Structural study of the transcriptional co-activator SAGA

Alexandre Durand

Université de Strasbourg

Ecole doctorale des Sciences de la Vie et de la Santé

Thèse

Présentée pour l'obtention du titre de

Docteur de l'Université de Strasbourg

Discipline : Aspects moléculaires et cellulaires de la Biologie

Par

Alexandre Durand

STRUCTURAL STUDY OF THE TRANSCRIPTIONAL CO-ACTIVATOR SAGA

Soutenue publiquement le 29 Avril 2014 devant le jury composé de :

Dr. Marc Timmers	Rapporteur
Dr. Patrick Bron	Rapporteur
Dr. Bruno Kieffer	Examinateur
Dr. Patrick Schultz	Directeur de thèse
Dr. Didier Devys	Membre invité

REMERCIEMENTS

To start, I would like to thank the members of the jury, Dr. Marc Timmers, Dr. Patrick Bron, Dr. Bruno Kieffer and Dr. Didier Devys, for accepting to comment and judge my work.

Je suis particulièrement reconnaissant à mon directeur de thèse, le Dr. Patrick Schultz, qui m'a accueilli dans son équipe, qui m'a formé à l'art de la microscopie électronique et qui m'a permis de mener à bien ce projet. Je le remercie pour ses conseils et son enthousiasme au quotidien.

Je remercie tous les membres de l'équipe de Patrick pour leur accueil, leurs conseils, leur disponibilité et leur bonne humeur : Merci en particulier à Gabor, Christine, Corinne, Grigory, Adam et Nicolas pour avoir partagé avec moi leurs connaissances, mais aussi de bons moments de convivialité qui permettent de surmonter tous les petits problèmes du quotidien. Je souhaite également beaucoup de succès à mon compagnon de thèse Grigory, dans la poursuite de son projet.

Ce projet a été réalisé au sein du Département de Biologie Structurale Intégrative, et j'ai eu la chance de pouvoir interagir avec de nombreuses personnes que je ne pourrais pas toutes citer ici. Je remercie en particulier Catherine Birck et tous les membres de la plateforme de biologie et génomique structurale, ainsi que Marc Ruff pour son aide et ses conseils. Je remercie Jean-François Ménétret et Alexander Myasnikov pour leur aide lors de longues sessions au microscope. Je souhaite également remercier le Dr. Anne-Catherine Dock-Bregeon, sans qui je n'aurai pas pu me lancer dans cette grande aventure que représente le travail de thèse. Merci également au Dr. Arnaud Poterszman, qui fut le premier à m'accueillir au sein du département.

Je remercie très sincèrement le Dr. Laszlo Tora et tous les membres de son équipe, qui m'ont beaucoup apporté par nos échanges et leurs conseils. Je souhaite remercier en particulier Jacques Bonnet, Didier Devys et également Marjorie Fournier pour leur aide et leur disponibilité.

Je remercie également Bertrand Séraphin et Céline Faux, pour leurs conseils et leur expertise. Un grand merci aux membres de la plateforme de protéomique, en particulier Adeline Page et Virginie Chavant.

J’ai une pensée pour tous mes amis qui m’ont précédé ici, et qui par leur exemple m’ont fortement influencé dans ma décision de commencer ce travail de thèse. Grâce à Wassim et nos innombrables discussions « profondes » autour d’un café, grâce à Chris et notre passion commune pour le ballon ovale, grâce à Martin, Nada, Serena, Massimo, Alastair, Laura et tous les autres, qui m’ont montré que les moments parfois un peu difficiles de ces années n’étaient pas insurmontables lorsque l’on était bien entouré, j’ai pu franchir le pas. Merci à vous tous.

Je dois également remercier mes parents et ma famille (et ma belle-famille !), qui m’ont permis d’arriver aussi loin, grâce à leur soutien et leurs encouragements. Je n’aurai probablement pas pu suivre ce chemin s’ils n’avaient pas été là pour mes premiers pas.

Pour finir, un immense merci à ma petite femme, Marie-Laure, qui m’a accompagnée, encouragée et soutenue au quotidien au cours de ces 4 années, et sans qui rien n’aurait pu être pareil. Merci infiniment pour tes efforts et ta patience, parfois sans limite !

ABBREVIATIONS

Ada	:	Alteration/Deficiency in activation
ATAC	:	Ada-Two-A-Containing
ATP	:	Adenosine triphosphate
Bp	:	base-pairs
BRE	:	TFIIB Recognition Element
BSA	:	Bovine Serum Albumin
CBB	:	Calmodulin Binding Buffer
CBP	:	Calmodulin-binding peptide
CEB	:	Calmodulin Elution Buffer
ChIP	:	Chromatin Immuno Precipitation
Cs	:	Spherical aberrations
CTD	:	C-terminal domain
CTF	:	Contrast Transfer Function
CWB	:	Calmodulin Washing Buffer
DCE	:	Downstream Core Element
DNA	:	Deoxyribonucleic acid
DPE	:	Downstream Promoter Element
DSIF	:	DRB Sensitivity Inducing Factor
DUB	:	Deubiquitination
FEG	:	Field-emission Gun
FRET	:	Fluorescence resonance energy transfer
FSC	:	Fourier Shell Correlation
Gcn	:	General control of amino acid biosynthesis
GTF	:	General Transcription Factor
H2B-Ub	:	Ubiquitination of H2B on lysine 123
H3K4-Me3/2	:	Di- or Tri-Methylation of Histone 3 Lysine 4
HAT	:	Histone Acetyl-transferase
HDAC	:	Histone deacetylase
HFD	:	Histone-fold Domain
HMT	:	Histone Methyl Transferase HMT
HRP	:	Horseradish Peroxidase
Inr	:	Initiator sequence
LS	:	Liquid Chromatography
MRA	:	Multi References Alignment

mRNA	:	Messenger RNA
MS	:	Mass Spectrometry
MS/MS	:	Tandem MS
MSA	:	Multivariate Statistical Analysis
MTE	:	Motif Ten Element
MudPIT	:	Multidimensional protein identification technology
NC2	:	Negative cofactor 2
NCP	:	Nucleosome Core Particule
NELF	:	Negative Elongation Factor
NER	:	Nucleotide Excision Repair
NMR	:	Nuclear Magnetic Resonance
NPC	:	Nuclear Pore Complex
NSAF	:	Normalized Spectral Abundance Factor
NTP	:	Nucleoside Tri-Phosphate
OD	:	Optical Density
ORF	:	Open Reading Frame
PAGE-SDS	:	Polyacrylamide Gel Electrophoresis in presence of Sodium Dodecyl Sulfate
PAP	:	Peroxidase anti-peroxidase
PCR	:	Polymerase Chain Reaction
PHD	:	Plant Homeo Domain
PIC	:	Pre-Initiation Complex
protA	:	protein A from *Staphylococcus aureus*
P-TEFb	:	Positive Transcription Elongation Factor b
RAP	:	RNA-pol II Associated Protein
RAR	:	Retinoic Acid Receptor
RNA pol	:	RNA polymerase
RNA	:	Ribonucleic acid
rRNA	:	Ribosomal RNA
RSC	:	Remodels Structure of Chromatin
RP	:	Regulatory Particle of the proteasome
RT	:	Room Temperature
SAGA	:	Spt-Ada-Gcn5 Acetyl-transferase
SALSA	:	SAGA altered Spt8 absent
SCA7	:	Spinocerebellar ataxia type 7
Sgf	:	SAGA-associated factor

SLIK	:	SAGA-like
SNR	:	Signal to noise ratio
Spt	:	Suppressor of Ty
STAGA	:	SPT3-TAFII31-GCN5-L acetyltransferase
TAF	:	TBP-associated factor
TAND	:	Taf N-terminal Domain
TAP	:	Tandem Affinity Purification
TBP	:	TATA-box Binding Protein
TCB	:	TEV Cleavage Buffer
TCR	:	Transcription-Coupled Repair
TEM	:	Transmission Electron Microscope
TEV	:	Tobacco Etch Virus
TRF	:	TBP-related factors
TFTC	:	TBP-free TAFIIs Complex
tRNA	:	Transfer RNA
TSS	:	Transcription Start Site
UAS	:	Upstream Activating Sequence
YPD	:	Yeast Peptone Dextrose
ZnF	:	Zinc-Finger

