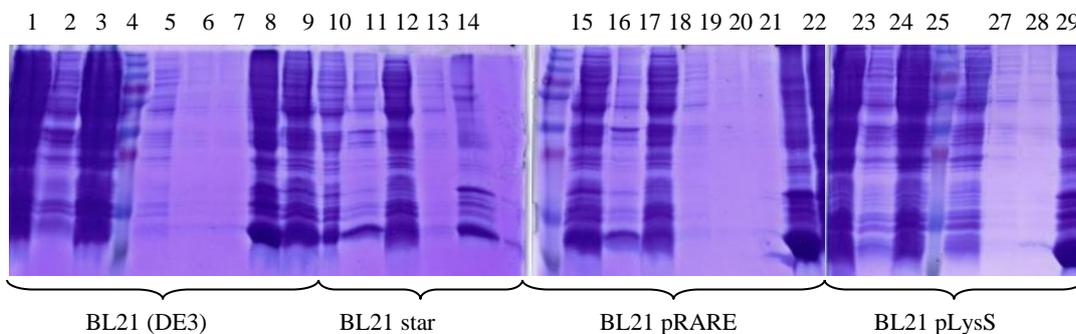
Comme présenté, les problèmes rencontrés lors de la purification ne permettaient pas la reproductibilité des cristaux et de plus ne permettaient pas de tester de nombreuses conditions. En effet, la protéine était sujette à la précipitation et l'agrégation lors des différentes étapes de la purification que ce soit lors de l'élution à l'imidazole lors de la chromatographie d'affinité nickel ou lors de la concentration lors de la concentration par ultrafiltration en amicon ou encore lors de l'ajout de l'ADN pour former les complexes. Tous ces problèmes ajoutés au faible rendement et à l'impossibilité de reproduire les cristaux m'ont poussé à revenir sur le protocole de purification. De plus on a pu voir que du fait de la présence de plusieurs sites de fixation plus d'un MBD3 2-72 pouvait se fixer à l'ADN entraînant potentiellement une hétérogénéité de l'échantillon.

## B. MBD3 2-72 $\Delta$ 6histag

### 1. Purification

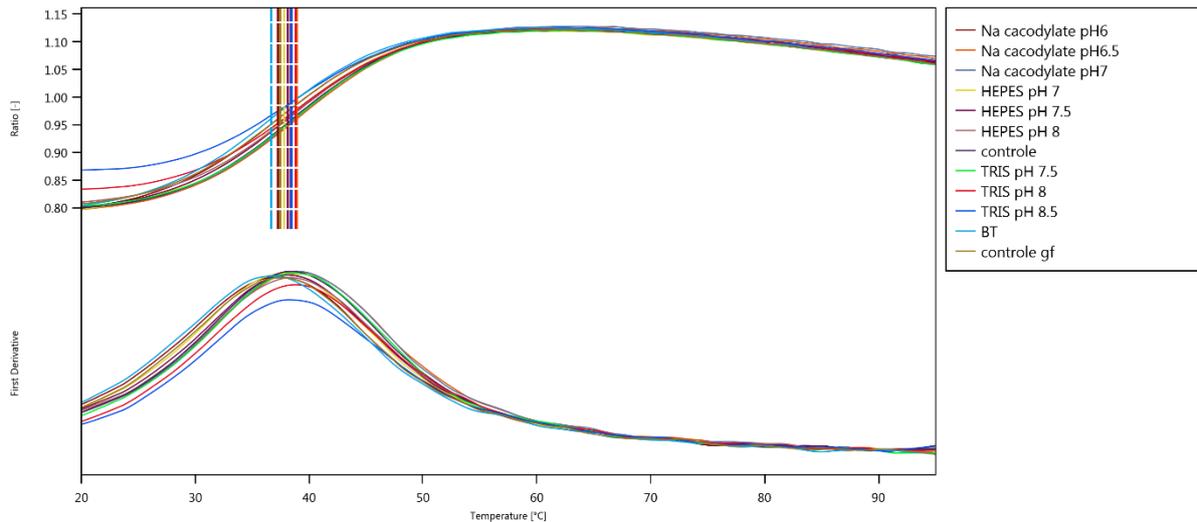
L'optimisation des conditions de purification du domaine MBD de MBD3 se sont réparties sur trois niveau ;

Tout d'abord des tests d'expression ont été réalisés en utilisant différentes souches bactériennes *E. coli*. Comme on peut l'observer en figure 75, bien que le domaine MBD soit surexprimé de façon soluble dans les différentes souches bactériennes, on remarque une légère augmentation de rendement pour la souche pRARE. Cette souche a donc été choisie pour l'expression du domaine MBD 2-72.



**Figure 75 Tests de d'expression de MBD.** Les fractions 9, 14, 22 et 30 correspondent aux billes Ni-NTA.

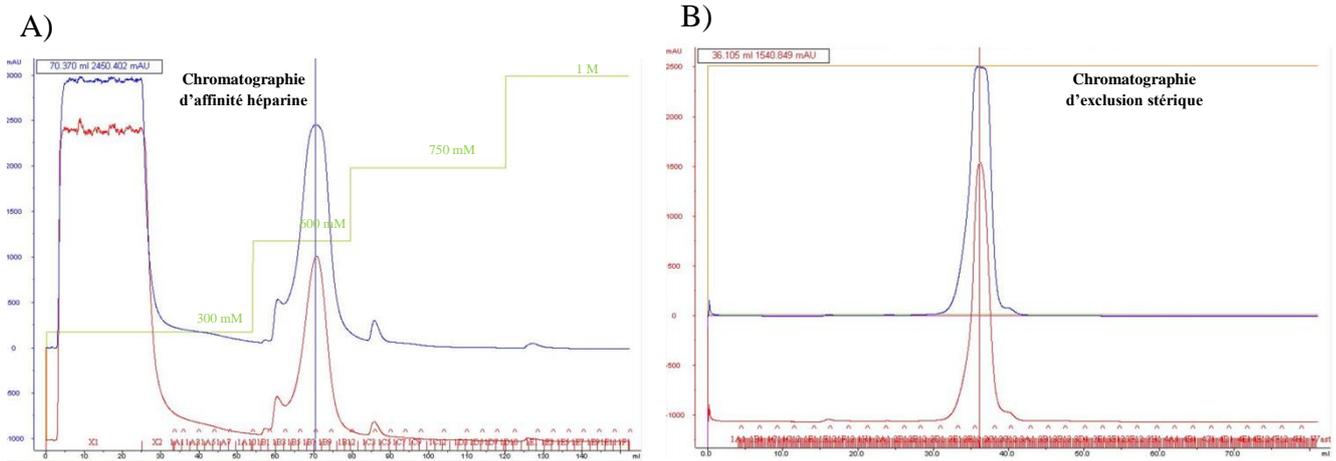
La technique de Thermofluor Assay permet d'estimer la stabilité d'une protéine en suivant la dénaturation des protéines au cours d'un gradient croissant de température, et d'en déterminer un paramètre  $T_m$ , aussi appelé température de demi dénaturation. Selon les conditions de solution tampon, plus le  $T_m$  est élevé, plus la protéine ou le complexe protéique est stable. Un screen de 12 solutions tampons a été mis au point, afin de tester des gammes pH allant respectivement en utilisant le Bis-Tris, le Tris, le MES, l'HEPES et le NaCaCo comme molécules tampon figure 76.



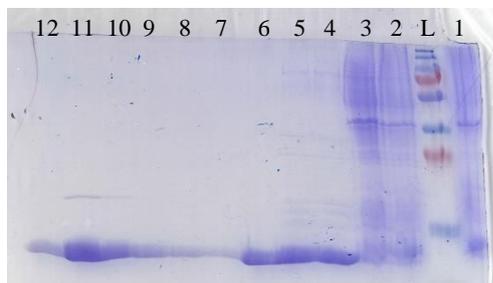
**Figure 76 Courbe de dénaturation du MBD.** Pour cette expérience l'échantillon de MBD3 2-72 purifié est concentré puis dilué dans différents buffers à différents pH entre 6,5 et 8,5.

Les résultats présentés sur la figure montrent le suivi de l'intensité de fluorescence en fonction de la température en degré Celsius (°C). Parmi tous les tampons testés, le Tris-HCl à pH 7,5 et 8 sont ceux pour lesquels les  $T_m$  du MBD sont les plus élevés, suggérant que ce tampon permet une meilleure stabilité du domaine. Ainsi, pour la suite des expériences de purification, le tampon choisi fut le suivant : 20 mM Tris-HCl pH 7,5. Le même type d'expérience avec une gamme de concentration saline a également été réalisé, mais n'aboutissant à aucun changement des concentrations en NaCl au cours de la purification.

Malgré ces changements, le domaine MBD de MBD3 était toujours extrêmement sujet à la précipitation, notamment lors des étapes de concentration en amicon mais également lors de la formation des complexes. Pour prévenir cette agrégation, la première chose fut de remplacer le type de concentrateur. En effet, lors de la concentration j'ai pu observer après lavage des membranes de cellulose au SDS la présence de protéine étudiée. J'ai donc opté pour l'utilisation du système vivaspin présentant une faible surface en contact avec une membrane composée de polyethersulfone caractérisées par une réactivité faible avec les protéines. Ce changement a été fondamental et m'a permis de concentrer le domaine MBD sans perte. En ce qui concerne la précipitation observée lors de l'ajout de l'ADN, de nombreux tests ont permis de mettre en évidence que celui-ci était dû à la présence du His-tag à l'extrémité N-terminale de MBD3. En effet, suite à son clivage par la protéase 3C aucune précipitation ne fut observée.

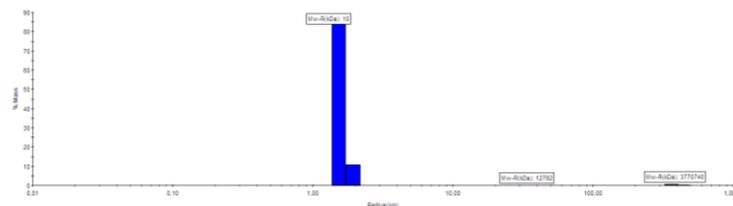


C)



<b>1</b>	Post-sonication
<b>2</b>	Echelle de poids moléculaire
<b>3</b>	Surnageant post-centrifugation
<b>4</b>	Elution nickel
<b>5</b>	Nickel inverse
<b>6</b>	Elution héparine
<b>7-12</b>	Elution chromatographie d'exclusion stérique

D)



Pic	Rayon (nm)	Masse moléculaire (kDa)	%Pd
<b>1</b>	1.6	10	8.1
<b>2</b>	35	12782	0
<b>3</b>	385	3770740	11.9

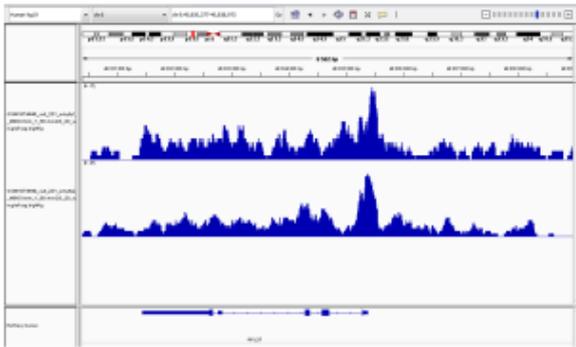
**Figure 77 Purification du domaine MBD de MBD3 2-72  $\Delta$ 6histag par chromatographie d'affinité puis d'exclusion stérique.** (A) Chromatogramme d'élution de la chromatographie d'affinité héparine. Le pic d'élution du domaine MBD est annoté 1. En vert sont représentées les concentrations croissantes en NaCl en mM. Le MBD est élué à 600 mM. (B) Chromatogramme d'élution de la chromatographie d'exclusion stérique. Le pic d'élution du MBD est annoté 2 et se situe à 36 mL. (C) Suivi de la purification du domaine MBD sur gel SDS-PAGE-17% (D) Mesures par DLS réalisées sur le MBD purifié. Le poids moléculaire relatif est de 10 kDa correspondant au domaine monomérique en prenant en compte l'hydratation.

Tous ces changements ont permis de purifier le domaine MBD de MBD3 2-72  $\Delta$ 6histag de façon reproductible et à un rendement entre 5 à 10 fois supérieur à précédemment. La figure 77 représente les différentes étapes de chromatographies et l'analyse par gel SDS et DLS de la pureté et l'homogénéité de la protéine purifiée.

Comme on peut l'observer, la protéine est pure et peut être produite en quantités importantes. De plus, les analyses de DLS mettent en évidence un faible pourcentage de polydispersité évalué à environ 8% indiquant que le domaine purifié est homogène en solution.

A la suite de ces modifications, il était également important de concevoir de nouveaux nucléotides pour la formation de complexes dans le but d'obtenir des échantillons le plus homogène possible. Pour cette conception je me suis inspirée du travail de Shimbo et al., cette étude par ChIPseq a permis de cartographier la présence de MBD3 sur le génome de cellules cancéreuses humaines (Shimbo et al., 2016). Parmi les différentes zones où MBD3 est localisé, j'ai choisi celles retrouvées au niveau de la région promotrice du gène RPL37 illustré en figure 78. Les longueurs en paire de base de ces nouveaux oligonucléotides correspondent à des multiples de tours d'hélices ou demi-tours d'hélices de l'ADN aidant à l'empilement cristallin. Ainsi les oligonucléotides de 11 et 12 bp sont caractéristiques d'un tour d'hélice et 20 bp de deux tours. Il est également intéressant d'essayer différentes extrémités d'ADN de type franches ou cohésives. Ces oligonucléotides sont synthétisés sous forme simple brin puis hybridés. L'hybridation est vérifiée ultérieurement sur gel acrylamide 20% en conditions natives.

A)



C)

GGATGCGGTAC 5CpG4 11bp  
 GGATGCGGTACC 5CpG5 12bp  
 GAGGATGCGGTACCAGGCAG 7CpG11 20bp

B)

**CpG Island Info**

**Position:** chr5:40835242-40835873  
**Genomic Size:** 632  
**Size:** 632  
**CpG count:** 53  
**C count plus G count:** 354  
**Percentage CpG:** 16.8%  
**Percentage C or G:** 56.0%  
**Ratio of observed to expected**  
**CpG:** 1.07

```
>SEQUENCE
CGAGGATGGAACGCGATTACAAACACCACAGTCACAACCTCAC
CATCTCGCTTCTGCGGCCGAGACCAGAAAGACCGGAAGAGAAG
GCACCTCCGCTATATCACGCTCAGAAGCTTCGCCCCAGGGAAG
CACTTCCCTGCGGGGCGCGGTGCAGCGTTCTCGAGCGTCCGTCG
TCGTAGTGTAGGTGGGGTCAAGAGGAGGATGCGGTACCAGGCA
GGGCCGATCGCGTTTCAGTGCCGGAAAGGCGGGCGGGACTCCC
AGTTTTCTGGCGACCGGATGGGCGAGTCTTCGTCCGGCCTGCT
TAGACCCTGAGGGCCGTTTAGATTACAGTTTTGCTTTTCGTGC
TGCTGTTGCTGGTGTACTGCTACTACTGATCGGTTTCGCACG
TTAGCAGTCAGATCTGTTGCTATACCGTCGATTACTGTCTTCT
AGGCACATGGAGAACCTGGAATGCGGGGCTGGCCCTGCCCAGG
GGTACCCGCACCGACTGAGCGTCCCGGCGTGTAATGACCGTG
TGCGCCTGGGATAATCCCTTTTAGGAGAAAGCAACTTCTCCAG
CAGCCTCCTCATTCCGATCCCGGAGGCTTTTGGCCGACGCTGA
CTTTCCCTGCGAGAGGTCGCTAGGAGTACG
```

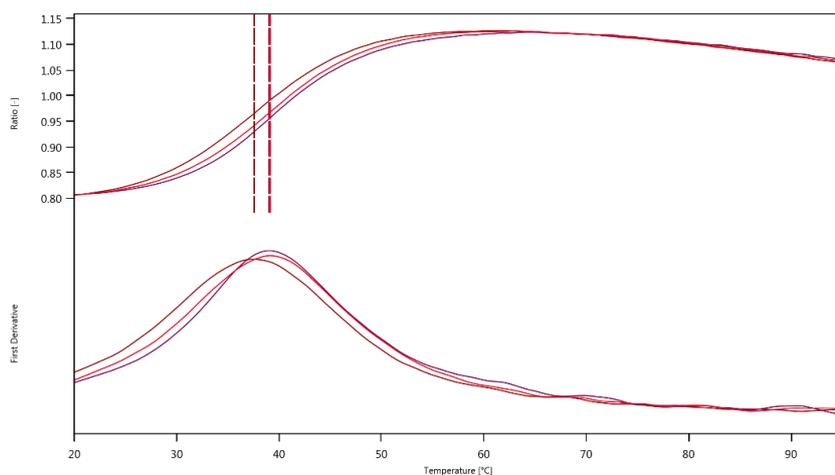
**Figure 78 Nouveaux oligonucléotides.** (A) Exemple des résultats de ChIP séquence localisé au niveau du gène de RPL37. (B) Détail en séquence et en pourcentage de dinucléotides CpG de l'îlot CpG dans la région promotrice du gène RPL37. Sur la séquence à droite est représentée en bleu la zone choisie avec en rouge le CpG. (C) représentation des oligonucléotides utilisés pour les études structurales et biophysiques du complexe MBD3:DNA. En rouge on notera la présence du dinucléotide CpG et en orange la présence de AT.

## 2. Caractérisation du complexe MBD3 2-72 $\Delta$ 6histag

De nouveaux complexes ont été réalisés et vérifiés par migration sur gel en conditions natives avec des gels 8% Tris-glycine. Un décalage de migration fut observé ainsi qu'un léger excès d'ADN. Une étape supplémentaire de purification par chromatographie d'exclusion stérique n'aboutissant qu'à la dilution et la dissociation du complexe dû à la faible stabilité du complexe, comme illustré par la présence de « smear » ou trainées révélés sur les gels en conditions natives. De plus, un léger excès d'ADN ne représente que rarement un problème en soit, du fait qu'il est nécessaire à la formation du complexe mais également du fait de la présence d'oligonucléotides simple brin.

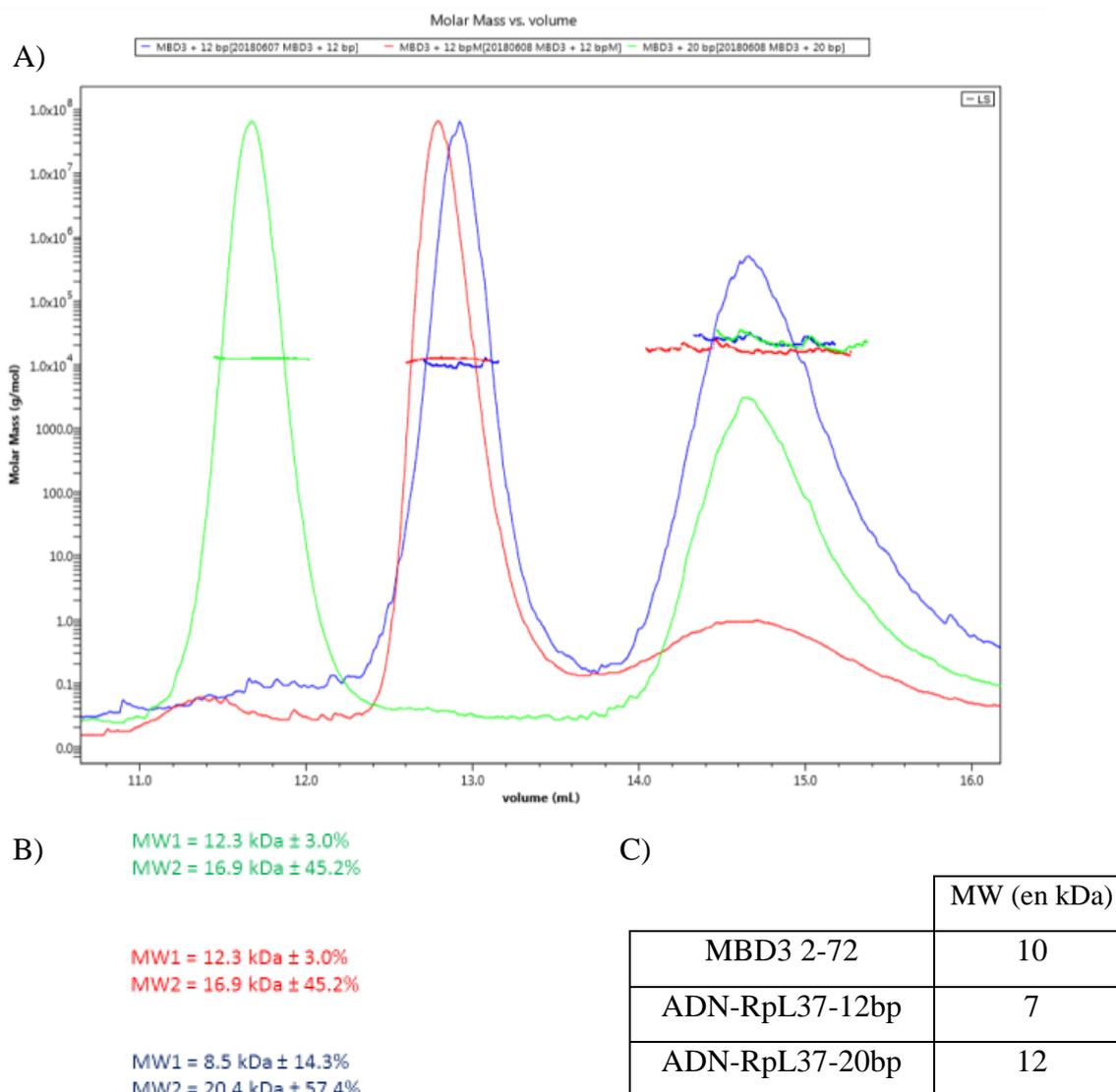
Pour suivre la stabilité du domaine MBD 2-72  $\Delta$ 6histag au sein des complexes avec l'ADN, des expériences de NanoDSF ont été réalisées afin de déterminer la température de demi-dénaturation  $T_m$  du MBD seul et en complexe avec l'oligonucléotide RPL37-11bp (Figure 79). Cette technique de fluorimétrie différentielle permet d'obtenir des informations sur la stabilité des protéines en utilisant la fluorescence intrinsèque du tryptophane et de la tyrosine. L'évolution de la fluorescence est représentée en fonction d'un gradient de température en °C.

Les courbes de dénaturation du MBD 2-72  $\Delta$ 6histag en complexe avec un oligonucléotide de 11 ou 20 bp présentent un décalage vers la droite, caractérisé par une température de dénaturation plus élevée. Cette augmentation significative du  $T_m$ , passant de 37°C à 40°C, suggère une meilleure stabilité et solubilité du MBD 2-72  $\Delta$ 6histag en présence d'ADN.



**Figure 79** Expérience de NanoDSF du MBD 2-72  $\Delta$ 6histag seul et en complexe avec un ADN de 11 ou 20 bp. La dénaturation du MBD (en marron) est comparée à celle du complexe MBD:ADN-RPL37-11bp (violet) et MBD:ADN-RPL37-20bp (rose). Le MBD est plus stable en présence d'ADN.

Pour compléter ces expériences de caractérisation de retard sur gel et de NanoDSF qui corroborent toutes la présence en solution de complexe dans l'échantillon du complexe MBD3 2-72  $\Delta$ 6histag:ADN-RPL37, des expériences de SEC-MALS ont été réalisées. Les résultats sont présentés en figure 80.



**Figure 80** Comparaison par chromatographie d'exclusion par la taille couplée à la détection de la diffusion de lumière laser multi-angles (SEC-MALS). (A) Superposition des trois profils de chromatographie. La courbe verte représente le chromatogramme du complexe MBD3 2-72  $\Delta$ 6histag:ADN-RPL37-20bp et la bleue celle du complexe MBD3 2-72  $\Delta$ 6histag:DNA-RPL-37-12bp. Enfin, la courbe rouge représente les résultats obtenus pour le complexe MBD3 2-72  $\Delta$ 6histag:ADN-RPL37-12bp méthylé. Le chromatogramme représente la masse moléculaire en g/mol en fonction du volume en mL. (B) Les tailles estimées en kDa des différents pics des expériences de SEC-MALS. (C) Tableau des tailles des différents composés utilisés.

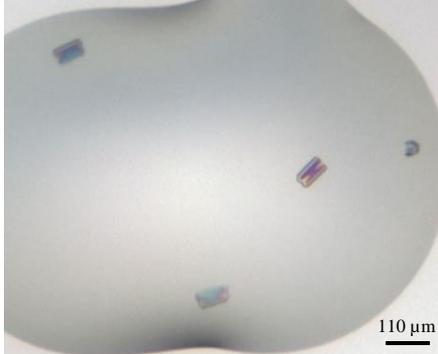
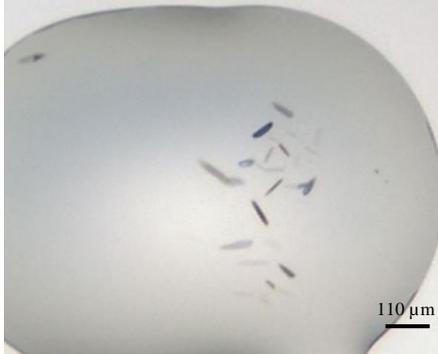
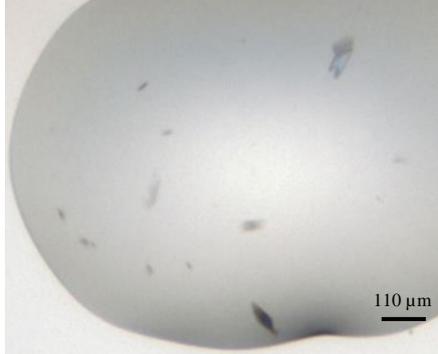
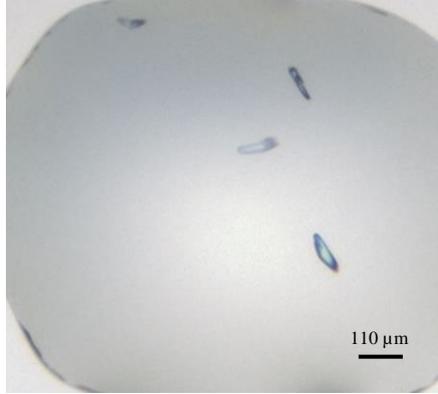
Lors de ces expériences de SEC-MALS, un échantillon supplémentaire a été réalisé ; il s'agit du complexe entre MBD3 2-72  $\Delta$ 6histag et un ADN méthylé. Cet échantillon a été utilisé pour comparer l'interaction du domaine MBD de MBD3 avec un dinucléotide de type 5'mCpG. En effet, au cours de mes travaux de thèse, une étude menée dans le laboratoire de Jinrong Min en 2018, a mise en évidence l'absence d'interaction entre le domaine MBD de MBD3 2-72  $\Delta$ 6histag et l'ADN non modifié (Liu et al., 2018b). Cette étude s'est intéressée à la spécificité de reconnaissance des différents domaines MBD pour l'ADN et plus précisément les dinucléotides CpG et CpA mais également sur l'impact de la base situé en 3' de ces dinucléotides. De façon surprenante, toutes les études réalisées ici ont mis en évidence la formation d'un complexe entre MBD3 2-72  $\Delta$ 6histag et l'ADN non méthylé de type CpG.

