

Institut de Génétique et de Biologie Moléculaire et Cellulaire  
IGBMC -CNRS UMR 7104 –INSERM U 1258

**THÈSE** Présentée par :

**Alexandre FRECHARD**

Soutenue le : **08 Avril 2021**

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

Discipline : Science de la vie

Spécialité : Biophysique et Biologie Structurale

**Etude structurale des coactivateurs NuA4 &  
SAGA par cryo-microscopie électronique**

**THÈSE dirigée par :**

Dr SCHULTZ Patrick

Directeur de recherche CNRS, IGBMC, Strasbourg

**RAPPORTEURS :**

Professeur VENIEN-BRYAN Catherine

Dr PLISSON-CHASTANG Célia

Directeur de recherche CNRS, IMPMC, Paris

Chargé de recherche CNRS, LBME, Toulouse

---

**AUTRES MEMBRES DU JURY :**

Dr HAMICHE Ali

Directeur de recherche CNRS, IGBMC, Strasbourg



À ma famille et mes amis

## Remerciements

Je tenais à remercier en premier Patrick Schultz, pour m'avoir permis de réaliser ma thèse dans son équipe. Grâce à lui, j'ai eu l'opportunité de travailler sur des sujets exceptionnels et avec des personnes qui le sont tout autant.

Je remercie également tous les membres de mon équipe, plus que jamais la science est un travail collaboratif et c'est grâce à eux que j'ai pu mener ma thèse à bien. Les longues discussions avec Adan Ben Shem, ont toujours été très instructives et m'ont permis d'avancer sur mon projet. C'est grâce à Corine Crucifix que j'ai pu apprendre à utiliser les microscopes électroniques et préparer les échantillons à visualiser. Le travail de Gabor Papaï sur la collecte des données Cryo-EM et l'aide qu'il a apporté pour leurs traitements. Toutes leurs aides m'ont grandement aidé au cours de ma thèse.

Je voulais également remercier Noëlle Potier et Luc Negroni ainsi que leurs équipes respectives pour les données de spectrométrie de masse, notamment l'équipe de Noëlle Potier, qui a consacré énormément de temps et d'effort sur leurs analyses.

Je remercie également Dominique Helmlinger et toute son équipe pour leur collaboration et leurs aide apportée sur l'études des mutations de NuA4.

De manière plus informelle, je voulais remercier Samuel et Joseph pour l'aide à la correction du manuscrit et également remercier toutes les personnes que j'ai pu croiser/discuter lors de ma thèse, c'est parfois au cours d'une pause/discussions entre collègues que des idées essentielles commencent à germer !





# Table des matières

Remerciements .....	4
Table des matières .....	5
Liste des tableaux .....	7
Liste des figures.....	8
I. Introduction .....	13
A. La Chromatine .....	15
1. La structure de la chromatine .....	16
2. Les coactivateurs de la transcription.....	20
B. Le coactivateur NuA4.....	21
1. Tra1 la sous-unité d'interaction avec les activateurs.....	24
2. Le module actine.....	29
3. Le module TINTIN .....	31
4. Le module d'interaction, Eaf1 .....	32
5. Le module d'acétylation des histones, Piccolo.....	33
6. Structure de Piccolo.....	34
7. Structure Cryo-EM avec le nucléosome .....	35
8. Piccolo dans NuA4 .....	37
9. Structure Cryo-EM du complexe TEEAA.....	38
10. Objectif.....	42
C. Le coactivateur SAGA .....	43
1. Le module d'interaction des activateurs : Tra1 .....	43
2. Le module TAF, de SAGA a TFIID.....	45
3. Le module de dé ubiquitination .....	45
4. Le module d'interaction avec TBP .....	47
5. Le module d'acétylation des histones.....	49
6. Structure Cryo-EM de SAGA.....	49
7. Objectif .....	51
D. La Microscopie électronique dans l'étude structurale des protéines.....	52
1. Le microscope électronique.....	52

2.	La préparation des échantillons pour la microscopie électronique.....	63
3.	L'analyse des images en particule isolée.....	68
II.	Matériels & Méthodes.....	83
A.	Préparation des échantillons et analyse en microscopie électronique.....	86
1.	Coloration négative.....	86
2.	Cryo microscopie électronique.....	89
B.	Test d'activité.....	92
1.	Acétylation.....	92
2.	Western blot.....	93
C.	Détermination de modèle atomique.....	95
III.	Résultats.....	97
A.	NuA4.....	99
1.	Détermination de la structure de NuA4 par cryo-EM.....	100
B.	NuA4 Mutant.....	103
1.	Modèle atomique de NuA4.....	105
C.	Conclusions sur le modèle atomique.....	116
1.	Flexibilité des modules d'interaction avec la chromatine.....	118
2.	Réticulation de NuA4 et analyse spectrométrique.....	125
3.	NuA4 et le nucléosome en microscopie électronique.....	130
D.	SAGA.....	132
1.	Résumé.....	132
2.	Article.....	135
IV.	Discussion.....	153
A.	NuA4.....	155
B.	SAGA et NuA4.....	159
V.	Conclusion & perspective.....	165
VI.	Bibliographies.....	170
	Résumé.....	180
	Résumé en anglais.....	180

## Liste des tableaux

<b>Tableau 1 : Conservation des sous-unités de NuA4 de la levure à l'humain .....</b>	<b>23</b>
<b>Tableau 2 : Conservation des différents modules/sous-unités de SAGA au cours de l'évolution.....</b>	<b>44</b>
<b>Tableau 3 : Réactifs pour la réaction d'acétylation.....</b>	<b>93</b>
<b>Tableau 4 : Solution utilisé pour le Westernblot .....</b>	<b>94</b>

## Liste des figures

<b>Figure 1 : Représentation schématique de l'organisation de l'ADN .....</b>	<b>15</b>
<b>Figure 2 : Structure et organisation du nucléosome .....</b>	<b>17</b>
<b>Figure 3 : Fibre de 11 &amp; 30 nm en microscopie électronique.....</b>	<b>18</b>
<b>Figure 4 : Fonction et recrutement des coactivateurs.....</b>	<b>21</b>
<b>Figure 5 : Schéma de l'organisation des protéines de la famille PIKK.....</b>	<b>24</b>
<b>Figure 6 : Structure de TRA1 et schéma de son organisation chez la levure .....</b>	<b>26</b>
<b>Figure 7 : Organisation générale de Tra1 .....</b>	<b>26</b>
<b>Figure 8 : Mutations de Tra1 interférant avec le ciblage des activateurs.....</b>	<b>27</b>
<b>Figure 9 : Structure cristallographique du complexe ternaire d'actine nucléaire .....</b>	<b>29</b>
<b>Figure 10 : Analyse structurale de la poche de liaison à l'ATP d 'Arp4 et de la N- actine.....</b>	<b>31</b>
<b>Figure 11 : Structure cristallographique du module Piccolo .....</b>	<b>34</b>
<b>Figure 12 : Changements conformationnels de HBL d'Esa1 .....</b>	<b>35</b>
<b>Figure 13 : Structure Cryo-EM de Piccolo avec le nucléosome .....</b>	<b>36</b>
<b>Figure 14 : Structure cryo-EM du complexe TEEAA .....</b>	<b>38</b>
<b>Figure 15 : Structure Cryo-EM du complexe TEEAA .....</b>	<b>39</b>
<b>Figure 16 : Surface d'interaction d'Eaf5 avec Tra1 .....</b>	<b>41</b>
<b>Figure 17 : Structure du module DUB de SAGA .....</b>	<b>46</b>
<b>Figure 18 : Structure de TBP et ses sites d'interaction avec Spt3 et Spt8 .....</b>	<b>48</b>
<b>Figure 19 : Structure Cryo-EM du coactivateur SAGA.....</b>	<b>50</b>
<b>Figure 20 : Schéma d'une lentille magnétique.....</b>	<b>55</b>
<b>Figure 21 : Aberration sphérique .....</b>	<b>56</b>
<b>Figure 22 : Aberration chromatique.....</b>	<b>57</b>
<b>Figure 23 DQE à 300 kV en fonction de la fréquence spatiale.....</b>	<b>60</b>
<b>Figure 24 : Signaux courants générés par l'interaction d'un faisceau d'électrons avec et un échantillon fin .....</b>	<b>61</b>
<b>Figure 25 : Contraste d'amplitude.....</b>	<b>62</b>
<b>Figure 26 : Les différentes étapes de la coloration négative.....</b>	<b>65</b>

<b>Figure 27 : Vitrifcation des échantillons .....</b>	<b>67</b>
<b>Figure 28 : Fonction de transfert de contraste et dé focus .....</b>	<b>71</b>
<b>Figure 29 : Méthode de gradient descendant pour la détermination de carte Cryo-EM .....</b>	<b>77</b>
<b>Figure 30 : Régularisation adaptative et invariante.....</b>	<b>80</b>
<b>Figure 31 : Affinement non uniforme .....</b>	<b>81</b>
<b>Figure 32 : Purification du complexe NuA4.....</b>	<b>99</b>
<b>Figure 33 : Analyse des images du NuA4 WT .....</b>	<b>102</b>
<b>Figure 34 : Analyse des images de NuA4 mutant .....</b>	<b>104</b>
<b>Figure 35 : Structure de NuA4 et modèle atomique.....</b>	<b>105</b>
<b>Figure 36 : Tracé d'Eaf1 dans NuA4 .....</b>	<b>107</b>
<b>Figure 37 : Interaction entre Eaf1, Act1 et Arp4 .....</b>	<b>108</b>
<b>Figure 38 : Interaction entre Eaf1 et Tra1 .....</b>	<b>109</b>
<b>Figure 39 : Réseau d'hélice formé par Eaf1, Swc4 et Epl1.....</b>	<b>111</b>
<b>Figure 40 : Interaction entre Eaf1 et Epl1 .....</b>	<b>113</b>
<b>Figure 41 : Interaction de Swc4 avec NuA4.....</b>	<b>114</b>
<b>Figure 42 : Poches de liaisons à l'ATP .....</b>	<b>115</b>
<b>Figure 43 : Régions désordonnées de Swc4 et du N-ter d'Eaf1.....</b>	<b>118</b>
<b>Figure 44 : Flexibilité du module Piccolo .....</b>	<b>120</b>
<b>Figure 45 : Région désordonné d'Epl1 .....</b>	<b>121</b>
<b>Figure 46 : Test d'activité du NuA4 mutant .....</b>	<b>122</b>
<b>Figure 47 : Analyse des cartes Cryo-EM NuA4 WT, <math>\Delta</math>364-414 et <math>\Delta</math>364-427 .....</b>	<b>123</b>
<b>Figure 48 : Soustraction de la densité de NuA4.....</b>	<b>124</b>
<b>Figure 49 : Analyse en spectrométrie de masse de NuA4 réticulé .....</b>	<b>126</b>
<b>Figure 50 : Analyse des pontages chimique avec le modèle atomique de NA4... </b>	<b>127</b>
<b>Figure 51 : Effet de différentes méthodes de fixation sur NuA4.....</b>	<b>128</b>
<b>Figure 52 : NuA4-NCP coloration négative .....</b>	<b>130</b>
<b>Figure 53 : Classes 2D de NuA4-NCP en Cryo-EM .....</b>	<b>131</b>
<b>Figure 54 : Structure et modèle atomique de SAGA .....</b>	<b>133</b>
<b>Figure 55 : Comparaison du module actine chez NuA4 et Ino80 .....</b>	<b>155</b>
<b>Figure 56 : Modèle de l'organisation structurale de NuA4 .....</b>	<b>157</b>

<b>Figure 57 : Modèle de fonctionnement de l'acétylation du NCP par NuA4 recruté par des activateurs de la transcription.....</b>	<b>158</b>
<b>Figure 58 : Comparaison des complexes SAGA et NuA4.....</b>	<b>159</b>
<b>Figure 59 : Comparaison de la structure de Tra1 chez SAGA et NuA4 .....</b>	<b>160</b>
<b>Figure 60 : Site de liaison avec Tra1 chez SAGA et comparaison avec NuA4 ...</b>	<b>161</b>
<b>Figure 61 : Comparaison des modèles de liaison des modules HAT de NuA4 et SAGA.....</b>	<b>163</b>

zz

## Abréviation

Act1 : Actine 1

ADN : Acide désoxyribonucléique

ARN : Acide ribonucléique

ARP4 : Actin Related Protein 4

ATP : Adénosine triphosphate

ADP : Adénosine diphosphate

CCD : charge-coupled device

CTF : Fonction de Transfert de Contraste

Cryo-EM : Cryo-microscopie électronique

DDC : Direct Detection Camera

DQE : Detective Quantum efficiency

Eaf1 : Esa1p-Associated Factor 1

Eaf3 : Esa1p-Associated Factor 3

Eaf5 : Esa1p-Associated Factor 5

Eaf7 : Esa1p-Associated Factor 7

Epl1 : Enhancer of Polycomb Like

FAT : FRAP, ATM, TRRAP

FEG : Field Emission Gun

FRB : FKBP12-rapamycin binding domain

FSC : Fourier Shell Correlation

HAT : Histone Acétyltransférase

HEAT : acronyme de **H**untingtine, facteur d'**E**longation eucaryotique 2, protéine phosphatase **2A** et la kinase **Tor1**

HFD : domaine histone-fold

HRP : peroxydase de raifort

HSA : **H**elicase **S**ANT domaine **A**ssociated

MET : Microscopes Electroniques à Transmission

NCP : Nucléosome

NuA4 : Nucléosome Acétyle transférase de l'histone H4

Piccolo : Pic NuA4

PIKK : Phosphoinositol 3-Kinases related Kinase

PTM : modification post traductionnelles

SAGA : Spt-Ada-Gcn5-Acétyltransférase

TAF : TBP Associated Factor

TBP : TATA Binding Protein

TINTIN : **T**ramer **I**ndependent of NuA4 involved in **T**ranscription **I**nteractions with **N**ucleosomes »

TEEAA : Tra1-Eaf1-Eaf5-Actin1-Arp4

WT : Wilde Type (type naturel)

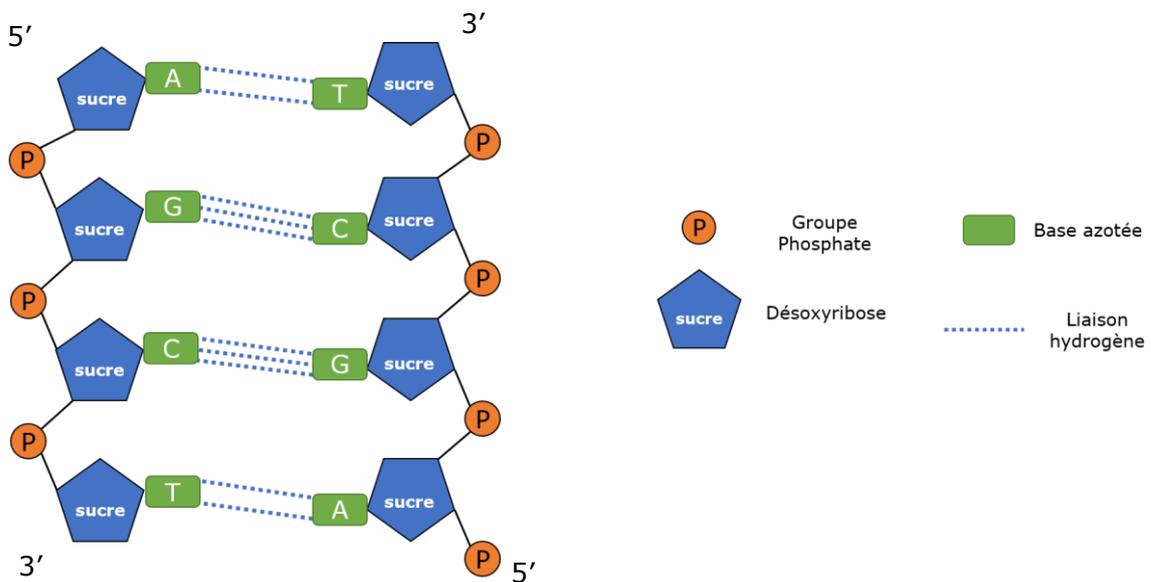


# I. Introduction



## A. La Chromatine

L'Acide désoxyribonucléique ou ADN est une macromolécule biologique, elle est le support de l'information génétique chez les êtres vivants. Elle est retrouvée dans tout type d'organisme, aussi bien chez les bactéries que chez les animaux ou les virus. Cette molécule est bicaténaire, ce qui veut dire qu'elle est formée de deux brins qui viennent s'entremêler pour aboutir à la formation d'une double hélice d'ADN. Chacun des brins est un polymère de nucléotides formant ainsi un monomère d'ADN. Un monomère est composé d'une des quatre bases azotées suivantes : l'Adénine, la Thymine, la Cytosine et la Guanine (A, T, G et C). Cette base est reliée de manière covalente à un sucre qui est le désoxyribose et ce dernier fait également une liaison covalente avec un groupement phosphate **Figure 1**. Le lien entre plusieurs nucléotides se fait par une liaison covalente entre le désoxyribose du premier nucléotide et le groupement phosphate du nucléotide suivant. La succession des nucléotides permet de former la séquence d'ADN.



**Figure 1 : Représentation schématique de l'organisation de l'ADN**

Les nucléotides de deux brins d'ADN forment des liaisons hydrogène au niveau de leurs bases, cet appariement suit des règles strictes. Une adénine ne peut s'apparier qu'avec une thymine et forment entre eux deux liaisons hydrogène, une cytosine s'apparie toujours avec une guanine en formant trois liaisons hydrogène (Watson & Crick, 1953). Il en résulte que pour deux brins d'ADN puissent s'apparier entre eux, ils doivent nécessairement être le complémentaire de l'autre.

Chez les organismes eucaryotes, l'ADN est situé dans un compartiment de la cellule qui est le noyau, pour qu'il puisse être contenu dans cet espace exiguë, il est condensé sous forme de chromatine. Ce terme vient du biologiste Walther Fleming qui l'utilisa pour la première fois en 1879 pour décrire les filaments fortement colorés qu'il pouvait observer dans le noyau. La chromatine est un état de l'ADN où celui-ci interagit avec des protéines, ce qui permet sa condensation. Elle ne permet pas que de contenir l'ADN dans le noyau, mais également d'assurer une plus grande stabilité de ce dernier et de réprimer l'accès aux gènes. Enfin elle est nécessaire à la mitose, puisque pour réaliser la division cellulaire il faut impérativement condenser l'ADN dans sa forme la plus dense qui est le chromosome afin de pouvoir partager le génome dupliqué entre les deux cellules filles.

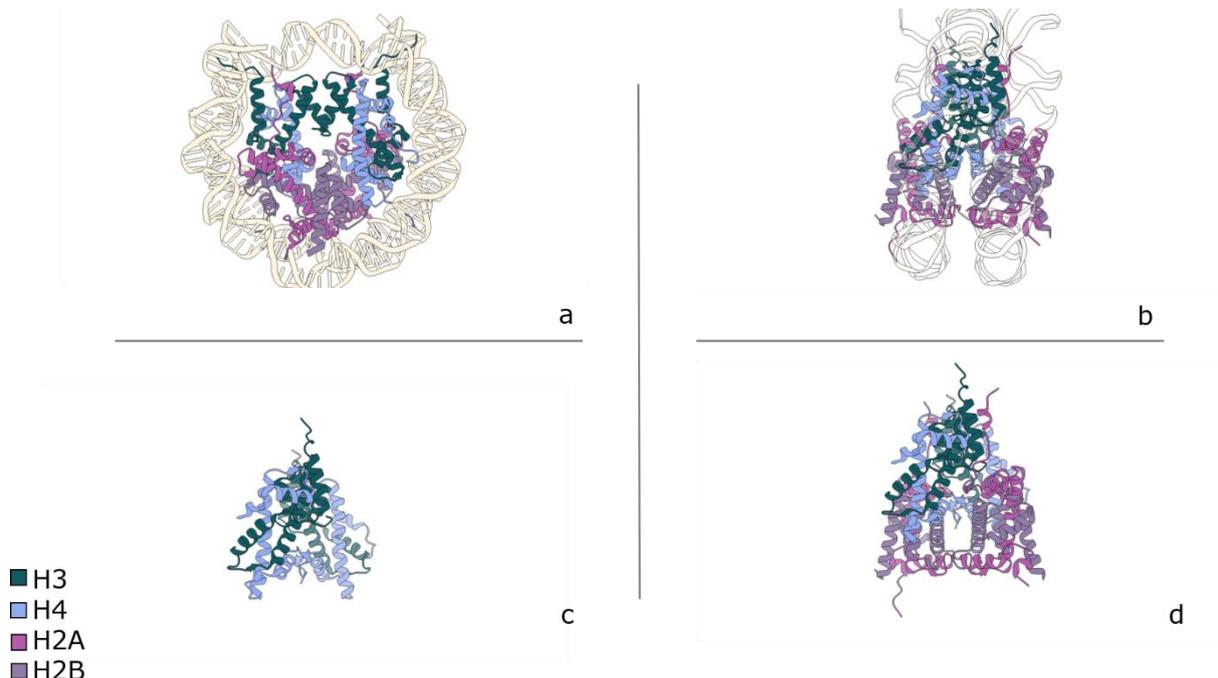
## 1. La structure de la chromatine

L'unité de base permettant de compacter l'ADN est le nucléosome abrégé maintenant NCP. C'est un octamère d'histones qui sont H2A, H2B, H3 et H4, chacune présente en deux copies, où vient s'enrouler 146 pdb d'ADN. La queue N-terminale des histones composant le NCP peut également être modifiée de manière post-traductionnelle sur leurs lysines par des mécanismes d'acétylation, d'ubiquitination, de méthylation ou de phosphorylation. Ces modifications sont connues sous le nom de PTM pour « post-translational modification » et sont importantes pour la régulation de la chromatine.

## a) La structure du nucléosome

En 1997, la première structure du NCP est obtenue par le groupe Richmond par cristallographie (Luger et al., 1997). Elle permet de comprendre comment est formé l'octamère d'histone et les interactions qu'il fait avec l'ADN.

L'octamère est formé par deux tétramères d'histones chacun composé de H2A, H2B, H3 & H4, entourés par ~146 pnb d'ADN **Figure 2 a et b**. L'interaction entre les deux tétramères se fait par leurs paires d'histones H3 & H4 **Figure 2 c et d**. La dimérisation des paires d'histones se fait grâce à leurs HFD (Histone Fold Domain) qui consiste en un motif de 3 hélices  $\alpha$ , dont deux courtes situées de part et d'autre de la troisième plus grande et sont reliés par des boucles.



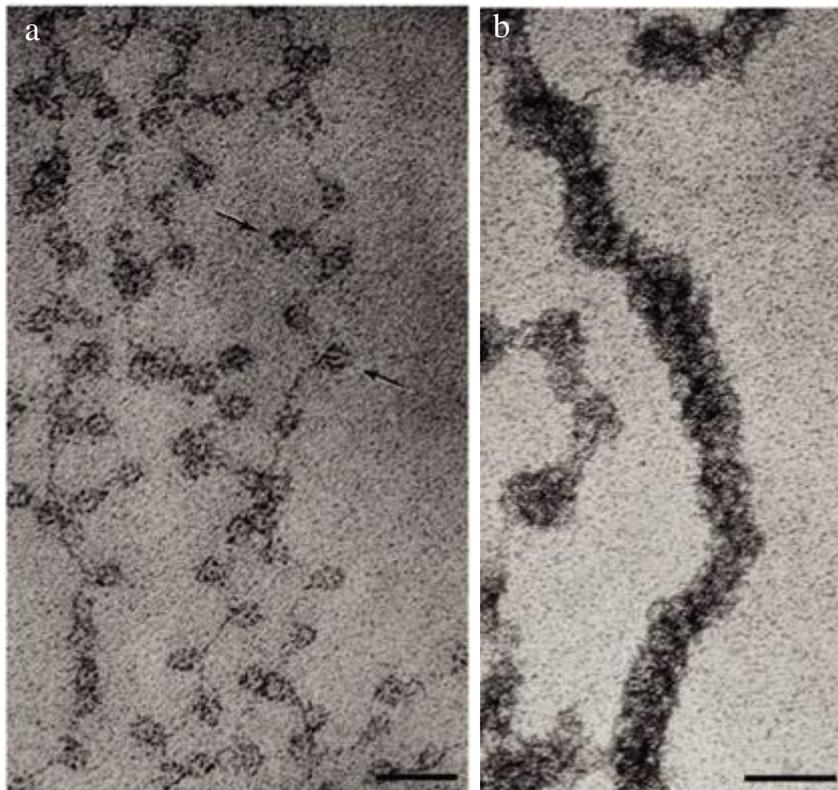
**Figure 2 : Structure et organisation du nucléosome**

(a) Vue de face de la structure du NCP où on peut voir un octamère d'histone dont chacun d'eux est coloré suivant le code couleur présent sur la figure, autour de l'octamère un ADN de 146 pnb vient s'enrouler formant le NCP. (b) Vue de côté du NCP où l'ADN a été mis en transparence pour faciliter la visualisation. (c) Seules les paires d'histones H3 & H4 de chaque tétramères sont ici visibles et montrent comment ils font le lien entre les deux tétramères

d'histones. (d) Tout l'octamère est visible ici, montrant effectivement que les interactions entre les deux tétramères sont réalisées par H3 & H4.

### **b) Différents degrés de condensation**

Chaque nucléosome compactant la chromatine est relié aux autres par un ADN appelé « ADN linker », ceci permet de former une structure ressemblant à un collier de perle **Figure 3 a**. Il a été montré qu'il peut exister des structures de compaction de l'ADN encore plus évoluées appelées fibre de 30 nm **Figure 3 b** en raison des observations



**Figure 3 : Fibre de 11 & 30 nm en microscopie électronique**

L'image (a) représente des nucléosomes, indiqués par des flèches, organisés en fibre de 11 nm la barre noire représente 30 nm. (b) Image d'une fibre de 30 nm, la barre noire représente 50 nm

Image tiré de (Olins & Olins, 2003)

faites par microscopie électronique où l'on voit une fibre ayant une épaisseur approximative de 30 nm (Olins & Olins, 2003). Bien que cette structure soit connue depuis de nombreuses années, l'observation par Cryo-microscopie ne montre aucune trace de la fibre de 30 nm dans le noyau des cellules (Maeshima et al., 2019), laissant penser que cette structure est un artéfact obtenu *in vitro* ou n'est présente que dans

certaines types de cellules et/ou à certains stades du cycle cellulaire et que, par conséquent, elle n'est pas aussi reproduite que ce que l'on pensait ou est un artefact obtenu *in vitro*.

### **c) L'euchromatine et l'hétérochromatine**

Dans le noyau de la cellule on peut caractériser la chromatine en deux états qui correspondent à leurs degrés de transcription. Le premier est l'hétérochromatine qui est une région de faible activité transcriptionnelle, alors que le second est l'euchromatine qui possède une forte activité transcriptionnelle.

Chez la levure, notamment *Saccharomyces cerevisiae*, la grande majorité du génome est sous forme d'euchromatine et seulement quelques régions sont sous formes d'hétérochromatine, comme par exemple celle des gènes impliqués dans le mécanisme de la reproduction sexuelle ou les télomères (Hickman et al., 2011). Dans le cas des eucaryotes supérieurs, la répartition entre eu/hétérochromatine est bien différente et plus complexe. Chez l'humain l'hétérochromatine peut être de deux types. Le premier est l'hétérochromatine facultative qui est une portion du génome dont la condensation varie d'un type cellulaire à un autre. La seconde est l'hétérochromatine constitutive qui est toujours condensée et silencieuse et correspond au centromère, télomère ou à des séquences d'ADN répétées.

On retrouve différentes modifications post-traductionnelles des histones PTM suivant le type de chromatine, les nucléosomes possédant des groupes acétyle sur la queue des histones H4/H3 sont associés avec une chromatine active transcriptionnellement.

### **d) Les modifications post traductionnelles des histones**

La partie N-terminale des histones est sujette à des modifications post traductionnelles. Ces modifications sont très dynamiques et sont régulées par des enzymes (acétyle transférase, ubiquitinase, méthyltransférase, kinase ...) qui sont capables d'inscrire ces modifications sur les histones et d'autres enzymes sont capables de les effacer (désacétylase, déubiquitinase ...). Ces modifications pourraient être importantes pour

déclencher une chaîne d'événements importante pour la régulation de la chromatine (Strahl & Allis, 2000).

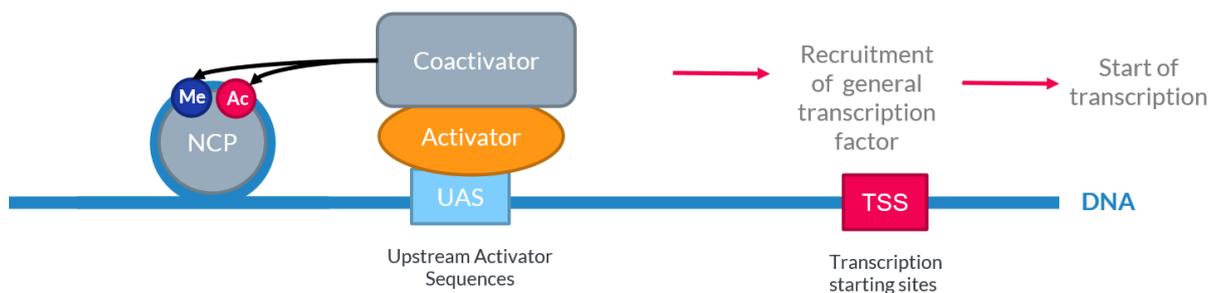
Les modifications d'histones permettent d'altérer l'état de la chromatine par deux mécanismes distincts. Le premier est l'altération des charges des histones (Rothbart & Strahl, 2014), ce qui va entraîner une perte d'interaction avec l'ADN. L'acétylation est une modification qui engendre ce genre d'effet, la charge positive des lysines situées sur la chaîne latérale va être neutralisée par l'ajout du groupement acétyle, ce qui réduit l'interaction avec l'ADN qui lui est chargé négativement (Barnes et al., 2019), ceci permettrait de déstabiliser les interactions entre le NCP et l'ADN, rendant l'ADN plus accessible aux facteurs de transcription. Le deuxième mécanisme impliquant les PTM serait leur capacité à être lu par des acteurs protéiques, ce qui donnerait lieu à une suite d'événements sur la chromatine.

## 2. Les coactivateurs de la transcription

La transcription est le mécanisme par lequel la cellule retranscrit l'information codante de son ADN en ARN qui pourra par la suite être traduit par le ribosome. Cependant pour que la transcription puisse démarrer il est nécessaire de réguler la compaction de la chromatine car cette dernière crée une barrière physique qui empêche la transcription des gènes, rendant impossible le recrutement des agents nécessaires au début et à la poursuite de la transcription. Pour ce faire la cellule possède des complexes appelés coactivateurs de la transcription qui sont des complexes multiprotéiques dont l'activité permet de réorganiser la compaction de la chromatine autour des promoteurs des gènes, ce qui permet le recrutement du complexe de pré initiation de la transcription (PIC).

Ces coactivateurs peuvent être recrutés sur ces régions spécifiques de l'ADN de différentes manières. La première est grâce à leurs interactions avec les activateurs de la transcription qui sont des protéines qui vont reconnaître et se lier à des sites qui leur sont spécifiques sur les promoteurs des gènes et ensuite interagir avec les coactivateurs ce qui va permettre de les recruter. La deuxième façon est par la reconnaissance de modifications épigénétiques des queues d'histones du nucléosome qui peuvent être des lysines acétylées, méthylées, di/tri méthylées ou encore ubiquitinées.

Une fois recrutés les coactivateurs vont pouvoir effectuer leurs activités, qui vont permettre de rendre la compaction de l'ADN plus lâche. Ceci favorisera le début de la transcription. Bien que ces complexes aient pour but de changer l'état de la chromatine, ils ne possèdent pas tous la même activité. Certains sont plus enclins à avoir une fonction qui va permettre de déplacer/évincer le nucléosome de l'ADN, c'est le cas par exemple des coactivateur Swi/snf & INO80. D'autres vont permettre l'incorporation de variants d'histones, par exemple chez la levure le complexe Swr1 échange H2A par le variant H2AZ. Des coactivateurs vont également « simplement » écrire des modifications épigénétiques : c'est le cas pour NuA4 & SAGA, qui sont responsables de l'acétylation des H2A & H4 pour le premier et H2B & H3 pour le second. Enfin les coactivateurs sont également capables d'interagir avec les facteurs de transcription comme SAGA, qui permet de délivrer la protéine de liaison de la boîte TATA où TBP.



**Figure 4 : Fonction et recrutement des coactivateurs**

Schémas du mode de fonctionnement des coactivateurs de la transcription. Le recrutement sur les régions promotrices des promoteurs peut se faire par deux méthodes : la première est par le biais d'activateurs, qui vont interagir avec les promoteurs de gènes puis recruter les coactivateurs ; la seconde et par les modifications épigénétiques du NCP qui vont être lu par les coactivateurs, ce qui permettra de les recruter. Une fois recrutés, les coactivateurs permettront le recrutement des facteurs généraux de la transcription et de son initiation.

## B. Le coactivateur NuA4

NuA4 est un complexe protéique impliqué dans la régulation de la transcription et la réparation des cassures double brin de l'ADN (Doyon & Côté, 2004), (Bennett & Peterson, 2015). Ce complexe est connu pour son activité d'histone acétyl transférase, c'est-à-dire qu'il est capable d'acétyler les lysines et spécifiquement les queues d'histone H4 et H2A2 du nucléosome, ainsi que le variant H2AZ. Il est retrouvé sur les régions promotrices des gènes où son activité d'acétyl transférase lui permet de réguler

la compaction de la chromatine. L'acétylation des histone H4 et H2A est associée à une chromatine active transcriptionnellement. Les autres sites où est retrouvé NuA4 sont les régions où l'ADN a subi une cassure double brin où il permet la mise en place de la réponse à ces dommages. De manière intéressante, contrairement à SAGA qui acétyle les NCP proche de son site de recrutement, NuA4 est capable d'acétyler les NCP sur une plus grande portée (Vignali et al., 2000).

Chez la levure, NuA4 est composé de 13 sous-unités qui peuvent être réparties en 5 modules : le module d'interaction avec les activateurs de la transcription qui est composé de Tra1, le module structural qui est composé de Eaf1 qui est la protéine qui permet de maintenir l'intégrité de la structure de NuA4 en effectuant les interactions avec les autres modules du coactivateur, le module Actine composé d'Actin1, Arp4, Swc4 & Yaf9 est impliqué dans l'interaction avec la chromatine, comme le quatrième module TINTIN qui peut reconnaître des modifications épigénétique du nucléosome qui sont la di et tri méthylation de queue d'histone et le dernier module est nommé Piccolo est le module catalytique qui porte l'activité d'histone acétyle transférase et permet d'acétyler le nucléosome.

Ce complexe est conservé au cours de l'évolution **Tableau 1** et possède un complexe homologue chez l'Homme (Tip60) et chez la levure (NuA4) (Doyon & Côté, 2004). En plus des activités que ce complexe partage avec son homologue chez la levure, il est impliqué dans l'échange de l'histones H2A par son variant H2AZ, qui est lui aussi souvent retrouvé dans les régions actives transcriptionnellement. Tip60 est impliqué dans plusieurs maladies comme le cancer et y joue un rôle bivalent tantôt répresseur de tumeur tantôt proto oncogène (Chevillard-briet et al., 2014)(Bassi et al., 2016). Bien qu'il soit connu pour acétyler la chromatine, ce complexe est également capable d'acétyler d'autre protéines comme les récepteurs aux androgènes, l'acétylation de ces derniers entrainant progressivement une résistance au traitement hormonaux. L'augmentation de l'activité de NuA4 entraine donc une prolifération du cancer (Shiota et al., 2009). Pourtant, une perte d'activité de NuA4 dans d'autres cancers augmente sa prolifération. Etant important pour la réparation de l'ADN une perte de l'activité de NuA4 entrainera une accumulation de mutations favorisant le développement du cancer (Mattera et al., 2009). Mieux connaitre comment NuA4/tip60 est capable d'effectuer

ces tâches est crucial pour mettre en lumière les mécanismes dans lesquels il est impliqué et permettrait de développer de nouvelles thérapies.

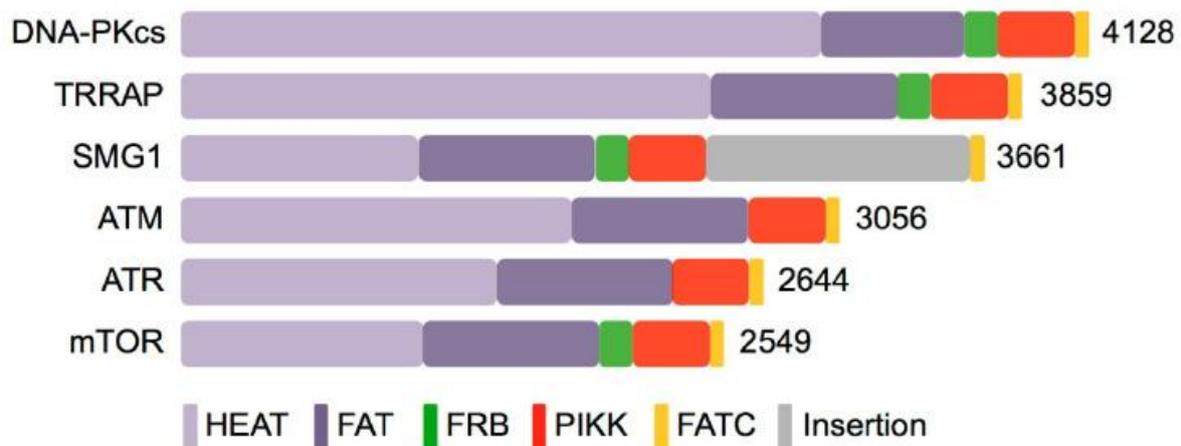
Sous-unités chez la levure	Homologue humain	Domaine(s)
Tra1	TRRAP	PI-3 kinase/ATM
Eaf1	hDomino? (p400)	SANT, HSA
Epl1	Enhancer of Polycomb (EPC1, EPC-like)	EpcA
Eaf2	DMAP1 (DNMT1-associated protein)	SANT
Arp4	BAF53a (BRG1-associated factor)	Actin related
Esa1	Tip60/Tip60b (Tat interactive protein)	Chromodomain/MYST HAT
Eaf3	Mortality factor related genes (MORF4, MRG15/X)	Chromodomains
Act1	Actin	
Yng2	Inhibitor of growth gene family (ING1 to -5)	PHD finger
Eaf5	?	
Yaf9	YEATS family (AF9, ENL, GAS41)	YEATS
Eaf6	hEaf6 (FLJ11730)	

**Tableau 1 : Conservation des sous-unités de NuA4 de la levure à l'humain**

D'après (Doyon & Côté, 2004)

## 1. Tra1 la sous-unité d'interaction avec les activateurs

Tra1 fait partie de la famille des protéines PIKK pour « Phosphoinositol 3-Kinases related Kinase » qui est une famille apparentée à un autre type de protéine : celle des PI3K pour « Phosphoinositol 3-Kinases ». Les PIKK se caractérisent par une très grande taille qui peut varier entre 2000 et 4000 acides aminés et par une activité serine/thréonine kinase.



**Figure 5 : Schéma de l'organisation des protéines de la famille PIKK**

D'après (Rivera-Calzada et al., 2015)

Le domaine kinase des PIKK est retrouvé de manière générale dans les 400 derniers acides aminés de la partie C-terminal de la protéine, ce domaine est flanqué par deux autres domaines : le domaine FAT dont le nom est dérivé de FRAP, ATM, & TRRAP et qui est situé dans les ~500 acides aminés du côté n-terminal du domaine kinase ainsi que le domaine FATC qui est composé des 35 derniers acides aminés du côté C-terminal **Figure 5**. Dans certaines protéines de la famille PIKK un domaine de liaison à la rapamycine est retrouvé entre le domaine kinase et FAT, il est appelé le FRB domaine pour FKBP12-rapamycin binding domain. La partie N-terminal quant à elle est constituée de régions répétées HEAT **Figure 5**. Les domaines HEAT sont constitués d'une répétition de deux hélices  $\alpha$  antiparallèles reliées par une boucle, aux positions 19 et 25 sont retrouvées respectivement une asparagine et une arginine. Ce motif se répète plusieurs fois ce qui

forme un solénoïde. Le nom de HEAT est un acronyme des différentes protéines où est retrouvé ce genre de domaine, l'huntingtine (H), le facteur d'élongation eucaryotique 2 (E), la protéine phosphatase 2A (A) et la kinase Tor1 (T).

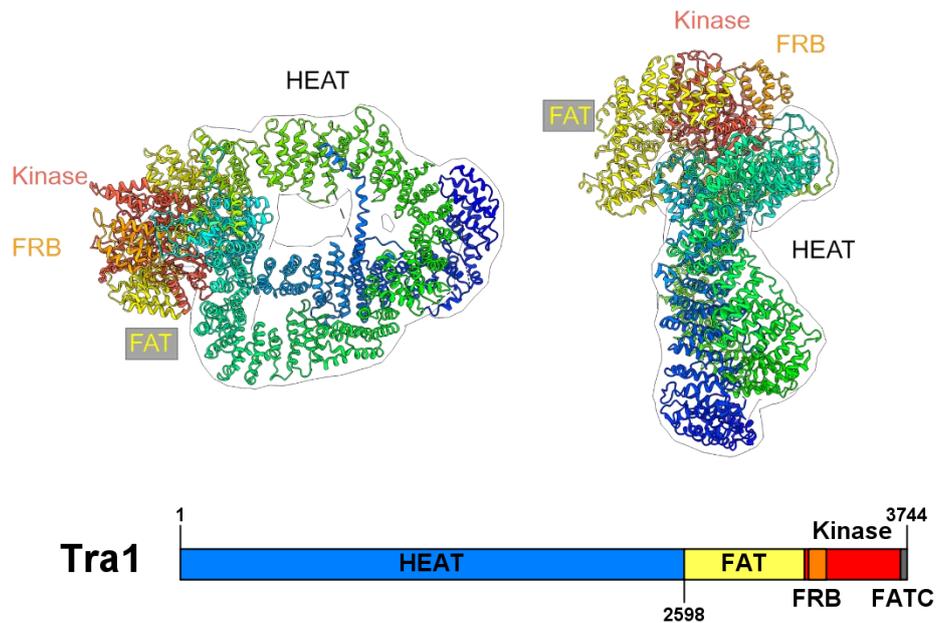
Tra1 est l'homologue de TRRAP chez la levure, tout comme son équivalent humain, mais à l'instar des autres membres de la famille PIKK le domaine kinase de Tra1 est inactif, ce qui explique le fait qu'il ne possède pas les motifs protéiques DXXXXN et DGF qui sont nécessaires pour réagir avec l'ATP.

Cette protéine est responsable de l'interaction de NuA4 avec les activateurs et permet donc de recruter le complexe sur des régions spécifiques du génome (Knutson & Hahn, 2011), il est aussi important de noter que l'on retrouve cette sous-unité dans un autre coactivateur qui est SAGA où elle possède également le même rôle. Cependant, les deux coactivateurs sont recrutés sur différentes portions du génome (Uprety et al., 2015)(Baptista et al., 2017), posant donc la question de comment les activateurs peuvent recruter de manière spécifique le Tra1 appartenant à NuA4 ou celui incorporé dans SAGA. Cette protéine est la plus grande sous-unité du complexe NuA4 chez la levure, elle possède un poids moléculaire de 433 KDa et représente à elle seule une grande partie du poids de NuA4.

### **a) Structure de Tra1**

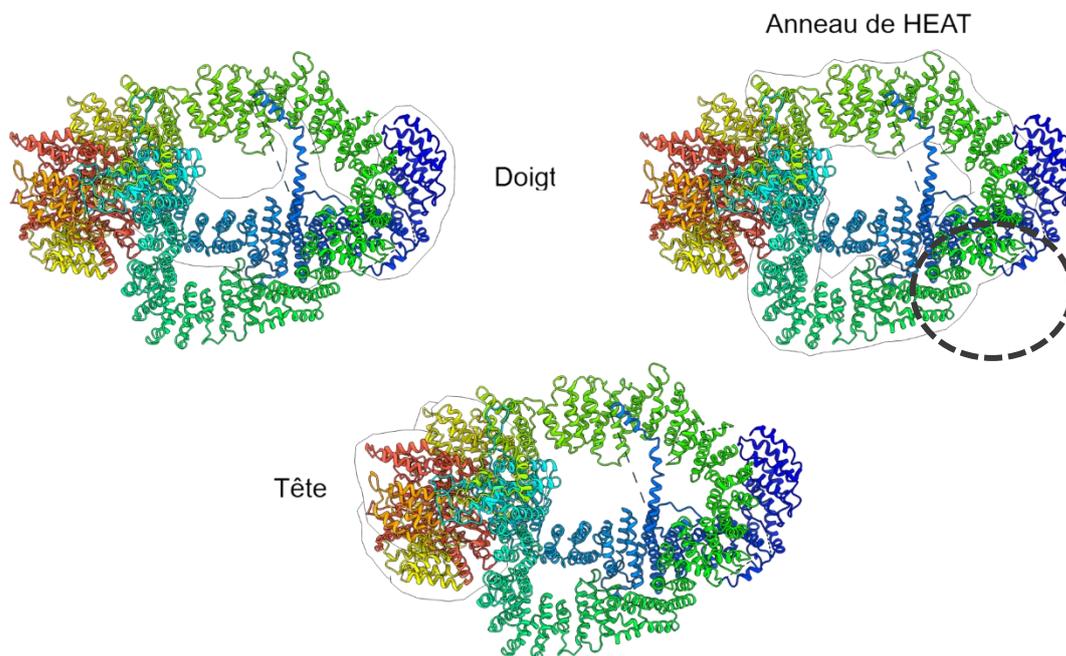
Ces dernières années plusieurs modèles atomiques de Tra1 ont pu être obtenus grâce à la cryo microscopie électronique (Sharov et al., 2017)(Díaz-Santín et al., 2017)(X. Wang et al., 2018). Ces modèles montrent que cette structure est typique de celle des autres membres des PIKK avec les domaines HEAT, FAT, FRB, Kinase et FATC qui sont arrangés de manière séquentielle du N au C-terminal **Figure 6**.

Le « doigt » et « l'anneau » sont des solénoïdes d'hélices alphas obtenus par l'association des domaines HEAT. Le doigt de Tra1 est constitué de 16 régions HEAT répétées et est l'équivalent de la région spirale des protéines mTOR, ATM et du bras de DNA-PKcs **Figure 7**. Quant à l'anneau, il est composé de 33 HEAT et représente le plus



**Figure 6 : Structure de TRA1 et schéma de son organisation chez la levure**

Le domaine HEAT contient 49 répétitions HEAT et commence du N-terminal de Tra1 jusqu'au domaine FAT, les 1000 premiers acides aminés forme un « doigt » puis le reste du domaine HEAT forme « un anneau » avant de rejoindre la « tête » de la protéine.



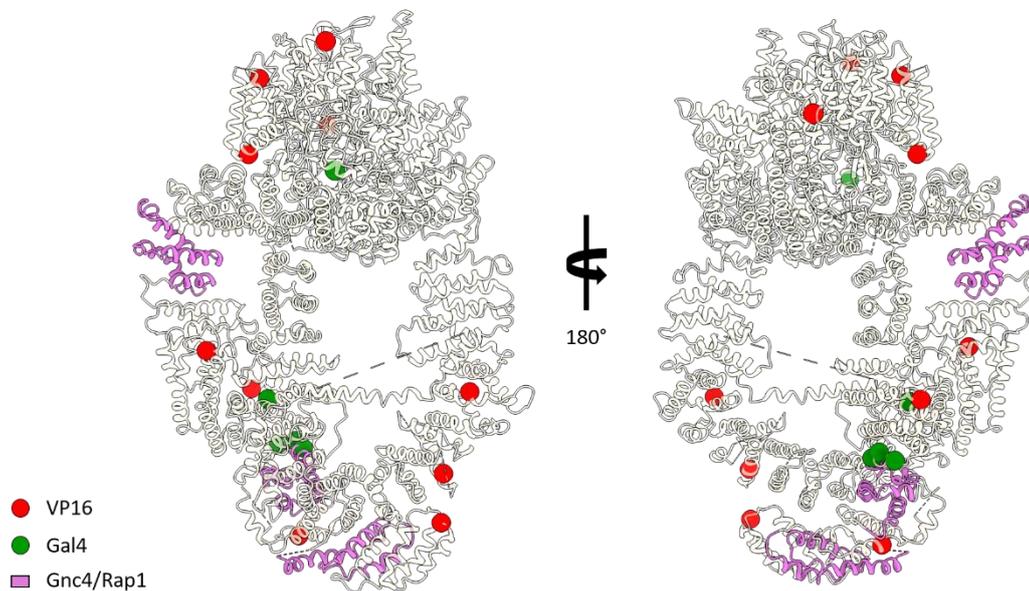
**Figure 7 : Organisation générale de Tra1**

La figure montre l'organisation globale de Tra1 en 3 parties qui sont le doigt un solénoïde de région HEAT répété, suivit par un anneau également composé de HEAT et la tête qui contient les domaine FAT, FRB, Kinase & FATC. L'ovale en traits discontinue indique une zone de croisement entre l'anneau de HEAT et la première moitié du doigt.

grand solénoïde d'hélices alphas de la protéine. La partie N-terminal et C-terminal de l'anneau permettent de le fermer avant d'arriver au domaine FAT **Figure 7**.

La tête de Tra1 est composée des domaines FAT, FRB, Kinase et FATC. Les domaines FAT et FRB viennent entourés le domaine kinase **Figure 6 & Figure 7**. Le domaine kinase est très semblable aux autres protéines des familles PIKK et PIK3, mais bien que les boucles responsables de la réaction catalytique, d'activation et de la liaison avec le phosphate soient présentes dans la structure, on observe que les résidus critiques pour l'interaction avec l'ATP/Magnésium et sa catalyse ne sont pas conservés. De plus le domaine de liaison possède une insertion de 18 acides aminés non retrouvée dans les autres membres de la famille. La position relative entre le domaine FRB et Kinase chez Tra1 est également différente par rapport aux structures de DNA-PK & mTOR. En effet, chez Tra1 le domaine FRB est positionné devant le site actif de la kinase alors que chez les deux autres protéines il est éloigné de ce dernier.

Le domaine FATC fait partie intégrante du domaine kinase et vient former un « bouchon » au-dessus d'une grande cavité hydrophobe de la kinase. La perturbation de



**Figure 8 : Mutations de Tra1 interférant avec le ciblage des activateurs**

Les mutations affectant les interactions avec VP16 sont affichées par les sphères rouges, les Mutations affectant les interactions avec Gal4 sont affichées par les sphères vertes et les régions importantes pour les activateurs Gcn4 et Rap1 sont coloriées en violet.

cet élément structural est susceptible d'entraîner une déstabilisation du domaine kinase ainsi que des domaines adjacents. Cette observation de la structure est corroborée par le fait que des mutations dans le domaine FATC entraînent souvent une perte de viabilité ou de stabilité de toute la protéine (Hoke et al., 2010). Il est supposé également que la mutation du domaine FATC est critique pour la régulation catalytique chez les PIKK. Or le fait que même chez Tra1, qui est une PIKK inactif, les mutations entraînent une perte de stabilité de la protéine suggère que ce domaine est également important pour la stabilité structurale de cette famille.

### **b) Mutation de Tra1 et site de liaison des activateurs**

Pour étudier les potentielles interactions entre Tra1 et les activateurs la **Figure 8** montre les mutations connues de NuA4 qui entraînent une disruption des interactions entre les activateurs et Tra1.

Des mutants de Tra1 qui contiennent des substitutions des acides aminés suivant montrent une perte d'interactions avec l'activateur Gal4 H400Y/D397N/S404F/D469N/V1115I. Ces acides aminés sont localisés dans la première partie du « doigt » de Tra1 et sont exposés au solvant ce qui semblerait indiquer un site de liaison de Gal4 à ce niveau de la protéine. Il est important de noter que ces mutations sont spécifiques à Gal4 et ne semblent jouer aucun rôle pour l'interaction avec d'autres activateurs.

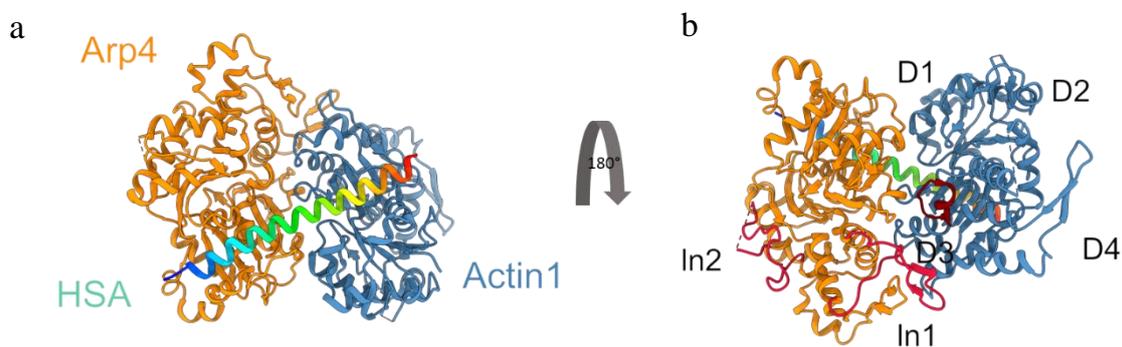
La délétion des résidus 88 à 165 & 319 à 399 entraîne également l'altération du mécanisme de recrutement des activateurs Gcn4 et Rap1, mais n'empêche cependant pas Gal4 d'être recruté. Ces délétions sont retrouvées dans la partie N-terminal de Tra1 mais sont bien spécifiques de ces deux activateurs, ce qui suggère que le doigt de Tra1 possède plusieurs sites de liaisons, chacun étant spécifique à certains activateurs.

Le doigt de Tra1 et son anneau interagissent entre eux, formant une sorte de croisement où plusieurs HEAT de l'anneau interagissent avec la première moitié de la région du doigt. Ce croisement a été mis en évidence, car plusieurs mutations de Tra1 disruptant l'interaction avec les activateurs y sont présentes **Figure 8**, ce qui suggère que cette zone de Tra1 est nécessaire pour le recrutement/ciblage des activateurs. Cependant il existe d'autres mutations de Tra1 qui entraînent une perte d'interaction avec les activateurs situés dans d'autre portion de la protéine, notamment la délétion des résidus

1424 à 1508 qui engendre de nouveau une disruption de l'interaction des activateur Gcn4 et Rap1. Les différentes mutations situées dans plusieurs régions de Tra1 altérant l'interaction avec VP16 suggèrent qu'il existe plusieurs sites importants pour le recrutement des activateurs.

## 2. Le module actine

Le module actine est composé de l'Actine1, Arp4, Yaf9 et swc4. Ce module est de manière intéressante présente chez NuA4 mais également dans le complexe Swr1 qui permet l'échange de l'histone H2A canonique par son variant H2AZ (Cao et al., 2016). Bien que ces deux complexes possèdent des fonctions différentes ce sont tous deux des coactivateurs impliqués dans la régulation de la compaction de la chromatine, ce qui suggère que dans les deux complexes le module actine possède une fonction similaire, sûrement liée à leurs interactions avec la chromatine.



**Figure 9 : Structure cristallographique du complexe ternaire d'actine nucléaire**

Sur l'image (a) on peut observer, que le domaine HSA qui est coloré de sa partie N-ter en bleu à sa partie C-ter en rouge, Arp4 qui est fixé sur la partie N-terminal de la HSA et l'Actine1 qui se positionne sur sa partie C-ter. La figure (b) montre l'autre côté du modèle atomique du complexe on peut y voir annoté les différents lobes de l'Actine D1, D2, D3 & D4 ; Dans le domaine D2 les pointillés indiquent la présence d'une boucle désordonnée importante pour l'interaction avec la chromatine.

D'après (Cao et al., 2016)

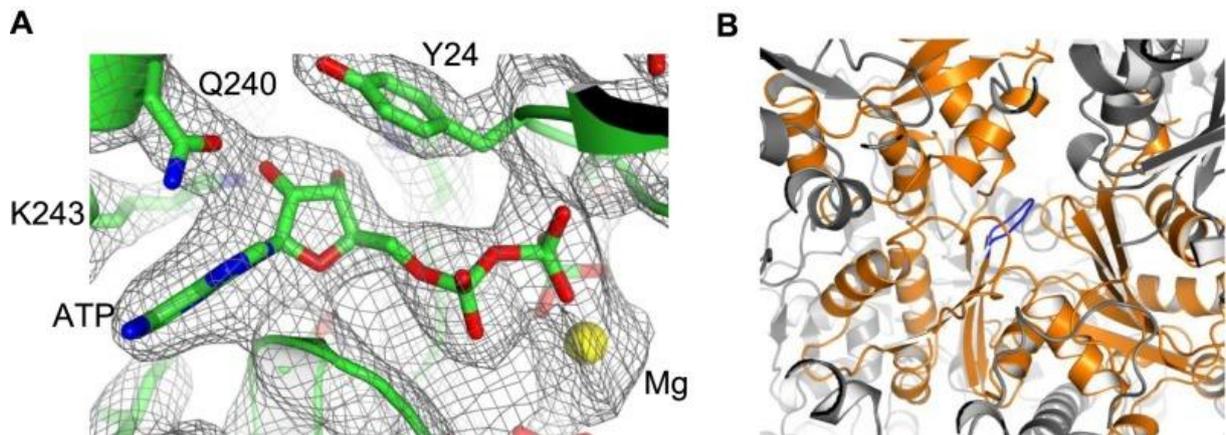
## a) Structure d'un complexe ternaire d'actine nucléaire Actin1/Arp4/ HSA

L'actine est connue pour être un acteur majeur de l'architecture et du mouvement cellulaire. Sous forme monomérique elle est appelée actine-G pour actine globulaire et sous forme filamenteuse elle est appelée actine-F. La liaison de l'ATP à l'actine aide à la formation de sa forme filamenteuse alors que sa liaison avec l'ADP favorise sa dépolymérisation. Il était autrefois contesté que l'actine soit présente dans le noyau, cependant elle a été retrouvée dans plusieurs complexes impliqués dans la régulation de la chromatine comme INO80, SWR1 & NuA4.

Bien que les protéines Actine1 et Arp4 soient très conservées au cours de l'évolution, leurs fonctions au sein des complexes régulant la chromatine restent peu connues. Cependant on sait que ces deux protéines sont incorporées par un domaine HSA (Helical SANT domaine Associated) dans les complexes NuA4 et Swr1 et que d'autres domaines HSA recrutent des structures similaires chez d'autres protéines.

Une structure cristallographique de Actin1, Arp4 et d'un domaine HSA a été obtenue par (Cao et al., 2016). Dans ce complexe ternaire Arp4, Actine1, l'assemblage est majoritairement guidé par le domaine HSA, sa partie N-terminale interagit avec Arp4 et sa partie C-terminal avec l'Actine1.

Actine1 interagit avec le domaine amphipathique du domaine HSA par le biais d'une fosse hydrophobe. Dans le domaine D2 de l'actine, la boucle affichée en pointillé sur l'image **Figure 9** est désordonnée. Cette boucle est supposée être importante pour l'interaction avec la chromatine. En effet la mutation A58T induit une perte d'interaction pour le complexe INO80 avec l'ADN libre et le nucléosome (Kapoor et al., 2013). Le domaine 3 de l'Actine est quant à lui la partie de la protéine qui fait le plus d'interaction avec Arp4. C'est le domaine HSA et Arp4 qui gêne physiquement les interactions que pourraient avoir Actine1 avec d'autres actines et préviennent donc la polymérisation de celle-ci (Tingting Cao et al.).



**Figure 10 : Analyse structurale de la poche de liaison à l'ATP d 'Arp4 et de la N-actine.**

(A) Superposition de la structure du complexe ternaire de l'actine avec la carte de densité électronique au niveau de la poche de liaison d'ATP de Arp4. L'ATP est représentée sous forme de bâtonnets, et le Mg sous forme d'une sphère jaune. (B) Environnement de la poche de l'actine se liant à l'ATP. La boucle Ser155-Thr160, qui est incompatible avec la liaison des nucléotides à l'actine, est colorée en bleu, et les molécules environnantes sont colorées en gris. La boucle Ser155-Thr160 est enfouie dans la fente du domaine principal.