TABLE OF CONTENT

Remerciements 1

Abbreviations 3

Table of content 6

List of Figures 10

Résumé 13

Chapitre 1. Introduction 13

Chapitre 2. Matériel et Méthodes 17

1. Structure du complexe SAGA observé par microscopie électronique 17

2. Purification du complexe SAGA endogène de levure 19

Chapitre 3. Résultats 21

1. Structure du complexe SAGA par microscopie électronique 21

2. Localisation du module DUB et des protéines Spt7 et Spt8 22

3. La délétion du module DUB favorise le clivage de Spt7 24

4. Le module DUB n'est pas toujours présent au sein du complexe SAGA 25

5. Interactions du complexe SAGA avec le nucléosome 26

Chapitre 4. Conclusions et perspectives 27

Chapter 1. Introduction 30

1.1 - Initiation of transcription by RNA pol II 31

1.1.1 The Promoter of the protein coding genes 31

1.1.2 The Pre-Initiation Complex 33

1.1.3 Structure of the Pre-Initiation Complex 46

1.2 - The Chromatin structure and Regulation of transcription 48

1.3 - Role of the Activators in transcription 54

1.4 - The Transcriptional Co-activator SAGA 63

1.4.1 Conservation of the SAGA complex during evolution 65

1.4.2 A subset of TAFs plays a structural role in SAGA 67

1.4.3 Role of the Spt proteins in structural integrity and interaction with TBP. 69

1.4.4 SAGA is a coactivator required for transcriptional activation 71

1.4.5 The Histone Acetyltransferase Activity of Gcn5..... 73

1.4.6 The deubiquitination module of the SAGA complex 76

1.4.7 The Structure of the SAGA complex reveals its modular organization 80

1.5 - Aims of this work 82

Chapter 2. Material and Methods.....86

2.1 - Structure of macromolecular complexes determined by Electron Microscopy 86

2.1.1 Principle of imaging with a Transmission Electron Microscope 87

2.1.2 Specimen preparation..... 89

2.1.3 Protocols for preparation of carbon-coated grids for Electron Microscopy 92

2.1.4 Data acquisition: Low-Dose Electron Microscopy..... 94

2.1.5 Image processing 95

2.2 - Purification of the SAGA complex.....103

2.2.1 Strategy for the purification of the SAGA complex: the TAP-tag method 103

2.2.2 Generation of the yeast strains..... 105

2.2.3 Culture of yeast..... 106

2.2.4 Preparation of the yeast extract..... 108

2.2.5 Purification of the SAGA complex 110

2.2.6 Reconstitution of the complex nucleosomes/SAGA 111

2.3 - Protein sample analysis.....112

2.3.1 Electrophoresis..... 112

2.3.2 Western-blot 113

2.3.3 Mass Spectrometry: the multidimensional protein identification technology 114

Chapter 3. Results.....117

3.1 - Purification of the endogenous SAGA complex from S. cerevisiae 117

3.1.1 Methods of lysis for optimal conditions of purification 117

3.1.2 Optimization of purification conditions 118

3.2 - Characterization of the complex by PAGE-SDS and mass spectrometry 120

3.3 - Structure of the SAGA complex revealed by Electron Microscopy 124

3.3.1 Description of the structure and the different conformations of SAGA . 124

3.3.2 Localization of the DUB module..... 128

3.3.3 A subpopulation of SAGA does not contain the DUB module..... 132

3.3.4 Localization of the Spt7 and Spt8 subunits..... 134

3.4 - Structure of the fully-hydrated SAGA complex.....135

3.5 - Interactions of the SAGA complex with the nucleosome 136

Chapter 4. Discussion 140

4.1 - Architecture of the SAGA complex.....140

4.2 - The DUB module clusters with Gcn5 and defines a nucleosome binding interface 142

4.3 - The SAGA complex forms a molecular clamp which might bind TBP 143

4.4 - The dynamic of the DUB module alters the TBP-binding surface of the SAGA complex.....145

Chapter 5. Conclusions and Perspectives 148

Chapter 6. References 153

Chapter 7. Annexes 175

Annexe 1 - Protocols.....176

7.1 - S. cerevisiae culture 176

7.1.1 Media composition 176

7.1.2 Yeast culture in liquid media..... 176

7.2 - Purification of the SAGA complex 178

7.2.1 Buffers 178

7.2.2 Preparation of the yeast extract 178

7.2.3 TAP-tag purification of the complex 179

7.3 - Protein complex analysis 180

7.3.1 PAGE-SDS 180

7.3.2 Western-Blot 181

7.4 - Sample preparation for Electron Microscopy analysis 182

7.4.1 Preparation of carbon-coated grids for Electron Microscopy 182

Annexe 2 - Publications 183

7.5 - Publication 1 183

7.6 - Publication 2 183

LIST OF FIGURES

Figure 1 : Organization of eukaryotic promoter 33

Figure 2 : Different models for PIC assmebly. 35

Figure 3 : Structure of yeast TBP bound to the TATA-box 37

Figure 4 : Structure of TBP-DNA-TFIIA complex 38

Figure 5 : Domain composition of yeast TFIIB 38

Figure 6 : Domains of the two subunits forming TFIIF 40

Figure 7 : Domains composition of TFIIE subunits α and β 41

Figure 8 : Molecular organization of TFIIH 43

Figure 9: Molecular architecture of RNA pol II 44

Figure 10: Structure of TFIIB-TBP-RNA pol II complex with opened DNA template 45

Figure 11 : Comparison of PIC structure in yeast and human 47

Figure 12 : Structure and compaction of the chromatin 49

Figure 13 : Structure of the nucleosome 50

Figure 14 : Histone modifications are marker of the transcriptional state genes. 53

Figure 15 : Cross-talk between H2B ubiquitilation and methylation of histone 3 54

Figure 16 : Summary of Tafs present in the different species and their nomenclatures 56

Figure 17 : Molecular architecture of TFIID 58

Figure 18 : Subunits compostion of yeast and human Mediator complex. 60

Figure 19 : Molecular organization of yeast and human Mediator complex 60

Figure 20 : Mechanism of Gal4 activation upon galactose induction. 62

Figure 21 : Composition of the SAGA complex 65

Figure 22 : Conserved structure of SAGA homologues 67

Figure 23 : Comparison of Tafs position in SAGA and TFIID 69

Figure 24 : Molecular architecture and nucleosome interaction site of NuA4 73

Figure 25 : Structure of the DUB module revealed by x-ray crystallography 77

Figure 26 : Conservation of domain organization in Sgf73 and Sgf11 78

Figure 27 : Sus1 links transcription and mRNA export 79

Figure 28 : Molecular organization of yeast SAGA 81

Figure 29 : Modular organization of the SAGA complex 82

Figure 30 : Schematic representation of a Transmission Electron Microscope and path of the electrons thought the optical elements 88

Figure 31 : Sample preparation techniques for single particle EM. 90

Figure 32 : Schematic representation of image analysis process. 96

Figure 33 : Sampling rate of viewing direction limits the resolution 97

Figure 34 : Euler angles convention. 99

Figure 35 : Principle of Random Conical Tilt method 100

Figure 36 : Schematic representation of the Tandem Affinity Purification. 104

Figure 37: Summary of the different strains used in this study, and their correspondiong genotype. 106

Table 3 : Composition of the buffers used during purification 109

Figure 38 : Comparison of different lysis methods 118

Figure 39 : Optimization of conditions for purification 119

Figure 40 : Control of the quality of the sample by Electron Microscopy. 120

Figure 41 : PAGE-SDS analysis of SAGA complexes on a 4-15% acrylamide gel gradient. 121

Figure 42 : Peptide coverage of Spt7 and Sgf73. 122

Figure 43 : Mass spectrometry analysis of the purified complexes. 123

Figure 44 : Characteristic views of the SAGA complex 125

Figure 45: 3D reconstruction of the SAGA complex 126

Figure 46: Different conformations obtained by maximum likelihood 3D classification. 127

Figure 47 : Independent 3D reconstructions of the two separated lobes 128

Figure 48 : Size and position of domain V in the SAGA complex 129

Figure 49 : Difference between the complex from the different strains. 130

Figure 50 : Fitting of the DUB module structure in the density map 131

Figure 51: Additional densities in the *Sgf73Δ* mutant strain compared to wild-type 132

Figure 52 : Separation of a DUB-containing and a DUB-free SAGA complex..... 133

Figure 53 : Localization of the Spt8 subunit by immuno-labelling..... 134

Figure 54 : Class average images of SAGA observed by cryo-EM 136

Figure 55 : Control of the interaction betwwen the purifed SAGA complex and the mono-nucleosome..... 137

Figure 56 : Visualization of the complex formed between SAGA and mono-nucleosome..... 138

Figure 57 : Position of TBP-SAGA interacting regions..... 144

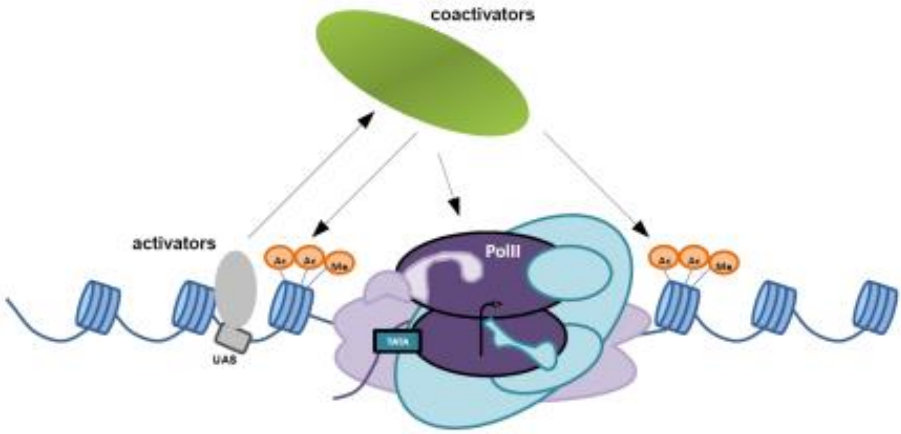
RÉSUMÉ

Chapitre 1. INTRODUCTION

La première étape de l'expression de l'information génétique contenue dans les gènes est la synthèse d'une molécule d'ARN à partir de la matrice d'ADN, lors d'un procédé appelé la transcription. Chez les eucaryotes, la transcription des gènes codant pour les protéines est réalisée par l'ARN polymérase II (pol II). Lors de la phase d'initiation de la transcription, l'ARN polymérase et des facteurs généraux de la transcription (GTF) vont être recrutés afin de former un complexe de pré-initiation (PIC). Les GTFs reconnaissent certains éléments spécifiques du promoteur proximal du gène, et permettent ainsi (i) la reconnaissance du site d'initiation de la transcription, (ii) le recrutement de la pol II et son positionnement précis au niveau du site de démarrage de la transcription au sein du promoteur du gène, et (iii) l'ouverture de la double hélice d'ADN au niveau du site d'initiation, étape nécessaire au démarrage de la synthèse de la molécule d'ARN.

La machinerie transcriptionnelle doit également passer la barrière physique formée par la structure compacte de la chromatine, et qui limite l'accessibilité de la matrice d'ADN. En effet, la chromatine est composée de la répétition d'un élément de base, le nucléosome, formé d'un octamère de protéines histones autour duquel s'enroule la molécule d'ADN. Le passage de la forme condensée de la chromatine à un état plus ouvert, et donc plus accessible, nécessite l'action de plusieurs classes de protéines, dont notamment les chaperonnes d'histone et les facteurs de remodelage de la chromatine.