Les expériences de SEC-MALS mettent en évidence pour les trois expériences réalisées la présence de deux populations. Pour les trois profils un premier pic correspondant à une masse moléculaire entre 8.5 et 12.3 kDa avec une marge d'erreur, qui correspond au domaine MBD3. Le second pic observé pour les trois profils illustré en figure 80-A, présentent une masse moléculaire estimée entre 17 et 20 kDa avec néanmoins pour les trois une marge d'erreur élevée. Ce deuxième pic est caractéristique de la formation des complexes, et comme on peut l'observer le pic caractéristique du complexe MBD3 2-72  $\Delta$ 6histag avec un ADN non méthylé (en bleu) présente un signal plus élevé que celui du complexe MBD3 2-72  $\Delta$ 6histag avec un ADN méthylé (en rouge). Cette information corrobore encore la formation du complexe avec un dinucléotide de type CpG.

### 3. Cristallisation des complexes oligoprotéiques MBD3 2-72

Suite à ces caractérisations biophysiques du complexe, des tests de cristallisation ont été menées sur la plateforme biologie structurale de l'IGBMC. Les rendements plus important de la purification ont permis de cribler un nombre plus important de conditions en modifiants également de nombreux paramètres tels que la température (20 et 4°C), les concentrations de complexe (2, 4 et 6 mg/mL), les ratios MBD3 2-72  $\Delta$ 6histag:ADN (1:1,1 et 1:2).

Après environ une semaine, des cristaux du complexe MBD3 2-72  $\Delta$ 6histag:ADN-RPL37-11bp sont apparus dans de nombreuses conditions du screen PACT. Les conditions sont relativement similaires, toutes composées d'un tampon Bis-Tris propane à un pH proche du pH physiologique ou

Cristaux	Complexe, ratio et concentration	Screen et conditions	Apparition
	<p>MBD3 2-72  <math>\Delta</math>6histag: RPL37-11bp            1 : 1,1            4 mg/mL</p>	<p>PACT C4            100 mM PCB            pH7            28% PEG 1500</p>	<p>J+6</p>
	<p>MBD3 2-72  <math>\Delta</math>6histag: RPL37-11bp            1 : 1,1            4 mg/mL</p>	<p>PACT F4            100 mM BTP            pH6,5            200 mM thiocyanate de potassium            20% PEG 3350</p>	<p>J+6</p>
	<p>MBD3 2-72  <math>\Delta</math>6histag: RPL37-11bp            1 : 1,1            4 mg/mL</p>	<p>PACT F10            100 mM BTP            pH6,5            200 mM sodium potassium phosphate            20% PEG 3350</p>	<p>J+6</p>
	<p>MBD3 2-72  <math>\Delta</math>6histag: RPL37-11bp            1 : 1,1            4 mg/mL</p>	<p>PACT G5            100 mM BTP            pH7,5            200 mM nitrate de sodium            20% PEG 3350</p>	<p>J+6</p>

**Tableau 11. Cristaux de complexes MBD3 2-72  $\Delta$ 6histag:ADN-RPL37-11bp obtenus avec le screen PACT à 4°C.**

légèrement acide, d'un sel et de PEG 3350. Les sels sont de type sodique tel que le sodium de bromide, de fluoride, d'iodide, de nitrate, de formate.

Une condition différentes de ce même screen commercial est également identifiée, composé de 100 mM de PCB pH7 et 28% de PEG 1500. Les cristaux présentant une taille raisonnable pour leur étude sont identifiés dans le tableau 11.

Il est intéressant de noter ici, aucun cristal ne fut observé dans les gouttes contrôles caractérisé par l'échantillon MBD sans ADN mais également que pour les tests réalisés avec le même ADN de 12 et 20 paires de bases aucun cristal ne fut observé. Cette observation met en exergue l'importance du choix de la longueur des oligonucléotides lors de la cristallisation de complexes de type oligo-protéique.

Une fois la croissance des cristaux achevée, il a été tenté de les récolter et le congeler dans l'azote liquide en vue de leur étude par rayons-X. Il s'est néanmoins avéré que les cristaux, présents dans toutes les gouttes, étaient accolés au support des plaques de cristallisation. Ce problème associé à la consistance interne des cristaux sous forme de gel à rendue leur récolte impossible. Dès lors le défis de ce projet fut de reproduire les cristaux de façon manuelle tout en augmentant la taille des gouttes et optimisant les conditions de cristallisation. Pour pallier au problème d'adhésion des cristaux, la géométrie des tests de cristallisation a été changée passant de cristallisation en goutte assise à goutte suspendue (Figure 81).

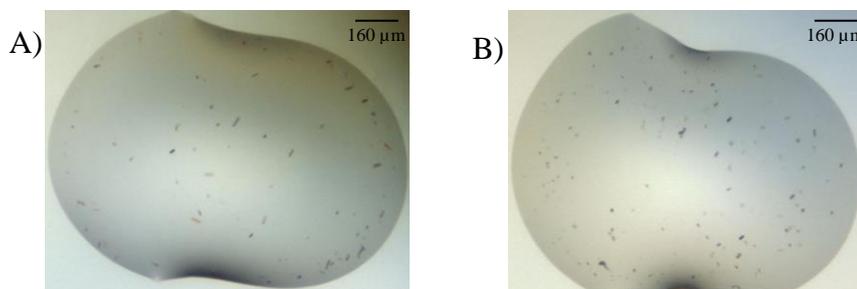


**Figure 81. Cristallisation en goutte assise et goutte suspendue et impact sur la localisation de la croissance cristalline.**

Pour ces tests deux conditions de cristallisations ont été choisie ; il s'agit des conditions C4 et G5. Lors du passage au criblage et optimisations manuelles, il s'est avéré que ces cristaux étaient bien reproductibles dans les mêmes conditions à 4°C et à une concentration de 4 mg/mL. Pour la conditions G5, les tests d'optimisation ont mis en évidence que la concentration en nitrate de sodium (150, 200 et 250 mM testés) n'avait pas un réel impact sur la croissance cristalline, à l'inverse du pH et de la concentration en précipitant. Les conditions criblées sont présentées dans le tableau 13 en annexes. Les conditions optimisées sont les suivantes :

- G5 : 100 mM Bis-Tris propane pH 7,5, 150 mM de nitrate de sodium et 22% de PEG 3350.
- C4 : 100 mM PCB pH7 et 30% PEG 1500.

Malgré leurs tailles plus importantes, les cristaux reproduits n'ont pas pu être récoltés pour leur étude par diffraction aux rayons-X. Deux raisons à cela ; premièrement malgré le changement de géométrie une partie des cristaux se situent toujours au niveau de la plaque de support de la goutte. La deuxième raison vient de la formation d'une peau à l'interface air-eau de la goutte dans laquelle les cristaux se sont développés. Cette peau en surface va isoler la goutte et ralentir et empêcher la diffusion et donc stopper la croissance cristalline, elle est principalement due à la précipitation en surface de protéines. Elle est souvent comme ici le siège d'une croissance cristalline due à la concentration localisée importante de protéines et ici de complexes. Toutes ces observations démontrent l'importance pour la croissance cristalline du complexe d'un support solide. Différentes stratégies ont été développées pour favoriser la cristallisation au sein de la goutte et permettre sa récolte. Ainsi des études de micro-ensemencement ont été réalisées. Pour cette technique, les cristaux obtenus précédemment ont été utilisés pour initier la cristallisation au sein de la goutte. Cette technique a permis de reproduire les cristaux déjà observés, cependant aucune amélioration concernant la croissance sur support solide ne fut observée. En parallèle des expériences de cross-seeding ont pu mettre en évidence au sein du screen commercial JCSG+ de nouvelles conditions de cristallisation présentée en figure 82 néanmoins caractérisées par le même problème récurrent.



**Figure 82. Nouveaux cristaux obtenus par la méthode de cross-seeding et le screen commercial JCSG+.** (A) Condition D6 à J+22 (nucléation J+5). (B) Condition H11 à J+30 (nucléation J+8)

Suite aux essais infructueux de cristallisation en gouttes suspendues, de micro-seeding et cross-seeding, mes recherches m'ont orienté vers deux techniques supplémentaires. La première le micro-batch se caractérise par la formation de gouttes dans un état de sursaturation recouverte d'huile ou d'un mélange d'huile. La présence de cette huile élimine l'interface air-eau et devrait de ce fait éliminer la formation de la peau en surface. Pour ces expériences des gouttes de 1+1  $\mu\text{L}$  ont été réalisées composé du complexe à plus haute concentration (7 mg/mL) et un nouveau screen de cristallisation a été conçu se basant sur le screen de départ et dont les concentrations en précipitant et en sel sont augmentées. Les gouttes sont recouvertes d'une fine couche d'huile à différents ratios paraffine:silicone pour plus ou moins permettre une évaporation. Ces expériences sont néanmoins toujours en cours, mais n'ont encore aboutit à aucun cristal.

Enfin, la dernière stratégie fut la technique du gel d'agarose. Pour cela le même screen de cristallisation que celui présenté dans le **tableau 13** en annexes a été utilisé. La majeure différence étant l'ajout dans la goutte de cristallisation de cristallisation d'un faible pourcentage (W/V) d'agarose LM (low-melting point). Utilisé à faible concentration finale, ici entre 0.05 et 0.3% (W/V), permet la formation d'un réseau au sein de la goutte qui pourra servir de support à la cristallisation. Cette technique a permis d'obtenir des cristaux pour le complexe MBD3 2-72  $\Delta 6$ histag:DNA-RPL37-11bp, qui sont apparus seulement après 5 à 6 jours dans les conditions similaires à celles déjà optimisées en présence de 0,1 et 0,2% (W/V) final d'agarose. Ces cristaux sont néanmoins de petites tailles et requiert encore des étapes d'optimisations. Au vu de la goutte à différentes focalisation, il a cependant été observé que les cristaux se trouvaient dans différents plans au sein du cristal. Cette technique semble être la plus adaptée à la croissance sur support pour le complexe étudié.

## C. MBD3 2-133

### 1. La construction 2-133

Les études de biologie structurale de complexes protéine:ADN présentent, comme on a pu le voir de nombreuses difficultés. Lors de ce chapitre j'ai pu montrer l'importance du choix de l'ADN en termes de séquence et de taille mais également de type d'extrémité. Néanmoins, le choix de la séquence de la protéine est également fondamental à la formation de cristaux du

complexe. En effet, les protéines connues pour interagir avec l'ADN sont très souvent caractérisées par la présence des régions désordonnées ou de boucles flexibles en N ou C-terminal qui rendent parfois impossible la cristallisation de tels complexes. La stratégie de base lors de l'étude de tels complexes est comme il a été présenté ici avec le domaine MBD de MBD3 de se focaliser sur le domaine minimal de liaison à l'ADN ordonné. Il est important de rappeler ici que la protéine MeCP2 est caractérisée en C-terminal à son domaine MBD d'une région désordonnée dans laquelle se retrouvent deux motifs de type RGR nommé crochets AT impliqué dans l'interaction à l'ADN au niveau de régions riches en dinucléotides AT. Une étude comparative de cette séquence désordonnée de MeCP2 avec les régions équivalentes pour MBD2 et MBD3 de différentes espèces a été réalisée avec le logiciel ClustalW. Les résultats sont présentés en Figure 8.



**Figure 83. La région IDR et le crochet AT chez les MBD protéines.** Les séquences des protéines MBD1 à 4 sont alignées ici et centrées au niveau des régions correspondant au crochet AT chez MeCP2. Pour plus de clarté les séquences de crochet AT ont été ajoutées à l'alignement sous le nom AT\_HOOK\_1 ou 2. En rouge sont entourés les motifs RGR caractéristiques du crochet AT. (détail en annexes : figure 89)

D'après l'alignement des séquences des différentes protéines MBD humaines, réalisé sur ClustalW, nous pouvons facilement remarquer que les deux crochets AT de MeCP2 ne sont pas retrouvés chez les autres représentants de la famille. Même si les séquences pour MBD2 et MBD3 équivalent au premier crochet AT de MeCP2 sont assez similaires en termes de famille d'acides aminés (KGKP pour MBD2 et 3 et RGRP pour MeCP2), cette séquence présente pour MBD2 et MBD3 ne peut pas être considérée comme un crochet AT, le motif GRP étant discriminant. Klose *et al.* ont ainsi prouvé que la protéine MeCP2 présentait une meilleure affinité pour l'ADN en présence de région riche en AT en 3' du dinucléotide CpG

mais qu'aucune similarité n'était détectée pour MBD2, suggérant une liaison de MBD2 sur l'ADN indépendante de la séquence adjacente (Klose et al., 2005). Néanmoins, de façon intéressante cette région prédite désordonnée (Figure 66) est conservée chez la plupart des êtres vivants (Figure 84). Cette même région caractérisée chez MBD2 a été montrée comme stabilisante pour la formation de complexe MBD2:ADN suggérant une coopération des deux domaines. En me basant sur ces résultats, il a été décidé d'étudier une deuxième construction de la protéine MBD3 possédant le domaine MBD mais également une partie de cette séquence IDR allant des résidus 2 à 133.

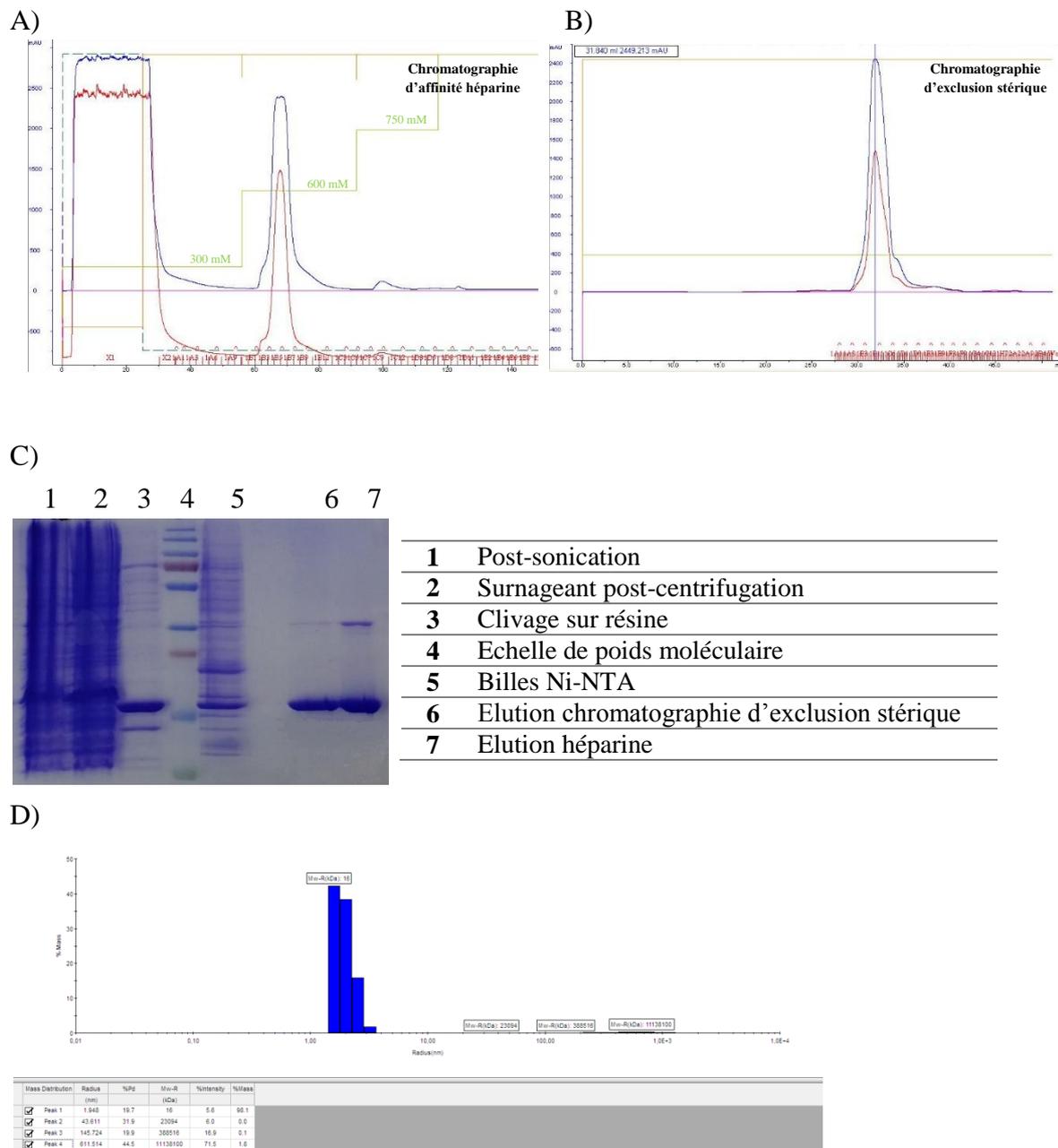


**Figure 84. La région IDR conservée chez MBD3.** La région IDR est délimité par la barre orange avec en amont la région MBD3. (détail en annexes : figure 90)

Pour cette étude la séquence codant pour la construction 2-133 de MBD3 a été cloné dans le vecteur pET28b. Les tests d'expression réalisés pour cette construction ont donné des résultats similaires que ceux réalisés pour MBD3 2-72. De ce fait l'expression a été faite après transformation dans des souches BL21pRARE, dans 6 l de LB2X induit avec 0.4 mM d'IPTG lorsque la DO atteint 0,6-0,8.

## 2. Purification de MBD3 2-133

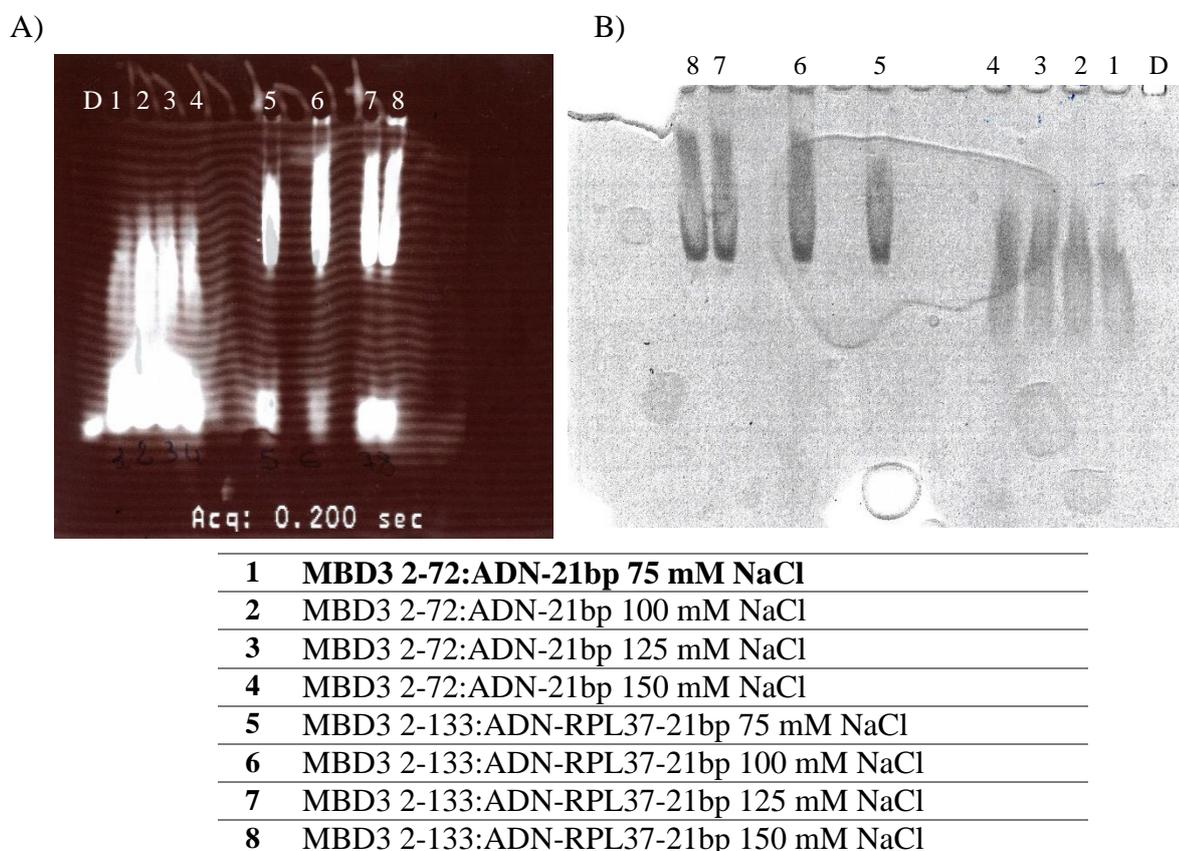
Le domaine MBD3 2-133 a été purifié par chromatographie liquide en trois étapes : une première étape de chromatographie d'affinité sur une résine chélatée au nickel a permis d'éliminer la majeure partie des contaminants protéiques puis une étape de chromatographie d'affinité sur résine héparine pour éliminer l'ADN et enfin une étape de chromatographie d'exclusion stérique a permis d'éliminer les éventuels agrégats et autres contaminants subsistant. Les profils d'élution pour les deux étapes finales de purification sont présentés sur la figure 85 et l'efficacité de chacune des étapes de purification est analysée par migration électrophorétique sur gel PAGE-15% en conditions dénaturantes. Lors de la première étape d'affinité au nickel, le MBD n'est pas élué mais clivé par ajout de la protéase 3C sur la nuit à 4°C lorsque la protéine est toujours en interaction avec la résine Ni-NTA. L'analyse sur gel SDS-PAGE indique que le MBD 2-133 est déjà très pur à l'issue de cette étape. Lors de l'étape de chromatographie d'affinité de type héparine, la protéine est éluee avec un gradient de NaCl, la concentration permettant l'élution est d'environ 600 mM de NaCl. Suite à sa concentration la protéine est injectée sur deux colonnes S200 10/300 en série. La migration sur gel d'un échantillon du pic d'élution montre clairement que le MBD 2-133 est très pur. L'homogénéité de l'échantillon a été vérifiée par DLS (Figure 85). La moyenne des paramètres mesurés sur les dix acquisitions réalisées montre un rayon hydrodynamique moyen de 1,9 nm pour une masse moléculaire relative de 16 kDa, caractéristique du MBD 2-133 sous forme monomérique. Le pourcentage de polydispersité évalué à 19% indique que la nouvelle construction purifiée est homogène en solution.



**Figure 85. Les étapes de purification de la construction MBD3 2-133.** (A) chromatogramme d'élution de la chromatographie d'affinité héparine. (B) chromatogramme d'élution de la chromatographie d'exclusion stérique. (C) Suivi de la purification du domaine MBD 2-133 sur gel dénaturant PAGE-15%. (D) Mesures de DLS réalisées sur le MBD 2-133 purifié qui montrent un profil monodisperse avec peu d'agrégation.

### 3. Complexes oligoprotéiques MBD3 2-133:ADN-RPL37-21bp et caractérisations biophysiques

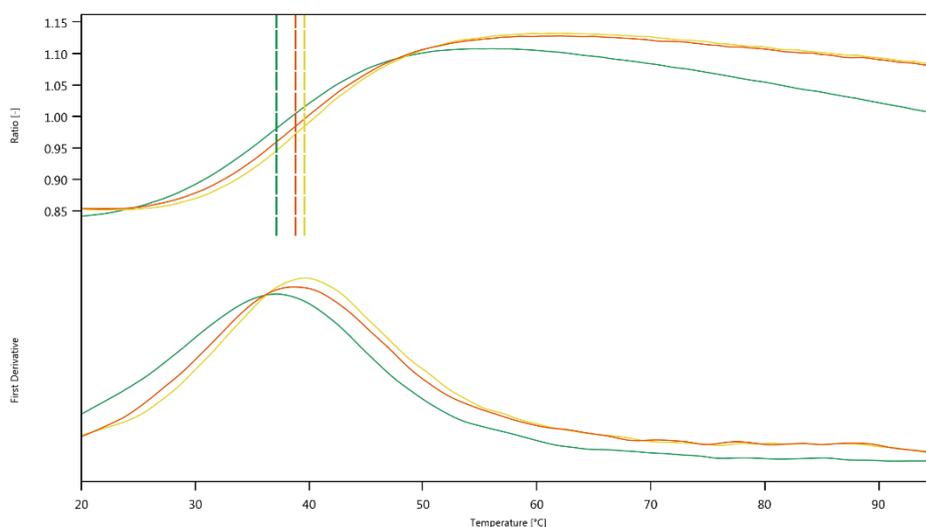
Au vu des résultats de la purification de cette nouvelle construction, des complexes ont été réalisés par ajout d'ADN à différents ratios. Les nouveaux complexes réalisés sont vérifiés par migration sur gel en conditions natives avec des gels 8% Tris-glycine. Cette étude a été réalisée avec l'ADN de RPL-37 21bp (Figure 78) et un gradient de concentration saline allant de 75 à 150 mM de NaCl. Un décalage de migration est observé ainsi qu'un léger excès d'ADN (Figure 86). Cette figure compare ainsi le profil de migration en gel natif entre la construction MBD3 2-72:ADN-21bp et MBD3 2-133:ADN-RPL37-21bp. Comme on peut l'observer quel que soit la construction un décalage de migration est observé pour les complexes ainsi qu'un excès d'ADN. Néanmoins, alors que le complexe avec la construction 2-72 du domaine MBD présente une migration en forme étalée, pour la nouvelle construction, la migration est une bande fine avec un léger étalement.



**Figure 86. Gels PAGE en condition natives.** Profil de migration sur gel natif-PAGE-8%, en (A) gel natif révélation aux UV. (B) Révélation au bleu de coomassie.

Sur le gel coloré en bleu de coomassie on observe plus précisément la bande caractéristique du complexe. Lors de la migration en gel natif, il semblerait que le complexe MBD3 2-133:ADN-RPL37-21bp soit plus stable. La seule différence étant ici les quelques acides aminés de la région IDR ajouté. Ces résultats sont en accord avec le postulat présenté précédemment.