D'après Tingting Cao et al. 2016

D'après la structure cristallographique **Figure 10** la poche où vient se loger l'ATP dans l'actine est altérée, bloquée par le mouvement de la boucle (S155-T160) qui est dans le domaine D3. En effet la chaîne latérale de l'asparagine 157 prend une position qui prévient la fixation de l'ATP dans la poche de liaison. Il est surprenant de noter cette absence de nucléotide avec l'actine car il est connu que cette absence provoque une rapide dénaturation de la protéine (Asakura, 1961).

En revanche Arp4 possède dans sa poche de liaison une molécule d'ATP, ce qui permet aux auteurs d'interpréter que le manque ATP chez l'Actin1 n'est pas un artéfact de purification ou de cristallisation, mais belle et bien une spécificité de l'actine quand elle est présente dans des complexes nucléaires.

### 3. Le module TINTIN

TINTIN est un sous complexe de NuA4 composé des 3 sous-unités Eaf3/5/7 et signifie « **T**ramer **I**ndependent of **N**uA4 involved in **T**ranscription **I**nteractions with

Nucleosomes ». À l'instar d'autres protéines de NuA4, Eaf3 est présente dans un autre complexe, qui est Rpd3S dont l'activité est opposée à NuA4 puisqu'il permet la désacétylation du nucléosome, ce qui engendre un repliement de la chromatine et réprime la transcription (Carrozza et al., 2005);(Joshi & Struhl, 2005);(Keogh et al., 2005), le chromodomaine de Eaf3 interagit avec H3K36me3 et recrute Rpd3S dans les régions codantes des gènes.

Chez NuA4 le rôle de TINTIN serait important pour la capacité du complexe à interagir avec la chromatine puisqu'une délétion de Eaf3/5/7 entraîne une diminution d'interaction avec le nucléosome (Sathianathan et al., 2016).

D'autres éléments montrent que le module TINTIN existe également comme complexe indépendant à NuA4 (Cheng & Côté, 2014) et est retrouvé dans les régions codantes actives transcriptionnellement et interagit avec l'ARN polymérase II. Cette interaction facilite l'élongation en favorisant la rupture des nucléosomes sur le passage de la polymérase, mais aussi à leurs replis après l'élongation (Rossetto et al., 2014).

D'après la littérature, chez la levure les 83 premiers acides aminés de la partie N-terminal de Eaf1 permettent l'interaction avec le module TINTIN (Eaf5/3/7) (Setiaputra et al., 2018) et son incorporation dans NuA4. Cette interaction se fait par le biais de Eaf5 qui permet d'ancrer le trimère. Eaf7 permet de faire le lien entre Eaf5 et Eaf3.

Chez l'humain Eaf3 et Eaf7 sont tous les deux conservés au cours de l'évolution sous le nom de MRG15 et MRGBP et sont retrouvés dans l'équivalent humain de NuA4 qui est Tip60, contrairement à Eaf5 qui n'est pas conservé.

#### 4. Le module d'interaction, Eaf1

Bien que la sous-unité Eaf1 n'est pas essentielle à la survie de la levure, c'est elle qui permet de garder l'intégrité structural du complexe : en effet elle est connue pour interagir avec les différentes sous-unités de NuA4 (Auger et al., 2008). Eaf1 possède trois domaines, un domaine HSA pour « Helical SANT domaine Associated »

qui permet l'association des protéines Actin1 et Arp4 avec le complexe, mais dont la délétion entraîne bien sur une perte d'interaction avec ces sous-unités, mais également la libération du module Piccolo de NuA4 montrant bien l'importance d'Eaf1 pour le maintien structurel du complexe (Auger et al., 2008). Le deuxième domaine est le SANT qui est un motif de trois hélices alpha contenant de grands résidus hydrophobes. Les domaines SANT sont également prédits pour interagir avec la chromatine, mais également faire des contacts protéine-protéine. Le dernier domaine est un domaine riche en glutamine dont la fonction reste inconnue.

## 5. Le module d'acétylation des histones, Piccolo

Esa1, Yng2 Eaf6 & EpL1 forment à eux quatre le module catalytique de NuA4 appelé Piccolo. C'est la protéine Esa1 qui est responsable de l'activité acétyl transférase, elle est cependant incapable d'acétyler la queue des histones du nucléosome seule. Il est nécessaire qu'Esa1 interagisse avec les trois autres sous-unités du module pour qu'elle soit capable d'acétyler correctement le nucléosome.

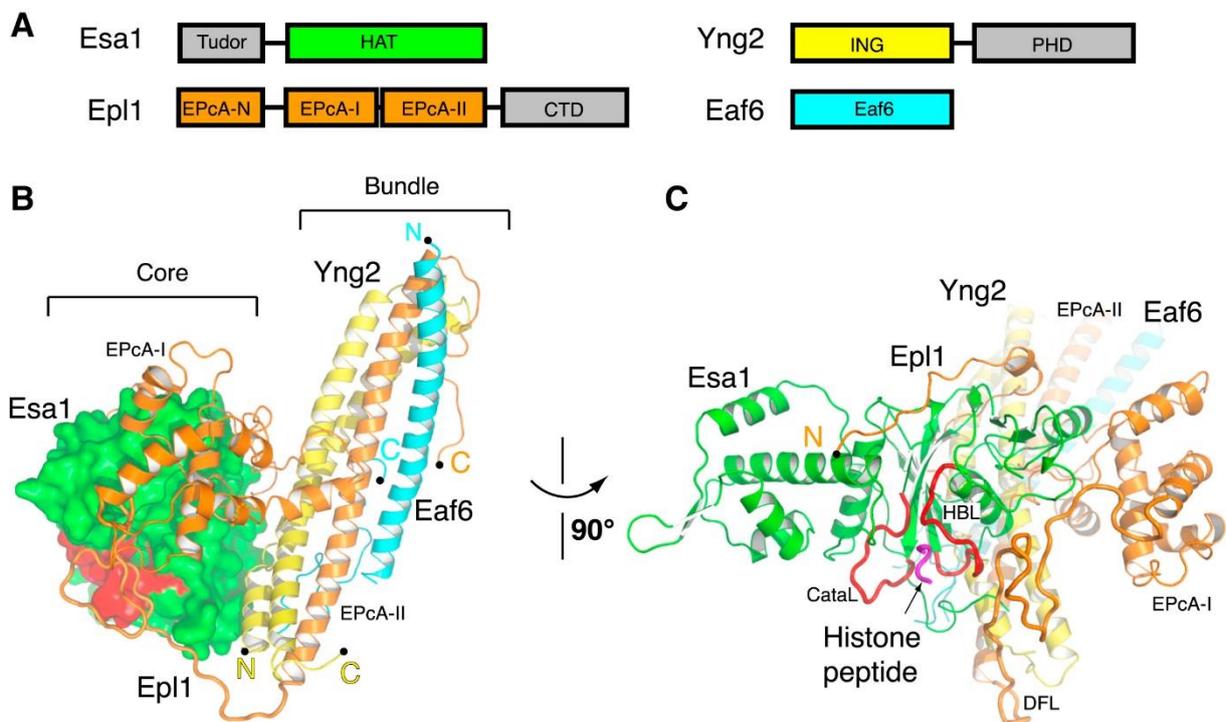
Dans le cas d'autres HAT comme GCN5, il est montré que l'acétylation de l'histone H3 se fait de manière spécifique grâce aux séquences situées de part et d'autre de la lysine 14. Dans le cas de Esa1 c'est seulement le résidu situé en position -1 de la lysine acétylée qui est important pour la réaction. En effet la présence d'une glycine ou d'une alanine en position -1 va permettre l'acétylation alors qu'un acide aminé d'un plus grand encombrement sphérique va quant à lui empêcher le fonctionnement de l'enzyme (Xu et al., 2016). Cependant NuA4 est spécifique des histone H4 et H2A, or H2B et H3 possèdent des lysines qui sont également précédées d'une glycine ou d'une alanine, alors pourquoi ces lysines ne sont-elles pas également acétylées par NuA4 ? Comment NuA4 acétyle de manière spécifique H4 ?

La spécificité de NuA4 à H4 ne se fait pas seulement par la séquence de son substrat mais également grâce à l'orientation avec laquelle il va interagir avec le nucléosome. Piccolo se fixe avec le nucléosome sur sa face plate, cette interaction permet de positionner le site catalytique d'Esa1 proche de la queue de l'histone H4 qui

va alors pouvoir être acétylée. L'acétylation d'H4 par NuA4 est donc un mécanisme qui utilise une double reconnaissance : l'une spatiale et l'autre de la séquence du substrat. C'est un nouveau mode de fonctionnement des HAT qui est proposé où l'orientation de Piccolo sur le nucléosome est ce qui donne la spécificité de NuA4 pour l'histone H4 (Xu et al., 2016).

## 6. Structure de Piccolo

La structure atomique obtenue par la diffraction aux rayons X a permis de comprendre l'architecture et le mode de fonctionnement du module catalytique de NuA4 (Xu et al., 2016). Cette structure se compose d'Esa1 dont le domaine Tudor a



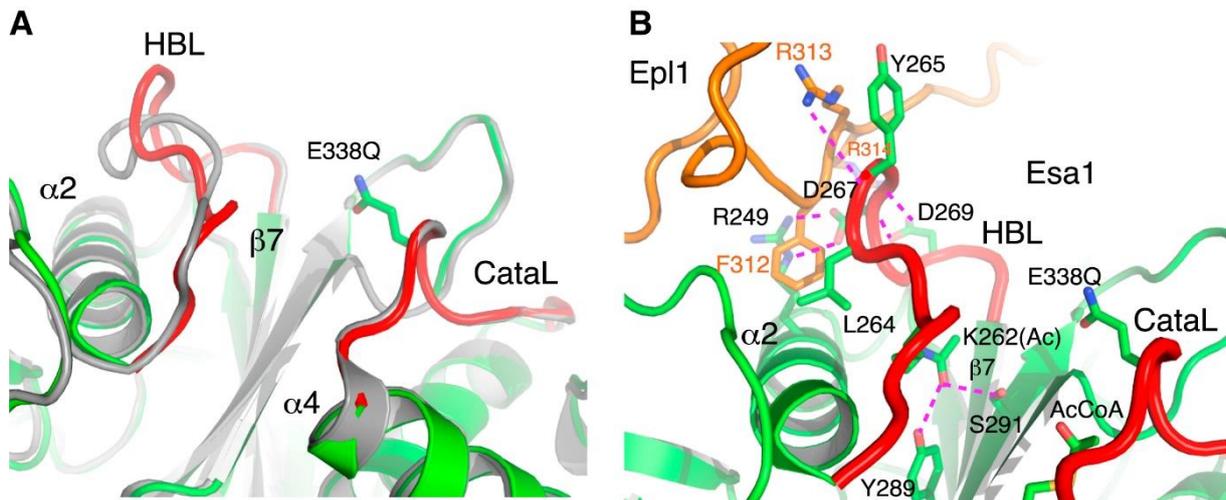
**Figure 11 : Structure cristallographique du module Piccolo**

(A) Les domaines coloriés en gris montrent les domaines absents dans la structure. (B) L'image montre que la structure peut être décomposée en deux parties, le Core composé d'Esa1 et la partie N-ter d'Epl1 et le bundle composé des hélices alphas de Yng2, Eaf6 et du reste d'Epl1. (C) La boucle notée DFL (Dual Fonction Loop) d'Epl1 interagit avec la boucle en rouge d'ESA1 notée HBL pour Histone Binding Loop, qui est importante pour l'interaction avec les histones. En rouge et noté Catal la boucle de Esa1 importante pour la réaction d'acétylation, en violet on voit le peptide d'histone pris en sandwich entre la HBL & Catal.

D'après (Xu et al., 2016)

été retiré, de Eaf6, Yng2 et d'Ep11 qui lui aussi a été tronqué puisque tout la partie C-ter non essentielle à la levure est supprimée.

La partie N-terminale d'EpL1 effectue le plus clair de ses interactions avec la sous-unité catalytique Esa1 et vient changer la conformation du site catalytique.



**Figure 12 : Changements conformationnels de HBL d'Esa1**

(A) Superposition des structures de la poche catalytique de l'Esa1 dans Piccolo et à l'état libre. Esa1 seule est colorée en gris (PDB : 3T09). Le Catal et le HBL sont colorés en rouge. (B) Interactions entre les résidus FRR d'Epl1 et le HBL de Esa1. Les interactions par liaison hydrogène sont représentées par des lignes en pointillées.

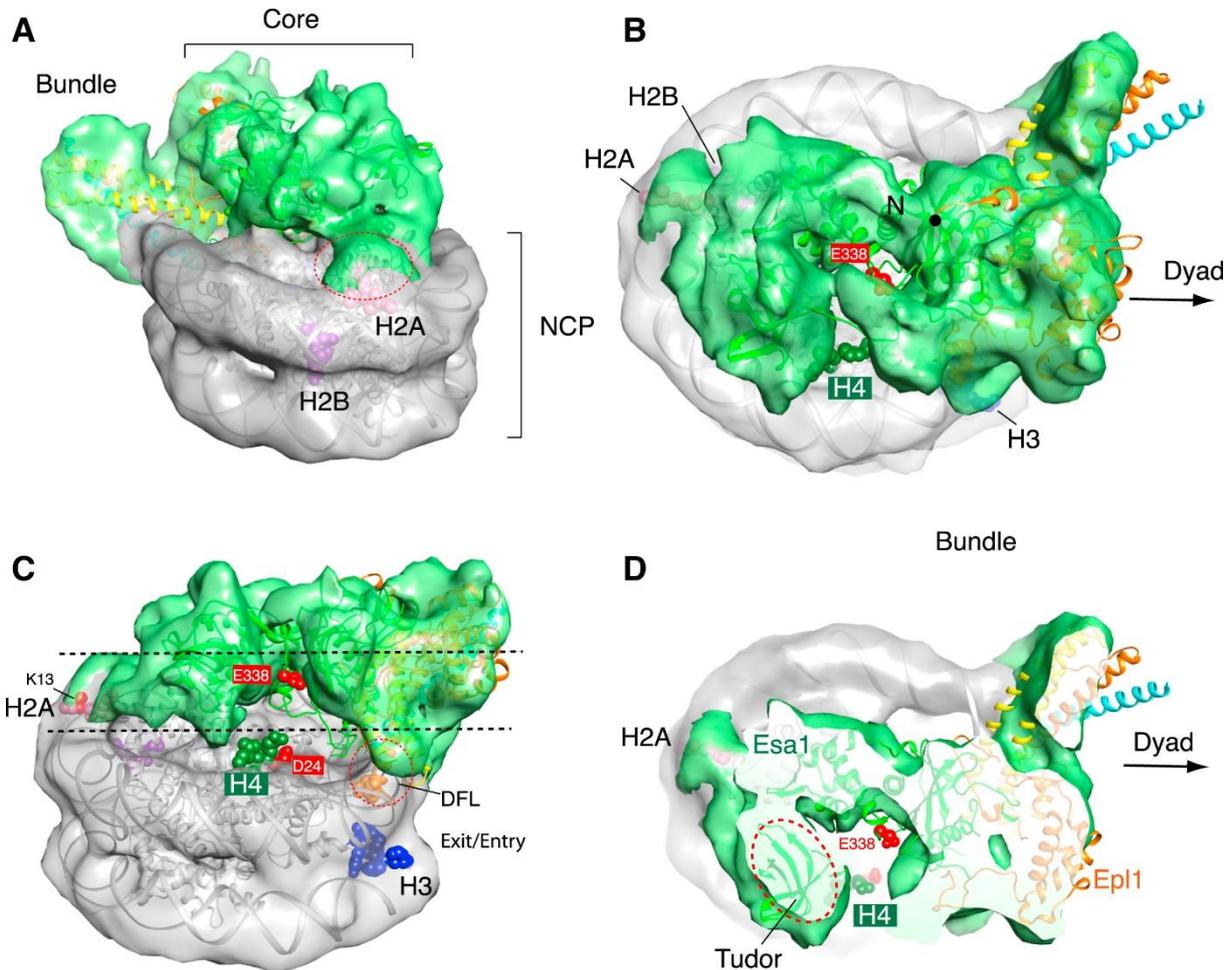
D'après (Xu et al., 2016)

Les résidus F312, R313 R314 de la DFL de EpL1 interagissent et déforment la HBL du site catalytique d'Esa1 ce qui a pour cause d'exposer au solvant de nouveaux résidus ce qui génère potentiellement une nouvelle interface de liaison.

## 7. Structure Cryo-EM avec le nucléosome

Une structure par Cryo-EM de Piccolo interagissant avec le nucléosome a été obtenue par Peng Xu et al en 2016. Cette structure a une résolution de  $\sim 8 \text{ \AA}$  et permet d'étudier la forme globale que prend le complexe. Ainsi, le modèle atomique du nucléosome, de Piccolo et du domaine Tudor d'Esa1 ont pu être placés dans la densité de la carte Cryo-EM (**Figure 13**). Cette carte permet de voir que lorsque Piccolo interagit avec le NCP celui-ci vient placer H4 juste devant le site catalytique. De plus

une partie de la DFL semble également importante pour l'interaction avec le nucléosome, notamment les résidus R321/K322, ce qui est vérifié par biochimie où l'on observe que la mutation de ces résidus réduit l'acétylation du nucléosome. Ceci démontre que ces deux acides aminés d'Ep11 sont essentiels pour le mécanisme de reconnaissance du nucléosome, soulignant l'importance d'EpL1 pour la levure.



**Figure 13 : Structure Cryo-EM de Piccolo avec le nucléosome**

Vue latérale de la carte de Cryo-EM ou les structures cristallines de Piccolo, de la séquence 601 NCP (PDB : 3MVD) et à la structure RMN du domaine Tudor d'Esa1 (PDB : 2RO0) ont été placés. La densité correspondant à Piccolo et le NCP sont respectivement coloriés en vert et gris. Les queues N-terminales de H4, H3, H2A et H2B sont présentées sous forme de sphères et respectivement colorées en vert foncé, bleu, rose et violet. Le site de contact du nucléosome de Piccolo se fait au-dessus du dimère H2A-H2B et est mis en évidence par un cercle rouge pointillé. (B) Vue du dessus de la structure Cryo-EM. Le résidu catalytique (E338) d'Esa1 se présente sous forme de sphères rouges. (C) Vue de la poche catalytique. Le site de contact du nucléosome-DFL (R321/K322) est mis en évidence par un cercle. (D) Vue d'une tranche présente dans (C) orientée de manière perpendiculaire.

D'après (Xu et al., 2016)

Un autre site de liaison entre Piccolo et le NCP se fait sur la région acide du dimère H2A/H2B, qui est connue pour être une surface d'interactions de plusieurs protéines interagissant avec le nucléosome (Luger et al., 2012)(McGinty et al., 2014), ce qui suggère que la reconnaissance du NCP se fait grâce à cette région qui interagirait avec le domaine EpcA-N. Les mesures de Kd entre le domaine EpcA-N et H2A/H2B montrent que dans le cas où les résidus acides du dimère d'histone sont mutés, le Kd change d'un facteur de 16 par rapport au WT. De plus il a été montré que l'arginine 58 d'Epl1 joue un rôle particulièrement important pour l'activité de l'enzyme (Huang & Tan, 2013) ce qui suggère qu'elle pourrait permettre l'ancrage de Piccolo au NCP.

## 8. Piccolo dans NuA4

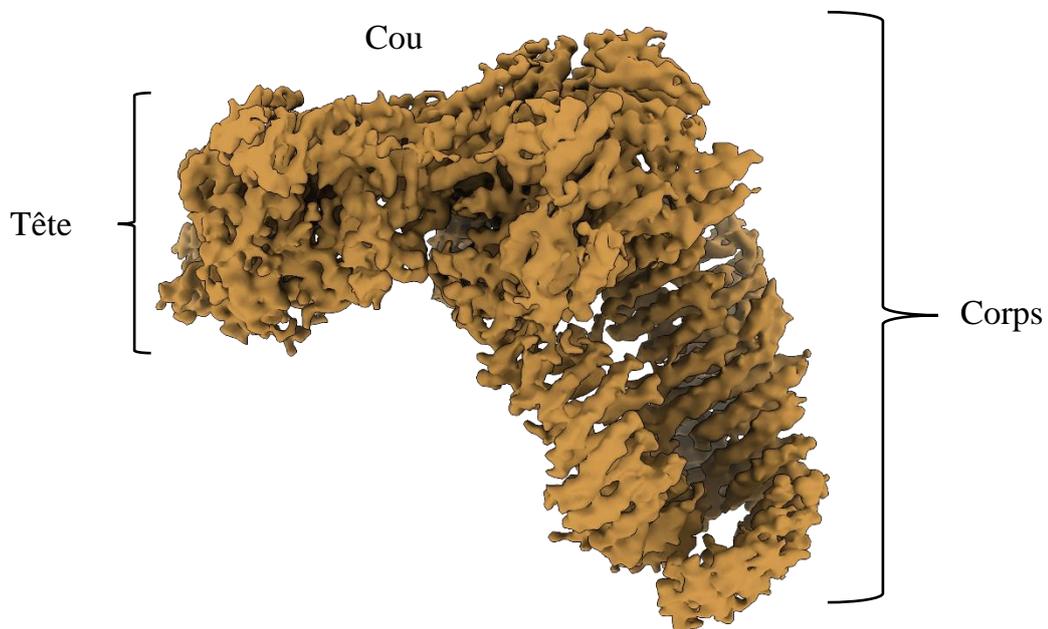
Il est proposé que Piccolo peut évoluer de manière indépendante à NuA4 et serait responsable de l'acétylation non spécifique de la chromatine, alors que lorsqu'il est incorporé dans NuA4 il permet une acétylation des nucléosomes plus spécifique. Cette idée vient du fait que lors de la purification de NuA4 les fractions obtenues contenaient un grand complexe capable d'acétyler le NCP, mais également un complexe beaucoup plus petit en taille mais ayant la même habilité à acétyler les histone H4 & H2A du NCP, ce sous-complexe fut appelé Piccolo pour Pic NuA4.

La sous unité qui permet d'ancrer Piccolo dans Nua4 est Epl1. Comme vu précédemment sa partie N-terminal joue un grand rôle dans ses interactions avec le reste des protéines de Piccolo, mais c'est sa partie C-terminal qui permet l'interaction avec NuA4. Dans la publication de (Boudreault et al., 2003) les auteurs se sont intéressés aux différentes parties de la séquence afin de déterminer leurs fonctions. Ainsi plusieurs mutants ont été obtenus : l'une dont toute la partie C-terminale est tronquée, et l'autre où la partie N-terminale est supprimée. Les résultats de ces mutations montrent que l'absence du N-terminal présent dans Piccolo est létale pour les levures, montrant que ce domaine est essentiel à leur survie. A l'inverse, le mutant dépourvu de la partie c-terminale est viable et dans ce cas la fraction contenant NuA4 lors de la purification perd son activité d'HAT alors que la fraction contenant Piccolo se retrouve enrichie par

ce dernier, ceci montre que c'est la partie C-ter de Epl1 qui permet l'incorporation de Piccolo dans le complexe NuA4.

## 9. Structure Cryo-EM du complexe TEEAA

La structure du complexe TEEAA est la première qui permet de comprendre l'architecture de base de NuA4 (X. Wang et al., 2018), cette structure obtenue par cryo-EM d'une résolution de 4,7 Å a permis de créer un premier modèle atomique de NuA4. Cependant ce modèle est incomplet : une bonne partie des sous-unités est absente,

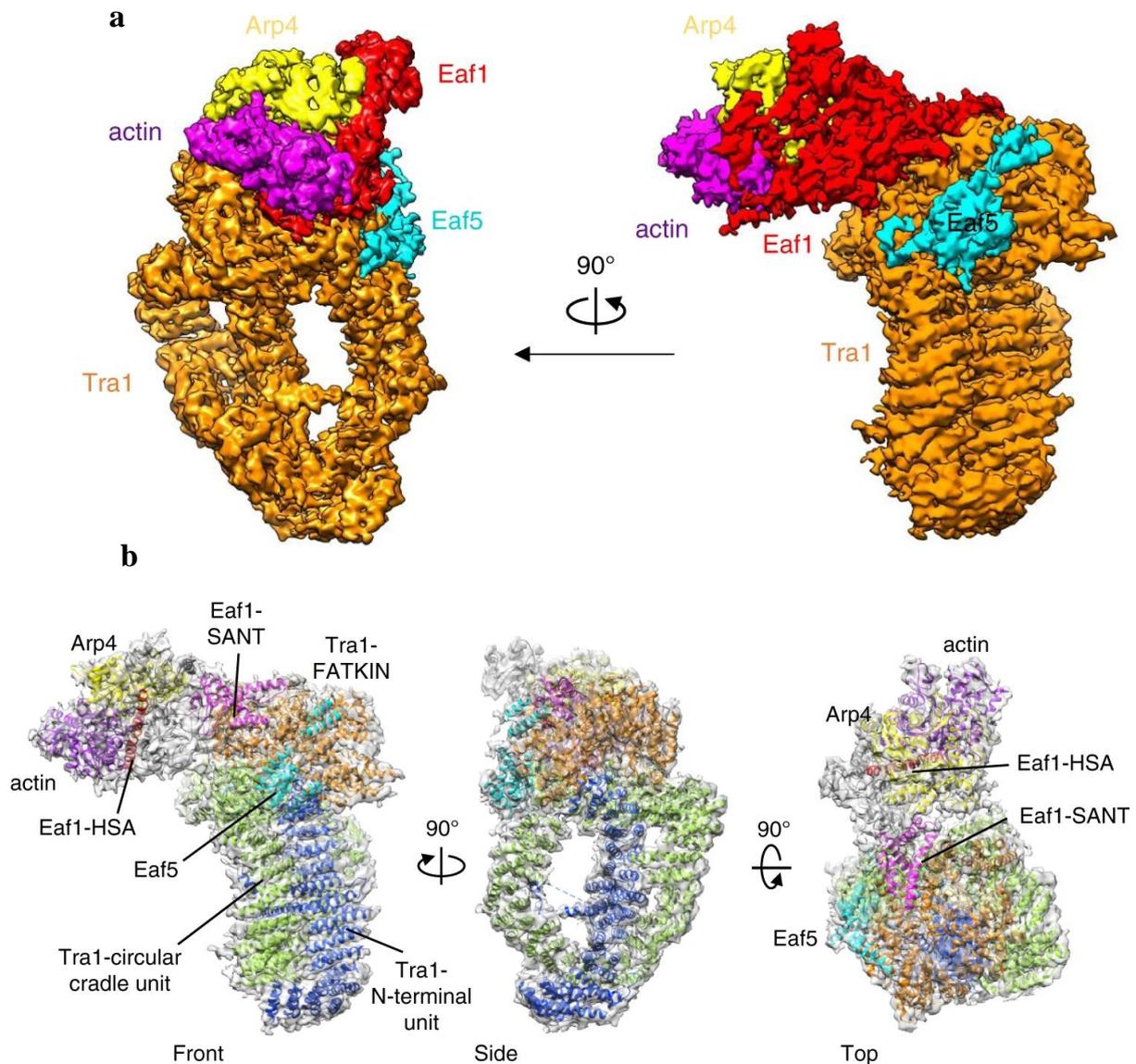


**Figure 14 : Structure cryo-EM du complexe TEEAA**

Le complexe se compose de trois domaines le corps, la tête et le cou

puisque seule 5 des 13 protéines de NuA4 ont pu être construites, Tra1, Eaf1, Eaf5, Arp4 & Actin1 dont l'acronyme forme le nom de la structure TEEAA. Le complexe TEEAA possède une tête, un cou et un corps. Les structures cristallographiques de Arp4 & Actin1, ont pu être placées dans la densité correspondant à la tête. Ce corps quant à lui est composé d'une grande majorité de Tra1 et de quelques hélices de Eaf5. Enfin, d'après les résultats le cou est exclusivement composé de la sous-unité essentielle à la

stabilité de NuA4 Eaf1. Cependant la résolution de la structure n'est pas suffisante pour permettre d'assigner avec certitude la séquence protéique à la carte de densité notamment pour Eaf1 & Eaf5.



**Figure 15 : Structure Cryo-EM du complexe TEEAA**

(a) La carte Cryo-EM est colorée suivant les sous-unités lui correspondant, on peut voir 5 protéines sur les 13 qui compose NuA4. (b) Modèle atomique de NuA4 placé dans la carte de Cryo-EM, différentes orientations sont affichées pour apprécier la structure sous différentes vues.

D'après (X. Wang et al., 2018)

### **a) Actine1 et Arp4 positionnés sur le domaine HSA d'Eaf1**

L'Actin1 et Arp4 sont retrouvés dans la région appelée « la tête de la structure » et viennent s'accrocher au reste de NuA4 par le biais du domaine HSA de Eaf1 de manière similaire à la structure du complexe ternaire d'actine nucléaire. Les résidus L362, M369 & F373 de la HSA font des contacts hydrophobes avec les résidus M1 et L462 de Arp4. Pour l'actine 1 ce sont les résidus F169, F375, M355, F352, I345 & L349 qui eux aussi forment des contacts hydrophobes avec I383, M387 & I391 du domaine HSA.

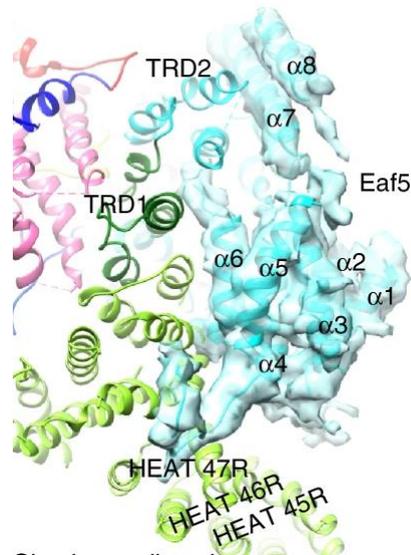
Il est connu qu'une boucle désordonnée, située dans le domaine D2 de l'Actin1, est importante pour l'interaction avec la chromatine (Kapoor et al., 2013). Dans le complexe TEEAA cette boucle est bien exposée au solvant, ce qui est en accord avec l'idée que cette région est importante pour l'interaction de NuA4 avec son substrat.

### **b) Implémentation d'Eaf1**

D'après la structure de TEEAA tout le cou de la structure appartient à Eaf1, qui pour rappel est la protéine qui est censé faire les liens structuraux avec les autres parties du complexe. Les interactions entre Eaf1 et Tra1 se font principalement par des interactions ioniques et hydrogène entre le domaine SANT de Eaf1 et les résidus E3512, D3605 & R3731 de Tra1, mais également avec une portion du domaine FAT, montrant qu'il existe de multiples interactions entre les deux protéines.

### c) Eaf5 ancre TinTin sur Tra1

La dernière protéine présente dans la structure du TEEAA est Eaf5, plusieurs hélices situées dans le N-ter de Eaf5 viennent interagir avec le domaine FAT de Tra1 ainsi qu'avec ses domaines HEAT. Le fait que Eaf5 effectue la majeure partie de ses



**Figure 16 : Surface d'interaction d'Eaf5 avec Tra1**

Eaf5 interagit avec le reste de NuA4 par le biais de la sous-unité Tra1, notamment avec son domaine FAT ainsi que quelques motifs HEAT

D'après (X. Wang et al., 2018)

interactions avec Tra1, mais puisse également interagir avec Eaf1 peut expliquer la capacité du sous complexe TINTIN, dont Eaf5 est un des membres, à se dissocier de NuA4. Cela pourrait également impliquer que des facteurs de transcription puissent moduler son interaction avec NuA4.

### d) L'incorporation du module catalytique Piccolo

D'après les auteurs du papier sur le complexe TEEAA, ils ont pu isoler une petite population de NuA4 avec Piccolo grâce à des étapes de classification 3D, ce qui leur a permis d'obtenir une carte de microscopie électronique à 7.6 Å de résolution. Cette carte montre que Piccolo devrait être situé entre Tra1 et les protéines Arp4/Actine1 dans la

région du cou de NuA4. Cependant la qualité de la carte ne permet pas d'être sûr avec exactitude que cette région est bien là où vient interagir Piccolo.

## 10. Objectif

Comme nous avons pu le voir il existe beaucoup de structures de différentes sous-unités/module de NuA4, mais il n'existe aucune structure à haute résolution du complexe, limitant grandement sa compréhension. Pour mieux comprendre l'architecture et l'interaction entre les différents constituants de NuA4, nous cherchons à obtenir une structure de ce dernier grâce à la cryo microscopie électronique. L'obtention de tels résultats permettrait d'établir le premier modèle atomique de NuA4. Pour étudier comment le complexe est capable d'acétyler le nucléosome nous chercherons également à déterminer une structure de NuA4 liée au nucléosome.

## C. Le coactivateur SAGA

SAGA pour Spt-Ada-Gcn5-Acétyletransférase est un complexe protéique impliqué dans les modifications d'histones et l'assemblage du complexe de pré initiation de la transcription. Il est composé de 19 sous-unités et est très conservé au cours de l'évolution, de la levure à l'humain, que ce soit en termes de fonction et de composition. Il possède deux activités enzymatiques qui sont l'acétylation des histones, par sa sous-unité Gcn5, qui acétyle spécifiquement H2B & H3, et la dé ubiquitination réalisée par le module DUB décrit plus tard dans les prochaines sections. Les 19 sous-unités qui forment le complexe peuvent être séparés en 5 modules : le module HAT composé de Gcn5, Ada2, Ada3 et Sgf29, le module DUB composé de Ubp8, Sus1, Sgf11 et de Sgf73, le module TAF comportant TAF5/6/9/10/12, le module Spt avec Spt3/7/8/20 et le module d'interaction avec les activateurs composés d'une unique protéine qui est Tra1. Chaque module sera décrit dans une partie qui lui sera dédiée. De la même manière que NuA4, SAGA possède des protéines qui sont également présentes dans d'autres complexes, Tra1 chez NuA4, mais également TAF5/6/9/10/12 qui sont aussi présentes chez TFIID.

### 1. Le module d'interaction des activateurs : Tra1

Tra1 a déjà été présenté dans le chapitre consacré à Nua4, mais on peut insister sur le fait que les deux complexes possèdent la même sous-unité pour interagir avec les activateurs alors qu'ils ne sont pas recrutés sur les mêmes portions du génome. La manière dont est intégré Tra1 dans SAGA & NuA4 joue peut-être un rôle, une différence d'incorporation pourrait peut-être permettre d'engendrer des surfaces d'intégrations différentes avec les activateurs. L'étude structurale de ces deux coactivateurs pourrait donc conduire à une meilleure compréhension du rôle de Tra1 en ce qui concerne sa fonction d'interaction avec les activateurs de la transcription.

<b>Module</b>	<b><i>Saccharomyces cerevisiae</i></b>	<b><i>Schizosaccharomyces pombe</i></b>	<b><i>Drosophila melanogaster</i></b>	<b><i>Homo sapiens</i></b>
HAT module	Gcn5p	Gcn5	dKAT2 (dGcn5)	KAT2A (GCN5)
	Ada2p	Ada2	dAda2b	TADA2b
	Ada3p (Ngg1p)	Ada3 (Ngg1)	dAda3	TADA3
	Sgf29p	Sgf29	Sgf29	SGF29 (CCDC101)
Core module	Taf5p	Taf5	Wda	TAF5L (PAF65β)
	Taf6p	Taf6	Saf6	TAF6L (PAF65α)
	Taf9p	Taf9	dE(y)1 (Taf9)	TAF9/TAF9b
	Taf10p	Taf10	Taf10b	TAF10 (STAF28)
	Taf12p	Taf12	Taf12	TAF12
	Ada1p	Ada1	Ada1	TADA1 (STAF42)
	Spt7p	Spt7	dSpt7	SUPT7L (STAF65γ)
	Spt20p	Spt20	Spt20	SUPT20H
	Spt3p	Spt3	dSpt3	SUPT3H
	Spt8p	Spt8	–	–
TF-binding module	Tra1p	Tra1	Nipped-A (dTra1)	TRRAP
DUB module	Ubp8p	Ubp8	dNonstop	USP22 (UBP22)
	Sgf11p	Sgf11	dSgf11	ATXN7L3
	Sgf73p	Sgf73	dATXN7	ATXN7 (SCA7)
	Sus1p	Sus1	dE(y)2	ENY2

**Tableau 2 : Conservation des différents modules/sous-unités de SAGA au cours de l'évolution**

## 2. Le module TAF, de SAGA a TFIID

A l'origine, les protéines Taf ont été découvertes en 1991 par (Dymlacht et al., 1991), dans le complexe TFIID et interagissaient avec la protéine TBP, un facteur de transcription permettant la mise en place du PIC. Ce n'est que quelques années plus tard, que des expériences menées par (Grant et al., 1998) ont montré que le complexe SAGA possédait également des sous-unités TAF, mais contrairement à TFIID qui possède 13 Taf, SAGA n'en possède que 5 qui sont Taf 5, Taf 6, Taf 9, Taf 10, Taf 12. Chaque taf possède un Histone Fold Domaine abrégé HDF qui sont responsables de la formation de l'octamère histone dans le cas du nucléosome. Chez TFIID il existe également un octamère d'histones formé par deux hétérodimères de TAF6/9-Taf4/Taf12. Cependant chez SAGA le dimère Taf4/Taf9 n'existe pas car Taf4 n'est pas présent dans le complexe et est remplacé par le HDF de Ada1. Il a été montré que les HDF de Taf10 et Spt7 dimérisent entre eux.

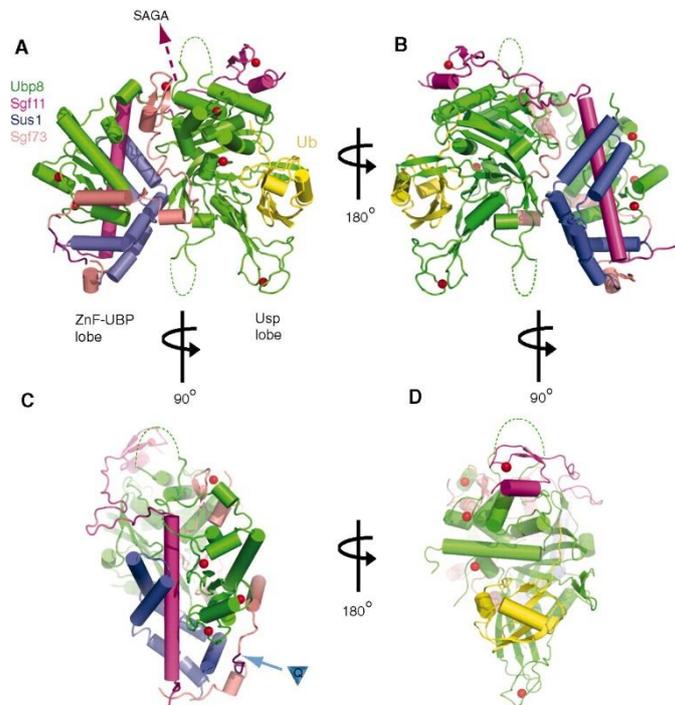
## 3. Le module de dé ubiquitination

SAGA est capable de réaliser la dé ubiquitination des histones grâce à son module DUB composé de Ubp8, Sgf73, Sgf11 et sus1. La protéine réalisant l'activité est Ubp8 et a été découverte en 2000 (Amerik et al., 2000), mais c'est seulement quelques années plus tard qu'elle a pu être identifiée comme composant du complexe SAGA où elle est capable de dé ubiquitiner les histones H2B *in vitro* (Henry et al., 2003)(Daniel et al., 2004). Sgf11 et Sgf73 sont découverts par des analyses biochimiques(Gavin et al., 2002; Helmlinger et al., 2004; Powell et al., 2004; Sanders et al., 2002) de même pour sus1(Rodríguez-Navarro et al., 2004). C'est les laboratoires Berger and Workman qui ont montré que Sgf11 était un composant de SAGA, qu'il faisait partie du module DUB et qu'il est également impliqué dans son activité (Köhler et al., 2006). Sgf11 possède également un doigt de zinc sur sa partie N-terminal (Zhao et al., 2008). Sgf73 est responsable de l'interaction du module DUB avec le reste du complexe (Köhler et al., 2008). De manière intéressante ubp8 a besoin d'être en présence des autres éléments du module pour être capable de dé ubiquitiner l'histone H2B.

En règle générale l'ubiquitination d'une protéine est associée à sa dégradation. Cependant dans le cas des histones elle est également utilisée comme marqueur signalétique comme les autres modifications épigénétiques que les histones peuvent subir (acétylation méthylation ect ...), on retrouve dans la chromatine active un enrichissement en histone H2B ubiquitinée.

### a) Structure du module DUB

La première structure du module DUB de SAGA a été obtenue en 2010 par l'équipe Wolberger (Samara et al., 2010) et a permis de mieux comprendre l'organisation et les connections internes de ce sous module de SAGA.



**Figure 17 : Structure du module DUB de SAGA**

(a) Vue de la structure cristallographique du DUB. Le lobe ZnF-UBP et le lobe Usp sont identifiés, les sous-unités sont colorées suivant la légende sur la gauche. Les boucles désordonnées des protéines sont indiquées par des pointillés. (b) Vue du DUB retourné à 180° (c) Vue du lobe Znf-UBP où l'on peut voir que Ubp8 et Sus1 interagissent de chaque côté d'une hélice de Sgf11 qui dirige la structure du lobe. (d) Vue du lobe Usp, la région à laquelle se lie l'ubiquitine est en jaune et le doigt de zinc de Sgf11 est en magenta.

D'après Nadine L. Samara et al., 2010

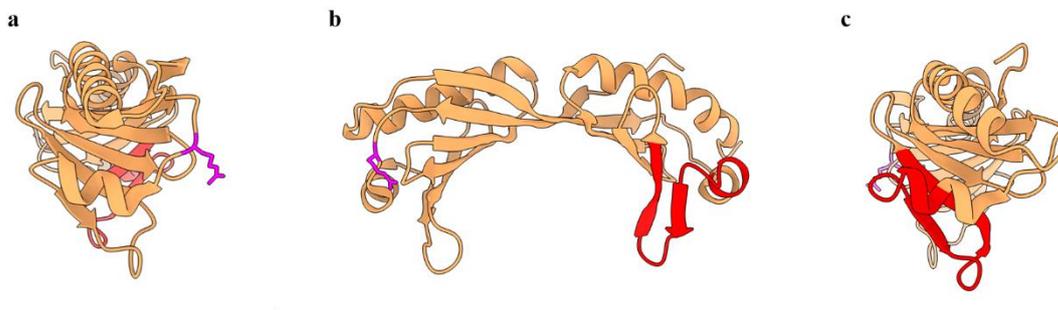
Les premiers éléments que nous montre la structure sont que les protéines du DUB sont fortement liées entre elles, expliquant pourquoi le module a besoin d'être au complet pour réaliser sa tâche. Le sous-complexe est organisé en deux lobes dont chacun d'eux possède un des domaines de ubp8. Le premier lobe est appelé Znf-Ubp, car c'est là où se trouve le doigt de zinc de ubp8 situé dans sa partie n-terminal, le deuxième lobe est appelé USP lobe, et contient cette fois-ci la partie C-ter de ubp8 là où est situé le domaine enzymatique. Le lobe Znf-Ubp est organisé le long d'une hélice de Sgf11, qui s'associe avec Ubp8 d'un côté et avec sus1 de l'autre **Figure 17 c**. Quant à Sgf73 ses ~40 premiers acides aminés interagissent avec le domaine du doigt de zinc de Ubp8 et avec Sus1 par le biais de petite hélice **Figure 17 a & c**. Dans le lobe Usp on peut voir que le doigt de zinc de Sgf11 interagit avec le domaine catalytique de Ubp8 et relie les deux lobes entre eux par une longue boucle d'acides aminés **Figure 17 b**. L'interaction entre les deux lobes est principalement faite par les résidus 45-95 de la protéine Sgf73 qui sont positionnés entre les deux lobes, seule la partie N-terminal de Sgf73 est présente dans la structure, ceci s'explique par le fait que ce sont les résidus C-terminaux qui font l'interaction avec le reste du complexe SAGA (Köhler et al., 2006; Lee et al., 2009).

#### 4. Le module d'interaction avec TBP

Le module d'interaction avec TBP ou module Spt est composé des 4 protéines Spt20, Spt8, Spt7 et Spt3. Le terme Spt vient de Suppressor of Ty, une famille de gènes identifiés à l'origine grâce au dépistage de mutations dans la levure *Saccharomyces cerevisiae*, qui rétablissent l'expression des gènes perturbés par l'insertion d'un transposon Ty (Yamaguchi et al., 2001). Dans cette famille on retrouve les gènes Spt11, Spt12 et Spt15 qui codent respectivement pour H2A, H2B et TBP.

Il a été montré que les protéines Spt8 et Spt3 interagissent toutes les deux directement avec le facteur de transcription TBP, impliquant SAGA dans les mécanismes visant à délivrer TBP. De plus, Spt3 possède deux HFD espacés de ~120 acides aminés : le premier HFD situé dans la partie N-terminale est homologue à Taf13 (Mengus et al., 1995), le second HFD situé dans la partie C-terminale est homologue

à Taf 11 (Birck et al., 1998). Ceci semble montrer que ces deux HFD forment un hétérodimère qui permet d'interagir avec TBP de la même manière que Taf11/Taf13 dans le complexe TFIID (Patel et al., 2018). Il a été également montré que Spt3 et Spt8 sont tous les deux capables d'interagir avec TBP in vitro et in vivo (Eisenmann et al., 1992; Laprade et al., 2007; Sermwittayawong & Tan, 2006; Warfield et al., 2004). La **Figure 18** montre les différents résidus de Spt3 et Spt8 impliqués dans l'interaction avec TBP. Spt3 seule est incapable de lier TBP, ce qui implique probablement que cette sous-unité doit être dans un contexte structural pour lui permettre de réaliser cette tâche.



**Figure 18 : Structure de TBP et ses sites d'interaction avec Spt3 et Spt8**

Structure cristallographique de TBP dont les résidus impliqués dans l'interaction avec Spt3 sont colorés en rouge et correspondent aux résidus 176-196. Le résidu impliqué dans l'interaction avec spt8 est quant à lui coloré en magenta et correspond à l'arginine 79. **b** montre la structure vue de face et **a** et **c** montrent la structure vue de côté.

D'autre part des études par crosslink, méthode qui permet d'étudier la proximité des différentes sous-unités composant un complexe protéique, ont permis de mettre en évidence que Spt7 réalise des crosslink avec TBP, ce qui implique qu'ils sont proches spatialement. Ces résultats montrent que ces 3 constituants du module Spt sont impliqués dans l'interaction avec le facteur de transcription TBP. Pour ce qui concerne la dernière sous-unité du module, Spt20, elle est, tout comme Spt7 est essentielle pour le maintien de l'intégrité du complexe.

Toutes les protéines de ce module sont également essentielles au recrutement de SAGA sur les promoteurs qui lui sont dédié et la mise en place du PIC (Mohibullah &

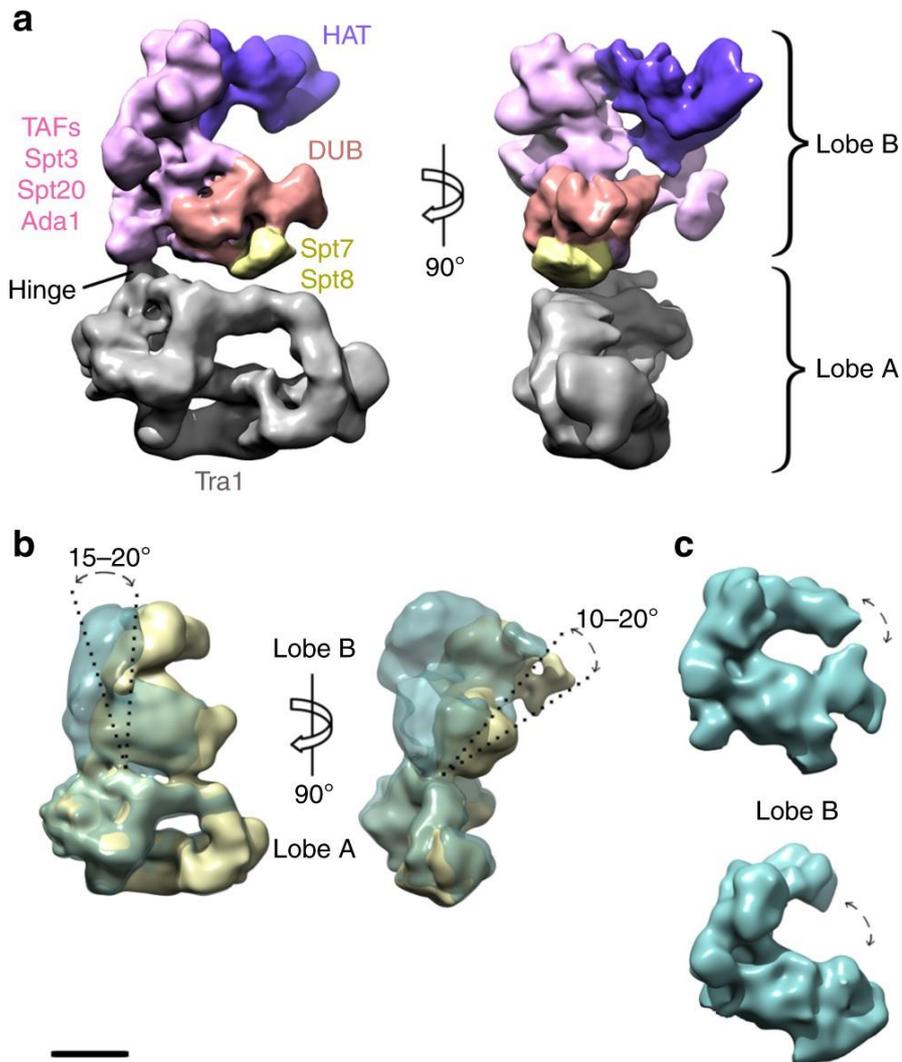
Hahn, 2008; Sterner et al., 1999) montrant leur importance qui peut être structurale et fonctionnelle.

## 5. Le module d'acétylation des histones.

Ce module est composé de 4 protéines : Gcn5 qui est la HAT du complexe, Ada3, Ada2 et Sgf29. Des analyses *in vitro* réalisées sur Gcn5 montrent qu'elle est capable d'acétyler les histones libres, mais qu'elle ne peut pas acétyler les histones dans le contexte du nucléosome (Kuo et al., 1996), pour ce faire il faut qu'elle soit incorporée dans le complexe SAGA (Grant et al., 1997). Gcn5 est composé d'un module catalytique et d'un bromo domaine capable de reconnaître des lysines monos acétylées, ces deux domaines sont conservés au cours de l'évolution chez les eucaryotes supérieurs. Plusieurs études qu'elles soient biochimiques ou génétiques montrent que Gcn5, Ada2 et Ada3 interagissent entre elle ce qui a permis de mettre en évidence l'existence d'un module HAT dans SAGA. La partie C-terminal de Ada3 fait le lien avec Ada2 qui lui-même fait le lien avec Gcn5. Sgf29 est quant à lui capable de reconnaître des histones méthylés notamment H3K4me2/3 (Bian et al., 2011) grâce à son domaine Tudor, ce qui est une plateforme de plus pour reconnaître le nucléosome. Le SANT domaine de Ada2 est indispensable à l'activité HAT du module (Sterner et al., 2002), mais contrairement à ce qui a été cru auparavant, le SANT domaine n'active pas l'activité de Gcn5 en affectant directement la liaison peptidique des histones. Au lieu de cela, le domaine SANT d'Ada2 améliore la liaison de Gcn5 à acétyl-CoA suggérant un nouveau mode de régulation par le contrôle du cosubstrat au lieu du substrat lui-même (Sun et al., 2018).

## 6. Structure Cryo-EM de SAGA

En 2017 l'équipe où j'ai effectué ma thèse a publié une structure Cryo-EM de SAGA à 11.7 Å permettant pour la première fois d'apprécier l'intégration de la sous-unité Tra1 dans le complexe, ainsi que d'autres éléments structuraux (Sharov et al., 2017). La structure de SAGA se compose de deux lobes A & B. Le lobe A est constitué de Tra1 qui est la protéine responsable de l'interaction avec les activateurs, et le lobe B



**Figure 19 : Structure Cryo-EM du coactivateur SAGA**

(a) Carte Cryo-EM du coactivateur SAGA résolue à 11.7 Å de moyenne. Les zones colorées de la carte représentent de manière approximative la position des différentes protéines de SAGA. La structure se compose de deux lobes A qui ne contient que la protéine Tra1 et B comportant le reste de la structure et des modules catalytiques. (b) Les classes obtenues en classifications 3D permettent de montrer le mouvement important qui existe entre les deux lobes. (c) Des mouvements à l'intérieur du lobe B sont également possible, impliquant une grande flexibilité du complexe.

De Sharov et al. 2017

qui comporte le reste des composants du complexe, notamment les modules catalytiques HAT et DUB **Figure 19 a**. Il existe une très forte flexibilité dans le lobe B, puisque suivant les classes on peut observer des variations d'angles entre le lobe A et le lobe B d'une dizaine à une vingtaine de degrés **Figure 19 b**. De plus on peut observer également qu'il existe des mouvements internes importants dans le lobe B **Figure 19 c**.

Pour pallier le mouvement du complexe chaque lobe a été analysé séparément, ce qui a permis d'améliorer grandement la résolution du lobe A en la portant à 5,7 Å, mais n'a malheureusement pas permis d'améliorer le lobe B. Ceci peut s'expliquer par les mouvements internes de ce dernier. La carte du lobe B a permis d'obtenir un premier modèle structural de Tra1 ressemblant à ce qui a été présenté dans la partie de Tra1 dans le chapitre consacré à NuA4.

## 7. Objectif

La faible résolution des structures cryo-EM de SAGA empêche d'établir un modèle atomique précis du complexe, notamment en ce qui concerne le lobe B. Notre équipe cherche à obtenir une structure à haute résolution du complexe dans le but d'en déduire un modèle atomique qui permettra d'appréhender l'architecture de SAGA et ses modes de fonctionnement. Du fait que SAGA et NuA4 partagent tous les deux la protéine Tra1, qui est impliquée dans l'interaction avec les activateurs, la détermination des structures des deux complexes permettrait de comparer l'implémentation de Tra1 dans les deux complexes et de mieux comprendre comment fonctionne ce module de recrutement des activateurs.

## **D. La Microscopie électronique dans l'étude structurale des protéines**

Dans cette partie je présente l'outil qu'est le microscope électronique à transmission et comment il permet d'être utilisé pour étudier la structure 3D des protéines. Pour ce faire je vais en premier lieu parler du fonctionnement du microscope électronique à transmission, comment les échantillons biologiques sont préparés pour passer sous le vide d'un microscope électronique, comment l'acquisition des images est faite au microscope et enfin comment ces images sont traitées pour déterminer la structure 3D des protéines

### **1. Le microscope électronique**

Les microscopes électroniques utilisent un faisceau d'électrons à haute énergie, qui va interagir avec l'échantillon afin de former une image de celui-ci. Leur développement a commencé dans les années 1930 par les deux ingénieurs allemand Ernst Ruska & Max Knoll. L'intérêt de développer de tels microscopes réside dans les propriétés physiques de l'électron. Dans le cas de la microscopie optique la résolution maximale atteignable est la moitié de la longueur d'onde du spectre visible soit,  $0.2 \mu\text{m}$  pour la longueur d'onde la plus petite qui est  $400 \text{ nm}$ . Pour les études cellulaires cette résolution est suffisante pour faire des observations, mais si l'on veut étudier des objets biologiques beaucoup plus petit comme les protéines, qui ont une taille proche de  $1\text{-}30 \text{ nm}$ , il faut utiliser d'autres méthodes. Les électrons peuvent atteindre des longueurs d'ondes beaucoup plus petites : accéléré à  $300 \text{ kV}$  un électron possède une longueur d'onde de  $1,97 \text{ pm}$ , ce qui permet d'atteindre des résolutions beaucoup plus hautes. Cependant bien qu'en théorie la résolution atteignable soit très élevée, les microscopes électroniques possèdent également leurs limites. Tout d'abord les lentilles électromagnétiques utilisées possèdent des aberrations qui limitent grandement la résolution et déforment l'image. En second lieu les électrons interagissent fortement avec la matière ce qui induit des dommages à l'échantillon dus aux radiations, mais également impose que le microscope soit sous vide car dans le cas contraire les électrons interagiraient avec l'air ambiant. De plus dans certains cas, notamment l'étude

d'échantillons biologiques, une étape de préparation des spécimens est nécessaire afin de les préparer pour résister à l'environnement sous vide et aux dommages qu'ils subiront lors de l'exposition au faisceau d'électron, mais nous y reviendrons dans le chapitre sur la préparation des échantillons.

### **a) Vue générale d'un microscope électronique à transmission**

A l'instar des microscopes optiques, qui ont besoins d'une source de photon, les microscopes électroniques à transmission maintenant abrégés en MET ont besoins d'une source d'émission d'électron qui est appelée le canon. Le faisceau émis par le canon est par la suite rendu parallèle par des lentilles appelées condenseurs. Une fois le rayon parallèle obtenu il traverse l'échantillon, puis l'objectif agrandit l'image de l'échantillon qui se forme sur un détecteur. Le principe reste très semblable à celui du microscope optique, les grandes différences étant que l'on utilise l'électron au lieu des photons, ce qui demande que le faisceau soit dans le vide et que les lentilles utilisées soient des lentilles électromagnétiques. Nous allons maintenant décrire plus en détails les différents composant du MET

### **b) Source d'électron**

Pour permettre le fonctionnement d'un MET il faut lui fournir dans un premier temps une source d'électron communément appelée le canon. Il existe deux grands types de canon, les thermoïoniques et les canons à émission par effet de champ (Field Emission Gun).

#### *(1) Les canons thermoïoniques*

Le principe de ces canons est de chauffer un filament de tungstène ayant une forme de V par un courant électrique pour que les électrons atteignent l'énergie nécessaire à leur travail de sortie ce qui leur permet de s'échapper de la surface du métal. En même temps un potentiel entre 20 kV et 300 kv est appliqué par une anode, ce qui permet d'accélérer les électrons dans la direction de l'échantillon à observer.

Les canons thermoïoniques ont comme désavantage que le faisceau produit n'est pas très cohérent, cela est dû au fait que les électrons émis ne possèdent pas tous la même énergie.

## (2) Les canons à émission de champs

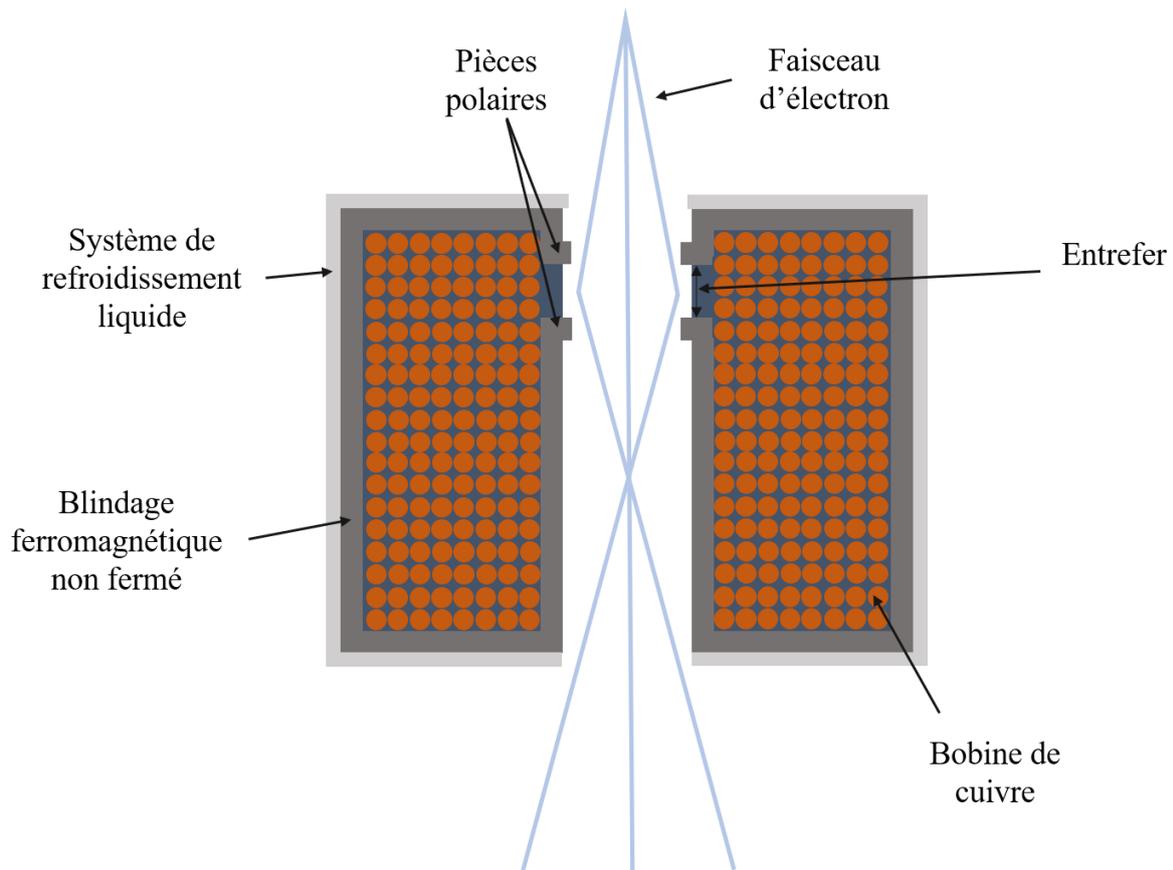
Dans ce type de canon on n'utilise pas de filament, mais une pointe de tungstène dont le diamètre est de 100 nm. La pointe n'est pas chauffée comme dans le cas des filaments des canons thermoïoniques, mais soumise à un fort champ magnétique de  $\approx 10^7$  V/cm et  $\approx 4$  kV ce qui permet d'extraire les électrons par effet de champs

Contrairement aux sources thermoïoniques les FEG permettent de produire un faisceau d'électron d'un diamètre plus faible, de cohérence meilleure et dont la variation énergétique est faible. Cependant ce genre de canon a besoin d'un vide plus poussé de l'ordre de  $10^{-10}$  mbar, afin d'empêcher l'oxydation de la pointe qui conduirait à une perte de l'émission de champs.