La formation du PIC est une étape très importante, qui nécessite l'action coordonnée d'un grand nombre de molécules. En particulier, les activateurs de la transcription reconnaissent des séquences spécifiques au sein ou en dehors du promoteur, et stimulent la transcription en favorisant la formation du PIC. Plusieurs complexes macromoléculaires, appelés co-activateurs transcriptionnels, sont recrutés par ces activateurs et font le lien avec les différents composants de la machinerie transcriptionnelle.



Les coactivateurs font le lien entre les activateurs de la transcription, qui lient certaines séquences d'ADN spécifiques à proximité du gène, et la machinerie transcriptionnelle. Notamment, les coactivateurs (i) interagissent avec les activateurs liés à leur séquence régulatrice, (ii) interagissent avec les différents composants du PIC pour faciliter leur recrutement, et (iii) peuvent lire et écrire des modifications spécifiques au niveau des histones, permettant l'altération de la structure de la chromatine.

Le complexe SAGA (Spt-Ada-Gcn5 acetyl transferase) est l'un de ces co-activateurs, conservé chez les eucaryotes au cours de l'évolution. Chez la levure *Saccharomyces cerevisiae*, il a été montré que SAGA est requis pour la transcription d'environ 10% des gènes, la plupart étant impliqués dans la réponse de la cellule aux stresses environnementaux (Basehoar et al., 2004; Huisinga and Pugh, 2004). SAGA est un large complexe comportant au moins 19 sous-unités, pour un poids moléculaire total d'environ 1.8 MDa. Le complexe SAGA est impliqué dans la modification post-traductionnelle des histones dans le contexte du nucléosome, et en particulier dans l'acétylation des lysines sur l'histone H3. L'acétylation des histones est généralement associée à la transcription active des gènes, et cette activité est portée au sein du complexe SAGA par l'histone acétyl-transférase (HAT) Gcn5, en association avec les protéines Ada2, Ada3 et Sgf29 qui forme un module d'acétylation au sein du complexe (Balasubramanian et al., 2002; Grant et al., 1997; Lee et al., 2011). De plus, une deuxième activité enzymatique impliquée dans la modification des histones a été identifié au sein du complexe SAGA. En effet, la protéine Ubp8, associée aux sous-unités Sgf73, Sus1 et Sgf11, est impliquée dans la coupure de l'ubiquitine liée à la lysine 123 de l'histone H2B (Daniel et al., 2004; Ingvarsdottir et al., 2005; Köhler et al., 2006; Lee et al., 2009). Cette étape est nécessaire au passage de la phase d'initiation à la phase d'élongation de la transcription (Wyce et al., 2007). SAGA est recruté au niveau du promoteur du gène par certains activateurs de la transcription, et ce recrutement se fait au travers de la protéine Tra1 (Bhaumik et al., 2004; Fishburn et al., 2005; Grant et al., 1998b). SAGA est nécessaire pour la formation du PIC au niveau du promoteur d'un ensemble de gènes (Bhaumik and Green, 2002), par le recrutement direct de la protéine TBP qui lie certains éléments de ces promoteurs et permet ainsi l'assemblage du PIC. Ce recrutement implique les protéines Spt3 et Spt8 du complexe SAGA qui interagissent directement avec TBP

(Laprade et al., 2007; Mohibullah and Hahn, 2008; Warfield et al., 2004). Enfin, le complexe SAGA contient un sous-ensemble de protéines Tafs (Grant et al., 1998a), partagée avec le facteur de transcription TFIID, et qui semblent jouer un rôle structural au sein du complexe. Finalement, SAGA contient également les protéines Spt20, Spt7 et Ada1, dont la délétion compromet l'intégrité du complexe et semble nécessaire pour le maintien de son architecture (Grant et al., 1997; Horiuchi et al., 1997).

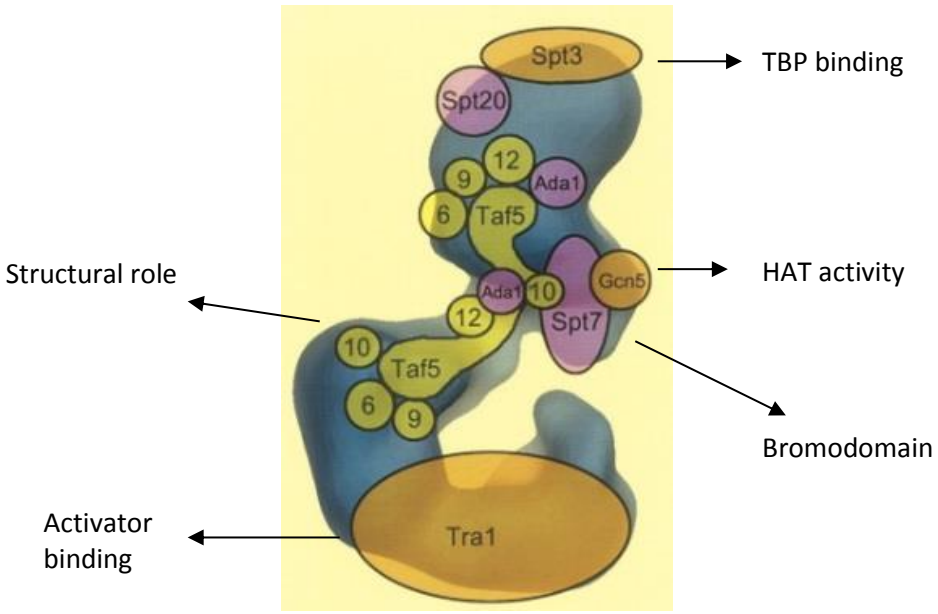
S. cerevisiae	Human	Drosophila	Function
Tra1	TRRAP	Tra1	Activator binding
Spt7	Spt7/SUPT7L/STAF65	Spt7	Structural role, bromodomain
Spt20	SPT20	Spt20	Structural role
Ada1	Tada1/STAF42	Ada1	Structural role, Histone Fold
Spt3	Spt3/SUPT3	Spt3	TBP binding
Spt8			TBP binding
Gcn5	GCN5L/KAT2A	Gcn5	Histone Acetyltransferase, bromodomain
Ada2	Tada2b	Ada2b	Histone Acetyltransferase module
Ada3	Tada3	Tada3	Histone Acetyltransferase module
Sgf29	Sgf29	Sgf29	Histone Acetyltransferase module, Tandem Tudor Domain
Taf5	Taf5L	Wda	Structural role, WD40
Taf6	Taf6L	Sda	Structural role, Histone Fold
Taf9	Taf9	Taf9	Structural role, Histone Fold
Taf10	Taf10	Taf10	Structural role, Histone Fold
Taf12	Taf12	Taf12	Structural role, Histone Fold
Ubp8	Usp22	Nonstop	Histone deubiquitination
Sgf73	Atxn7	CG9866	Histone deubiquitination
Sgf11	Atxn7L3		Histone deubiquitination
Sus1	ENY2	E(y)2	Histone deubiquitination

Liste des sous-unités du complex SAGA chez la levure, l'homme et la drosophile. Les sous-unités sont regroupés en fonction de leur fonction au sein du complexe.

Une forme alternative (nommée SLIK ou SALSA) du complexe SAGA a été identifiée, et se caractérise notamment par le clivage de la partie C-terminal de Spt7, et la perte de la sous-unité Spt8, et semble impliquée dans la régulation de certains gènes (Pray-Grant et al., 2002; Sterner et al., 2002)

Des études structurales et de protéomiques ont décrit l'organisation modulaire du complexe, dont les différentes sous-unités contribuant à une même fonction semblent s'associer en modules fonctionnels au sein du complexe (Lee et al., 2011; Wu et al., 2004). L'architecture moléculaire du complexe est organisée en 5 domaines, qui regroupent les différentes fonctions de SAGA. Ainsi, la fonction de liaison aux activateurs est située au sein d'un premier domaine, à l'opposé du site d'interaction avec TBP qui a été identifié au niveau d'un domaine très flexible. Les sous-unités Tafs, dont la stœchiométrie exacte n'est pas encore connue, sont regroupées au sein de deux domaines centraux, en accord avec leur rôle structural. Enfin, Gcn5 est présent au sein d'un autre domaine avec la protéine Spt7 (capable

de lier certaines lysines acétylées sur les queues d'histone), et définit une région potentielle d'interaction avec les nucléosomes.



Le complexe SAGA est organisé de manière modulaire. La structure du complexe revèle l'existance de 5 domaines, où se regoupent les différentes sous-unités se regourpent selon leurs différentes fonctions.

Le complexe SAGA est donc au centre des mécanismes d'initiation de la transcription, en intégrant différentes fonctions nécessaire à cette étape. Ainsi, le complexe SAGA fait le lien entre la machinerie transcriptionnelle et les activateurs, permet le recrutement de TBP et la formation du PIC, et modifie la chromatine par la lecture et l'écriture de modifications spécifiques sur les queues des histones. L'importance du rôle du complexe SAGA lors de la transcription est illustrée par la nécessité du bon fonctionnement de ses différentes fonctions pour le développement normal de l'embryon chez les eucaryotes supérieurs (Carre et al., 2005; Guelman et al., 2006; Weake et al., 2009; Xu et al., 2000), et par son rôle lors du développement de certains cancers (McMahon et al., 1998; Zhang et al., 2008).

Au cours de cette étude, nous nous sommes intéressés au module de déubiquitination (DUB), formé chez la levure par les protéines Sgf73, Ubp8, Sgf11 et Sus1. L'activité enzymatique du module est portée par la sous-unité Ubp8, dont l'homologue chez l'humain (USP22) a été identifié comme un marqueur de la résistance au traitement et à la formation de métastases au cours de certains cancers (Zhang et al., 2008, p. 22). De plus, l'homologue humain de Sgf73, l'ATXN7 peut présenter une extension poly-glutamine dans sa partie N-terminale, responsable de la maladie neurodégénérative SCA7 (Helmlinger et al., 2004). Afin de comprendre le rôle d'Ubp8 au sein du complexe SAGA, ainsi que l'effet de l'altération du module au cours de ces maladies, nous avons localisé le module DUB au sein du complexe. De plus, il semblerait que le module DUB puisse être détaché du reste du complexe par la

particule régulatrice 19S du protéasome (Lim et al., 2013), participant ainsi à la régulation fonctionnelle de SAGA. Nous avons donc analysé l'effet de la perte du module sur la structure du complexe. Enfin, le profil de transcription de certains gènes est altéré chez les modèles de souris de la maladie SCA7. Nous avons donc étudié les possibles interactions entre le module DUB et la fonction de reconnaissance du promoteur du complexe SAGA.