Pour suivre la stabilité de cette nouvelle construction 2-133 $\Delta$ 6histag seul et en complexe avec l'ADN, des expériences de NanoDSF ont été réalisées afin de déterminer le  $T_m$  du MBD seul et en complexe avec l'oligonucléotide RPL37-11bp et 21bp (Figure 87). Les courbes de dénaturation du MBD en complexe avec un oligonucléotide de 11 ou 21 bp présentent un décalage vers la droite, caractérisé par une température de dénaturation plus élevée. Cette augmentation significative du  $T_m$ , passant de 37°C à 40°C, suggère une meilleure stabilité et solubilité du MBD3 2-133 $\Delta$ 6histag en présence d'ADN. Aucune réelle différence significative est à noter avec les résultats obtenus avec MBD3 2-72 $\Delta$ 6histag.



**Figure 87. Expérience de NanoDSF du MBD3 2-133 $\Delta$ 6histag seul et en complexe avec un ADN de 11 ou 21 bp** La dénaturation du MBD (en vert) est comparée à celle du complexe MBD3 2-133 $\Delta$ 6histag:ADN-RPL37-11bp (orange) et MBD3 2-133 $\Delta$ 6histag:ADN-RPL37-21bp (en jaune). Le MBD est plus stable en présence d'ADN.

Il est néanmoins intéressant de noter une légère différence du  $T_m$  entre l'ADN de 11 et 21bp. Cette différence bien que minime pourrait provenir de la stabilisation de la région IDR sur la double hélice d'ADN.

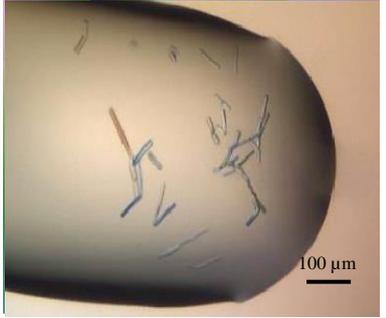
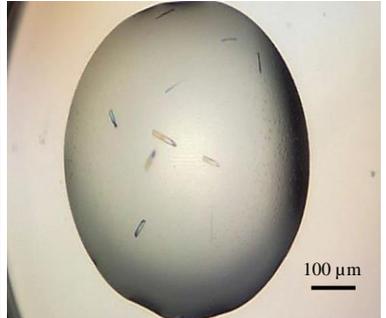
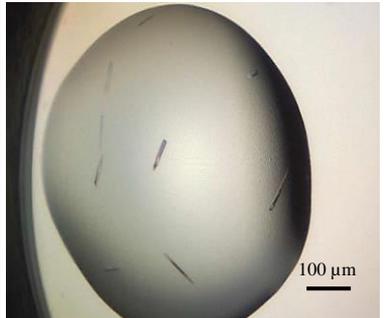
#### 4. Cristallisation

Suite à ces caractérisations biophysiques du complexe, des tests de cristallisation ont encore été menés sur la plateforme biologie structurale de l'IGBMC. Comme précédemment, de nombreux paramètres ont été testés ; la température (20 et 4°C), les concentrations de complexe (2, 4 et 6 mg/mL), les ratios MBD3 2-133Δ6histag:ADN (1:1,1, 1:2).

Après environ quatre jours des cristaux du complexe MBD3 2-133Δ6histag:ADN-RPL37-21bp sont apparus dans de nombreuses conditions du screen Nucleix. Les conditions sont relativement similaires, toutes composées d'un tampon de type sodium cacodylate à un pH 6 ou 6,5. On notera également la présence de sels de type chlorure tel que le  $MgCl_2$  et  $CaCl_2$  en faible concentration (Tableau 12).

Une fois la croissance des cristaux achevée, il s'est avéré que ces cristaux comme ceux des complexes MBD3 2-72Δ6histag:ADN-RPL37-21bp, étaient accolés au support des plaques de cristallisation. Ce problème associé à la consistance interne des cristaux sous forme de gel à rendue leur récolte impossible. De plus, bien que les cristaux soient reproductibles (Tableau 14 en annexes), les conditions de cristallisation sont caractérisées par la présence d'isopropanol. Ce composé volatile rend les cristaux difficiles à manipuler ; en effet, lors de l'ouverture des puits de cristallisation, l'évaporation importante de l'isopropanol rend les cristaux fragiles et cause la dissolution de certains.

Les conditions optimales lors de la reproduction des cristaux semblent être 50 mM sodium cacodylate pH 7, 10 mM  $MgCl_2$ , 12 mM  $CaCl_2$ , 15% isopropanol et 2.5 mM spermine. Le criblage en vue d'optimiser les conditions de cristallisation on montrés qu'une baisse du pH ou une augmentation de l'isopropanol au-dessus de 15% (V/V) augmentaient la vitesse de nucléation mais aboutissaient à des cristaux de tailles inférieur à 20  $\mu m$ . Enfin, toutes les gouttes dans ces conditions qui dérivent de la condition F4 du screen Nucleix sont caractérisées par la présence en surface d'une membrane du très probablement à précipitation du complexe à l'interface eau-air.

Cristaux	Complexe, ratio et concentration	Screen et conditions	Apparition
	MBD3 2-133 $\Delta$ 6histag :RPL37-21bp 1 : 1,1 4,4 mg/mL	Nucleix H12 50 mM sodium cacodylate pH6,5 10 mM MgCl <sub>2</sub> 50 mM spermine	J+4
	MBD3 2-133 $\Delta$ 6histag:RPL37-21bp 1 : 1,1 4,4 mg/mL	Nucleix E4 50 mM sodium cacodylate pH6,5 20 mM MgCl <sub>2</sub> 2.25 mM spermine <b>9% isopropanol</b>	J+5
	MBD3 2-133 $\Delta$ 6histag:RPL37-21bp 1 : 1,1 4,4 mg/mL	Nucleix F10 50 mM sodium cacodylate pH6,5 20 mM CaCl <sub>2</sub> 2.25 mM spermine <b>9% isopropanol</b>	J+5
	MBD3 2-133 $\Delta$ 6histag:RPL37-21bp 1 : 1,1 4,4 mg/mL	Nucleix F4 50 mM sodium cacodylate pH6 10 mM MgCl <sub>2</sub> 10 mM CaCl <sub>2</sub> 2.25 mM spermine <b>10% isopropanol</b>	J+5

**Tableau 12. Cristaux de complexes MBD 2-133  $\Delta$ 6histag:ADN-RPL37-11bp obtenus avec le screen PACT à 4°C.**

### III. Conclusion

Lors des études sur la sous-unité MBD3 du complexe, des interactions avec un ADN non méthylé ont été caractérisées. De plus l'intérêt de la région désordonnée en C-terminale du domaine MBD dans cette interaction a pu être mis en évidence. Ces données qualitatives seront bientôt complétées par des données quantitatives par des méthodes de MST. Les cristaux obtenus récemment pour les deux constructions de la protéine MBD3 en complexe avec de l'ADN sont reproductibles bien que nécessitant encore des étapes d'optimisation. Leur étude par diffraction aux rayons-X devrait permettre l'obtention d'informations sur les bases structurales de l'interaction entre MBD3 et un ADN non-méthylé.

Bien que préliminaires, ces résultats pourront servir de base aux futures études réalisées dans l'équipe.

# Discussion

Le complexe NuRD est un complexe important impliqué dans le processus de remodelage de la chromatine. Comme présenté lors de l'introduction, il a été découvert en 1998 et s'est montré être un complexe singulier au sein de la superfamille des complexes de remodelage ATP dépendant, étant le seul à ce jour à porter deux activités indépendantes avec de effets opposés sur la chromatine : une activité ATP-dépendante et une activité déacétylase.

Avec l'évolution des découvertes sur les fonctions différentes de NuRD dans l'activation et la répression de la transcription, la communauté penche de plus en plus pour décrire NuRD comme une famille de complexes caractérisés par la présence d'un cœur conservé et non plus comme une entité unique à caractère répressif.

C'est dans ce contexte qu'une collaboration entre les équipes du Dr. Ali Hamiche et du Dr. Bruno Klaholz, a été mis en place pour étudier les bases structurales du complexe NuRD acteur majeur de la régulation de la transcription, et plus particulièrement l'étude du complexe endogène entier, purifié à partir de cellules humaines. L'objectif principal du projet fut donc ici, d'une part la caractérisation du complexe et l'optimisation des conditions de stabilité en cryo-EM en vue d'une étude structurale et par ailleurs l'étude de sa sous-unité MBD3 qui constitue l'élément clef des interactions du complexe NuRD avec l'ADN. En effet, malgré les 20 années qui se sont écoulées depuis sa découverte, des questions fondamentales restent toujours sans réponse. Notamment, la stœchiométrie de chaque sous-unité reste encore en partie inconnue mais surtout aucune information sur la structure globale du complexe n'est à ce jour connue.

Le but de mon travail était donc d'obtenir de nouvelles informations structurales sur la topologie du complexe, la stœchiométrie des sous-unités et leurs interactions. Pour cela, mon travail s'est basé sur l'étude de deux type de complexe NuRD ; un complexe soluble retrouvé dans la fraction soluble du noyau et un complexe insoluble en interaction avec le nucléosome. L'intérêt majeur étant ici d'accroître nos connaissances à la fois sur le complexe NuRD mais également de le caractériser dans un contexte fonctionnel en interaction avec un composant de la chromatine, le nucléosome. Les différentes sous-unités sont impliquées dans différents types d'interactions (ADN, modification des histones, protéines-protéines) ; comprendre comment collaborent les sous-unités pour reconnaître la cible du complexe qu'est le nucléosome pourrait permettre d'apporter une nouvelle vision du mode de fonctionnement du

complexe. Il n'y a en effet aujourd'hui aucune structure à haute résolution connue du complexe NuRD. La publication d'une structure à une résolution de 33 Å obtenue par coloration négative par l'équipe de Mackay cette année met en avant la difficulté d'étude de ce complexe (Silva *et al.*, 2018). Le titre même de leur futur article « The Nucleosome Remodeling and Deacetylase complex has an asymmetric, dynamic and modular architecture », attire l'attention sur le côté dynamique et asymétrique du complexe rendant son étude un véritable défi aujourd'hui même avec l'évolution des techniques de biologie structurale et en cryo-EM en particulier.

Au vu des faibles quantités produites de façon endogène, et du caractère flexible et dynamique du complexe NuRD, la méthode de cristallographie aux rayons-X ne semblait pas adaptée. Les réelles avancées des dernières années dans le domaine de la microscopie électronique en font une méthode plus adaptée pour l'étude de complexes macromoléculaire, permettant en même temps de caractériser et optimiser l'échantillon.

Une grande partie de mon travail de thèse a été d'optimiser la purification et la préparation des échantillons de NuRD, les rendant adéquat à une étude en microscopie. Il s'agissait donc de modifier les différents composés de manière à gagner en résolution et contraste tout en ne déstabilisant pas le complexe et permettre le maintien de sa structure tridimensionnelle. Bien que les objectifs soient clairs, le complexe NuRD du fait de son instabilité notamment en conditions cryogéniques a rendu l'obtention de résultats structuraux compliquée, mais le travail réalisé a permis d'établir des conditions de purification et d'observation en cryo-EM.

En effet, j'ai pu étudier l'impact de différents types de tampon sur la stabilité du complexe. Il a été très tôt identifié que le complexe NuRD du fait de sa flexibilité et de son instabilité présentait une forte propension à s'agréger et se dissocier en ses sous-unités lors de son étude en conditions cryogéniques. La modification des conditions dans lesquelles se trouvent les particules, présentent un réel intérêt. Ainsi, les variations en termes de pH vont générer une modification des charges en surface des protéines pouvant influencer sur la solubilité et donc l'agrégation du complexe. Pour cette raison différentes solutions tampon, permettant de modifier la force ionique du tampon, ont été testées, néanmoins aucune réelle amélioration ne fut observée. Le choix c'est donc porté sur l'HEPES à pH 8,0. Deux avantages à ce choix, tout d'abord à l'inverse de tampon de type TRIS, l'HEPES ne créait que peu de bruit de fond et il permit également de mettre en place des études de réticulation. Tout comme le pH, la

composition en sels des échantillons peut impacter sur la stabilisation du complexe en fonction des ions présent en solution. La présence de 150 mM de NaCl semble un bon compromis entre stabilité du complexe et contraste suffisant des images. Ces variations de force ionique n'ont dans aucune des expériences améliorées la répartition des particules.

Au-delà de cet aspect biochimique, de nombreuses autres optimisations plus techniques ont été apportés ; notamment en ce qui concerne la préparation des grilles avec le vitrobot. De façon général, les meilleurs résultats ont été obtenu avec l'utilisation d'un temps d'incubation faible sur la grille et une température d'environ 5°C permettant de limiter les interactions des particules avec l'interface air-tampon à l'origine de la dissociation des particules. Récemment, il a été développé une nouvelle méthode pour le dépôt des échantillons, qui permet d'une part d'éviter l'utilisation d'un filtre automatique entraînant une pression sur l'échantillon, mais également un temps de congélation extrêmement rapide. L'avantage de cette méthode, utilisant un robot Spotiton, système non-commercialisé pour le moment, est principalement dans la conservation de la stabilité des échantillons mais également dans la diminution d'orientations préférentielles. L'utilisation d'une telle méthode permettrait de réduire le temps pendant lequel le complexe NuRD rencontre la surface, cause de dissociation. De plus, le temps réduit entre le dépôt de l'échantillon et sa congélation sur grille limiterait également le déplacement des particules sur le carbone. Il s'agit d'une approche qui pourrait être exploré dans l'avenir.

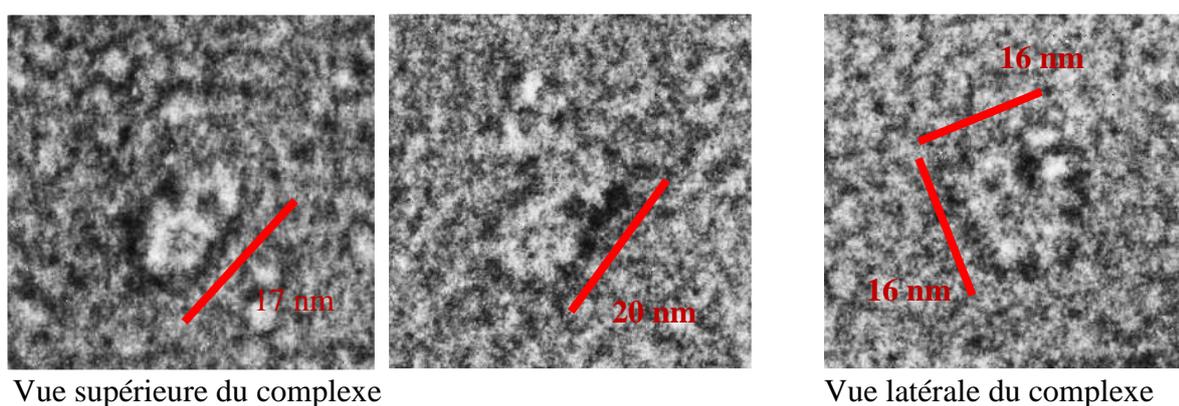
Pour continuer ce projet et dans l'espoir d'obtenir une structure de ce complexe, il serait intéressant d'initier la purification avec des quantités plus importantes de cellules. Lors des études de complexes exprimés dans des cellules humaines et donc exprimés de façon endogènes, il est commun de commencer avec des quantités de l'ordre de la centaine de litres de cellules, ce qui est difficile à réaliser pour cellules HeLa. J'ai donc essayé de compenser la faible quantité de cellules par l'utilisation de différentes techniques. Ainsi l'utilisation d'agents de pontage a été utilisée pour stabiliser le complexe, réduire l'hétérogénéité de l'échantillon et stabiliser les particules. L'idée ici était de stabiliser le complexe de façon à avoir moins de dissociation et donc plus de particules sur les grilles. L'utilisation de ces agents a été couplé à un gradient de type GraDeR, pour diminuer progressivement le détergent, et à l'utilisation de grilles en carbone continue. L'utilisation de ces grilles R2/2 recouverte d'une fine couche de carbone a été une réelle avancée pour l'étude du complexe. Du fait du caractère hydrophile du carbone créé par la décharge luminescence, le complexe ne se retrouvait pas dans les ménisques de glace des grilles, mais s'accumulait sur le support en

carbone fin. L'utilisation du carbone a permis de potentialiser la présence des complexes au niveau des trous recouverts de carbone.

Au final, de nombreuses optimisations ont été nécessaires pour stabiliser l'échantillon lors de son observation en cryo-EM (expériences menées sur le microscope Polara). Bien que des complexes soient observés, les optimisations ne sont pas encore totalement adéquates à la collecte de données sur le microscope à haute résolution Titan Krios et pour le traitement des données comme en atteste les données tests collectées. Différents problèmes ont pu être détectés lors de la collecte de données, tout d'abord l'utilisation d'une couche de carbone rendait difficile la distinction des particules. Pour pallier ce problème il serait intéressant d'une part d'utiliser des couches de carbone plus fines sur les grilles pour réduire au maximum le bruit généré. D'autre part, il pourrait être intéressant de tester d'autres types de support comme l'oxyde de graphène. L'oxyde de graphène est un dérivé hydrophile du graphène sur lequel les macromolécules biologiques se fixent facilement, avec des propriétés supérieures à celles des films de carbone amorphe couramment utilisés en microscopie électronique. Le réseau cristallin monocouche de carbone est pour l'oxyde de graphène hautement transparent aux électrons et présente une conductivité supérieure à celle du carbone amorphe. Il va donc réduire les effets de charge et de distorsions d'images. Ainsi, l'oxyde de graphène est un substrat particulièrement prometteur pour l'examen de matériaux biologiques par microscopie électronique et pourrait résoudre le problème de contraste observé. De plus, du fait de sa haute conductivité, ce support devrait réduire les effets de charge qui créent des distorsions dans les images (Figure 58-D) (Pantelic et al., 2010).

Au fur et à mesure des échecs et des réussites, lors de ces différents tests, certaines conditions ont pu être caractérisées comme favorables au complexe NuRD avec parfois des différences entre les fractions soluble et insoluble. Ainsi, pour la fraction insoluble, l'utilisation de L-glutaraldéhyde comme agent de pontage semble plus adéquat (Figure 56), alors que pour la fraction soluble, la présence de BS3 a semble-t-il permis de stabiliser le complexe (Figure 58). Néanmoins, différents agents de pontage pourraient encore être testés et utilisés, notamment le formaldéhyde, recommandé dans l'étude de complexes nucléosomiaux. Une autre méthode de pontage est aussi envisagée, elle consisterait à ajouter le composé réticulant lorsque que le complexe est fixé sur la résine. Un simple lavage de la résine permettrait ainsi d'éliminer le surplus. Le complexe serait dès lors stabilisé à l'étape précédant le gradient.

Mon travail de thèse a permis d'établir une partie des conditions de préparations de grilles pour le complexe NuRD, qui cependant ont encore besoin d'être optimisés. Mais la visualisation de complexes en coloration négative ainsi qu'en conditions cryogénique sont une base importante dans la faisabilité de l'étude structurale de NuRD. Ainsi les complexes visualisés, de formes ou variables selon les orientations, présentent une taille d'environ 17 à 22 nm en fonction de l'orientation des particules, qui est une taille cohérente avec le poids moléculaire estimé de 1 MDa de NuRD. Les travaux réalisés ici ont permis de mettre en avant pour la première fois une topologie sous forme de domaines du complexe NuRD. Comme présenté en **Figure 88**, les différentes vues du complexe semblent en effet mettre en évidence, pour la première fois une organisation sous forme de deux domaines mais également d'émettre des hypothèses sur leurs orientations.



**Figure 88. Organisation du complexe NuRD en domaines.**

Bien qu'aucune structure ne fût obtenue pendant ces quatre années d'étude du complexe, le projet a permis d'en établir les bases. Il reste aujourd'hui une quantité importante d'hypothèses à tester. Parmi ces hypothèses, et hormis les avancées technologiques déjà présentées, de nombreux composés pourraient stabiliser un peu plus le complexe sans pour autant réduire le contraste :

En termes de détergents, il serait intéressant de tester la famille des NDSB. Ces composés zwitterionique sont connus pour réduire l'agrégation et intervenir dans la stabilisation du repliement des protéines. Ils sont classés comme composés sulfobetaines de

type non-détergents, possédant donc un groupement sulfobétaique hydrophile ainsi qu'un groupement hydrophobique court ne possédant pas de capacité d'agrégation et donc celle de former des micelles. L'ajout de tels composés lors de la purification du complexe et lors de la formation des grilles pourrait aider au maintien de la solubilité ainsi que de l'intégrité du complexe.

La majeure partie des grilles réalisées contenait environ 1% de glycérol jouait le rôle de stabilisant limitant l'agrégation et évitait une dissociation des sous-unités du complexe. Comme on a pu le voir le contraste des grilles du complexe ne permettaient pas toujours de discerner les particules. Ceci n'est pas dû au glycérol seul, sa concentration étant faible mais son association avec la présence de détergent, du film de carbone continu, de la concentration en sels entraîne un faible contraste par effet cumulatif. Il pourrait être intéressant de remplacer le glycérol par du tréhalose. Ce sucre, comme le saccharose va réduire les dégâts causés par le faisceau d'électrons sur l'échantillon, mais de façon surprenante son impact sur le contraste sera moindre. Ainsi, des études ont montré que 3 à 5% de tréhalose était une quantité suffisante pour réduire les artéfacts de dessèchement et d'évaporation de l'eau qui surviennent couramment lors de la vitrification de l'échantillon tout en conservant un contraste adéquat (De Carlo *et. al.*, 2001) (Chiu *et. al.*, 2011).

Enfin, le complexe NuRD est un complexe enzymatique qui requiert l'intervention de ligands ou cofacteurs. Comme on a pu le voir dans l'introduction, de façon similaire à la structure du complexe HDAC3:SMRT, on retrouve à l'interface HDAC1:SANT une poche extrêmement basique capable d'accommoder un dérivé inositol de type Ins(1,4,5,6)P4. La présence d'IP4 augmente drastiquement l'activité de désacétylase du complexe et est considéré dès lors comme un cofacteur du complexe NuRD. De plus, lors de la cristallisation de ce complexe, la formation des cristaux n'a pu être obtenue qu'en présence de quantités significatives de sulfate d'ammonium. Cette étude a pu mettre en évidence, lors de la résolution structurale, que les ions sulfates venaient se substituer à l'IP4 dans la poche basique et masquaient donc cette poche en réduisant son exposition. L'ajout d'IP4 lors de la purification ou lors de la préparation de l'échantillon, pourrait stabiliser le complexe en masquant cette poche basique et réduire l'hétérogénéité ou les mouvements du complexe (Millard *et al.*, 2013a) (Watson *et al.*, 2016).

Cette discussion ne se voulait pas une liste non exhaustive des conditions à tester pour le complexe NuRD, mais juste mettre en évidence que malgré les difficultés rencontrées lors de cette étude, des pistes restent à être explorées. Ces alternatives peuvent être directement

appliquées à l'échantillon lors de sa préparation, ou être une conséquence directe de nouvelles méthodes développées. La cryo-EM reste une technique qui évolue rapidement actuellement et qui s'adapte aux difficultés rencontrés par les utilisateurs en vue de l'obtention de structures à haute résolution pour des particules plus complexes et dynamiques, non symétriques, et relativement fragile.

Parmi les sous-unités du complexe NuRD, la sous-unité MBD3 est particulièrement intéressante. En effet, ce membre unique de la famille des MBPs est caractérisé par deux mutations au sein de son domaine de liaison à l'ADN, le MBD, critique dans la reconnaissance de l'ADN méthylé. Cette mutation n'est pas observée chez les vertébrés inférieurs, suggérant dès lors un rôle redondant des protéines MBD2 et MBD3 chez ces organismes. Des études de localisation génomique par séquençage ChIP ont mise en évidence la localisation de MBD3 au niveau de région riches en CpG non méthylées des îlots CpG, alors que MBD2 est distribué le long du génome de façon méthyle-dépendante (Shimbo *et al.*, 2013) (Günther *et al.*, 2013). Néanmoins, l'étude des spécificités d'interaction de la sous-unité du complexe NuRD bien qu'étudié depuis plus de 15 ans, fait encore aujourd'hui l'objet de nombreuses controverses. Cramer et al en 2014 ont montré par des expériences de RMN que MBD3 présentait une faible préférence pour l'ADN méthylé, de façon néanmoins moins marqué que MBD2 (Cramer *et al.*, 2014). Leur expérience a montré également une absence totale de reconnaissance pour les 5'mCpG oxydés, contrairement à un article aujourd'hui assez controversé sur la capacité de MBD3 à interagir de façon préférentielle avec les dinucléotides 5'hmCpG (Yildirim *et al.*, 2011).

L'objectif de ce projet fut l'étude des bases moléculaires de l'interaction entre la protéine MBD3 et le dinucléotide CpG non modifié présent au niveau des îlots dit CpGs. Durant mes travaux, j'ai pu purifier et caractériser le domaine MBD de MBD3 et mettre en évidence par des caractérisations biophysiques tels que des expériences de retard sur gel, de SEC-MALLS et nanoDSF l'interaction de ce dernier avec des ADN présentant le dinucléotide CpG non méthylé. Des cristaux de ce complexe ont été obtenus mais leur faible résolution n'a pas permis de mettre en évidence les bases moléculaires de cette interaction. Ces résultats préliminaires sont néanmoins en opposition avec un papier paru cette année de Ke Liu *et al.* Leur étude sur les bases structurales des différents domaines MBD couplée à des études d'interactions par ITC, ont mis en évidence une absence d'interaction entre le domaine MBD de MBD3 et un ADN CpG non méthylé. (Liu et al., 2018b). Cette étude semble être en désaccord avec les caractérisations réalisées ici mais également avec les études de localisation

génomique par séquençage CHIP qui ont mis en évidence la localisation de MBD3 au niveau de régions riches en CpG présentant un niveau extrêmement faible de modifications au niveau des cytosines (Shimbo *et al.*, 2013) (Günther *et al.*, 2013). Ces données soutiennent le modèle selon lequel MBD3 chez les mammifères ne reconnaît pas la méthylation ou l'hydroxyméthylation de l'ADN (Baubec *et al.*, 2013). Les deux mutations décrites dans le domaine MBD de MBD3, K30H et Y34F, deux résidus impliqués dans la stabilisation du complexe de MBD2 avec l'ADN de type mCpG, expliquent la différence de spécificité pour l'ADN entre MBD3 et les autres membres de la famille. Néanmoins, elles n'expliquent pas la perte d'interaction observée par Ke Liu *et al.*

De nombreux cristaux ont été obtenus durant mon travail de thèse. Ces cristaux bien que reproductibles de taille suffisante à leur étude par diffraction des rayons-X, présentaient de nombreux inconvénients. Ces cristaux étaient en effet élastiques qui présente un frein pour leur récolte, de plus ils nécessitaient un support solide à leur nucléation et leur croissance. Les mêmes résultats ont été observés lors des tests en gouttes suspendues et micro ensemencement mettant en avant l'importance de ce support. Ce problème a été résolu par la technique de croissance cristalline en gel agarose ; ainsi, les conditions de cristallisation identifiées ont permis par l'ajout d'une faible concentration de l'ordre de 0,1% (W/V) la formation de cristaux au sein de la goutte en suspension. Aujourd'hui, ces cristaux n'ont pas encore été testés mais devraient être récoltés et congelés dans l'azote liquide en vue de leur étude par diffraction aux rayons-X. Probablement qu'il serait nécessaire d'affiner les conditions afin d'obtenir des cristaux diffractant à haute résolution afin d'identifier clairement les bases impliquées dans cette interaction mais également de réaliser une étude comparative avec la structure publiée cette année du domaine MBD de MBD3 avec un ADN 5'mCpG (Liu *et al.*, 2018).