### c) **Les lentilles**

Les électrons sont des particules chargées négativement et si nous leur appliquons un champ magnétique, il est possible d'altérer leurs trajectoires. C'est en 1926 que Hans Busch développa pour la première fois l'idée d'utiliser des champs magnétiques radiaux pour influencer sur la trajectoire des électrons.

Les lentilles magnétiques utilisent une bobine torique parcourue par un courant électrique pour engendrer un champ magnétique axial permettant de focaliser les électrons, un peu comme les lentilles en verre le feraient avec les photons. La bobine est entourée d'un système permettant le refroidissement par eau et d'un blindage ferromagnétique qui permet de canaliser les lignes de champs au voisinage d'une ouverture grâce à des excroissances appelés pièces polaires **Figure 20**. Lors de son entrée dans la lentille un électron divergeant va subir une force qui va le ramener vers l'axe optique tout (**Figure 20**) en effectuant un mouvement de spirale.



**Figure 20 : Schéma d'une lentille magnétique**

Schéma d'une coupe de lentille de microscope électronique. La lentille est composée d'une bobine torique constituée de cuivre entouré par un blindage ferromagnétique possédant une ouverture, l'entrefer, bordé d'excroissance appelé pièce polaire, qui permet de condenser les lignes de champs produit par le champ magnétique de la bobine. Le champ magnétique produit est plus faible à mesure que l'on s'approche de l'axe optique. La lentille est entourée d'un système de refroidissement par eau liquide afin de limiter sa chauffe.

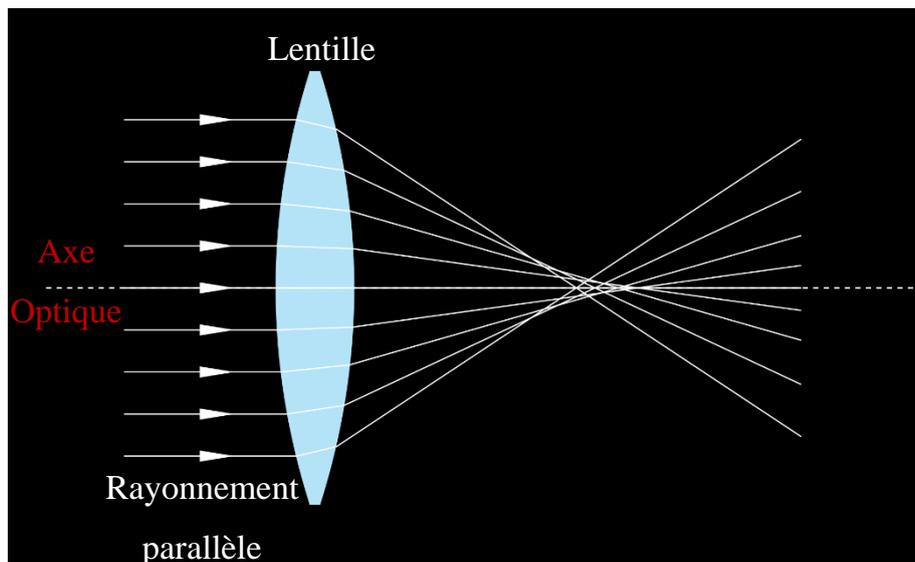
#### **d) Les aberrations des lentilles magnétiques**

Bien que l'on compare souvent les lentilles magnétiques et optiques, car toute deux permettent (dans le cas de la lentille convergente) de focaliser les électrons/photons de manière semblable, il existe une grande différence entre ces deux types de lentille : ce sont leurs qualités. Tous deux possèdent des aberrations qui seront décrites dans les sections suivantes, mais les lentilles magnétiques sont de loin les plus imparfaites. Une analogie pour montrer à quel point leurs imperfections sont importantes est que les utiliser reviendrait à regarder un objet à travers le fond d'une

bouteille de soda, ou encore que si notre œil possédait les mêmes imperfections, nous pourrions être qualifié d'aveugle.

### (1) Aberration sphérique

L'aberration sphérique est due au fait que les électrons ne sont pas tous focalisés au foyer, mais de manières différentes suivant leurs entrées par rapport à l'axe optique (**Figure 21**). Dans le cas d'une lentille parfaite les électrons éloignés et proches de l'axe optique passent tous par le foyer de la lentille, or malheureusement il est à l'heure actuelle impossible de créer des lentilles magnétiques parfaites. Il en résulte que les rayons qui rentrent de manière parallèle dans la lentille ne convergent pas tous vers le même foyer, cependant il existe maintenant des correcteurs d'aberration sphérique qui permettent de la limiter.

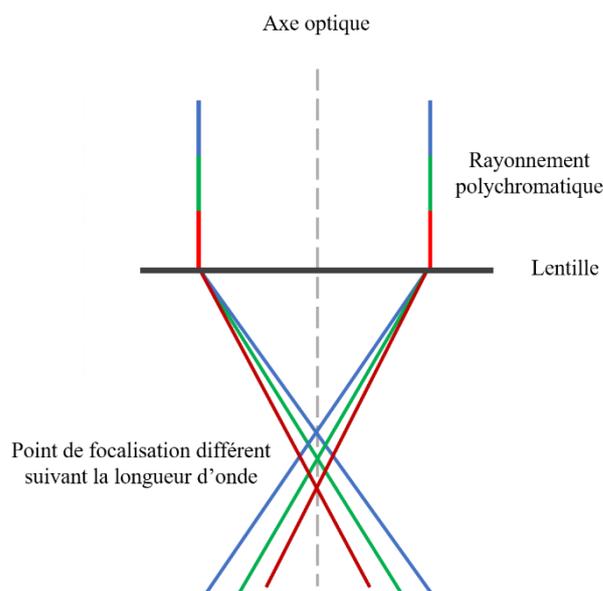


**Figure 21 : Aberration sphérique**

Schéma montrant comment l'aberration sphérique d'une lentille influence la focalisation du rayonnement qui la traverse. Un rayonnement parallèle et focalisé par la lentille, on voit que les rayons éloignés de l'axe optique sont plus fortement focalisés que ceux situés près de l'axe optique.

## (2) Aberration chromatique

Le nom de cette aberration est reliée à la couleur du rayonnement ou plus précisément aux longueurs d'ondes qui le composent. En effet, l'altération de la trajectoire d'une onde par une lentille est fonction de sa longueur d'onde. Si un rayonnement possédant plusieurs longueurs d'onde passe à travers une lentille, chaque rayonnement ne sera pas focalisé au même point.



**Figure 22 : Aberration chromatique**

Les aberrations chromatiques sont causées par le fait que la focalisation d'une onde par une lentille dépend également de la longueur d'onde. Dans le cas où un rayonnement est polychromatique les longueurs qui le composent auront chacune un point de focalisation différent.

Dans la théorie on utilise souvent le cas où le faisceau d'électron est monochromatique, ce qui n'est pas vraiment le cas, mais si on utilise un FEG, sa variation d'énergie et de  $\sim 0.3$  eV et donc la variation de la longueur est tellement faible que l'on pourrait ignorer cette aberration. Le fait d'introduire un échantillon dans le système optique va cependant créer une grande variation d'énergie dans les électrons qui le traverseront, car certains électrons interagissent avec l'échantillon et perdent de l'énergie : plus l'échantillon est épais plus ce phénomène peut prendre de l'importance. Une première méthode pour limiter ce problème est donc d'utiliser des échantillons très fins. Une

méthode autre consiste à utiliser un filtre à énergie permettant de filtrer les électrons qui ont perdu de l'énergie pour limiter l'altération de l'image.

### (3) *Astigmatisme*

L'astigmatisme est dû au fait que le champ magnétique des lentilles n'est pas uniforme. La raison est qu'il est impossible d'usiner parfaitement les composants de la lentille : il y aura toujours une légère asymétrie ou encore de légères variations dans la structure des matériaux employés. Même dans le cas où ces difficultés pourraient être éliminées, le fait d'introduire un diaphragme qui n'est pas parfaitement centré autour de l'axe optique peut perturber le champ magnétique. De plus, s'il n'est pas parfaitement propre, la contamination qui s'est déposée dessus, peut se charger et également perturber le champ produit par les lentilles. Il existe donc un grand nombre d'éléments qui peuvent induire de l'astigmatisme dans un microscope. Fort heureusement cette aberration n'est pas difficile à corriger. On peut utiliser des stigmatiseurs qui créent de faibles champs magnétiques en comparaison des lentilles qui permettent de corriger l'astigmatisme.

#### **e) Visualisation des images formées par le microscope**

En règle générale les écrans/cameras qui permettent l'affichage/l'enregistrement sont situés au bas de la colonne du MET. La méthode la plus basique pour observer une image est l'utilisation d'écrans fluorescents, qui sont recouvert de ZnS, une substance qui peut être excitée par les électrons lorsque ceux-ci viennent frapper l'écran, ce qui produit un rayonnement vert. Ce genre d'appareil est classique dans les MET, bien qu'il ne permet pas d'enregistrer des données il est très utile pour les opérations de base du microscope. Pour être capable d'enregistrer des images il existe plusieurs méthodes qui vont être décrites plus en détail dans les prochaines sections : la première est l'utilisation de films, la seconde de caméra dite CCD pour Charge Coupled Device et pour finir les caméras DDC pour Direct Detection Camera, qui sont à l'heure actuelle les meilleurs instruments pour enregistrer des images à haute résolution.

### (1) Les films photographiques

Les films sont les premiers moyens utilisés pour pouvoir enregistrer des images. Ils ont une taille de 9 par 6 cm et sont composés d'halogénure d'argent pouvant interagir avec les électrons. Une fois les films exposés ils doivent être développés puis scannés afin d'obtenir une image numérique pouvant être utilisée. Le problème de cette méthode est le temps qu'elle demande pour être mise en place, sa complexité et les avancées en matière de cameras DDC. Toutes ces restrictions ont fait que cette technologie est devenue complètement obsolète de nos jours.

### (2) Les cameras CCD

Le premier appareil permettant d'enregistrer directement des images depuis le microscope sont les caméras CCD, cependant si on les exposait directement aux électrons elles seraient endommagées par ces derniers. Pour pallier à ce problème les cameras CCD sont reliées par une fibre optique à un écran scintillateur, ceci permet de convertir l'électron qui vient frapper le scintillateur en photon, qui est transporté vers le détecteur, qui convertit le signal reçu en image.

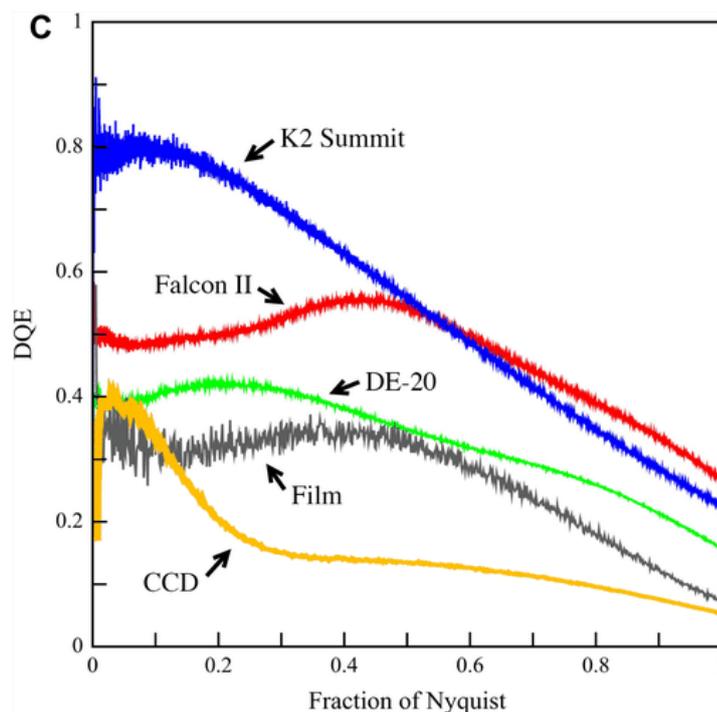
Bien que ces caméras représentent une grande facilité d'utilisation par rapport au film et permettent de collecter des données de façon automatisée elles possèdent de gros défauts. Elles sont sujettes à la rétrodiffusion des électrons (backscattering en anglais) et la nature des électrons fait que la caméra peut enregistrer un mauvais nombre d'événements, ce qui ajoute encore plus de bruit dans l'image enregistrée.

### (3) Les cameras à détection directe des électrons

Le développement des cameras à détection directe des électrons a connu un essor spectaculaire ces dernières années et est responsable en partie de la popularisation de la cryo microscopie. Les DDC (Direct Detection Camera) sont capables comme leur nom l'indique de détecter de manière directe un électron arrivant sur le capteur et sont conçus pour réduire les effets de rétrodiffusion, ce qui réduit grandement le bruit. La rapidité de lecture de ces nouveaux capteurs permet également d'acquérir un grand nombre

d'images durant une exposition courte, ce qui permet de réduire les effets de mouvements induit par le faisceau d'électrons sur l'échantillon (Correction du mouvement induit par le faisceau d'électrons).

Un autre paramètre à prendre en compte est que ce type de caméra possède une DQE (Detective Quantum efficiency ou en français efficacité quantique de détection) beaucoup plus élevée que les films photographiques qui eux-mêmes ont un DQE plus élevé que les camera CCD **Figure 23**. Cette mesure permet d'exprimer le ratio entre le nombre d'électrons détectés sur le nombre d'électrons émis et permet de mesurer à quel point le détecteur ajoute du bruit dans l'image : plus la DQE est élevée, c'est-à-dire proche de 1, qui est le cas d'un détecteur parfait, plus le détecteur est performant. La valeur de DQE pour un détecteur change également en fonction de la fréquence spatiale voir **Figure 23**, où dans les hautes fréquences spatiales là où les détails sont les plus fins, la DQE a tendance à décroître fortement.

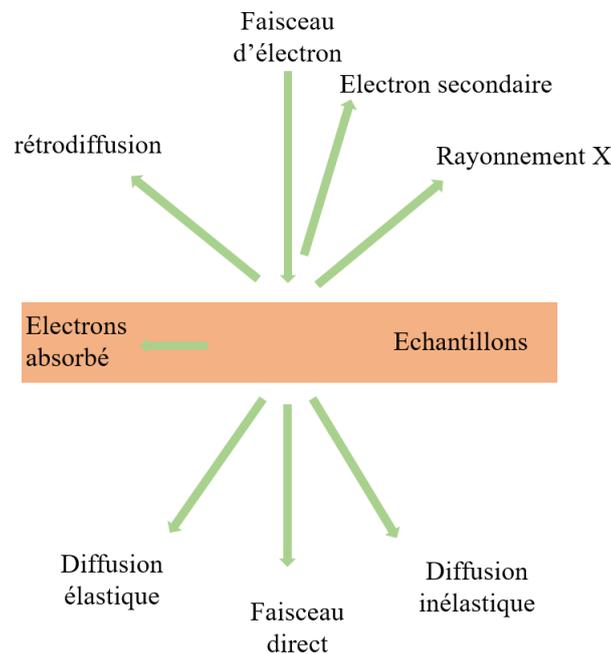


**Figure 23 DQE à 300 kV en fonction de la fréquence spatiale**

L'image montre l'évolution de la DQE en fonction de la fréquence spatiale pour différentes caméras et film photographique. On peut voir que les caméras DDC ont une meilleure DQE que les Films et que la DQE a tendance à chuter avec les hautes fréquences spatiales.

## f) Interaction des électrons avec l'échantillon

En temps normal dans un microscope optique le contraste de l'image est dû à la



**Figure 24 : Signaux courants générés par l'interaction d'un faisceau d'électrons avec et un échantillon fin**

La plupart des électrons ne sont pas diffusés et leurs trajectoires restent donc inchangées. Les électrons diffusés le seront de deux manières différentes, inélastique ou élastique. C'est le second type de diffusion qui nous intéresse dans le cas d'études d'échantillons biologiques par MET.

coloration que l'on donne à l'échantillon. Ainsi différents éléments de la cellule ne seront pas colorés de la même manière, ce qui donne lieu à des absorptions de photon différentes créant un contraste. Dans le cas des microscopes électroniques si les échantillons sont constitués d'atomes légers, comme c'est le cas pour les échantillons biologiques composés en grande partie d'azote, de carbone et d'oxygène, l'absorption est faible ce qui conduit à un faible contraste d'absorption. Lorsqu'un faisceau d'électron traverse un échantillon mince il peut se produire deux types de diffusion :

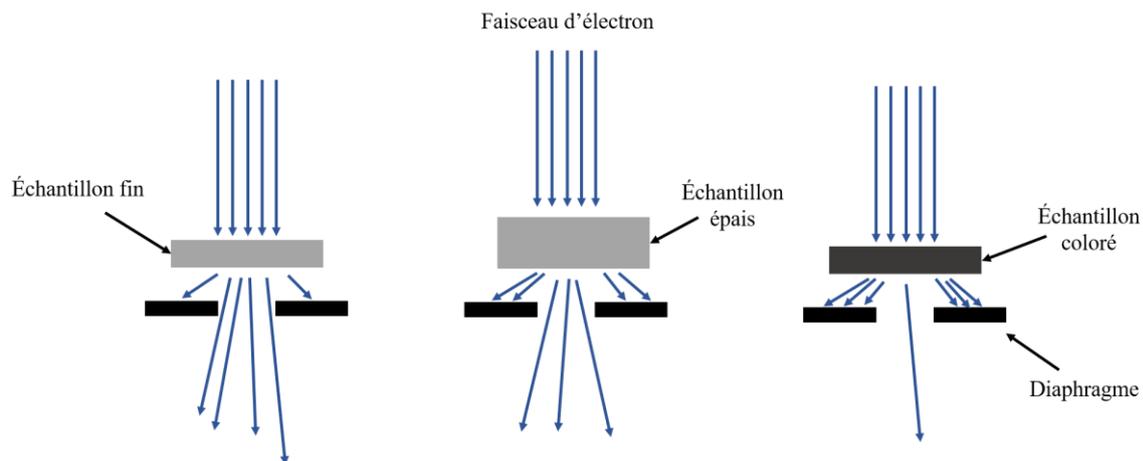
Une diffusion élastique, qui correspond à un électron qui interagit avec le noyau des atomes composant l'échantillon, ce qui conduit à une déviation de l'électron. Cette déviation est d'autant plus grande que le numéro atomique de l'atome est élevé. Dans

ce cas-là bien que dévié l'électron ne perd pas d'énergie. C'est ce type de diffusion qui est responsable du contraste dans les MET

Une diffusion inélastique, qui est le cas où l'électron effectue des interactions électrostatiques avec les électrons de l'échantillon. Elle conduit à une perte d'énergie et une déviation plus faible de l'électron. Ceci produit des électrons secondaires et un rayonnement X qui vont ioniser et briser des liaisons chimiques, dégrader l'échantillon et l'information à haute résolution. Un autre cas est que le faisceau va simplement traverser l'échantillon sans être altéré.

### (1) Contraste de Phase et d'amplitude

Le contraste dans un MET est composé de deux types : le contraste d'amplitude et le contraste de phase, tous deux produits par la diffusion élastique des électrons.



**Figure 25 : Contraste d'amplitude**

Le contraste d'amplitude est généré par les interactions élastiques des électrons du faisceau avec l'échantillon. L'introduction d'un diaphragme permet de sélectionner une partie des électrons diffusés, ce qui augmente le contraste, mais diminue la résolution. Plus l'échantillon est épais plus le contraste d'amplitude est fort. De même l'ajout d'un sel d'atome lourd pour colorer l'échantillon permet également son augmentation.

Le contraste d'amplitude dépend de la masse de l'échantillon, mais également de son épaisseur **Figure 25**, plus l'échantillon est épais plus il interagit avec les électrons. Il est produit par la diffusion élastique des électrons et peut être augmenté en

ajoutant un diaphragme de contraste afin de sélectionner une partie des électrons diffusés **Figure 25**. Dans le cas de la coloration négative, une méthode de préparation des échantillons utilisant des sels d'atomes lourds, le contraste d'amplitude peut représenter près de 30 % du contraste total de l'image (Erickson & Lug, 1971). Cependant dans le cas de la cryo-microscopie, du fait que les échantillons biologiques sont majoritairement composés d'éléments légers comme l'azote, le carbone ou l'oxygène et que l'on utilise aucun colorant, le contraste d'amplitude contribue très peu au contraste total de l'image.

Le contraste de phase résulte de l'interférence des phases des électrons diffusés (diffusion élastique) et des électrons transmis (électron qui n'ont pas interagi avec l'échantillon). Pour comprendre ces effets il faut considérer les électrons comme des ondes qui possèdent une amplitude et une phase. Quand l'onde arrive sur l'échantillon celui-ci va décaler sa phase à cause de son interaction avec l'échantillon. Les interférences produites entre les ondes diffusées et transmises vont ainsi produire un contraste de phase. Cependant dans le plan focal il n'y a pas de contraste de phase, il est donc nécessaire de convertir ces changements de phases en changement d'amplitudes afin de pouvoir les observer. Pour ce faire on défocalise l'image, ce qui change le plan focal. Dans le cas d'électrons transmis non diffusés ceci ne change rien à leurs phases, cependant dans le cas des électrons diffusés la variation des phases varie beaucoup en fonction des valeurs de focus, les ondes diffusées vont donc être en phases ou non avec les ondes transmises, ce qui permet de faire apparaître une différence d'amplitude. En MET ce contraste contribue bien plus au contraste de l'image que le contraste d'amplitude.

## 2. La préparation des échantillons pour la microscopie électronique

Au cours des dernières décennies l'utilisation de microscope électronique pour visualiser, caractériser et déterminer la structure des protéines est devenue très importante dans le domaine de la biologie structurale. Cependant l'étude d'échantillons biologiques dans le vide d'un MET n'est pas chose aisée, il convient de préparer les échantillons pour qu'ils puissent être visualisés dans le vide d'un microscope tout en

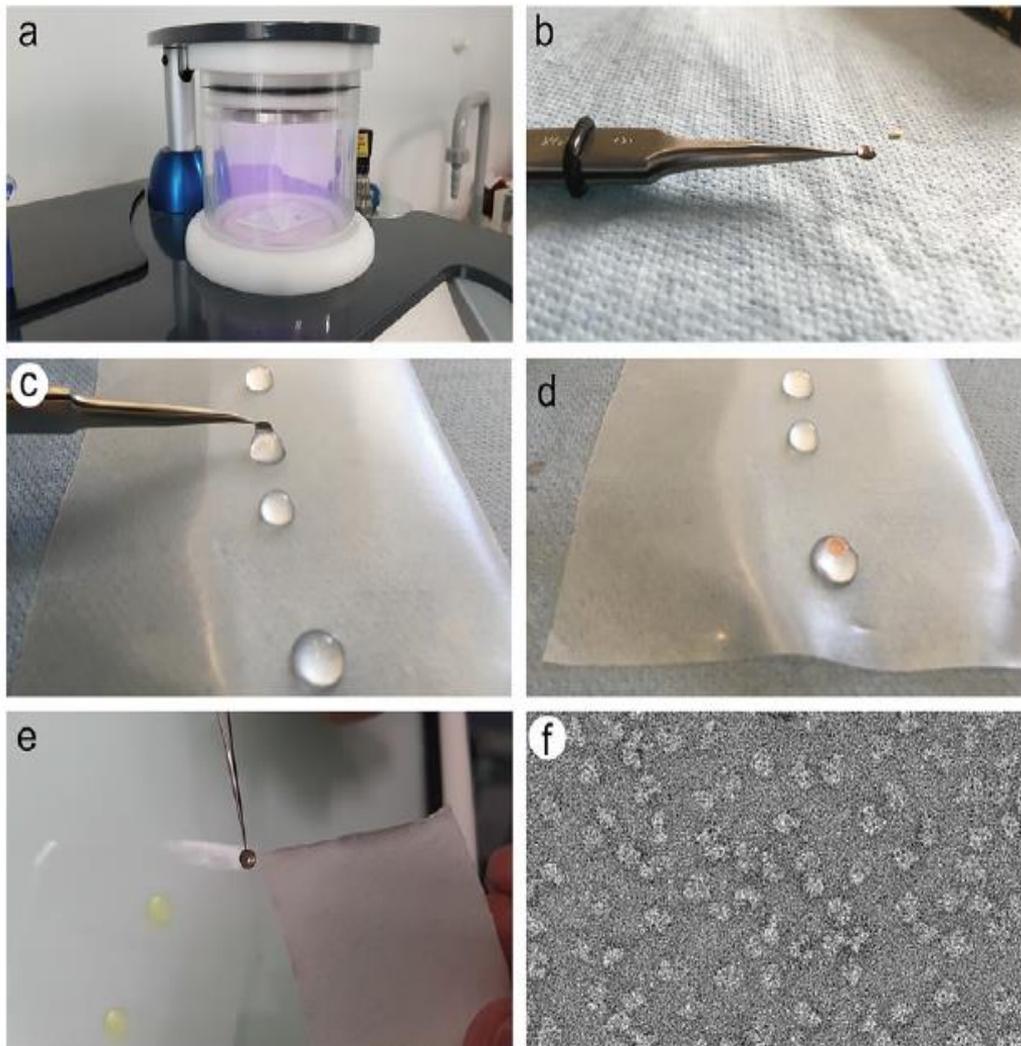
préservant les informations sur la structure de la protéine. Dans cette partie nous allons voir deux grandes méthodes de préparation des échantillons qui permettent d'étudier les protéines dans un microscope électronique, la cryo-EM et la coloration négative.

Dans la plupart des cas l'échantillon est appliqué sur une grille de microscopie électronique d'une taille de 3 mm ce qui correspond à peu de choses près à la tête d'une allumette. Il existe plusieurs types de grilles et leurs compositions peuvent varier, passant du cuivre à l'or ou à des alliages tel que le cuivre/rhodium. Dans le cas de la coloration négative un film de carbone continu est appliqué sur la grille afin de pouvoir agir comme un support pour que les protéines puissent s'absorber dessus, alors que dans le cas de la cryo-EM le carbone est perforé. En temps normal la surface du film de carbone est hydrophobe, pour permettre une meilleure dispersion de l'échantillon et de son absorption, il est nécessaire de rendre cette surface hydrophile. L'étape qui permet de rendre la surface du carbone hydrophile est appelée éfluvage ou glow-discharge en anglais. Elle consiste à exposer la grille à un plasma ionisant obtenu en appliquant un courant à un gaz à faible pression, le type de gaz utilisé est responsable du type de charge obtenue sur la surface du carbone. Cette étape d'éfluvage est commune aux deux méthodes de préparation d'échantillons que nous allons voir dans les deux prochaines parties.

### **a) Coloration négative**

La première méthode simple de coloration négative a été décrite par Brenner & Horne en 1959 (Brenner & Horne, 1959). Contrairement à la cryo-EM la coloration négative ne permet pas d'obtenir des structures à haute résolution, cependant associée à l'analyse de particules isolées, elle est capable de fournir des informations préliminaires, comme la qualité de l'échantillon, son homogénéité, sa tendance à agréger et peut donc être utilisée pour optimiser un protocole de purification en vue d'obtenir une structure à haute résolution en Cryo-EM. Il est important de noter que dans certains cas toutes les informations obtenues ne sont pas toujours transposables à la Cryo-EM, mais dans l'ensemble la coloration négative constitue une base solide pour commencer un projet.

Le principe de la coloration négative est d'incorporer la protéine dans une matrice d'atomes de sels lourds ce qui permet à la fois de prévenir l'effondrement de la protéine lors de son assèchement dans le vide du microscope, mais également d'augmenter le contraste d'amplitude de l'échantillon. Les colorants les plus couramment utilisés sont l'acétate d'uranyle, formiate d'uranyle et l'acide phosphotungstique, leurs masses



**Figure 26 : Les différentes étapes de la coloration négative**

(a) La grille est effluée par un plasma d'air ionisé pour rendre le support hydrophile. (b) 3  $\mu$ L de protéines purifiées et solubilisées dans un tampon sont adsorbées sur le support de la grille pendant 1 min. (c) L'excès d'échantillon et les composants nuisibles du tampon sont dilués par plusieurs passages sur les gouttes de tampon ou d'eau. (d) Coloration de l'échantillon : la grille est déposée sur le dessus d'une goutte de colorant d'acétate d'uranyle à 2% pendant 30 s. (e) La solution de coloration est éliminée par tamponnage à l'aide de papier filtre. (f) Micrographie électronique d'un échantillon bien coloré, les particules sont bien dispersées, montrent peu d'agrégation et une coloration optimale.

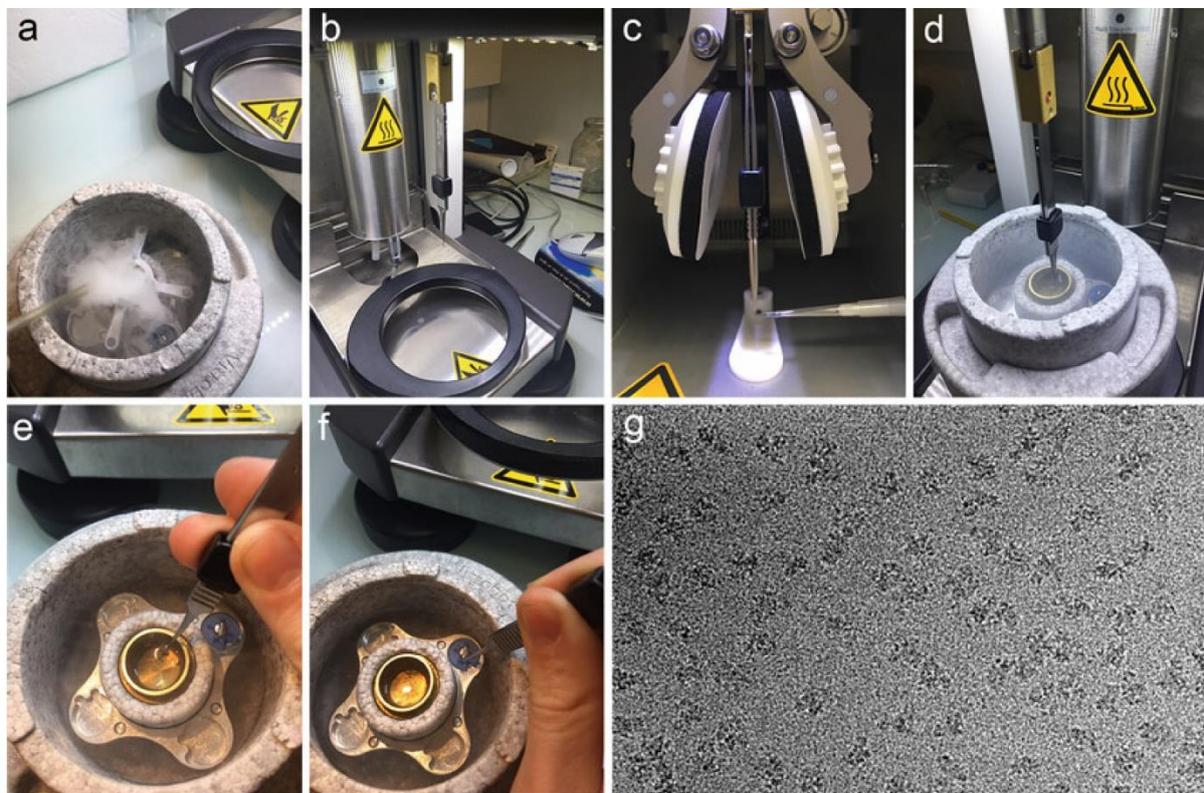
importantes permettent une diffusion plus importante des électrons. Le désavantage de cette technique est qu'elle ne permet d'étudier que la surface/forme des protéines et la résolution maximal atteignable après une analyse des images n'est que de 17-20 Å. Mais l'avantage de la coloration négative est qu'elle est facile à mettre en place et peut donner beaucoup d'informations pour la suite de l'étude structurale et le fort contraste permet d'interpréter plus simplement les images obtenues au MET.

## **b) La Cryo microscopie électronique**

L'un des majeurs désavantages de la coloration négative est qu'il est nécessaire de sécher l'échantillon, or les protéines ont besoin d'être hydratées pour être dans un état natif. Pour surmonter ce problème il est nécessaire de garder l'échantillon hydraté tout au long de sa préparation et également lors de son introduction ainsi qu'à l'acquisition des images dans un MET. Congeler les protéines dans de la glace serait un moyen efficace de les garder hydratées, cependant dans des conditions de pression standard la glace forme une structure cristalline, ce qui gêne énormément la visualisation et rend impossible l'étude des échantillons. Pour pallier ce problème on cherche à congeler les protéines dans une glace vitreuse, c'est-à-dire dépourvue de structure cristalline. L'avantage d'une telle glace est qu'elle est transparente aux électrons et permet donc l'observation au MET. De plus les protéines restent hydratées, ce qui permet leurs études dans un environnement plus proche de celui de la cellule et la température cryogénique permet de réduire les dommages causés par l'exposition au faisceau d'électrons. L'obtention de glace amorphe à pression atmosphérique fût pendant de nombreuses années jugée impossible, mais les travaux de Jacques Dubochet ont permis de mettre en œuvre cette technique et d'ouvrir la voie à la Cryo-EM( J Dubochet et al 1988).

Pour obtenir une glace amorphe il est nécessaire de refroidir l'échantillon très rapidement et de le garder à une température inférieure à -170 °C. Pour ce faire on utilise un plongeur qui retient une grille de microscopie électronique préalablement effluvé où l'on vient déposer quelques µl d'échantillon, l'excédent d'eau est retiré par absorption grâce à du papier filtre ce qui forme un fin film d'eau sur la grille. La grille est ensuite

plongée dans de l'éthane liquide refroidi par azote liquide, cette étape permet de congeler l'eau de manière instantanée et d'obtenir une glace amorphe. L'utilisation d'éthane liquide est cruciale car dans le cas où l'on plonge directement une grille dans l'azote liquide, le delta de température est tellement élevé qu'une gaine d'azote gazeux se forme autour de la grille, ce qui l'empêche de se refroidir rapidement (effet Leidenfrost). L'image **Figure 27** montre les différentes étapes de la cryo-EM



**Figure 27 : Vitrification des échantillons**

(a) support en polystyrène rempli d'azote liquide, on peut voir au centre un réservoir en métal contenant l'éthane liquide. (b) La grille de microscopie électronique est attrapée par une pince reliée au plongeur. (c) 3  $\mu$ L d'échantillon est déposé sur la grille qui sera soulevée jusqu'aux papiers filtres ou ils seront automatiquement pressés contre la grille pour enlever l'excès de liquide. (d) Le plongeur immerge rapidement la grille dans l'éthane liquide. (e,f) L'échantillon vitrifié est transféré dans une boîte cryogénique pour y être stocké. (g) Micrographie électronique d'un échantillon hydraté congelé montrant des particules dispersées, une faible agrégation, une glace fine et aucune contamination.

D'après (Frechard et al., 2021)

### 3. L'analyse des images en particule isolée

Lors de l'étude d'un échantillon de protéines par microscopie électronique, les images possèdent un très faible ratio signal sur bruit. Pour répondre à ce problème, comme les échantillons étudiés sont ceux de protéines purifiées, c'est-à-dire d'un même objet présent en de centaines de milliers de copies. On peut faire la moyenne des images de ces objets ce qui va éliminer le bruit et donc augmenter grandement le ratio signal sur bruit et permet de restaurer les informations à haute résolution. Il faut cependant prendre en compte que la protéine, qui est un objet 3D, peut adopter des orientations différentes dans l'espace lors de la préparation des échantillons, que ce soit en coloration négative ou en cryo-EM. Par conséquent l'image d'une protéine obtenue par MET qui est une projection 2D d'un objet 3D et possède une orientation. Chaque projection 2D de protéines contenues dans une micrographie sont appelées particules et ont leurs propres orientations. Pour être capable de rétroprojeter les particules afin d'obtenir une structure il est nécessaire d'en connaître l'orientation. La méthode qui combine les différentes étapes qui permettent l'obtention d'une structure à partir de protéines purifiées s'appelle l'analyse de particules isolées ou single particles analysis (SPA) en anglais et sera décrite dans les sections suivantes. L'explication de cette technique se fera dans un ordre « chronologique », des étapes de pré analyse qui se font en sortie de microscope aux étapes finales d'affinement des structures.

#### **a) Correction du mouvement induit par le faisceau d'électrons**

Quand l'échantillon est illuminé par le faisceau d'électrons cela produit un mouvement qui va induire un flou dans l'image, limitant grandement la résolution atteignable avec de telles données. Ce mouvement peut être décomposé en deux composantes, la première est un mouvement uniforme de l'ensemble des particules de l'image et la deuxième est un mouvement non uniforme variant non pas sur l'ensemble de l'image mais de manière locale. Le mouvement induit par le faisceau d'électrons est également plus intense au début de l'exposition et diminue rapidement par la suite.

L'avènement des caméras à détection directe des électrons a cependant permis de pallier à ce problème. Ces caméras sont capables d'enregistrer 10 à 40 images par seconde, rendant possible la collecte d'un « film ». L'avantage de collecter des données sous cette forme est que chaque image du film est prise dans un intervalle de temps très faible ce qui limite le mouvement d'une image à une autre.

Cependant chaque image du film est faiblement exposée, ce qui induit un faible ratio signal sur bruit et rend difficile les étapes d'affinements d'une structure 3D, en particulier pour des objets de petites tailles. L'idée de la correction de mouvement est de corriger le mouvement entre toutes les images acquises dans un film pour obtenir à la fin une seule image nette obtenue par l'alignement de chaque image du film.

Dans un souci de clarté dorénavant nous appellerons les images collectées sous forme de film « film », chaque image composant le film sera appelée une trame et l'alignement des différentes trames du film permet d'obtenir une image dont le mouvement des particules induit par le faisceau est corrigé et qui est appelé micrographie.

### (1) Motion Corr

Le programme Motion Corr (Li et al., 2013) a pour but de corriger le mouvement uniforme d'un film, ce mouvement se traduit par un mouvement des particules dans une même direction. Le programme utilise des mesures redondantes entre le décalage de chaque trame qui compose le film pour obtenir une estimation du mouvement relatif entre chaque trame par la méthode des moindres carrés, afin de corriger le mouvement uniforme des différentes trames du film pour obtenir une micrographie nette.

Dans la pratique toutes les trames ne sont pas corrigées. En effet les premières images du film sont retirées car le mouvement induit par le faisceau est plus fort dans les premiers instants de l'acquisition et donc bien que les premières trames acquises sont en théorie les moins endommagées, car la dose d'électrons reçue sur l'échantillon n'est pas encore élevée, le fort mouvement qu'elles subissent limite leurs résolutions et rend impossible leurs corrections.

## (2) Le Programme MotionCorr2

Ce Programme permet de corriger les mouvements uniformes et non uniformes induits par le faisceau d'électrons en utilisant un modèle validé expérimentalement qui décrit le mouvement de l'échantillon comme une déformation locale qui varie de façon régulière tout au long de l'exposition (Zheng et al., 2017).

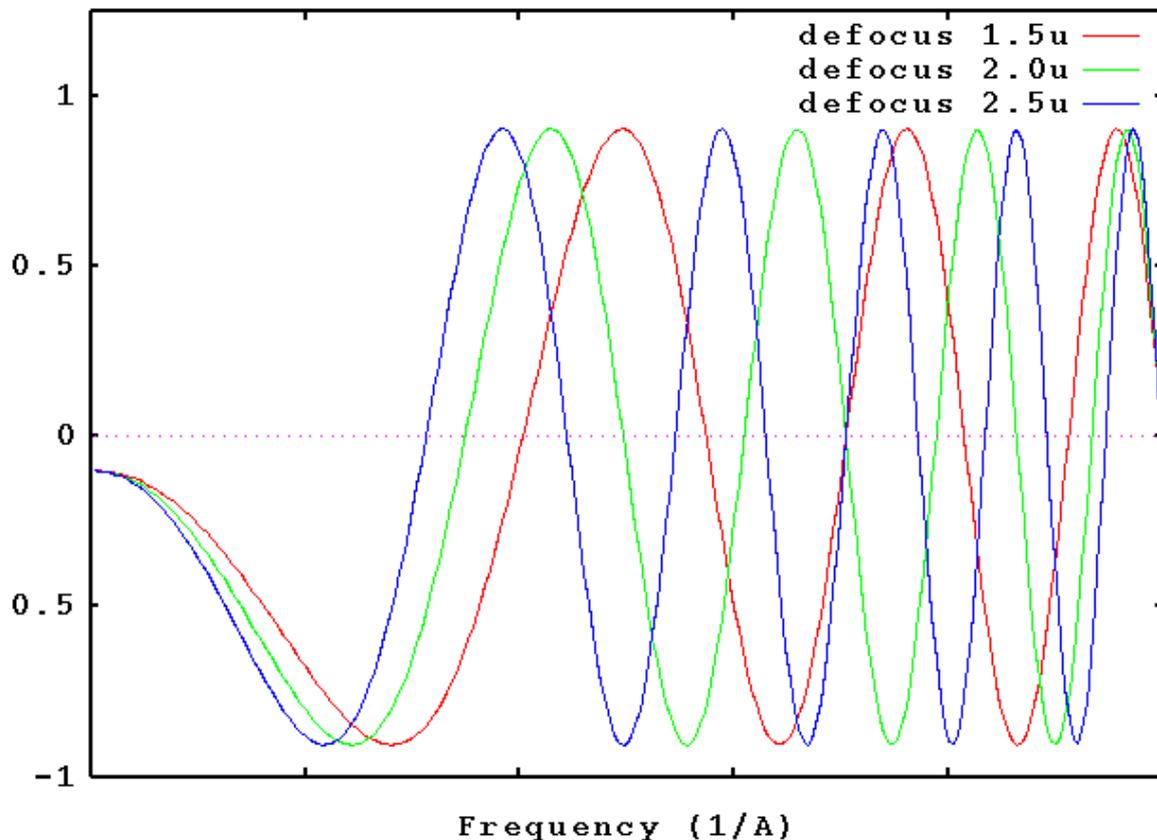
La nouvelle approche de cette deuxième version consiste non pas d'essayer de comparer les trames de chaque film entre elles pour déterminer leurs mouvements par rapport les unes aux autres, mais de comparer des sous régions des trames à une somme approximative pour chaque sous-région, puis d'itérer la détermination des mouvements de translation pour améliorer la somme.

### **b) Correction de la fonction de transfert de contraste**

La fonction de transfert de contraste d'un microscope est une fonction qui exprime la valeur d'intensité du contraste en fonction des fréquences spatiales associées. Elle dépend de l'aberration sphérique des lentilles et également de la valeur de défocalisation **Figure 28**.

Comme indiqué dans la section dédiée au contraste, du fait que les échantillons biologiques possèdent peu de contraste d'amplitude on convertit le contraste de phase en contraste d'amplitude en défocalisant l'image. Cette défocalisation va induire l'inversion du contraste pour certaines gammes de fréquences (valeur de CTF négative) qui dépend de l'importance de la défocalisation. Il est possible de corriger cette inversion du contraste en inversant la valeur pour les intensités négatives. Cependant, à chaque fois que la courbe de la CTF passe par 0 l'information est perdue. Pour récupérer

l'information on collecte à des valeurs de défocus différentes de manière à couvrir toutes les gammes de fréquences manquantes.



**Figure 28 : Fonction de transfert de contraste et dé focus**

La valeur de dé focalisation influe sur le comportement de la CTF. L'utilisation de plusieurs valeurs de dé focus permet de récupérer les informations qui sont perdues à chaque fois que la CTF passe par 0

### c) Sélection des particules

Après la correction de mouvement des micrographies, il est essentiel de sélectionner et d'extraire chaque particule de ces dernières, pour ce faire il existe plusieurs méthodes. La première est celle de comparaison des modèles ou « Template Matching » en anglais. Elle consiste à utiliser un modèle 2D ayant la forme de notre particule qui sera comparée avec chaque particule des micrographies par cross corrélation afin de sélectionner automatiquement toutes les particules suivant un seuil. Il existe plusieurs moyens d'obtenir ce genre de modèle :

- Des images fortement filtrées de nos particules, obtenues par classification 2D (voir classification 2D)
- La génération d'un blob gaussien ayant les différentes tailles de notre particule
- La génération d'un disque gaussien ayant le diamètre attendu des particules

Bien que pendant longtemps ces méthodes étaient très utilisées pour sélectionner les particules, elles comportent plusieurs problèmes : elles ont tendance à sélectionner beaucoup de contaminants comme la glace, des particules agrégées et dans certains cas si le seuil de corrélation n'est pas assez élevé elles peuvent également sélectionner du bruit, ce qui peut grandement impacter la qualité de l'analyse des images.

Depuis plusieurs années le développement des intelligences artificielles notamment dans les domaines de l'apprentissage automatique et plus précisément l'apprentissage profond ont permis de créer des programmes capables d'apprendre et de réaliser certaines tâches bien précises, notamment dans le domaine de la reconnaissance des images. Des programmes de sélection de particules s'inspirant de ces technologies ont vu le jour et sont capables de reconnaître les particules, d'éviter les agrégats et le carbone avec une précision semblable à celle d'un opérateur humain. Comme exemple de programme on peut citer crYOLO (Wagner et al., 2019), Warp (Tegunov & Cramer, 2019) ou encore Topaz (Bepler et al., 2019). Dans le cas de crYOLO le programme se base sur l'algorithme de reconnaissance d'image YOLO, un programme qui permet de reconnaître les différents éléments dans une image tel qu'une personne, un chien, un cheval etc ...

### (1) Principe des algorithmes d'apprentissage profond

Les algorithmes d'apprentissage profond se basent sur l'utilisation de neurones artificielles, qui sont des fonctions mathématiques qui imitent grossièrement le fonctionnement des neurones. Chaque neurone possède des « entrées » qui sont les informations qui lui sont données et permet de les traiter pour donner un résultat qui est la « sortie ». Dans la pratique l'utilisation d'un seul neurone n'est pas suffisante pour

traiter des informations complexes. C'est pourquoi on associe plusieurs neurones artificiels entre eux, afin de former un réseau. Une fois ce réseau formé il doit être entraîné en donnant des valeurs d'entrée et les valeurs de sortie correspondantes, par exemple la circonférence du tronc d'arbre et sa taille. Dans notre exemple la phase d'entraînement du réseau consiste donc à ce qu'il trouve la relation entre la circonférence des arbres et leur taille. Une fois le réseau entraîné il sera capable de prévoir les valeurs de sortie (taille de l'arbre) sachant les valeurs d'entrée (circonférence de l'arbre).

## (2) CrYOLO

Pour que crYOLO soit capable de reconnaître une particule il doit être entraîné en fournissant une première sélection de particules faite manuellement (~200-2500). Avec l'approche YOLO, la micrographie complète est prise comme entrée et lorsque l'image est passée à travers le réseau de neurones, elle est spatialement sous-échantillonnée en une grille. Ensuite, YOLO prédit pour chaque cellule de la grille si elle contient le centre d'une boîte contenant une particule. Si c'est le cas, il estime la position relative du centre de la particule à l'intérieur de la cellule, ainsi que la largeur et la hauteur de la boîte. L'intérêt de cette méthode est qu'il n'est nécessaire de fournir en entrée que des particules manuellement sélectionnées pour entraîner le réseau. De plus, comme le réseau voit le micrographe complet, il apprend le contexte de répartition des particules. Une fois les particules sélectionnées par l'un ou l'autre programme elles sont extraites et normalisées.

### **d) Alignement des particules**

Avant de pouvoir faire la moyenne des images contenant les particules pour récupérer les informations à hautes résolutions, il est nécessaire de les aligner entre elles par des mouvements de rotation et de translation. Il existe deux grandes méthodes pour effectuer cette tâche : la corrélation croisée maximale et le maximum de vraisemblance.

Dans le cas de la corrélation croisée une référence est utilisée et comparée avec des particules qui possèdent une certaine orientation qui est inconnue. Pour chaque

particule on calcule une fonction qui donne la corrélation entre l'image de référence et les particules à aligner suivant les rotations et translations qu'on leur applique. On cherche ensuite pour quelles orientations leur coefficient de corrélation est maximum et on utilise cet alignement pour faire la moyenne des particules qui est utilisée ensuite comme nouvelle référence pour réitérer le processus pour améliorer l'alignement de manière itérative. De cette manière après plusieurs cycles on va converger vers un meilleur alignement et une augmentation du rapport signal sur bruit. Il est important de noter que vu que les images de MET sont très bruitées et possèdent un ratio signal sur bruit très faible, l'alignement peut trouver dans le bruit des valeurs de corrélation élevées ce qui empêchera la convergence de l'algorithme vers un résultat probant.

En ce qui concerne le maximum de vraisemblance, l'idée est que à la place de donner une seule orientation optimale à une particule, on explore toutes les orientations possibles qu'elle peut avoir et chacune de ces orientations se verra associer une probabilité. Dans ce cas-là une particule ne contribue pas qu'une seule fois dans la moyenne finale, mais autant de fois que d'orientations explorées et chaque contribution est pondérée par un poids qui est la probabilité que la particule soit dans cette orientation. Cette méthode possède l'avantage d'être moins sensible aux références initiales utilisées et donc de moins biaiser l'alignement, mais en contrepartie elle est plus gourmande en temps de calcul.

Il est également intéressant de noter que dans le cas où le rapport signal sur bruit est bon, les méthodes de maximum de vraisemblance et de corrélation croisée maximale sont équivalentes. Cependant, dans le cadre de la SPA les derniers programmes développés se basent beaucoup plus sur le maximum de vraisemblance.

#### **e) Classification 2D**

Après la sélection des particules et leurs extractions une première étape de classification en deux dimensions permet d'effectuer un premier tri et de retirer les « mauvaises » particules qui sont en générale : des agrégats, de la contamination, de la glace ou du carbone. Pour ce faire les particules sont alignées entre elles par des mouvements de rotations et translations, et sont réparties dans différentes classes suivant leurs orientations. Ainsi les particules dans la même disposition et conformation seront

placées dans une même classe. La classification 2D se fait sans donner de référence, la première étape est la génération d'un nombre de référence  $X$ , dont la valeur est définie par l'expérimentateur, générées par des particules dont l'orientation est aléatoire. Les particules sont alignées contre ces références et de manière itérative : l'alignement et les références sont améliorées. Cette méthode permet d'éliminer les potentiels biais que pourrait introduire une référence fournie par l'expérimentateur. La moyenne des classes permet grandement d'améliorer le rapport signal sur bruit, ce qui permet d'apprécier la qualité ou l'hétérogénéité du jeu de données. En règle générale plusieurs cycles de classification 2D sont faites les uns à la suite des autres pour retirer toutes les mauvaises particules.

#### **f) Détermination d'un premier modèle 3D**

Les prochaines étapes de l'analyse d'image que sont la classification 3D et l'affinement nécessitent l'obtention d'un modèle 3D qui servira de référence lors de l'alignement des images pour trouver leurs orientations. Si des informations structurelles comme un modèle atomique sont déjà connues, alors il est facile d'en dériver une référence qui servira de modèle initial. Le cas le plus compliqué est quand aucune structure de la protéine n'est connue, il faut alors trouver un moyen de générer un premier modèle 3D.

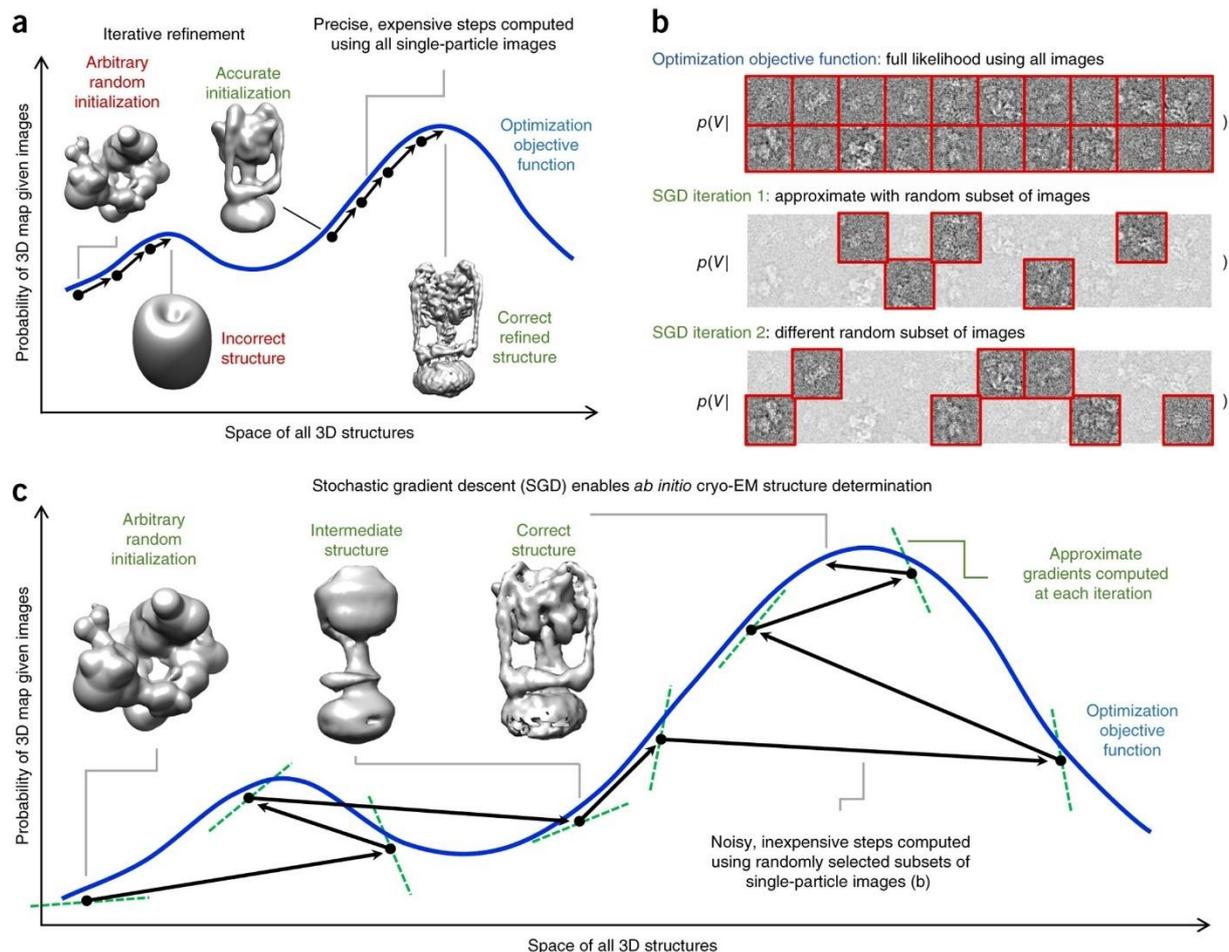
Pour ce faire il existe plusieurs méthodes algorithmiques comme les programmes se basant sur les lignes communes ou des méthodes expérimentales comme le random conical tilt (RCT). Ici nous allons décrire une méthode qui a été développée ces dernières années et est basée sur la descente de gradient. Elle permet d'obtenir une structure 3D sans référence initiale en utilisant simplement comme entrée les particules sélectionnées.

##### *(1) Méthode par descente de gradient*

La détermination d'un modèle 3D est un problème non convexe, c'est-à-dire que la fonction à optimiser possède de nombreux minimum/maximum locaux et les algorithmes classiques parcourant la fonction à la recherche d'un minimum ou maximum ont de fortes chances d'être bloqués dans l'un de ces optimums locaux si la recherche n'est pas correctement initialisée **Figure 29**. Les algorithmes utilisés en Cryo-EM n'échappent pas à ce problème : si une référence est trop éloignée de la structure réelle de la protéine, les programmes risquent d'être bloqués dans un optimum local conduisant à la détermination d'une structure erronée. Fort heureusement de nouveaux algorithmes comme la méthode de descente de gradient ont été développés et sont particulièrement efficaces pour parcourir des fonctions non convexes sans pour autant être bloqués dans un optimum local (Punjani et al., 2017).

La détermination de structure par cryo-EM est un problème d'optimisation qui peut être décrit dans le cadre de probabilité bayésien, le but de cette optimisation est de trouver un volume 3D qui explique le mieux les images observées. La méthode par descente de gradient consiste en l'optimisation d'une fonction objectif qui décrit la probabilité qu'un volume explique les particules observées, en sélectionnant à chaque fois un petit sous-groupe de particules prises aléatoirement. À chaque itération on cherche à optimiser un volume avec le sous-groupe de particules sélectionnées. À la fin d'une itération le volume est mis à jour, on peut passer à la prochaine itération en utilisant des nouvelles particules **Figure 29**. Du fait qu'à chaque itération on n'analyse qu'un petit groupe de particules le coût en calcul de cette méthode est peu élevé.

Cette méthode permet de déterminer des modèle 3D ab initio mais également de faire des classifications 3D ab initio sans structure de référence qui pourrait biaiser la découverte de nouvelles conformations.



**Figure 29 : Méthode de gradient descendant pour la détermination de carte Cryo-EM**

(a) Dans le cadre d'utilisation d'algorithme itératif, on peut voir que s'ils ne sont pas initialisés correctement, ils peuvent être bloqués dans un sous optimum local. (b) Dans le cas d'une approximation classique toutes les images sont utilisées alors que dans le cas de gradient stochastique seul un petit ensemble de particules est utilisé. Entre chaque itération le sous ensemble de particules est modifié pour utiliser des particules différentes. (c) L'algorithme utilisant la descente de gradient stochastique permet de déterminer des structures *ab-initio* tout en étant insensible à la structure de départ. On commence par l'initialisation d'une structure aléatoire arbitraire qui est améliorée progressivement lors des nombreuses itérations. Le succès de cette méthode peut être expliqué par le fait qu'elle permet d'explorer toute les cartes 3D pour finalement arriver à la structure correcte.

D'après Ali Punjani & al. 2017

### **g) Reconstruction des cartes de Cryo-EM**

Dans le cadre de la Cryo-EM pour pouvoir reconstruire un objet 3D d'après ses projections 2D, il est nécessaire de déterminer 5 paramètres : Les 3 angles d'Euler et deux mouvements de translation situés dans le plan. C'est pourquoi il est important de posséder une structure 3D initiale puisque dans ce cas on peut l'utiliser afin d'aligner les particules. Dans ce cas on commence par projeter la structure initiale pour obtenir des références 2D, chaque image expérimentale est comparée avec chacune de ces projections et un poids est appliqué suivant la probabilité que la particule soit dans l'orientation comparée. Une fois que l'orientation est connue on peut l'utiliser pour la rétroprojeter et obtenir un modèle 3D amélioré. Ces étapes combinées s'appellent l'affinement.