Pour cela, nous avons résolu la structure du complexe purifié à partir d'une souche sauvage, et de deux souches mutantes où la partie N-terminale de Sgf73 (Sgf73Δ1-104), ou la protéine entière (Sgf73Δ), a été supprimée. Les complexes ont été analysés par spectrométrie de masse et leur structure ont été comparée afin d'observer les différences dues à la perte du module.

Chapitre 2. MATERIEL ET METHODES

1. Structure du complexe SAGA observé par microscopie électronique

La plupart des processus biologiques dans la cellule nécessite l'action de protéines qui n'agissent pas de manière isolée, mais plutôt au sein de larges complexes macromoléculaires qui permettent l'action coordonnée de ces différents composants. Ces complexes sont souvent peu abondants et fragiles, et la réalisation d'étude structurale sur ces complexes s'avèrent généralement une tache compliquée.

La microscopie électronique est la méthode de choix pour l'étude structurale de ces larges assemblages macromoléculaires. En effet, la microscopie électronique permet l'observation directe de complexes de plusieurs mégadaltons, dans un état proche de celui présent dans la cellule. De plus, une étude par microscopie électronique nécessite une plus faible quantité de matériel que d'autres méthodes comme la résonnance magnétique nucléaire (RMN) ou la cristallographie aux rayons X. La microscopie électronique a le potentiel pour résoudre la structure d'un complexe biologique et de ces différentes conformations à une résolution proche de l'atome.

La visualisation du complexe en microscopie électronique est basée sur l'interaction des électrons émis par la source avec l'objet observé. Les électrons vont être déviés par les atomes qui composent le spécimen, puis focalisés par un complexe jeu de

lentilles électromagnétiques sur un détecteur qui enregistrera une série d'image du spécimen. Sur ces images, le contraste nécessaire à la bonne visualisation de l'objet est fourni par un diaphragme qui bloque les électrons fortement déviés par le spécimen (contraste d'amplitude), et par l'interférence entre les électrons non-diffusés et diffusés, ces derniers ayant subi un changement de phase due à la défocalisation du faisceau d'électron et aux aberrations des lentilles (contraste de transfert).

La préparation du spécimen se fait en plusieurs étapes. D'abord, le spécimen est adsorbé sur un film de carbone posé sur une grille de microscopie et rendu hydrophile par une décharge dans l'air. Puis, l'échantillon peut être coloré par un sel de métal lourd (tel que l'acétate d'uranyl), qui se dépose autour de l'échantillon et dévie fortement les électrons produisant ainsi un fort contraste d'amplitude. Cette technique, appelé coloration négative, a pour avantage de produire un fort contraste facilitant l'observation, mais génère des artefacts liés aux conditions salines élevées et au pH acide du colorant, ainsi qu'à la déshydratation complète du spécimen, ce qui altère la structure du complexe observé.

Une alternative à la coloration consiste à vitrifier le spécimen en le plongeant rapidement dans un bain à très basse température, tel que l'éthane liquide. Cette technique, appelée cryo-microscopie électronique, permet d'observer un spécimen complètement hydraté et proche de son état physiologique. Cependant, le contraste est ici très faible, et rend l'analyse des images particulièrement difficile.

L'irradiation du spécimen par le faisceau d'électron provoque des ruptures des liaisons moléculaires au sein du spécimen et altère sa structure. Pour limiter ces effets, plusieurs images sont enregistrées à très faible dose (nombre d'électron par unité de surface), ce qui est responsable d'un très fort bruit sur les images mais permet de répartir la dose nécessaire à la visualisions des particules sur plusieurs images. Les différentes vues du spécimen sur les différentes images seront ensuite additionnées, dans le but de réduire ce bruit.

Les images contiennent les vues (idéalement d'une même molécule) sous différents angles d'observation. Pour reconstituer un volume tridimensionnelle du spécimen, ces vues devront être combinées. Cela nécessite un long processus de traitement des images, dont le but est la détermination des angles de projections de ces différentes vues. Pour cela, les images seront regroupées en fonction de leur similarité afin de réduire le niveau de bruit présent dans les images. Ce regroupement nécessite que les images soient dans le même registre, c'est-à-dire alignées précisément les unes par rapport aux autres. Cet alignement, et

la classification des images permettant le regroupement des images similaires (c'est-à-dire correspondant à des vues proches), est un procédé itératif qui permettra d'obtenir des vues caractéristiques de l'objet présentant un rapport signal-bruit élevé. Ces vues pourront être utilisées pour reconstruire un modèle tridimensionnelle de l'objet, après l'assignation des angles de projections. De plus, la classification des différentes vues permettra de séparer les éventuelles différentes conformations du complexe.

2. Purification du complexe SAGA endogène de levure

Pour commencer une étude en microscopie électronique, il est indispensable de produire un échantillon suffisamment homogène et concentré. Le complexe SAGA est constitué de 19 sous-unités, ce qui rend la production du complexe par un système recombinant particulièrement difficile. C'est pourquoi nous avons mis en place un protocole de purification du complexe SAGA endogène chez la levure.

La méthode de TAP-tag (Puig et al., 2001) a été développé dans le but de purifier des complexes peu abondants dans des conditions douces, afin de préserver les interactions entre les différentes sous-unités du complexe. Elle est basée sur l'utilisation de deux étiquettes de purification par affinités possédant une forte affinité et une bonne spécificité.

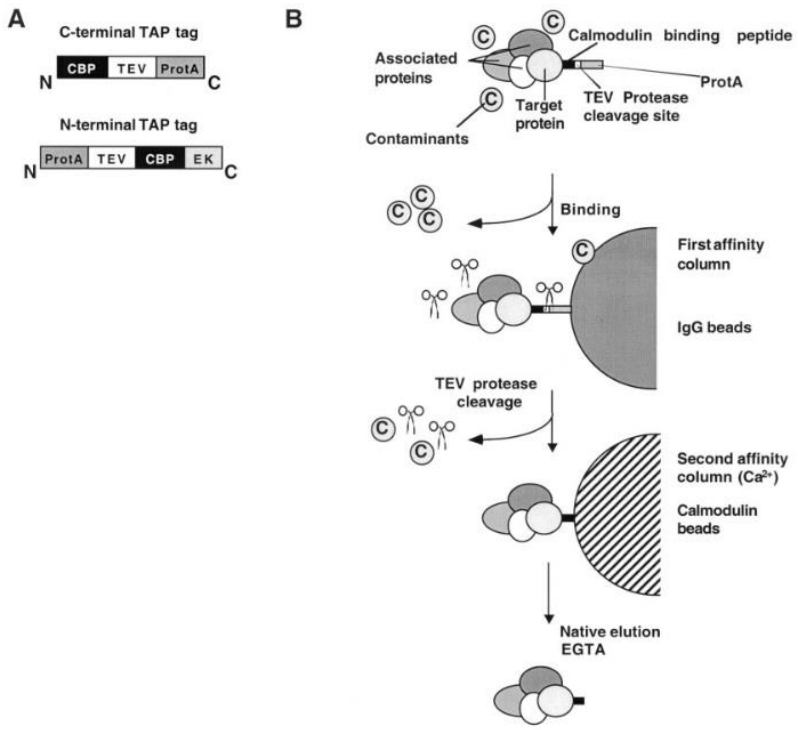


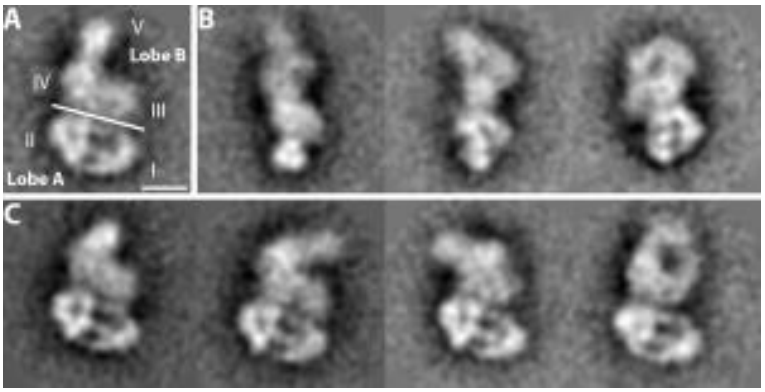
Schéma récapitulatif de la méthode Tap-tag. Le TAP-tag se compose de deux étiquettes en tandem, l'étiquette d'afiinité protA et l'étiquette CBP, et permet la purification de la cible par étapes de purification simple en conditions natives.

Les levures sont d'abord modifiées pour exprimer le tag fusionné (généralement sur la partie C-terminal) avec une des sous-unités du complexe. Les levures sont cultivées en milieu riche YPD, puis lysées par une méthode douce. Le choix de la méthode de lyse est crucial, et doit permettre de produire un extrait cellulaire concentré où le complexe peut être purifié avec une efficacité maximum. Puis le complexe est purifié en suivant un protocole optimisé afin de permettre de préserver la stabilité du complexe. Enfin, la qualité de l'échantillon sera contrôlée par une analyse sur gel de polyacrylamide en présence de SDS (PAGE-SDS) coloré à l'argent, et par spectrométrie de masse afin de vérifier la composition protéique du complexe.