Mon étude ne s'est pas limitée à l'étude du domaine MBD, je me suis également intéressée à la région adjacente à ce domaine caractérisée de région IDR. Il a en effet été montré en 2015, dans l'équipe de DC. Williams que, malgré son désordre inhérent, cette région présente chez MBD2 une affinité de liaison globale de MBD2 pour l'ADN méthylé. Suite à la comparaison des séquences protéiques, il a été mis en évidence la présence de cette même région adjacente pour MBD3, conservée chez toutes les espèces, mettant en avant l'importance probablement structurale et fonctionnelle de cette dernière. Une nouvelle construction de MBD3 2-133 a donc été clonée et purifiée. Des complexes oligoprotéiques ont été réalisés et caractérisés mettant en avant une stabilité plus importante de ces complexes.

Les cristaux obtenus pour ce complexe présentaient les mêmes besoins que ceux du domaine MBD seul en complexe avec l'ADN à savoir un support solide pour la nucléation et la croissance cristalline. En plus de cela les conditions de cristallisation identifiées sont caractérisées par la présence d'isopropanol hautement volatile entraînant suite à son évaporation lors de l'ouverture des puits une détérioration immédiate des cristaux que l'on n'a pu éviter malgré les diverses essais.

Bien qu'aucun des cristaux présentés ici n'ait pu être testé par diffraction des rayons-X, les perspectives pour ce projet sont multiples. Dans un premier temps il serait intéressant de tester les cristaux du domaine MBD 2-72 en complexe avec l'ADN non méthylé de 11 paires de bases qui sont apparus récemment en présence du réseau d'agarose au sein de la goutte. Néanmoins, au vu des premiers tests il a pu être constaté que ces cristaux étaient extrêmement fragiles et leur constitution en gel nécessiterait des étapes de déshydratation en amont de la cryoprotection.

Pour la construction MBD 2-133 en complexe avec un ADN non méthylé de 21 paires de bases, des expériences en microbatch sont envisagées. En effet, les types d'huiles utilisées dans ces expériences bien que connues pour ne laisser évaporer quasi aucun composé, permettent la diffusion de composés tel que l'isopropanol. Ainsi, cette technique consiste à réaliser des gouttes de cristallisation en batch sous une couche d'huile composée d'un ratio 50:50 de paraffine et silicone équilibré avec un puits composé d'isopropanol qui va saturer la couche d'huile et permettre la cristallisation. Le second avantage ici étant que les cristaux peuvent être récoltés directement en passant au travers de l'huile avec la boucle de montage. Cette huile va pouvoir ainsi servir de cryoprotectant mais également empêchera l'évaporation de l'isopropanol (Mortuza et al., 2004).

Enfin, des expériences de thermophorèse à micro-échelle pourraient être menées pour détecter les interactions des différentes constructions du MBD avec de l'ADN non-méthylé CpG et apporter des informations quantitatives aux données structurales. Ces expériences pourront également mettre en évidence l'impact de la région IDR de MBD3 sur la fixation de la protéine.

Comme on a pu le voir, l'interaction entre MBD3 et l'ADN se réalise principalement par le domaine dit de liaison à l'ADN, MBD, mais les régions adjacentes et la topologie de la protéine entière peuvent impacter sur l'interaction avec l'ADN. C'est dans cette optique que d'autres constructions de MBD3 ont été réalisées et clonées, notamment la construction 2-260

qui présente le domaine de super-hélice « coil-coiled » qui pourrait être co-exprimée avec la sous unité P66 $\alpha$  (également clonée). Ce complexe MBD3-P66 $\alpha$ -ADN pourrait faire l'objet d'une étude structurale par cryo-EM au vue de la taille plus importante du complexe (130 kDa environ) et apporter des informations structurales supplémentaires sur le complexe NuRD.

Dans l'avenir, l'étude du complexe NuRD permettra de resituer le complexe dans le contexte des connaissances sur la chromatine, la régulation de la transcription et les multiples activités catalytiques de NuRD et de reconnaissance à l'ADN.

# Annexes

**Tableau 13 Screen des conditions de cristallisation du complexe MBD3 2-72 Δ6Histag : ADN-RPL37-11bp**

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
<b>PACT C4</b>	19% PEG 1500	21% PEG 1500	23% PEG 1500	25% PEG 1500	27% PEG 1500	29% PEG 1500
	100 mM PCB pH7					
<b>PACT F4</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	26% PEG 3350
	100 mM BTP pH6,5 200 mM thyocyanate de potassium					
<b>PACT F4</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	26% PEG 3350
	100 mM BTP pH6,5 250 mM thyocyanate de potassium					
<b>PACT F4</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	26% PEG 3350
	100 mM BTP pH6,5 150 mM thyocyanate de potassium					
<b>PACT F5</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	26% PEG 3350
	100 mM BTP pH6,5 200 mM Na/K phosphate					
<b>PACT F5</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	26% PEG 3350
	100 mM BTP pH6,5 250 mM Na/K phosphate					
<b>PACT F5</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	26% PEG 3350
	100 mM BTP pH6,5 150 mM Na/K phosphate					

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
<b>PACT F10</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	26% PEG 3350
	100 mM BTP pH6,5 200 mM nitrate de sodium					
<b>PACT F10</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	66% PEG 3350
	100 mM BTP pH6,5 250 mM nitrate de sodium					
<b>PACT F10</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	26% PEG 3350
	100 mM BTP pH6,5 150 mM nitrate de sodium					
<b>PACT G5</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	26% PEG 3350
	100 mM BTP pH7,5 200 mM nitrate de sodium					
<b>PACT G5</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	26% PEG 3350
	100 mM BTP pH7,5 250 mM nitrate de sodium					
<b>PACT G5</b>	16% PEG 3350	18% PEG 3350	20% PEG 3350	22% PEG 3350	24% PEG 3350	26% PEG 3350
	100 mM BTP pH7,5 200 mM nitrate de sodium					

\* En rouge sont indiquées les conditions ayant données des cristaux lors du changement de géométrie en gouttes suspendues.

Tableau 14 Screen des conditions de cristallisation du complexe MBD3 2-133  $\Delta$ 6Histag : ADN-RPL37-21bp

	1	2	3	4	5	6	7	8
A	8% isopropanol	10% isopropanol	12% isopropanol	14% isopropanol	16% isopropanol	18% isopropanol	20% isopropanol	22% isopropanol
	50 mM Sodium cacodylate pH7 10 mM MgCl <sub>2</sub> 10 mM CaCl <sub>2</sub> 2.5 mM Spermine							
B	8 mM MgCl <sub>2</sub>	10 mM MgCl <sub>2</sub>	12 mM MgCl <sub>2</sub>	15 mM MgCl <sub>2</sub>	8 mM MgCl <sub>2</sub>	10 mM MgCl <sub>2</sub>	12 mM MgCl <sub>2</sub>	15 mM MgCl <sub>2</sub>
	10% Isopropanol				15% Isopropanol			
50 mM Sodium cacodylate pH7 10 mM CaCl <sub>2</sub> 2.5 mM Spermine								
C	8 mM CaCl <sub>2</sub>	10 mM CaCl <sub>2</sub>	12 mM CaCl <sub>2</sub>	15 mM CaCl <sub>2</sub>	8 mM CaCl <sub>2</sub>	10 mM CaCl <sub>2</sub>	12 mM CaCl <sub>2</sub>	15 mM CaCl <sub>2</sub>
	10% Isopropanol				15% Isopropanol			
50 mM Sodium cacodylate pH7 2.5 mM Spermine								
D	8% isopropanol	10% isopropanol	12% isopropanol	14% isopropanol	16% isopropanol	18% isopropanol	20% isopropanol	22% isopropanol
	50 mM Sodium cacodylate pH6 10 mM MgCl <sub>2</sub> 10 mM CaCl <sub>2</sub> 2.5 mM Spermine							
E	8 mM MgCl <sub>2</sub>	10 mM MgCl <sub>2</sub>	12 mM MgCl <sub>2</sub>	15 mM MgCl <sub>2</sub>	8 mM MgCl <sub>2</sub>	10 mM MgCl <sub>2</sub>	12 mM MgCl <sub>2</sub>	15 mM MgCl <sub>2</sub>
	10% Isopropanol				15% Isopropanol			
50 mM Sodium cacodylate pH6 10 mM CaCl <sub>2</sub> 2.5 mM Spermine								
F	8 mM CaCl <sub>2</sub>	10 mM CaCl <sub>2</sub>	12 mM CaCl <sub>2</sub>	15 mM CaCl <sub>2</sub>	8 mM CaCl <sub>2</sub>	10 mM CaCl <sub>2</sub>	12 mM CaCl <sub>2</sub>	15 mM CaCl <sub>2</sub>
	10% Isopropanol				15% Isopropanol			
50 mM Sodium cacodylate pH6 10 mM MgCl <sub>2</sub> 2.5 mM Spermine								

\* En rouge sont indiquées les conditions ayant données des cristaux lors du changement de géométrie en gouttes suspendues.

\* En bleu sont indiquées les conditions ayant données des cristaux de petites tailles < 10  $\mu$ m lors du changement de géométrie en gouttes suspendues.

Figure 89. Détail figure 83 page 185

```

MBD4_HUMAN 99 -----HWIKRSESNPLQEPASAGFGATGTECRKSYPCGHERNVVQRLLF
MBD1_HUMAN 24 -----HEDMILD-CPALGPQMKRREVEFRKS
MBD2_HUMAN 24 -----MERKRM-CPALPQMEREEVFRS
AT_HOOK_2 168 -----ESGKRMID-CPALPPQMKKEEVRKS
MECP2 0
AT_HOOK_1 113 -----PAVPEASASPIQRSTIIR---DRGPMYD-DPTLPEGTRKLLQRKS
0
MBD4_HUMAN 243 -----SEL---QESRGLSNFTSTHLLKLEDEGVDDVFRKVRKPKGVKVTILKGIPI
MBD1_HUMAN 140 -----PAP8C---CE-----NGGISF
MBD3_HUMAN 167 -----FNEKFKLSGLNAFDIAEELVK---TMDLPIGLQG
MBD2_HUMAN 311 -----FNEKFKLSGLNAFDIAEELVK---TMDLPIGLQG
AT_HOOK_2 0
MECP2 234 -----VLEKSPGKLLKMKHPQTS
AT_HOOK_1 13 -----P-GGIA

MBD4_HUMAN 331 -----SLSSGNIKFCSEQVTSG
MBD3_HUMAN 235 -----FKCERRRCLIVERIRGCGVCRGCTQE
MBD2_HUMAN 258 -----RLEELMADMLAHVE---ELARDG
AT_HOOK_2 399 -----KLEELMADILSRAA---DTE
MECP2 16 -----AEADPQATPKRGRKIP
AT_HOOK_1 334 -----QHWIKRPRKRYKAEADPQATPKRGRKIPSYVAIAAAAEAKKAVKESSTRSVQETVPIKRRKTRVTSIEWEVVIRLLVSTLGEIS
13 -----GK

```





# Publications et conférences

Orlov I, Myasnikov AG, Andronov L, Natchiar SK, Khatter H, Beinsteiner B, Ménétret JF, Hazemann I, Mohideen K, Tazibt K, **Tabaroni R**, Kratzat H, Djabeur N, Bruxelles T, Raivoniaina F, Pompeo LD, Torchy M, Billas I, Urzhumtsev A, Klaholz BP., 2017. The integrative role of cryo electron microscopy in molecular and cellular structural biology. *Biol Cell*. Feb;109(2):81-93.

Instruct Biennial Structural Biology Meeting - Brno, Czech Republic – May 2017 - Poster  
**Tabaroni R.**, Hamiche A., Klaholz BP., Structural and functional studies of the MBD3 subunit of NuRD



# The integrative role of cryo electron microscopy in molecular and cellular structural biology

Igor Orlov\*†‡§, Alexander G. Myasnikov\*†‡§, Leonid Andronov\*†‡§, S. Kundhavai Natchiar\*†‡§, Heena Khatter\*†‡§<sup>2</sup>, Brice Beinstener\*†‡§, Jean-François Ménétret\*†‡§, Isabelle Hazemann\*†‡§, Kareem Mohideen\*†‡§, Karima Tazibt\*†‡§, Rachel Tabaroni\*†‡§, Hanna Kratzat\*†‡§, Nadia Djabeur\*†‡§, Tatiana Bruxelles\*†‡§, Finaritra Raivoniaina\*†‡§, Lorenza di Pompeo\*†‡§, Morgan Torchy\*†‡§, Isabelle Billas\*†‡§, Alexandre Urzhumtsev\*†‡§ and Bruno P. Klaholz\*†‡§<sup>1</sup>

\*Centre for Integrative Biology (CBI), Department of Integrated Structural Biology, IGBMC (Institute of Genetics and of Molecular and Cellular Biology), Illkirch, France, †Centre National de la Recherche Scientifique (CNRS) UMR 7104, Illkirch, France, ‡Institut National de la Santé et de la Recherche Médicale (INSERM) U964, Illkirch, France, and §Université de Strasbourg, Strasbourg, France

After gradually moving away from preparation methods prone to artefacts such as plastic embedding and negative staining for cell sections and single particles, the field of cryo electron microscopy (cryo-EM) is now heading off at unprecedented speed towards high-resolution analysis of biological objects of various sizes. This ‘revolution in resolution’ is happening largely thanks to new developments of new-generation cameras used for recording the images in the cryo electron microscope which have much increased sensitivity being based on complementary metal oxide semiconductor devices. Combined with advanced image processing and 3D reconstruction, the cryo-EM analysis of nucleoprotein complexes can provide unprecedented insights at molecular and atomic levels and address regulatory mechanisms in the cell. These advances reinforce the integrative role of cryo-EM in synergy with other methods such as X-ray crystallography, fluorescence imaging or focussed-ion beam milling as exemplified here by some recent studies from our laboratory on ribosomes, viruses, chromatin and nuclear receptors. Such multi-scale and multi-resolution approaches allow integrating molecular and cellular levels when applied to purified or *in situ* macromolecular complexes, thus illustrating the trend of the field towards cellular structural biology.

## Introduction

The key event in cryo electron microscopy (cryo-EM) has been the introduction of cryo methods that allow preserving the biological sample in a hydrated

and functional state (Dubochet et al., 1988). One of the reasons for the importance of cryo methods is that artefacts due to sample dehydration, fixation, adsorption and staining can be avoided, thus allowing the sample to be observed in a functionally relevant state. Sample preservation is thus an advantage of cryo methods that is worth considering both for single particle analysis of complexes extracted from the cellular context and the analysis of cell sections (*i.e.* in contrast to plastic embedding, fixation etc.). The second good reason to use cryo methods is that they allow performing high-resolution analysis, whereas conventional methods intrinsically limit the

<sup>1</sup>To whom correspondence should be addressed (email: klaholz@igbmc.fr)

<sup>2</sup>Present address: European Molecular Biology Laboratory (EMBL), Structural and Computational Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

**Key words:** cryo electron microscopy, cryo electron tomography, Crystallography, Super-resolution microscopy, Structural biology.

**Abbreviations:** cryo-EM, cryo electron microscopy; cryo-ET, cryo electron tomography; CCD, charge-coupled device; CMOS, complementary metal oxide semiconductor; DDD, direct detection device; DQE, detective quantum efficiency; EcR, ecdysone receptor; FIB, focussed-ion beam; FRISBI, French Infrastructure for Integrated Structural Biology; Instruct, Integrated Structural Biology Infrastructure for Europe; kDa, kilo Dalton; MSA, multi-variate statistical analysis; mRNA, messenger RNA; rRNA, ribosomal RNA; SEM, scanning electron microscopy; SMLM, single-molecule localisation microscopy; tRNA, transfer RNA; 2D, two dimensional; 3D, three dimensional.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

attainable resolution. This second aspect was rather clear early on when cryo techniques were introduced in the field, as illustrated by the stunning amount of molecular structures visible by eye on cryo-EM images of various samples such as viruses, chromatin and ribosomes (Dubochet et al., 1988). However, it became obvious that the main limitation in seeing high-resolution details and being able to reconstruct them computationally in three dimensions (3D) was the ability to record appropriate images with good contrast and at the same time preserve the high-resolution information (*i.e.* avoid defocussing the microscope too much). For decades, the recording medium was photographic film (negatives, *e.g.* SO-163 from Kodak) which had the advantage of being able to record a large field of the specimen ( $\sim 8 \times 10$  cm film support size) and at high resolution (grain size around  $10 \mu\text{m}$ ), and have a good detective quantum efficiency (DQE) as compared with charge-coupled devices (CCDs). Nevertheless, CCDs had the advantage of direct data acquisition, that is no need for digitising negatives on a scanning device, an aspect that facilitates automation of image processing. However, the strongest and latest breakthrough is the introduction of direct electron detectors (based on a complementary metal oxide semiconductor (CMOS) detector with a direct detection device (DDD) sensor) which have much higher sensitivity than film or CCD thanks to their direct measurement of electron events, requiring no amplification of the signal nor fibre or lens optics (*e.g.* CCDs operate through an electron to light conversion by a scintillator followed by coupled fibre optics). Latest-generation variants of these direct electron detectors comprise (i) back-thinning of the silicon chip to avoid multiple electron scattering events, (ii) counting-mode to measure individual electron events at high read-out speeds (tens to hundreds of frames per second), (iii) localisation of the electron impact position with sub-pixel precision (*i.e.* super-resolution mode). Specific characteristics of these cameras have been described (McMullan et al., 2009; Ruskin et al., 2013; McMullan et al., 2014; Kuijper et al., 2015; Spear et al., 2015) and include an overall high DQE across the entire frequency range, wherein the particularly increased amplitudes in the low frequency range provide a much improved image contrast. An additional feature is the high read-out speed that allows dose fractionation and movie processing (Brilot et al., 2012; Campbell et al., 2012;

Li et al., 2013; Veesler et al., 2013; Allegretti et al., 2014; Scheres, 2014), that is beam-induced specimen drift can be motion-corrected by aligning a series of low-dose images taken on one given area of the specimen (Kunath et al., 1984); in addition, sub-frames with optimised dose can be selected for image processing (exposure filtering; Grant et al., 2015). It is these major technological developments of new-generation detectors that have recently introduced a 'revolution in resolution' in the cryo-EM field, analogous to the impact of the Pilatus & Eiger pixel detectors in the field of X-ray crystallography (Broennimann et al., 2006; Rajendran et al., 2011; Casanas et al., 2016). This has greatly contributed to a strong increase of the amount of structures determined by cryo-EM within the last few years (Figure 1).

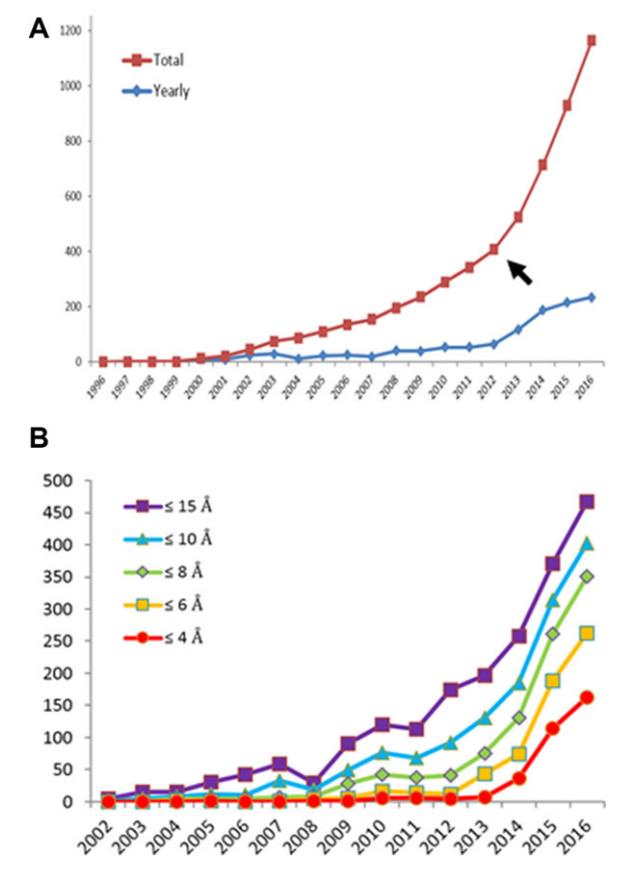
If the sample permits it, such cameras in principle allow – in synergy with advanced image processing informatics tools – to determine 3D structures in the  $3 \text{ \AA}$  resolution range. Such resolution values allow building and refining atomic models against the cryo-EM maps. To illustrate this in more detail, we will now provide some examples of recent high-resolution cryo-EM analysis from projects of our laboratory (Figure 2) and we will discuss the integrative role of cryo-EM in synergy with other complementary methods. This reflects the general trend of the field towards cellular structural biology which relies more and more on multi-resolution and multi-scale approaches to address the molecular and cellular organisation of living cells (Figure 3).

### Single particle cryo-EM analysis

The strength of current cryo-EM relies on the combination of cutting-edge instrumentation (high-resolution cryo electron microscopes equipped with direct electron detectors, such as those made available through the European and French infrastructures Instruct and FRISBI, see <https://www.structuralbiology.eu/> and <http://frisbi.eu/>) and advanced image processing. Image processing and 3D reconstruction represent important components of modern cryo-EM because they allow getting unique and detailed 3D insights into the object of interest, much beyond a visual 2D description at low resolution. Images obtained on a transmission electron microscope are 2D projection images of the object and therefore contain all internal features of the 3D

**Figure 1 | Recent increase of cryo-EM and cryo-ET studies as illustrated from the number of map depositions in the EMDB as a function of time and resolution**

(A, B) Largely thanks to developments of high-sensitivity detectors, the year 2013 (arrow) marks a transition of the exponentially growing number of structures determined by cryo-EM or cryo-ET, that is the curve became bimodal with a steep increase in the last 4 years which is likely to continue considering the vast amount of biological objects that now become amenable to cryo-EM and cryo-ET analysis. This is also visible in the growing contribution of high-resolution cryo-EM structures over the four last years (B). The data are taken from the <http://www.ebi.ac.uk/pdbe/emdb/> and <http://www.rcsb.org/pdb/> websites (as of September 26, 2016).