Il est connu que l'affinement itératif d'une structure peut conduire à un suraffinement causé par le bruit qui engendre une augmentation artificielle de la résolution et biaise la carte obtenue. Pour éviter ce phénomène une approche appelée gold-standard est utilisée. Lors du début de l'affinement, les particules sont séparées aléatoirement en deux moitiés et sont alignées contre une référence fortement filtrée afin d'effacer tout détail de haute résolution qui pourrait biaiser l'affinement. Chaque moitié est affinée séparément et la FSC (La FSC sera décrite dans le paragraphe suivant) entre les deux cartes est calculée. Dans le cas où l'affinement converge les deux moitiés sont combinées pour obtenir une seule carte de densité qui est filtrée suivant la résolution obtenue par la FSC. Si l'affinement n'a pas convergé alors on effectue un nouveau cycle.

### (1) La mesure de résolution

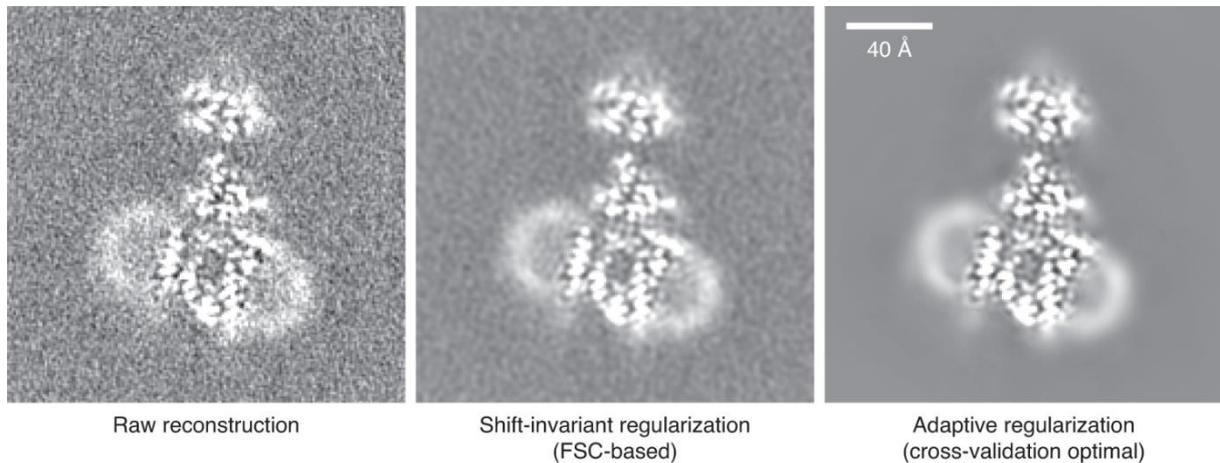
Pour mesurer la résolution d'une carte de cryo-EM on utilise la Fourier Shell correlation (FSC). La FSC représente un coefficient de corrélation en fonction de la fréquence spatiale de deux cartes de même résolution. Elle est obtenue en comparant dans l'espace de Fourier deux cartes de cryo-EM et permet de montrer comment les deux cartes sont corrélées suivant la fréquence spatiale. Typiquement la FSC est calculée à partir des deux cartes de la méthode gold standard. Elle permet l'estimation de la résolution, en lisant la valeur de la courbe au critère de résolution 0.143 (Rosenthal & Henderson, 2003). La valeur n'est pas une résolution absolue de la carte mais une résolution moyenne de celle-ci. Dans une carte à une résolution de 3,4 Å on peut donc quand même trouver des régions de meilleure et de pire résolution.

### **Le paramètre de régularisation**

La qualité de la reconstruction des cartes cryo-EM dépend fortement du paramètre de régularisation. La régularisation est une méthode très utilisée en informatique et en statistique qui consiste à s'appuyer sur des connaissances préalables pour pénaliser la complexité inutile d'un modèle afin d'éviter la sur interprétation. Dans le cas de la reconstruction 3D, elle est nécessaire pour atténuer les effets du bruit afin que seul le signal de la protéine soit présent/utilisé dans reconstruction de la densité 3D.

Dans la plupart des programmes d'affinement de structure un paramètre de régularisation sous forme d'un filtre linéaire est déduit de la FSC et utilisé pour filtrer la structure obtenue. Ce paramètre filtre toute la structure de la même manière, ce qui implique que certaines régions de la protéine se retrouvent sur filtrés et d'autre sous

filtrés ce qui engendre une perte de détail de haute résolution et une augmentation du bruit de manière local.

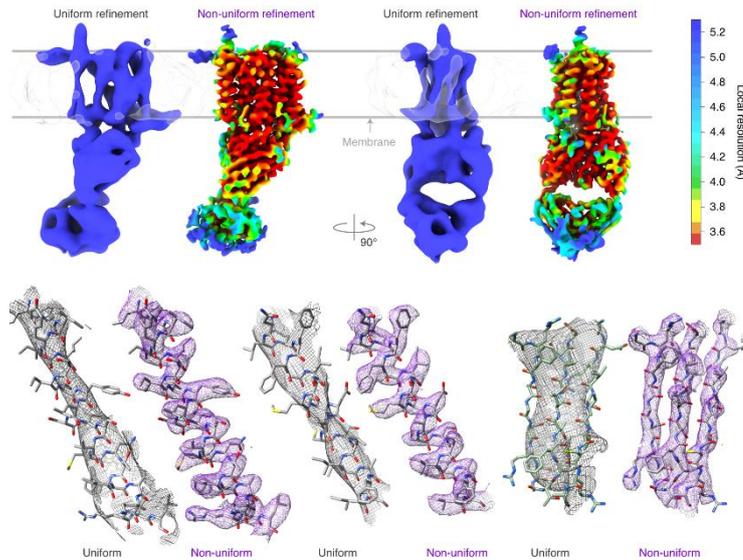


**Figure 30 : Régularisation adaptative et invariante**

Coupe centrale d'une reconstruction d'une protéine membranaire après 9 itérations : l'image de droite montre la reconstruction de base qui est très bruitée. L'image du centre montre la même reconstruction avec cette fois-ci l'utilisation d'un filtre adaptatif. L'image de droite est toujours la même reconstruction, mais où l'on a effectué une régularisation adaptative et montre que le bruit est bien plus atténué et les informations à haute résolutions conservées.

## (2) L'affinement non uniforme

En temps normal les algorithmes d'affinement assument qu'une protéine est uniforme et rigide. Cependant il existe de nombreux cas où certaines régions des protéines soient très désordonnées ou flexibles. Dans ce cas-ci, si une régularisation uniforme est appliqué, elle va introduire du bruit et atténuer certaines informations limitant la résolution atteignable. Pour pallier ce problème il est nécessaire de trouver une méthode où la régularisation est adaptative. Ali Punjani & al 2020 ont développé un paramètre de régularisation adaptatif qui permet de tenir compte des régions variable des protéines ce qui permet de retirer le bruit venant de ces régions tout en gardant le reste des informations structurale **Figure 30**. Cet algorithme a été implémenté dans un programme d'affinement non uniforme qui permet de drastiquement améliorer la résolution de certaines protéines notamment membranaire qui sont connue pour posséder d'importantes régions variables **Figure 31** (Punjani et al., 2020).



**Figure 31 : Affinement non uniforme**

Comparaison entre un affinement uniforme et non uniforme, on peut constater que dans ce cas d'étude de protéines membranaire la résolution est fortement améliorée lors de l'utilisation d'affinement non uniforme.

### **h) Séparation et détection de l'hétérogénéité**

La SPA analyse un échantillon purifié de protéines, ce qui implique que l'on étudie un même objet présent en de nombreuses copies cependant, l'échantillon est loin d'être homogène. Tout d'abord il est connu que les protéines peuvent posséder plusieurs conformations introduisant un premier niveau d'hétérogénéité. De plus, la qualité de l'échantillon joue également un rôle, car sa propension à agréger ou que des sous-unités/partenaires soit manquantes/dissociés vont accroître l'hétérogénéité. Pour reconstruire des structures à haute résolution, il est donc nécessaire de pouvoir classer les particules en fonction de leurs conformations, sous-unités manquantes/présentes et de leurs agrégations/altérations. La méthode qui est classiquement utilisée pour permettre de gérer l'hétérogénéité est la classification en 3D non supervisées, qui utilise des algorithmes de vraisemblance régularisé (Scheres, 2012). L'approche consiste en l'utilisation d'un modèle 3D initial fortement filtré qui sera utilisé pour l'alignement des particules. Les particules sont séparées en X groupes suivant le nombre de classes final voulue. Les structures sont affinées simultanément et permettent à la fois de

déterminer l'angle des particules et leurs appartenances à une classe sans avoir à connaître à l'avance les différentes conformations. Une fois terminée les différentes classes obtenues sont inspectées visuellement pour sélectionner les classes d'intérêts et retirer les classes qui correspondent à du bruit/particules agrégés/endommagés.

## **II. Matériels & Méthodes**



La purification de NuA4 est réalisée par le Dr Adam Ben-Shem . Le complexe est purifié à partir d'extraits nucléaires de la levure *Pichia pastoris* en utilisant un Strep tag sur la sous-unité Eaf1 ou Epl1 dans les cas des mutants. Deux litres de levure ont été cultivées à 24 °C avec du glycérol comme source de carbone et collectées lorsque la densité optique à 600 nm a atteint 16-18. Les cellules ont été lavées dans de l'eau glacée puis traitées avec 10 mM de DTT. La paroi cellulaire a été digérée par addition de lyticase et les sphéroplastes ont été centrifugés à 5 500 g pendant 20 minutes. Toutes les autres étapes ont été effectuées entre 0 et 4 °C. Des inhibiteurs de protéase ont été ajoutés à tous les tampons. Les sphéroplastes ont été lavés deux fois dans du sorbitol 1,1 M et ont ensuite été lysés par suspension dans un tampon hypotonique (15-18% Ficoll 400, 0,6 mM MgCl<sub>2</sub>, 20 mM K-phosphate tampon pH 6,6. Du saccharose (0,1 M) et du MgCl<sub>2</sub> (5 mM) ont ensuite été ajoutés. Les noyaux (et certains débris) ont été précipités en centrifugeant à 33 000 g pendant 37 minutes, remis en suspension dans un tampon de lavage (saccharose 0,6 M, polyvinylpyrrolidone 8%, MgCl<sub>2</sub> 1 mM, tampon phosphate 20 mM pH 6,6) et centrifugés à nouveau à 34 000 g pendant 50 minutes. Les noyaux ont été remis en suspension dans un tampon d'extraction faiblement ionique (40 mM HEPES pH 8,0, 20% de saccharose, 8 mM MgCl<sub>2</sub>, 5-6 mM DTT). Après 20 minutes d'incubation, 300 mM de NaCl, 2 mM de CaCl<sub>2</sub> et 150 µl de solution de  $\alpha$ -amylase (MegaZyme) ont été ajoutés. Après 30 minutes d'incubation supplémentaires, les débris ont été précipités à 33 000 g pendant 38 minutes. Le surnageant a été recueilli et 1 à 2 % de PEG 20 000 ainsi que 5 mM de MgCl<sub>2</sub> supplémentaires ont été ajoutés pour précipiter certains organites et parties de membrane restants par une courte étape de centrifugation à 33 000 g pendant 10 minutes. Le culot a été solubilisé dans un volume minimal et de l'avidine a été ajoutée pour bloquer les protéines biotinylés endogènes. La suspension a été incubée avec des billes de streptavidine pendant 4 h dans un tampon A (40 mM HEPES pH 8,0, 250 mM de chlorure de sodium, 10% de saccharose, 2 mM MgCl<sub>2</sub>, 2 mM DTT) lavé 5 fois et élué avec un tampon A contenant 10 mM de biotine. L'éluat a été concentré avec du Millipore Amicon-Ultra (coupure de 50 kDa) et passé dans un gradient de 10 à 30 % de saccharose avec le tampon B (20 mM HEPES pH 8,0, 150 mM acétate de potassium, 2 mM DTT, 6 mM MgCl<sub>2</sub>) dans le

rotor SW60 (38 300 tr/min pendant 13,5 h). NuA4 a été fractionné à environ 25 % de saccharose et concentré avec de l'Amicon-Ultra à ~2 mg/ml.

Dans le cas d'études par microscopie électronique l'échantillon est fixé par méthode Grafix qui consiste à déposer l'échantillons sur un gradient de sucrose et de Glutaraldéhyde pour le fixer tout en retirant les agrégats. Il s'avère que cette méthode de fixation est essentielle pour le maintien de l'intégrité du complexe NuA4 (voir Réticulation de NuA4 et analyse spectrométrique). Une fois le gradient terminé l'échantillon est dialysé, pour retirer le sucrose.

Pour les études/tests avec le nucléosome, les histones de *Xenopus laevis* ont été clonées, exprimées chez *E. coli* et purifiées comme décrit dans (Luger K & al, 1999 ; Dyer PN, et al, 2004). En résumé, le protocole de purification comporte trois étapes : lyse des cellules avec dénaturation des histones, une gel filtration dans des conditions dénaturantes et chromatographie par échange d'ions HPLC dans des conditions dénaturantes. Une fois les histones purifiées, ils sont utilisés pour reformer l'octamère d'histone, puis le nucléosome est reconstitué par l'ajout de fragments d'ADN et en modifiant la concentration de sel. La purification est achevée par une chromatographie d'échanges d'ions qui permet de s'éparer les octamères et ADN libre et les nucléosomes formés.

## **A. Préparation des échantillons et analyse en microscopie électronique**

### 1. Coloration négative

La technique de colorations négative a été utilisé pour déterminer les conditions de fixation optimum de NuA4. Son interaction avec le nucléosome, mais également comme contrôle avant chaque congélation de grille pour vérifier la qualité et la concentration de l'échantillons.

## **a) Matériel**

- Grille de type mesh 300 en cuivre (Electron Microscopy Science) recouverte d'un carbone continue (grille commerciale)
- Pince de précision pour manipuler les grilles de microscopie électronique
- Plasma cleaner, instrument servant à produire un plasma gazeux pour charger la surface des grilles par un potentiel électrostatique pour rendre leurs surfaces hydrophiles afin d'absorber les protéines. Dans notre cas nous avons utilisé un Nanoclean modèle 1070 de la marque Fischione.
- Un sel d'atome lourd pour colorer l'échantillon qui est ici l'Acétate d'uranyle en solution à 2% (p/v)
- Buffer pour les étapes de lavages
- Papier filtre. Le Protocole nécessite plusieurs étapes de séchages, afin de retirer les excès de solution/colorant
- Parafilm (VWR international SAS). Le protocole utilise plusieurs étapes de lavage. Le parafilm est utilisé pour déposer de petites gouttes de solution sur lesquelles la grille EM peut être placée.

## **b) Protocole**

1. Effluvage des grilles recouvertes de carbone pour les rendre hydrophiles. Dans le cas du Fischione Nanoclean 1070 un plasma à 30 % et appliqué et le mélange de gaz est de 20 % oxygène 80 % Argon, les grilles sont placées dans un étui en métal percé qui permet au plasma d'y accéder. Le temps d'effluvage est de 90 s.
2. Une grille effluée est saisie à l'aide d'une pince de précision, qui est clampée pour éviter que la grille ne tombe. 3 µl d'échantillon de NuA4 sont déposées à une concentration de 0.07 mg/ml. L'échantillon est absorbé pendant 1 minute.

3. Un morceau rectangulaire de parafilm est posé sur la paillasse et 3 gouttes de 40  $\mu$ l de buffer sont déposées décus, de même qu'une goutte de 40  $\mu$ l acétate d'uranyle à 2%.
4. Une fois que l'échantillon a fini d'être absorbée sur la grille, elle passe successivement d'une goutte de buffer à une autre de manière à retirer l'excédent de matériel qui n'est pas absorbé. Une fois les 3 étapes de lavage faites, la grille est déposée face carbone sur la goutte d'acétate d'uranyle pendant 30 s.
5. L'étape de coloration terminée, la grille est séchée en utilisant du papier filtre qui vient absorber l'excédent de liquide depuis le bord de la grille de manière à ne pas endommager le carbone.
6. La grille est stockée dans une boîte numérotée en attendant d'être analysée

### **c) Microscope utilisé**

Le microscope utilisé pour l'acquisition d'image en coloration négative fut un Tecnai F20 de l'entreprise FEI, il opère à 200kV et le grandissement utilisé est de 66 000 ce qui correspond à une taille de pixel de 2,12 Å/pixel. Le défocus utilisé est de -1.5 micron. La caméra utilisée est une Gatan Ultrascan 1000 CCD d'une taille de pixel physique 14  $\mu$ m

### **d) Analyse d'images**

Les étapes de sélections des particules ont été faites à l'aide du logiciel Gautomatch qui utilise des algorithmes de template matching pour reconnaître automatiquement les particules à sélectionner, mais également à l'aide de crYOLO qui fonctionne d'après l'algorithme de reconnaissance d'image YOLO. Le calcul de la CTF est fait par le logiciel Gcft.

Le reste des étapes de classification 2D/3D et affinement sont faites à l'aide du logiciel Relion ou Cryosparc.

## 2. Cryo microscopie électronique

Une fois que les échantillons ont été optimisés/vérifiés par coloration négative ils sont vitrifiés suivant le protocole suivant

### a) Matériels

- Grille de microscopie électronique composée d'un alliage de cuivre et de rhodium recouvert par un carbone à trou dont le diamètre est de 2  $\mu\text{m}$ . Quantifoil R2/2.
- Pince de précision pour manipuler les grilles.
- Plasma cleaner, le même dispositif a été utilisé que pour de la coloration négative
- Azote liquide, il est utilisé à plusieurs étapes de la préparation, d'abord pour refroidir et maintenir l'éthane liquide ensuite pour garder les échantillons vitrifiés à une température de  $-196^{\circ}\text{C}$  afin d'éviter qu'il se réchauffe et que la glace se transforme et ainsi permettre le stockage à long terme des grilles.
- Un Vritobot produit par l'entreprise FEI qui est le dispositif qui permet de vitrifier les échantillons sur une grille EM
- Cryo-boxes. Ces boîtes rondes sont utilisées pour stocker les échantillons vitrifiés dans l'azote liquide.
- Bouteille de gaz d'éthane, L'éthane sera refroidi de manière à le rendre liquide pour qu'il serve à congeler rapidement les échantillons garantissant l'obtention de glace vitreuse.

### b) Protocole

1. Effluvage des grilles recouvertes de carbone à trous. La même machine et les mêmes paramètres ont été utilisés qu'en coloration négative

2. La température de la chambre du Vitrobot a été réglée à 6°C et le pourcentage d'humidité à 100%, pour garder un taux d'humidité constant des papiers filtre imbibés d'eau ont été rajoutés dans la chambre du Vitrobot.
3. Le conteneur d'éthane est refroidi par l'azote liquide. Une fois à bonne température un flux d'éthane gazeux est introduit dans le conteneur froid avec pour but de le liquéfier. L'éthane est maintenu à bonne température grâce à l'azote liquide.
4. La grille préalablement effluée est attrapée sur ses bords par une pince qui est ensuite chargée sur le Vitrobot et introduite dans la chambre. Un temps d'attente de 30 s est respecté de manière à ce que la grille atteigne la température de la chambre.
5. Une fois la grille chargée et à température, 3 µl d'échantillon de NuA4 à une concentration de 0.225 mg/ml ont été déposées sur la grille. La force d'absorption du Vitrobot a été ajustée à 8, le temps d'absorption à 1s et le temps d'attente à 0s.
6. La grille est plongée automatiquement dans l'éthane liquide par le Vitrobot puis récupérée et stockée manuellement dans une cryobox.

### **c) Microscopes utilisés**

#### *(1) Screening*

L'analyse des grilles vitrifiées est faite par deux microscopes : le microscope FEI Polara de l'institut opérant à 100 kV qui est équipé d'une caméra Falcon II et le microscope Glacios opérant à 200 kV. Un atlas est effectué avec un faible grandissement pour évaluer la qualité de la glace, des images à plus grand grandissement (x78 000) sont également prises pour vérifier l'état de l'échantillon suivant l'épaisseur de la glace.

## (2) Collecte de donnée

Tous les jeux de données ont été collectées au sein de l'institut sur le microscope électronique Titan Krios possédant un correcteur de  $C_s$ , opérant à 300 kV en mode nano sonde à l'aide du logiciel serialEM pour la collecte automatique des données. Les images des films ont été enregistrées sur une caméra à détection directe des électrons Gatan K2 de  $4k \times 4k$  après un filtre d'énergie Quantum Ls 967 utilisant une largeur de fente de 20 eV. Les images ont été acquises en mode super-résolution à un grossissement de 105 000, ce qui a donné une taille de pixel de 0,55 Å. Quarante images de film ont été enregistrées à une dose de  $1,32 \text{ e}/\text{Å}^2$  par image, ce qui correspond à une dose totale de  $52,8 \text{ e}/\text{Å}^2$ , mais seules les 38 dernières images ont été conservées pour le traitement d'image.

### **d) Analyse des images**

La correction au mouvement induit par le faisceau et la pondération en fonction de la dose est faite par les logiciels motionCorr2 ou WARP. Le calcul de la CTF est fait par le logiciel Gctf (Zhang, 2016) d'après les images non pondérées, alors que le reste des étapes d'analyse d'images se font grâce aux images pondérées. Après un examen manuel, les images présentant une mauvaise CTF, une agrégation de particules ou une contamination par la glace ont été rejetées. Les particules ont ensuite été sélectionnées par le programme crYOLO. Par la suite les programmes de traitement d'image Relion et Cryosparc V2 ont été utilisés. Les jeux de données subissent dans un premier temps plusieurs classification 2D pour retirer tout agrégats ou cristaux de glace, tout en évitant au maximum de perdre des particules. Une première étape de détermination de structure initiale/classification 3D ab-initio est faite et les modèles obtenus sont utilisés pour les étapes de classification 3D. Un premier cycle de classification 3D est effectué et seul les classes présentant des structures secondaires visible sont gardées. La classification 3D sur l'ensemble de NuA4 ne permet pas de classifier toute l'hétérogénéité du complexe, pour ce faire des étapes de classification sans alignement sont faite en utilisant un masque sur la région d'intérêt. Les étapes d'affinements se sont faites sur le logicielle Cryosparc avec le mode non uniforme. La partie la plus stable de NuA4 a été

affinée en utilisant un masque autour d'elle avec le module d'affinement local de Cryosparc. La réalisation du masque s'est faite grâce à l'utilisation des logiciels Chimera, Relion et Cryosparc. Une fois la région d'intérêt déterminée, le reste des densités sont effacées manuellement par la commande erase du logiciel de visualisation Chimera. La carte obtenue est ensuite utilisée par les logiciels Relion ou cryosparc pour générer un masque binaire, qui sera utilisé pour les étapes de classifications / affinement3D. Après avoir généré un masque, un affinement local des particules est réalisé. Il consiste à utiliser l'alignement des particules obtenu lors de l'étape d'affinement précédente et l'ajout d'un masque pour focaliser l'alignement sur une portion du complexe. Ceci a pour but d'améliorer la qualité/résolution de la région présente dans le masque. L'estimation de la résolution est déterminée en utilisant le critère de corrélation de Fourier (FSC) = 0,143.

## **B. Test d'activité**

NuA4 est une HAT, j'ai pu réaliser des tests d'activité pour vérifier que notre complexe était actif ainsi que les mutants de NuA4 que j'ai pu étudier. SAGA a été utilisé comme contrôle positif.

### 1. Acétylation

Pour 100 µl de volume réactionnelle A :

- 90 µl de f NCP
- 10 µl de NuA4 ou 7.5 µl of SAGA

Pour 200 µL de volume réactionnelle B :

- 1,5 µl de BSA stock
- 15 µl Ac-CoA (1mM) ou 0.8 µl de solution stock
- 60 µl de glycérol v/v 50%
- 30 µl de 10X buffer

- QSP 200  $\mu$ l de H<sub>2</sub>O

Produit	Concentration
NuA4	1mg/ml (1 $\mu$ M)
BSA stock	20 mg/ml
NCP	3,5 $\mu$ M
Ac-CoA	20 mM

**Tableau 3 : Réactifs pour la réaction d'acétylation**

1. 20  $\mu$ l de volume réactionnel A sont placés dans un tube Eppendorf
2. 10  $\mu$ l de volume réactionnel B sont ensuite ajoutés dans l'Eppendorf
3. La réaction est incubée pendant 10 minutes à température ambiante ou 30 minutes sur glace
4. La réaction est stoppée par l'ajout de SDS et bouillie pendant 3 minute
5. Le tube est congelé pour l'analyse ultérieure par western blot

## 2. Western blot

Solution	Composition
TBST	0,137 M NaCl, 20 mM Tris ph7,5, 0,1% tween 20
Solution de Blocage	5 % w/v BSA dans du TBST
Tampon de transfert	100 ml de tampon western blot, 200 ml of ethanol and 700 ml of water

Anticorps Primaire	Solution de blocage avec Ac primaire dilué 1/1000
Anticorps Secondaire	Solution de blocage avec Ac secondaire dilué 1/10 000 ou 1/ 5 000

**Tableau 4 : Solution utilisé pour le Westernblot**

Pour vérifier si les nucléosomes ont bien été acétylés, j'ai réalisé un western blot avec un anticorps primaire qui reconnaît les lysines acétylées. L'anticorps a été acheté chez l'entreprise Cell Signaling Technology sa référence est #9441.

1. Faire tremper la membrane PVDF pendant 1 minute avec de l'éthanol
2. Lavage rapide à l'eau distillée
3. Équilibrer la membrane en utilisant un tampon de transfert pendant 10 min
4. Placer la membrane contre le Gel et de chaque côté ajouté des papiers filtres et des éponges préalablement imbibées de tampon de transfert.
5. Lancer le transfert pendant 1H à 70 V
6. Après le transfert, bloquer la membrane en utilisant la solution de blocage pendant 1H à température ambiante avec une agitation douce (Pour cette étape, la BSA peut être remplacée par du lait en poudre).
7. Faire un lavage grossier à l'aide de TBST 1X pour éliminer la majeure partie de la solution de blocage.
8. Effectuer 3 étapes de lavage, chaque étape utilise 15 ml de TBST 1X en agitant 5 min.
9. Enlever tout le tampon et incuber la membrane avec l'anticorps primaire pendant une nuit à 4°C.
10. Après incubation, récupérer l'anticorps et le stocker à -20°C.

11. Refaire les étapes de lavage 7 et 8.
12. Incuber la membrane avec l'anticorps secondaire dans une solution de blocage pendant 1H à température ambiante.
13. Refaire les étapes de lavage 7 et 8.
14. Mélanger 2 ml de chaque réactif pour la réaction HRP
15. Mettre le mélange sur la membrane et attendre 1 minute
16. Exposer la membrane (la première exposition peut être acquise en mode automatique, après avoir mis manuellement le temps d'exposition pour éviter la surexposition).

### **C. Détermination de modèle atomique**

Les logiciels I-tasser (Yang & Zhang, 2015) et Swiss-model (Bertoni et al., 2017) ont été utilisés pour modéliser des protéines par homologie. Les modèles générés ont ensuite été placés dans les cartes de densité à l'aide de la fonction « fit » du logiciel de visualisation Chimera dans sa version 1.14.

Pour les protéines ne pouvant pas être modélisées par homologie de structure existante, le logiciel Coot (Emsley & Cowtan, 2004) a été utilisé pour établir des modèles atomiques à l'aide des différentes carte Cryo-EM. Pour nous aider dans la construction de modèle atomique, nous avons utilisé des prédictions de structures secondaires et de régions désordonnées à l'aide des serveurs en ligne de RaptorX (Z. Wang et al., 2011) et Psi-Pred (Buchan & Jones, 2019). Les logiciels Phenix (Terwilliger, 2010) et Isolde (Croll, 2018) ont ensuite été utilisés pour affiner le modèle atomique déterminé.



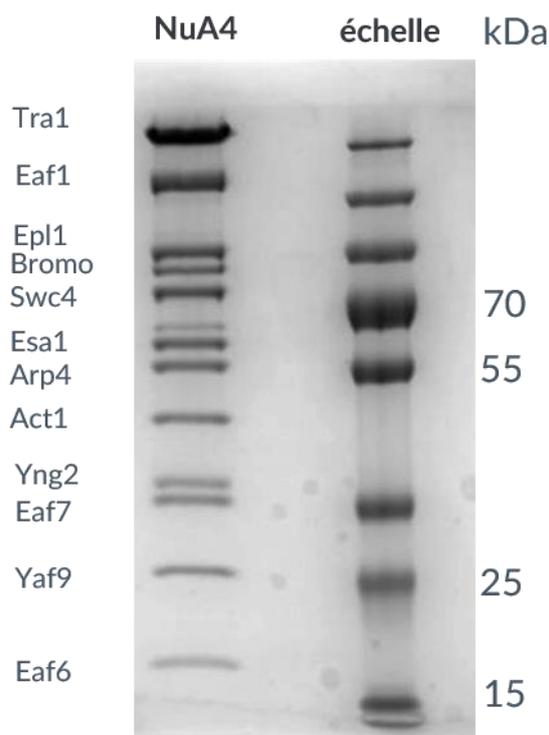
# **III. Résultats**



Au cours de ma thèse j'ai pu travailler sur les coactivateurs NuA4 et SAGA. Dans cette partie je vais vous parler des résultats que j'ai pu obtenir, dans un premier temps sur NuA4 et dans un second sur SAGA.

En ce qui concerne mon premier projet sur NuA4 mon rôle a été essentiellement la congélation des échantillons, leur évaluation, la collecte de données en cryo microscopie, la collecte de données en coloration négative, l'analyse des images obtenues et pour finir la détermination d'un modèle atomique de NuA4. Dans le cadre du projet SAGA ma principale contribution fut l'élaboration du modèle atomique et son affinement à partir des cartes déterminées par Gabor Papai.

### A. NuA4



**Figure 32 : Purification du complexe NuA4**

Gel SDS page du complexe NuA4 purifié, Les 12 sous-unité, qui compose le complexe sont visible.

Après les étapes de purifications, l'échantillon est analysé par Gel SDS page. Cette étape est cruciale car elle permet de nous assurer que le complexe purifié comporte bien toutes les sous-unités de NuA4 **Figure 32**.

## 1. Détermination de la structure de NuA4 par cryo-EM

Dans cette partie je vais vous présenter les différentes structures que j'ai pu obtenir de NuA4. Trois structures de NuA4 ont été obtenues : l'une d'un NuA4 classique appelé maintenant WT et de deux autres NuA4 mutants possédant chacun une délétion plus ou moins courte dans Epl1, l'intérêt de ces délétions sera expliqué plus tard.

### a) L'analyse des images

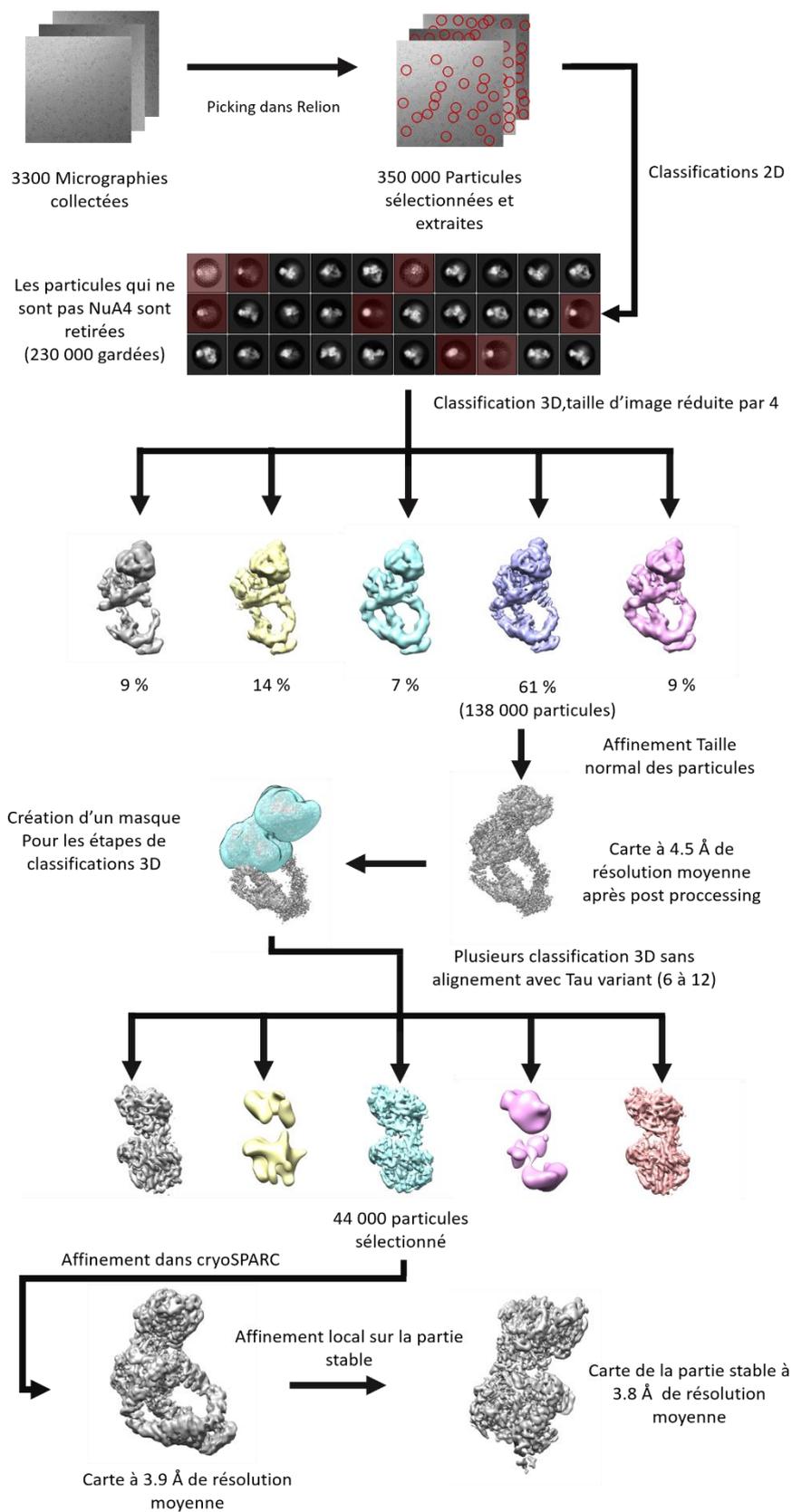
#### (1) NuA4 WT

Le premier jeu de données que j'ai analysé au début de ma thèse concerne un échantillon WT de NuA4. Le microscope FEI titan a permis d'obtenir 3300 micrographies. Les premières étapes de pré analyse ont été réalisées de manière automatique par le pipeline de la plateforme qui utilise les programmes motioncor2 et Gctf pour aligner et calculer la CTF de chaque image. Une fois ces premières étapes terminées, j'ai sélectionné les images dont la résolution était supérieure à 5 Å et celles dont l'image et la CTF n'étaient pas altérées, mais également les images dépourvues d'agrégation et de contaminant.

La sélection des particules et la suite de l'analyse des images s'est faite à l'aide du logiciel RELION. Plus de 350 000 particules ont été sélectionnées et extraites en utilisant comme référence un disque gaussien. Pour rendre les temps de calculs plus courts les premières étapes de classification 2D/3D sont faites en réduisant la taille des particules par 4. Plusieurs cycles de classification 2D ont été effectués dans le but d'éliminer les particules n'étant pas notre échantillon (agrégation, glace, contaminant ...). Au terme de ces classifications 2D, 230 000 particules ont été gardées pour la suite de l'analyse. Nous disposions déjà d'un modèle 3D de NuA4 obtenu par moi-même lors de mon stage de fin de Master. Grâce à ce modèle 3D, j'ai effectué directement un premier affinement et obtenu une structure à 8.9 Å. Compte tenu de la taille des pixels qui était alors de 4,4 Å la limite de résolution pour ces images réduites a été atteinte. J'ai ensuite utilisé l'alignement des particules obtenu pour effectuer une

classification 3D sans alignement. Le principe est de ne pas réaligner les particules lors de la classification 3D, mais d'utiliser l'alignement obtenu lors de l'affinement, afin de limiter les temps de calculs. Une fois la classification 3D terminée, j'ai obtenu 5 classes parmi lesquelles j'ai sélectionné celles qui étaient majoritaires. Ces classes étaient également celles qui avaient la meilleure résolution et dont on pouvait commencer à voir les hélices alphas.

Les particules sélectionnées ont ensuite été réextraites cette fois-ci dans leur taille d'origine correspondant à une taille de pixel de 1,1 Å. Un nouvel affinement a été réalisé et j'ai obtenu une structure à 4.5 Å. Pour tenter d'améliorer la résolution sur une partie du complexe j'ai créé un masque autour de la région la plus stable, qui a été utilisé pour des étapes de classifications 3D sans alignement, tout en faisant varier le paramètre de régularisation (entre 6 et 12). Cette étape m'a permis d'isoler 44 000 particules qui ont ensuite été affinées à l'aide du logiciel cryoSPARC, ce qui a permis d'obtenir une structure à 3.9 Å de résolution moyenne. Un affinement local a également été réalisé sur la même partie qui a été classifiée en utilisant le même masque et le nouvel alignement des particules obtenu après l'affinement. Cette méthode permet, en donnant un masque à l'algorithme d'affinement et l'alignement de particules obtenue précédemment, de se focaliser sur une région du complexe afin d'essayer d'obtenir une résolution plus élevée par un alignement plus précis des particules. Après l'étape d'affinement local j'ai obtenu une carte de 3.8 Å de résolution moyenne



**Figure 33 : Analyse des images du NuA4 WT**

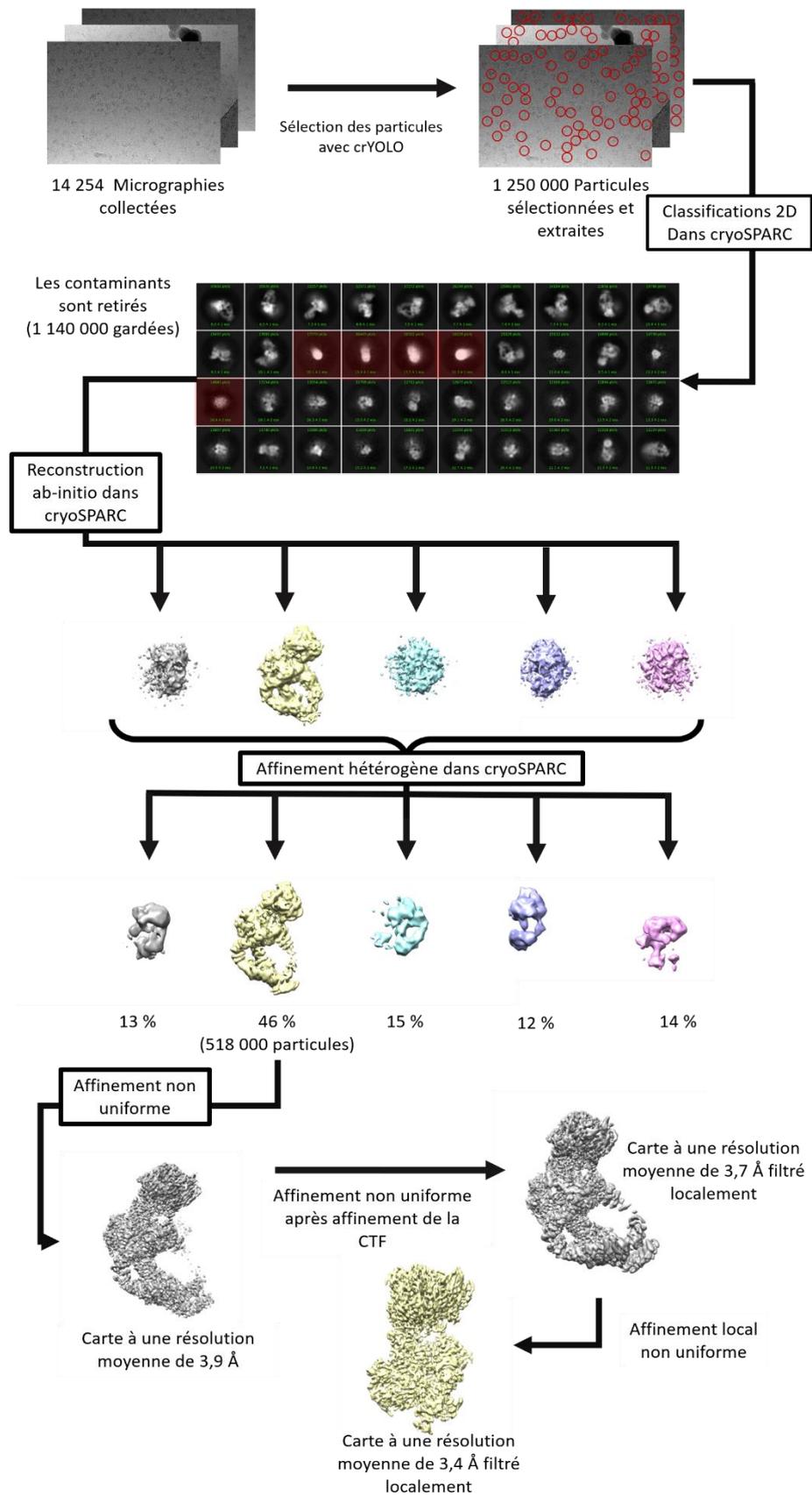
## B. NuA4 Mutant

Ma méthode d'analyse des images a par la suite évolué aux vues des nouveaux programmes et de leurs mises à jour.

Les premières étapes de pré analyse sont maintenant faites par le logiciel WARP qui permet l'alignement des micrographies et le calcul de la CTF. Ce logiciel est également capable de réaliser la sélection automatique de particules, mais j'ai choisi d'utiliser le logiciel crYOLO pour cette tâche. J'ai utilisé une vingtaine de micrographies de références pris aléatoirement dont j'ai sélectionné les particules que j'ai utilisé pour entraîner le programme. Le logiciel m'a permis de sélectionner 1.3 million de particules et ont été extraites à l'aide du logiciel RELION.

Les étapes de classification 2D ont ensuite été réalisées par le logiciel cryoSPARC. La détermination 3D de modèle ab-initio de cryoSPARC a permis de reconstruire 5 références qui sont ensuite utilisées pour les étapes de classification 3D. La classe 2 qui est la classe majoritaire et qui possède la plus haute résolution 6.2 Å est ensuite utilisée pour continuer l'analyse des images.

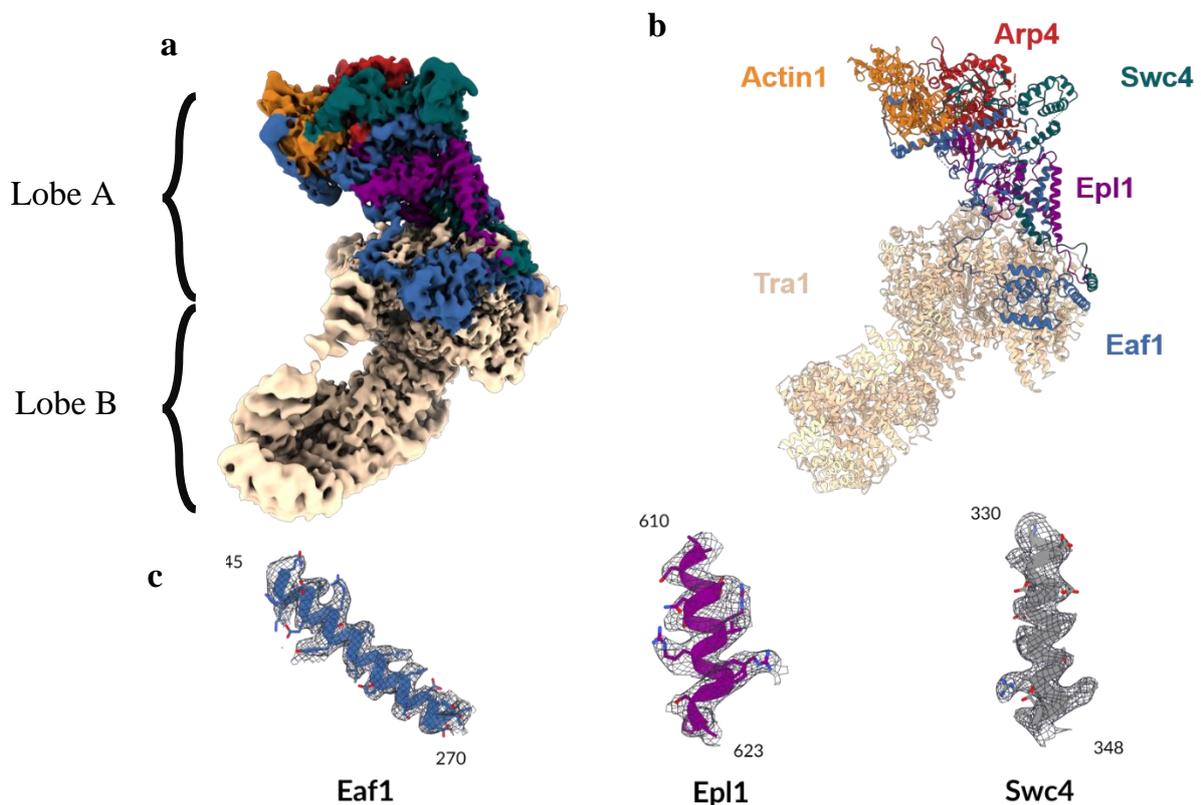
Un premier affinement non uniforme est fait et a permis d'obtenir une structure avec une résolution moyenne de 3.9 Å. On affine localement la CTF, ce qui permet d'améliorer l'affinement non uniforme à 3.7 Å. Un affinement local d'une portion de cette structure a ensuite été fait. L'idée est que l'une des régions du complexe semblait être plus stable et en masquant cette région, à l'aide d'un masque binaire, le programme d'affinement va pouvoir se focaliser en ne prenant en compte que le signal présent dans le masque. Ceci va affiner l'alignement des particules sur cette portion du complexe, en utilisant comme alignement de départ celui obtenue précédemment. Grâce à cette stratégie j'ai pu obtenir une carte de cette portion du complexe à une moyenne de 3.4 Å ce qui représente une nette amélioration de la résolution et facilitera l'interprétation des résultats.



**Figure 34 : Analyse des images de NuA4 mutant**

## 1. Modèle atomique de NuA4

Grâce aux cartes obtenues j'ai pu déterminer un nouveau modèle atomique de NuA4 **Figure 35 a & b**. La structure se compose de deux lobes A et B reliés entre eux par un important réseau d'hélices formé par Swc4, Eaf1 et Arp4 **Figure 35 a**. Le lobe A est composé d'Act1, Arp4, Swc4, la partie C-terminale d'Epl1 et N-terminale d'Eaf1. Le lobe B est quant à lui composé d'une grande majorité de Tra1, du C-terminale d'Eaf1 et d'une hélice située en C-terminale de Swc4.



**Figure 35 : Structure de NuA4 et modèle atomique**

(a) Carte cryo-EM du complexe NuA4 à 3.7 Å de résolution moyenne filtrée localement. La carte est colorée suivant les sous-unités lui correspondant. (b) Modèle atomique de NuA4, le modèle se compose de Tra1, Actin1, Arp4, Swc4, Epl1 et Eaf1. (c) Exemple de densité montrant que la résolution de la carte est suffisante pour établir un modèle atomique

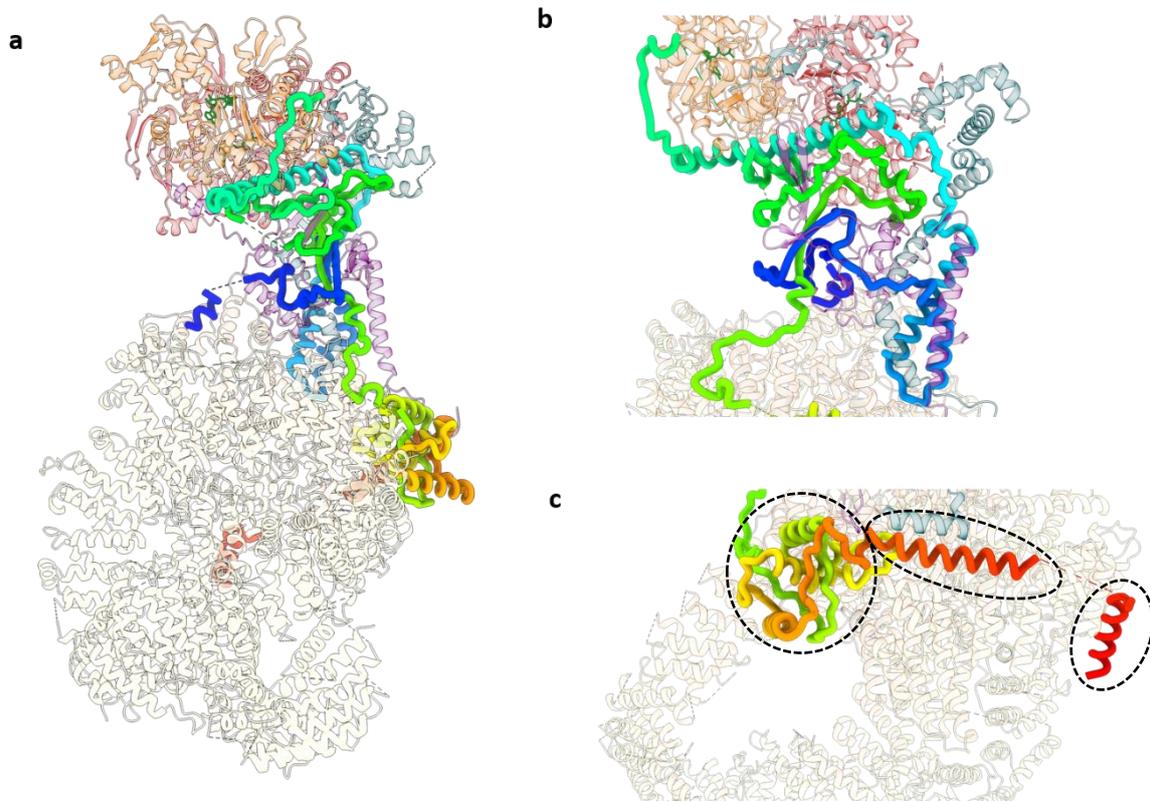
La suite de ce chapitre va être consacrée aux détails structuraux du modèle atomique de NuA4 obtenu en commençant par la protéine Eaf1 qui est pour rappel essentielle à l'intégrité du complexe.

### a) Eaf1 une protéine éclatée

Dans la structure du complexe TEEAA, Eaf1 est présentée comme étant une protéine très ordonnée, compacte et faisant le lien structurel entre les actines et Tra1 (X. Wang et al., 2018). Ce cœur très organisé agirait comme une plateforme où viendraient s'intégrer les différents modules de NuA4. Elle est également décrite comme occupant intégralement l'espace entre les actines et Tra1. Cependant, la résolution de ce complexe n'est que de 4,7 Å ce qui appelle à la prudence et demande la détermination d'une structure à plus haute résolution pour établir précisément un modèle de NuA4.

Grâce à notre modèle atomique obtenu à l'aide de cartes allant jusqu'à 3.4 Å de résolution moyenne, nous pouvons enfin décrire avec exactitude les protéines et les interactions présentes dans NuA4 **Figure 35 c**. Il en résulte que le lobe A n'est pas constitué d'Act1, Arp4 et Eaf1 mais de Act1, Arp4, Swc4, Eaf1 et Epl1, de plus Eaf1 est loin d'être une protéine compacte où viendrait s'intégrer le reste des protéines constituant NuA4 **Figure 35 a & b**.

Eaf1 est une protéine qui est « éclatée » au sein de la structure et qui réalise de nombreux points d'interaction avec différentes protéines/parties du complexe **Figure 36 a**. La partie N-terminal d'Eaf1 est dédiée à interagir et maintenir l'intégrité du lobe A **Figure 36 b**. Eaf1 forme une boucle parcourant tout le lobe A. Elle descend vers Tra1 avant de remonter vers les actines pour enfin se diriger vers le lobe B où sa partie C-terminale effectue les dernières interactions avec Tra1 **Figure 36 c**. Notre modèle propose donc une nouvelle vision d'Eaf1 qui nous montre qu'elle parcourt le complexe afin d'établir des interactions entre les différents constituant de NuA4.



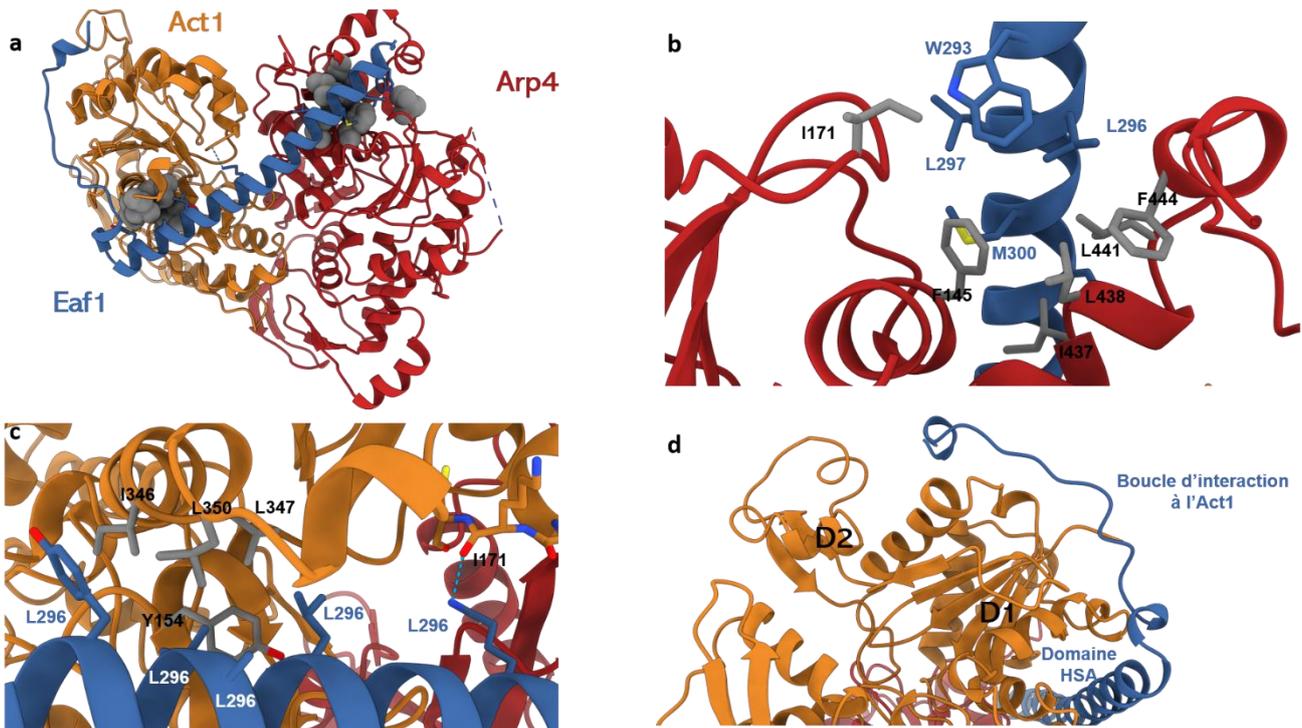
**Figure 36 : Tracé d'Eaf1 dans NuA4**

(a) La protéine Eaf1 est coloriée de sa partie N-terminale (en bleu) à sa partie C-terminale (en rouge). (b) Tracé d'Eaf1 dans le lobe A de NuA4, on peut voir que la protéine passe par le bas du lobe avant de remonter interagir avec Actin1 et Arp4, puis de redescendre pour rejoindre le lobe B. (c) Lobe B de NuA4 on voit Eaf1 arriver et interagir avec Tra1 en trois points indiqués par les pointillés.

(1) *Le domaine HSA, un site interaction pour Actin1 et Arp4*

Il existe plusieurs structures atomiques dans lesquels un domaine HSA s'associe à Act1 et Arp4 (Brahma et al., 2018; Cao et al., 2016), notre structure n'échappe pas à la règle et le domaine HSA d'Eaf1 sert de plateforme pour l'interaction des actines avec NuA4. La première moitié N-terminale de la HSA intègre Arp4 et la seconde moitié Act1. Dans les deux cas, à chaque extrémité de la HSA des interactions hydrophobes sont formées avec Act1 et Arp4 et garantissent leur maintien dans le complexe **Figure 37 a**.

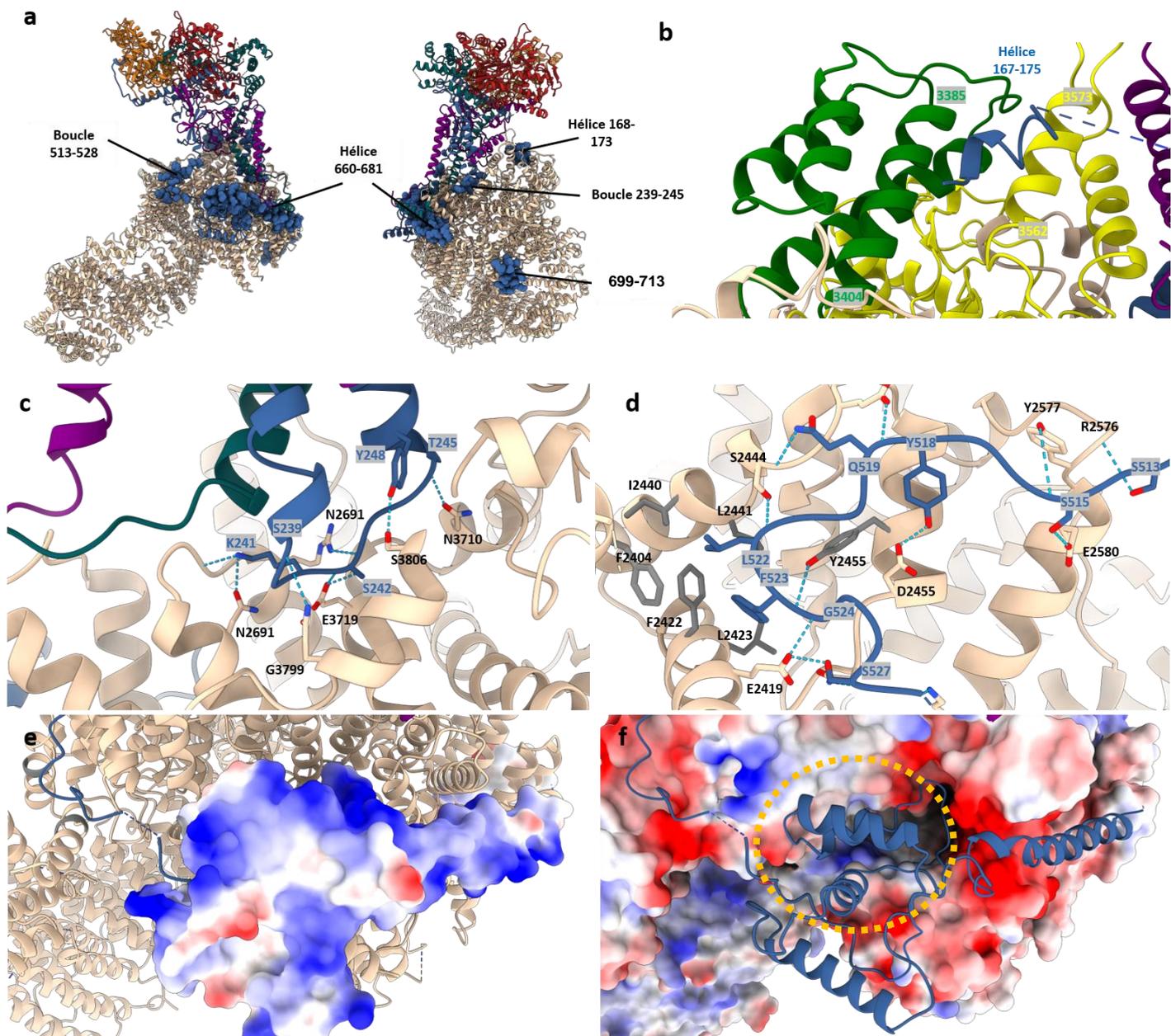
Les interactions hydrophobes entre Eaf1 et Arp4 sont formées par les résidus W293, L296, L297, M300 et L303 de la HSA avec les résidus F145, I171, L437, L438, L441 et F444 d'Arp4 **Figure 37 b**. Dans le cas d'Eaf1 et Act1, les résidus L321, A324, V 325 et Y328 de la HSA interagissent avec les résidus hydrophobes L347, L 350, I346 et Y144 d'Act1 **Figure 37 c**. Au milieu de la HSA la K314 fait une liaison hydrogène avec la chaîne principale de la K374 d'Act1 **Figure 37 c**.