Chapitre 3. RESULTATS

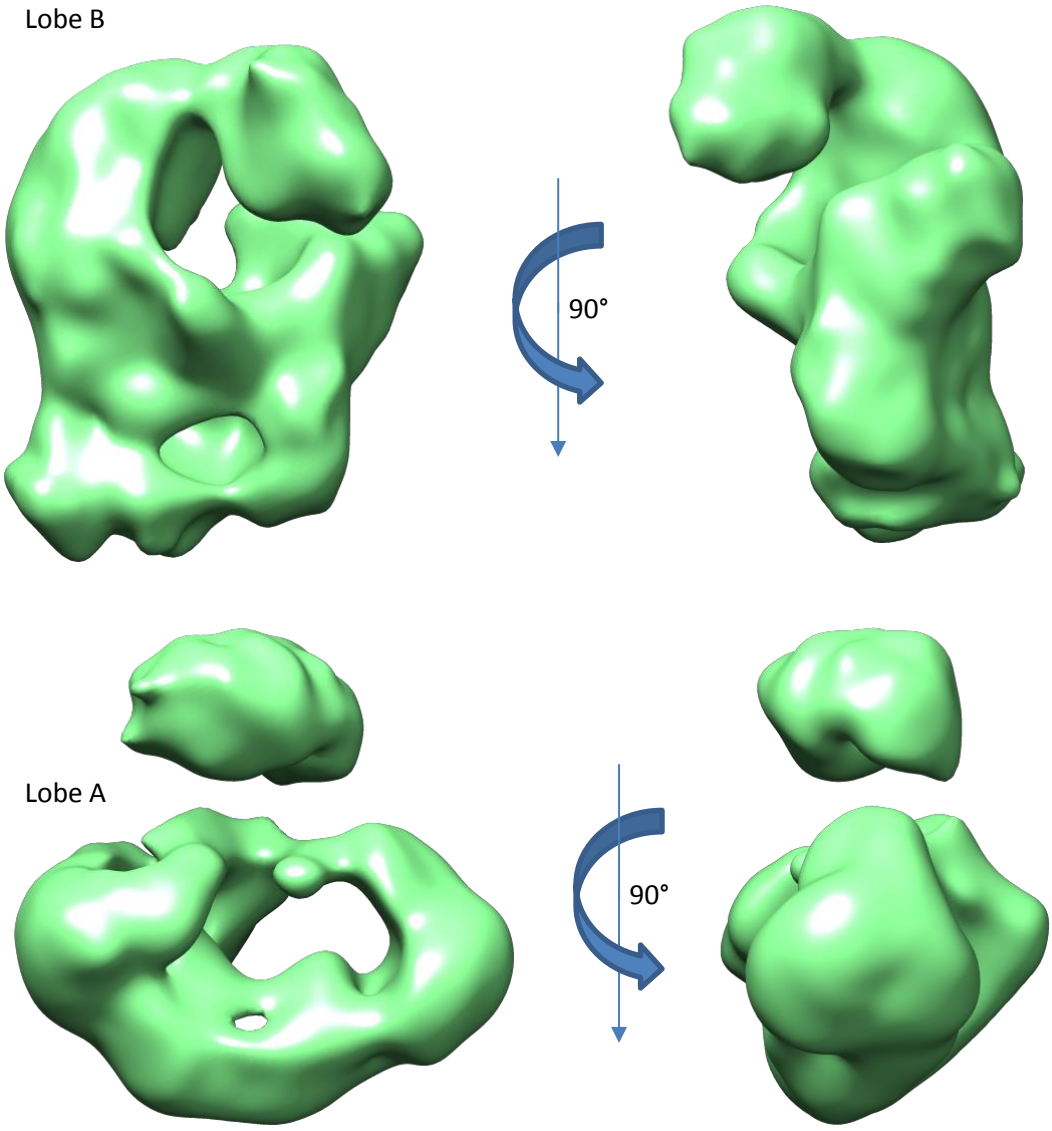
1. Structure du complexe SAGA par microscopie électronique

L'analyse par microscopie électronique du complexe produit à partir de la souche sauvage a permis une description plus précise de la structure du complexe. Ainsi, les 5 domaines précédemment décrits 2 peuvent être regroupés en deux lobes. Le lobe A apparaît très homogène au sein des images et se compose du domaine I, contenant la protéine Tra1 et du domaine II. On observe une forte homologie entre le lobe A et la structure du complexe NuA4, qui contient également Tra1 (Chittuluru et al., 2011), permettant l'identification de cette sous-unité sans ambiguïté. Le lobe B contient les domaines III, IV et V, et présente une structure en forme de pince moléculaire adoptant des conformations ouvertes ou fermées.



Le complexe est organisé en 2 lobes, le lobe A étant très homogène et le lobe B montrant un haut degré de flexibilité.

Enfin, l'analyse des images semble suggérer une flexibilité dans le positionnement des deux lobes l'un par rapport à l'autre. L'extraction et l'analyse des deux lobes de manière séparée et indépendante a permis d'améliorer la définition des deux lobes, en accord avec leur flexibilité relative. Il a été ainsi possible de mettre en valeur la présence du domaine V, dans une position intermédiaire plus fortement représentée au sein du complexe sauvage, ainsi que l'apparition d'une densité supplémentaire au sein du domaine III.



L'analyse séparée des lobes A et B permet de mettre en évidence la présence du domaine III et du domaine V, formant un pince moléculaire, qui n'avait jamais été observé auparavant.

2. Localisation du module DUB et des protéines Spt7 et Spt8

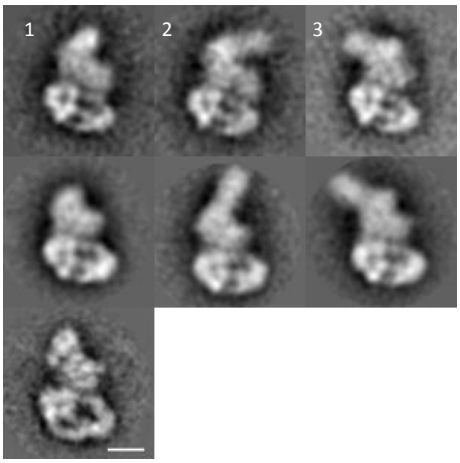
La résolution des structures des deux complexes provenant des mutants de délétion a permis d'observer une structure allongée de taille similaire à celle observée dans le cas du complexe sauvage, suggérant que l'intégrité du complexe n'a pas été affectée par la délétion. Le lobe A ne montre pas de différence majeure avec celui du complexe issu du sauvage, permettant d'exclure la présence du module au sein de ce lobe. Dans le cas du mutant Sgf73Δ1-104, le lobe B adopte toujours une forme de pince, excluant la présence du module au sein du domaine V. Cependant, une proportion accrue de conformation où ce

domaine n'est pas visible est observée, et cette observation est encore plus forte dans le cas du mutant Sgf73Δ. Cela semble indiquer que les propriétés dynamiques du domaine V sont affectées par le module DUB, et en particulier par la partie C-terminal de Sg73.

(A) WT

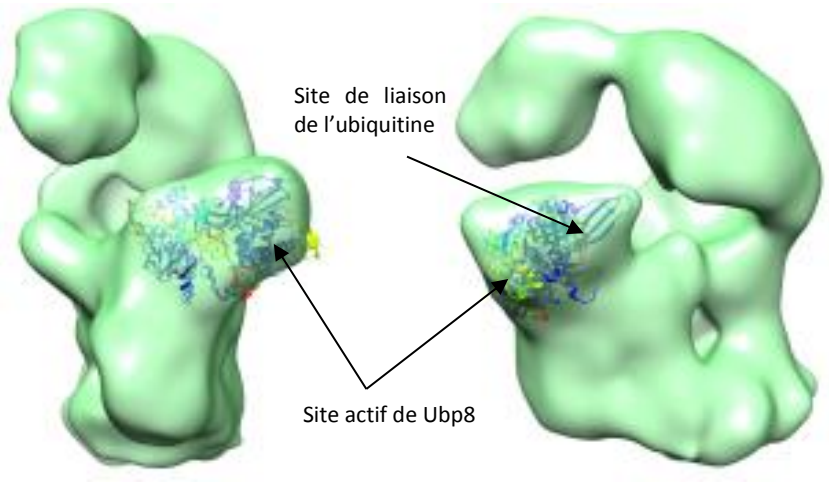
(B) Sgf73Δ1-104

(C) ΔSgf73



Présence et position du domaine V au sein des complexes purifiés à partir du souche sauvage, et de deux souches de délétion totale ou partielle de la sous-unité Sgf73.

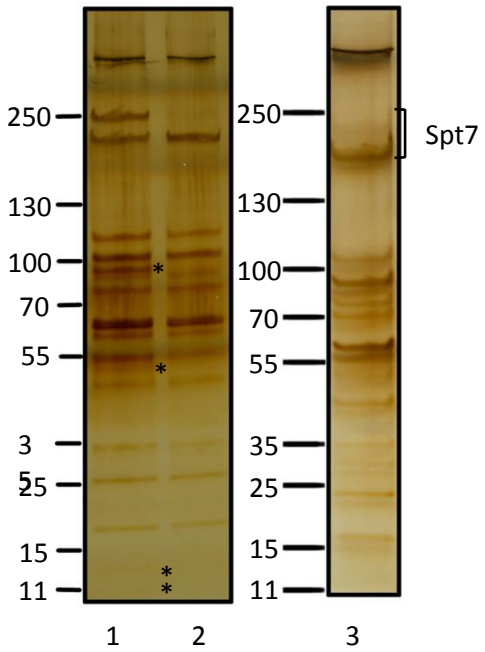
L'analyse séparée des deux lobes, comme dans le cas du complexe issu de la souche sauvage, a permis de mettre en évidence une densité manquante au niveau du domaine III, dont la taille est compatible avec la structure cristallographique du module DUB (Köhler et al., 2010). Ceci positionne donc le module DUB à proximité du module d'acétylation des histones contenant Gcn5, et de la protéine Spt7 contenant un domaine d'interaction potentiel avec les nucléosomes, et définit ainsi un domaine au sein du complexe SAGA regroupant les fonctions de modification des histones et d'interaction avec la chromatine. Un immuno-marquage de Spt8 a également permis de mettre en évidence la présence de cette protéine dans cette région du complexe, en accord avec la présence de Spt7 et son accessibilité accrue lors de la perte du module DUB.