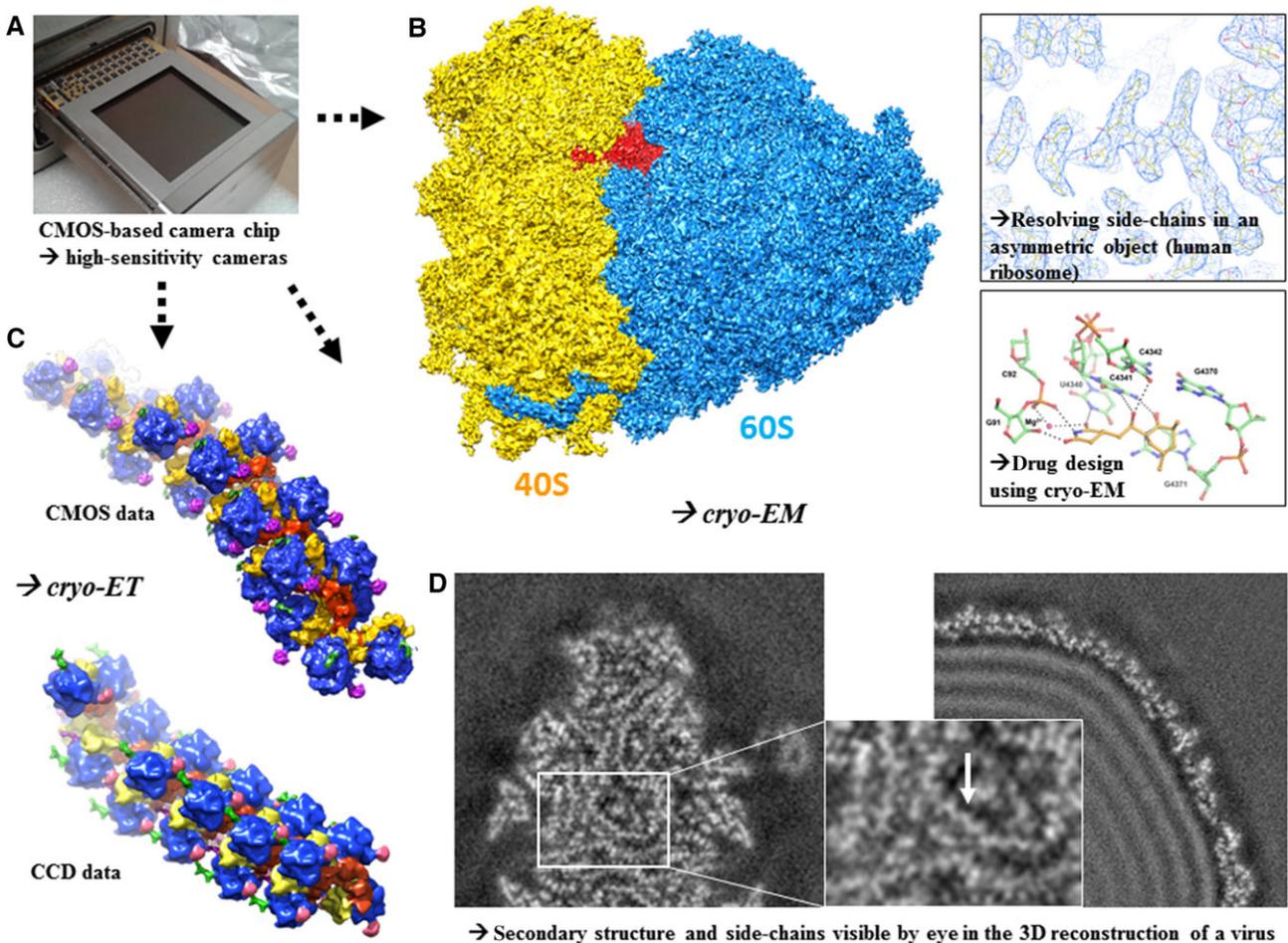


object which is seen under different orientations. Thus, reversely, a 3D map of the object can be obtained from 2D projection images by back-projection, that is the views of the 3D object seen under different orientations (viewing angles) can be combined into a single 3D reconstruction; this con-

cept applies to both single particle cryo-EM and cryo electron tomography (cryo-ET; for detailed basics on image processing and 3D reconstruction see for example reviews by van Heel et al., 2000; Ray et al., 2003; Briggs et al., 2013; Lučić et al., 2013; Milne et al., 2013; Scheres, 2014; Carazo et al., 2015; Carroni et al., 2016). An assumption in the 3D reconstruction process is that the input images describe the same object that it is functionally, structurally and conformationally homogenous. While this is the case for a single tomogram (recorded on one single object), it is rarely the case when averaging techniques are used, that is sub-tomogram averaging and single particle reconstruction using images describing (and tentatively merging) physically different objects. Sample heterogeneity can make the interpretation of 3D maps difficult or even impossible and might significantly limit the attainable resolution of the 3D reconstruction. Therefore, methods for 3D classifications to enable structure sorting are becoming an essential tool for the high-resolution analysis of single particle cryo-EM data, allowing the simultaneous analysis of several structures that are in equilibrium with each other. Examples of these are approaches based on cross correlation analysis using reference structures (template-based supervised classification, Gao et al., 2004) or based on multivariate statistical analysis (MSA) including local variance analysis in the particle images (White et al., 2004; Klaholz et al., 2004; Orlova & Saibil, 2010), resampling and bootstrapping methods to identify flexible regions in a macromolecular complex and perform 3D classifications (Penczek et al., 2006; Simonetti et al., 2008; Fischer et al., 2010; Klaholz, 2015; Liao et al., 2015), unsupervised classification (Fu et al., 2006) and maximum-likelihood (ML) based 3D classifications (Sigworth 1998; Scheres et al., 2010; Scheres et al., 2010; Lyumkis et al., 2013). The three categories of 2D/3D classification methods thus comprise (i) template-based methods which are intrinsically reference-biased, (ii) classification based on statistical analysis using MSA and bootstrapping methods and (iii) ML-based sorting; methods (ii) and (iii) are now commonly used as they turn out to be more robust during cryo-EM image processing of variable structures. These methods start from a low-resolution identification of larger conformational changes of the macromolecule of interest and then iteratively extend towards high-resolution sorting and

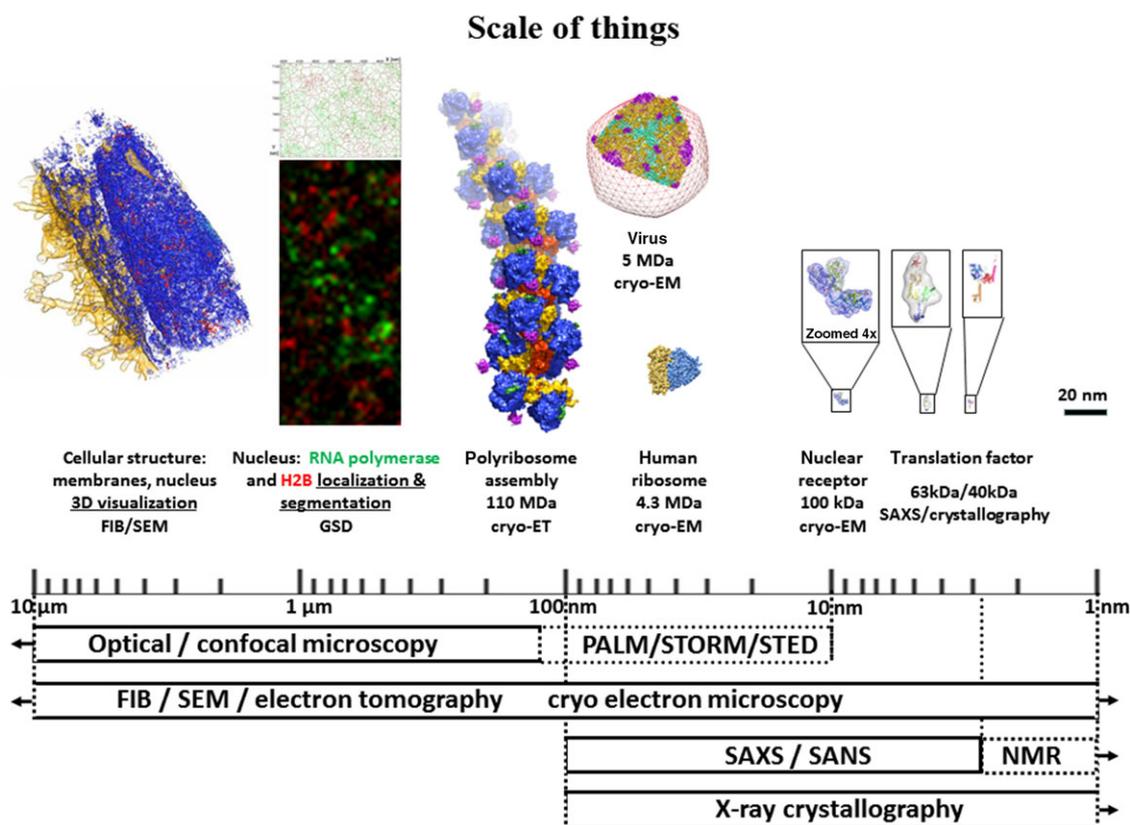
**Figure 2 | Examples of high-resolution cryo-EM and cryo-ET studies using direct electron detectors**

(A) The introduction of CMOS-based direct electron detectors in the EM field has led to a ‘revolution in resolution’ thanks to their much increased sensitivity (as an example, here the Falcon camera from the FEI company; setup installed on the Polara and Titan Krios cryo electron microscopes at the Centre for Integrative Biology, IGBMC; available through infrastructure access at <https://www.structuralbiology.eu/> and <http://frisbi.eu/>). (B) First high-resolution structure determination of the human 80S ribosome (Khatler et al., 2015). The map obtained from single particle cryo-EM image processing and 3D reconstruction (40S and 60S ribosomal subunits are labelled, the exit site tRNA is colored in red) allows visualising side-chains of amino acids and nucleic acids whose position needs to be known with precision to perform structure-based drug design (panels on the right; including a first human 80S complex with an antibiotic; Myasnikov et al., 2016). Two aspects can be highlighted there: (i) a resolution in the 3 Å range that is required to derive detailed atomic models can nowadays be obtained even for asymmetric objects, and (ii) cryo-EM has the potential to be used for structure-based drug design. (C) Comparison of cryo-ET reconstructions (after sub-tomogram averaging) obtained using either a CCD camera or a CMOS camera, illustrating the increased amount of structural features that can be visualised (poly-ribosome assembly, 60S and 40S ribosomal subunits are labelled in blue and yellow, respectively; Myasnikov et al., 2014). (D) Single particle cryo-EM reconstruction of the 135S bacteriophage capsid (cross-section through the 3D reconstruction; Orlov et al., unpublished) in which secondary structure elements such as  $\alpha$ -helices and residue details (side-chains; white arrow) can be visualised.



**Figure 3 | Scale of things and the integrative role of cryo-EM in molecular and cellular structural biology**

Schematic representation of the multi-scale and multi-resolution approach in structural biology. Various examples of macromolecular complexes studied in our group are shown to illustrate the integrative role of cryo-EM/cryo-ET at the interface with complementary methods such as X-ray crystallography, NMR, SAXS (towards the molecular and atomic levels on the right) and FIB/SEM, super-resolution fluorescence microscopy (PALM, STORM/GSDIM, STED etc.) and optical/confocal microscopy (towards the cellular level on the left). Right to left: crystal structure of translation initiation factor IF2 and SAXS analysis of IF2 (Simonetti et al., 2013a), single particle cryo-EM structure of the 100 kDa nuclear receptor complex USP/EcR (Maletta et al., 2014); the structures are zoomed 4× for visualisation; middle: single particle cryo-EM analysis of the human 80S ribosome (Khatter et al., 2015) and the 1358 bacteriophage (Orlov et al., unpublished) and cryo-ET analysis of eukaryotic polyribosomes (Myasnikov et al., 2014); left: super-resolution imaging (GSDIM) of RNA polymerase (green) and histone H2B (red) and cluster analysis using new tools (Andronov et al., 2016a, 2016b); FIB/SEM analysis of HeLa cells (Orlov et al., 2015).



structure refinement (discussed in more detail in Klaholz, 2015).

Recent examples of high-resolution single particle cryo-EM analysis include work on various complexes such as viruses, ribosomes, proteasome,  $\beta$ -galactoside and so on (among many others: Chen et al., 2009; Zhang et al., 2013; Bai et al., 2013; Wong et al., 2014; Banerjee et al., 2016; Earl et al., 2016; Fischer et al., 2015; Khatter et al., 2015; Greber et al., 2015). Among the macromolecular complexes that we recently studied in our laboratory to

high resolution (3 Å resolution range) using cryo-EM are the human ribosome and a virus (phage 1358); moreover, we also analysed a macromolecular assembly of poly-ribosomes by cryo-ET. These projects all benefited from the direct electron detector technology, high-resolution cryo electron microscopes and appropriate image processing (Figure 2).

Analysing the structure of the human ribosome can be considered to be the ultimate goal in structural studies of eukaryotic ribosomes, but it had to overcome the common *a priori* thinking that human

ribosomes would be too inhomogeneous and difficult to analyse. In order to address this issue, we developed biochemical protocols for the high-quality preparation of human ribosomes (Khatter et al., 2014) and their high-resolution structural analysis using advanced high-resolution cryo-EM and atomic model building (Figure 2; Khatter et al., 2015). The combination of data acquisition on a DDD camera, high-resolution image processing including 3D classifications and atomic model building and refinement using tools from the field of X-ray crystallography [PHENIX (Afonine et al., 2012), Chimera (Yang et al., 2012) and Coot (Emsley et al., 2010), among other available tools (*e.g.* Buster, REFMAC, CNS etc.; Smart et al., 2012; Brown et al., 2015; Adams et al., 2013; Natchiar et al., in prep.)] allowed determining the first atomic structure of the human ribosome. In this context, it is worth noting that the procedures for atomic model building and refinement into cryo-EM maps are evolving a lot recently, illustrating a strong synergy between crystallographic methods and cryo-EM (Brown et al., 2015; Barad et al., 2015; Sobolev et al., 2015; Natchiar et al., in prep.). The structural analysis of the human ribosome enabled full integration down to the atomic level, thus providing unprecedented insights into nucleotide and amino acid side-chains which can be analysed with respect to sequence conservation using a new tool for integrated sequence and 3D structure analysis of large macromolecular complexes (Beinsteiner et al., 2015). For example, it uncovered specific molecular interactions of the 28S rRNA with the universally conserved CCA end of the tRNA in the exit site. The structure opens the possibility of studying the molecular basis of interactions of drugs with the human ribosome, as well as functional complexes with mRNA, tRNA and protein factors. A follow-up study is the first structure of a ligand complex with the human ribosome that we recently determined, highlighting the potential of the human ribosome as a drug target for protein synthesis deregulations such as cancer (proliferating cells are dependent on elevated protein synthesis rates) and illustrating the capacity of high-resolution cryo-EM for structure-based drug design studies (Figure 2; Myasnikov et al., 2016). For comparison, our previous studies of bacterial translation initiation complexes were limited to a much lower resolution (8–10 Å; Simonetti et al., 2008; Simonetti et al., 2013a, 2013b; Eiler et al., 2013; Simonetti

et al., 2016) but nevertheless allowed addressing the localisation of initiation factors and tRNA, thus providing a solid basis for synergies with other methods including X-ray crystallography and solution SAXS.

Another example of a high-resolution single particle cryo-EM study is our ongoing structural analysis of the capsid of the lactococcal siphophage 1358 virion. Previously, negative staining EM and cryo-EM were used to determine the structure of the different parts of this tailed bacteriophage from the *Siphoviridae* family and to assemble an overall composite structure of the phage (Spinelli et al., 2014). The building units comprise the capsid, the connector, the tail and the base-plate. The structural analysis of the capsid using high-resolution cryo-EM now provides unprecedented insights in the 3–4 Å range (refinement ongoing; Orlov et al., unpublished). The quality of the structure is illustrated by the wealth of structural details that can be seen even by eye in the cross-section of the 3D reconstruction including secondary structure elements and side-chains. Indeed, zig-zag shapes in the densities reveal  $\alpha$ -helices and protruding densities correspond to the amino acid side-chains (Figure 2, bottom right). This highlights the fact that a resolution has been reached that is sufficient for tracing the peptide backbone and positioning side-chains from the amino acid sequence, thus allowing an atomic model refinement against the cryo-EM map. Considering that phage 1358 infects *L. lactis* strains and has some similarity to *Listeria* phages, these structural investigations may be useful for medical and agricultural implications.

### Cryo-ET analysis and sub-tomogram averaging

Cryo-ET is particularly useful to analyse unique cellular structures. Rather than obtaining a 3D reconstruction from different orientational views of physically different objects like in single particle cryo-EM, it is obtained from images acquired while rotating the sample with predefined angles. The images of a tilt series can then be aligned to reconstruct a tomogram (see for example Briggs, 2013; Lučić et al., 2013; Dubrovsky et al., 2015; Asano et al., 2016). If common sub-structures exist these can be extracted as sub-tomograms, aligned and averaged to form sub-tomogram averages with an improved signal-to-noise ratio (Wan et al., 2016; Galaz-Montoya et al., 2016).

## Integrative role of cryo-EM in structural biology

Such approaches have been used in the past to analyse, for example, the cellular structure of the nuclear pore complex (NPC), of membrane-bound ribosomes, of actin filaments in combination with segmentation tools and template-based identification of the components (Frangakis et al., 2002; Medalia et al., 2002; Beck et al., 2004; Ortiz et al., 2006). Recently, stunning insights into cellular sub-structures have been obtained (Hagen et al., 2015; Nans et al., 2015; Chang et al., 2016; Irobalieva et al., 2016; Kosinski et al., 2016; Lin et al., 2016; Mahamid et al., 2016), achieving in a first case even side-chain resolution (Schur et al., 2016) using optimised cryo-ET, sub-tomogram averaging and a new dose-symmetric tilt acquisition scheme that preserves high-resolution data more isotropically (Hagen et al., 2016). We used cryo-ET in combination with sub-tomogram averaging and molecular modelling to address the supramolecular organisation of eukaryotic polyribosomes which can form large macromolecular assemblies (Brandt et al., 2010; Myasnikov et al., 2013; Afonina et al., 2014; Myasnikov et al., 2014; Afonina et al., 2015). This allowed deriving the 3D structure of one of the largest asymmetric complexes to date (~100 MDa, comprising over 20 ribosomes on the same mRNA molecule; Myasnikov et al., 2014). The structure allowed the visualisation of the three functional parts of the polysome assembly: the central core region that forms a rather compact left-handed supramolecular helix and the more open regions that harbour the initiation and termination sites at either ends. The helical region forms a continuous mRNA channel where the mRNA strand bridges neighbouring exit and entry sites of the ribosomes and prevents the mRNA from looping between ribosomes. This structure provides unprecedented insights into protein- and RNA-mediated inter-ribosome contacts that involve conserved sites through 40S ribosomal subunits and long protruding RNA expansion segments. These findings shed new light on the molecular machinery of the ribosome and its mode of action in the cellular context. The impact of direct electron detectors is illustrated by the improved resolution of the tomograms (Figure 2). In the future, contrast-increasing phase plates may be helpful to address more of the molecular details within polysomes, possibly in combination with double tilt cryo-ET which provides 3D reconstructions with less reconstruction artefacts (Myasnikov et al., 2013; one of the first ex-

amples of single particle cryo-ET together with some other examples: Dudkina et al., 2010; Murata et al., 2010; Wang et al., 2011).

## Integrative role of cryo-EM and current trends in complementary approaches

Another aspect of cryo-EM is its integrative role in multi-scale and multi-resolution approaches that integrate molecular and cellular levels by combining various structural methods such as X-ray crystallography, fluorescence imaging or focussed-ion beam milling as exemplified here by some recent studies from our laboratory on ribosomes, viruses, chromatin and nuclear receptors (Figure 3). An important consideration is that challenging objects require the right choice of the method (Ménétret et al., 2013), each having intrinsic limitations, but when combined these methods can create synergies at the interfaces that overcome the individual limits. For example, in X-ray crystallography a typical bottleneck is crystallisation, but it is not much restricted to a particular object size as illustrated by the possibility of crystallising small organic compounds and large complexes such as ribosomes and viruses (including very large viruses with a diameter in the 1000 Å range). However, in cryo-EM, a typical limitation is the lower size of a complex because the lower the molecular weight of the macromolecule the smaller the image contrast is, which limits the possibility of accurately determining the structure. Nevertheless, there are clear trends in the cryo-EM field to move towards smaller complexes. An example of this is our recent study of a 100 kDa complex, the first structure of a full-length nuclear receptor bound to an inverted DNA repeat, the USP/EcR complex in insects where EcR is the ecdysone receptor (Maletta et al., 2014); it revealed an asymmetric organisation of the complex although the DNA is almost symmetric (Figure 3); note that the data were collected at a reduced acceleration voltage (100 kV) to increase the image contrast. This represents a technical advance in the field with respect of studying relatively small complexes by cryo-EM, thus opening the possibility to study many other biological complexes and drug targets previously believed to be too small for cryo-EM. Other examples in the field are the studies of a 50 kDa RNA (Baird et al., 2010) and most recently the 3.8 Å resolution structure of a 93 kDa protein (Merk et al., 2016),

highlighting the fact that high-resolution structures can also be obtained on relatively small complexes, provided the biological samples behave well and can be imaged well (*i.e.* low aggregation, stability in relatively low-salt or low-detergent concentrations, good particle distribution etc.).

Among various structural biology methods that use averaging techniques, a common limitation is the sample homogeneity. While crystallisation (and re-crystallisation) can contribute to the high-level purification of a chemical compound or even a large macromolecular complex (*e.g.* PEG10000 precipitation of yeast and human ribosomes; Ben-Shem et al., 2010; Khatter et al., 2014; Khatter et al., 2015), sample homogeneity often limits crystallisation. However, the problem of sample heterogeneity can be turned into an advantage when using single-particle cryo-EM considering that with advanced image processing and 3D classification tools (see section Single particle cryo-EM analysis) different structures can be sorted and separated into distinct sub-populations, allowing to improve the refinement of a cryo-EM structure by local masking and refinement to high resolution and at the same time describe several conformational states of a complex that are in equilibrium with each other. Similarly, when using cryo-ET and sub-tomogram averaging of similar structures observed *in situ* in the cell (*e.g.* polyribosomes in Figure 2), it is possible to classify the aligned tomograms and address several structures at the same time (Heumann et al., 2011; Stölken et al., 2011; Frank et al., 2012; Kuybeda et al., 2013; Xu et al., 2013; Chen et al., 2014; Castaño-Díez et al., 2016; Obbineni et al., 2016). However, sorting is not possible on unique structures such as large regions of the cell, in which case methods such as focussed-ion beam/scanning electron microscopy (FIB/SEM) are more appropriate (*e.g.* chromatin in Figure 3; Orlov et al., 2015). FIB is a method originally emanating from material sciences, which, when applied to biological specimens, appears to be one of the most promising structural biology methods for future cell biology studies (Villa et al., 2013; Kizilyaprak et al., 2014; Rigort et al., 2015). In contrast to ultramicrotome sectioning of cells, which can exhibit knife cutting and compression artefacts, FIB milling allows obtaining high-quality cuts through the cell that can be observed by SEM (section by section to reconstitute a full tomogram, in the literal sense of

τομὸς *tomos* (Greek: cut). Alternatively, thin lamella can be prepared by FIB milling and observed by cryo-ET (*e.g.* Mahamid et al., 2015; Schaffer et al., 2016; Zhang et al., 2016). Such analyses have the great advantage of being *in situ* (Lučić et al., 2013) and thus in the functional cellular context, but they require the identification of the complexes of interest by complementary methods such as fluorescence labelling. In an effort to allow the identification of chromatin structures, we have explored the possibilities of molecular imaging in cells through single-molecule localisation microscopy (SMLM; *e.g.* dSTORM, PALM; we used GSDIM, ground state depletion microscopy followed by individual molecule return). To make best use of super-resolution data, we developed an integrated software pipeline for image reconstruction, drift and chromatic aberration correction, co-localisation, resolution estimation, segmentation, clustering and quantification of labelled complexes (SharpViSu & ClusterViSu; Andronov et al., 2016a, 2016b). In addition to be very useful in the field of super-resolution microscopy, in the future such informatics tools may become particularly useful in the context of correlative light and electron microscopy (CLEM) approaches (Koning et al., 2014; Schirra et al., 2014; Arnold et al., 2016; Karreman et al., 2016; Schorb et al., 2016) to address the cellular fine structure, identify and localise protein complexes and visualise them at high resolution using cryo-EM and cryo-ET, that is perform an identification of the molecule-of-interest (MOI) rather than only the overall region-of-interest (ROI). Examples including SMLM (*e.g.* Kim et al., 2015) are studies of complexes such as RNA polymerase, nucleosomes, viruses or NPCs (Szymborska et al., 2013; Löscherberger et al., 2014; Laine et al., 2015) some of which were reconstructed from super-resolution data using cryo-EM 3D reconstruction methods (Szymborska et al., 2013; Andronov et al., in prep.).

Current and near-future challenges in cryo-EM developments comprise key questions such as (i) how to push the resolution to the atomic level, (ii) how to analyse flexible complexes and (iii) how to integrate towards the cellular level. Two important areas to address these are (i) the instrumentation and (ii) software developments. Instrumentation (i) comprises for example high-resolution electron microscopes, direct electron detector cameras with higher DQE at high frequencies, energy filters to

## Integrative role of cryo-EM in structural biology

remove inelastically scattered electrons and reduce the background in the images, correction of the spherical aberration (Cs) to improve the optical system in the column of the microscope, microelectron diffraction to determine the structure from small 3D crystals (Nederlof et al., 2013; Sawaya et al., 2016; Shi et al., 2016), spraying of small amounts on cryo-EM grids (Chen et al., 2015; Razinkov et al., 2016), cryo transfer between FIB and electron microscope for cryo-ET (Schaffer et al., 2016) and contrast-increasing phase plates placed in the back-focal plane of the microscope (Danev et al., 2014; Dai et al., 2014; Frindt et al., 2014; Walter et al., 2015; Chua et al., 2016; Danev et al., 2016; Glaeser, 2016; Rhinow 2016; Khoshouei et al., 2016a); phase plates could facilitate structure determination of relatively small complexes (Khoshouei et al., 2016b), which were difficult to address previously and usually limited to lower resolution (Baird et al., 2010; Orlov et al., 2012; Maletta et al., 2014) even though DDDs have helped a lot moving forward (Merk et al., 2016), suggesting that synergies will appear for example between phase plates, energy filters and high-sensitivity cameras to enable high-contrast high-resolution image acquisition (a feature that usually contradicts itself considering the requirement of defocussing during data collection to get some reasonable amount of image contrast required for image processing). Further software developments (ii) will be required for automatic data acquisition for massive data collection for single particle cryo-EM and cryo-ET (e.g. reviewed in Tan et al., 2016; automatic beam alignments including for Cs-corrected microscopes; remote control), and on-the-fly image processing during data collection, particle sorting and 3D classifications to address sample heterogeneity, automation in cryo-EM/cryo-ET structure determination, automation of backbone tracing like in X-ray crystallography, atomic model building and refinement into cryo-EM/cryo-ET maps of large macromolecular complexes to move towards larger cellular structures. Taken together, the scientific community is currently experiencing a very exciting era that is moving more and more towards the integration of various structural and imaging methods among which cryo-EM, cryo-ET and FIB will play a key role in multi-resolution integration. In close synergy with functional studies, this evolution towards cellular structural biology will lead in the coming

years to unprecedented insights into cellular function and drug targets, including the analysis of the dynamic changes of macromolecular complexes that reflect their functional transitions in the cellular environment.

## Funding

This work was supported by the European Research Council (ERC Starting Grant N.243296 TRANSLATIONMACHINERY), the Fondation pour la Recherche Médicale (FRM), the Association pour la Recherche sur le Cancer (ARC), the Alsace Region, the Agence National pour la Recherche, Université de Strasbourg and Investissement d'Avenir (IDEX), the Institut National du Cancer (INCa), the Centre National pour la Recherche Scientifique (CNRS), the Institut National pour la Recherche Médicale (Inserm), by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01, and Instruct as part of the European Strategy Forum on Research Infrastructures (ESFRI).

## Acknowledgements

We apologise to our colleagues for any publications not being cited which is due to space limitations and the cross-disciplinary nature of this review article. We would like to thank our colleagues with whom we collaborate on the projects mentioned in this review, notably Jean-Luc Vonesch, Alexander Spirin, Vladimir Shirokov, Dino Moras, Jean-François Peyron, Danièle Spehner and Christian Cambillau, and the members of our platforms and facilities for their support.

## Conflict of interest statement

The authors have declared no conflict of interest.