**Figure 37 : Interaction entre Eaf1, Act1 et Arp4**

(a) Act1 et Arp4 sont maintenus dans le complexe NuA4 par la grande hélice bleue qui est le domaine HSA de Eaf1. Arp4 est du côté N-terminale de la HSA et Act1 du côté C-terminale. Les résidus affichés en sphère et coloriés en gris représentent les zones hydrophobes d'Act1 et Arp4. (b) Les résidus hydrophobes d'Arp4 qui interagissent avec la HSA sont affichés en gris. (c) Comme pour b les résidus d'Act1 importants pour les interactions hydrophobes sont affichés en gris. (d) La boucle située après le C-terminal de la HSA d'Eaf1 entoure le domaine D1 d'Act1 et continue vers le domaine D2.

Nous avons pu également observer une nouvelle zone d'interaction entre Eaf1 et Act1. Après la fin C-terminal de la HSA d'Eaf1, la protéine continue et étreint le lobe D1 d'Act1 avant de se diriger vers son lobe D2.



**Figure 38 : Interaction entre Eaf1 et Tra1**

(a) modèle atomique de NuA4, les régions d'Eaf1 qui interagissent avec Tra1 sont affichées en sphères colorées en bleu et nommées. (b) L'hélice 167-175 d'Eaf1 est située entre le domaine FRB et Kinase de Tra. (c) Site d'interaction de la boucle 239-245 réalisant des liaisons hydrogènes avec Tra1 qui stabilisent l'incorporation du lobe A.(d) la boucle 513-528 d'Eaf1 vient interagir avec les domaines HEAT de Tra1 par des liaisons hydrogènes et des interactions hydrophobes réalisées par les résidus colorés en gris de Tra1. (e) Représente le potentiel électrostatique du SANT domaine et de l'hélice 660-681 d'eaf1. (f) Représente le potentiel électrostatique de Tra1(le bleu indique un potentiel positif, le rouge un potentiel négatif). Le domaine SANT de Eaf1 est entouré de pointillés orange.

## (2) Interaction entre Eaf1 et Tra1

Bien que la majeure partie du N-terminal d'Eaf1 est présente dans le lobe A de NuA4, la première hélice d'Eaf1 que nous voyons dans la structure est située sur Tra1 **Figure 38 a**. L'hélice 168-173 d'Eaf1 vient s'intercaler entre l'hélice 3385-3404 du domaine FRB et l'hélice 3562-3573 du domaine kinase avant de rejoindre le lobe A **Figure 38 b**.

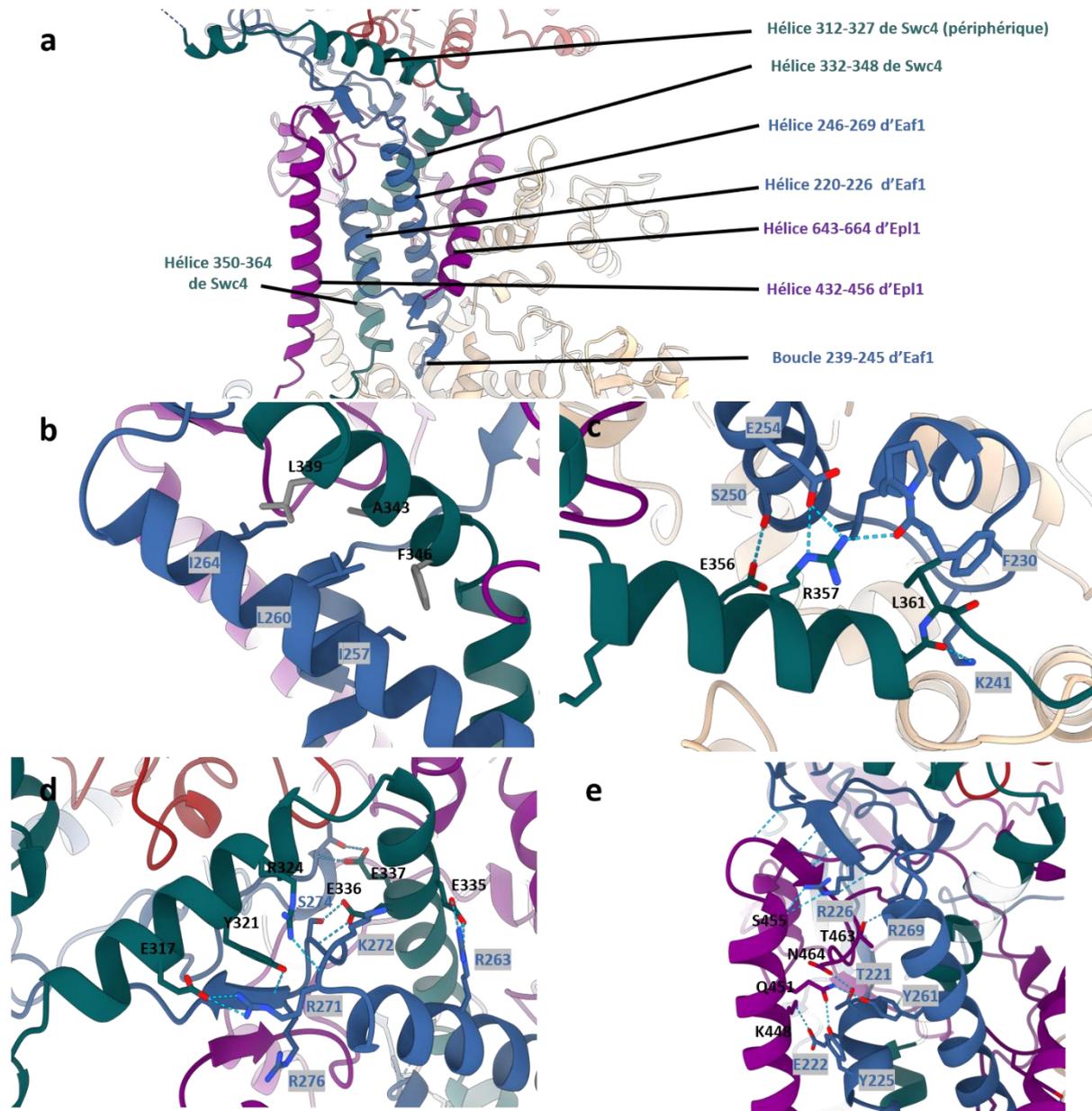
Après son entrée dans le lobe A Eaf1 va contacter Tra1 par le biais de la boucle 239-245 **Figure 38 a**. Ce contact est essentiel car il permet d'ancrer le lobe A sur Tra1 par l'intermédiaire de nombreuses liaisons hydrogènes. La S239 d'Eaf1 établit une liaison hydrogène avec Q3799 de Tra1, K241 deux avec D2691 et A22685, S242 deux liaisons avec E3719 et R2683, T245 une seule avec N3710 et pour finir Y258 avec S3806 **Figure 38 c**. Un peu plus en aval, K256 réalise une liaison hydrogène avec E3590 de Tra1. Après ces interactions Eaf1 se dirige vers Act1 et Arp4, avant de redescendre vers le lobe B pour y faire ses dernières interactions avec Tra1.

Lors de l'entrée d'Eaf1 dans le lobe B les premières interactions avec Tra1 sont faites par la boucle 513-528 sur une partie des domaines HEAT. Cette boucle réalise tout d'abord des liaisons hydrogènes avec Tra1 **Figure 38 c**, puis par la suite les résidus L522 et F523 d'Eaf1 plongent dans une poche hydrophobe réalisée par F2404, F2422, L2423, I2440, L2441 et Y2455 de Tra1 **Figure 38 c**.

Après la boucle 513-528, Eaf1 forme un domaine SANT et une longue hélice alpha. Les interactions entre Tra1 et ces éléments se fait par des interactions électrostatiques sur le domaine FAT. La **Figure 38 e & f** montre que la surface d'interaction de Tra1 est fortement chargée négativement, alors que la surface d'interaction d'Eaf1 est chargée positivement.

A l'instar de l'hélice (168-173) la dernière hélice visible d'Eaf1 est située sur Tra1 et est très isolée.

(3) *Eaf1, Swc4 et Epl1 forment un cœur structural d'hélices alphas*



**Figure 39 : Réseau d'hélice formé par Eaf1, Swc4 et Epl1**

(a) Réseau d'hélices formé autour d'Eaf1 permettant le maintien structural du lobe A avec le lobe B. Les différentes hélices sont nommées suivant leurs débuts et leurs fins. (b) Résidus impliqués dans l'interaction hydrophobe des hélices 246-269 d'Eaf1 et 332-348 de Swc4. (c) Liaisons hydrogène formées entre l'hélices 350-364 de Swc4 et Eaf1. (d) Liaisons hydrogène formées par les hélices 312-327 et 332-348 de Swc4 et Eaf1. (e) Liaisons hydrogène entre les hélices 432-456 d'Epl1 et 220-226 d'Eaf1. Au-dessus de l'hélice 432-456 d'Epl1, Eaf1 et Epl1 interagissent pour former un petit feuillet beta

De part et d'autre de la boucle 239-245 **Figure 39 a**, Eaf1 forme des hélices alphas qui interagissent avec d'autres hélices formées par Epl1 et Swc4. Ensemble elles forment un réseau de connexions qui permet de faire la jonction structurale entre le lobe A et le lobe B **Figure 39 a**.

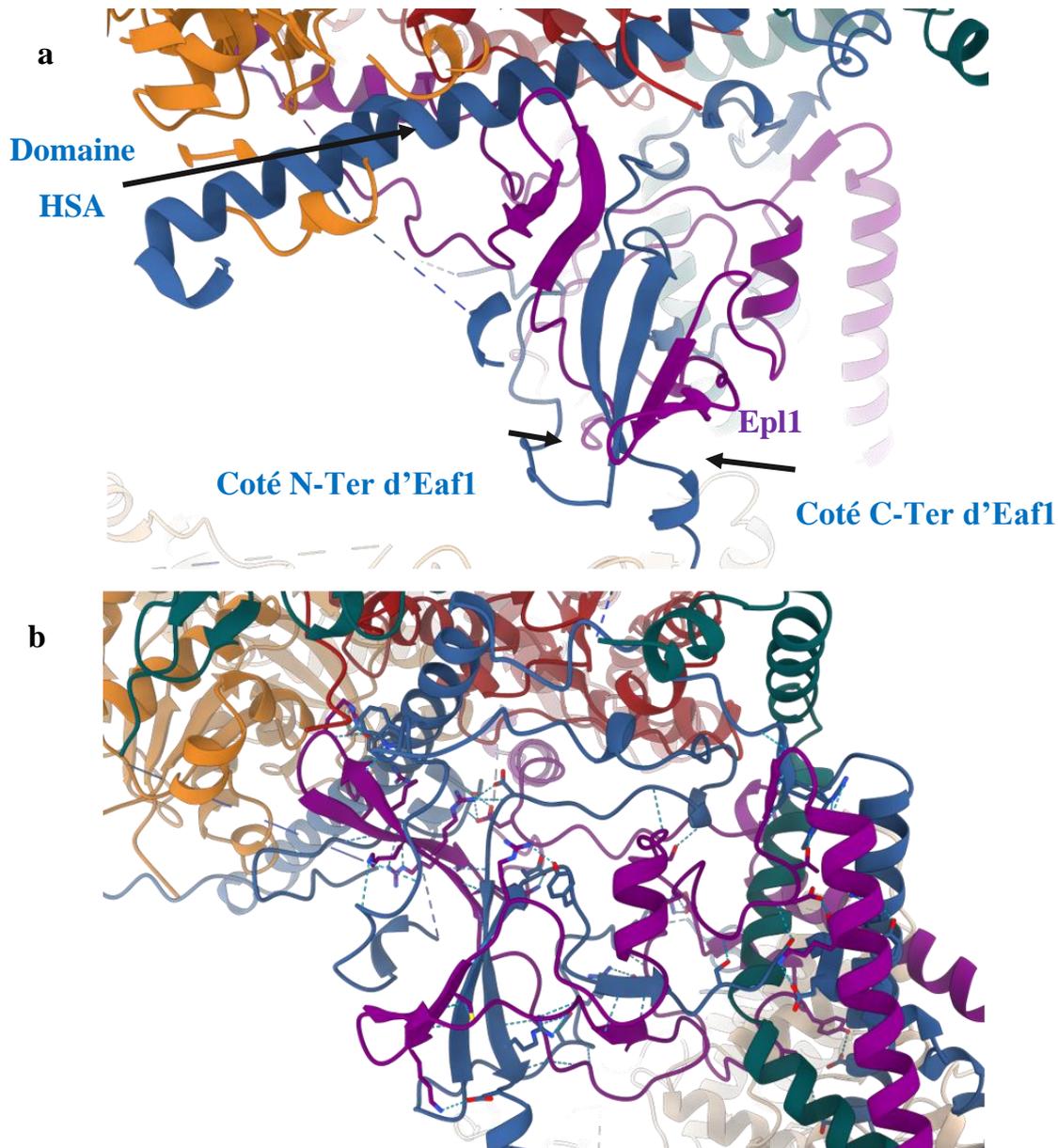
Les interactions dans ce réseau d'hélices sont majoritairement axées autour d'Eaf1 et composées de liaisons hydrogène et de quelques interactions hydrophobes. L'hélice 332-348 de Swc4 possède des résidus hydrophobes qui sont L339, A343 et F346 et qui viennent interagir avec les résidus I264, L260 et I257 de l'hélice 246-269 d'Eaf1 **Figure 39 b**. Swc4 effectue également des liaisons hydrogènes au cœur du réseau par son hélice 350-364 avec Eaf1 grâce à ses résidus E356, R357 et L363 et aux résidus E254, S250, F230 et K241 **Figure 39 c**. Plus en périphérie, l'hélice 312-327 de Swc4 effectue des liaisons hydrogène avec Eaf1 à l'aide de ses résidus E317, Y321, R324, E335, E336, E337 sur les résidus R271, R276, R271, R263, S274 et K272, T489 d'Eaf1 **Figure 39 d**.

Eaf1 et Swc4 sont entourés par deux hélices qui appartiennent à Epl1 (432-456 et 643-664) **Figure 39 a**. C'est l'hélice 220-226 d'Eaf1 qui permet de stabiliser l'hélice 432-456 d'Epl1 par le biais de liaisons hydrogène. De plus après l'hélice 432-456 d'Epl1, la protéine vient continuer à interagir avec Eaf1 en formant un petit feuillet beta entre les deux protéines **Figure 39 e**.

#### (4) Eaf1 et Epl1 des protéines essentielles pour l'intégrité du lobe A

Dans le lobe A Eaf1 réalise de nombreuses interactions avec Epl1 au point d'entremêler certaines de leurs structures secondaires. Le lobe A de NuA4 possède un grand feuillet beta formé par Epl1 et Eaf1 **Figure 40 a**. La partie N-terminal d'Eaf1 arrivant dans le lobe A et sa partie C-terminal le quittant forment un feuillet antiparallèle. De part et d'autre de ce petit feuillet formé par Eaf1, deux feuillets beta appartenant à Epl1 composés de deux brins chacun viennent se greffer. Il en résulte un feuillet beta antiparallèle issu de l'hybridation d'Eaf1 et Epl1. Ce feuillet occupe une

grande partie de la zone située sous le domaine HSA d'Eaf1 **Figure 40**. Le reste de la densité est composée d'autres petites structures secondaires et de boucles appartenant à Eaf1 ou à Epl1.

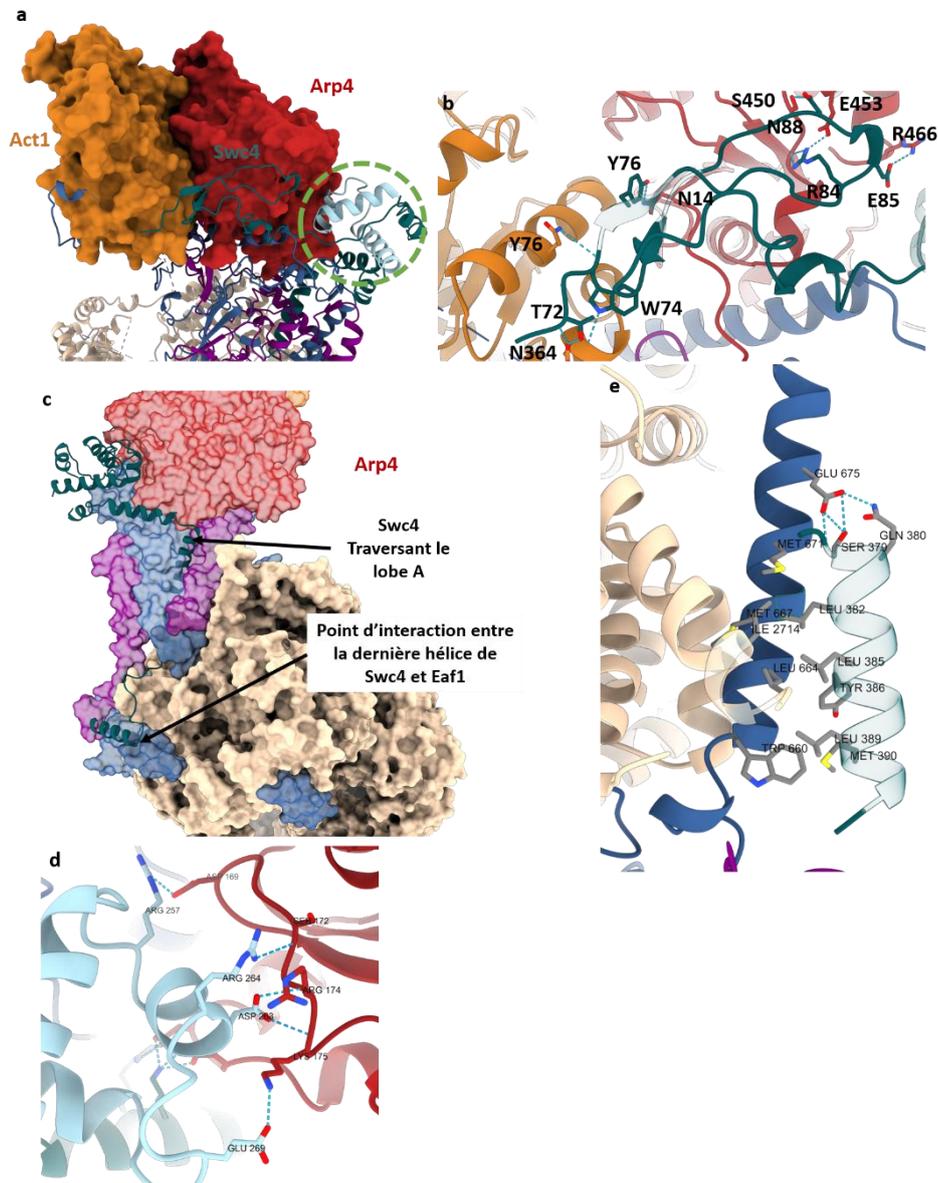


**Figure 40 : Interaction entre Eaf1 et Epl1**

(a) Feuille beta hybride formé par Eaf1 (bleu) et Epl1 (violet), en dessous de la HSA d'Eaf1. (b) Liaisons hydrogène formés entre Eaf1 et Epl1 dans le lobe A

L'affichage des liaisons hydrogènes entre Epl1 et Eaf1 montre à quel point elles interagissent entre elles et sont importantes pour la stabilité de cette région **Figure 40 b**. Quasiment tous les résidus d'Epl1 allant de 419 à 665 ont pu être construits et son haut niveau d'interaction avec Eaf1 montre que toute cette partie C-terminal d'Epl1 est

fortement ancrée dans NuA4. Or c'est Epl1 qui permet de faire le lien entre Piccolo et NuA4. La partie C-terminal d'Epl1 serait donc importante pour son interaction avec le cœur de NuA4 alors que la partie N-terminale serait essentielle pour



**Figure 41 : Interaction de Swc4 avec NuA4**

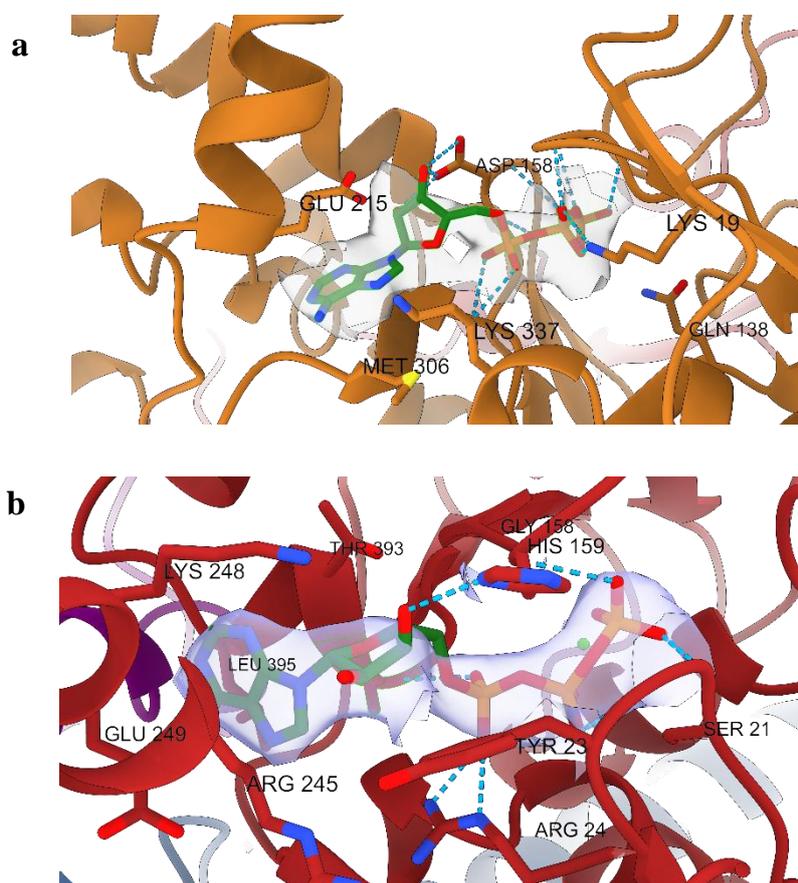
(a) Interaction entre Swc4 et Act1/Arp4, la partie N-terminale de Swc4 se pose contre la surface générée par Act1 et Arp4. Le SANT domaine de Swc4 est coloré en bleu clair et entouré de pointillés. (b) Détails des liaisons hydrogènes faites entre le N-ter de Swc4 et Act1/Arp4. (c) Les surfaces de NuA4 sont affichées pour permettre d'observer comment Swc4 traverse le lobe A de NuA4 avant de rejoindre le Lobe B où il vient s'accrocher sur une hélice d'Eaf1. (e) Détails des interactions hydrophobes entre la dernière hélice de Swc4 et Eaf1, ainsi que leurs liaisons hydrogènes.

l'activité/interaction de Piccolo.

## (5) Le module Actine

Pour rappel le module actine est composé d'Arp4, Actin1, Swc4 et Yaf9. Nous avons pu placer les structures atomiques de Arp4, Act1 et construire une bonne partie de Swc4, mais Yaf9 semble absent. Un chapitre lui sera dédié, de même que les autres éléments manquants de NuA4.

La partie N-terminale de Swc4 interagit avec Act1 et Arp4 en venant recouvrir leurs surfaces et en effectuant des liaisons hydrogènes **Figure 41 a & b**. En continuant vers le C-terminale de Swc4 on retrouve son domaine SANT qui est fixé sur Arp4 par des liaisons hydrogène **Figure 40 a & d**. R257, 4264, N263 et E269 interagissent avec les résidus N169, S172, R174 et K175 situés sur une boucle d'Arp4 **Figure 40 d**. Par la suite Swc4 traverse le lobe A de NuA4 pour aller se connecter à l'hélice 660-681 d'Eaf1,



**Figure 42 : Poches de liaisons à l'ATP**

(a) La figure représente l'ATP situé dans la poche de liaisons de Act1 avec ses résidus de contacts et ses liaisons hydrogène. (b) La figure représente également un ATP dans une poche de liaison, qui appartient cette fois-ci à Arp4.

elle-même fixée sur Tra1 **Figure 40 c**. L'interaction des deux hélices de Swc4 et celles d'Eaf1 est majoritairement faite par des résidus hydrophobes (L382, L385, Y386, L389, M390 avec W660, L664, M667, M671 ainsi que I2714 de Tra1) **Figure 40 e**. Des interactions hydrogènes sont également faites entre Q380, S279 de Swc4 et E675 d'Eaf1.

De manière similaire à la structure cristallographique de Arp4 Actin1 de Ino80, Act1 et Arp4 possèdent tous les deux un ATP dans leurs poches de liaisons **Figure 42**. Il est à noter que ces ATP sont co-purifiés avec notre complexe et ne sont pas ajoutés lors de la purification. Dans le cas d'Arp4 il semblerait que la protéine soit catalytiquement inactive (Gerhold et al., 2012), l'ATP serait donc recrutée par Arp4, mais jamais hydrolysé et resterait attaché à la protéine. Pour Actin1 peut être que son incorporation dans le complexe empêche l'hydrolyse de l'ATP, qui serait lui aussi piégé dans la structure.

Chez Ino80, Actin1/Arp4 et le domaine HSA interagissent avec l'ADN linker du NCP (Jungblut et al., 2020), Chez NuA4, la HSA est couverte par Epl1 et Eaf1 ce qui exclue toute interaction avec l'ADN. De plus, dans le cas de Ino80, Arp4/Act1 et HSA sont assez flexibles et peu définies dans les structures obtenues, ce qui est le contraire chez NuA4 où ils sont solidement fixés. Nous supposons que chez NuA4 la fonction de ces protéines/domaine est différente.

### C. Conclusions sur le modèle atomique

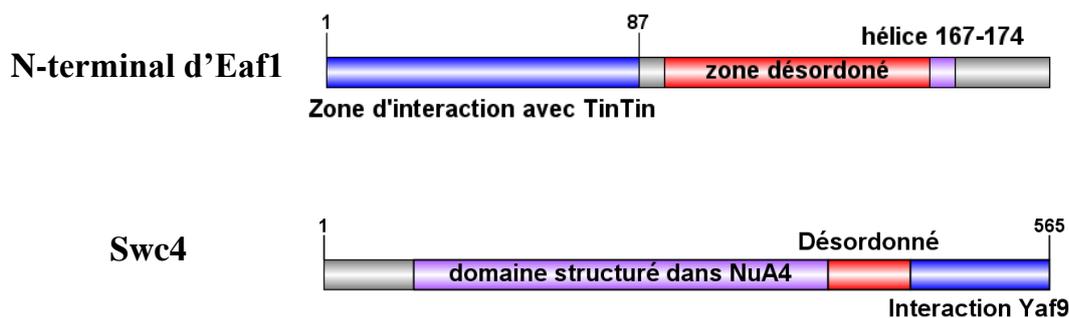
En résumé, en ce qui concerne le modèle atomique de NuA4, nous avons réussi à assigner tous les emplacements de la carte cryo-EM de manière non ambiguë grâce à nos cartes allant de 3.9 à 3.4 Å de résolution **Figure 35**. Il en résulte que le complexe est composé de deux lobes A et B. Le lobe A est constitué de Arp4, Act1, le N-terminale de Eaf1, Le C-terminale de Epl1 et Swc4 **Figure 35**. Le lobe B est majoritairement composé de Tra1, mais également du C-terminal d'Eaf1 et d'une hélice de Swc4 **Figure 36 & Figure 41 c**.

La majorité des interactions entre les sous-unités sont faites par Eaf1 qui est connue pour être responsable de la stabilité du complexe. Eaf1 est loin d'être une protéine très ordonnée, ainsi elle parcourt toute la structure en cheminant vers les différentes protéines du complexe **Figure 36**. Ainsi elle permet le maintien de l'intégrité du lobe A et sa fixation sur Tra1. Eaf1 et ses nombreuses interactions avec Epl1 faites par l'hybridation de leurs feuillets beta et les liaisons hydrogènes de leurs boucles permettent d'ancrer solidement le C-terminale d'EPL1 dans NuA4 **Figure 40**, montrant qu'il est essentiel pour le maintien de Piccolo dans le complexe, mais également pour l'intégrité du lobe A. Le domaine HSA d'Eaf1 est également crucial pour l'interaction entre Act1 et Arp4 **Figure 37**.

Chez la levure *Pichia pastoris*, NuA4 est censé être composé de 12 sous-unités, mais notre modèle a plusieurs protéines/modules absent. Le module TinTin est complètement absent de notre structure, de même que Yaf9 alors que le reste des sous-unités du module actine sont présentes. Pour Piccolo bien qu'on ait pu construire quasiment toute la partie C-terminale d'Epl1, le reste des sous-unités de Piccolo sont absentes. Notre hypothèse pour expliquer le manque de toutes ses sous-unités est qu'elles sont toutes trop flexibles pour pouvoir obtenir une structure et que ce n'est pas due à une perte des sous-unités lors de la purification. En effet, les gels SDS obtenus montrent que toutes les sous-unités de NuA4 sont présentes (voir **Figure 32** en début de chapitre) et l'analyse par spectrométrie de masse vient corroborer ces résultats et indique que la stœchiométrie des protéines de NuA4 est respectée et qu'aucune sous-unité n'est présente en faible quantité.

## 1. Flexibilité des modules d'interaction avec la chromatine

Les protéines/modules manquants dans notre structure sont tous impliqués dans l'interaction avec la chromatine. Le bromo domaine situé dans TinTin est capable de reconnaître l'acétylation des histones, le domaine YEATS de Yaf9 reconnaît les histones méthylés et Piccolo est le module réalisant l'acétylation des histones. Pour essayer de comprendre leur absence, nous avons dans un premier temps effectué des prédictions des séquences désordonnées des protéines impliquées dans l'interaction des protéines manquantes dans NuA4.



**Figure 43 : Régions désordonnées de Swc4 et du N-ter d'Eaf1**

Les zones d'interactions entre les modules absents de notre structure sont coloriées en bleu, les zones structurées sont coloriées en mauve, les régions désordonnées sont coloriées en rouge.

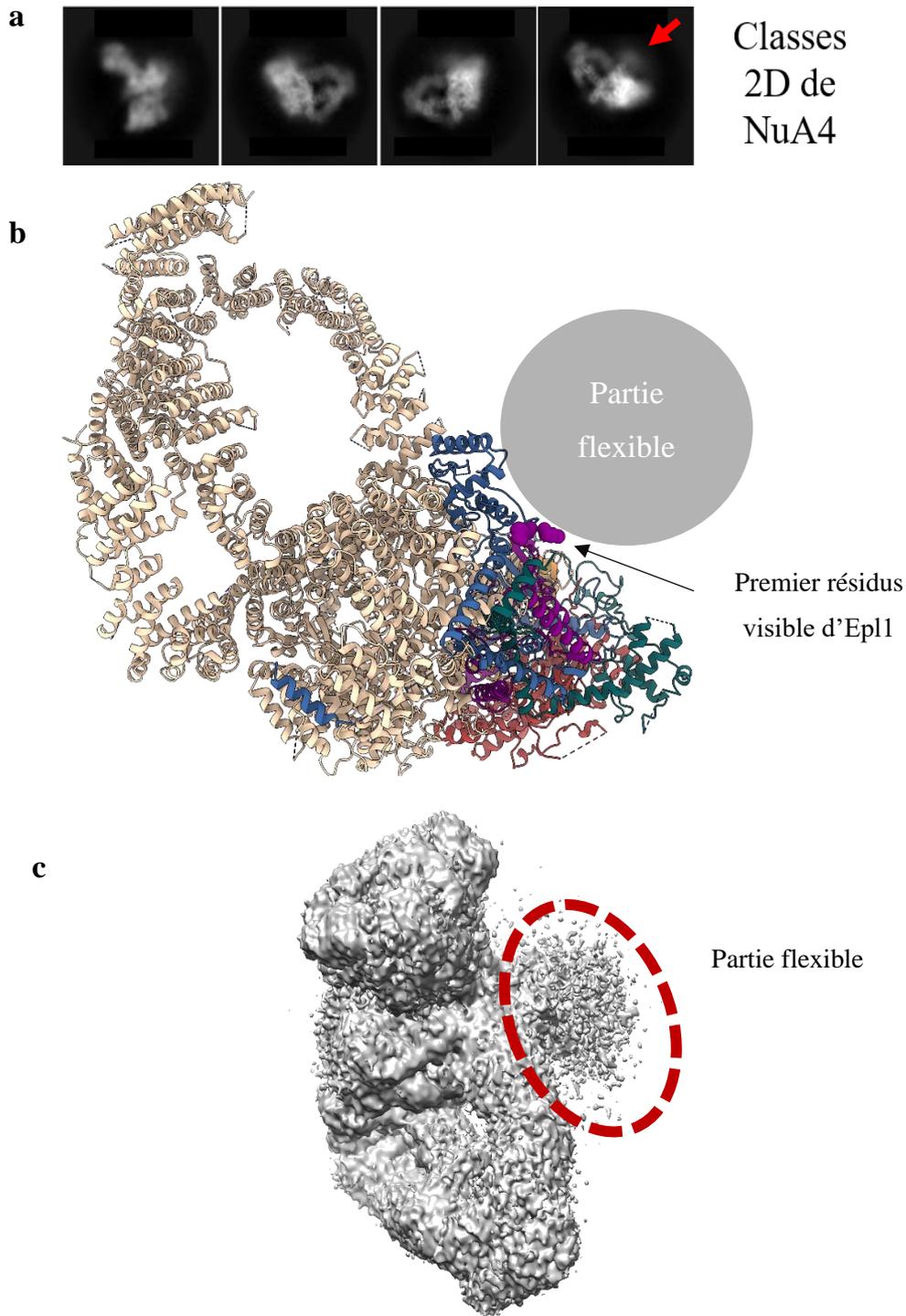
Dans le cas du module TinTin, des expériences biochimiques (X. Wang et al., 2018) montrent qu'il est incorporé dans NuA4 grâce à son interaction avec les 87 premiers acides aminés d'Eaf1. Or d'après la prédiction des régions désordonnées, toute la région entre les 87 premiers acides aminés et l'hélice 167-174, qui est le premier élément structural visible d'Eaf1 dans notre structure, est prédite comme désordonnée **Figure 43**. Ceci appuie l'idée que TinTin est bien présent dans notre structure, mais le lien l'intégrant au complexe est désordonné, créant une grande flexibilité qui rend impossible sa visualisation.

Il en va de même pour l'interaction Swc4 et Yaf9. Ce sont les 100 derniers acides aminés de Swc4 qui interagissent avec les 44 derniers acides aminés d'Yaf9 et permettent d'attacher Yaf9 à NuA4. Or comme pour TinTin, la zone précédant le site

d'interaction d'Yaf9 et prédite comme désordonnée **Figure 43**. La flexibilité induite par la boucle rend également impossible toute visualisation d'Yaf9.

(1) *Le module catalytique Piccolo*

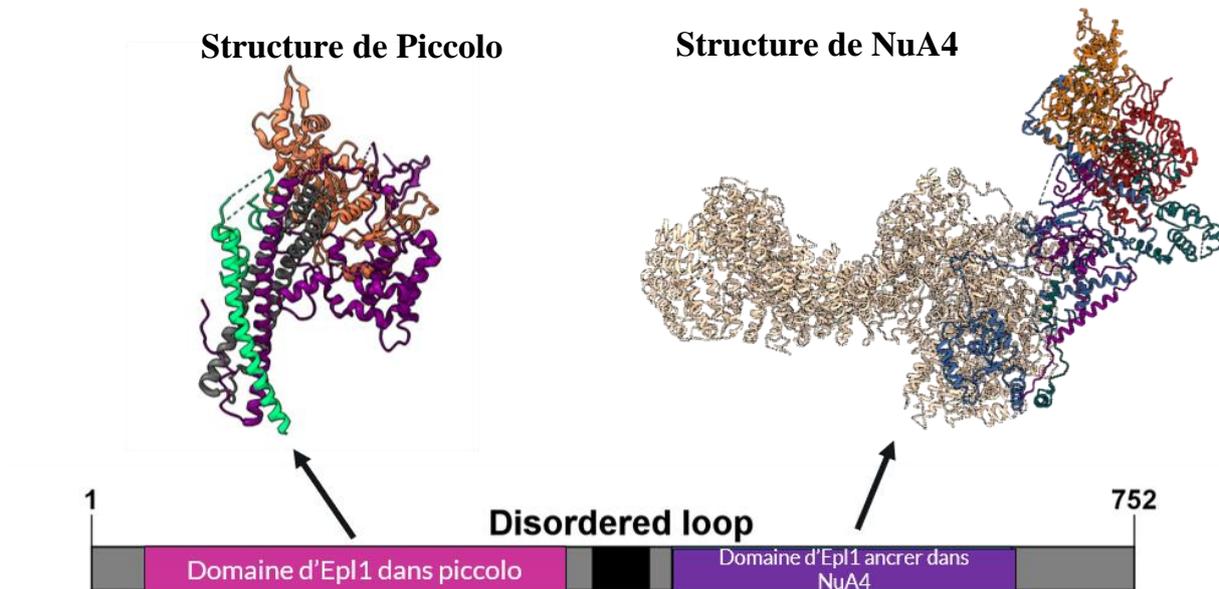
Le cas de Piccolo est également similaire à celui d'Yaf9 et de TinTin. Dans le chapitre précédent nous avons expliqué que toute une partie du C-terminal d'Ep11 a pu être construite et qu'elle est responsable de l'incorporation de Piccolo dans NuA4. Cependant la structure de Piccolo est connue, mais il est impossible de la placer dans notre carte. De la même manière que Yaf9 et TinTin nous avons effectué des prédictions des séquences désordonnées pour vérifier s'il existe également des régions désordonnées qui pourraient rendre le module catalytique flexible. De plus l'idée de flexibilité du module HAT est appuyée par nos classifications 2D, les images moyennes de classes montrent une densité floue sur la région périphérique de NuA4 juste à côté de l'hélice 432-456 d'Ep11 **Figure 44 a & b**, la zone où l'on commence à voir Ep11 dans notre structure.



### Figure 44 : Flexibilité du module Piccolo

(a) Sur les images de classes 2D de NuA4 on peut voir une zone floue indiquée par une flèche rouge. (b) Modèle atomique de NuA4 dans la même orientation que la classe 2D où la zone flexible est visible. Un cercle gris indique la position de la zone flexible qui serait juste à côté des premiers résidus visible d'Ep11. (c) Structure Cryo-EM de NuA4, le seuil du volume est baissé pour laisser apparaître le bruit dû à la partie flexible

Ce flou traduit un fort mouvement présent dans cette zone et devrait être le module catalytique. Cette région est également fortement bruitée dans la reconstruction 3D ce qui est caractéristique des régions fortement flexibles en cryo-EM **Figure 44 c**.

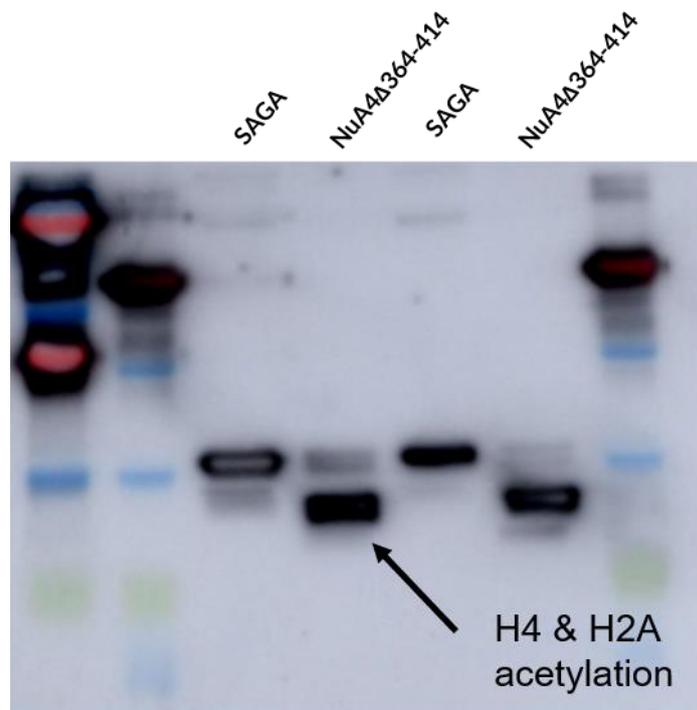


**Figure 45 : Région désordonnée d'Epl1**

La région en rose correspond à la partie structurée d'Epl1 dans Piccolo, en violet celle présente dans NuA4. On peut voir en noir qu'une section désordonnée relie les deux domaines structurés entre eux.

L'analyse des séquences nous révèle qu'il existe une boucle désordonnée de 56 acides aminés entre la partie C-terminale d'Epl1 visible dans notre carte et la partie N-terminale d'Epl1 présente dans la structure de Piccolo **Figure 45**. Cette boucle est responsable de la flexibilité du module HAT. Après avoir identifié cette région désordonnée, nous avons voulu réduire sa taille afin de stabiliser le module catalytique dans notre structure. Pour ce faire un mutant qui possède une délétion des résidus 364-414 d'Epl1 a été obtenu. Le Dr Adam Ben-Shem responsable de la mutation et purification du complexe a remarqué que les mutants ont plus de mal à se développer en culture, montrant qu'il y a un impact sur le phénotype des levures. Pour vérifier si la délétion de la région désordonnée empêche la capacité d'acétylation de NuA4, nous avons fait des tests d'activités.

Les résultats montrent que NuA4 qui possède un Epl1 mutant est toujours capable d'acétyler fortement le nucléosome *in vitro* **Figure 46**. Notre hypothèse est que



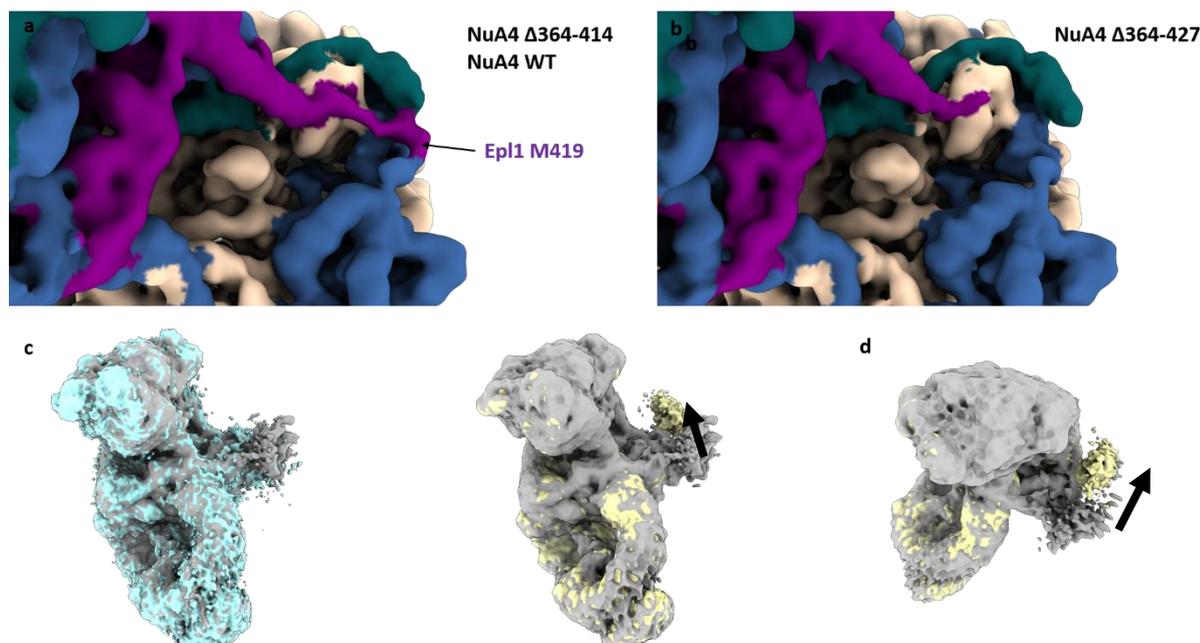
**Figure 46 : Test d'activité du NuA4 mutant**

Test d'activité montrant que notre mutant est toujours capable d'acétyler le NCP, le complexe SAGA est utilisé comme contrôle positif. Les colonnes de gauche contenant SAGA et NuA4 ont été incubés pendant 10 min à T° ambiante, celle de droite à 4°C pendant 30 min. La flèche indique la bande correspondante aux histones H4 et H2A

la délétion de cette boucle n'empêche pas NuA4 d'acétyler le NCP, mais va réduire la flexibilité du module catalytique, qui aura du mal à trouver/acétyler le NCP dans un contexte chromatinien. Les travaux de Marissa Vignali montrent que quand NuA4 est recruté par le biais de l'activateur Gal4-VP16 sur un site spécifique, il est capable d'acétyler les NCP sur une distance élevée (Vignali et al., 2000). La flexibilité du module catalytique est peut-être à l'origine de cette habilité d'acétylation à longue portée de NuA4.

## (2) Structure Cryo-EM des mutants de NuA4

Pour réduire la flexibilité du modèle catalytique deux mutants de NuA4 ayant chacun une délétion différente de la région désordonnée ont pu être analysés en Cryo-



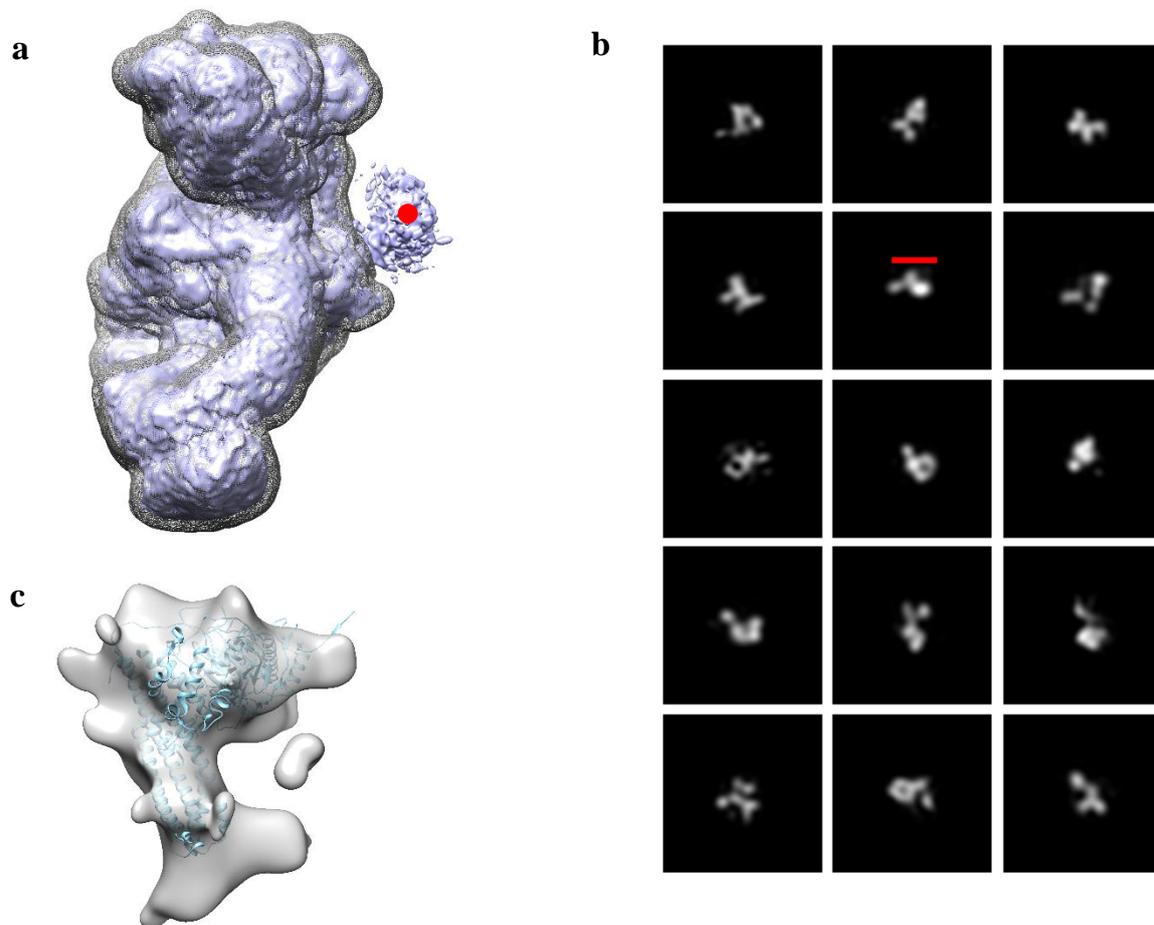
**Figure 47 : Analyse des cartes Cryo-EM NuA4 WT,  $\Delta 364-414$  et  $\Delta 364-427$**

Carte Cryo-EM colorée suivant les sous-unités de NuA4. La délétion 364-414 ne se voit pas sur la carte dans le cas du mutant  $\Delta 364-414$  car elle se situe en amont des résidus visibles sur notre structure. Les cartes du mutant et du WT sont identiques à ce niveau. **(b)** Ici la délétion  $\Delta 364-427$  est visible car elle touche à des résidus en aval des premiers résidus visibles d'Epl1 sur la structure. On voit clairement le manque de densité par rapport à **(a)**, le volume de **(a)** et **(b)** est placé au même seuil pour éviter toute mauvaise interprétation. **(c)** Comparaison du bruit correspondant à Piccolo, les cartes de NuA4 WT (bleu clair) et SL-1 (gris) sont superposées et on constate que le bruit est toujours dans la même position. Les cartes SL-1 (gris) et SL-2 (jaune) sont superposées, on constate que dans le cas de SL-2 le bruit à cette fois-ci a été déplacé dans le sens indiqué par la flèche. **(d)** Autre angle d'observation du déplacement du bruit entre SL-1 et SL-2.

EM. Le premier mutant appelé SL-1 pour Short Loop 1 possède la délétion d'Epl1 la moins longue  $\Delta 364-414$ . Le deuxième mutant SL-2 possède une délétion d'Epl1 plus importante  $\Delta 364-427$ , allant même retirer certains résidus qui interagissent avec Tra1 et qui sont visibles dans notre structure. Le mutant SL-1 a pu être résolu à 3.7 Å de moyenne, la délétion de la boucle n'a pas été suffisante pour stabiliser le module catalytique. En ce qui concerne SL-2 les résultats sont similaires : le module catalytique reste trop flexible, même après des classifications 3D avec un masque autour de la région d'intérêt. Cependant, dans SL-2 la délétion de la boucle contient des résidus qui étaient visibles sur notre structure WT et SL-1 et ne le sont plus maintenant **Figure 47**

**a & b.** Ceci montre de manière claire que cette densité correspond bien à Ep11. De plus en analysant le bruit sur la périphérie de NuA4, on peut constater que par rapport à SL-1 et WT, le bruit a été déplacé montrant que notre délétion a belle et bien un effet sur la flexibilité de Piccolo **Figure 47 c & d.**

Pour tenter de reconstruire Piccolo, nous avons soustrait la densité de NuA4 sur chaque image et centré sur la région bruitée en périphérie de NuA4. Des étapes de classification 2D ont ensuite été faites et montrent clairement un objet de la taille



**Figure 48 : Soustraction de la densité de NuA4**

(a) Masque et carte utilisés pour la soustraction de la densité de NuA4, le point rouge indique le nouveau centre de la boîte après la soustraction. En théorie seulement le signal de Piccolo devrait rester (b) Classes 2D après soustraction de la densité correspondant à NuA4. La barre rouge représente 130 Å ce qui correspond à la taille de Piccolo. (c) Reconstruction faite à partir des images soustraites, l'objet 3D a la taille de Piccolo et sa structure cristallographique peut y être placée.

approximative de Piccolo (~130 Å). Nous avons par la suite tenté d'établir une structure 3D de l'objet repéré en utilisant cryoSPARC pour générer des modèles 3D à partir des images soustraites. Grâce à cette méthode nous avons pu reconstruire cette carte où l'on peut clairement placer la structure cristallographique de Piccolo.

### (3) Résumé des mutants et de la flexibilité de NuA4

Grâce à l'analyse des séquences, notamment des régions désordonnées, nous avons pu observer que plusieurs constituants de NuA4 sont reliés au complexe par des régions désordonnées. Ces régions introduisent une forte flexibilité dans TINTIN, Yaf9 et Piccolo, qui sont tous les trois impliqués dans l'interaction avec la chromatine. Pour tenter de réduire la flexibilité de Piccolo nous avons fait une délétion afin de retirer la région désordonnée responsable de la flexibilité du module catalytique. Cette délétion a pour effet de ralentir la croissance des levures, ceci suppose que la flexibilité induite par cette région désordonnée est peut-être fonctionnellement importante. Cette hypothèse est en accord avec l'idée que NuA4 peut effectuer une acétylation à longue distance par rapport à son site de recrutement. Une diminution de la taille de la région désordonnée réduirait peut-être le champ d'acétylation de NuA4. Les tests d'activité montrent que NuA4 est toujours capable d'acétyler le NCP in vitro et que la mutation n'a pas affecté l'activité du complexe. Cette boucle serait donc importante in vivo où le contexte chromatinien rend peut-être difficile l'acétylation de NCP par un module catalytique peu flexible. L'analyse des images obtenues en Cryo-EM des mutants et du WT montre que la délétion de la région ne suffit pas à stabiliser le module catalytique, mais dans le cas du mutant SL-2 le bruit généré par la flexibilité de Piccolo a été déplacé, montrant que cette mutation influe sur la flexibilité de Piccolo. Nous avons également pu obtenir un modèle à basse résolution de Piccolo où l'on peut placer sa structure cristallographique en soustrayant des images la densité de NuA4.

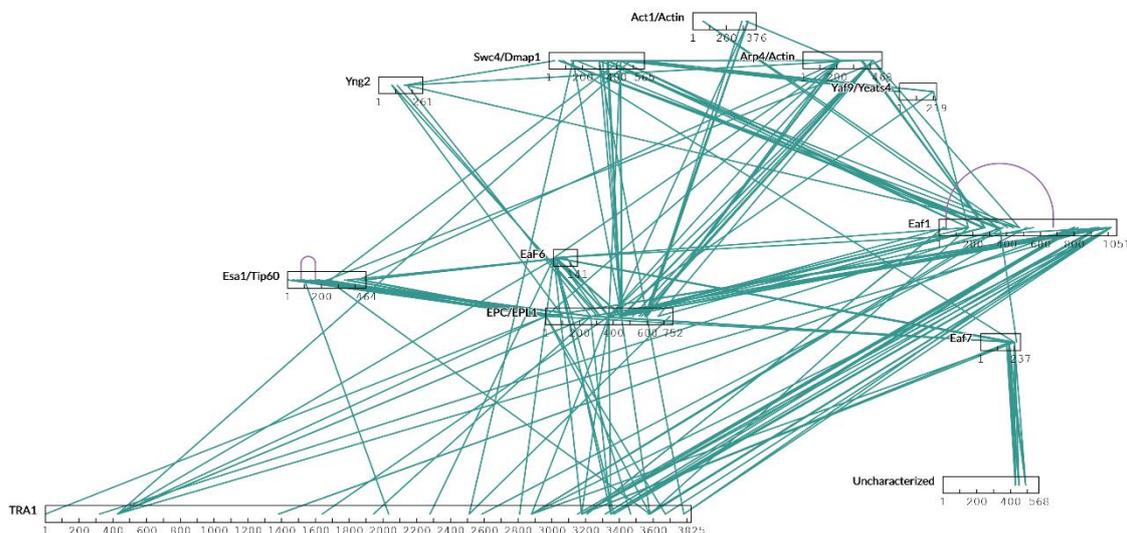
## 2. Réticulation de NuA4 et analyse spectrométrique

Lors de la thèse, nous avons utilisé des agents pontant chimiquement les lysines. Ce pontage permet deux choses. La première est la stabilisation de l'échantillon, qui

permet de maintenir l'intégrité structurale de NuA4 lors de la préparation de grille pour l'étude par Cryo-EM ou coloration négative. La deuxième est que l'analyse par spectrométrie de masse de complexe protéique réticulé permet d'étudier les interactions proches des protéines le constituant.

### a) Etude spectrométrique de NuA4 ponté par la BS3

Un échantillon de NuA4 fixé par la BS3 a été utilisé pour acquérir des données de spectrométrie de masse par la plateforme de spectrométrie de masse de l'institut dirigé par Luc Negroni. L'analyse des données générées a été faite en collaboration avec l'équipe de Noëlle Potier et exploitée par nous.

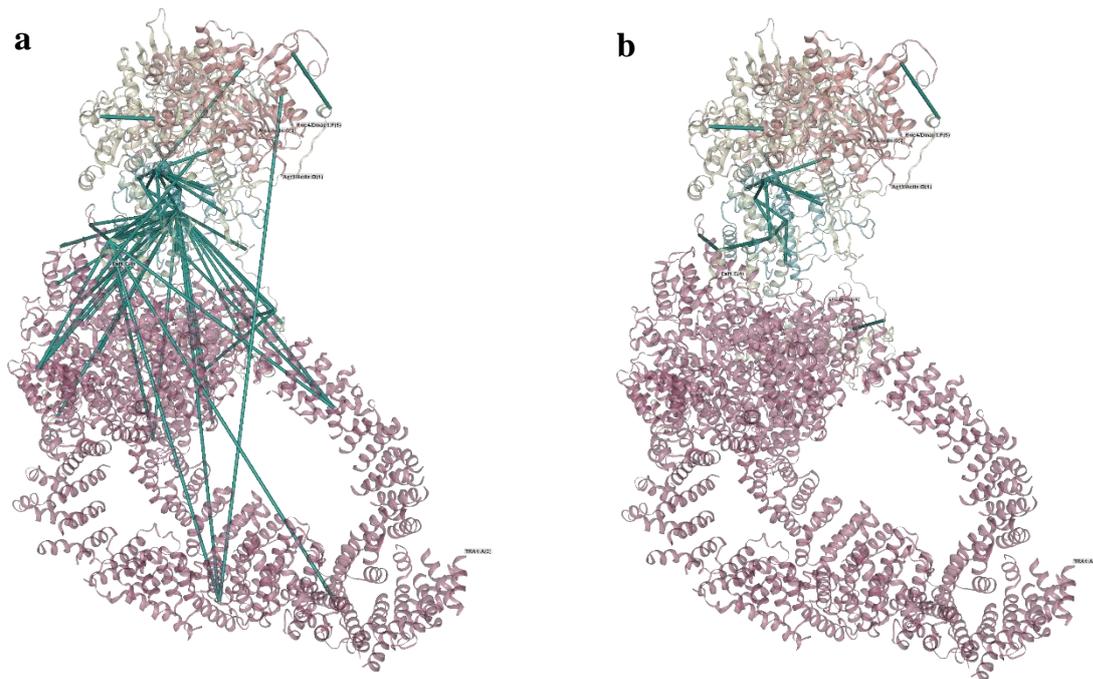


**Figure 49 : Analyse en spectrométrie de masse de NuA4 réticulé**

Les résultats montrent qu'un grand nombre de pontages chimiques entre les différentes lysines de NuA4 a pu être identifié **Figure 49**. Nous avons utilisé notre structure atomique de NuA4 afin de visualiser les pontages identifiés. L'idée était de découvrir des sites d'interactions potentiels entre des sous-unités de NuA4, notamment celles flexibles qui n'apparaissent pas dans la structure.