Position du module de déubiquitination au sein de la structure du compexe SAGA.

3. La délétion du module DUB favorise le clivage de Spt7

La caractérisation des complexes par PAGE-SDS et spectrométrie de masse (Mudpit) a permis de confirmer la perte totale du module DUB au sein des complexes obtenus à partir des deux souches mutantes. De manière surprenante, ces complexes purifiés contiennent en majorité une forme tronquée de Spt7. La protéine Spt8 est également peu présente dans ces échantillons. Ces 2 particularités sont caractéristiques de la forme alternative de SAGA, nommée SLIK ou SALSA, présentant des propriétés de liaison avec TBP différentes du complexe SAGA, et pourrait être impliquée dans la régulation de l'expression de certains gènes (Pray-Grant et al., 2002; Sterner et al., 2002). L'augmentation de l'efficacité du clivage dans le cas du mutant de délétion pourrait être liée à une augmentation de l'accessibilité de Spt7 en absence du module DUB. Cela suggère une proximité de ces sous-unités au sein du complexe.

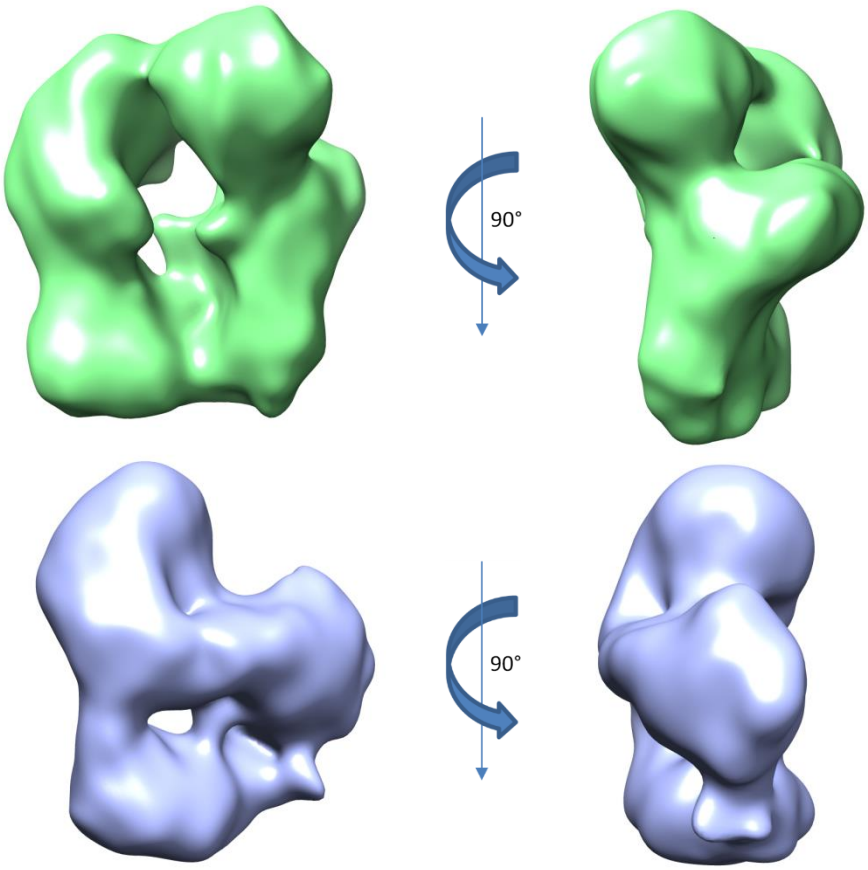


Analyse par PAGE-SDS de la composition du complexe SAGA purifié à partir de la souche sauvage (piste 1), d'une souche contenant une délétion de la partie N-terminale de Sgf73 (piste 2) et d'une souche contenant une délétion totale de Sgf73 (piste 3). La disparation des bandes marquées par une étoile (*) correspond à la perte des sous-unités du module de déubiquitination.

De plus, comme montré précédemment, la position du domaine V portant le site d'interaction avec TBP semble être différente lorsque le module DUB, et particulièrement Sgf73, est absent. Cela suggère un possible rôle inattendu pour cette protéine dans la régulation de l'activité d'activation de latranscription, mais ce résultat nécessite d'être confirmer.

4. Le module DUB n'est pas toujours présent au sein du complexe SAGA

Il a été proposé que le module DUB pouvait être dissocié de manière fonctionnelle du complexe SAGA (Lim et al., 2013). La séparation du jeu de données basée sur la similarité des images avec un modèle du complexe comportant ou non ce module, a permis de mettre en évidence une sous-population de complexe SAGA où le module DUB est absent. De manière surprenante, il semble que la proportion de complexe ne contenant pas le module est assez élevée, et coïncide avec la proportion de complexe ayant perdu Spt8.

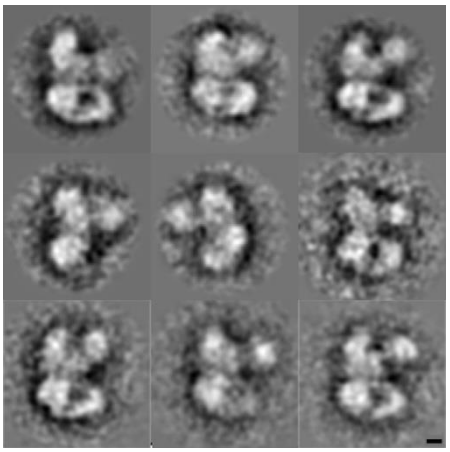


La séparation du jeu de données du complexe issu de la souche sauvage selon la présence ou non du domaine III met en évidence la présence d'une large population (50% des particules observées) où le module de déubiquitination semble absent.

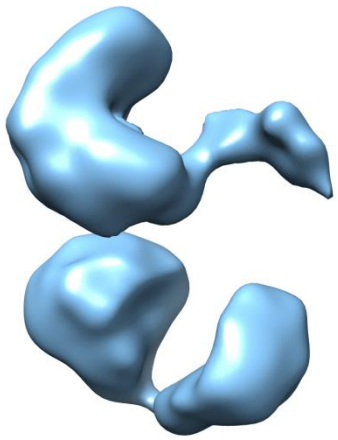
5. Interactions du complexe SAGA avec le nucléosome

La localisation du module de déubiquitination au sein du complexe permet de définir une possible zone d'interaction avec le nucléosome. Pour vérifier si le nucléosome se lie au complexe au travers de cette interface, nous avons reconstitué un complexe formé entre ces deux éléments et la résolution de la structure de ce complexe est en cours.

A



B



Images moyennes de classes (A) et reconstruction tri-dimensionnelle du complexe (B) formé entre SAGA et le nucléosome et montrant la présence d'une densité proche du domaine III, compatible avec la présence d'un nucléosome.

Nos résultats préliminaires indiquent que l'interaction se situe bien au niveau du domaine III, mais que la liaison semble se faire au travers d'une partie flexible du complexe et/ou du nucléosome. Il est possible que des contacts additionnels soient présents dans le contexte de la chromatine, permettant une liaison plus stable du complexe.

Chapitre 4. CONCLUSIONS ET PERSPECTIVES

L'étude structurale du complexe SAGA réalisée au cours de ces travaux de thèse a permis d'obtenir une meilleure description de la structure du complexe, et de mettre en évidence l'existence d'un possible mécanisme de régulation de la fonction de reconnaissance de promoteur du complexe SAGA par une de ces sous-unités. Ainsi, la localisation du module DUB à proximité de Gcn5 permet de définir une région du complexe impliquée dans l'interaction avec la chromatine, en accord avec les études génétiques, biochimiques et fonctionnelles décrivant la structure modulaire du complexe. De plus, le complexe contient un lobe formant une pince moléculaire, dont les deux extrémités comportent une sous-unité interagissant directement avec TBP. Le module DUB, notamment la protéine Sgf73, semble impliquée dans la régulation des conformations de ce lobe, et pourrait donc permettre la régulation de la fonction de reconnaissance du promoteur de SAGA.

L'amélioration des conditions de purification du complexe a permis d'initier une étude par cryo microscopie électronique, dans le but d'obtenir des données structurales à haute résolution du complexe, nécessaire à l'étude des mécanismes de régulation du complexe. Afin de mieux comprendre comment se lie TBP avec le complexe, et quelle est la

contribution des différentes sous-unités à cette liaison et à sa régulation, il serait important d'obtenir des données structures sur des complexes fonctionnels.

Une publication scientifique reprenant ces résultats est en cours de préparation et un article de revue a été publié.

CHAPTER I :

INTRODUCTION

Chapter 1. INTRODUCTION

In all organisms, the genetic information of each cell is encoded in a set of deoxyribonucleic acid (DNA) molecules which composes the genome of each individual. The genes contain the information required by the cell to ensure all the biological functions and processes that will take place during its life cycle. In order to achieve this goal, this information needs to be extracted and translated in order to form active biological molecules, such as proteins or ribonucleic acid (RNA), which serves as the molecular tools of the cell. All the biological processes that lead to synthesis of these macromolecules is called gene expression.

Gene expression is a complex mechanism that comprises several steps which are tightly regulated in order to achieve basic cellular functions as well as appropriate responses to stimuli from the environment. For many years, the central dogma in molecular biology has been the conversion of the DNA into RNA, which will in turn be exported from the nuclei to the cytoplasm in order to be translated to protein.