## References

- Adams, P.D., Baker, D., Brunger, A.T., Das, R., DiMaio, F., Read, R.J., Richardson, D.C., Richardson, J.S. and Terwilliger, T.C. (2013) Advances, interactions, and future developments in the CNS, Phenix, and Rosetta structural biology software systems. *Annu. Rev. Biophys.* **42**, 265–287
- Afonina, Z.A., Myasnikov, A.G., Shirokov, V.A., Klaholz, B.P. and Spirin, A.S. (2015) Conformation transitions of eukaryotic polyribosomes during multi-round translation. *Nucleic Acids Res.* **43**, 618–628
- Afonina, Z.A., Myasnikov, A.G., Shirokov, V.A., Klaholz, B.P. and Spirin, A.S. (2014) Formation of circular polyribosomes on

- eukaryotic mRNA without cap-structure and poly(A)-tail: a cryo electron tomography study. *Nucleic Acids Res.* **42**, 9461–9469
- Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D. (2012) Toward automated crystallographic structure refinement with phenix.refine. *Acta Cryst.* **D68**, 352–367
- Allegretti, M., Mills, D.J., McMullan, G., Kühlbrandt, W. and Vonck, J. (2014) Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. *Elife* **3**:e01963. doi: 10.7554/eLife.01963.
- Andronov, L., Lutz, Y., Vonesch, J.-L. and Klaholz, B.P. (2016a) SharpViSu: integrated analysis and segmentation of super-resolution microscopy data. *Bioinformatics* **32**, 2239–2241
- Andronov, L., Orlov, I., Lutz, Y., Vonesch, J.-L. and Klaholz, B.P. (2016b) ClusterViSu, a method for clustering of protein complexes by Voronoi tessellation in super-resolution microscopy. *Sci. Rep.* **6**, 24084
- Arnold, J., Mahamid, J., Lucic, V., de Marco, A., Fernandez, J.J., Laugks, T., Mayer, T., Hyman, A.A., Baumeister, W. and Plitzko, J.M. (2016) Site-Specific Cryo-focused Ion Beam Sample Preparation Guided by 3D Correlative Microscopy. *Biophys. J.* **110**, 860–869
- Asano, S., Engel, B.D. and Baumeister, W. (2016) In Situ Cryo-Electron Tomography: A Post-Reductionist Approach to Structural Biology. *J. Mol. Biol.* **428**, 332–343
- Bai, X.C., Fernandez, I.S., McMullan, G. and Scheres, S.H. (2013) Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* **2**:e00461. doi: 10.7554/eLife.00461.
- Baird, N.J., Ludtke, S.J., Khant, H., Chiu, W., Pan, T. and Sosnick, T.R. (2010) Discrete structure of an RNA folding intermediate revealed by cryo-electron microscopy. *J. Am. Chem. Soc.* **132**, 16352–163523
- Banerjee, S., Bartesaghi, A., Merk, A., Rao, P., Bulfer, S.L., Yan, Y., Green, N., Mroczkowski, B., Neitz, R.J., Wipf, P., Falconieri, V., Deshaies, R.J., Milne, J.L., Huryn, D., Arkin, M. and Subramaniam, S. (2016) 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science* **351**, 871–875
- Barad, B.A., Echols, N., Wang, R.Y., Cheng, Y., DiMaio, F., Adams, P.D. and Fraser, J.S. (2015) EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946.
- Beck, M., Förster, F., Ecke, M., Plitzko, J.M., Melchior, F., Gerisch, G., Baumeister, W. and Medalia, O. (2004) Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* **306**, 1387–1390
- Beinsteiner, B., Michalon, J. and Klaholz, B. P. (2015) IBISS, a versatile and interactive tool for integrated sequence and 3D structure analysis of large macromolecular complexes. *Bioinformatics* **31**, 3339–3344
- Ben-Shem, A., Jenner, L., Yusupova, G. and Yusupov, M. (2010) Crystal structure of the eukaryotic ribosome. *Science* **330**, 1203–1209
- Brandt, F., Carlson, L.A., Hartl, F.U., Baumeister, W. and Grünewald, K. (2010) The three-dimensional organization of polyribosomes in intact human cells. *Mol. Cell* **39**, 560–569
- Briggs, J.A. (2013) Structural biology in situ - the potential of subtomogram averaging. *Curr. Opin. Struct. Biol.* **23**, 261–267
- Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., Henderson, R. and Grigorieff, N. (2012) Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.* **177**, 630–637
- Broennimann, C., Eikenberry, E.F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H. and Wagner, A. (2006) The PILATUS 1M detector. *J. Synchrotron Radiat.* **13**, 120–130
- Brown, A., Long, F., Nicholls, R.A., Toots, J., Emsley, P. and Murshudov, G. (2015) Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 136–153
- Campbell, M.G., Cheng, A., Brilot, A.F., Moeller, A., Lyumkis, D., Veessler, D., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B. and Grigorieff, N. (2012) Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* **20**, 1823–1828
- Carazo, J.M., Sorzano, C.O., Otón, J., Marabini, R. and Vargas, J. (2015) Three-dimensional reconstruction methods in Single Particle Analysis from transmission electron microscopy data. *Arch. Biochem. Biophys.* **581**, 39–48
- Caroni, M. and Saibil, H.R. (2016) Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods* **95**, 78–85
- Casanas, A., Warshamanage, R., Finke, A.D., Panepucci, E., Olieric, V., Nöll, A., Tampé, R., Brandstetter, S., Förster, A., Mueller, M., Schulze-Briese, C., Bunk, O. and Wang, M. (2016) ELGER detector: application in macromolecular crystallography. *Acta Crystallogr. D Struct. Biol.* **72**, 1036–1048
- Castaño-Diez, D., Kudryashev, M. and Stahlberg, H. (2016) Dynamo Catalogue: Geometrical tools and data management for particle picking in subtomogram averaging of cryo-electron tomograms. *J. Struct. Biol.* **16**, 30111–30113
- Chang, Y.W., Rettberg, L.A., Treuner-Lange, A., Iwasa, J., Søgaard-Andersen, L. and Jensen, G.J. (2016) Architecture of the type IVa pilus machine. *Science* **351**, 6278:aad2001
- Chen, B., Kaledhonkar, S., Sun, M., Shen, B., Lu, Z., Barnard, D., Lu, T.M., Gonzalez, R.L. Jr. and Frank, J. (2015) Structural dynamics of ribosome subunit association studied by mixing-spraying time-resolved cryogenic electron microscopy. *Structure* **23**, 1097–1105
- Chen, J.Z., Settembre, E.C., Aoki, S.T., Zhang, X., Bellamy, A.R., Dormitzer, P.R., Harrison, S.C. and Grigorieff, N. (2009) Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10644–10648
- Chen, Y., Pfeffer, S., Fernández, J.J., Sorzano, C.O. and Förster, F. (2014) Autofocused 3D classification of cryoelectron subtomograms. *Structure* **22**, 1528–1537
- Chua, E.Y., Vogirala, V.K., Inian, O., Wong, A.S., Nordenskiöld, L., Plitzko, J.M., Danev, R. and Sandin, S. (2016) 3.9 Å structure of the nucleosome core particle determined by phase-plate cryo-EM. *Nucleic Acids Res.* **44**, 8013–8019
- Dai, W., Fu, C., Khant, H.A., Ludtke, S.J., Schmid, M.F. and Chiu, W. (2014) Zernike phase-contrast electron cryotomography applied to marine cyanobacteria infected with cyanophages. *Nat. Protoc.* **9**, 2630–2642
- Danev, R. and Baumeister, W. (2016) Cryo-EM single particle analysis with the Volta phase plate. *Elife* **5**: pii: e13046
- Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J.M. and Baumeister, W. (2014) Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **44**, 15635–15640
- Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowell, A.W. and Schultz, P. (1988) Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228
- Dubrovsky, A., Sorrentino, S., Harapin, J., Sapra, K.T. and Medalia, O. (2015) Developments in cryo-electron tomography for in situ structural analysis. *Arch. Biochem. Biophys.* **581**, 78–85
- Dudkina, N.V., Kouril, R., Bultema, J.B. and Boekema, E.J. (2010) Imaging of organelles by electron microscopy reveals protein-protein interactions in mitochondria and chloroplasts. *FEBS Lett.* **584**, 2510–2515

- Earl, L.A. and Subramaniam, S. (2016) Cryo-EM of viruses and vaccine design. *Proc. Nat. Acad. Sci. U.S.A.* **113**, 8903–8905
- Eiler, D., Lin, J., Simonetti, A., Klaholz, B.P. and Steitz, T.A. (2013) IF2 Initiation factor 2 crystal structure reveals a different domain organization from eukaryotic initiation factor 5B and mechanism among translational GTPases. *Proc. Nat. Acad. Sci. U.S.A.* **110**, 15662–15667
- Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
- Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V. and Stark, H. (2010). Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329–333
- Fischer, N., Neumann, P., Konevega, A.L., Bock, L.V., Ficner, R., Rodnina, M.V. and Stark, H. (2015) Structure of the E. coli ribosome-EF-Tu complex at <math>3\text{ \AA}</math> resolution by Cs-corrected cryo-EM. *Nature* **520**, 567–570
- Frangakis, A.S., Böhm, J., Förster, F., Nickell, S., Nicastro, D., Typke, D., Hegerl, R. and Baumeister, W. (2002) Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14153–14158
- Frank, G.A., Bartesaghi, A., Kuybeda, O., Borgnia, M.J., White, T.A., Sapiro, G. and Subramaniam, S. (2012) Computational separation of conformational heterogeneity using cryo-electron tomography and 3D sub-volume averaging. *J. Struct. Biol.* **178**, 165–176
- Frindt, N., Oster, M., Hettler, S., Gamm, B., Dieterle, L., Kowalsky, W., Gerthsen, D. and Schröder, R.R. (2014) In-focus electrostatic Zach phase plate imaging for transmission electron microscopy with tunable phase contrast of frozen hydrated biological samples. *Microsc. Microanal.* **1**, 175–183
- Fu, J., Gao, H. and Frank, J. (2006) Unsupervised classification of single particles by cluster tracking in multi-dimensional space. *J. Struct. Biol.* **157**, 226–239
- Galaz-Montoya, J.G., Hecksel, C.W., Baldwin, P.R., Wang, E., Weaver, S.C., Schmid, M.F., Ludtke, S.J. and Chiu, W. (2016) Alignment algorithms and per-particle CTF correction for single particle cryo-electron tomography. *J. Struct. Biol.* **194**, 383–394
- Gao, H., Valle, M., Ehrenberg, M. and Frank, J. (2004) Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *J. Struct. Biol.* **147**, 283–290
- Glaeser, R.M. (2016) Protein complexes in focus. *Elife* **5**: pii: e16156
- Grant, T. and Grigorieff, N. (2015) Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* **4**:e06980
- Greber, B.J., Bieri, P., Leibundgut, M., Leitner, A., Aebersold, R., Boehringer, D. and Ban, N. (2015) The complete structure of the 55S mammalian mitochondrial ribosome. *Science* **348**, 303–308
- Hagen, C., Dent, K.C., Zeev-Ben-Mordehai, T., Grange, M., Bosse, J.B., Whittle, C., Klupp, B.G., Siebert, C.A., Vasishtan, D., Bäuerlein, F.J., Chelieski, J., Werner, S., Guttman, P., Rehbein, S., Henzler, K., Demmerle, J., Adler, B., Kosziniowski, U., Schermelleh, L., Schneider, G., Enquist, L.W., Plitzko, J.M., Mettenleiter, T.C. and Grünewald, K. (2015) Structural Basis of Vesicle Formation at the Inner Nuclear Membrane. *Cell* **163**, 1692–1701
- Hagen, W.J., Wan, W. and Briggs, J.A. (2016) Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **16**, 30113–30117
- Heumann, J.M., Hoenger, A. and Mastronarde, D.N. (2011) Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *J. Struct. Biol.* **175**, 288–299
- Irobalieva, R.N., Martins, B. and Medalia, O. (2016) Cellular structural biology as revealed by cryo-electron tomography. *J. Cell Sci.* **129**, 469–476
- Karreman, M.A., Mercier, L., Schieber, N.L., Solecki, G., Allio, G., Winkler, F., Ruthensteiner, B., Goetz, J.G. and Schwab, Y. (2016) Fast and precise targeting of single tumor cells in vivo by multimodal correlative microscopy. *J. Cell Sci.* **129**, 444–456
- Khatter, H., Myasnikov, A.G., Mastio, L., Billas, I.M.L., Birk C., Stella, S. and Klaholz, B.P. (2014) Purification, characterization and crystallization of the human 80S ribosome. *Nucleic Acids Res.* **42**, 1–11
- Khatter, H., Myasnikov, A.G., Natchiar, K. and Klaholz, B.P. (2015) Structure of the human 80S ribosome. *Nature* **520**, 640–645
- Khoshouei, M., Pfeffer, S., Baumeister, W., Förster, F. and Danev, R. (2016a) Subtomogram analysis using the Volta phase plate. *J. Struct. Biol.* **16**, 30103–30104
- Khoshouei, M., Radjainia, M., Phillips, A.J., Gerrard, J.A., Mitra, A.K., Plitzko, J.M., Baumeister, W. and Danev, R. (2016b) Volta phase plate cryo-EM of the small protein complex Prx3. *Nat. Commun.* **7**, 10534
- Kim, D., Deerinck, T.J., Sigal, Y.M., Babcock, H.P., Ellisman, M.H. and Zhuang, X. (2015) Correlative stochastic optical reconstruction microscopy and electron microscopy. *PLoS One* **10**, e0124581.
- Kizilyaprak, C., Daraspe, J. and Humbel, B.M. (2014) Focused ion beam scanning electron microscopy in biology. *J. Microsc.* **254**, 109–114
- Klaholz, B.P. (2015) Structure sorting of multiple macromolecular states in heterogeneous cryo-EM samples by 3D multivariate statistical analysis. *Open J. Stat.* **5**, 820–836
- Klaholz, B.P., Myasnikov, A.G. and van Heel, M. (2004) Visualization of release factor 3 on the ribosome during termination of protein synthesis. *Nature* **427**, 862–865
- Koning, R.I., Celler, K., Willemsse, J., Bos, E., van Wezel, G.P. and Koster, A.J. (2014) Correlative cryo-fluorescence light microscopy and cryo-electron tomography of *Streptomyces*. *Methods Cell Biol.* **124**, 217–239
- Kosinski, J., Mosalaganti, S., von Appen, A., Teimer, R., DiGuilio, A.L., Wan, W., Bui, K.H., Hagen, W.J., Briggs, J.A., Glavy, J.S., Hurt, E. and Beck, M. (2016) Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365
- Kuijper, M., van Hoffen, G., Janssen, B., Geurink, R., De Carlo, S., Vos, M., van Duinen, G., van Haeringen, B. and Storms, M. (2015) FEI's direct electron detector developments: Embarking on a revolution in cryo-TEM. *J. Struct. Biol.* **192**, 179–187
- Kunath, W., Weiss, K., Sack-Kongehl, H., Kessel, M., Zeitler, E. (1984). Time-resolved low-dose microscopy of glutamine synthetase molecules. *Ultramicroscopy* **13**, 241–252
- Kuybeda, O., Frank, G.A., Bartesaghi, A., Borgnia, M., Subramaniam, S. and Sapiro, G.A. (2013) Collaborative framework for 3D alignment and classification of heterogeneous subvolumes in cryo-electron tomography. *J. Struct. Biol.* **181**, 116–127
- Laine, R.F., Albecka, A., van de Linde, S., Rees, E.J., Crump, C.M. and Kaminski, C.F. (2015) Structural analysis of herpes simplex virus by optical super-resolution imaging. *Nat. Commun.* **6**, 5980
- Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A. and Cheng, Y. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590
- Liao, H.Y., Hashem, Y. and Frank, J. (2015) Efficient estimation of three-dimensional covariance and its application in the analysis of heterogeneous samples in cryo-electron microscopy. *Structure* **23**, 1129–1237
- Lin, D.H., Stuwe, T., Schilbach, S., Rundlet, E.J., Perriches, T., Mobbs, G., Fan, Y., Thierbach, K., Huber, F.M., Collins, L.N., Davenport, A.M., Jeon, Y.E. and Hoelz, A. (2016) Architecture of the symmetric core of the nuclear pore. *Science* **352**, 6283
- Löschberger, A., Franke, C., Krohne, G., van de Linde, S. and Sauer, M. (2014) Correlative super-resolution fluorescence and electron microscopy of the nuclear pore complex with molecular resolution. *J. Cell Sci.* **127**, 4351–4355

- Lučić, V., Rigort, A. and Baumeister, W. (2013) Cryo-electron tomography: the challenge of doing structural biology in situ. *J. Cell Biol.* **202**, 407–419
- Lyumkis, D., Brilot, A.F., Theobald, D.L. and Grigorieff, N. (2013) Likelihood-based classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* **183**, 377–388
- Mahamid, J., Pfeffer, S., Schaffer, M., Villa, E., Danev, R., Cuellar, L.K., Förster, F., Hyman, A.A., Plitzko, J.M. and Baumeister, W. (2016) Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969–972
- Mahamid, J., Schampers, R., Persoon, H., Hyman, A.A., Baumeister, W. and Plitzko, J.M. (2015) A focused ion beam milling and lift-off approach for site-specific preparation of frozen-hydrated lamellas from multicellular organisms. *J. Struct. Biol.* **192**, 262–269
- Maletta, M., Orlov, I.M., Roblin, P., Beck, Y., Moras, D., Billas, I.M.L. and Klaholz, B.P. (2014) The palindromic DNA-bound USP/EcR nuclear receptor adopts an asymmetric organization with allosteric domain positioning. *Nat. Commun.* **5**, 4139
- McMullan, G., Chen, S., Henderson, R. and Faruqi, A.R. (2009) Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy* **109**, 1126–1143
- McMullan, G., Faruqi, A.R., Clare, D. and Henderson, R. (2014) Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* **147**, 156–163
- Medalia, O., Weber, I., Frangakis, A.S., Nicastro, D., Gerisch, G. and Baumeister, W. (2002) Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science* **298**, 1209–1213
- Ménétre, J.-F., Khatter, H., Simonetti, A., Orlov, I., Myasnikov, A.G., Vidhya, K.V., Manicka, S., Torchy, M., Mohideen, K., Humm, A.-S., Hazemann, I., Urzhumtsev, A., Klaholz, B.P. (2013) Integrative structure-function analysis of large nucleoprotein complexes. In: D. Klostermeier and C. Hammann (Eds.). *RNA structure and folding*. de Gruyter.
- Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M.I., Pragani, R., Boxer, M.B., Earl, L.A., Milne, J.L. and Subramaniam, S. (2016) Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **165**, 1698–1707
- Milne, J.L., Borgnia, M.J., Bartesaghi, A., Tran, E.E., Earl, L.A., Schauder, D.M., Lengyel, J., Pierson, J., Patwardhan, A. and Subramaniam, S. (2013) Cryo-electron microscopy - a primer for the non-microscopist. *FEBS J.* **280**, 28–45
- Murata, K., Liu, X., Danev, R., Jakana, J., Schmid, M.F., King, J., Nagayama, K. and Chiu, W. (2010) Zernike phase contrast cryo-electron microscopy and tomography for structure determination at nanometer and subnanometer resolutions. *Structure* **18**, 903–912
- Myasnikov, A.G., Afonina, Z.A., Ménétre, J.-F., Shirokov, V.A., Spirin, A.S. and Klaholz, B.P. (2014) The molecular structure of the left-handed supra-molecular helix of eukaryotic polyribosomes. *Nat. Commun.* **5**, 5294
- Myasnikov, A.G., Afonina Z. and Klaholz, B.P. (2013) Single particle and molecular assembly analysis of polyribosomes by single- and double-tilt cryo electron tomography. *Ultramicroscopy* **126**, 33–39
- Myasnikov, A.G., Natchiar, S.K., Nebout, M., Hazemann, I., Imbert, V., Khatter, H., Peyron, J.-F. and Klaholz, B.P. (2016) Structure-function insights reveal the human ribosome as a cancer target for antibiotics. *Nat. Commun.* **7**, 12856
- Nans, A., Kudryashev, M., Saibil, H.R. and Hayward, R.D. (2015) Structure of a bacterial type III secretion system in contact with a host membrane in situ. *Nat. Commun.* **6**, 10114
- Nederlof, I., Li, Y.W., van Heel, M. and Abrahams, J.P. (2013) Imaging protein three-dimensional nanocrystals with cryo-EM. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 852–859
- Obbineni, J.M., Yamamoto, R. and Ishikawa, T. (2016) A simple and fast approach for missing-wedge invariant classification of subtomograms extracted from filamentous structures. *J. Struct. Biol.* **16**, 30172
- Orlov, I., Rochel, N., Moras, D. and Klaholz, B.P. (2012) Structure of the full human RXR/VDR nuclear receptor heterodimer complex with its DR3 target DNA. *EMBO J.* **31**, 291–300
- Orlov, I., Schertel, A., Zuber, G., Klaholz, B.P., Drillien, R., Weiss, E., Schultz, P. and Spehner, D. (2015) Live cell immunogold labelling of RNA polymerase II. *Sci. Rep.* **5**, 8324
- Orlova, E.V. and Saibil, H.R. (2010) Methods for three-dimensional reconstruction of heterogeneous assemblies. *Methods Enzymol.* **482**, 321–41
- Ortiz, J.O., Förster, F., Kürner, J., Linaroudis, A.A. and Baumeister, W. (2006) Mapping 70S ribosomes in intact cells by cryoelectron tomography and pattern recognition. *J. Struct. Biol.* **156**, 334–341
- Penczek, P.A., Frank, J. and Spahn, C.M. (2006) A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *J. Struct. Biol.* **154**, 184–194
- Rajendran, C., Dworkowski, F.S., Wang, M. and Schulze-Briese, C. (2011) Radiation damage in room-temperature data acquisition with the PILATUS 6M pixel detector. *J. Synchrotron Radiat.* **18**, 318–328
- Ray, P., Klaholz, B.P., Finn, R.D., Orlova, E.V., Burrows, P.C., Gowen, B., Buck, M. and van Heel, M. (2003) Determination of Escherichia coli RNA polymerase structure by single particle cryoelectron microscopy. *Methods Enzymol.* **370**, 24–42
- Razinkov, I., Dandey, V.P., Wei, H., Zhang, Z., Melnekoff, D., Rice, W.J., Wigge, C., Potter, C.S. and Carragher, B. (2016) A new method for vitrifying samples for cryoEM. *J. Struct. Biol.* **195**, 190–198
- Rhinow, D. (2016) Towards an optimum design for thin film phase plates. *Ultramicroscopy* **160**, 1–6
- Rigort, A. and Plitzko, J.M. (2015) Cryo-focused-ion-beam applications in structural biology. *Arch. Biochem. Biophys.* **581**, 122–130
- Ruskin, R.S., Yu, Z. and Grigorieff, N. (2013) Quantitative characterization of electron detectors for transmission electron microscopy. *J. Struct. Biol.* **184**, 385–393
- Sawaya, M.R., Rodriguez, J., Cascio, D., Collazo, M.J., Shi, D., Reyes, F.E., Hattne, J., Gonen, T. and Eisenberg, D.S. (2016) Ab initio structure determination from prion nanocrystals at atomic resolution by MicroED. *Proc. Natl. Acad. Sci. U.S.A.* **2016**. pii: 201606287
- Schaffer, M., Mahamid, J., Engel, B.D., Laugks, T., Baumeister, W. and Plitzko, J.M. (2016) Optimized cryo-focused ion beam sample preparation aimed at in situ structural studies of membrane proteins. *J. Struct. Biol.* **16**, 30151–30154
- Scheres, S.H., Valle, M., Nuez, R., Sorzano, C.O., Marabini, R., Herman, G.T. and Carazo, J.M. (2005) Maximum likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.* **22**, 139–149
- Scheres, S.H. (2010) Classification of structural heterogeneity by maximum-likelihood methods. *Methods Enzymol.* **482**, 295–320
- Scheres, S.H. (2014) Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife* **3**, e03665.
- Schirra, R.T. Jr. and Zhang, P. (2014) Correlative fluorescence and electron microscopy. *Curr. Protoc. Cytom.* **70**, 1–10
- Schorb, M., Gaechter, L., Avinoam, O., Sieckmann, F., Clarke, M., Bebeacua, C., Bykov, Y.S., Sonnen, A.F., Lihl, R. and Briggs, J.A. (2016) New hardware and workflows for semi-automated correlative cryo-fluorescence and cryo-electron microscopy/tomography. *J. Struct. Biol.* **16**, 30135–30136
- Schur, F.K., Obr, M., Hagen, W.J., Wan, W., Jakobi, A.J., Kirkpatrick, J.M., Sachse, C., Kräusslich, H.G. and Briggs, J.A. (2016) An

- atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506–508
- Shi, D., Nannenga, B.L., de la Cruz, M.J., Liu, J., Sawtelle, S., Calero, G., Reyes, F.E., Hattne, J. and Gonen, T. (2016) The collection of MicroED data for macromolecular crystallography. *Nat. Protoc.* **11**, 895–904
- Sigworth, F.J. (1998) A maximum-likelihood approach to single-particle image refinement. *J. Struct. Biol.* **122**, 328–339
- Simonetti, A., Marzi, S., Fabbretti, A., Hazemann, I., Jenner, L., Urzhumtsev, A., Gualerzi, C.O. and Klaholz, B.P. (2013a) Structure of the protein core of translation initiation factor 2 in apo, GTP-bound and GDP-bound forms. *Acta Cryst.* **D69**, 925–933
- Simonetti, A., Marzi, S., Billas, I.M.L., Tsai, A., Fabbretti, A., Myasnikov, A.G., Roblin, P., Vaiana, A.C., Hazemann, I., Eiler, D., Steitz, T.A., Puglisi, J.D., Gualerzi, C.O. and Klaholz, B.P. (2013b) Involvement of IF2 N domain in ribosomal subunit joining revealed from architecture and function of the full-length initiation factor. *Proc. Nat. Acad. Sci. U.S.A.* **110**, 15656–15661
- Simonetti, A., Ménétret, J-F., Martin, F., Myasnikov, A.G., Vicens, Q., Prongidi-Fix, L., Natchiar, S.K., Klaholz, B.P. and Eriani, G. (2016) Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. *Nat. Commun.* **7**, 12622
- Simonetti, A., Marzi, S., Myasnikov, A.G., Fabbretti, A., Yusupova, G., Yusupov, M., Gualerzi, C.O. and Klaholz, B.P. (2008) Structure of the 30S translation initiation complex. *Nature* **455**, 416–420
- Smart, O.S., Womack, T.O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vornheim, C. and Bricogne, G. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368–380
- Sobolev, O.V., Afonine, P.V., Adams, P.D. and Urzhumtsev, A. (2015) Programming new geometry restraints: parallelism of atomic groups. *J. Appl. Crystallogr.* **48**, 1130–1141
- Spear, J.M., Noble, A.J., Xie, Q., Sousa, D.R., Chapman, M.S. and Stagg, S.M. (2015) The influence of frame alignment with dose compensation on the quality of single particle reconstructions. *J. Struct. Biol.* **192**, 196–203
- Spinelli, S., Bebeacua C., Orlov, I., Tremblay, D., Klaholz, B.P., Moineau, S. and Cambillau, C. (2014) CryoEM structure of the lactococcal siphophage 1358 virion. *J. Virol.* **88**, 8900–8910
- Stölken, M., Beck, F., Haller, T., Hegerl, R., Gutsche, I., Carazo, J.M., Baumeister, W., Scheres, S.H. and Nickell, S. (2011) Maximum likelihood based classification of electron tomographic data. *J. Struct. Biol.* **173**, 77–85
- Szyzborska, A., de Marco, A., Daigle, N., Cordes, V.C., Briggs, J.A. and Ellenberg, J. (2013) Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* **341**, 655–658
- Tan, Y.Z., Cheng, A., Potter, C.S. and Carragher, B. (2016) Automated data collection in single particle electron microscopy. *Microscopy (Oxf.)* **65**, 43–56
- van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M. and Patwardhan, A. (2000) Single-particle electron cryo-microscopy: towards atomic resolution. *Q. Rev. Biophys.* **33**, 307–369
- Veesler, D., Campbell, M.G., Cheng, A., Fu, C.Y., Murez, Z., Johnson, J.E., Potter, C.S. and Carragher, B. (2013) Maximizing the potential of electron cryomicroscopy data collected using direct detectors. *J. Struct. Biol.* **184**, 193–202
- Villa, E., Schaffer, M., Plitzko, J.M. and Baumeister, W. (2013) Opening windows into the cell: focused-ion-beam milling for cryo-electron tomography. *Curr. Opin. Struct. Biol.* **23**, 771–777
- Walter, A., Steltenkamp, S., Schmitz, S., Holik, P., Pakanavicius, E., Sachser, R., Huth, M., Rhinow, D. and Kühlbrandt, W. (2015) Towards an optimum design for electrostatic phase plates. *Ultramicroscopy* **153**, 22–31
- Wan, W. and Briggs, J.A. (2016) Cryo-electron tomography and subtomogram averaging. *Methods Enzymol.* **579**, 329–367
- Wang, K., Fu, C.Y., Khayat, R., Doerschuk, P.C. and Johnson, J.E. (2011) In vivo virus structures: simultaneous classification, resolution enhancement, and noise reduction in whole-cell electron tomography. *J. Struct. Biol.* **174**, 425–433
- White, H.E., Saibil, H.R., Ignatiou, A. and Orlova, E.V. (2004) Recognition and separation of single particles with size variation by statistical analysis of their images. *J. Mol. Biol.* **13**, 453–460
- Wong, W., Bai, X.C., Brown, A., Fernandez, I.S., Hanssen, E., Condrón, M., Tan, Y.H., Baum, J. and Scheres, S.H. (2014) Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *Elife* **3**. doi: 10.7554/eLife.03080
- Xu, M. and Alber, F. (2013) Automated target segmentation and real space fast alignment methods for high-throughput classification and averaging of crowded cryo-electron subtomograms. *Bioinformatics* **29**, 274–282
- Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C.C., Pettersen, E.F., Goddard, T.D., Meng, E.C., Sali, A. and Ferrin, T.E. (2012) UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *J. Struct. Biol.* **179**, 269–278
- Zhang, J., Ji, G., Huang, X., Xu, W. and Sun, F. (2016) An improved cryo-FIB method for fabrication of frozen hydrated lamella. *J. Struct. Biol.* **194**, 218–223
- Zhang, X., Ge, P., Yu, X., Brannan, J.M., Bi, G., Zhang, Q., Schein, S. and Zhou, Z.H. (2013) Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. *Nat. Struct. Mol. Biol.* **20**, 105–110