Lors de l'analyse des pontages chimiques avec le modèle atomique de NuA4, 54 sur les 488 pontages chimiques ont pu être visualisés. Les données montrent que sur ces 54 pontages seul 17 ont une taille inférieure à 30 Å, qui est la taille maximale que peut

avoir un pontage chimique à l'aide de la BS3. Identifier quelques pontages incohérents peut arriver lors du traitement des données à cause des erreurs d'identification des peptides, mais le fait que la majorité des pontages aient une taille incohérente exclue cette hypothèse et pointe plutôt vers un problème d'échantillon. Il est connu que dans certains cas la fixation d'échantillons par des molécules de pontages va déstabiliser le complexe au lieu de le stabiliser. Nous allons dans la prochaine partie vous présenter



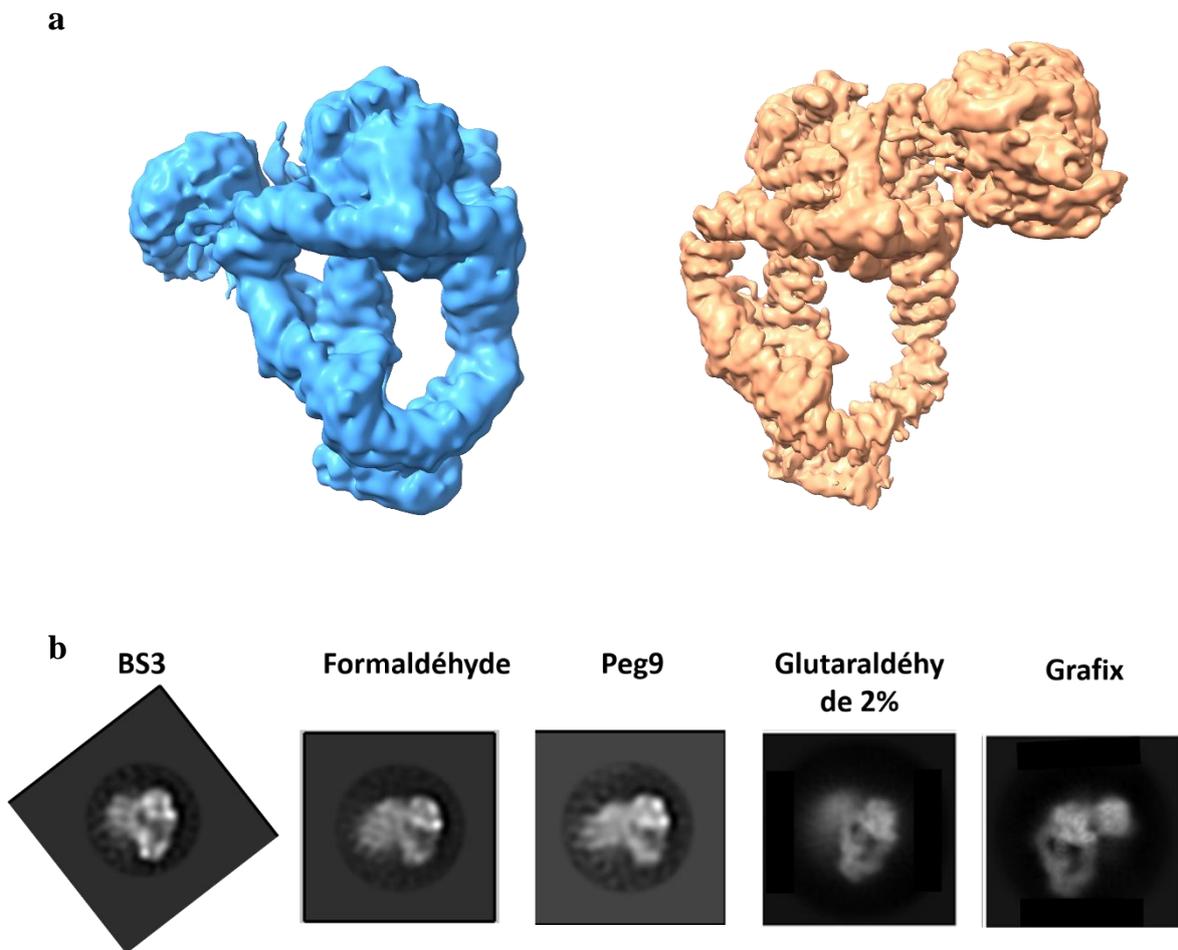
**Figure 50 : Analyse des pontages chimiques avec le modèle atomique de NuA4**

(a) Sont affichés sur la structure, tous les pontages chimiques détectés par spectrométrie de masse. (b) Seuls les pontages d'une taille inférieure à 30 Å sont affichés.

des analyses faites par Cryo-EM et coloration négative de NuA4, mise en présence de différents agents pontants et méthodes de fixations, pour voir leurs impacts sur l'intégrité du complexe.

## **b) Les agents de réticulation et NuA4**

Lors de l'étude de NuA4 en microscopie électronique, il est nécessaire de fixer le complexe. Dans le cas contraire, on observe sa dissociation et seule la protéine Tra1



**Figure 51 : Effet de différentes méthodes de fixation sur NuA4**

(a) La carte bleue est obtenue lors de la fixation de NuA4 par l'ajout directe de glutaraldéhyde à une concentration final de 0.2%, et montre que le lobe A perd son interaction avec Tra1, contrairement à NuA4 fixé à l'aide de la méthode Grafix (carte en orange). (b) Comparaison des classes 2D de NuA4 fixé par la BS3, le formaldéhyde et le peg9 en coloration négative avec des classes 2D de NuA4 obtenu en cryo sur des échantillons fixés directement au glutaraldéhyde à 0,2% et par méthode Grafix. On voit clairement un effet délétère sur le lobe A du complexe s'il n'est pas fixé par Grafix.

reste visible **Figure 51 a**. Cependant, nous avons observé des inconsistances dans la fixation de NuA4 par le glutaraldéhyde. Dans un premier temps, nous avons toujours fixé NuA4 par la méthode Grafix qui consiste à faire passer l'échantillon sur un gradient de sucrose et de glutaraldéhyde. Ceci permet de fixer en douceur le complexe et nous a permis d'obtenir nos structures à haute résolution. Dans un second temps pour éviter de consommer trop d'échantillons de NuA4, car le Grafix consomme toute une préparation. Nous avons tenté de réaliser la fixation d'une manière plus classique, c'est-

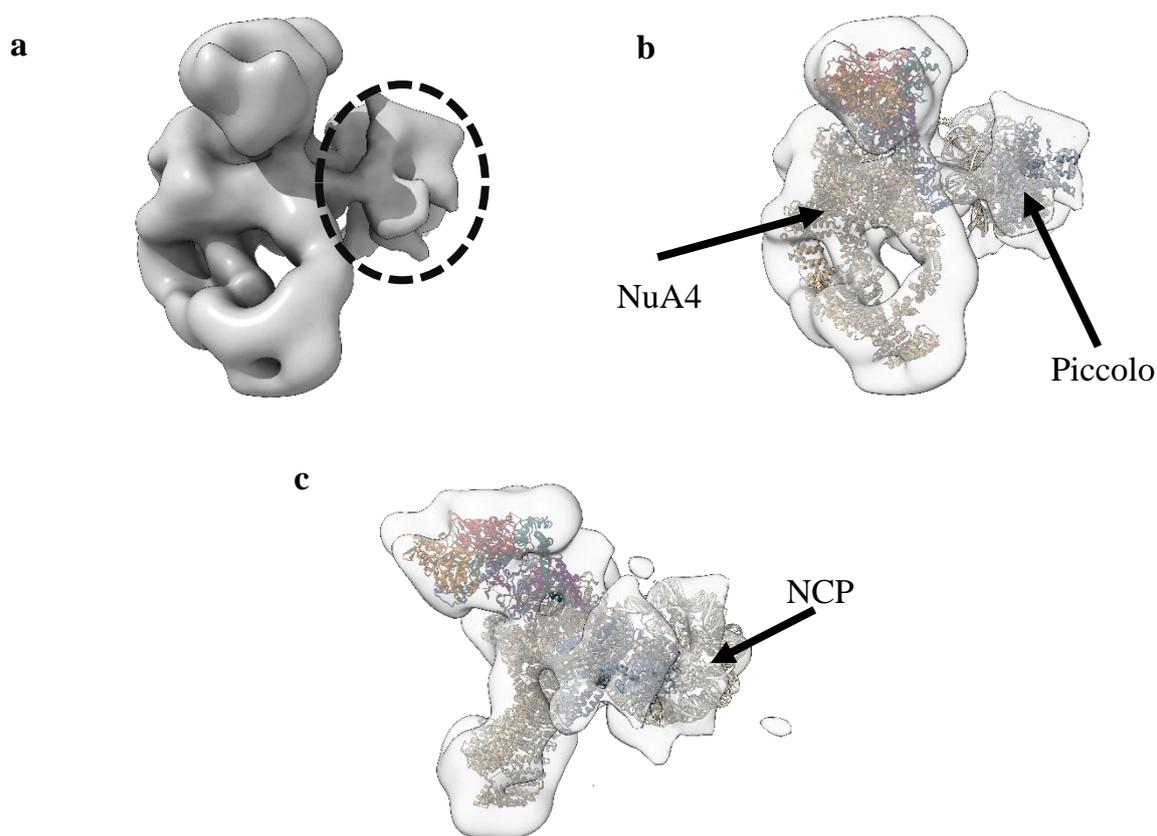
à-dire par l'ajout direct de glutaraldéhyde dans l'échantillon à une concentration finale de 0.2 %. Après analyse des images obtenues par cette nouvelle méthode de fixation nous observons que tout le lobe A de NuA4 a disparu, mais qu'un fort bruit apparaît à côté de Tra1 correspondant sûrement au lobe A désordonné, mais toujours maintenu sur le complexe. Dans les deux cas, la fixation de NuA4 est réalisée par les mêmes agents réticulant, la seule différence est le protocole de fixation employé.

Ayant remarqué ces résultats nous avons voulu vérifier si le complexe était également endommagé lors de la fixation à la BS3, ce qui pourrait expliquer les résultats incohérents de l'analyse par masse spectrométrie. Pour ce faire, nous avons réalisé des tests en coloration négative avec la BS3 à la même concentration que celle utilisée en spectrométrie de masse (3mM). D'autres agents réticulants ont également été testés pour vérifier si cela est spécifique ou non de la glutaraldéhyde/BS3. Les résultats montrent tous que les agents pontants induisent une perte de l'intégrité du complexe. En effet la comparaison des images 2D montre clairement que le lobe A perd son interaction avec Tra1, se désordonne et vient former un nuage flou sur la gauche de Tra1, contrairement à un NuA4 fixé par Grafix où même en classe 2D, on continue à voir le lobe A sur Tra1 **Figure 51 b**.

Ceci explique donc les incohérences observées lors de l'analyse par spectrométrie et montre que les résultats générés ne peuvent pas être utilisés pour l'étude des interactions des sous-unités de NuA4. Le fait que tous les agents réticulant induisent l'altération de NuA4, suggère que peut-être certaines lysines sont essentielles pour l'interaction avec Tra1. Lors de l'analyse de la structure atomique de NuA4, nous avons pu voir que la K241 et K256 d'Eaf1 interagissent avec Tra1. Dans le cas d'une fixation trop rapide de NuA4, la perte des interactions de ces deux lysines pourrait peut-être induire la dissociation du lobe A. Alors que lors d'une fixation plus « lente » comme dans le Grafix, d'autres lysines sont en premier lieu pontées stabilisant le complexe avant que les lysines 241 et 256 soit attaquées par le glutaraldéhyde.

### 3. NuA4 et le nucléosome en microscopie électronique

Nous avons tenté d'obtenir une structure de NuA4 avec le NCP, pour ce faire un test en coloration négative a été réalisé. Les grilles ont été préparées suivant le protocole



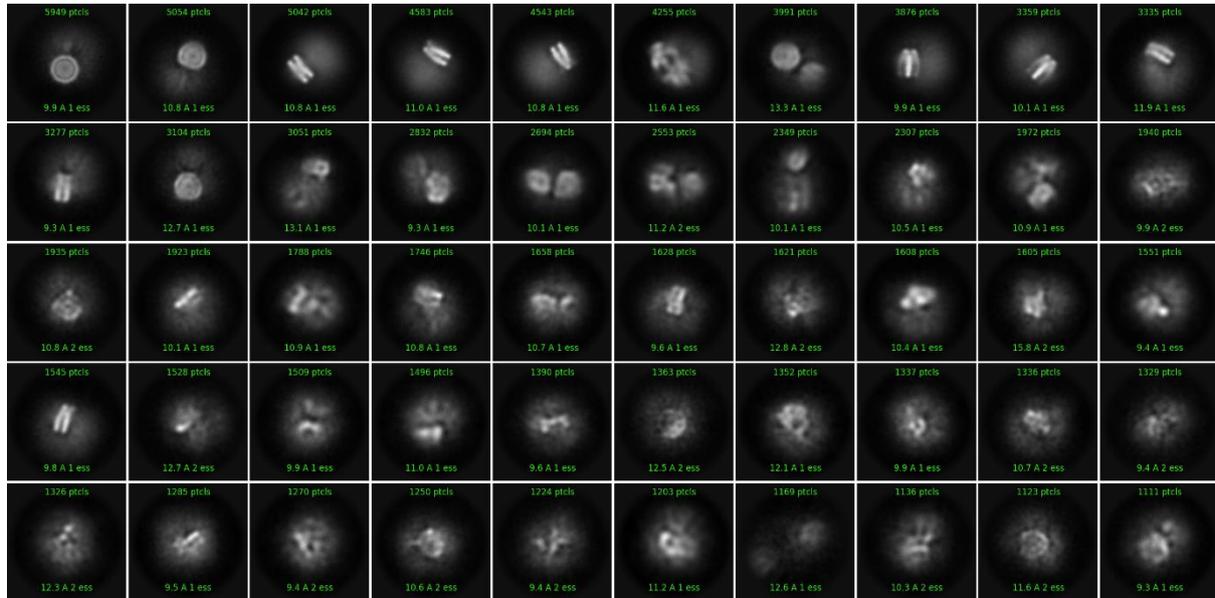
**Figure 52 : NuA4-NCP coloration négative**

(a) Structure à 25 Å de résolution de NuA4 obtenue par coloration négative. Les pointillés indiquent une densité supplémentaire par rapport à la structure de NuA4 sans le NCP (b & c) les modèles atomiques de NuA4, de Piccolo et du NCP sont placés dans la carte. Dans la densité supplémentaire il est possible de placer les modèles atomiques de Piccolo et du NCP.

indiqué dans la partie matériel & méthode. Une fois les premières étapes d'analyse faites, nous avons pu obtenir une carte à 25 Å de résolution **Figure 52 a**. NuA4 est bien visible et son modèle atomique peut être placé dans la carte **Figure 52 b & c**. Une forte densité située sur la région où est attendu Piccolo est présente et correspondrait au nucléosome lié au module catalytique. Ce premier résultat montre la faisabilité de

l'obtention d'une structure NuA4/NCP et est assez encourageant pour justifier le passage à la cryo-EM.

Nous avons donc tenté de collecter un jeu de données en cryo-EM. Les premiers résultats de classification 2D montrent un problème dans l'échantillon. Le NCP est bien visible et facilement identifiable, cependant la forme caractéristique de NuA4 n'est



**Figure 53 : Classes 2D de NuA4-NCP en Cryo-EM**

Classes 2D de NuA4 liée au NCP, la plupart des classes ne présentent aucun éléments structuraux de NuA4, cependant le NCP est bien visible avec un flou autour de lui.

pas présente et seule une zone floue autour du NCP est visible **Figure 53**. Il est clair que notre échantillon est endommagé, ce qui conduit à une dénaturation du complexe. Après discussion avec le Dr Adam Ben-Shem qui s'occupe de la purification de l'échantillon, nous avons réalisé que lors d'une étape de concentration le magnésium était 6 fois inférieur à celle utilisée normalement, ce qui expliquerait pourquoi le complexe a été dénaturé. De nouvelles expériences tenant compte de ce paramètre doivent donc être réalisées pour obtenir une structure Cryo-EM de NuA4 lié au NCP.

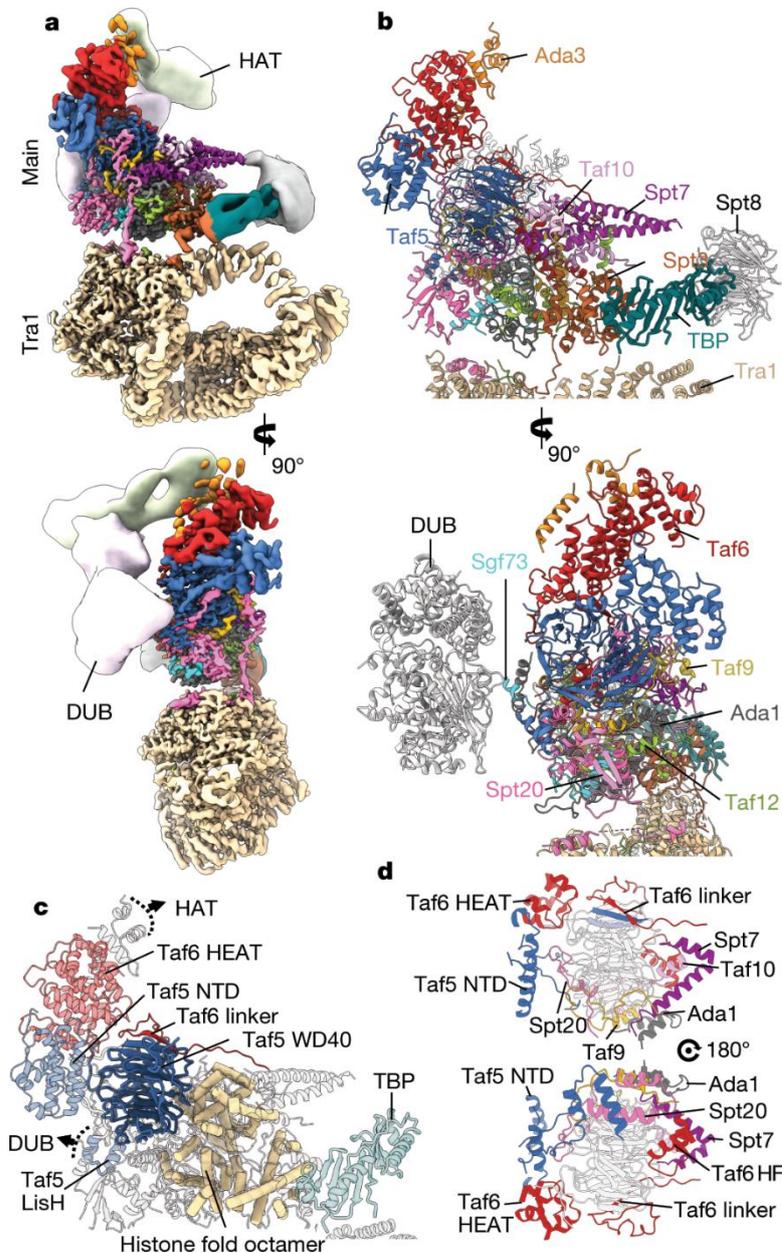
## D. SAGA

### 1. Résumé

La structure de SAGA se compose d'un lobe principal contenant les activités enzymatiques ainsi que la plupart des protéines composant le complexe et d'un autre lobe majoritairement composé de Tra1 et d'une partie de Taf12 et Spt20 **Figure 54**. Le module d'acétylation est hautement flexible engendrant une faible résolution de ce dernier, mais son site de liaison a pu être identifié. Ada3 vient ancrer le module HAT sur les domaines HEAT de Taf6, par l'intermédiaire de plusieurs hélices alpha. Le deuxième module catalytique est le DUB et est également flexible, mais son modèle atomique a pu être placé dans la densité. Le lien entre le DUB et SAGA est assuré par Sgf73 qui plonge dans la structure au niveau du domaine SEP de Spt20 et du domaine lys de Taf5.

La structure du lobe principal de SAGA est organisée autour du beta-propeller de Taf5 qui permet de connecter les différents modules fonctionnels entre eux. Dans le cas du module HAT, le beta-propeller de Taf5 vient encastrier le domaine N-terminal de Taf5 et Taf6, qui lui-même incorpore le module HAT par Ada3. De son côté le beta-propeller va contacter les protéines responsables de la formation d'un octamère d'histones déformé et impliqué dans l'interaction de SAGA et de TBP. Ceci montre le rôle central de ce domaine de Taf5, qui est la protéine organisatrice du lobe principal de SAGA.

De manière intéressante, l'interaction entre Tra1 et le lobe principal est faite par des boucles faisant intervenir 3 protéines : Spt3, Spt20 et Taf12. Spt3 effectue l'interaction la plus ponctuelle, au travers d'une boucle partant du lobe principal dont quelques résidus viennent contacter une poche hydrophobe de Tra1. Spt20 interagit avec Tra1 par le biais de deux hélices alphas. Taf12 forme la plus grande interaction avec Tra1, au travers d'une boucle parcourant sa surface et s'insinuant dans l'une des crevasses de la protéine.



**Figure 54 : Structure et modèle atomique de SAGA**

(a) Deux vues de la carte Cryo-EM composite du complexe SAGA, les cartes d'affinement local du lobe principal et de Tra1 ont été utilisés ainsi que des cartes issues de classification 3D filtré de manière à voir les autres éléments du complexe. (b) Modèle atomique correspondant au lobe principal vu en a. (c) Le domaine Taf5 WD40 organise structurellement le lobe principal. (d) Réseau complexe des contacts protéiques avec le domaine Taf5 WD40 (blanc)

Il a été proposé depuis plusieurs années la présence d'un octamère d'histones au sein des complexes SAGA et TFIID. Notre structure permet de proposer un modèle atomique de cette structure au sein de SAGA. Il s'avère que l'octamère est déformé par rapport à celui du NCP et offre à Spt3 une interface unique pour la liaison à TBP. La flexibilité engendrée par la déformation de l'octamère pourrait être importante pour permettre des changements conformationnels conduisant à la libération de TBP de Spt3. D'autres protéines de SAGA sont impliquées dans l'interaction avec TBP, notamment le domaine WD40 de Spt3.

En temps normal, quand TBP est incorporé dans SAGA l'encombrement stérique autour de TBP empêche toute liaison avec l'ADN. Or, le facteur de transcription TFIIA permet à TBP de se lier à l'ADN formant un trimère TFIIA-TBP-ADN qui est éjecté de SAGA. Cette éjection se fait probablement par des changements conformationnels impliquant Spt3, probablement rendus possibles par la flexibilité induite par l'octamère déformé.

## Article

# Structure of SAGA and mechanism of TBP deposition on gene promoters

<https://doi.org/10.1038/s41586-020-1944-2>

Received: 20 June 2019

Accepted: 19 November 2019

Published online: 22 January 2020

Gabor Papai<sup>1,2,3,4</sup>, Alexandre Frechard<sup>1,2,3,4</sup>, Olga Kolesnikova<sup>1,2,3,4</sup>, Corinne Crucifix<sup>1,2,3,4</sup>, Patrick Schultz<sup>1,2,3,4\*</sup> & Adam Ben-Shem<sup>1,2,3,4\*</sup>

SAGA (Spt–Ada–Gcn5–acetyltransferase) is a 19-subunit complex that stimulates transcription via two chromatin-modifying enzymatic modules and by delivering the TATA box binding protein (TBP) to nucleate the pre-initiation complex on DNA, a pivotal event in the expression of protein-encoding genes<sup>1</sup>. Here we present the structure of yeast SAGA with bound TBP. The core of the complex is resolved at 3.5 Å resolution (0.143 Fourier shell correlation). The structure reveals the intricate network of interactions that coordinate the different functional domains of SAGA and resolves an octamer of histone-fold domains at the core of SAGA. This deformed octamer deviates considerably from the symmetrical analogue in the nucleosome and is precisely tuned to establish a peripheral site for TBP, where steric hindrance represses binding of spurious DNA. Complementary biochemical analysis points to a mechanism for TBP delivery and release from SAGA that requires transcription factor IIA and whose efficiency correlates with the affinity of DNA to TBP. We provide the foundations for understanding the specific delivery of TBP to gene promoters and the multiple roles of SAGA in regulating gene expression.

Transcription of protein-encoding genes begins with the formation of a pre-initiation complex (PIC) comprising RNA polymerase II and several general transcription factors<sup>2</sup>. PIC assembly is nucleated by loading TBP onto promoter DNA<sup>3</sup>, a focal point for regulated gene expression<sup>4</sup>. Two multiprotein complexes, transcription factor IID (TFIID) and SAGA, can deliver TBP to gene promoters<sup>5,6</sup> and are required for global gene expression in yeast<sup>7,8</sup>.

The 1.6-MDa SAGA complex also stimulates transcription via its two chromatin-modifying enzymatic activities. The complex comprises 19 subunits organized in four modules with distinct functions<sup>2,9</sup>: a histone acetyltransferase (HAT) module, a histone deubiquitinase (DUB) module, the 430-kDa Tra1 subunit that serves as a docking platform for transcription factors that recruit SAGA to activating DNA sequences upstream of the promoter, and a 10-subunit central module that is physically connected to all other modules and is responsible for recruiting TBP to SAGA (Extended Data Fig. 1a).

Structures of several individual isolated components of SAGA, notably the DUB module<sup>10,11</sup>, the catalytic subunit of the HAT module (Gcn5)<sup>12</sup> and Tra1<sup>13,14</sup> have been obtained. Previous electron microscopy studies positioned Tra1 in one lobe of the SAGA complex, but all other subunits appeared together as a main lobe that could only be described at low resolution<sup>15,16</sup>.

The central module of yeast SAGA shares five TBP-associated factors (TAFs) with TFIID. Both complexes also contain multiple histone-fold domains that associate into heterodimers and higher-order oligomers<sup>17,18</sup> (Extended Data Fig. 1a, b). The central module is therefore expected to be structurally related to TFIID, suggesting that TBP binding in both complexes shares some common features. Progress in defining the

architecture of TFIID was reported recently; however, the machinery responsible for TBP binding and delivery remained poorly resolved<sup>19,20</sup>.

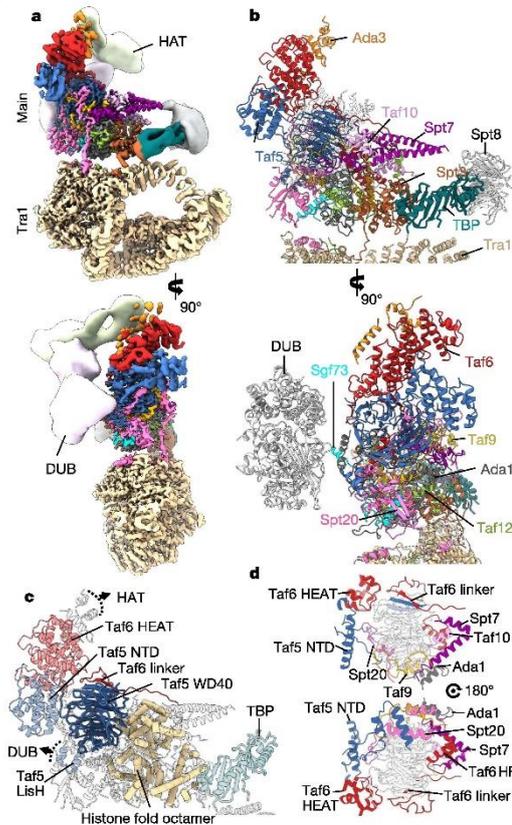
Here we report the 3.5 Å-resolution structure of the full SAGA complex bound to TBP, determined from cryo-electron microscopy (cryo-EM) maps that show clear density for the majority of the protein side chains. Our structure describes in atomic detail the network of interactions that coordinate the functional domains of SAGA as well as a machinery precisely geared to binding TBP and preventing aberrant transcription initiation. With complementary experimental findings, our structure points to a mechanism of TBP delivery and release from SAGA that depends on TFIIA and the affinity of promoter DNA to TBP, providing new insights into regulation of PIC assembly.

## Overall architecture of SAGA

We reconstituted the complex with SAGA purified from the yeast *Komagataella phaffii* (also known as *Pichia pastoris*), full length TBP and ubiquitin aldehyde and collected a dataset of more than one million particles. Image analysis resulted in a map of the full SAGA complex at 3.8 Å resolution, calculated from 338,850 selected particles (Fig. 1a, Extended Data Figs. 1c, d, 2, Supplementary Video 1, Extended Data Table 1). Focused refinement of the main lobe and of Tra1 further improved their average resolution to 3.5 Å, with many regions reaching as high as 3 Å. Hence, Tra1 and the central module, as well as the domains of Sgf73 that anchor the DUB to the central module, show clear densities for the majority of side chains (Fig. 1b, Extended Data Fig. 2b, c).

At the periphery of SAGA, exposed next to Spt3, we identified the distinctive saddle-shaped density of the conserved core of TBP (cTBP),

<sup>1</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire, Integrated Structural Biology Department, Equipe labellisée Ligue Contre le Cancer, Illkirch, France. <sup>2</sup>Centre National de la Recherche Scientifique, UMR7104, Illkirch, France. <sup>3</sup>Institut National de la Santé et de la Recherche Médicale, U1258, Illkirch, France. <sup>4</sup>Université de Strasbourg, Illkirch, France. \*e-mail: patrick.schultz@igbmc.fr; adam@igbmc.fr



**Fig. 1 | Structure of the yeast SAGA–TBP complex.** **a**, Front (top) and side (bottom) views of a composite cryo-EM reconstruction of the yeast SAGA–TBP complex. Maps from focused refinements of Tra1 (tan), DUB (light pink), Spt8 (white), the main lobe and TBP (blue–green) were combined. HAT density (light green) is taken from the low-pass-filtered map of the full complex. **b**, Corresponding views of the atomic model of the SAGA–TBP complex focusing on the main lobe. **c**, The Taf5 WD40 domain structurally orchestrates the main lobe. Orientation as in **a** (top). **d**, Intricate network of protein contacts with the Taf5 WD40 domain (white).

which was absent in a separate map of SAGA lacking TBP (Extended Data Fig. 3a–d). Local classification revealed sufficient secondary structural features to guide the precise positioning of cTBP. Site-specific cross-linking and documented mutations resolved any ambiguity in orientation arising from the pseudosymmetric fold of cTBP<sup>21,22</sup>. A disc-shaped density neighbouring TBP was assigned to the WD40 protein Spt8, which was shown to interact with TBP<sup>23</sup> (Extended Data Fig. 3e, f). At the opposite pole of the main lobe, the DUB crystal structure fits into and occupies a large but not well-resolved density that is nearly completely detached from and independent of SAGA. The DUB module is connected to the main lobe by a narrow, elongated density that contacts the surface of SAGA proximal to the Spt20 SEP domain and the LisH domain of Taf5 (Fig. 1a, b, Extended Data Fig. 1). The HAT enzymatic module adopts multiple conformations and orientations. Accordingly, the bulk of this module is visible only as an elongated featureless density at the tip of our map. Nevertheless, we could identify its docking site

on SAGA, where two helical domains—attributable to Ada3—lie at the surface of the Taf6 HEAT-repeats domain.

**Taf5 orchestrates the main lobe**

The main lobe is dominated by the seven-blade WD40 propeller of Taf5 (Fig. 1c). This domain occupies a central position, serving as a structural hub that connects all major SAGA functions. The flat top face of the propeller disc clamps together the Taf5 N-terminal domain (NTD) and the Taf6 HEAT repeats that bind the HAT module. The opposite bottom face of WD40 holds the histone-fold machinery that ultimately binds TBP and connects to the DUB module and Tra1.

The positions of the Taf5 NTD and Taf6 HEAT repeats are markedly different in SAGA compared with TFIID<sup>19</sup>, as these domains assume different roles in establishing connections to the enzymatic DUB (the LisH extension of Taf5 NTD) and HAT (Taf6) modules (Fig. 1c, Extended Data Fig. 4). Compared with TFIID, the Taf5 NTD in SAGA makes a rotation of almost 180 degrees around the WD40 domain, and the Taf6 HEAT repeats, which are remote in TFIID, are relocated to the vicinity of the top face of WD40 in SAGA. Notably, unlike SAGA, TFIID harbours two copies of Taf5 and Taf6 that are held together by homodimerization of Taf5 NTDs and Taf6 HEAT repeats. As a result of all these variations and additional interactions with SAGA-specific subunits, the Taf5 WD40 environment is very different in SAGA compared with TFIID.

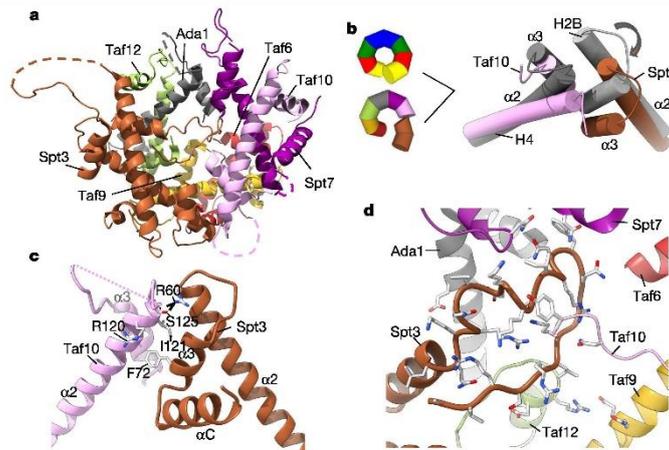
The ability of Taf6 HEAT repeats to occupy three different positions in SAGA and TFIID relative to the histone fold is afforded by a 136-residue linker between these two domains. The Taf6 linker partially encircles the Taf5 WD40 domain (Fig. 1c), and its position is stabilized by several interactions along its trajectory. Most notably, the Taf6 linker residues T135–L141 add a  $\beta$ -strand to the last blade of the WD40 propeller, forming an inter-protein  $\beta$ -sheet.

WD40 domains are recognized as potential hubs that mediate protein–protein interactions. The Taf5 WD40 in SAGA is a noteworthy demonstration of this capacity, as it associates with at least 11 different protein domains (Fig. 1c, d). Taf5 is thus a centrally located master hub that orchestrates the architecture of the entire main lobe.

**A deformed histone-fold octamer**

An octamer consisting of four histone-fold pairs arranged in a disc-shaped spiral, similar to the nucleosomal histone octamer, has been proposed to reside in TFIID and SAGA<sup>24,25</sup>. Our structure presents an atomic model of the full histone-fold octamer that resides at the core of SAGA and, most probably with minor variations, in the TBP-binding lobe of TFIID (Fig. 2a). It reveals a disparity between histone-fold interactions in SAGA compared with the nucleosome, particularly regarding the incorporation into SAGA of the TBP-binding subunit Spt3, a homologue of the TFIID Taf11–Taf13 histone-fold pair<sup>24</sup> (Supplementary Video 2).

In the nucleosome, the interaction between neighbouring histone pairs is mediated by a tightly packed four-helix bundle composed of helices  $\alpha$ 2 and  $\alpha$ 3 of one histone from each pair<sup>26</sup>. The SAGA histone-fold pairs are organized along the spiral in the following order: Taf6–Taf9, Taf12–Ada1, Spt7–Taf10 and Spt3, which contains two histone folds, one at each of its ends: nSpt3-HF and cSpt3-HF. Whereas all other histone-fold pairs are oriented as in the nucleosome, the Spt3 pair is tilted by 20 degrees compared with its analogous nucleosome H2A–H2B histone pair (Fig. 2b). This tilt prevents participation of the Taf10  $\alpha$ 3-helix in the interaction, shears the four-helix bundle and therefore nearly completely frees Spt3 from its association with the histone-fold octamer. The remaining contact is facilitated by a slight tilt in Taf10 and maintained by only very few bonds between Spt3 and Taf10 residues (Fig. 2c). To secure incorporation of Spt3 into SAGA, the C-terminal tail of Spt3 inserts into the centre of the spiral disc (a void volume in the nucleosome), makes a near complete circle and forms single bonds with each of the other histone folds in SAGA (Fig. 2d).



**Fig. 2 | The octamer of histone folds.** **a**, Structural organization of the four pairs of histone folds forming an octamer analogous to the nucleosomal histones. **b**, Left, schematic representation of the histone-fold octamer in the nucleosome (top) and in SAGA (bottom). Right, the four-helix bundle connecting the Spt7–Taf10 histone-fold pair (pink) to the partially dislodged

nSpt3–cSpt3 histone-fold pair (brown) compared with the analogous nucleosome interaction (grey). **c**, Detailed view of the interaction between the Spt7–Taf10 histone-fold pair and the nSpt3–cSpt3 histone-fold pair. **d**, The C-terminal tail of Spt3 inserts into the centre of the histone-fold octamer.

The Spt3 deviation from the canonical octamer does not originate from a single structural determinant. The atypical sequences of the histone folds (in particular Taf10 and Spt3), the long Taf10 loop between its  $\alpha 2$ - and  $\alpha 3$ -helices and the inherent asymmetry arising from the eight different histone-fold proteins involved all contribute. In this respect, it is noteworthy that the deviation from the nucleosome pattern is already apparent at the preceding four-helix bundle where the association almost entirely lacks hydrophobic packing (Extended Data Fig. 5). We propose that this assembly of histone folds, which deviates considerably from the symmetric arrangement in the nucleosome, is geared specifically towards a delicate balance between a rigid Spt3 that is firmly embedded into the octamer and an overly flexible Spt3 that associates with SAGA through loops alone. Below we examine these ideas in the context of TBP binding and release from SAGA.

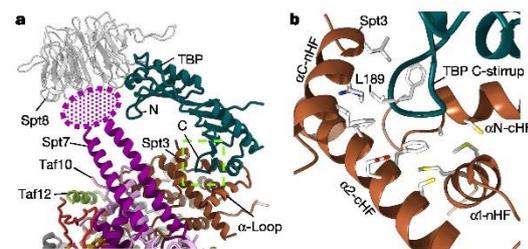
### TBP binding and release machinery

SAGA presents TBP at the extremity of the main lobe embracing it by binding to two opposite ends of its surface and orienting the DNA-binding cleft of TBP towards the main body of SAGA (Fig. 3a). The principal component in coupling TBP to SAGA and determining its position is Spt3. The C-terminal stirrup of TBP is the primary binding site for Spt3, similar to the interaction between TBP and TFIIIB<sup>27</sup>. In SAGA, the stirrup is nearly completely buried and forms multiple contacts in a large pocket created by helices  $\alpha N$  and  $\alpha 2$  from cSpt3-HF and by helix  $\alpha C$ , which is joined to nSpt3-HF via a conserved two-amino-acid rigid turn (Figs. 2c, 3b). The association of the C-terminal half of the pseudosymmetric cTBP with Spt3 is in line with previous biochemical and genetic data<sup>21,22</sup> (Extended Data Fig. 6). Indeed, accommodation of the cTBP N-terminal half by Spt3 is unfavourable, as this would involve the negatively charged D81 facing a negatively charged patch on Spt3 and place the bulky charged R98 where the C-terminal stirrup L189 is tightly packed against a hydrophobic crevice (Fig. 3b).

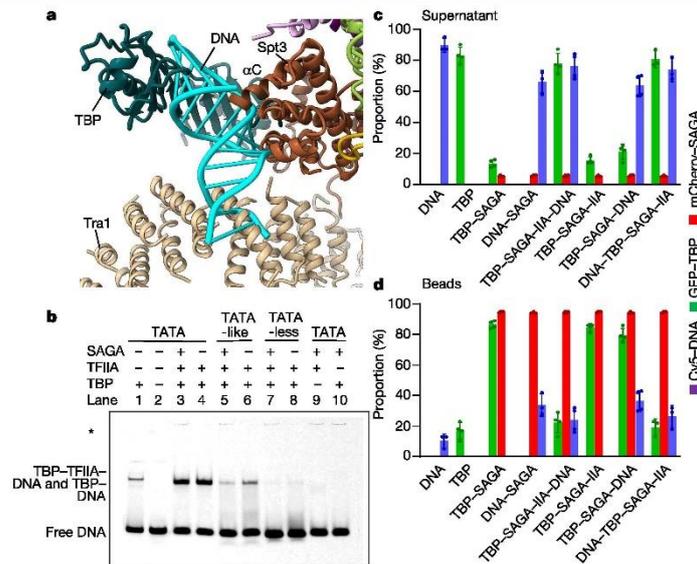
In contrast to Spt3, we find that Spt8 is not part of the central module (Fig. 3a). It is flexibly tethered via a fuzzy density to the outward tips of two long helices, extensions of each end of the Spt7 histone fold. The fuzzy density corresponds, at least in part, to a 30-residue-long

unstructured region in the C terminus of Spt7, the deletion of which results in loss of Spt8<sup>28</sup>. The connection of Spt8 to the N-terminal half of cTBP is flexible and appears to have a secondary role in positioning TBP. This finding is consistent with deletion mutants of Spt3 in yeast showing a sharp decline in global levels of newly synthesized mRNA, whereas impairment of Spt8 had almost no effect<sup>28</sup>. We suggest that Spt8 has an auxiliary role in tuning the orientation of TBP and in competing with other TBP-associated proteins<sup>29–31</sup>. The flexible tethering of Spt8 to the SAGA core contributes to small movements of TBP that are reflected in a lower local resolution. This ‘breathing’ of TBP might be important in transiently enabling initial access to DNA or regulatory factors.

The space between TBP and the main lobe of SAGA is just wide enough to accommodate a double-stranded DNA bent by TBP (Fig. 4a). However, the path of DNA is obstructed by structural elements located distally to the TBP DNA-binding cleft, notably the peripheral ring of the Tra1 HEAT repeats, as well as the nSpt3-HF helix  $\alpha C$ , a major component of the pocket that envelops the C-terminal TBP stirrup. Gel-shift assays (Fig. 4b, Extended Data Fig. 7b) and pull-down assays (Fig. 4c, d,



**Fig. 3 | Architecture of the TBP-docking site.** **a**, cTBP interacts with Spt3 mainly through the stirrup of its C-terminal half (labelled C), whereas its N-terminal half (N) binds to the Spt8 WD40 domain. A poorly resolved Spt7 domain (pink ellipse) connects Spt8 to the central module. **b**, Enlargement of the boxed area in **a**, depicting the interaction between the C-terminal stirrup of TBP and a pocket created by Spt3 helices.



**Fig. 4 | Steric hindrance for TBP binding to DNA and TFIIA-dependent TBP delivery.** **a**, DNA (light blue) was modelled into the DNA-binding cleft of TBP by superposing the structure of cTBP–DNA (PDB: 1YTF) on SAGA-bound cTBP. DNA clashes with Tra1 HEAT repeats and with Spt3 helix  $\alpha$ C. **b**, Gel-shift assay using a Cy5-labelled TATA-containing DNA (TATA), a TATA-like DNA and a non-specific DNA (TATA-less). DNA probes and TFIIA (as indicated) were added to a preformed SAGA–TBP complex. The asterisk indicates minor non-specific DNA

association with SAGA. **c, d**, Pull-down assays with beads that capture SAGA. Components were labelled with different fluorophores and incubated as for gel-shift experiments. The proportion of fluorescent signal that is due to the unbound (**c**) and bound (**d**) complexes is presented. The rightmost bars represent experiments in which TBP was first incubated with DNA and TFIIA before the addition of SAGA. Experiments were repeated three to four times. Data are shown as dots for individual experiments and mean  $\pm$  s.d.

Extended Data Fig. 7a) of DNA binding to SAGA–TBP demonstrate that this mild steric hindrance is nonetheless sufficient to impede binding of even high-affinity DNA containing a consensus TATA box (Fig. 4b, lane 10, Fig. 4c, d).

The general transcription factor TFIIA enhances DNA binding to TBP, competes with Spt8 for occupation of the N-terminal half of cTBP and is found in PICs nucleated by SAGA<sup>32–34</sup>. When a consensus TATA-box-containing DNA was incubated with the SAGA–TBP complex together with TFIIA, TBP was released from SAGA as a ternary DNA–TBP–TFIIA complex (Fig. 4b, lane 3, Extended Data Fig. 7d). The binding of DNA to TBP is coupled with the release of TBP, as we did not observe any DNA associated with SAGA-bound TBP. Indeed, the release of TBP from SAGA can now be understood as being essential for the subsequent steps in assembling the PIC because the clasping of the TBP C-terminal stirrup by Spt3 and TFIIA are mutually exclusive. TBP was also detached from SAGA, although less efficiently, when DNA bearing the TATA element containing two mismatches was used (Fig. 4b, lanes 5 and 6), whereas no release of TBP was observed when the DNA did not contain any TATA or TATA-like elements (Fig. 4b, lanes 7, 8). These experiments show that the detachment of TBP from SAGA correlates with the affinity of TBP for its DNA substrate.

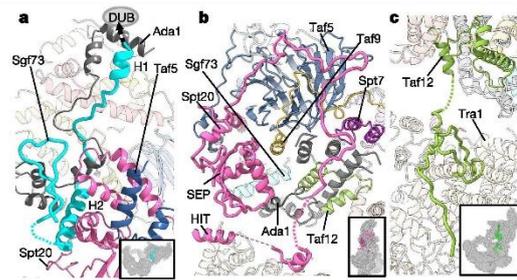
Our structure suggests that the TFIIA-stimulated mechanism for the concerted DNA binding and TBP release from SAGA includes the following steps: TFIIA first displaces Spt8 from TBP; it then promotes an initial binding of DNA to TBP, thus shifting helix  $\alpha$ C of Spt3, which obstructs the path of DNA. Notably, as helix  $\alpha$ C is firmly joined to helix  $\alpha$ 3 in nSpt3-HF, this crucial step is made possible because the deformed octamer does not immobilize this histone fold via a tight four-helix bundle. Shifting helix  $\alpha$ C breaks the pocket that clasps the C-terminal stirrup; TBP can then tilt, allowing DNA to pass in the gap between

Tra1 and the main body of SAGA, completing the binding to DNA. Final release of TBP from SAGA might still require a further conformational change in Spt3. Spt3 alone is unable to bind TBP<sup>23</sup>, and it is tempting to speculate that by shifting helices  $\alpha$ C– $\alpha$ 3 of nSpt3-HF, partially loosening Spt3 ties with SAGA, a conformation similar to isolated Spt3 ensues. Following the release of TBP and DNA, helix  $\alpha$ C can regain its position, enabling Spt3 to restore its TBP-binding conformation.

It has been suggested that in TFIIID, the association of TBP with spurious DNA is prevented by a helical protein domain that mimics TATA DNA and binds at the DNA cleft of TBP<sup>35</sup>. Remarkably, our maps do not show any substantial density attached at the DNA-binding cleft of TBP. Instead, our structure reveals a mechanism for preventing association of non-specific DNA based on limited steric hindrance, distal to the DNA cleft in TBP. Only the synergy of a cognate TATA or TATA-like DNA with TFIIA can overcome this hindrance.

### Firmly tethered flexible enzymatic modules

The DUB and HAT modules are highly dynamic (Fig. 1a, b, Extended Data Fig. 8a, b), permitting the exploration of a large conformational space in search of their chromatin substrates. The structure shows how these modules maintain a robust binding to SAGA despite this flexibility. The DUB module consists of three subunits and the Sgf73 N-terminal end. The C-terminal part of Sgf73 was implicated in attaching DUB to SAGA, and deletion analysis narrowed down the anchoring region to residues 350–400 in *Saccharomyces cerevisiae* (227–277 in *K. phaffii*), as removal of this region leads to dissociation of an intact DUB module from SAGA<sup>36,37</sup>. We found that these residues form an elongated domain comprising two helices, Sgf73-H1 and Sgf73-H2, connected by a stretched loop and embedded in the central module (Fig. 5a). This



**Fig. 5 | Connecting the DUB module and Tra1 to the central module.** a, The Sgf73 domain that contains helices Sgf73-H1 and Sgf73-H2 (cyan) anchors the DUB module within the central module. Sgf73-H2 is embedded in a chamber formed by Spt20 (pink) and Ada1 (dark grey). The insert highlights the depicted region within the overall structure of SAGA. b, c, Bridges between the central module and Tra1 (tan) established by Spt20 (pink) (b) and Taf12 (green) (c).

DUB-anchoring domain traverses the core, entering proximal to the Taf9 histone fold and exiting next to the Spt20 SEP domain. Sgf73-H2 lies in a chamber-like structure composed of several helices from the Spt20 SEP domain and from an elongated domain of Ada1. The chamber consists exclusively of SAGA-specific elements with no contribution from subunits shared with TFIID. Thus, a dedicated environment is created at the core of SAGA, partially occupied in TFIID by the Taf5 NTD, to house the helix that anchors the DUB module into SAGA. The DUB module is tethered to the SAGA core through a 164-residue poorly resolved linker that connects the anchoring domain to the N-terminal end of Sgf73 (Extended Data Fig. 8c, d).

In contrast to the DUB, no parts of the HAT module are embedded in the SAGA core (Fig. 1b). The HAT densities emerge, harbouring two Ada3 helical domains, at the surface of the Taf6 HEAT repeats that are used for homodimerization in TFIID (Extended Data Fig. 8e, f).

Thus, the two enzymatic modules of SAGA adopt very different strategies to remain bound to SAGA while maintaining their freedom of movement. The DUB module is a stably folded structure<sup>10,11</sup> that gains independence through an unstructured long linker that has little contact with the SAGA surface until it connects to an extension firmly embedded in the SAGA core. The HAT module, on the other hand, is itself a flexible structure<sup>12</sup>, with several submodules as well as a large proportion of intrinsically unfolded sequences, and it binds only to the surface of the SAGA core, using two helical domains of Ada3.

### Network of unstructured protein elements

Three distinct bridges serve to physically and possibly functionally couple Tra1 to the central module. The first bridge is established by a highly dispersed domain of Spt20 that originates as a strand joining the Spt20 SEP  $\beta$ -sheet, followed by a helix-turn-helix that also interacts extensively with the Spt20 SEP domain (Fig. 5b). It continues as a loop of around 70 residues, touching several subunits as it traverses the surface of the SAGA core on its 160 Å-long way to Tra1. There, it forms a helix that associates with the surface of the FAT (FRAP, ATM and TRRAP) domain and is followed by a second helix, which is clamped within a concave structure in Tra1. This second helix corresponds to the Spt20 region that is denoted as the HIT domain and the deletion of which prevents assembly of Tra1 into SAGA<sup>38</sup>. An exceptionally long loop that precedes the histone fold of Taf12 is responsible for establishing the second bridge (Fig. 5c). Residues 430–474 from this loop trace a lasso-shaped thread at the surface of the Tra1 FAT domain within the deep groove formed between two tetratricopeptide repeat-containing domains (TRDs). The third bridge is formed by the loop connecting the

second and third helices of cSpt3-HF as it infiltrates between two HEAT repeats of the Tra1 ring (Extended Data Fig. 9a, b).

We also observe several protein stretches that lack secondary structure elements but contact multiple subunits and contribute considerably to an intricate intertwined network of interactions within the main lobe of SAGA. For example, the C-terminal tail in Taf9 and the short loop preceding it contact four histone folds and six other domains (Extended Data Fig. 9c). This network of interactions is necessary for stabilizing an octamer that could not be reconstituted *in vitro* from its constituent histone folds and underlies the effects of some mutations on functions that do not directly involve the altered proteins<sup>39</sup>.

### Conclusion

The high-resolution description of SAGA unravels how the central module acts as a scaffold that coordinates an activator-interaction platform and two enzymatic modules, as well as TBP binding and release. Our structure reveals that a deformed histone-fold octamer, aided by components of Spt7 and Spt8, is finely tuned to define a TBP-docking site in which distal steric hindrance represses DNA binding, thus preventing unwanted transcription initiation. However, with the assistance of TFIIA, TBP can productively bind DNA to form a ternary TBP–TFIIA–DNA complex that is released from SAGA. The efficiency and probability of TBP detachment is commensurate with the affinity of TBP for the DNA sequence.

These findings imply a short residency time for SAGA on the minimal promoter, in accordance with genome-wide analyses that locate SAGA mainly at upstream activating sequences (UASs)<sup>7,40</sup>. The peripheral position of TBP and its binding mode in SAGA also suggest that transcription regulatory factors such as negative cofactor 2 (NC2) or Mot1 could gain access to TBP with relative ease. These results are consistent with a highly dynamic scenario for the control of TBP turnover<sup>41</sup>.

Sequence homology strongly suggests that the histone-fold machinery in TFIID assumes a similar function, which implies that the mechanics of depositing TBP on DNA by the two complexes share common features. However, several crucial aspects of TBP binding and delivery vary between the two complexes, including the capacity of the TFIID subunits Taf2 and Taf4 to bind promoter DNA, the ability of the Taf1 NTD to interact with the DNA-binding interface of TBP and the stronger affinity of Taf11–13, compared with Spt3, for TBP<sup>42</sup>. Thus, access of transcription regulatory factors to TBP and residency time at the minimal promoter are expected to be significantly different between TFIID and SAGA.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1944-2>.

1. Helmlinger, D. & Toral, L. Shering the SAGA. *Trends Biochem. Sci.* **42**, 850–861 (2017).
2. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327–335 (1996).
3. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 129–143 (2015).
4. Hahn, S. & Young, E. T. Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189**, 705–736 (2011).
5. Larschan, E. & Winston, F. The *S. cerevisiae* SAGA complex functions *in vivo* as a coactivator for transcriptional activation by Gal4. *Genes Dev.* **15**, 1946–1956 (2001).
6. Bhaumik, S. R. & Green, M. R. Differential requirement of SAGA components for recruitment of TATA-box-binding protein to promoters *in vivo*. *Mol. Cell. Biol.* **22**, 7365–7371 (2002).
7. Baptista, T. et al. SAGA is a general cofactor for RNA polymerase II transcription. *Mol. Cell* **68**, 130–143 (2017).
8. Warfield, L. et al. Transcription of nearly all yeast RNA polymerase II-transcribed genes is dependent on transcription factor TFIID. *Mol. Cell* **68**, 116–129 (2017).

## Article

9. Lee, K. K. et al. Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. *Mol. Syst. Biol.* **7**, 503 (2011).
10. Köhler, A., Zimmermann, E., Schneider, M., Hurt, E. & Zheng, N. Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. *Cell* **141**, 606–617 (2010).
11. Samara, N. L. et al. Structural insights into the assembly and function of the SAGA deubiquitinating module. *Science* **328**, 1025–1029 (2010).
12. Sun, J. et al. Structural basis for activation of SAGA histone acetyltransferase Gcn5 by partner subunit Ada2. *Proc. Natl Acad. Sci. USA* **115**, 10010–10015 (2018).
13. Diaz-Santín, L. M., Lukoyanova, N., Aciyan, E. & Cheung, A. C. Cryo-EM structure of the SAGA and NuA4 coactivator subunit Trt1 at 3.7 angstrom resolution. *eLife* **6**, e28384 (2017).
14. Sharov, G. et al. Structure of the transcription activator target Trt1 within the chromatin modifying complex SAGA. *Nat. Commun.* **8**, 1556 (2017).
15. Setiawati, D. et al. Conformational flexibility and subunit arrangement of the modular yeast Spt-Ada-Gcn5 acetyltransferase complex. *J. Biol. Chem.* **290**, 10057–10070 (2015).
16. Wu, P. Y., Ruhlmann, C., Winston, F. & Schultz, P. Molecular architecture of the *S. cerevisiae* SAGA complex. *Mol. Cell* **15**, 199–209 (2004).
17. Gangloff, Y. G., Romier, C., Thuault, S., Werten, S. & Davidson, I. The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem. Sci.* **26**, 250–257 (2001).
18. Grant, P. A. et al. A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. *Cell* **94**, 45–53 (1998).
19. Kolesnikova, O. et al. Molecular structure of promoter-bound yeast TFIID. *Nat. Commun.* **9**, 4666 (2018).
20. Patel, A. B. et al. Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362**, eaau8872 (2018).
21. Eisenmann, D. M., Arndt, K. M., Ricupero, S. L., Rooney, J. W. & Winston, F. SPT3 interacts with TFIID to allow normal transcription in *Saccharomyces cerevisiae*. *Genes Dev.* **6**, 1319–1331 (1992).
22. Mohibullah, N. & Hahn, S. Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3. *Genes Dev.* **22**, 2994–3006 (2008).
23. Sermwittayawong, D. & Tan, S. SAGA binds TBP via its Spt8 subunit in competition with DNA: implications for TBP recruitment. *EMBO J.* **25**, 3791–3800 (2006).
24. Birck, C. et al. Human TAF<sub>28</sub> and TAF<sub>18</sub> interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. *Cell* **94**, 239–249 (1998).
25. Hoffmann, A. et al. A histone octamer-like structure within TFIID. *Nature* **380**, 356–359 (1996).
26. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **399**, 251–260 (1997).
27. Nikolov, D. B. et al. Crystal structure of a TFIIB–TBP–TATA-element ternary complex. *Nature* **377**, 119–128 (1995).
28. Wu, P. Y. & Winston, F. Analysis of Spt7 function in the *Saccharomyces cerevisiae* SAGA coactivator complex. *Mol. Cell. Biol.* **22**, 5367–5379 (2002).
29. Geiger, J. H., Hahn, S., Lee, S. & Sigler, P. B. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* **272**, 830–836 (1996).
30. Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**, 127–134 (1996).
31. Wollmann, P. et al. Structure and mechanism of the Swi2/Snf2 remodeller Mot1 in complex with its substrate TBP. *Nature* **475**, 403–407 (2011).
32. Imbalzano, A. N., Zaret, K. S. & Kingston, R. E. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *J. Biol. Chem.* **269**, 8280–8286 (1994).
33. Warfield, L., Ranish, J. A. & Hahn, S. Positive and negative functions of the SAGA complex mediated through interaction of Spt8 with TBP and the N-terminal domain of TFIIA. *Genes Dev.* **18**, 1022–1034 (2004).
34. Petrenko, N., Jin, Y., Dong, L., Wong, K. H. & Struhl, K. Requirements for RNA polymerase II preinitiation complex formation in vivo. *eLife* **8**, e43654 (2019).
35. Anandapadmanabhan, M. et al. High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. *Nat. Struct. Mol. Biol.* **20**, 1008–1014 (2013).
36. Han, Y., Luo, J., Ranish, J. & Hahn, S. Architecture of the *Saccharomyces cerevisiae* SAGA transcription coactivator complex. *EMBO J.* **33**, 2534–2546 (2014).
37. Kamata, K. et al. C-terminus of the Sgf73 subunit of SAGA and SLIK is important for retention in the larger complex and for heterochromatin boundary function. *Genes Cells* **18**, 823–837 (2013).
38. Elias-Villalobos, E., Toullac, D., Faux, C., Séveno, M. & Helmlinger, D. Chaperone-mediated ordered assembly of the SAGA and NuA4 transcription co-activator complexes. *Nat. Commun.* **10**, 5237 (2019).
39. Saint, M. et al. The TAF9 C-terminal conserved region domain is required for SAGA and TFIID promoter occupancy to promote transcriptional activation. *Mol. Cell. Biol.* **34**, 1547–1563 (2014).
40. Huisings, K. L. & Pugh, B. F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol. Cell* **18**, 573–585 (2004).
41. Ravarani, C. N., Chalancón, G., Breker, M., de Groot, N. S. & Babu, M. M. Affinity and competition for TBP are molecular determinants of gene expression noise. *Nat. Commun.* **7**, 10417 (2016).
42. Gupta, K. et al. Architecture of TAF1/TAF13/TBP complex suggests novel regulation properties of general transcription factor TFIID. *eLife* **6**, e30395 (2017).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

No statistical methods were used to predetermine sample size.

### Purification of SAGA

The SAGA complex was purified from nuclear extracts of the budding yeast *K. phaffii* (also known as *P. pastoris*) using the 38-amino-acid streptavidin-binding peptide (SBP) affinity tag that was fused to the C terminus of the endogenous Sgf73 subunit (Extended Data Figs. 1, 7e). We introduced several modifications into our original procedure<sup>34</sup>, with the combined effect of increasing the yield by three times. These improvements enabled us to reduce the harmful effects of proteases and to keep SAGA more concentrated along the different steps of the procedure.

Two litres of yeast cells (four flat-bottom flasks with 0.5 l in each) were grown at 24 °C with glycerol as carbon source and collected when optical density at 600 nm ( $OD_{600\text{nm}}$ ) reached 16–18. Cells were washed in ice-cold water and then treated with 10 mM DTT. The cell wall was digested by addition of lyticase and spheroplasts were pelleted at 5,500g for 20 min. All further steps were performed at 0 to 4 °C. Protease inhibitors were added to all buffers. Spheroplasts were washed twice in 1.1 M sorbitol and were then disrupted by suspension in a hypotonic buffer (15–18% Ficoll 400, 0.6 mM  $MgCl_2$ , 20 mM K-phosphate buffer pH 6.6) using a ULTRA-TURRAX disperser. Sucrose (0.1 M) and  $MgCl_2$  (5 mM) were then added. Nuclei (and some debris) were pelleted at 33,000g for 37 min, resuspended in a wash buffer (0.6 M sucrose, 8% polyvinylpyrrolidone, 1 mM  $MgCl_2$ , 20 mM phosphate buffer pH 6.6) and pelleted again at 34,000g for 50 min. Nuclei were resuspended in low ionic extraction buffer (40 mM HEPES pH 8.0, 20% sucrose, 8 mM  $MgCl_2$ , 5–6 mM DTT) with 30 strokes using a tight pestle in a dounce homogenizer. Very few molecules of SAGA are liberated under the low ionic conditions, leaving time for protease inhibitors to bind their targets before SAGA is extracted.