The process of production of a RNA molecule from a DNA template, termed transcription, is a highly coordinated process mediated by RNA polymerases (RNA pols). While in prokaryote a unique polymerase ensure this process, in eukaryotes, three RNA pols (I, II, and III) were identified. RNA pol I is primarily involved in transcribing 18S and 28S ribosomal RNAs (rRNA), while RNA pol II transcribes messenger RNAs (mRNA), and RNA pol III is responsible for synthesis of cellular 5S rRNA and transfer RNAs (tRNA). The RNA pol I is localized within nucleoli, the sites for rRNA synthesis, whereas RNA pol II and III are normally present in the nucleoplasm. Transcription of protein coding genes by the RNA pol II is composed of three steps, each of them being at the center of numerous regulation processes in order to control gene expression. During initiation, the RNA pol II will be positioned on the gene promoter. Then during elongation, RNA pol II will read the coding sequence of the transcribed gene and produce a pre-mRNA which will need to be processed in several subsequent steps in order to give raise to a mature mRNA. These maturation steps are usually performed co-transcriptionally and involve the capping of the 5’ end of the mRNA, the removal of introns and the addition of a polyA tail at the 3’ end of the transcript. Eventually, in the termination step, the mRNA molecule will be released by the transcription machinery, which can resume to another cycle of transcription, and will be exported from the nucleus to the cytoplasm where the translation by the ribosomes will take place.

In eukaryote, a major obstacle for the transcription machinery is formed by the structure of the chromatin. In the nucleus, the DNA molecule is wrapped around nucleosomes, an octamer of histone proteins, in a compact structure. The RNA pol needs to gain access to the DNA template, with the help of different classes of cofactors which will alter the chromatin. These cofactors can act by adding post-translational modification on the histone proteins, thus weakening the protein-DNA interactions which bind the DNA template to the nucleosome, or by directly remodeling of the chromatin itself in an ATP-dependent manner.

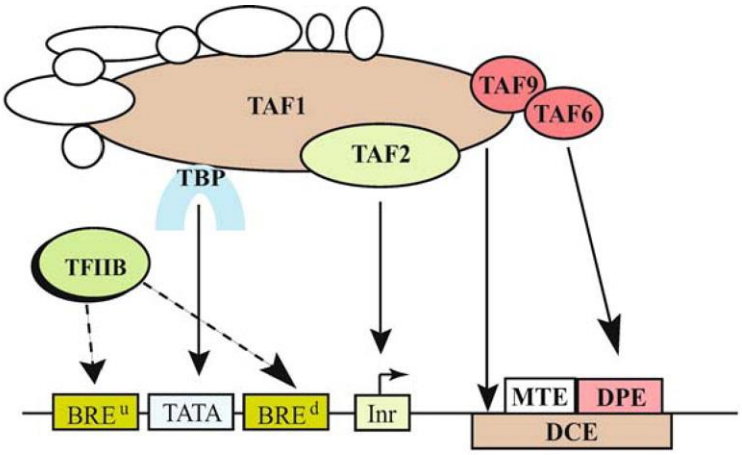
1.1 - INITIATION OF TRANSCRIPTION BY RNA POL II

1.1.1 The Promoter of the protein coding genes

To initiate transcription of the coding sequence of the gene, the transcription machinery needs first to be recruited. In this regard, it requires signals to precisely define where to start the synthesis of the RNA molecule. The promoter is the DNA region in the vicinity of a gene which integrates a set of DNA sequences (or elements) and will define the transcription start site (TSS) and regulate the frequency of the initiation event. The proximal promoter is composed of a combination of different core promoter elements (see below) in the close vicinity of the TSS which are recognized by a set of transcription factors (see section 1.1.2). These transcriptions factors will guide the polymerase to the TSS (defined as the +1 nucleotide) and allow correct initiation of the transcription. The binding of these factors to the promoter is also regulated by more distal sequence elements, termed enhancers. Enhancers are DNA sequences located upstream or downstream of the proximal promoter, in some cases far from the TSS, which can stimulate transcription upon binding of transcription factors (called activators). In yeast, these regulatory sequences are generally limited to Upstream Activation Sequences (UAS), which are located a few hundred bases upstream of the TSS.

In higher eukaryotes, seven core promoter elements have been identified, and characterized mainly in *Drosophila* while they are still poorly characterized in yeast (reviewed in (Thomas and Chiang, 2006) and (Juven-Gershon and Kadonaga, 2010), summarized in Figure 1). The first core promoter element that was identified is the TATA-box, an A/T-rich sequence located upstream of the TSS, in position -31 to -24 (Corden et al., 1980), with the striking exception of the yeast *Saccharomyces cerevisiae* (-40 to -120). The consensus sequence of the TATA-box is TATA(A/T)A(A/T)(A/G) in metazoan, and is recognized by the TATA-box Binding

Protein (TBP) to nucleate the formation of a Pre-Initiation Complex (PIC, see below). The second core promoter element is the Initiator sequence (Inr) and seems to be the most commonly occurring element (Juven-Gershon and Kadonaga, 2010; Yang et al., 2007). The Inr is a pyrimidine-rich sequence of consensus (T/C)(T/C)A_{+1}N(T/A)(T/C)(T/C) in human and slightly differs in *Drosophila*. The Inr is able to drive transcription initiation alone or in association with the TATA-box. The third core promoter element is the Downstream Promoter Element (DPE), which is located in position +28 to +34 downstream of the TSS. Its consensus sequence is (A/G)G(A/T)CGTG and it functions in synergy with the Inr, exhibiting a strict dependence on the spacing between these two elements (Kutach and Kadonaga, 2000). The DPE is conserved from Drosophila to human, but seems to be absent from yeast. Similarly to the DPE, two others core promoter elements are located downstream of the TSS: the Motif Ten Element (MTE, consensus C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C)) and the Downstream Core Element (DCE). The MTE is found between +18 and +29 and works cooperatively with the Inr, whereas it can act either independently or in synergy with the DPE and the TATA-box (Lim et al., 2004). In contrast, the DCE (composed of the three motifs CTTC, CTGT, and AGC, respectively, spanning from +6 to +34) is exclusively found in the absence of the DPE. All these core promoter elements are recognition sites for proteins of the General Transcription Factor (GTF, see below part 1.1.2) TFIID. In contrast, the last two core promoter elements, namely the downstream and upstream TFIIB Recognition Elements (BREu and BREd), are recognition sites of a second GTF called TFIIB. They are located at both sides of the TATA-box, and their consensus sequences are (G/C)(G/C)(G/A)CGCC (BREu) and (G/A)T(T/G/A)(T/G)(G/T)(T/G)(T/G) (BREd). BRE elements are found in both TATA-containing and TATA-less promoters, and define the orientation of the transcription. Here again, BRE motifs seems to be largely absent in yeast promoters (Yang et al., 2007).



Core Promoter Element	Position	Consensus Sequence (5' to 3')	Bound Protein
BRE^u	-38 to -32	(G/C)(G/C)(G/A)CGCC	TFIIB
TATA	-31 to -24	TATA(A/T)A(A/T)(A/G)	TBP
BRE^d	-23 to -17	(G/A)T(T/G/A)(T/G)(G/T)(T/G)(T/G)	TFIIB
Inr	-2 to +5	PyPyAN(T/A)PyPy	TAF1/TAF2
MTE	+18 to +29	C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C)	n.a.
DPE	+28 to +34	(A/G)G(A/T)CGTG	TAF6/TAF9
DCE	3 subelements +6 to +11 +16 to +21 +30 to +34	core sequence: S_I CTTC S_{II} CTGT S_{III} AGC	TAF1

Figure 1 : Organization of eukaryotic promoter

The Core Promoter Elements and their respective binding partners and consensus sequence in human. From (Thomas and Chiang, 2006)

Interestingly, none of the core promoter elements identified thus far is ubiquitous or universally required for transcription. For example, the TATA-box was originally thought to be present at the promoter of all genes but surprisingly, genome-wide studies have revealed that only a small subset, in the range of 20% in mammals and yeast (Basehoar et al., 2004), possesses this element. TATA-less promoters are often housekeeping genes which are tightly regulated, and the variability of the core promoter elements used in those promoters seems to play a critical role in this regulation. Thus, transcriptional regulation is achieved by the diversity of enhancers which regulate spatially and temporally the expression of the genes, but also by diversity in core promoter structure.

1.1.2 The Pre-Initiation Complex

On its own, none of the three eukaryotic RNA pols can initiate transcription at the TSS. The GTFs, as well as transcription factors and coactivators (see section 1.3 - Role of the Activators in transcription) are required for the specific recognition of the promoter, the

recruitment of the RNA polymerase, the interaction with regulatory factors, the unwinding of the chromatin and the recognition of the TSS. During transcription of protein coding genes by RNA pol II, several GTFs (namely TFIIA, TBP, TFIIB, TFIIF, TFIIE and TFIIH) and the pol II itself, assemble into a large PIC regrouping more than 60 different proteins. The formation of the PIC is an important and a highly regulated step of the transcription. The different GTFs present in the PIC will form multiple interactions with specific elements of the core promoter, in order to anchor the RNA pol II to the promoter DNA. In the latest steps of PIC formation, the transcription factor TFIIH will open the double-stranded DNA molecule in the close vicinity of the pol II. Then, TFIIB will locate the TSS and position the RNA pol II to the TSS, in order to start transcription. After a few rounds of abortive transcription, where short, non-productive RNA are produced, the RNA pol II pauses a few nucleotides downstream of the TSS due to the action of negative elongation factors (see section 1.1.2.7 below). This major restriction point is a key rate-limiting step in the transition from initiation to elongation, and require additional positive factors to alleviate the RNA pol II pausing and transition to the processive elongation where the RNA molecule will be fully synthetized.

Two models exist for the assembly of the PIC. In the sequential assembly pathway (Figure 2A), the different GTFs are added sequentially in a stepwise manner. TFIID is the first GTF to bind to the core promoter (at TFIID-dependent promoter), and this binding is followed by the sequential addition of TFIIA and TFIIB which stabilize the interaction of TFIID with the promoter (Buratowski et al., 1989). This allows the recruitment of the RNA pol II associated to TFIIF, and place the enzyme at the center of the PIC (Flores et al., 1991). After binding of TFIIE, TFIIH is the last factor to be recruited (Maxon et al., 1994). The RNA pol II Holoenzyme Pathway (Figure 2B) comes from the observation that PIC components, with the exception of TFIID and TFIIA, can be co-purified with the RNA pol II (Koleske and Young, 1994). In this model, TFIID will bind to the core promoter and facilitate the positioning of the RNA pol II holoenzyme to the promoter, similarly of the mechanism in the prokaryotic system. Others proteins were shown to copurify with the holoenzyme, and in particular proteins from the Mediator complex (Gustafsson et al., 1997). Mediator was identified as a protein complex which is required for bridging activator proteins to the transcription machinery, and thus belong to a second class of transcription factor, called co-activators, that will be discuss later (section 1.3.1.2).