Received: 12 July 2016; Revised: 4 October 2016; Accepted: 5 October 2016; Accepted article online: 12 October 2016





## Références

- Allard, S., Utley, R.T., Savard, J., Clarke, A., Grant, P., Brandl, C.J., Pillus, L., Workman, J.L., Côté, J., 1999. NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. *EMBO J.* 18, 5108–5119.
- Allen, H.F., Wade, P.A., Kutateladze, T.G., 2013. The NuRD architecture. *Cell. Mol. Life Sci.* 70, 3513–3524.
- Allfrey, V.G., Mirsky, A.E., 1964. Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis. *Science* 144, 559.
- Alqarni, S.S.M., Murthy, A., Zhang, W., Przewloka, M.R., Silva, A.P.G., Watson, A.A., Lejon, S., Pei, X.Y., Smits, A.H., Kloet, S.L., Wang, H., Shepherd, N.E., Stokes, P.H., Blobel, G.A., Vermeulen, M., Glover, D.M., Mackay, J.P., Laue, E.D., 2014. Insight into the Architecture of the NuRD Complex: STRUCTURE OF THE RbAp48-MTA1 SUBCOMPLEX. *Journal of Biological Chemistry* 289, 21844–21855.
- Ambrosi, C., Manzo, M., Baubec, T., 2017. Dynamics and Context-Dependent Roles of DNA Methylation. *J. Mol. Biol.* 429, 1459–1475.
- Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., Zoghbi, H.Y., 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* 23, 185–188.
- Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., Shirakawa, M., 2008. Recognition of hemimethylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* 455, 818–821.
- Armache, K.-J., Garlick, J.D., Canzio, D., Narlikar, G.J., Kingston, R.E., 2011. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* 334, 977–982.
- Ayer, D.E., 1999. Histone deacetylases: transcriptional repression with SINers and NuRDs. *Trends Cell Biol.* 9, 193–198.
- Baker, S.A., Chen, L., Wilkins, A.D., Yu, P., Lichtarge, O., Zoghbi, H.Y., 2013. An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. *Cell* 152, 984–996.
- Banting, G.S., Barak, O., Ames, T.M., Burnham, A.C., Kardel, M.D., Cooch, N.S., Davidson, C.E., Godbout, R., McDermid, H.E., Shiekhhattar, R., 2005. CECR2, a protein involved in neurulation, forms a novel chromatin remodeling complex with SNF2L. *Hum. Mol. Genet.* 14, 513–524.
- Barak, O., Lazzaro, M.A., Lane, W.S., Speicher, D.W., Picketts, D.J., Shiekhhattar, R., 2003. Isolation of human NURF: a regulator of Engrailed gene expression. *EMBO J* 22, 6089–6100.
- Bartesaghi, A., Aguerrebere, C., Falconieri, V., Banerjee, S., Earl, L.A., Zhu, X., Grigorieff, N., Milne, J.L.S., Sapiro, G., Wu, X., Subramaniam, S., 2018. Atomic Resolution Cryo-EM Structure of  $\beta$ -Galactosidase. *Structure* 26, 848-856.e3.
- Basta, J., Rauchman, M., 2015. The nucleosome remodeling and deacetylase complex in development and disease. *Transl Res* 165, 36–47.
- Baubec, T., Ivánek, R., Lienert, F., Schübeler, D., 2013. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* 153, 480–492.
- Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A.R., Papai, G., Reymer, A., Syed, S.H., Lone, I.N., Tonchev, O., Crucifix, C., Menoni, H., Papin, C., Skoufias, D.A., Kurumizaka, H., Lavery, R., Hamiche, A., Hayes, J.J., Schultz, P., Angelov, D., Petosa, C., Dimitrov, S., 2017. Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Molecular Cell* 66, 384-397.e8.

- Bergs, J.W., Neuendorff, N., van der Heijden, G., Wassenaar, E., Rexin, P., Elsässer, H.-P., Moll, R., Baarends, W.M., Brehm, A., 2014. Differential expression and sex chromosome association of CHD3/4 and CHD5 during spermatogenesis. *PLoS ONE* 9, e98203.
- Biertümpfel, C., Basquin, J., Suck, D., Sauter, C., 2002. Crystallization of biological macromolecules using agarose gel. *Acta Crystallographica Section D Biological Crystallography* 58, 1657–1659.
- Bostick, M., Kim, J.K., Esteve, P.-O., Clark, A., Pradhan, S., Jacobsen, S.E., 2007. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science* 317, 1760–1764.
- Boyer, L.A., Latek, R.R., Peterson, C.L., 2004. The SANT domain: a unique histone-tail-binding module? *Nat. Rev. Mol. Cell Biol.* 5, 158–163. Brachmann, C.B., Sherman, J.M., Devine, S.E., Cameron, E.E., Pillus, L., Boeke, J.D., 1995. The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability. *Genes Dev.* 9, 2888–2902.
- Brackertz, M., Boeke, J., Zhang, R., Renkawitz, R., 2002. Two highly related p66 proteins comprise a new family of potent transcriptional repressors interacting with MBD2 and MBD3. *J. Biol. Chem.* 277, 40958–40966.
- Brackertz, M., Gong, Z., Leers, J., Renkawitz, R., 2006. p66alpha and p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction. *Nucleic Acids Res.* 34, 397–406.
- Brahma, S., Udugama, M.I., Kim, J., Hada, A., Bhardwaj, S.K., Hailu, S.G., Lee, T.-H., Bartholomew, B., 2017. INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers. *Nat Commun* 8, 15616.
- Brakke, M.K., 1951. Density Gradient Centrifugation: A New Separation Technique <sup>1</sup>. *Journal of the American Chemical Society* 73, 1847–1848.
- Brasen, C., Dorosz, J., Wiuf, A., Boesen, T., Mirza, O., Gajhede, M., 2017. Expression, purification and characterization of the human MTA2-RBBP7 complex. *Biochim. Biophys. Acta* 1865, 531–538. Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., Allis, C.D., 1996. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* 84, 843–851.
- Brunmeir, R., Lagger, S., Seiser, C., 2009. Histone deacetylase HDAC1/HDAC2-controlled embryonic development and cell differentiation. *Int. J. Dev. Biol.* 53, 275–289.
- Buck-Koehntop, B.A., Martinez-Yamout, M.A., Dyson, H.J., Wright, P.E., 2012. Kaiso uses all three zinc fingers and adjacent sequence motifs for high affinity binding to sequence-specific and methyl-CpG DNA targets. *FEBS Lett.* 586, 734–739.
- Callebaut, I., Courvalin, J.C., Mornon, J.P., 1999. The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. *FEBS Lett.* 446, 189–193.
- Cao, L.-L., Song, X., Pei, L., Liu, L., Wang, H., Jia, M., 2017. Histone deacetylase HDAC1 expression correlates with the progression and prognosis of lung cancer. *Medicine (Baltimore)* 96.
- Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T.C., Qin, J., Zoghbi, H.Y., 2008. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* 320, 1224–1229.
- Chia, J.Y., Tan, W.S., Ng, C.L., Hu, N.-J., Foo, H.L., Ho, K.L., 2016. A/T Run Geometry of B-form DNA Is Independent of Bound Methyl-CpG Binding Domain, Cytosine Methylation and Flanking Sequence. *Scientific Reports* 6.
- Chiu, P.-L., Kelly, D.F., Walz, T., 2011. The use of trehalose in the preparation of specimens for molecular electron microscopy. *Micron* 42, 762–772.
- Clapier, C.R., Cairns, B.R., 2009. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* 78, 273–304.

- Clapier, C.R., Iwasa, J., Cairns, B.R., Peterson, C.L., 2017. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* 18, 407–422.
- Clouaire, T., de las Heras, J.I., Merusi, C., Stancheva, I., 2010. Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic Acids Research* 38, 4620–4634.
- Corona, D.F.V., Clapier, C.R., Becker, P.B., Tamkun, J.W., 2002. Modulation of ISWI function by site-specific histone acetylation. *EMBO Rep.* 3, 242–247.
- Cramer, J.M., Scarsdale, J.N., Walavalkar, N.M., Buchwald, W.A., Ginder, G.D., Williams, D.C., 2014. Probing the Dynamic Distribution of Bound States for Methylcytosine-binding Domains on DNA. *J Biol Chem* 289, 1294–1302.
- Cukier, H.N., Lee, J.M., Ma, D., Young, J.I., Mayo, V., Butler, B.L., Ramsook, S.S., Rantus, J.A., Abrams, A.J., Whitehead, P.L., Wright, H.H., Abramson, R.K., Haines, J.L., Cuccaro, M.L., Pericak-Vance, M.A., Gilbert, J.R., 2012. The expanding role of MBD genes in autism: identification of a MECP2 duplication and novel alterations in MBD5, MBD6, and SETDB1. *Autism Res* 5, 385–397.
- De Carlo, Adrian, Kälin, Mayer, Dubochet, 1999. Unexpected property of trehalose as observed by cryo-electron microscopy: unexpected property of trehalose as observed by cryo-electron microscopy. *Journal of Microscopy* 196, 40–45.
- De Meyer, T., Mampaey, E., Vlemmix, M., Denil, S., Trooskens, G., Renard, J.-P., De Keulenaer, S., Dehan, P., Menschaert, G., Van Criekinge, W., 2013. Quality evaluation of methyl binding domain based kits for enrichment DNA-methylation sequencing. *PLoS ONE* 8, e59068.
- Deans, C., Maggert, K.A., 2015. What do you mean, “epigenetic”? *Genetics* 199, 887–896.
- Denslow, S.A., Wade, P.A., 2007. The human Mi-2/NuRD complex and gene regulation. *Oncogene* 26, 5433–5438.
- Desai, M.A., Webb, H.D., Sinanan, L.M., Scarsdale, J.N., Walavalkar, N.M., Ginder, G.D., Williams, D.C., 2015. An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. *Nucleic Acids Res.* 43, 3100–3113.
- dos Santos, R.L., Tosti, L., Radzishewska, A., Caballero, I.M., Kaji, K., Hendrich, B., Silva, J.C.R., 2014. MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. *Cell Stem Cell* 15, 102–110.
- Dowling, D.P., Gantt, S.L., Gattis, S.G., Fierke, C.A., Christianson, D.W., 2008. Structural Studies of Human Histone Deacetylase 8 and Its Site-Specific Variants Complexed with Substrate and Inhibitors <sup>†</sup> · <sup>‡</sup>. *Biochemistry* 47, 13554–13563.
- Dyballa, N., Metzger, S., 2009. Fast and sensitive colloidal coomassie G-250 staining for proteins in polyacrylamide gels. *J Vis Exp*.
- Eustermann, S., Schall, K., Kostrewa, D., Lakomek, K., Strauss, M., Moldt, M., Hopfner, K.-P., 2018. Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. *Nature* 556, 386–390.
- Feng, Q., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., Zhang, Y., 2002. Identification and functional characterization of the p66/p68 components of the MeCP1 complex. *Mol. Cell Biol.* 22, 536–546.
- Finnin, M.S., Donigian, J.R., Cohen, A., Richon, V.M., Rifkind, R.A., Marks, P.A., Breslow, R., Pavletich, N.P., 1999. Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* 401, 188–193.
- Fischer, D.D., Cai, R., Bhatia, U., Asselbergs, F.A.M., Song, C., Terry, R., Trogani, N., Widmer, R., Atadja, P., Cohen, D., 2002. Isolation and characterization of a novel class II histone deacetylase, HDAC10. *J. Biol. Chem.* 277, 6656–6666.
- Forneris, F., Binda, C., Battaglioli, E., Mattevi, A., 2008. LSD1: oxidative chemistry for multifaceted functions in chromatin regulation. *Trends Biochem. Sci.* 33, 181–189.

- Fraga, M.F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P.A., Esteller, M., 2003. The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res.* 31, 1765–1774.
- Frye, R.A., 1999. Characterization of five human cDNAs with homology to the yeast SIR2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity. *Biochem. Biophys. Res. Commun.* 260, 273–279.
- Fu, J., Qin, L., He, T., Qin, J., Hong, J., Wong, J., Liao, L., Xu, J., 2011. The TWIST/Mi2/NuRD protein complex and its essential role in cancer metastasis. *Cell Research* 21, 275–289.
- Gao, L., Cueto, M.A., Asselbergs, F., Atadja, P., 2002. Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family. *J. Biol. Chem.* 277, 25748–25755.
- Gnanapragasam, M.N., Scarsdale, J.N., Amaya, M.L., Webb, H.D., Desai, M.A., Walavalkar, N.M., Wang, S.Z., Zu Zhu, S., Ginder, G.D., Williams, D.C., 2011. p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7487–7492.
- Gong, F., Clouaire, T., Aguirrebengoa, M., Legube, G., Miller, K.M., 2017. Histone demethylase KDM5A regulates the ZMYND8–NuRD chromatin remodeler to promote DNA repair. *J Cell Biol* 216, 1959–1974.
- Gong, Z., Brackertz, M., Renkawitz, R., 2006. SUMO Modification Enhances p66-Mediated Transcriptional Repression of the Mi-2/NuRD Complex. *Molecular and Cellular Biology* 26, 4519–4528.
- Gregorette, I.V., Lee, Y.-M., Goodson, H.V., 2004. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. *J. Mol. Biol.* 338, 17–31.
- Grozinger, C.M., Hassig, C.A., Schreiber, S.L., 1999. Three proteins define a class of human histone deacetylases related to yeast Hda1p. *Proc. Natl. Acad. Sci. U.S.A.* 96, 4868–4873.
- Günther, K., Rust, M., Leers, J., Boettger, T., Scharfe, M., Jarek, M., Bartkuhn, M., Renkawitz, R., 2013. Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. *Nucleic Acids Res.* 41, 3010–3021.
- Hakimi, M.-A., Bochar, D.A., Schmiesing, J.A., Dong, Y., Barak, O.G., Speicher, D.W., Yokomori, K., Shiekhhattar, R., 2002. A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature* 418, 994–998.
- Harms, K.L., Chen, X., 2007. Histone deacetylase 2 modulates p53 transcriptional activities through regulation of p53-DNA binding activity. *Cancer Res.* 67, 3145–3152.
- Hashimoto, H., Horton, J.R., Zhang, X., Bostick, M., Jacobsen, S.E., Cheng, X., 2008. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* 455, 826–829.
- Hashimoto, H., Zhang, X., Cheng, X., 2012. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res* 40, 8276–8284.
- Hauer, F., Gerle, C., Fischer, N., Oshima, A., Shinzawa-Itoh, K., Shimada, S., Yokoyama, K., Fujiyoshi, Y., Stark, H., 2015. GraDeR: Membrane Protein Complex Preparation for Single-Particle Cryo-EM. *Structure* 23, 1769–1775.
- He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., Xu, G.-L., 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307.
- Hendrich, B., Bird, A., 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* 18, 6538–6547.
- Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., Bird, A., 2001. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev* 15, 710–723.

- Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J., Bird, A., 1999. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* 401, 301–304.
- Hendrich, B., Tweedie, S., 2003. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* 19, 269–277.
- Ho, K.L., McNae, I.W., Schmiedeberg, L., Klose, R.J., Bird, A.P., Walkinshaw, M.D., 2008. MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell* 29, 525–531.
- Hoffman, E.A., Frey, B.L., Smith, L.M., Auble, D.T., 2015. Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J. Biol. Chem.* 290, 26404–26411.
- Hoffmeister, H., Fuchs, A., Erdel, F., Pinz, S., Gröbner-Ferreira, R., Bruckmann, A., Deutzmann, R., Schwartz, U., Maldonado, R., Huber, C., Dendorfer, A.-S., Rippe, K., Längst, G., 2017. CHD3 and CHD4 form distinct NuRD complexes with different yet overlapping functionality. *Nucleic Acids Research* 45, 10534–10554.
- Hotchkiss, R.D., 1948. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.* 175, 315–332.
- Huang, J., Sengupta, R., Espejo, A.B., Lee, M.G., Dorsey, J.A., Richter, M., Opravil, S., Shiekhatar, R., Bedford, M.T., Jenuwein, T., Berger, S.L., 2007. p53 is regulated by the lysine demethylase LSD1. *Nature* 449, 105–108.
- Inoue, A., Fujimoto, D., 1969. Enzymatic deacetylation of histone. *Biochem. Biophys. Res. Commun.* 36, 146–150.
- Jain, T., Sheehan, P., Crum, J., Carragher, B., Potter, C.S., 2012. Spotiton: A prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *Journal of Structural Biology* 179, 68–75.
- Javaid, N., Choi, S., 2017. Acetylation- and Methylation-Related Epigenetic Proteins in the Context of Their Targets. *Genes (Basel)* 8.
- Jørgensen, H.F., Ben-Porath, I., Bird, A.P., 2004. Mbd1 Is Recruited to both Methylated and Nonmethylated CpGs via Distinct DNA Binding Domains. *Molecular and Cellular Biology* 24, 3387–3395.
- Kaczmarek, Z., Ortega, E., Goudarzi, A., Huang, H., Kim, S., Márquez, J.A., Zhao, Y., Khochbin, S., Panne, D., 2017. Structure of p300 in complex with acyl-CoA variants. *Nature Chemical Biology* 13, 21–29.
- Kahl, P., Gullotti, L., Heukamp, L.C., Wolf, S., Friedrichs, N., Vorreuther, R., Solleder, G., Bastian, P.J., Ellinger, J., Metzger, E., Schüle, R., Buettner, R., 2006. Androgen receptor coactivators lysine-specific histone demethylase 1 and four and a half LIM domain protein 2 predict risk of prostate cancer recurrence. *Cancer Res.* 66, 11341–11347.
- Kaji, K., Nichols, J., Hendrich, B., 2007. Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. *Development* 134, 1123–1132.
- Kao, H.-Y., Lee, C.-H., Komarov, A., Han, C.C., Evans, R.M., 2002. Isolation and characterization of mammalian HDAC10, a novel histone deacetylase. *J. Biol. Chem.* 277, 187–193.
- Kapoor, P., Chen, M., Winkler, D.D., Luger, K., Shen, X., 2013. Evidence for monomeric actin function in INO80 chromatin remodeling. *Nat Struct Mol Biol* 20, 426–432.
- Kapoor, P., Shen, X., 2014. Mechanisms of Nuclear Actin in Chromatin Remodeling Complexes. *Trends Cell Biol* 24, 238–246.
- Kashiwagi, M., Morgan, B.A., Georgopoulos, K., 2007. The chromatin remodeler Mi-2beta is required for establishment of the basal epidermis and normal differentiation of its progeny. *Development* 134, 1571–1582.
- Kelly, R.D.W., Cowley, S.M., 2013. The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. *Biochem. Soc. Trans.* 41, 741–749.
- Kim, J., Sif, S., Jones, B., Jackson, A., Koipally, J., Heller, E., Winandy, S., Viel, A., Sawyer, A., Ikeda, T., Kingston, R., Georgopoulos, K., 1999. Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity* 10, 345–355.

- Kloet, S.L., Baymaz, H.I., Makowski, M., Groenewold, V., Jansen, P.W.T.C., Berendsen, M., Niazi, H., Kops, G.J., Vermeulen, M., 2015. Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *FEBS Journal* 282, 1774–1785.
- Klose, R.J., Sarraf, S.A., Schmiedeberg, L., McDermott, S.M., Stancheva, I., Bird, A.P., 2005. DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell* 19, 667–678.
- Koh, D.-I., Han, D., Ryu, H., Choi, W.-I., Jeon, B.-N., Kim, M.-K., Kim, Y., Kim, J.Y., Parry, L., Clarke, A.R., Reynolds, A.B., Hur, M.-W., 2014. KAISO, a critical regulator of p53-mediated transcription of CDKN1A and apoptotic genes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15078–15083.
- Kuo, M.H., Allis, C.D., 1998. Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* 20, 615–626.
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., Reinberg, D., 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* 16, 2893–2905.
- Kwan, A.H.Y., Gell, D.A., Verger, A., Crossley, M., Matthews, J.M., Mackay, J.P., 2003. Solution structure of the 2nd PHD domain from Mi2b with C-terminal loop replaced by corresponding loop from WSTF.
- Lai, A.Y., Wade, P.A., 2011. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat. Rev. Cancer* 11, 588–596.
- Lauberth, S.M., Rauchman, M., 2006. A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex. *J. Biol. Chem.* 281, 23922–23931.
- Lauffer, B.E.L., Mintzer, R., Fong, R., Mukund, S., Tam, C., Zilberleyb, I., Flicke, B., Ritscher, A., Fedorowicz, G., Vallero, R., Ortwine, D.F., Gunzner, J., Modrusan, Z., Neumann, L., Koth, C.M., Lupardus, P.J., Kaminker, J.S., Heise, C.E., Steiner, P., 2013. Histone Deacetylase (HDAC) Inhibitor Kinetic Rate Constants Correlate with Cellular Histone Acetylation but Not Transcription and Cell Viability. *Journal of Biological Chemistry* 288, 26926–26943.
- Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A.M., Cohen, A., Lasonder, E., Stunnenberg, H.G., 2006. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol. Cell. Biol.* 26, 843–851.
- Lee, M.G., Wynder, C., Bochar, D.A., Hakimi, M.-A., Cooch, N., Shiekhatter, R., 2006a. Functional Interplay between Histone Demethylase and Deacetylase Enzymes. *Molecular and Cellular Biology* 26, 6395–6402.
- Lee, M.G., Wynder, C., Bochar, D.A., Hakimi, M.-A., Cooch, N., Shiekhatter, R., 2006b. Functional interplay between histone demethylase and deacetylase enzymes. *Mol. Cell. Biol.* 26, 6395–6402.
- Lee, M.G., Wynder, C., Cooch, N., Shiekhatter, R., 2005. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* 437, 432–435.
- Leipe, D.D., Landsman, D., 1997. Histone deacetylases, acetoin utilization proteins and acetylpolyamine amidohydrolases are members of an ancient protein superfamily. *Nucleic Acids Res.* 25, 3693–3697.
- Lejon, S., Thong, S.Y., Murthy, A., AlQarni, S., Murzina, N.V., Blobel, G.A., Laue, E.D., Mackay, J.P., 2011. Insights into Association of the NuRD Complex with FOG-1 from the Crystal Structure of an RbAp48-FOG-1 Complex. *Journal of Biological Chemistry* 286, 1196–1203.
- Leo, C., Chen, J.D., 2000. The SRC family of nuclear receptor coactivators. *Gene* 245, 1–11.

- Li, B., Carey, M., Workman, J.L., 2007. The role of chromatin during transcription. *Cell* 128, 707–719.
- Liu, K., Xu, C., Lei, M., Yang, A., Loppnau, P., Hughes, T.R., Min, J., 2018a. Structural basis for the ability of MBD domains to bind methyl-CG and TG sites in DNA. *J. Biol. Chem.* 293, 7344–7354
- Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y., Xue, Y., Ren, J., 2015. IBS: an illustrator for the presentation and visualization of biological sequences: Fig. 1. *Bioinformatics* 31, 3359–3361.
- Lorch, Y., Zhang, M., Kornberg, R.D., 1999. Histone octamer transfer by a chromatin-remodeling complex. *Cell* 96, 389–392.
- Low, J.K.K., Webb, S.R., Silva, A.P.G., Saathoff, H., Ryan, D.P., Torrado, M., Brofelth, M., Parker, B.L., Shepherd, N.E., Mackay, J.P., 2016. CHD4 Is a Peripheral Component of the Nucleosome Remodeling and Deacetylase Complex. *J. Biol. Chem.* 291, 15853–15866.
- Lyst, M.J., Connelly, J., Merusi, C., Bird, A., 2016. Sequence-specific DNA binding by AT-hook motifs in MeCP2. *FEBS Letters* 590, 2927–2933.
- Ma, P., Schultz, R.M., 2016. HDAC1 and HDAC2 in mouse oocytes and preimplantation embryos: Specificity versus compensation. *Cell Death Differ* 23, 1119–1127.
- Mansfield, R.E., Musselman, C.A., Kwan, A.H., Oliver, S.S., Garske, A.L., Davrazou, F., Denu, J.M., Kutateladze, T.G., Mackay, J.P., 2011. Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J. Biol. Chem.* 286, 11779–11791.
- Marmorstein, R., Trievel, R.C., 2009. Histone modifying enzymes: structures, mechanisms, and specificities. *Biochim. Biophys. Acta* 1789, 58–68.
- Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., Bird, A.P., 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 58, 499–507.
- Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M.I., Pragani, R., Boxer, M.B., Earl, L.A., Milne, J.L.S., Subramaniam, S., 2016. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* 165, 1698–1707.
- Millard, C.J., Fairall, L., Schwabe, J.W.R., 2014. Towards an understanding of the structure and function of MTA1. *Cancer Metastasis Rev* 33, 857–867.
- Millard, C.J., Varma, N., Saleh, A., Morris, K., Watson, P.J., Bottrill, A.R., Fairall, L., Smith, C.J., Schwabe, J.W., 2016. The structure of the core NuRD repression complex provides insights into its interaction with chromatin [WWW Document]. *eLife*.
- Millard, C.J., Watson, P.J., Celardo, I., Gordiyenko, Y., Cowley, S.M., Robinson, C.V., Fairall, L., Schwabe, J.W.R., 2013. Class I HDACs Share a Common Mechanism of Regulation by Inositol Phosphates. *Molecular Cell* 51, 57–67.
- Mohd-Sarip, A., Teeuwssen, M., Bot, A.G., Herdt, M.J.D., Willems, S.M., Jong, R.J.B. de, Looijenga, L.H.J., Zatreanu, D., Bezstarosti, K., Riet, J. van, Oole, E., Ijcken, W.F.J. van, Werken, H.J.G. van de, Demmers, J.A., Fodde, R., Verrijzer, C.P., 2017. DOC1-Dependent Recruitment of NURD Reveals Antagonism with SWI/SNF during Epithelial-Mesenchymal Transition in Oral Cancer Cells. *Cell Reports* 20, 61–75.
- Mori, T., Li, Y., Hata, H., Ono, K., Kochi, H., 2002. NIRF, a novel RING finger protein, is involved in cell-cycle regulation. *Biochem. Biophys. Res. Commun.* 296, 530–536.
- Murzina, N.V., Pei, X.-Y., Zhang, W., Sparkes, M., Vicente-Garcia, J., Pratap, J.V., McLaughlin, S.H., Ben-Shahar, T.R., Verreault, A., Luisi, B.F., Laue, E.D., 2008. Structural Basis for the Recognition of Histone H4 by the Histone-Chaperone RbAp46. *Structure* 16, 1077–1085.
- Musselman, C.A., Mansfield, R.E., Garske, A.L., Davrazou, F., Kwan, A.H., Oliver, S.S., O’Leary, H., Denu, J.M., Mackay, J.P., Kutateladze, T.G., 2009. Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem. J.* 423, 179–187.