After 20 min of incubation, 300 mM NaCl, 2 mM  $CaCl_2$  and 150  $\mu$ l of  $\alpha$ -amylase solution (MegaZyme) were added. Following another 30 min of incubation, debris was precipitated at 33,000g for 38 min. The supernatant was collected and 1–2% PEG 20,000 as well as additional 5 mM  $MgCl_2$  were added to precipitate some remaining organelles and membrane parts by a short centrifugation step at 33,000g for 10 min. The increase in magnesium concentration (final 13 mM) proved crucial to prevent precipitation of SAGA at this step. The PEG 20,000 concentration was then increased to 5.8% and SAGA precipitated in a second short centrifugation step. The pellet was solubilized in a minimal volume and avidin was added to block endogenously biotinylated proteins. The suspension was incubated with streptavidin beads for 4 h in buffer A (40 mM HEPES pH 8.0, 250 mM sodium chloride, 10% sucrose, 2 mM  $MgCl_2$ , 2 mM DTT) washed 5 times and eluted with buffer A containing 10 mM biotin. The eluate was concentrated with Millipore Amicon-Ultra (50 kDa cut-off) and spun in a 10–30% sucrose gradient with buffer B (20 mM HEPES pH 8.0, 150 mM potassium acetate, 2 mM DTT, 6 mM  $MgCl_2$ ) in rotor SW60 (38,300 rpm for 13.5 h). SAGA was fractionated at approximately 25% sucrose and concentrated with Amicon-Ultra to 2 mg  $ml^{-1}$ . It is important to note that mass spectrometry (Extended Data Fig. 7) and SDS-PAGE analysis<sup>34</sup> show no cleavage of the Spt7 subunit or sub-stoichiometric amounts of Spt8. It is also noteworthy that very little endogenous TBP is present in our final SAGA sample.

### Purification of TBP and TFIIA

Full-length yeast *S. cerevisiae* TBP was expressed as N-terminal 6 $\times$ His-GFP fusion protein in *Escherichia coli* strain BL21(DE3)RIL. Cells were grown in LB medium with ampicillin until an  $OD_{600\text{nm}}$  of 0.7 was reached. Expression was induced with 1 mM IPTG for 3 h at 37 °C. The cells were collected by centrifugation and resuspended in lysis buffer (20 mM HEPES-KOH (pH 7.5), 0.5 M NaCl, 10% glycerol). Cells were treated with 1 mg  $ml^{-1}$  lysozyme (Sigma) for 30 min on ice, and then

lysed by sonication after the addition of cComplete, EDTA-free Protease Inhibitor Cocktail (Roche) and PMSF to a final concentration of 1 mM. The resulting lysate was clarified by centrifugation and incubated with equilibrated cComplete His-Tag Purification Resin (Roche) for 1 h at 4 °C. After the beads were collected, they were washed three times with lysis buffer supplemented with 5 mM imidazole, and the bound protein was eluted with 20 mM HEPES-KOH (pH 7.5), 0.2 M NaCl, 10% glycerol supplemented with 250 mM imidazole. The recombinant protein was further purified on a Heparine HiTrap 1 ml column (GE Healthcare) equilibrated with 20 mM HEPES-KOH (pH 7.5), 0.2 M NaCl, 20% glycerol and eluted with a gradient from 200 mM to 1 M NaCl. Fractions were analysed by Coomassie blue staining, pooled and concentrated in Amicon-Ultra with a 10,000 (10K)-molecular-weight cut-off (Millipore) to a final concentration of 3 mg  $ml^{-1}$ . DTT (5 mM) was added to all buffers immediately before use. The yeast *S. cerevisiae* TFIIA was expressed in *E. coli*, purified from inclusion bodies and reconstituted as described earlier<sup>33</sup>.

### Gel-shift DNA-binding assays

DNA probes were obtained by annealing of equimolar amounts of complementary oligonucleotides (Eurofins Genomics for short DNA, SIGMA for long DNA) at a final concentration of 25  $\mu$ M in 20 mM HEPES-KOH, 50 mM NaCl by heating the mixture to 100 °C for 5 min and cooling slowly down to room temperature. To estimate the quality of annealing, the obtained DNA fragments were loaded on a 12% polyacrylamide gel and analysed using a Typhoon FLA9500 Imager (GE Healthcare) to follow Cy5 fluorescence signal. No Cy5-labelled single-stranded DNA was detected (Extended Data Fig. 7c). To further ascertain that we removed all effects of single-stranded DNA, all DNA samples were treated with Exonuclease I (NEB). Binding assays (10  $\mu$ l) contained 3 mM Tris-HCl (pH 8.0), 6 mM HEPES-KOH (pH 8.0), 45 mM KAc, 20–42 mM KCl, 5 mM  $MgCl_2$ , 3% glycerol, 6% sucrose, DNA probe (Cy5-labelled), 0.01% NP40, 25  $\mu$ g  $ml^{-1}$  bovine serum albumin (BSA) (NEB), 1 mM DTT and proteins as indicated. TBP was incubated first with a fourfold excess of SAGA complex for 1 h at 4 °C. Under these conditions almost all TBP was bound to SAGA. Then DNA and TFIIA were added, and the reaction mixtures were incubated for 20 min at room temperature before they were loaded onto agarose-polyacrylamide native gels (1% agarose, 1.3% acrylamide with 1 $\times$  Tris-glycine buffer, 5 mM  $MgCl_2$  in the gel and running buffer). The final reaction mix contained 0.075  $\mu$ M GFP-TBP, 0.3  $\mu$ M SAGA, 0.075  $\mu$ M DNA and 0.15  $\mu$ M TFIIA complex. Binding assays with long (100-bp) DNA fragments were done under similar conditions except that a threefold excess (with respect to SAGA) of long non-labelled TATA-less DNA was also added because SAGA was shown to bind long DNA non-specifically<sup>44</sup>.

The gel-shift experiments were repeated several times, all showing consistent results. Gels were analysed using a Typhoon FLA9500 Imager to follow Cy5 fluorescence signal. TBP-DNA and TBP-TFIIA-DNA migrate similarly in native agarose-polyacrylamide gels due to the limited resolving power of these gels for relatively low-molecular-weight complexes. We ascertained the composition of the shifted bands by mass spectrometry (Extended Data Fig. 7d). Sequences of DNA molecules used: 'TATA' (yeast *CYC1* TATA), 5'-Cy5-TGCTCTGTATGATATATAAAACCTCTG-3'; 'TATA-like' (*CYC1* TATA with two mutations decreasing the affinity to TBP), 5'-Cy5-TGCTCTGTATGATAGAGAAACTCTG-3'; 'TATA-less' (non-specific DNA, *CYC1* UAS), 5'-Cy5-GGCCGGGTTTACGGACGATGGCAG-3'; 'TATA<sub>1</sub>' (*Pichia pastoris* pGAP promoter fragment), TGTCATGAGATATTGGAAACCACCAGAATCGAATATAAAGCGCAACACCTTTCCCAATTTGGTTTCTCCTGACCCAAAGACTTTAAATTTAATTTA; 'TATA-like<sub>2</sub>' (TATA<sub>1</sub> with two mutations decreasing the affinity to TBP), TGTCATGAGATATTGGAAACCACCAGAATCGAAGAGAAAAGCGCAACACCTTTCCCAATTTGGTTTCTCCTGACCCAAAGACTTTAAATTTAATTTA; 'TATA-less<sub>2</sub>' (non-specific DNA), GACTCATGTCATGAGATCATTGGACACCACCAGAATCGCGTATCGAAGCGAACACCTGTCTCAGCTCTGGTGTCTCCTGACCCACAGACTTCGAA

## Article

CGTA. The position of the TATA sequence is indicated by bold letters and the two mismatches in the TATA-less sequences are underlined.

### Pull-down DNA-binding assays

Binding assays (20  $\mu$ l) contained 3 mM Tris-HCl (pH 8.0), 5 mM HEPES-KOH (pH 8.0), 35 mM KAc, 46 mM KCl, 5 mM MgCl<sub>2</sub>, 3% glycerol, 6% sucrose, 0.01% NP40, 25  $\mu$ g ml<sup>-1</sup> BSA (NEB), 1 mM DTT, proteins and DNA probe (Cy5-labelled 'TATA' DNA) as indicated. TBP was incubated first with a fourfold excess of SAGA complex for 1 h at 4 °C. Then DNA and TFIIA were added (as indicated), and the reaction mixtures were incubated for 20 min at room temperature. The final reaction mix contained 0.075  $\mu$ M GFP-TBP, 0.3  $\mu$ M SAGA (with Sgf11 subunit fused to mCherry), 0.075  $\mu$ M Cy5-DNA and 0.15  $\mu$ M TFIIA complex. Ten microlitres of streptavidin-sepharose beads (equilibrated with the binding buffer) were added to each reaction mix and tubes were incubated for 2 h at 4 °C with shaking using ThermoMixerC (Eppendorf, shaking frequency 1,220 rpm). Each reaction mix containing beads was transferred into an Ultrafree-MC centrifugal filter unit and centrifuged for 5 min at 12,000g at 4 °C. Beads with retained protein or DNA-protein complexes and the flowthrough with unbound material were analysed using a PHERAstar Plus microplate reader (BMG LABTECH) to quantify the fluorescence signal of GFP (corresponding to TBP), mCherry (corresponding to SAGA complex) and Cy5 (corresponding to DNA). A complementary experiment with the reverse order of incubations was also carried out. Hence, TBP was first incubated with Cy5-labelled DNA and TFIIA for 20 min at room temperature, then a fourfold excess of SAGA complex was added and reactions were incubated for 1 h at 4 °C. Experiments were repeated three to four times.

### Cryo-EM sample preparation and data acquisition

Concentrated SAGA was incubated overnight on ice with an excess of ubiquitin aldehyde (UbAl) at a molar ratio of 1:4. Complete binding and inhibition was ascertained by following the activity of SAGA using Ubiquitin-AMC (Enzo). SAGA-UbAl complexes were precipitated with PEG to remove sucrose<sup>14</sup>, suspended at a concentration of 1 mg ml<sup>-1</sup> and incubated overnight on ice with a 5:1 molar excess of full-length *S. cerevisiae* TBP. SAGA-TBP was then diluted to 0.2 mg ml<sup>-1</sup> and cross-linked with 0.1% glutaraldehyde for 30 min on ice. About 3  $\mu$ l of sample was applied onto a holey carbon grid (Quantifoil R2/2.300 mesh) rendered hydrophilic by a 90-s treatment in a Fischione 1070 plasma cleaner operating at 30% power with a gas mixture of 80% argon:20% oxygen. The grid was blotted for 1 s (blot force 8) and flash-frozen in liquid ethane using a Vitrobot Mark IV (FEI) at 4 °C and 95–100% humidity. To verify that the occupancy of GFP-TBP on SAGA was approaching 100%, we loaded the cross-linked complex on a superose 6 10/300 GL column (GE Healthcare) and compared the fluorescent signal at the peak fraction with that of cross-linked GFP-TBP of known concentrations. The concentrations of SAGA were estimated by UV absorbance. Images were acquired on a Cs-corrected Titan Krios (Thermo Fisher) microscope operating at 300 kV in nanoprobe mode using the serialEM software for automated data collection<sup>45</sup>. Movie frames were recorded on a 4k  $\times$  4k Gatan K2 summit direct electron detector after a Quantum LS 967 energy filter using a 20-eV slit width in zero-loss mode. Images were acquired in super-resolution mode at a nominal magnification of 105,000, which yielded a pixel size of 0.55 Å. Forty movie frames were recorded at a dose of 1.32 e<sup>-</sup> Å<sup>-2</sup> per frame, corresponding to a total dose of 52.8 e<sup>-</sup> Å<sup>-2</sup>, but only the last 38 frames were kept for further processing.

### Image processing

Movie frames were aligned, dose-weighted, binned by 2 and averaged using Motioncor2<sup>46</sup> to correct for beam-induced specimen motion and to account for radiation damage by applying an exposure-dependent filter. Non-weighted movie sums were used for contrast transfer function (CTF) estimation with the Gctf<sup>47</sup> program, while dose-weighted sums

were used for all subsequent steps of image processing. After manual screening, images with poor CTF, particle aggregation or ice contamination were discarded. Particles were picked using crYOLO (<https://doi.org/10.1101/356584>). These datasets were analysed in RELION 3.0<sup>48</sup>, cryoSPARC<sup>49</sup> and cisTEM<sup>50</sup> according to standard protocols. In brief, three rounds of reference-free 2D classification of the individual particle images were performed in RELION to remove images, like ice contaminations or deformed particles, that failed to enter into high-resolution class averages. The selected images were aligned against a low-pass-filtered starting model<sup>14</sup>. Three rounds of 3D classification with increasing regularization parameter *T* were performed and classes showing high-resolution features were selected for the subsequent steps. These images were imported, and refined in cryoSPARC v.2. Three-dimensional classification of the entire dataset could not clearly separate distinct conformations of the SAGA complex. Therefore, we carried out a focused refinement of the separate lobes using the masked lobes as references. The Tra1 lobe was refined with a mask covering the Tra1 densities using the 'Local refinement' module. The structure of the main lobe was obtained by using signal subtraction, in which the densities corresponding to the Tra1 lobe were subtracted from the images. This operation was followed by local refinement with a mask covering the main lobe densities. Global resolution estimates were determined using the Fourier shell correlation (FSC) = 0.143 criterion after a gold-standard refinement. Local resolutions were estimated with ResMap<sup>51</sup> and cryoSPARC. Maps to fit TBP, Spt8 and the DUB module were obtained after 3D classifications using RELION 3 by applying masks covering the region of interest.

To describe the flexibility of the DUB and HAT modules, the selected SAGA images were analysed using the IMAGIC software package<sup>52</sup>. The selected images were aligned against 75 equidistant projections of the 3D model and clustered to form 1,000 2D classes. Images from classes corresponding to two different views were grouped and aligned separately against their corresponding reference. Local classification into 50 classes was performed for each group by using an adapted mask and the best class averages were selected to show the extent of protein density variation within each view.

### Model building

Pairwise alignment was performed with Needle<sup>53</sup>. Homology models for Tra1, TBP, Spt8, DUB module, all eight histone folds, Taf6 HEAT repeats, Taf5 WD40, Spt20 SEP domain and Taf5 NTD domains were generated using I-TASSER<sup>54</sup> or Swiss-Model and docked into the maps manually. The rigid-body-docking of DUB and Spt8 was guided by the overall shape of the local density, which lacked clear secondary structure features, probably owing to the peripheral position and flexible nature of their connection to the SAGA core. The position of the DUB is consistent with previous studies, notably XL/MS<sup>56</sup> and negatively stained particles with GFP-labelled subunits of the module<sup>55</sup>. As the DUB is composed of two lobes with similar overall shape we cannot exclude at the current resolution an inverted (upside-down) orientation for the DUB. The location and orientation of Spt8 is validated by cross-linking mass spectrometry<sup>54</sup>, specific single-site photo-activated cross-linking<sup>23,33</sup> and biochemical data<sup>16</sup>.

The other homology models, with the notable exception of TBP, for which only a minor tilt of the C-terminal stirrup was applied, were rebuilt according to density using Coot<sup>55</sup> and then extended from their N' and C' termini until a significant discontinuity in the map was encountered. Such discontinuities normally corresponded to sites where RaptorX<sup>56</sup> and PSIPRED<sup>57</sup> predictions indicated a large disordered region. To assign densities that were not clearly connected to any of these extended starting models, we considered the local map properties including secondary structure features, proteins in the vicinity, biochemical data, cross-linking information for surrounding domains and, most importantly, the fit between protein side chains and the density. These parts of the structure were then built de novo.

For a large proportion of the side chains, density was clear and practically all densities in the core of SAGA and Tra1 could be assigned with confidence. The single exception to this procedure is the Ada3 domain at the docking site of the HAT module, where density showed secondary structure elements but could not resolve individual side chains. We cannot therefore exclude errors in amino-acids register for this domain. Tra1 and the main lobe were separately geometry-optimized in PHENIX<sup>58</sup> followed by real-space refinement with secondary structure restraints against the corresponding post-processed maps. This procedure was repeated after a round of final manual corrections to the structure in Coot. All display images were generated using UCSF Chimera<sup>59</sup> and ChimeraX<sup>60</sup>.

#### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

#### Data availability

The cryo-EM maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes EMD-10438 (SAGA-TBP), EMD-10440 (SAGA-TBP, refined for Tra1 lobe), EMD-10441 (SAGA-TBP, refined for main lobe), EMD-10446 (SAGA-TBP, low-pass-filtered), EMD-10448 (SAGA-TBP, focused classification on TBP), EMD-10447 (SAGA without TBP). The model coordinates for SAGA-TBP were deposited in the Protein Data Bank (PDB) under the accession codes 6TB4 and 6TBM (including Spt8 and DUB).

43. Ranish, J. A., Lane, W. S. & Hahn, S. Isolation of two genes that encode subunits of the yeast transcription factor IIA. *Science* **255**, 1127–1129 (1992).
44. Mittal, C., Culbertson, S. J. & Shogren-Kneek, M. A. Distinct requirements of linker DNA and transcriptional activators in promoting SAGA-mediated nucleosome acetylation. *J. Biol. Chem.* **283**, 13736–13749 (2018).
45. Mastrorade, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **182**, 36–51 (2005).
46. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
47. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
48. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
49. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
50. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
51. Kucukelbir, A., Sigworth, F. J. & Tagara, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
52. van Heel, M., Harauz, G., Orlov, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).

53. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
54. Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174–B1 (2015).
55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
56. Källberg, M. et al. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522 (2012).
57. Buchan, D. W., Minnici, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **41**, W349–57 (2013).
58. Terwilliger, T. C. Rapid model building of alpha-helices in electron-density maps. *Acta Crystallogr. D* **66**, 268–275 (2010).
59. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
60. Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
61. Zhou, Q. & Berk, A. J. The yeast TATA-binding protein (TBP) core domain assembles with human TBP-associated factors into a functional TFIID complex. *Mol. Cell. Biol.* **15**, 534–539 (1995).
62. Kalluher, R. J., III et al. Yeast and human TFIIDs are interchangeable for the response to acidic transcriptional activators in vitro. *Genes Dev.* **6**, 296–303 (1992).
63. Ozer, J., Lezina, L. E., Ewing, J., Audi, S. & Lieberman, P. M. Association of transcription factor IIA with TATA binding protein is required for transcriptional activation of a subset of promoters and cell cycle progression in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **18**, 2559–2570 (1998).
64. Laprade, L., Rose, D. & Winston, F. Characterization of new Spt3 and TATA-binding protein mutants of *Saccharomyces cerevisiae*: Spt3 TBP allele-specific interactions and bypass of Spt8. *Genetics* **177**, 2007–2017 (2007).

**Acknowledgements** We thank D. Devys and L. Tora for carefully reading the manuscript and for advice; R. Wagner and B. Séraphin for advice during the initial stages of this project; and R. Ben-Sherm for graphics expertise. We acknowledge support from the Institut National de la Santé et de la Recherche Médicale (INSERM), the Centre National pour la Recherche Scientifique (CNRS), the Association pour la Recherche sur le Cancer (ARC), the Ligue contre le Cancer, ANR-15-CE11-0022-01 and ANR-10-LABX-0030-INRT, a French State fund managed by the Agence Nationale de la Recherche under the frame program investissements d'avenir ANR-10-IDEX-0002-02. We acknowledge the use of resources of the French Infrastructure for Integrated Structural Biology FRISBI ANR-10-INBS-05 and of Instruct-ERIC.

**Author contributions** P.S. and A.B.-S. designed the study; A.B.-S. designed the SAGA purification method and reconstituted the SAGA-TBP complex; A.B.-S. and C.C. defined conditions for grid preparation and freezing; C.C. and G.P. froze grids; G.P. collected and analysed cryo-EM data; A.F., G.P. and A.B.-S. interpreted the maps by fitting crystal coordinates and model building; O.K. purified TBP, TFIIA and SAGA and used them to perform pull-down and gel-shift assays; P.S. and A.B.-S. supervised the work; G.P. and O.K. prepared figures; and A.B.-S. and P.S. wrote the manuscript with input from all authors.

**Competing interests** The authors declare no competing interests.

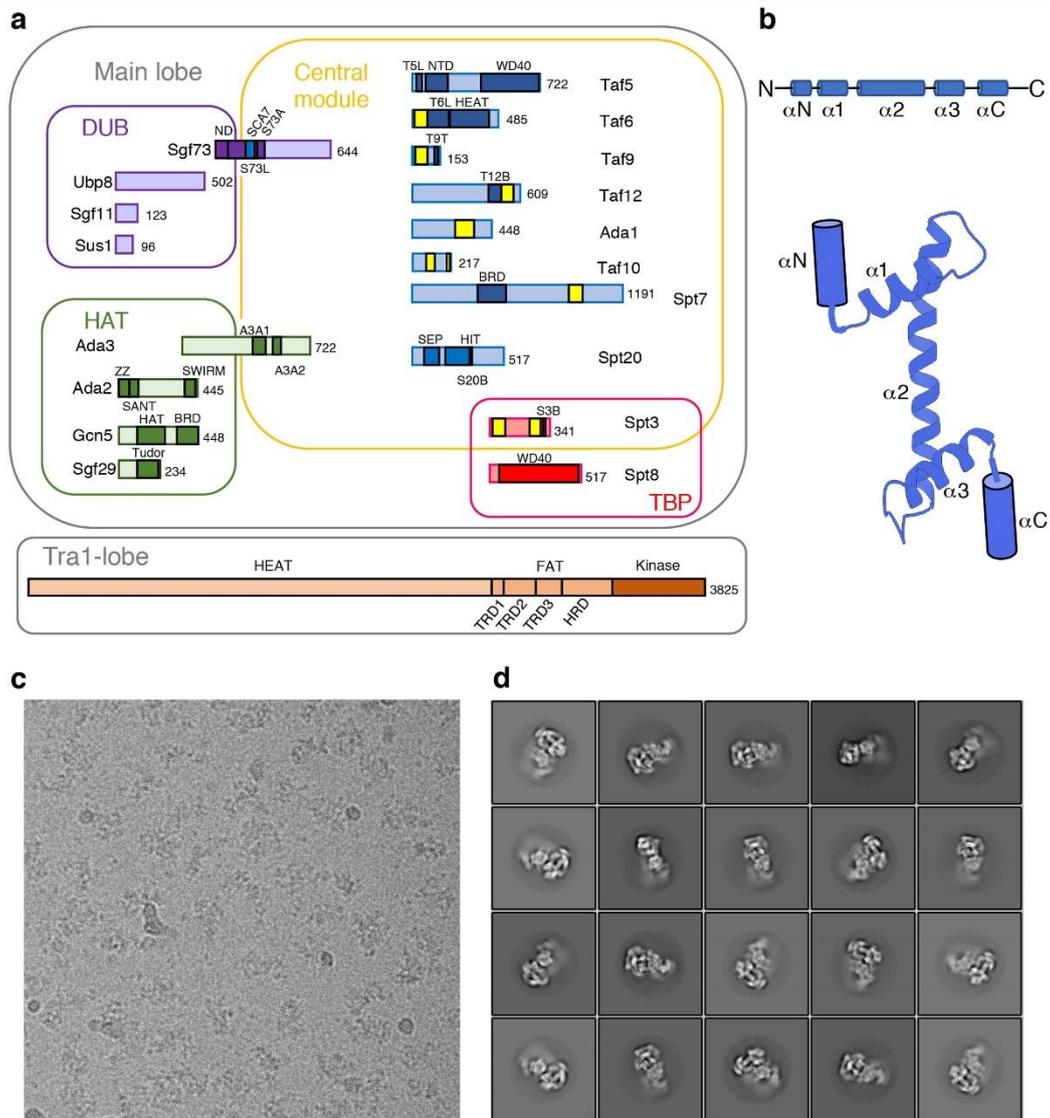
#### Additional information

**Supplementary Information** is available for this paper at <https://doi.org/10.1038/s41586-020-1844-2>.

**Correspondence and requests for materials** should be addressed to P.S. or A.B.-S.

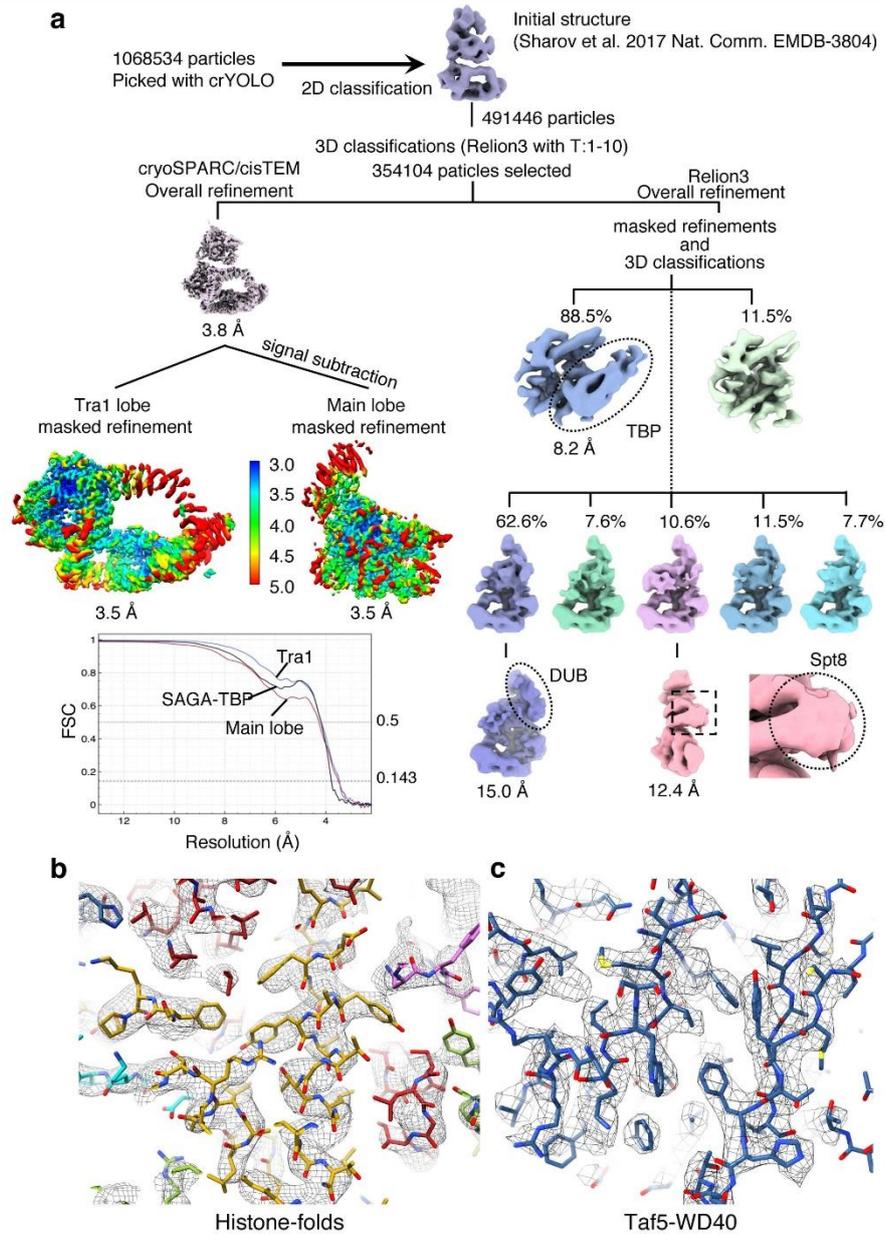
**Peer review Information** Nature thanks Steve Hahn and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



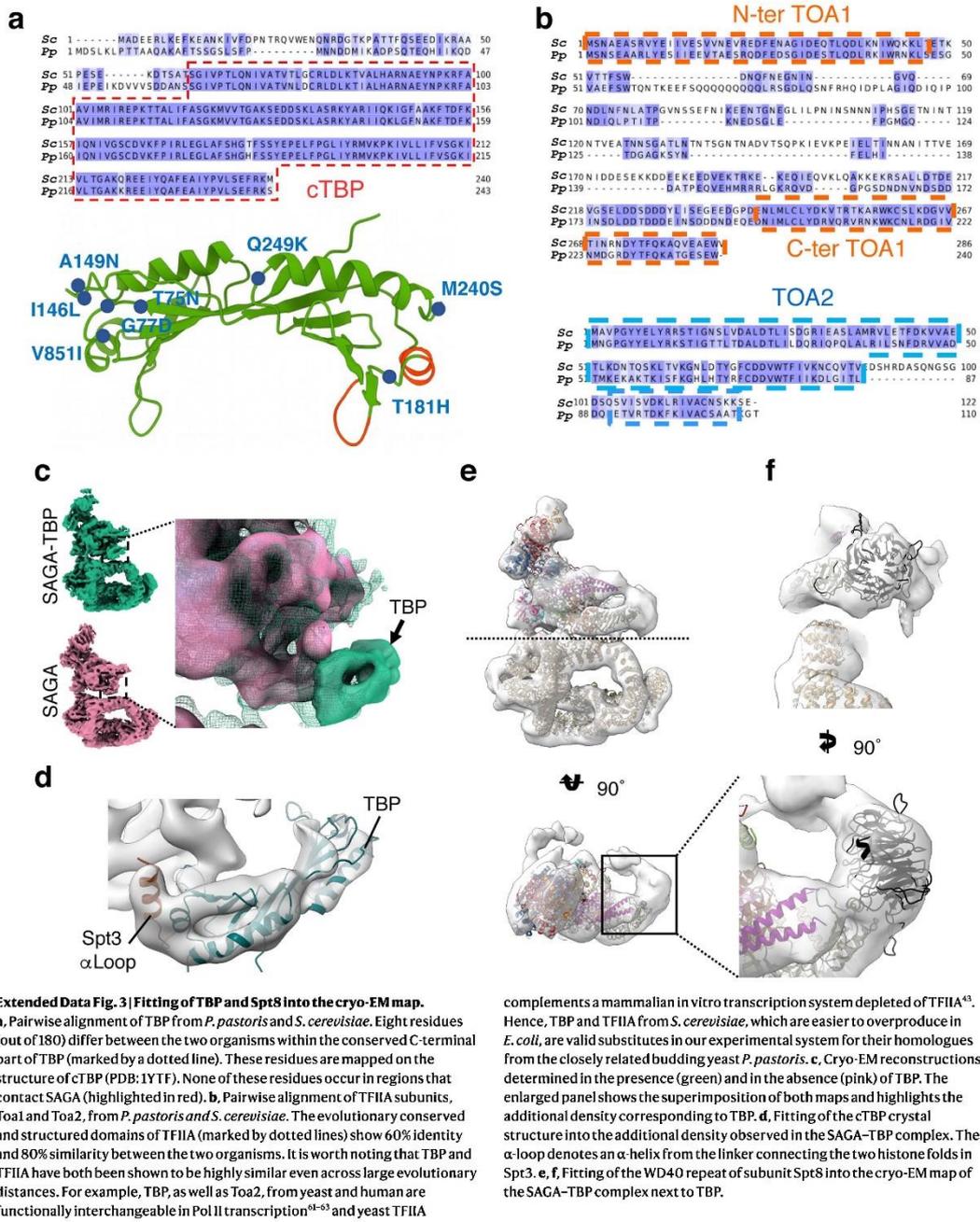
**Extended Data Fig. 1 | Subunit and domain organization of the yeast SAGA complex.** **a**, SAGA subunits are organized into the Tra1 lobe and the main lobe. The main lobe is divided into enzymatic HAT and DUB modules, a central module and a TBP-binding module. Overlaps appear as the central module is a scaffold for the assembly of all other modules. Abbreviations used owing to space limitations: Taf5 lissencephaly homology (LisH) domain (TSL), Taf6 linker domain (N83–S218, T6L), Taf9 C tail (N110–L139, T9T), Sgf73 NTD (M1–K69, ND), Sgf73 linker region (I70–H234, S73L), Sgf73 anchoring domain (E235–L277, S73A), Spt20 HIT domain interacting with Tra1 (M325–L340, HIT), Ada3 anchoring domains (K398–L471, A3A1) and (Y509–E557, A3A2) and bromodomain (BRD). The SEP domain found in Spt20 is named after

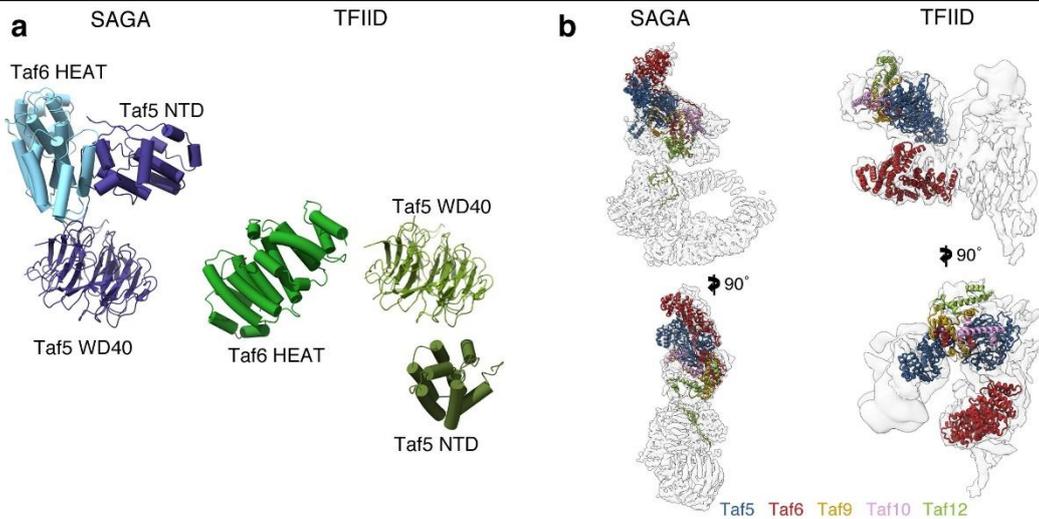
*S. cerevisiae* Shp1, *Drosophila melanogaster* eyes closed gene (*Eyc*) and vertebrate P47. The FAT repeats are divided into three TRD domains and one HEAT-repeats domain (HRD). We defined three bridge-forming domains involved in connecting Tra1 to the central module: the Spt20 bridge (K184–K336, S20B), the Taf12 bridge (T430–K502, T12B) and the Spt3 bridge (E268–I304, S3B). The histone-fold domains, present in seven subunits, are highlighted as yellow boxes. **b**, Schematic representation and helix nomenclature for histone-fold domains. **c**, Original micrograph of frozen-hydrated *P. pastoris* SAGA–TBP complex. **d**, Two-dimensional class averages showing high-resolution structural features.



**Extended Data Fig. 2 | Cryo-EM data-analysis strategy and resolution assessment of the cryo-EM structures. a.** Image-processing strategy used to obtain the 3.5 Å resolution maps of the SAGA-TBP complex. Maps coloured in rainbow represent the local resolution of the reconstructions. FSC curves are depicted as a function of resolution in angstrom for the entire SAGA-TBP complex (black), the Tra1 lobe (blue) and the main lobe (red). cryoSPARC v.2 and cisTEM were used to generate the 3.5 Å resolution maps of the Tra1 and main lobes. RELION 3 was used for masked refinements and classifications of

flexible regions, corresponding to TBP, the DUB module and Spt8. Dotted circles show the regions that were used to dock the crystal structures of TBP (PDB: 1TBP), Spt8 (homology model by Swiss-Model using PDB 4J87 as template) and the DUB module (PDB: 6AQR). **b, c.** Representative regions illustrating the quality of the cryo-EM map and the high-resolution structural features. Cryo-EM map and atomic model showing that side chains are clearly identified. **b.** Histone-fold domain of Taf9. **c.** Part of the Taf5WD40 domain.

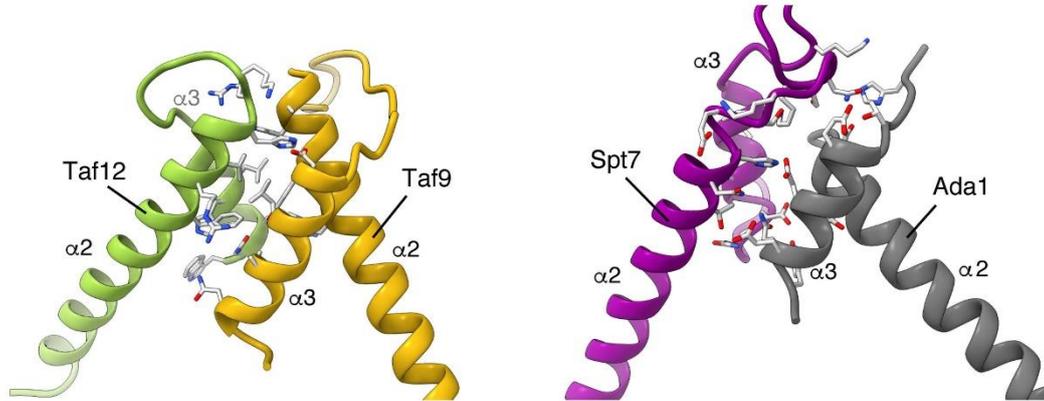




**Extended Data Fig. 4 | Comparing the arrangement of common SAGA and TFIID subunits.** **a**, The Taf5 NTD and Taf6 HEAT repeats have different positions in SAGA and in TFIID. The structures of Taf5 NTD, Taf6 HEAT repeats and Taf5 WD40 domain in SAGA (left) and in the twin lobe of TFIID<sup>19</sup> (right) are depicted. For the sake of clarity, all other parts of SAGA and TFIID were removed. To compare the position of the Taf5 NTD and Taf6 HEAT repeats between SAGA and TFIID, both complexes were aligned to have the same orientation of the Taf5 WD40 domain. The top face of the WD40 repeat, which is to a large extent free in TFIID, is nearly completely engaged in SAGA. The

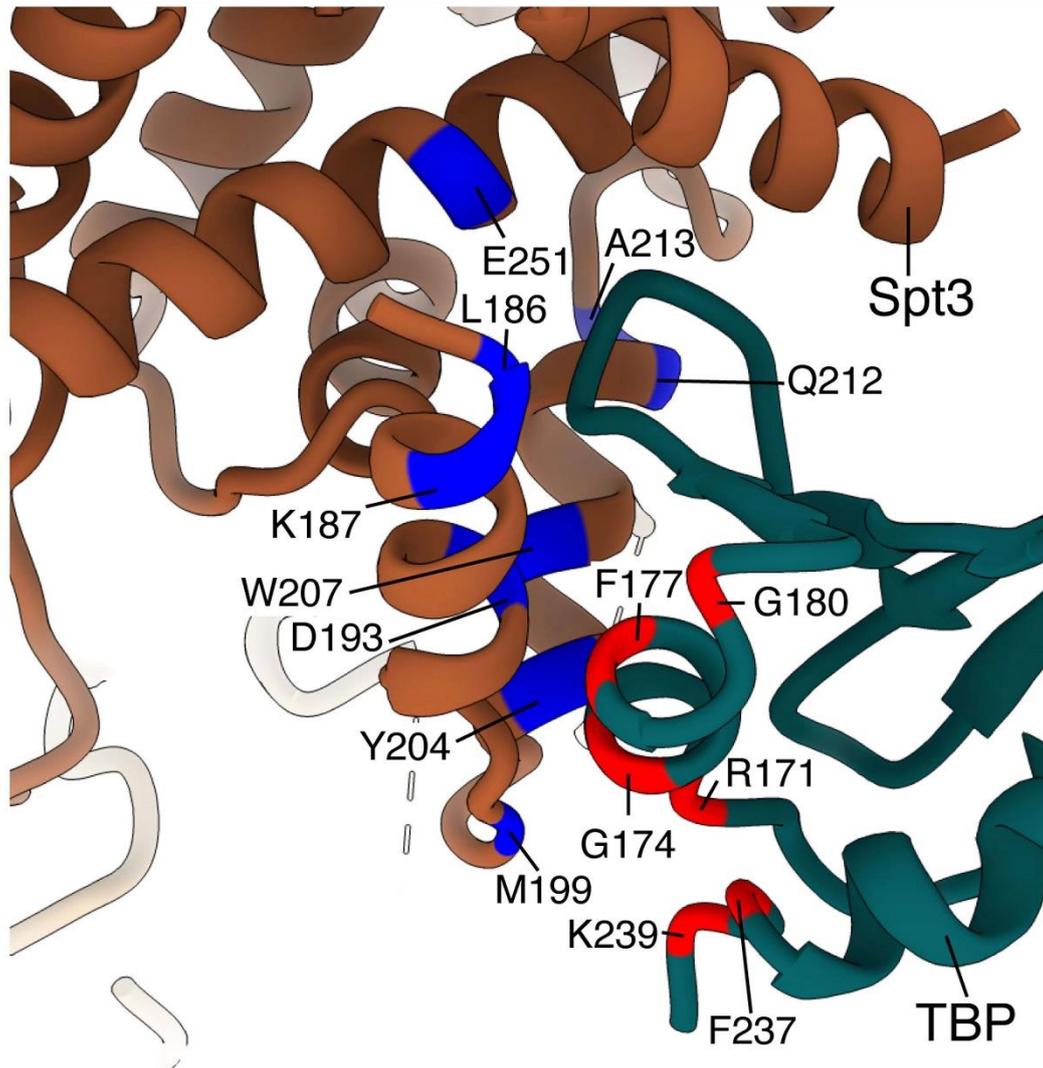
positions of the Taf5 NTD and Taf6 HEAT repeats are defined by direct interactions of the two domains between themselves, with the WD40 surface, and with the  $\alpha$ -helix A352–D364 of Taf5. **b**, SAGA and TFIID share five subunits (Taf5 and the four histone fold-containing subunits Taf6, Taf9, Taf10 and Taf12). These subunits are present as one and two copies in SAGA and TFIID, respectively. The relative positions of these subunits are depicted in both complexes. Note that in TFIID, the TBP-binding lobe was poorly resolved<sup>19,20</sup> and therefore the second copy of only the Taf6 HEAT repeats and Taf5 NTD is shown.

## Article



**Extended Data Fig. 5 | Interactions between histone-fold pairs in SAGA deviate from the nucleosome pattern already at the four-helix bundle preceding the weak association between Spt3 and Taf10.** The binding between the first two histone-fold pairs along the spiral (left), Taf6–Taf9 with Taf12–Ada1, shows little difference from the nucleosome, possibly reflecting the fact that these histone folds have the highest sequence similarity to

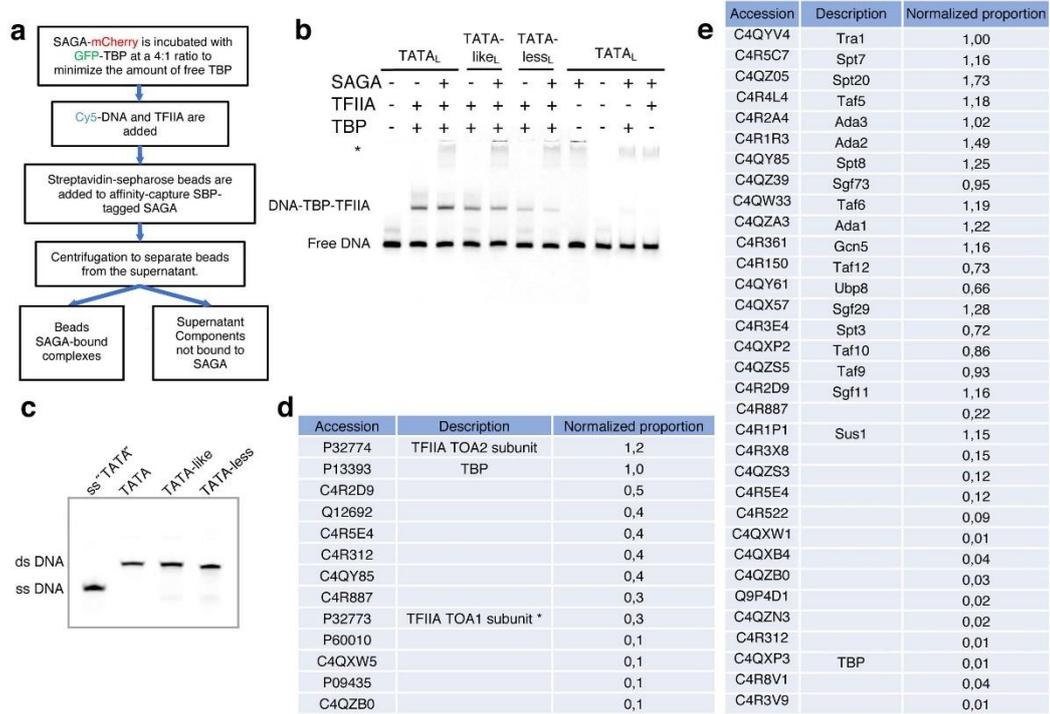
histones. As in the nucleosome analogue, hydrophobic interactions have a very important role in this binding. By contrast, the four-helix bundle preceding the association between Spt3 and the Spt7–Taf10 pair, linking Taf12–Ada1 to Spt7–Taf10, is maintained almost solely by hydrogen and electrostatic bonds. Only side chains of residues that participate in the interactions are shown.



**Extended Data Fig. 6 | Mapping previously isolated mutations in Spt3 and TBP.** Several classes of Spt3 and TBP mutations were shown to interact genetically with respect to transcription. Mutations from two major studies<sup>21,64</sup>

are mapped on the structure of Spt3 and TBP in the SAGA-TBP complex. The mutations appear to cluster at the contact sites between TBP and Spt3.

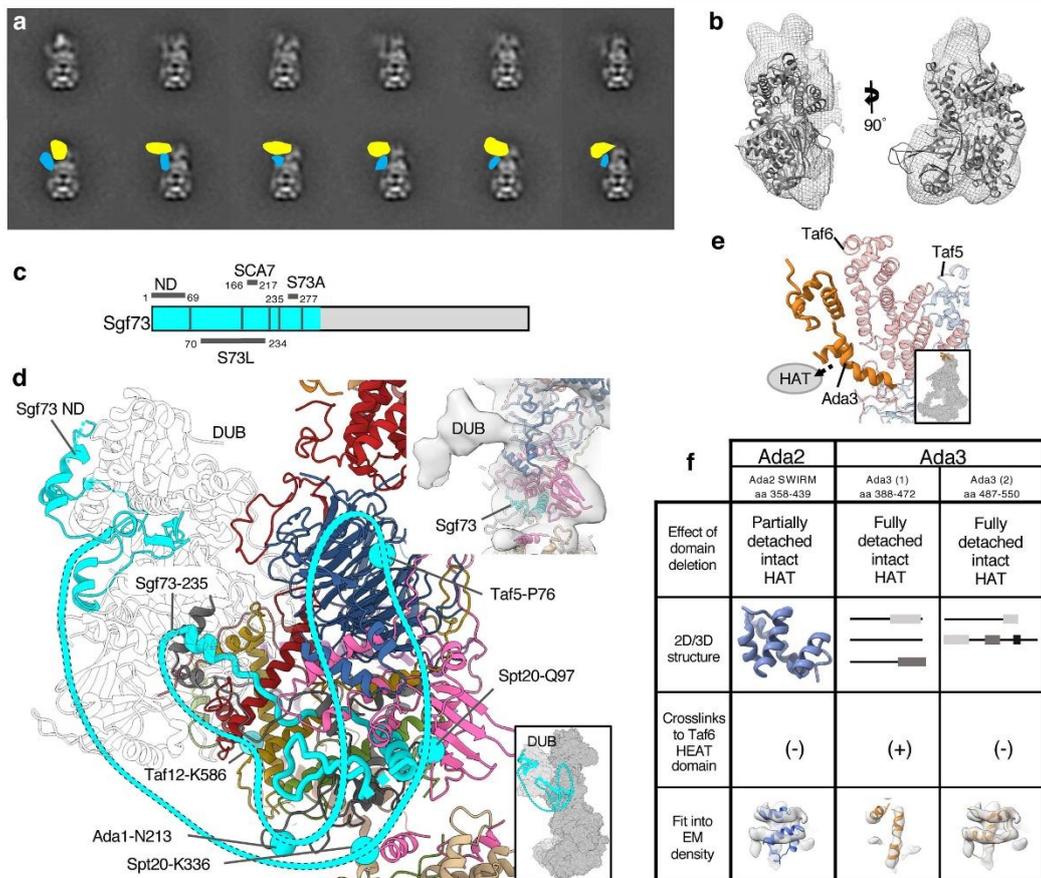
## Article



### Extended Data Fig. 7 | TFIIA-dependent TBP release from SAGA.

**a**, Experimental set-up for pull-down assays that investigate TBP and DNA interactions with SBP-tagged SAGA. **b**, Gel-shift assay as presented in Fig. 4b but using a longer (100-bp) Cy5-labelled DNA: TATA-containing promoter (TATA<sub>L</sub>), a TATA-like DNA probe (TATA-like<sub>L</sub>) and a non-specific DNA probe (TATA-less<sub>L</sub>). The asterisk indicates non-specific DNA association with SAGA. **c**, Cy5-labelled DNA was almost exclusively double-stranded. Annealed Cy5-labelled DNA used in gel-shift and pull-down assays was loaded on a polyacrylamide 12% gel. Single stranded Cy5-labelled TATA DNA was loaded for

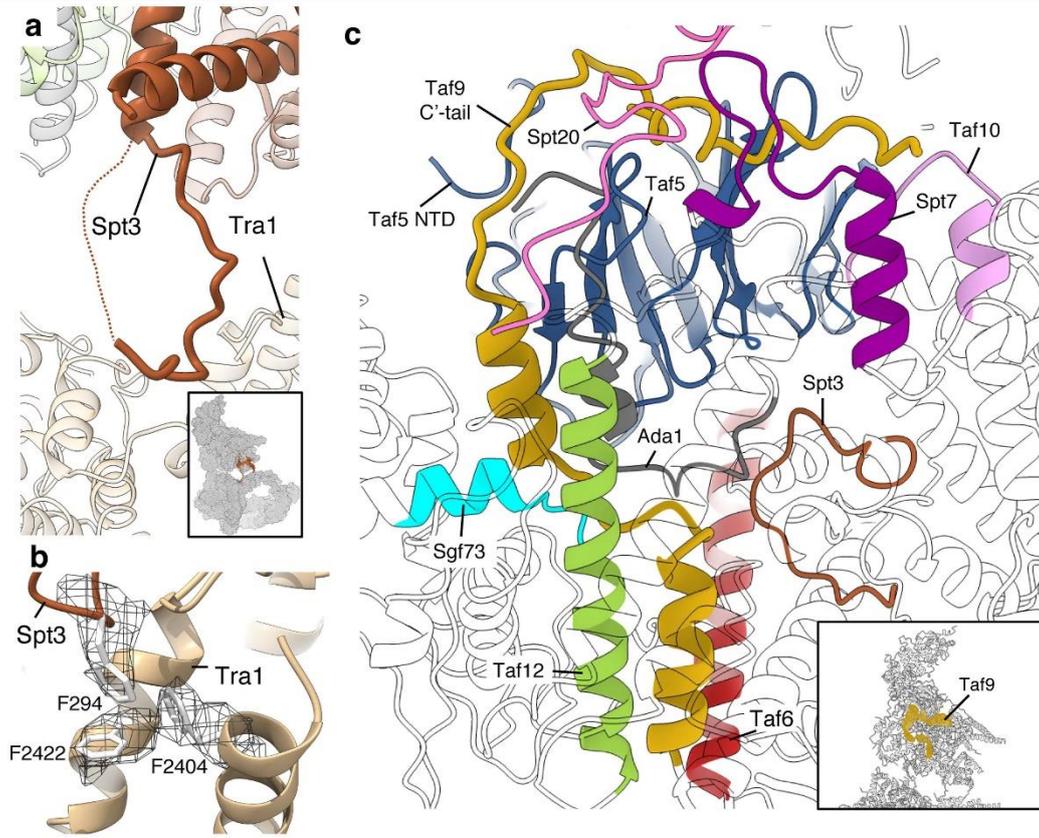
comparison. **d**, Summary of mass-spectrometry data for the major shifted DNA band (Fig. 4, lane 3) showing that it contains both TBP and TFIIA as its major constituents. The spectral count divided by the length in amino acids (PSM/AA) of each protein was normalized with respect to PSM/AA of TBP (raw data are provided in Supplementary Table 1). **e**, Summary of mass-spectrometry data of purified SAGA. To estimate stoichiometry, PSM/AA was calculated for each protein and normalized with respect to PSM/AA of subunit Tra1 (raw data are provided in Supplementary Table 2). Experiments were repeated three to four times.



**Extended Data Fig. 8 | Flexibility and anchoring of the enzymatic modules.**  
**a.** Two-dimensional classes of SAGA side views revealing different positions of the enzymatic HAT and DUB modules. Bottom, approximate positions of the modules are highlighted in yellow for the HAT module and in blue for the DUB module. The HAT module seems more diffuse and flexible than the DUB module, whose position fluctuates less. **b.** Docking of the DUB crystal structure (PDB: 6AQR) into the 15 Å resolution cryo-EM map reconstructed from a small class of particles that probably share a similar position for the DUB. **c.** Domain organization of the Sgf73 subunit. **d.** The Sgf73 linker tethers DUB to the central module. The linker (S73L) connecting the N-terminal end (ND), which is part of the DUB, and the anchoring domain (S73A) embedded in the central module, includes a SCA7 domain but additionally includes roughly 90 amino acids predicted to be unstructured. All residues that cross-link<sup>36</sup> to the Sgf73 linker can be mapped on the surface of SAGA. These residues are depicted with

light-blue spheres. A putative path for the 164-residue-long linker is delineated (dashed) according to the cross-linking sites and scarce traces of the linker found in the cryo-EM map. The top insert shows the density that forms contact between DUB and the central module. This contact point is located next to the Spt20 SEP and Taf5 LisH domains and is the only trace of the flexible Sgf73 linker in our maps. **e.** Two helical domains (ochre) anchor the HAT module at the surface of the Taf6 HEAT repeats (red). **f.** These helical domains are assigned to Ada3. Domain-deletion analysis identified two domains in Ada3 and one in Ada2 that serve to dock HAT on SAGA and whose deletion results in the release of intact HAT from SAGA<sup>36</sup>. Fit between the maps and secondary structure predictions (PSIPRED 3.3) or 3D structure of the Ada3 and Ada2 domains is presented in addition to available cross-linking data between these domains and the Taf6 HEAT repeats.

Article



**Extended Data Fig. 9 | Protein loops and tails contribute to an intricate network of interactions and to bridges between the central module and Tra1.** **a.** Bridge established by the loop connecting the second and third helices of the cSpt3 histone fold (brown). **b.** The loop between helices  $\alpha 2$  and  $\alpha 3$  of the cSpt3 histone fold inserts between two helices of Tra1. Aromatic stacking interactions are formed between Spt3 F294 of this loop and F2404 and F2422 of Tra1. The map density for these three phenylalanine residues is depicted. **c.** The

C-terminal tail of Taf9 (R101–L139) and the short preceding loop illustrate the importance of regions lacking secondary structure elements in SAGA. These parts of Taf9 interact with many subunits and domains, including the Spt3 C' tail, Spt20 bridge domain, Taf12 histone fold, Ada1, Sgf73 anchoring domain, Taf6 histone fold, Spt7 histone fold, Taf5 WD40, Taf5 NTD and Taf10 histone fold.

**Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics**

	SAGA EMD-10447	SAGA-TBP EMD-10438 (overall) EMD-10440 (Tra1) EMD-10441 (main lobe)	Modelled coordinates PDB 6TB4 *
<b>Data collection and processing</b>			
Magnification	105000	105000	
Voltage (kV)	300	300	
Electron exposure (e <sup>-</sup> /Å <sup>2</sup> )	52.8	52.8	
Defocus range (µm)	0.6-4.6	0.5-4.9	
Pixel size (Å)	1.09	1.09	
Symmetry imposed	C1	C1	
Initial particle images (no.)	426839	1068534	
Final particle images (no.)	218114	338850	
Map resolution (Å)			
Overall	5.92	3.80	
Main Lobe	4.87	3.52	
Tra1 Lobe	4.52	3.50	
FSC threshold	0.143	0.143	
Map resolution range (Å)			
<b>Refinement</b>			
Model composition			
Non-hydrogen atoms			39144
Protein residues			5144
Ligands			
R.m.s. deviations			
Bond lengths (Å)			0.008
Bond angles (°)			1.38
Validation			
MolProbity score			1.86
Clashscore			5.53
Poor rotamers (%)			0.46
Ramachandran plot			
Favored (%)			89.5
Allowed (%)			10.4
Disallowed (%)			0.16

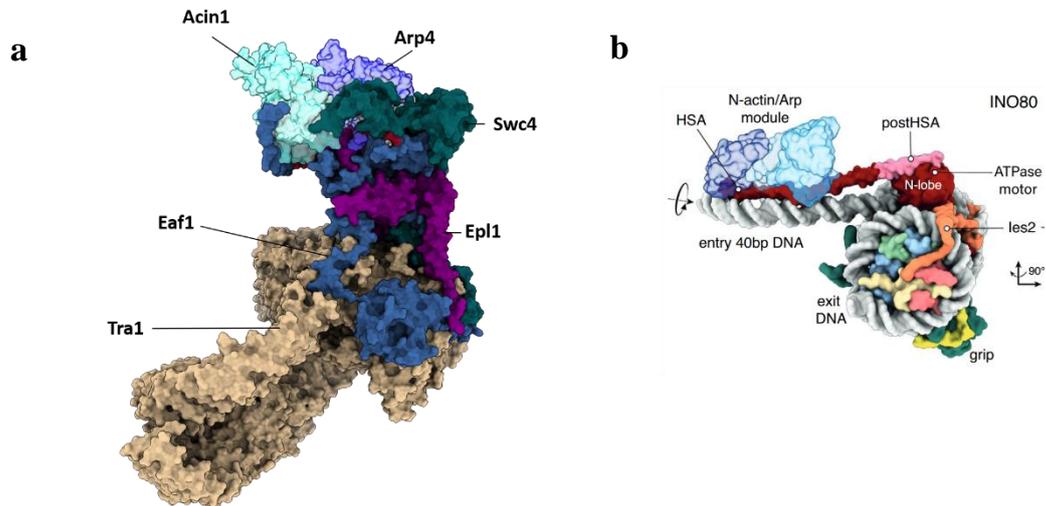
\*In addition a model including Spt8 and DUB fitted into lower-resolution maps was deposited under PDB accession code 6TBM.

## **IV. Discussion**



## A. NuA4

L'étude de la structure à haute résolution de NuA4 nous montre/révèle que le cœur stable du complexe est formé par : Eaf1, Epl1, Swc4, Act1, Arp4 et Tra1. Les interactions inter sous-unités ont pu également être étudiées, validant ainsi le rôle d'Eaf1 comme protéine organisatrice de NuA4.



**Figure 55 : Comparaison du module actine chez NuA4 et Ino80**

La structure de NuA4 (a) et (b) une structure composite de Ino80 obtenue en Cryo-EM et du module actine obtenue en cristallographie. Dans les deux complexes Actin1 et Arp4 sont colorés de la même manière pour faciliter la comparaison. Chez Ino80, le module Arp/act est situé en périphérie et interagit avec le linker du NCP par le biais du domaine HSA. Chez NuA4 le domaine HSA est complètement caché dans la structure rendant impossible toute liaison avec l'ADN.