- Musselman, C.A., Ramírez, J., Sims, J.K., Mansfield, R.E., Oliver, S.S., Denu, J.M., Mackay, J.P., Wade, P.A., Hagman, J., Kutateladze, T.G., 2012. Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. *Proc. Natl. Acad. Sci. U.S.A.* 109, 787–792.
- Nair, S.S., Li, D.-Q., Kumar, R., 2013. A core chromatin remodeling factor instructs global chromatin signaling through multivalent reading of nucleosome codes. *Mol. Cell* 49, 704–718.
- Nan, X., Meehan, R.R., Bird, A., 1993. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Research* 21, 4886–4892.
- Natchiar, S.K., Myasnikov, A.G., Kratzat, H., Hazemann, I., Klaholz, B.P., 2017. Visualization of chemical modifications in the human 80S ribosome structure. *Nature*.
- Nitarska, J., Smith, J.G., Sherlock, W.T., Hillege, M.M.G., Nott, A., Barshop, W.D., Vashisht, A.A., Wohlschlegel, J.A., Mitter, R., Riccio, A., 2016. A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. *Cell Rep* 17, 1683–1698.
- Oh, Y.M., Kwon, Y.E., Kim, J.M., Bae, S.J., Lee, B.K., Yoo, S.J., Chung, C.H., Deshaies, R.J., Seol, J.H., 2009. Chfr is linked to tumour metastasis through the downregulation of HDAC1. *Nat. Cell Biol.* 11, 295–302.
- Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., Shirakawa, M., 2001. Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* 105, 487–497.
- Orlov, I., Myasnikov, A.G., Andronov, L., Natchiar, S.K., Khatter, H., Beinsteiner, B., Ménétret, J.-F., Hazemann, I., Mohideen, K., Tazibt, K., Tabaroni, R., Kratzat, H., Djabeur, N., Bruxelles, T., Raivoniaina, F., Pompeo, L. di, Torchy, M., Billas, I., Urzhumtsev, A., Klaholz, B.P., 2017. The integrative role of cryo electron microscopy in molecular and cellular structural biology: Integrative role of cryo-EM in structural biology. *Biology of the Cell* 109, 81–93.
- Otani, J., Arita, K., Kato, T., Kinoshita, M., Kimura, H., Suetake, I., Tajima, S., Ariyoshi, M., Shirakawa, M., 2013. Structural Basis of the Versatile DNA Recognition Ability of the Methyl-CpG Binding Domain of Methyl-CpG Binding Domain Protein 4. *J. Biol. Chem.* 288, 6351–6362.
- Pantelic, R.S., Meyer, J.C., Kaiser, U., Baumeister, W., Plitzko, J.M., 2010. Graphene oxide: A substrate for optimizing preparations of frozen-hydrated samples. *Journal of Structural Biology* 170, 152–156.
- Parthun, M.R., Widom, J., Gottschling, D.E., 1996. The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. *Cell* 87, 85–94.
- Prokhortchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., Prokhortchouk, E., 2001. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev* 15, 1613–1618.
- Qian, Y.W., Wang, Y.C., Hollingsworth, R.E., Jones, D., Ling, N., Lee, E.Y., 1993. A retinoblastoma-binding protein related to a negative regulator of Ras in yeast. *Nature* 364, 648–652.
- Qian, Y.-Y., Liu, Z.-S., Pan, D.-Y., Li, K., 2017. Tumoricidal activities of pterostilbene depend upon destabilizing the MTA1-NuRD complex and enhancing P53 acetylation in hepatocellular carcinoma. *Exp Ther Med* 14, 3098–3104.
- Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., Maza, I., Mor, N., Baran, D., Weinberger, L., Jaitin, D.A., Lara-Astiaso, D., Blecher-Gonen, R., Shipony, Z., Mukamel, Z., Hagai, T., Gilad, S., Amann-

- Zalcenstein, D., Tanay, A., Amit, I., Novershtern, N., Hanna, J.H., 2013. Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* 502, 65–70.
- Reddy, B.A., Bajpe, P.K., Bassett, A., Moshkin, Y.M., Kozhevnikova, E., Bezstarosti, K., Demmers, J.A.A., Travers, A.A., Verrijzer, C.P., 2010. Drosophila Transcription Factor Tramtrack69 Binds MEP1 To Recruit the Chromatin Remodeler NuRD. *Molecular and Cellular Biology* 30, 5234–5244.
- Reichert, N., Choukrallah, M.-A., Matthias, P., 2012. Multiple roles of class I HDACs in proliferation, differentiation, and development. *Cell. Mol. Life Sci.* 69, 2173–2187.
- Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O’Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., Costello, I., Humphreys, P., Mansfield, W., Nakagawa, K., Strouboulis, J., Behrens, A., Bertone, P., Hendrich, B., 2012. NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* 10, 583–594.
- Roche, A.E., Bassett, B.J., Samant, S.A., Hong, W., Blobel, G.A., Svensson, E.C., 2008. The zinc finger and C-terminal domains of MTA proteins are required for FOG-2-mediated transcriptional repression via the NuRD complex. *Journal of Molecular and Cellular Cardiology* 44, 352–360.
- Sarraf, S.A., Stancheva, I., 2004. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell* 15, 595–605.
- Sasai, N., Defossez, P.-A., 2009. Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. *Int. J. Dev. Biol.* 53, 323–334.
- Schmidberger, J.W., Sharifi Tabar, M., Torrado, M., Silva, A.P.G., Landsberg, M.J., Brillault, L., AlQarni, S., Zeng, Y.C., Parker, B.L., Low, J.K.K., Mackay, J.P., 2016. The MTA1 subunit of the nucleosome remodeling and deacetylase complex can recruit two copies of RBBP4/7. *Protein Sci.* 25, 1472–1482.
- Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., Bunker, R.D., Wirth, U., Bouwmeester, T., Bauer, A., Ly-Hartig, N., Zhao, K., Chan, H., Gu, J., Gut, H., Fischle, W., Müller, J., Thomä, N.H., 2011. Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell* 42, 330–341.
- Schuetz, A., Min, J., Allali-Hassani, A., Schapira, M., Shuen, M., Loppnau, P., Mazitschek, R., Kwiatkowski, N.P., Lewis, T.A., Maglathin, R.L., McLean, T.H., Bochkarev, A., Plotnikov, A.N., Vedadi, M., Arrowsmith, C.H., 2008. Human HDAC7 Harbors a Class IIa Histone Deacetylase-specific Zinc Binding Motif and Cryptic Deacetylase Activity. *Journal of Biological Chemistry* 283, 11355–11363.
- Seelig, H.P., Renz, M., Targoff, I.N., Ge, Q., Frank, M.B., 1996. Two forms of the major antigenic protein of the dermatomyositis-specific Mi-2 autoantigen. *Arthritis Rheum.* 39, 1769–1771.
- Segré, C.V., Chiocca, S., 2011a. Regulating the Regulators: The Post-Translational Code of Class I HDAC1 and HDAC2. *J Biomed Biotechnol* 2011.
- Serebryanny, L.A., Cruz, C.M., de Lanerolle, P., 2016. A Role for Nuclear Actin in HDAC 1 and 2 Regulation. *Sci Rep* 6.
- Seto, E., Yoshida, M., 2014. Erasers of Histone Acetylation: The Histone Deacetylase Enzymes. *Cold Spring Harb Perspect Biol* 6.
- Shahbazian, M.D., Zoghbi, H.Y., 2002. Rett syndrome and MeCP2: linking epigenetics and neuronal function. *Am. J. Hum. Genet.* 71, 1259–1272.
- Shen, E., Wang, Q., Rabe, H., Liu, W., Cantor, H., Leavenworth, J.W., 2018. Chromatin remodeling by the NuRD complex regulates development of follicular helper and regulatory T cells. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6780–6785.
- Shen, X., Mizuguchi, G., Hamiche, A., Wu, C., 2000. A chromatin remodelling complex involved in transcription and DNA processing. *Nature* 406, 541–544.

- Shi, Y., 2007. Histone lysine demethylases: emerging roles in development, physiology and disease. *Nat. Rev. Genet.* 8, 829–833.
- Shi, Y.-J., Matson, C., Lan, F., Iwase, S., Baba, T., Shi, Y., 2005. Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell* 19, 857–864.
- Shi, Yujiang, Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., Shi, Yang, 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119, 941–953.
- Shimbo, T., Du, Y., Grimm, S.A., Dhasarathy, A., Mav, D., Shah, R.R., Shi, H., Wade, P.A., 2013. MBD3 localizes at promoters, gene bodies and enhancers of active genes. *PLoS Genet.* 9, e1004028.
- Shimbo, T., Takaku, M., Wade, P.A., 2016. High-quality ChIP-seq analysis of MBD3 in human breast cancer cells. *Genom Data* 7, 173–174.
- Singh, S.K., Sigworth, F., 2015. Cryo-EM: spinning the micelles away. *Structure* 23, 1561.
- Smith, Z.D., Meissner, A., 2013. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* 14, 204–220.
- Song, J., Teplova, M., Ishibe-Murakami, S., Patel, D.J., 2012. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* 335, 709–712.
- Spruijt, C.G., Bartels, S.J.J., Brinkman, A.B., Tjeertes, J.V., Poser, I., Stunnenberg, H.G., Vermeulen, M., 2010. CDK2AP1/DOC-1 is a bona fide subunit of the Mi-2/NuRD complex. *Mol. BioSyst.* 6, 1700–1706.
- Spruijt, C.G., Luijsterburg, M.S., Menafrá, R., Lindeboom, R.G.H., Jansen, P.W.T.C., Edupuganti, R.R., Baltissen, M.P., Wiegant, W.W., Voelker-Albert, M.C., Matarese, F., Mensinga, A., Poser, I., Vos, H.R., Stunnenberg, H.G., van Attikum, H., Vermeulen, M., 2016. ZMYND8 Co-localizes with NuRD on Target Genes and Regulates Poly(ADP-Ribose)-Dependent Recruitment of GATAD2A/NuRD to Sites of DNA Damage. *Cell Reports* 17, 783–798.
- Sridharan, R., Smale, S.T., 2007. Predominant interaction of both Ikaros and Helios with the NuRD complex in immature thymocytes. *J. Biol. Chem.* 282, 30227–30238.
- Stark, H., 2010. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Meth. Enzymol.* 481, 109–126.
- Strohner, R., Nemeth, A., Jansa, P., Hofmann-Rohrer, U., Santoro, R., Längst, G., Grummt, I., 2001. NoRC--a novel member of mammalian ISWI-containing chromatin remodeling machines. *EMBO J.* 20, 4892–4900.
- Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M.P., Walker, J.R., Hogenesch, J.B., 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6062–6067.
- Szerlong, H., Hinata, K., Viswanathan, R., Erdjument-Bromage, H., Tempst, P., Cairns, B.R., 2008. The HSA domain binds nuclear actin-related proteins to regulate chromatin-remodeling ATPases. *Nat. Struct. Mol. Biol.* 15, 469–476.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., Rao, A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B., Zhao, Y., 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146, 1016–1028.
- Tan, Y.Z., Aiyer, S., Mietzsch, M., Hull, J.A., McKenna, R., Grieger, J., Samulski, R.J., Baker, T.S., Agbandje-McKenna, M., Lyumkis, D., 2018. Sub-2 Å Ewald curvature corrected structure of an AAV2 capsid variant. *Nat Commun* 9, 3628. Taplick, J., Kurtev, V., Kroboth,

- K., Posch, M., Lechner, T., Seiser, C., 2001. Homo-oligomerisation and nuclear localisation of mouse histone deacetylase 1. *J. Mol. Biol.* 308, 27–38.
- Tencer, A.H., Cox, K.L., Di, L., Bridgers, J.B., Lyu, J., Wang, X., Sims, J.K., Weaver, T.M., Allen, H.F., Zhang, Y., Gatchalian, J., Darcy, M.A., Gibson, M.D., Ikebe, J., Li, W., Wade, P.A., Hayes, J.J., Strahl, B.D., Kono, H., Poirier, M.G., Musselman, C.A., Kutateladze, T.G., 2017. Covalent Modifications of Histone H3K9 Promote Binding of CHD3. *Cell Reports* 21, 455–466.
- Todd, R., McBride, J., Tsuji, T., Donoff, R.B., Nagai, M., Chou, M.Y., Chiang, T., Wong, D.T., 1995. Deleted in oral cancer-1 (doc-1), a novel oral tumor suppressor gene. *FASEB J.* 9, 1362–1370.
- Tong, J.K., Hassig, C.A., Schnitzler, G.R., Kingston, R.E., Schreiber, S.L., 1998. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* 395, 917–921.
- Torigoe, S.E., Urwin, D.L., Ishii, H., Smith, D.E., Kadonaga, J.T., 2011. Identification of a rapidly formed nonnucleosomal histone-DNA intermediate that is converted into chromatin by ACF. *Mol. Cell* 43, 638–648.
- Tosi, A., Haas, C., Herzog, F., Gilmozzi, A., Berninghausen, O., Ungewickell, C., Gerhold, C.B., Lakomek, K., Aebersold, R., Beckmann, R., Hopfner, K.-P., 2013. Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* 154, 1207–1219.
- Tricarico, R., Cortellino, S., Riccio, A., Jagmohan-Changur, S., Van der Klift, H., Wijnen, J., Turner, D., Ventura, A., Rovella, V., Percesepe, A., Lucci-Cordisco, E., Radice, P., Bertario, L., Pedroni, M., Ponz de Leon, M., Mancuso, P., Devarajan, K., Cai, K.Q., Klein-Szanto, A.J.P., Neri, G., Møller, P., Viel, A., Genuardi, M., Fodde, R., Bellacosa, A., 2015. Involvement of MBD4 inactivation in mismatch repair-deficient tumorigenesis. *Oncotarget* 6, 42892–42904.
- Uchimura, Y., Ichimura, T., Uwada, J., Tachibana, T., Sugahara, S., Nakao, M., Saitoh, H., 2006. Involvement of SUMO Modification in MBD1- and MCAF1-mediated Heterochromatin Formation. *J. Biol. Chem.* 281, 23180–23190.
- Vannini, A., Volpari, C., Gallinari, P., Jones, P., Mattu, M., Carfi, A., De Francesco, R., Steinkühler, C., Di Marco, S., 2007. Substrate binding to histone deacetylases as shown by the crystal structure of the HDAC8–substrate complex. *EMBO reports* 8, 879–884.
- Verreault, A., Kaufman, P.D., Kobayashi, R., Stillman, B., 1998. Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Curr. Biol.* 8, 96–108.
- Verreault, A., Kaufman, P.D., Kobayashi, R., Stillman, B., 1996. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* 87, 95–104.
- von Loeffelholz, O., Papai, G., Danev, R., Myasnikov, A.G., Natchiar, S.K., Hazemann, I., Ménétret, J.-F., Klaholz, B.P., 2018. Volta phase plate data collection facilitates image processing and cryo-EM structure determination. *Journal of Structural Biology* 202, 191–199.
- Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., Skalnik, D.G., 2000. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol. Cell. Biol.* 20, 2108–2121.
- Wade, P.A., Jones, P.L., Vermaak, D., Wolffe, A.P., 1998. A multiple subunit Mi-2 histone deacetylase from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPase. *Curr. Biol.* 8, 843–846.
- Walavalkar, N.M., Gordon, N., Williams, D.C., 2013. Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine binding domain 2 (MBD2) homologues and GATA zinc finger domain containing 2A (GATAD2A/p66 $\alpha$ ). *J. Biol. Chem.* 288, 3419–3427.

- Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G.G., Krones, A., Ohgi, K.A., Zhu, P., Garcia-Bassets, I., Liu, F., Taylor, H., Lozach, J., Jayes, F.L., Korach, K.S., Glass, C.K., Fu, X.-D., Rosenfeld, M.G., 2007. Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* 446, 882–887.
- Wang, Y., Zhang, H., Chen, Y., Sun, Y., Yang, F., Yu, W., Liang, J., Sun, L., Yang, X., Shi, L., Li, R., Li, Y., Zhang, Y., Li, Q., Yi, X., Shang, Y., 2009. LSD1 Is a Subunit of the NuRD Complex and Targets the Metastasis Programs in Breast Cancer. *Cell* 138, 660–672.
- Watson, P.J., Millard, C.J., Riley, A.M., Robertson, N.S., Wright, L.C., Godage, H.Y., Cowley, S.M., Jamieson, A.G., Potter, B.V.L., Schwabe, J.W.R., 2016. Insights into the activation mechanism of class I HDAC complexes by inositol phosphates. *Nature Communications* 7, 11262.
- Wen, Y.D., Perissi, V., Staszewski, L.M., Yang, W.M., Krones, A., Glass, C.K., Rosenfeld, M.G., Seto, E., 2000. The histone deacetylase-3 complex contains nuclear receptor corepressors. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7202–7207.
- Wenzel, D., Palladino, F., Jedrusik-Bode, M., 2011. Epigenetics in *C. elegans*: Facts and challenges. *genesis* 49, 647–661.
- Whitehouse, I., Flaus, A., Cairns, B.R., White, M.F., Workman, J.L., Owen-Hughes, T., 1999. Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature* 400, 784–787.
- Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., Young, R.A., 2012. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482, 221–225.
- Wilczewski, C.M., Hepperla, A.J., Shimbo, T., Wasson, L., Robbe, Z.L., Davis, I.J., Wade, P.A., Conlon, F.L., 2018. CHD4 and the NuRD complex directly control cardiac sarcomere formation. *Proceedings of the National Academy of Sciences* 115, 6727–6732.
- Winter, J., Pantelis, A., Reich, R., Jepsen, S., Allam, J.P., Novak, N., Wenghoefer, M., 2011. Risk Estimation for a Malignant Transformation of Oral Lesions by S100A7 and Doc-1 Gene Expression. *Cancer Investigation* 29, 478–484.
- Wion, D., Casadesús, J., 2006. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* 4, 183–192.
- Witt, O., Deubzer, H.E., Milde, T., Oehme, I., 2009. HDAC family: What are the cancer relevant targets? *Cancer Letters* 277, 8–21.
- Wood, K.H., Zhou, Z., 2016. Emerging Molecular and Biological Functions of MBD2, a Reader of DNA Methylation. *Frontiers in Genetics* 7.
- Wu, L.-C., Chen, Y.-L., Wu, W.-R., Li, C.-F., Huang, H.-Y., Lee, S.-W., Chang, S.-L., Lin, C.-Y., Chen, Y.-H., Hsu, H.-P., Lu, P.-J., Shiue, Y.-L., 2012. Expression of cyclin-dependent kinase 2-associated protein 1 confers an independent prognosticator in nasopharyngeal carcinoma: a cohort study. *J. Clin. Pathol.* 65, 795–801.
- Wu, X., Zhang, Y., 2017. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* 18, 517–534.
- Xiao, H., Sandaltzopoulos, R., Wang, H.M., Hamiche, A., Ranallo, R., Lee, K.M., Fu, D., Wu, C., 2001. Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell* 8, 531–543.
- Xu, C., Liu, K., Lei, M., Yang, A., Li, Y., Hughes, T.R., Min, J., 2018. DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. *Structure* 26, 85–95.e3.
- Yamaguchi, T., Cubizolles, F., Zhang, Y., Reichert, N., Kohler, H., Seiser, C., Matthias, P., 2010. Histone deacetylases 1 and 2 act in concert to promote the G1-to-S progression. *Genes Dev.* 24, 455–469.

- Yan, Y., Barlev, N.A., Haley, R.H., Berger, S.L., Marmorstein, R., 2000. Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases. *Mol. Cell* 6, 1195–1205.
- Yang, T., Jian, W., Luo, Y., Fu, X., Noguchi, C., Bungert, J., Huang, S., Qiu, Y., 2012. Acetylation of Histone Deacetylase 1 Regulates NuRD Corepressor Complex Activity. *J Biol Chem* 287, 40279–40291.
- Yang, W.M., Inouye, C., Zeng, Y., Bearss, D., Seto, E., 1996. Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. *Proc. Natl. Acad. Sci. U.S.A.* 93, 12845–12850.
- Yang, Y., Huang, W., Qiu, R., Liu, R., Zeng, Y., Gao, J., Zheng, Y., Hou, Y., Wang, S., Yu, W., Leng, S., Feng, D., Wang, Y., 2018. LSD1 coordinates with the SIN3A/HDAC complex and maintains sensitivity to chemotherapy in breast cancer. *J Mol Cell Biol*.
- Yildirim, O., Li, R., Hung, J.-H., Chen, P.B., Dong, X., Ee, L.-S., Weng, Z., Rando, O.J., Fazzio, T.G., 2011. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147, 1498–1510.
- Yoon, H.-G., Chan, D.W., Reynolds, A.B., Qin, J., Wong, J., 2003. N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Mol. Cell* 12, 723–734.
- You, A., Tong, J.K., Grozinger, C.M., Schreiber, S.L., 2001. CoREST is an integral component of the CoREST- human histone deacetylase complex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1454–1458.
- Yuan, H., Marmorstein, R., 2013. Histone acetyltransferases: Rising ancient counterparts to protein kinases. *Biopolymers* 99, 98–111.
- Zhang, W., Tyl, M., Ward, R., Sobott, F., Maman, J., Murthy, A.S., Watson, A.A., Fedorov, O., Bowman, A., Owen-Hughes, T., El Mkami, H., Murzina, N.V., Norman, D.G., Laue, E.D., 2013. Structural plasticity of histones H3-H4 facilitates their allosteric exchange between RbAp48 and ASF1. *Nat. Struct. Mol. Biol.* 20, 29–35.
- Zhang, Y., LeRoy, G., Seelig, H.P., Lane, W.S., Reinberg, D., 1998. The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* 95, 279–289.
- Zhou, T., Xiong, J., Wang, M., Yang, N., Wong, J., Zhu, B., Xu, R.-M., 2014. Structural Basis for Hydroxymethylcytosine Recognition by the SRA Domain of UHRF2. *Molecular Cell* 54, 879–886.
- Zou, X., Ma, W., Solov'yov, I.A., Chipot, C., Schulten, K., 2012. Recognition of methylated DNA through methyl-CpG binding domain proteins. *Nucleic Acids Res.* 40, 2747–2758.
- Zupkovitz, G., Tischler, J., Posch, M., Sadzak, I., Ramsauer, K., Egger, G., Grausenburger, R., Schweifer, N., Chiocca, S., Decker, T., Seiser, C., 2006. Negative and Positive Regulation of Gene Expression by Mouse Histone Deacetylase 1. *Molecular and Cellular Biology* 26, 7913–7928.



# Etude structurale du complexe de remodelage de la chromatine NuRD et sa sous-unité MBD3 liée à l'ADN

## Résumé

La régulation de la transcription est un processus dynamique faisant intervenir le recrutement de complexes protéiques impliqués dans le remodelage de la chromatine. Parmi eux, mon travail s'est focalisé sur le complexe NuRD (Nucleosome Remodeling and histone Deacetylation) et sa sous-unité de liaison à l'ADN CpG MBD3. Pour cela une approche de biologie structurale intégrative combinant la préparation biochimique, la caractérisation biophysique et l'étude structurale par cryo-EM et cristallographie aux rayons-X a été mise en place.

Les caractérisations biophysiques de MBD3 ont permis de mettre en évidence son interaction avec un ADN non-modifié CpG et des cristaux diffractant jusqu'à 3.9 Å ont été obtenus. De plus la région désordonnée en aval du domaine de liaison a été identifiée et son impact dans la formation de complexe caractérisé. Des cristaux pour les différentes constructions en complexe avec l'ADN ont été obtenus et sont actuellement optimisés.

Enfin l'optimisation de la purification et la préparation du complexe, ont permis la visualisation du complexe NuRD et mettent en avant pour la première fois une organisation en domaines du complexe.

Mots-clés : NuRD, MBD3, cristallographie aux rayons-X, cryo-EM

## Résumé en anglais

Transcription regulation of chromatin is a very dynamic process regulated through the recruitment of chromatin-remodeling complexes. My work focuses on NuRD for Nucleosome remodeling and histones deacetylation complex a 1 MDa multi-subunit protein complex and its subunit MBD3 a CpG-binding protein and more precisely on an integrated biology approach of this molecular assembly and its interaction with DNA. It combines biochemical preparation, biophysical characterization, single particle cryo-electron microscopy and x-ray crystallography.

Biophysical analysis show that MBD domain of MBD3 interacts with unmodified CpG DNA, a crystal diffracting up to 3.9 Å were obtained. Moreover a C-terminal intrinsically disordered region of MBD3 were identified and despite is inherent disorder seems to increase the binding affinity of MBD3 for DNA. Crystals were obtained for both constructs in complex with DNA and are currently optimized.

Cryo-EM study of NuRD complex allows us to develop and optimized purification and grids preparation for the visualization of the complex. The present results reveal a domain organization of the complex never identify before.

Keywords: NuRD, MBD3, X-ray crystallography, Cryo-EM, remodelling