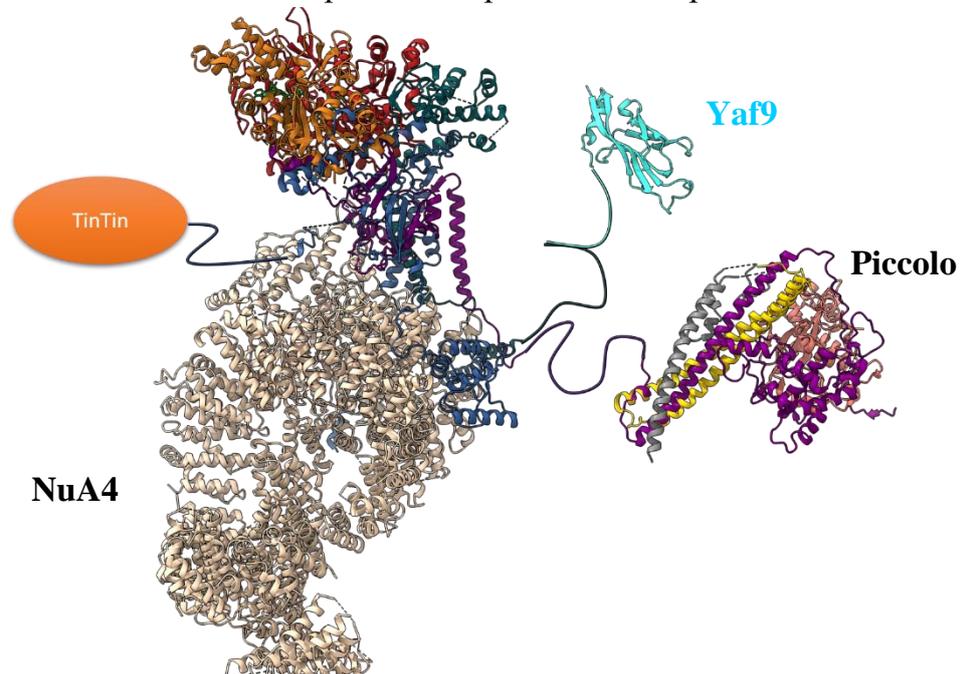
Chez Ino80 les protéines Act1, Arp4 et un domaine HSA sont impliqués dans l'interaction avec l'ADN **Figure 55**. La HSA vient reconnaître l'ADN linker du nucléosome et permet d'établir une communication allostérique avec le reste du complexe, notamment avec l'ATPase, afin de pouvoir entrainer le glissement du nucléosome (Jungblut et al., 2020). Chez Swi/Snf, Act1 et Arp4 sont remplacés par Arp7 et Arp9 qui son structurellement très similaires, mais le fonctionnement reste le même : la HSA et les deux Arp permettent l'interaction avec l'ADN linker du NCP pour permettre au complexe de réaliser sa tâche. Chez NuA4 il est peu probable que le domaine HSA d'Eaf1, Arp4 et Actin1 soit capable de reconnaître l'ADN. La HSA est fortement enfuie dans le complexe et est cachée par Epl1 et le reste d'Eaf1, empêchant

toute liaison à l'ADN **Figure 55**. Ceci peut suggérer qu'il existerait une autre fonction chez ce trimère de protéines dans NuA4. De plus le trimère Arp4/act1/HSA n'est pas implémenté de la même manière que chez Ino80. Arp4/Act1/HSA sont très flexibles et situés en périphérie du complexe Ino80, alors que chez NuA4 ils sont solidement ancrés dans le cœur du complexe, ce qui va dans le sens d'une fonction qui est autre, par rapport à Ino80.

Tous les modules/protéines possédant des domaines importants pour l'interaction avec la chromatine sont reliés à NuA4 par des régions désordonnées, les rendant trop flexibles pour les voir apparaître dans notre structure. Il est tentant de supposer que cette flexibilité facilite la reconnaissance et le recrutement du NCP, surtout qu'en essayant de réduire la taille de la région désordonnée reliant Piccolo à NuA4, nous obtenons un mutant dont la vitesse de croissance est réduite. En effet nous avons montré que le module catalytique Piccolo est attaché à NuA4 par la partie C-terminal d'Epl1, alors que sa partie N-terminal est intégrée à Piccolo. Les parties N-terminal et C-terminal d'Epl1 sont reliées par une région désordonnée d'une cinquantaine d'acides aminés. La délétion de cette région entraîne un phénotype différent du NuA4 WT chez *Pichia pastoris*, suggérant un rôle fonctionnel pour cette région. Cependant, notre système ne permet pas de comprendre avec exactitude cet aspect. Lors de l'introduction de la séquence mutée d'Epl1, l'ancienne séquence WT d'Epl1 est déplacée en aval et forme un ORF qui pourrait permettre la faible expression d'un Epl1 WT. Le fait que les levures soient toujours viables et ont une vitesse de croissance moins élevée pourrait donc être dû à la faible présence d'un NuA4 WT qui garantit une activité minimale permettant la survie de la levure, mais pas suffisante pour garantir un phénotype normal. Pour étudier de manière plus rigoureuse et précise l'aspect fonctionnel de la région désordonnée d'Epl1, une collaboration est en cours avec l'équipe de Dominique Helmlinger. Ses premiers résultats montrent que chez *S. cerevisiae* la délétion de la région désordonnée équivalente à celle identifiée chez *Pichia pastoris* n'entraîne aucun problème de croissance/viabilité. Cependant, ces résultats sont à remettre en perspective par la présence de Piccolo libre chez *S. cerevisiae*, ce qui n'est pas sûr pour *Pichia pastoris*. Chez *S. cerevisiae* il existe une faible quantité de Piccolo libre (non lié à NuA4) qui assure une activité d'acétylation basale rendant plus

difficile l'interprétation des résultats. Dans ce cas Piccolo n'est plus relié à NuA4 et sa flexibilité est « infinie », car il n'est plus attaché au complexe et l'altération due à la mutation ne va donc pas altérer l'activité et ceci garanti la survie et la croissance de la levure mutée. Nous avons cependant pu montrer que la longueur de la région entre le C-terminal et le N-terminal influe sur la croissance de la levure et la flexibilité du module Piccolo chez *Pichia pastoris*, car bien que nous n'arrivions pas à stabiliser le module Piccolo, le bruit dans la densité de la carte qu'il produit est déplacé par rapport au WT. De plus amples recherches sur cette boucle doivent être menées afin de comprendre son rôle qui pourrait être de donner une flexibilité au module HAT assurant la capacité de NuA4 à acétyler le NCP dans un contexte chromatidien.

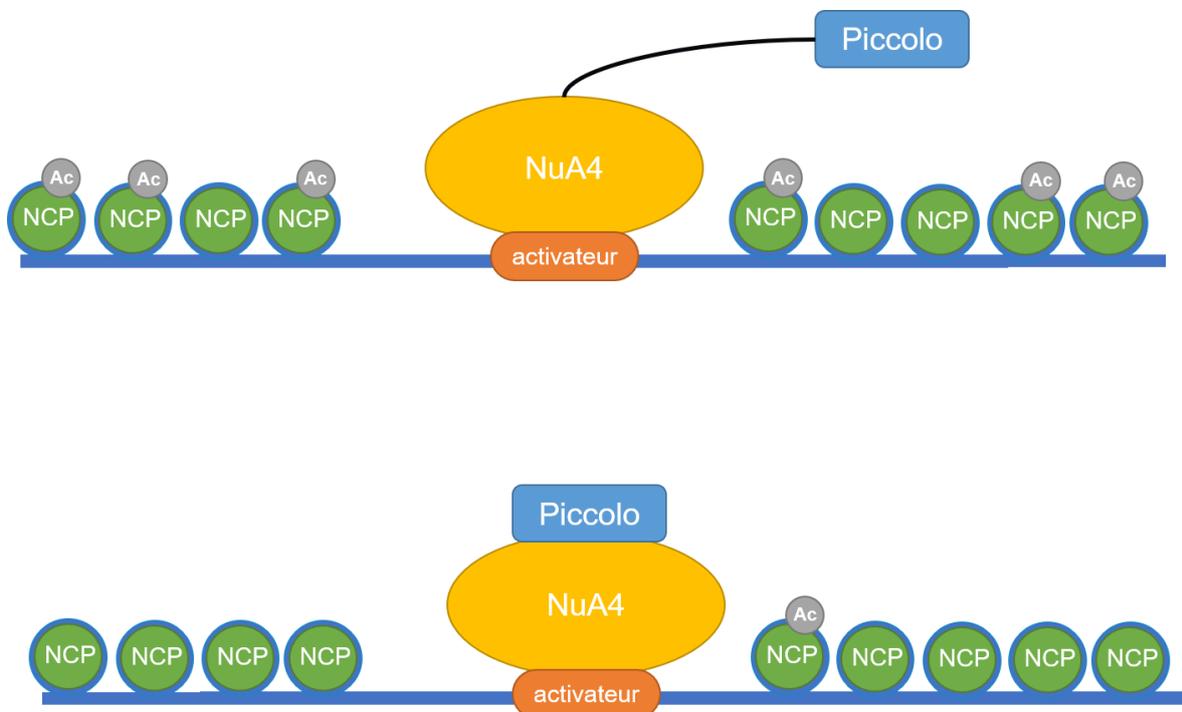
Dans la structure de 2018 du complexe TEEAA, les auteurs avaient réussi à isoler une faible population de particules qui comporterait Piccolo lié à NuA4. Nos résultats viennent contredire cette observation, en effet la partie C-terminal a été construite dans notre structure et les nombreuses interactions qu'effectue Epl1 avec NuA4 montrent qu'il est solidement ancré dans le complexe et ne peut surement pas s'en dissocier.



**Figure 56 : Modèle de l'organisation structurale de NuA4**

Le cœur stable de NuA4 servirait de plateforme d'ancrage aux différents modules, qui sont reliés par des régions désordonnées.

Le modèle actuel que nous proposons de NuA4 est un cœur structural bien ordonné composé de Eaf1, Swc4, Act1, Arp4, le C-terminal d'Epl1 et Tra1, dont le reste des sous-unités sont reliées par des régions désordonnées qui permettent l'interaction avec le NCP. Ainsi la figure **Figure 56** représente schématiquement à l'aide de notre modèle atomique de NuA4 la disposition relative des protéines/modules manquants grâce aux données biochimiques disponibles et la figure **Figure 57** comment la flexibilité du module Piccolo de NuA4 pourrait être importante pour acétyler le NCP. Ce modèle expliquerait pourquoi NuA4 est capable d'acétyler le génome sur de longues distances par rapport à son site de recrutement.

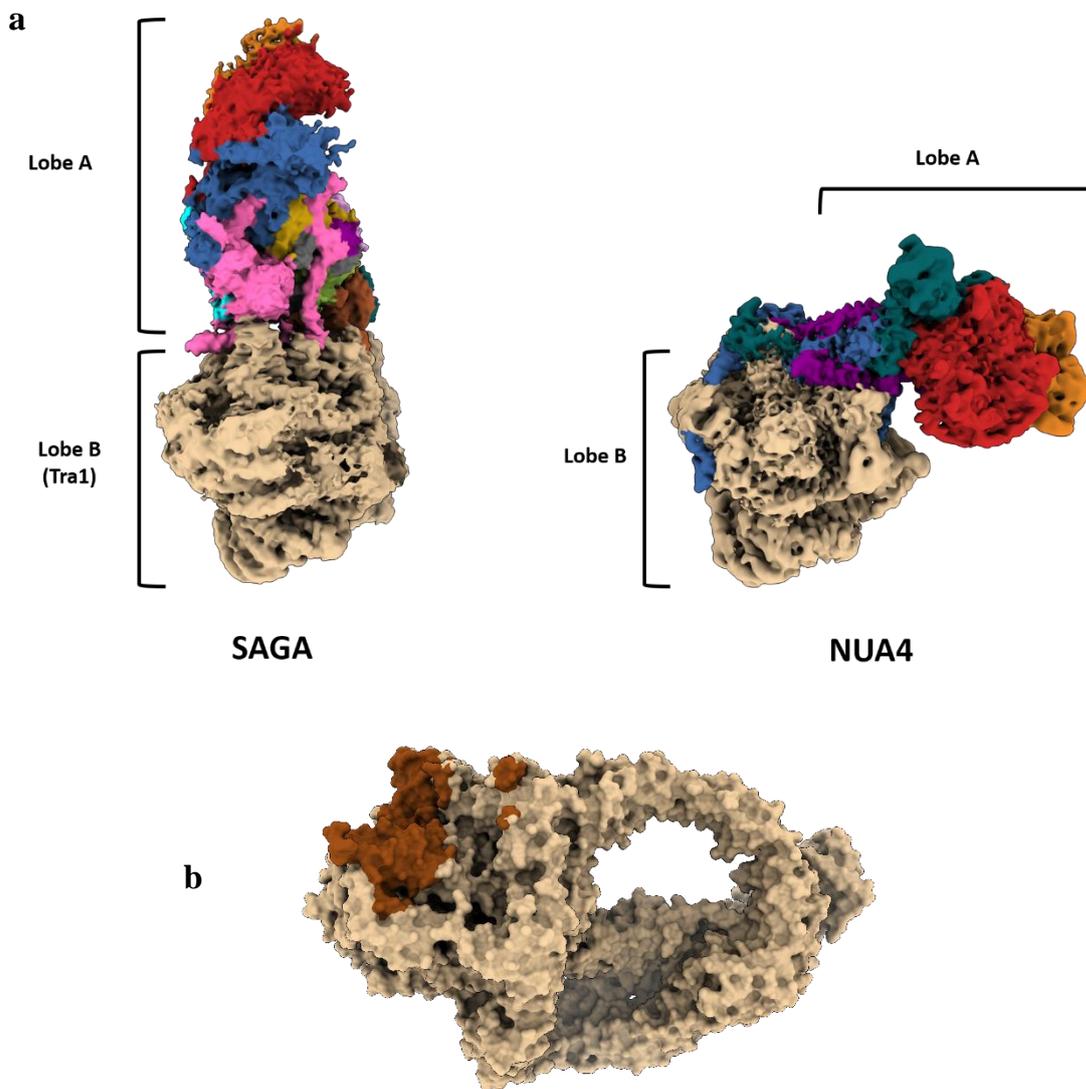


**Figure 57 : Modèle de fonctionnement de l'acétylation du NCP par NuA4 recruté par des activateurs de la transcription**

La flexibilité de Piccolo permettrait à NuA4 d'acétyler les NCP sur une longue distance. Si l'on vient à réduire cette flexibilité, Piccolo ne peut plus interagir avec des NCP éloignés.

## B. SAGA et NuA4

SAGA et NuA4 sont deux coactivateurs impliqués dans l'acétylation des histones. Les deux complexes possèdent tous les deux la protéine Tra1 qui permet l'interaction avec les activateurs. Nous avons obtenu des structures à haute résolution

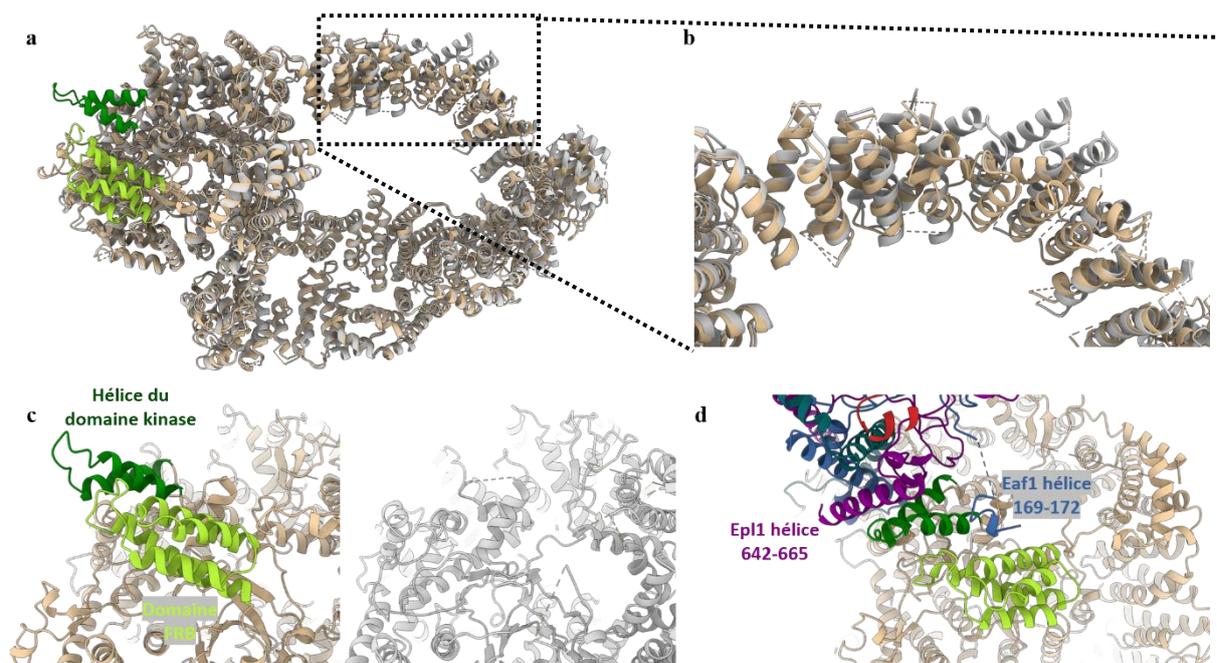


**Figure 58 : Comparaison des complexes SAGA et NuA4**

(a) Les deux complexes sont divisés en deux lobes A et B. Dans SAGA et NuA4 le lobe B est composé majoritairement de Tra1 colorié en beige dans les deux structures. Le lobe A est quant à lui composé du reste des sous-unités. Comparé à SAGA le lobe A de NuA4 est basculé de 90°. (b) La surface de Tra1 est affichée, les zones colorées en marron représentent la surface de Tra1 caché par le lobe A de NuA4 par rapport à SAGA.

des deux complexes qui nous permettent de comparer leurs incorporations avec Tra1 avec le reste de leurs sous-unités.

Dans les deux cas les deux coactivateurs possèdent chacun deux lobes, le lobe A contenant le reste des sous-unités des complexes et le lobe B qui lui est majoritairement composé de Tra1 **Figure 58 a**. L'orientation du lobe A de NuA4 par rapport à celui de SAGA est très différente. En prenant comme référence SAGA, le lobe A de NuA4 est incliné à 90° et possède une accroche différente de celui de SAGA **Figure 58 a**. Le fait que Tra1 soit implémenté de manière différente dans les deux complexes engendre que la surface accessible au solvant n'est pas la même **Figure 58 b**. Ceci pourrait permettre aux activateurs de faire une différence entre les deux complexes. Cette hypothèse permettrait d'expliquer le fait que SAGA et NuA4 ne soient pas recrutés sur les mêmes portions du génome. On peut également penser qu'une incorporation différente du lobe

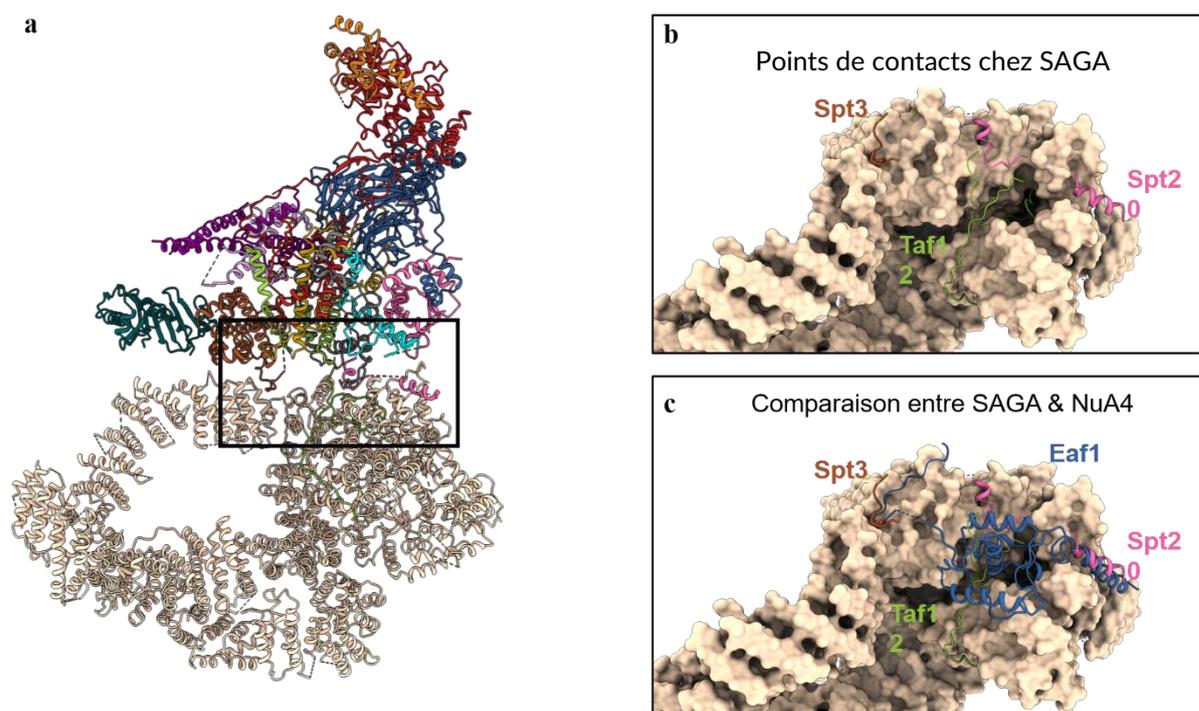


**Figure 59 : Comparaison de la structure de Tra1 chez SAGA et NuA4**

(a) La structure atomique de Tra1 et NuA4 sont superposées, la couleur beige est pour NuA4 et la grise pour SAGA. (b) Les HEAT situés en amont du domaine FAT divergent entre la structure de SAGA et de NuA4. (c) Des hélices du domaine kinase et du domaine FRB sont visibles chez NuA4, mais pas chez SAGA. (d) Chez NuA4 le domaine FRB et les hélices visibles sont stabilisés par les hélices d'Epl1 et de Eaf1. s

A entrainerait des encombrements stériques/sites d'interactions autres qui engendrent un comportement différent avec la chromatine.

Le RMSD entre le Tra1 de NuA4 et celui de SAGA est de 1 Å traduisant une structure globalement similaire, cependant il existe quelques changements dans deux régions de la protéine. La première est une variabilité de la position des HEAT situés avant le domaine FAT chez NuA4 par rapport à SAGA **Figure 59 a & b**. Cette région est faiblement résolue dans les deux complexes, mais encore plus chez NuA4. Nous expliquons la flexibilité de ces régions HEAT par l'hypothèse que du fait de la méthode de purification, des activateurs soient co-purifiés et entraînent peut-être des changements de conformation de Tra1 dans ces régions. La seconde zone à varier est le domaine FRB et deux hélices du domaine kinase situé entre les résidus 3558-3598.



**Figure 60 : Site de liaison avec Tra1 chez SAGA et comparaison avec NuA4**

(a) Le rectangle noire indique la zone d'interaction entre les deux lobes de SAGA. (b) La surface de Tra1 est affichée en beige, on peut observer les sites de liaisons de Spt20, Spt3 et Taf12 avec Tra1. (c) La sous-unité Eaf1 de NuA4 est affichée pour comparer ses interactions avec Tra1 par rapport à Spt20, Spt3 et Taf12, on peut clairement voir qu'Eaf1 vient recouvrir les mêmes sites de liaisons que Spt20, Spt3 et Taf12.

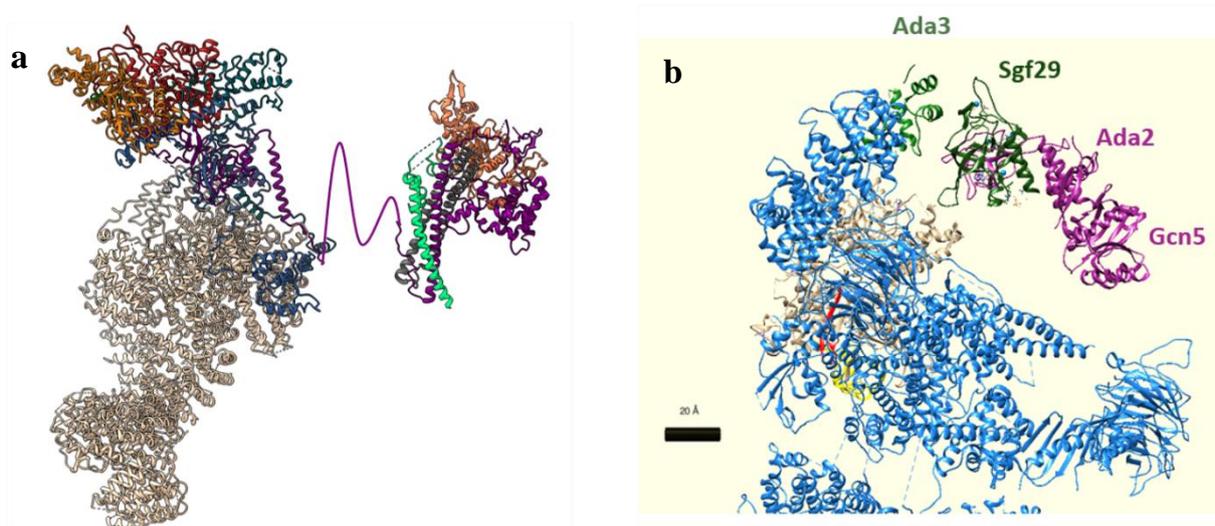
Dans le cas de SAGA, le FRB et les hélices ne sont pas visibles **Figure 59 c** et sont plus

flexibles que chez la levure *S. cerevisiae*, alors que dans NuA4 bien que faiblement résolu, il est possible de placer les éléments de structures secondaires **Figure 59 c**. Ceci est probablement dû aux hélices 643-665 de Epl1 et 167-172 d'Eaf1 qui stabilisent les deux hélices du domaine kinase et les hélices du domaine FRB **Figure 59 d**. Il est à noter que dans le cas de la structure de Tra1 chez *S. cerevisiae* cette région n'est pas flexible, il semblerait que ce soit spécifique à notre souche de levure.

Bien que Tra1 soit implémentée de manière très différente dans les 2 complexes, il s'avère que la protéine possède de nombreux points d'interactions similaires entre SAGA et NuA4. SAGA est accroché à Tra1 par le biais de 3 protéines qui sont Taf12, Spt20 et Spt3 qui viennent s'accrocher sur/près du domaine FAT de Tra1 **Figure 60 a & b**. Chez NuA4 les mêmes sites de liaison à Tra1 sont retrouvés, mais seulement par une seule sous-unité, qui est Eaf1 **Figure 60 c**. Le fait qu'Eaf1 recouvre les mêmes positions chez Tra1 que Taf12, Spt20 et Spt3 permet probablement d'empêcher l'hybridation de ces derniers avec le Tra1 de NuA4.

Chez les SAGA et NuA4 le module HAT est trop flexible pour être visible dans les structures obtenues, mais la flexibilité n'est pas induite de la même manière. Dans le cas de NuA4 c'est une région désordonnée qui permet l'extrême flexibilité de Piccolo **Figure 61 a**, alors que pour SAGA ce sont les différentes sous-unités composant le module catalytique qui créeraient cette flexibilité **Figure 61 b**. Chez SAGA à basse résolution il est possible de visualiser la région où est présent le module catalytique, alors que chez NuA4 seul du bruit engendré par ce dernier est visible. La différence de flexibilité pourrait être expliquée par le mode de fonctionnement des HAT des deux complexes. Gcn5, la sous-unité réalisant l'acétylation chez SAGA, est une enzyme dite processive, c'est-à-dire qu'une fois liée à son substrat qui est la queue d'une histone, elle va acétyler une première lysine puis une seconde, puis une troisième, jusqu'à la dernière avant de se dissocier. Esa1 est une enzyme dite dissociative : à chaque fois qu'elle réalise l'acétylation d'une lysine, elle se dissocie et doit donc se réassocier avant de pouvoir acétyler les autres lysines d'une même queue d'histones. Il paraît donc logique que dans le cas de SAGA la flexibilité pour permettre l'interaction avec le NCP est moins importante, car le complexe n'a besoin de reconnaître le NCP qu'une seule fois. Pour NuA4 c'est le contraire : pour acétyler toutes les lysines possibles il faudra

s'associer plusieurs fois au NCP et donc posséder un module HAT flexible permettant de faciliter la liaison au NCP, ce qui est important pour l'activité du complexe.



**Figure 61 : Comparaison des modèles de liaison des modules HAT de NuA4 et SAGA**

(a) Modèle de l'incorporation du module HAT dans NuA4, c'est une région désordonnée d'Ep11 qui engendre la flexibilité. (b) Modèle de l'incorporation du module HAT dans SAGA, ici la flexibilité est due à l'interaction de plusieurs protéines entre elles.



# **V. Conclusion & perspective**



Lors de ma thèse nous avons réussi à obtenir des structures à haute résolution des coactivateurs SAGA et NuA4. Ceci nous a permis de mieux comprendre l'architecture de ces complexes et d'établir des modèles atomiques. Pour SAGA c'est la première fois qu'il est possible d'obtenir une structure à haute résolution, cela représente une avancée majeure dans la compréhension de ses modes de fonctionnement, notamment sur son habilité à délivrer le facteur de transcription TBP. Pour NuA4 une structure à une résolution moyenne avait été publiée par une équipe concurrente lors du début de la thèse, mais la résolution n'était pas suffisante pour établir un modèle atomique précis et non ambiguë. La structure que nous avons obtenue représente la première structure à haute résolution et permet de mettre à jour les connaissances structurales sur le complexe et de corriger la précédente structure publiée. Nos résultats montrent une incroyable flexibilité des modules devant interagir avec la chromatine. Cette flexibilité pourrait posséder une importance fonctionnelle afin de permettre au complexe de trouver et d'acétyler son substrat qu'est le NCP dans le contexte de la chromatine. Cependant il reste de nombreuses étapes à réaliser afin de tester l'aspect fonctionnel des modules flexibles. Le fait que la réduction de flexibilité du module catalytique n'entraîne aucun phénotype chez la levure *S. cerevisiae*, mais une croissance plus lente chez *P. pastoris* est étrange et doit être plus étudié. La vérification de la présence ou absence d'un Piccolo libre chez *P. pastoris* permettrait d'expliquer cette différence et une étude sur ce sujet devrait être menée. Des tests sur l'habilité de NuA4 à acétyler la chromatine sur de grandes distances, suivant des mutations rendant son module catalytique plus ou moins flexible doivent être réalisées, afin de savoir si de telles mutations impactent la distance d'acétylation de NuA4. Nous avons tenté de réaliser une structure avec le NCP et NuA4, les résultats de coloration négative ont montré la faisabilité du projet, mais il est nécessaire d'optimiser les conditions de préparation de l'échantillon pour l'étude en cryo-EM. Maintenant que nous sommes capables d'obtenir des structures à haute résolution des coactivateurs SAGA et NuA4, il serait intéressant d'ajouter des activateurs aux échantillons et de visualiser leurs sites de liaison à Tra1, ce qui permettrait de mieux comprendre le mécanisme de recrutement de SAGA et NuA4 par les activateurs.





# **VI. Bibliographies**

- Amerik, A. Y., Nowak, J., Swaminathan, S., & Hochstrasser, M. (2000). The Doa4 deubiquitinating enzyme is functionally linked to the vacuolar protein-sorting and endocytic pathways. *Molecular Biology of the Cell*, *11*(10), 3365–3380. <https://doi.org/10.1091/mbc.11.10.3365>
- Asakura, S. (1961). The interaction between G-actin and ATP. *Archives of Biochemistry and Biophysics*, *92*(1), 140–149. [https://doi.org/10.1016/0003-9861\(61\)90228-4](https://doi.org/10.1016/0003-9861(61)90228-4)
- Auger, A., Galarneau, L., Altaf, M., Nourani, A., Doyon, Y., Utley, R. T., Cronier, D., Allard, S., & Cote, J. (2008). Eaf1 Is the Platform for NuA4 Molecular Assembly That Evolutionarily Links Chromatin Acetylation to ATP-Dependent Exchange of Histone H2A Variants. *Molecular and Cellular Biology*, *28*(7), 2257–2270. <https://doi.org/10.1128/mcb.01755-07>
- Baptista, T., Grünberg, S., Minoungou, N., Koster, M. J. E., Timmers, H. T. M., Hahn, S., Devys, D., & Tora, L. (2017). SAGA Is a General Cofactor for RNA Polymerase II Transcription. *Molecular Cell*, *68*(1), 130–143.e5. <https://doi.org/10.1016/j.molcel.2017.08.016>
- Barnes, C. E., English, D. M., & Cowley, S. M. (2019). Acetylation and Co: An expanding repertoire of histone acylations regulates chromatin and transcription. In *Essays in Biochemistry* (Vol. 63, Issue 1, pp. 97–107). Portland Press Ltd. <https://doi.org/10.1042/EBC20180061>
- Bassi, C., Li, Y. T., Khu, K., Mateo, F., Baniasadi, P. S., Elia, A., Mason, J., Stambolic, V., Pujana, M. A., Mak, T. W., & Gorrini, C. (2016). The acetyltransferase Tip60 contributes to mammary tumorigenesis by modulating DNA repair. *Cell Death and Differentiation*, *23*(7), 1198–1208. <https://doi.org/10.1038/cdd.2015.173>
- Bennett, G., & Peterson, C. L. (2015). SWI/SNF recruitment to a DNA double-strand break by the NuA4 and Gcn5 histone acetyltransferases. *DNA Repair*, *30*, 38–45. <https://doi.org/10.1016/j.dnarep.2015.03.006>
- Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J., & Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*, *16*(11), 1153–1160. <https://doi.org/10.1038/s41592-019-0575-8>
- Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., & Schwede, T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports*, *7*(1), 1–15. <https://doi.org/10.1038/s41598-017-09654-8>
- Bian, C., Xu, C., Ruan, J., Lee, K. K., Burke, T. L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B. O., Fleharty, B. E., Paulson, A., Allali-Hassani, A., Zhou, J. Q., Mer, G., Grant, P. A., Workman, J. L., Zang, J., & Min, J. (2011). Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO Journal*, *30*(14), 2829–2842. <https://doi.org/10.1038/emboj.2011.193>
- Birck, C., Poch, O., Romier, C., Ruff, M., Mengus, G., Lavigne, A. C., Davidson, I., & Moras, D. (1998). Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. *Cell*, *94*(2), 239–249. [https://doi.org/10.1016/S0092-8674\(00\)81423-3](https://doi.org/10.1016/S0092-8674(00)81423-3)
- Boudreault, A. A., Cronier, D., Selleck, W., Lacoste, N., Utley, R. T., Allard, S., Savard, J., Lane, W. S., Tan, S., & Côté, J. (2003). Yeast Enhancer of Polycomb defines global Esa1-dependent acetylation of chromatin. *Genes and Development*, *17*(11), 1415–1428. <https://doi.org/10.1101/gad.1056603>
- Brahma, S., Ngubo, M., Paul, S., Udugama, M., & Bartholomew, B. (2018). The Arp8 and Arp4 module acts as a DNA sensor controlling INO80 chromatin remodeling. *Nature Communications*, *9*(1), 3309. <https://doi.org/10.1038/s41467-018-05710-7>

- Brenner, S., & Horne, R. W. (1959). A negative staining method for high resolution electron microscopy of viruses. *BBA - Biochimica et Biophysica Acta*, 34(C), 103–110. [https://doi.org/10.1016/0006-3002\(59\)90237-9](https://doi.org/10.1016/0006-3002(59)90237-9)
- Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research*, 47(W1), W402–W407. <https://doi.org/10.1093/nar/gkz297>
- Cao, T., Sun, L., Jiang, Y., Huang, S., Wang, J., & Chen, Z. (2016). Crystal structure of a nuclear actin ternary complex. *Proceedings of the National Academy of Sciences*, 113(32), 8985–8990. <https://doi.org/10.1073/pnas.1602818113>
- Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P., & Workman, J. L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, 123(4), 581–592. <https://doi.org/10.1016/j.cell.2005.10.023>
- Cheng, X., & Côté, J. (2014). A new companion of elongating RNA polymerase II: TINTIN, an independent sub-module of NuA4/TIP60 for nucleosome transactions. *Transcription*, 5(5). <https://doi.org/10.1080/21541264.2014.995571>
- Chevillard-briet, M., Quaranta, M., Grézy, A., Mattera, L., Courilleau, C., Philippe, M., Mercier, P., Corpet, D., Lough, J., Ueda, T., Fukunaga, R., Trouche, D., & Escaffit, F. (2014). Interplay between chromatin-modifying enzymes controls colon cancer progression through Wnt signaling. *Human Molecular Genetics*, 23(8), 2120–2131. <https://doi.org/10.1093/hmg/ddt604>
- Croll, T. I. (2018). ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D: Structural Biology*, 74(6), 519–530. <https://doi.org/10.1107/S2059798318002425>
- Daniel, J. A., Torok, M. S., Sun, Z. W., Schieltz, D., Allis, C. D., Yates, J. R., & Grant, P. A. (2004). Deubiquitination of Histone H2B by a Yeast Acetyltransferase Complex Regulates Transcription. *Journal of Biological Chemistry*, 279(3), 1867–1871. <https://doi.org/10.1074/jbc.C300494200>
- Dyer, P. N. et al. Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. *Methods Enzymol.* 375, 23–44 (2003).
- Díaz-Santín, L. M., Lukyanova, N., Aciyan, E., & Cheung, A. C. M. (2017). Cryo-EM structure of the SAGA and NuA4 coactivator subunit Tra1 at 3.7 angstrom resolution. *ELife*, 6, 1–20. <https://doi.org/10.7554/eLife.28384>
- Doyon, Y., & Côté, J. (2004). The highly conserved and multifunctional NuA4 HAT complex. *Current Opinion in Genetics and Development*, 14(2), 147–154. <https://doi.org/10.1016/j.gde.2004.02.009>
- Dynlacht, B. D., Hoey, T., & Tjian, R. (1991). Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell*, 66(3), 563–576. [https://doi.org/10.1016/0092-8674\(81\)90019-2](https://doi.org/10.1016/0092-8674(81)90019-2)
- Eisenmann, D. M., Arndt, K. M., Ricupero, S. L., Rooney, J. W., & Winston, F. (1992). SPT3 interacts with TFIID to allow normal transcription in *Saccharomyces cerevisiae*. *Genes and Development*, 6(7), 1319–1331. <https://doi.org/10.1101/gad.6.7.1319>
- Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60(12 I), 2126–2132. <https://doi.org/10.1107/S0907444904019158>
- Erickson, H. P., & Lug, A. K. (1971). Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 261(837), 105–118.

<https://doi.org/10.1098/rstb.1971.0040>

- Frechard, A., Sharov, G., Werderer, M., & Schultz, P. (2021). Optimization of Sample Preparation for the Observation of Macromolecular Complexes by Electron (cryo-)Microscopy. In *Methods in Molecular Biology* (Vol. 2247, pp. 243–256). Humana Press Inc. [https://doi.org/10.1007/978-1-0716-1126-5\\_13](https://doi.org/10.1007/978-1-0716-1126-5_13)
- Gavin, A. C., Bösch, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., ... Superti-Furga, G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, *415*(6868), 141–147. <https://doi.org/10.1038/415141a>
- Gerhold, C. B., Winkler, D. D., Lakomek, K., Seifert, F. U., Fenn, S., Kessler, B., Witte, G., Luger, K., & Hopfner, K. P. (2012). Structure of Actin-related protein 8 and its contribution to nucleosome binding. *Nucleic Acids Research*, *40*(21), 11036–11046. <https://doi.org/10.1093/nar/gks842>
- Grant, P. A., Duggan, L., Côté, J., Roberts, S. M., Brownell, J. E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C. D., Winston, F., Berger, S. L., & Workman, J. L. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: Characterization of an ada complex and the saga (spt/ada) complex. *Genes and Development*, *11*(13), 1640–1650. <https://doi.org/10.1101/gad.11.13.1640>
- Grant, P. A., Schieltz, D., Pray-Grant, M. G., Steger, D. J., Reese, J. C., Yates, J. R., & Workman, J. L. (1998). A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. *Cell*, *94*(1), 45–53. [https://doi.org/10.1016/S0092-8674\(00\)81220-9](https://doi.org/10.1016/S0092-8674(00)81220-9)
- Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguët, L., Potier, N., Van-Dorsselaer, A., Wurtz, J. M., Mandel, J. L., Tora, L., & Devys, D. (2004). Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Human Molecular Genetics*, *13*(12), 1257–1265. <https://doi.org/10.1093/hmg/ddh139>
- Henry, K. W., Wyce, A., Lo, W. S., Duggan, L. J., Emre, N. C. T., Kao, C. F., Pillus, L., Shilatifard, A., Osley, M. A., & Berger, S. L. (2003). Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes and Development*, *17*(21), 2648–2663. <https://doi.org/10.1101/gad.1144003>
- Hickman, M. A., Froyd, C. A., & Rusche, L. N. (2011). Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage. In *Eukaryotic Cell* (Vol. 10, Issue 9, pp. 1183–1192). American Society for Microbiology Journals. <https://doi.org/10.1128/EC.05123-11>
- Hoke, S. M. T., Irina Mutiu, A., Genereaux, J., Kvas, S., Buck, M., Yu, M., Gloor, G. B., & Brandl, C. J. (2010). Mutational analysis of the C-terminal FATC domain of *Saccharomyces cerevisiae* Tra1. *Current Genetics*, *56*(5), 447–465. <https://doi.org/10.1007/s00294-010-0313-3>
- Huang, J., & Tan, S. (2013). Piccolo NuA4-Catalyzed Acetylation of Nucleosomal Histones: Critical Roles of an Esa1 Tudor/Chromo Barrel Loop and an Epl1 Enhancer of Polycomb A (EPcA) Basic Region. *Molecular and Cellular Biology*, *33*(1), 159–169. <https://doi.org/10.1128/mcb.01131-12>
- Joshi, A. A., & Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to pol II elongation. *Molecular Cell*, *20*(6), 971–978. <https://doi.org/10.1016/j.molcel.2005.11.021>
- Jungblut, A., Hopfner, K. P., & Eustermann, S. (2020). Megadalton chromatin remodelers: common principles for versatile functions. In *Current Opinion in Structural Biology* (Vol.

- 64, pp. 134–144). Elsevier Ltd. <https://doi.org/10.1016/j.sbi.2020.06.024>
- Kapoor, P., Chen, M., Winkler, D. D., Luger, K., & Shen, X. (2013). Evidence for monomeric actin function in INO80 chromatin remodeling. *Nature Structural and Molecular Biology*, *20*(4), 426–432. <https://doi.org/10.1038/nsmb.2529>
- Keogh, M. C., Kurdistani, S. K., Morris, S. A., Ahn, S. H., Podolny, V., Collins, S. R., Schuldiner, M., Chin, K., Punna, T., Thompson, N. J., Boone, C., Emili, A., Weissman, J. S., Hughes, T. R., Strahl, B. D., Grunstein, M., Greenblatt, J. F., Buratowski, S., & Krogan, N. J. (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell*, *123*(4), 593–605. <https://doi.org/10.1016/j.cell.2005.10.025>
- Knutson, B. A., & Hahn, S. (2011). Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and NuA4 complexes. *Molecular and Cellular Biology*, *31*(4), 818–831. <https://doi.org/10.1128/MCB.00687-10>
- Köhler, A., Pascual-García, P., Llopis, A., Zapater, M., Posas, F., Hurt, E., & Rodríguez-Navarro, S. (2006). The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5 acetyltransferase-mediated H2B deubiquitinylation through its interaction with Ubp8 and Sgf11. *Molecular Biology of the Cell*, *17*(10), 4228–4236. <https://doi.org/10.1091/mbc.E06-02-0098>
- Köhler, A., Schneider, M., Cabal, G. G., Nehrbass, U., & Hurt, E. (2008). Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export. *Nature Cell Biology*, *10*(6), 707–715. <https://doi.org/10.1038/ncb1733>
- Kuo, M. H., Brownell, J. E., Sobel, R. E., Ranalli, T. A., Cook, R. G., Edmondson, D. G., Roth, S. Y., & Allis, C. D. (1996). Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature*, *383*(6597), 269–272. <https://doi.org/10.1038/383269a0>
- Laprade, L., Rose, D., & Winston, F. (2007). Characterization of new Spt3 and TATA-binding protein mutants of *Saccharomyces cerevisiae*: Spt3-TBP allele-specific interactions and bypass of Spt8. *Genetics*, *177*(4), 2007–2017. <https://doi.org/10.1534/genetics.107.081976>
- Lee, K. K., Swanson, S. K., Florens, L., Washburn, M. P., & Workman, J. L. (2009). Yeast Sgf73/Ataxin-7 serves to anchor the deubiquitination module into both SAGA and Slik(SALSA) HAT complexes. *Epigenetics & Chromatin*, *2*(1), 2. <https://doi.org/10.1186/1756-8935-2-2>
- Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A., & Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*, *10*(6), 584–590. <https://doi.org/10.1038/nmeth.2472>
- Luger, K., Dechassa, M. L., & Tremethick, D. J. (2012). New insights into nucleosome and chromatin structure: An ordered state or a disordered affair? In *Nature Reviews Molecular Cell Biology* (Vol. 13, Issue 7, pp. 436–447). NIH Public Access. <https://doi.org/10.1038/nrm3382>
- Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods in molecular biology* (Clifton, N.J.) vol. 119 1–16 (1999).
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. <https://doi.org/10.1038/38444>
- Maeshima, K., Ide, S., & Babokhov, M. (2019). Dynamic chromatin organization without the 30-nm fiber. In *Current Opinion in Cell Biology* (Vol. 58, pp. 95–104). Elsevier Ltd. <https://doi.org/10.1016/j.ceb.2019.02.003>

- Mattera, L., Escaffit, F., Pillaire, M. J., Selves, J., Tyteca, S., Hoffmann, J. S., Gourraud, P. A., Chevillard-Briet, M., Cazaux, C., & Trouche, D. (2009). The p400/Tip60 ratio is critical for colorectal cancer cell proliferation through DNA damage response pathways. *Oncogene*, *28*(12), 1506–1517. <https://doi.org/10.1038/onc.2008.499>
- McGinty, R. K., Henrici, R. C., & Tan, S. (2014). Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature*, *514*(7524), 591–596. <https://doi.org/10.1038/nature13890>
- Mengus, G., May, M., Jacq, X., Staub, A., Tora, L., Chambon, P., & Davidson, I. (1995). Cloning and characterization of hTAF(II)18, hTAF(II)20 and hTAF(II)28: Three subunits of the human transcription factor TFIID. *EMBO Journal*, *14*(7), 1520–1531. <https://doi.org/10.1002/j.1460-2075.1995.tb07138.x>
- Mohibullah, N., & Hahn, S. (2008). Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3. *Genes and Development*, *22*(21), 2994–3006. <https://doi.org/10.1101/gad.1724408>
- Olins, D. E., & Olins, A. L. (2003). Chromatin history: Our view from the bridge. In *Nature Reviews Molecular Cell Biology* (Vol. 4, Issue 10, pp. 809–814). European Association for Cardio-Thoracic Surgery. <https://doi.org/10.1038/nrm1225>
- Patel, A. B., Louder, R. K., Greber, B. J., Grünberg, S., Luo, J., Fang, J., Liu, Y., Ranish, J., Hahn, S., & Nogales, E. (2018). Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science*, *362*(6421). <https://doi.org/10.1126/science.aau8872>
- Powell, D. W., Weaver, C. M., Jennings, J. L., McAfee, K. J., He, Y., Weil, P. A., & Link, A. J. (2004). Cluster Analysis of Mass Spectrometry Data Reveals a Novel Component of SAGA. *Molecular and Cellular Biology*, *24*(16), 7249–7259. <https://doi.org/10.1128/mcb.24.16.7249-7259.2004>
- Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, *14*(3), 290–296. <https://doi.org/10.1038/nmeth.4169>
- Punjani, A., Zhang, H., & Fleet, D. J. (2020). Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods*, *17*(12), 1214–1221. <https://doi.org/10.1038/s41592-020-00990-8>
- Rivera-Calzada, A., López-Perrote, A., Melero, R., Boskovic, J., Muñoz-Hernández, H., Martino, F., & Llorca, O. (2015). Structure and assembly of the PI3K-like protein kinases (PIKKs) revealed by electron microscopy. *AIMS Biophysics*, *2*(2), 36–57. <https://doi.org/10.3934/biophy.2015.2.36>
- Rodríguez-Navarro, S., Fischer, T., Luo, M. J., Antúnez, O., Brettschneider, S., Lechner, J., Pérez-Ortín, J. E., Reed, R., & Hurt, E. (2004). Sus1, a Functional Component of the SAGA Histone Acetylase Complex and the Nuclear Pore-Associated mRNA Export Machinery. *Cell*, *116*(1), 75–86. [https://doi.org/10.1016/S0092-8674\(03\)01025-0](https://doi.org/10.1016/S0092-8674(03)01025-0)
- Rosenthal, P. B., & Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, *333*(4), 721–745. <https://doi.org/10.1016/j.jmb.2003.07.013>
- Rossetto, D., Cramet, M., Wang, A. Y., Steunou, A., Lacoste, N., Schulze, J. M., Côté, V., Monnet-Saksouk, J., Piquet, S., Nourani, A., Kobor, M. S., & Côté, J. (2014). Eaf5/7/3 form a functionally independent NuA4 submodule linked to RNA polymerase II-coupled nucleosome recycling. *The EMBO Journal*, *33*(12), 1397–1415. <https://doi.org/10.15252/emj.201386433>
- Rothbart, S. B., & Strahl, B. D. (2014). Interpreting the language of histone and DNA modifications. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* (Vol.

- 1839, Issue 8, pp. 627–643). Elsevier. <https://doi.org/10.1016/j.bbagr.2014.03.001>
- Samara, N. L., Datta, A. B., Berndsen, C. E., Zhang, X., Yao, T., Cohen, R. E., & Wolberger, C. (2010). Structural insights into the assembly and function of the SAGA deubiquitinating module. *Science*, 328(5981), 1025–1029. <https://doi.org/10.1126/science.1190049>
- Sanders, S. L., Jennings, J., Canutescu, A., Link, A. J., & Weil, P. A. (2002). Proteomics of the Eukaryotic Transcription Machinery: Identification of Proteins Associated with Components of Yeast TFIID by Multidimensional Mass Spectrometry. *Molecular and Cellular Biology*, 22(13), 4723–4738. <https://doi.org/10.1128/mcb.22.13.4723-4738.2002>
- Sathianathan, A., Ravichandran, P., Lippi, J. M., Cohen, L., Messina, A., Shaju, S., Swede, M. J., & Ginsburg, D. S. (2016). The Eaf3/5/7 Subcomplex Stimulates NuA4 Interaction with Methylated Histone H3 Lys-36 and RNA Polymerase II. *The Journal of Biological Chemistry*, 291(40), 21195–21207. <https://doi.org/10.1074/jbc.M116.718742>
- Scheres, S. H. W. (2012). A bayesian view on cryo-EM structure determination. *Journal of Molecular Biology*, 415(2), 406–418. <https://doi.org/10.1016/j.jmb.2011.11.010>
- Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics*, 21(2), 129–228. <https://doi.org/10.1017/S0033583500004297>
- Sermwittayawong, D., & Tan, S. (2006). SAGA binds TBP via its Spt8 subunit in competition with DNA: Implications for TBP recruitment. *EMBO Journal*, 25(16), 3791–3800. <https://doi.org/10.1038/sj.emboj.7601265>
- Setiaputra, D., Dalwadi, U., Ross, J. D., Yip, C. K., Ahmad, S., Steunou, A.-L., Côté, J., Lu, S., & Dong, M.-Q. (2018). Molecular architecture of the essential yeast histone acetyltransferase complex NuA4 redefines its multimodularity. *Molecular and Cellular Biology*, 38(9), e00570-17. <https://doi.org/10.1128/MCB.00570-17>
- Sharov, G., Voltz, K., Durand, A., Kolesnikova, O., Papai, G., Myasnikov, A. G., Dejaegere, A., Shem, A. Ben, & Schultz, P. (2017). Structure of the transcription activator target Tra1 within the chromatin modifying complex SAGA. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-01564-7>
- Shiota, M., Yokomizo, A., Masubuchi, D., Tada, Y., Inokuchi, J., Eto, M., Uchiumi, T., Fujimoto, N., & Naito, S. (2009). Tip60 promotes prostate cancer cell proliferation by translocation of androgen receptor into the nucleus. *The Prostate*, n/a-n/a. <https://doi.org/10.1002/pros.21088>
- Sterner, D. E., Belotserkovskaya, R., & Berger, S. L. (2002). SALSAs, a variant of yeast SAGA, contains truncated Spt7, which correlates with activated transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 99(18), 11622–11627. <https://doi.org/10.1073/pnas.182021199>
- Sterner, D. E., Grant, P. A., Roberts, S. M., Duggan, L. J., Belotserkovskaya, R., Pacella, L. A., Winston, F., Workman, J. L., & Berger, S. L. (1999). Functional Organization of the Yeast SAGA Complex: Distinct Components Involved in Structural Integrity, Nucleosome Acetylation, and TATA-Binding Protein Interaction. *Molecular and Cellular Biology*, 19(1), 86–98. <https://doi.org/10.1128/mcb.19.1.86>
- Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. In *Nature* (Vol. 403, Issue 6765, pp. 41–45). Nature. <https://doi.org/10.1038/47412>
- Sun, J., Paduch, M., Kim, S. A., Kramer, R. M., Barrios, A. F., Lu, V., Luke, J., Usatyuk, S., Kossiakoff, A. A., & Tan, S. (2018). Structural basis for activation of SAGA histone acetyltransferase Gcn5 by partner subunit Ada2. *Proceedings of the National Academy of Sciences of the United States of America*, 115(40), 10010–10015. <https://doi.org/10.1073/pnas.1805343115>
- Tegunov, D., & Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing

- with Warp. *Nature Methods*, 16(11), 1146–1152. <https://doi.org/10.1038/s41592-019-0580-y>
- Terwilliger, T. C. (2010). Rapid model building of  $\alpha$ -helices in electron-density maps. *Acta Crystallographica Section D: Biological Crystallography*, 66(3), 268–275. <https://doi.org/10.1107/S0907444910000314>
- Uprety, B., Sen, R., & Bhaumik, S. R. (2015). Eaf1p Is Required for Recruitment of NuA4 in Targeting TFIID to the Promoters of the Ribosomal Protein Genes for Transcriptional Initiation In Vivo. *Molecular and Cellular Biology*, 35(17), 2947–2964. <https://doi.org/10.1128/mcb.01524-14>
- Vignali, M., Steger, D. J., Neely, K. E., & Workman, J. L. (2000). Distribution of acetylated histones resulting from Gal4-VP16 recruitment of SAGA and NuA4 complexes. *EMBO Journal*, 19(11), 2629–2640. <https://doi.org/10.1093/emboj/19.11.2629>
- Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B., Schubert, E., Shaikh, T. R., Lill, P., Gatsogiannis, C., & Raunser, S. (2019). SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology*, 2(1), 1–13. <https://doi.org/10.1038/s42003-019-0437-z>
- Wang, X., Ahmad, S., Zhang, Z., Côté, J., & Cai, G. (2018). Architecture of the *Saccharomyces cerevisiae* NuA4/TIP60 complex. *Nature Communications*, 9(1), 1147. <https://doi.org/10.1038/s41467-018-03504-5>
- Wang, Z., Zhao, F., Peng, J., & Xu, J. (2011). Protein 8-class secondary structure prediction using conditional neural fields. *Proteomics*, 11(19), 3786–3792. <https://doi.org/10.1002/pmic.201100196>
- Warfield, L., Ranish, J. A., & Hahn, S. (2004). Positive and negative functions of the SAGA complex mediated through interaction of Spt8 with TBP and the N-terminal domain of TFIIA. *Genes and Development*, 18(9), 1022–1034. <https://doi.org/10.1101/gad.1192204>
- Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737–738. <https://doi.org/10.1038/171737a0>
- Xu, P., Li, C., Chen, Z., Jiang, S., Fan, S., Wang, J., Dai, J., Zhu, P., & Chen, Z. (2016). The NuA4 Core Complex Acetylates Nucleosomal Histone H4 through a Double Recognition Mechanism. *Molecular Cell*, 63(6), 965–975. <https://doi.org/10.1016/j.molcel.2016.07.024>
- Yamaguchi, Y., Narita, T., Inukai, N., Wada, T., & Handa, H. (2001). SPT Genes: Key Players in the Regulation of Transcription, Chromatin Structure and Other Cellular Processes. *Journal of Biochemistry*, 129(2), 185–191. <https://doi.org/10.1093/oxfordjournals.jbchem.a002842>
- Yang, J., & Zhang, Y. (2015). I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*, 43(W1), W174–W181.
- Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*, 193(1), 1–12. <https://doi.org/10.1016/j.jsb.2015.11.003>
- Zhao, Y., Lang, G., Ito, S., Bonnet, J., Metzger, E., Sawatsubashi, S., Suzuki, E., Le Guezennec, X., Stunnenberg, H. G., Krasnov, A., Georgieva, S. G., Schüle, R., Takeyama, K. I., Kato, S., Tora, L., & Devys, D. (2008). A TFTC/STAGA Module Mediates Histone H2A and H2B Deubiquitination, Coactivates Nuclear Receptors, and Counteracts Heterochromatin Silencing. *Molecular Cell*, 29(1), 92–101.
- Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., & Agard, D. A. (2017). MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. In *Nature Methods* (Vol. 14, Issue 4, pp. 331–332). Nature Publishing Group. <https://doi.org/10.1038/nmeth.4193>



## Chapitre dans un ouvrage imprimé

Frechard, A., Sharov, G., Werderer, M., & Schultz, P. (2021). **Optimization of Sample Preparation for the Observation of Macromolecular Complexes by Electron (cryo-)Microscopy**. In *Methods in Molecular Biology* (Vol. 2247, pp. 243–256). Humana Press Inc. [https://doi.org/10.1007/978-1-0716-1126-5\\_13](https://doi.org/10.1007/978-1-0716-1126-5_13)

## Articles de périodiques imprimés

Papai, G., Frechard, A., Kolesnikova, O., Crucifix, C., Schultz, P., & Ben-Shem, A. (2020). **Structure of SAGA and mechanism of TBP deposition on gene promoters**. *Nature*, 577(7792), 711–716. <https://doi.org/10.1038/s41586-020-1944-2>

## Communication dans un congrès

### **Structure of the transcriptional coactivator NuA4**

Alexandre Frechard, Adam Ben Shem, Olga Kolesnikova, Gabor Papai, Corine Crucifix, Patrick Schultz

LMB-IGBMC Graduate Life Science Symposium, 11-13 juillet 2018, Cambridge Royaume-Unis

### **Cryo-EM structure of the transcriptional coactivator NuA4**

Alexandre Frechard, Unistra Israel Académie des Sciences, Octobre 25 2018, Strasbourg France

### **Structure Cryo-EM du coactivateur transcriptionnel NuA4**

Alexandre Frechard, Adam Ben Shem, Olga Kolesnikova, Gabor Papai, Corine Crucifix, Patrick Schultz

XVI<sup>ème</sup> colloque de la Société Française des Microscopies 2-5 juillet 2019 Poitiers France

# Etude structurale des coactivateurs NuA4 & SAGA par cryo-microscopie électronique

## Résumé

La transcription est un mécanisme essentiel par lequel la cellule exprime l'information génétique contenue sur son ADN sous forme d'ARN afin qu'il soit par la suite traduit en protéine. Chez les eucaryotes, l'ADN est compacté en chromatine ce qui empêche la mise en place de la transcription. Il est nécessaire pour la cellule de réguler cette compaction, afin d'être capable de répondre au stress environnemental ou simplement d'exprimer les gènes nécessaires pour le métabolisme cellulaire. Pour ce faire, la cellule possède des coactivateurs qui sont des complexes protéiques recrutés en amont des gènes, sur leurs régions promotrices et qui permettent de réguler la compaction de l'ADN et par conséquent l'expression génique. NuA4 & SAGA sont deux coactivateurs impliqués dans l'acétylation du nucléosome et sont importants pour l'expression de nombreux gènes. Dans cette thèse, je présente une étude structurale où mon équipe et moi-même avons réussi pour la première fois à obtenir une structure à haute résolution de ces deux complexes grâce à la Cryo-microscopie. Ces nouvelles structures nous ont permis d'établir des modèles atomiques précis, nous renseignant sur les mécanismes fonctionnels et l'architecture de ces coactivateurs.

## Résumé en anglais

Transcription is an essential mechanism by which the cell expresses the genetic information contained in DNA as RNA so that it can later be translated into protein. In eukaryotes, the DNA is compacted into chromatin, which prevents transcription to occur. It is necessary for the cell to regulate this compaction, to be able to respond to environmental stress or simply to express the genes necessary for cellular metabolism. To do this, the cell has coactivators, which are protein complexes recruited upstream of the genes, on their promoter regions, and which allow the regulation of DNA compaction and therefore gene expression. NuA4 & SAGA are two coactivators that are involved in nucleosome acetylation and are important for the expression of many genes. In this thesis, I present a structural study where my team and I succeeded for the first time in obtaining a high-resolution structure of these two complexes using cryomicroscopy. These new structures obtained allowed us to establish precise atomic models, informing us about the functional mechanisms of these coactivators and their architectures.