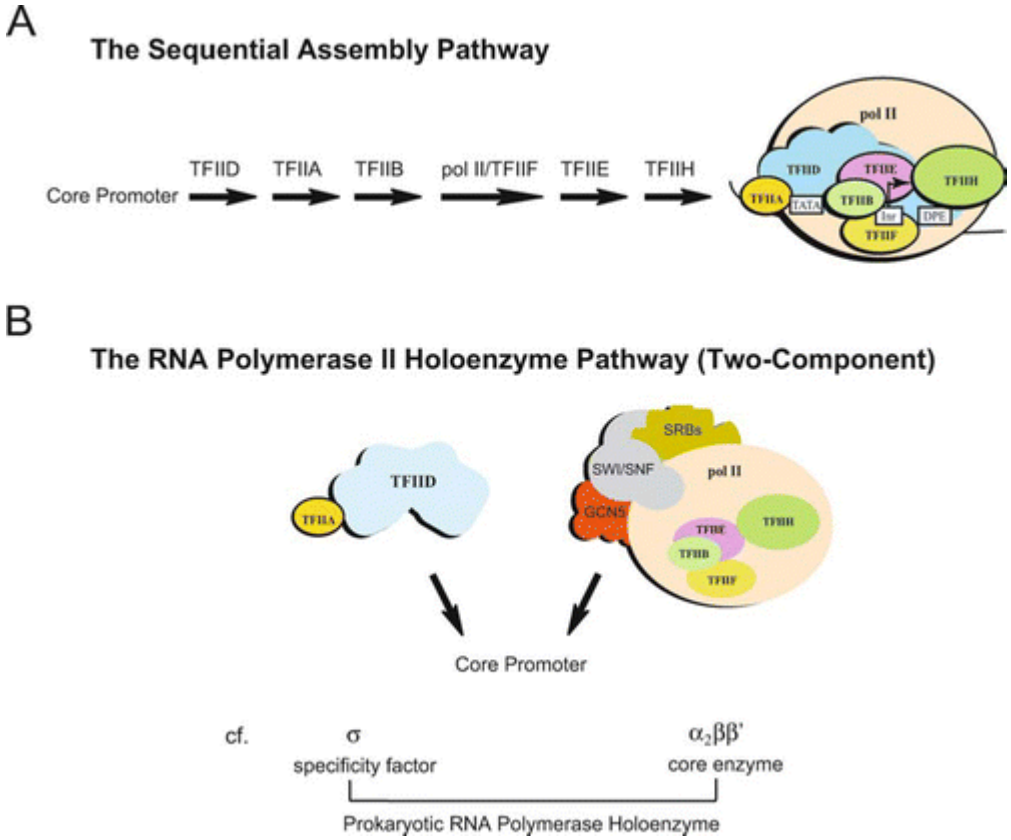


Figure 2 : Different models for PIC assmebly.
From (Thomas and Chiang, 2006)

1.1.2.1 The TATA-binding protein

In the model of the sequential assembly, TBP is the first transcription factor to bind to the promoter and nucleates formation of the PIC. In yeast, TBP is associated to two large complexes in order to stimulate transcription: (i) TFIID, which is composed of TBP and 14 TBP-associated factors (TAFs) and is responsible for expression of about 90% of the genes in yeast. Those genes are mainly housekeeping genes, which show a high basal transcription level (Basehoar et al., 2004; Huisinga and Pugh, 2004). As already mentioned, TFIID is involved in promoter recognition through the interaction of some Tafs with core promoter elements. However, TFIID also belongs to the class of the transcriptional coactivator, as it bridges the activators to the transcription machinery. The role and functions of TFIID will be discussed in section 1.3.1.1. (ii) The SAGA complex, a large coactivator complex involved in the transcription of approximately 10% of genes in yeast, which are mainly stress-induced genes and are highly regulated (Basehoar et al., 2004; Huisinga and Pugh, 2004). The SAGA complex will be the topic of section 1.4 - .

TBP is a protein of 240 amino acids in yeast (and 339 in human) highly conserved in sequence and function during evolution. The TATA-box is recognized by the C-terminal part

of TBP, and its binding to the promoter element results in the bending of the DNA by approximately 90 degrees, forming an asymmetric platform for PIC assembly. TBP structure resembles a molecular saddle, which sits on the DNA (J. L. Kim et al., 1993; Y. Kim et al., 1993; Nikolov et al., 1996). The convex face of this saddle forms an interface for interactions with transcription factors, while the concave face contains phenylalanine residues which will provoke a sharp kink of the DNA by widening the minor groove of the double helix (Figure 3). TBP is able to bind aspecifically DNA, particularly A/T rich region (Coleman and Pugh, 1995). To prevent transcription initiation at non-promoter sequence, several mechanisms of TBP-binding inhibition exist. TBP can form a homodimer through extensive contacts between the concave face of the two monomers, thus blocking the DNA-binding surface of TBP and consequently TBP function (Coleman et al., 1995; Jackson-Fisher et al., 1999; Kato et al., 1994). However, this dimerization seems dependent on TBP concentration and buffer composition, in particular for the presence of magnesium, and so far it is not clear if TBP dimerization is relevant *in vivo* (Vanathi et al., 2003). Similarly, TAF1 N-terminal domain (TAND) can interact with TBP to inhibit its function (Kokubo et al., 1993). The N-terminal region of the TAND (TAND1) binds the concave face of TBP, blocking protein DNA-interaction, while a second region (TAND2) interacts with the convex face and compete for TFIIA binding (see below, paragraph TFIIA) (Kokubo et al., 1998; Liu et al., 1998). An additional mechanism by which TBP function is regulated, is by the binding with the protein BTAF1 (in human) or Mot1 (in yeast), which bind to the concave face of TBP in a similar manner as TAF1, blocking TBP-promoter binding (Pereira et al., 2001). In addition, BTAF1 possesses an enzymatic activity which is able to disrupt TBP-DNA complex in an ATP (adenosine triphosphate)-dependent manner (Auble et al., 1997; Chicca et al., 1998). However, this activity was also proposed in yeast to positively acts on transcription by regulating TBP-binding to non-promoter binding site (Muldrow et al., 1999). Finally, the TBP-TATA complex can be recognized and bound by the negative cofactor 2 (NC2), which is composed of NC2α and NC2β subunits interacting via two Histone-Fold Domain (HFD) and forms a molecular clamp which block the convex surface of TBP (Goppelt and Meisterernst, 1996; Kamada et al., 2001). Here again, blocking of this surface results in the inhibition of TFIIA and TFIIB interaction with TBP, and thus prevent formation of the PIC. However, a positive role for NC2 in transcription initiation has been described, but the mechanisms involved are not fully understood (Geisberg et al., 2001)

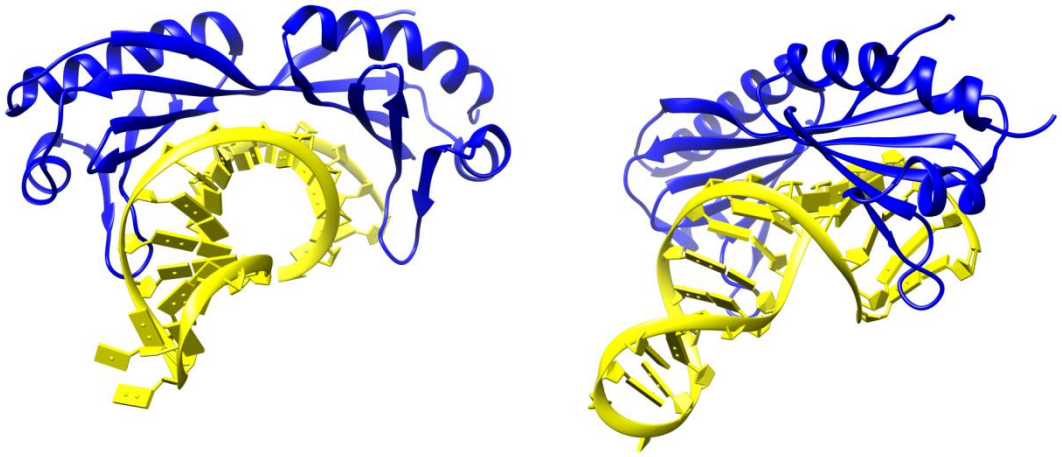


Figure 3 : Structure of yeast TBP bound to the TATA-box
Yeast TBP (in blue) sits on the minor groove of the DNA (in yellow) and provokes a strong kink of the DNA molecule. Second image is turned roughly 45° compared to the first. PDB file 1YTB (J. L. Kim et al., 1993)

Although TBP is unique in yeast, higher eukaryotes can have one or two copies of genes encoding TBP-related factors (TRF) (reviewed in (Davidson, 2003)). TRFs promotes transcription of a subset of genes in a cell-type specific manner (Kopytova et al., 2006; Martianov et al., 2002).

1.1.2.2 TFIIA

TFIIA is a heterodimer in yeast, but is composed of three proteins in human, namely TFIIAα, TFIIAβ and TFIIAγ (respectively 35 kDa, 19 kDa and 12 kDa). TFIIAα AND β derive from a single gene those product is processed by a protease cleavage to form the two subunits. TFIIA functions as a derepressor for transcription, by alleviating inhibition of TBP described before, such as TBP dimerization, binding by BTAF1/Mot1 or TAF1 in TFIID, or binding of NC2. TFIIA binds the TBP-TATA complex and makes contacts with DNA upstream of the TATA-box, in order to stabilize protein-DNA interactions. In addition, it competes for the binding of the convex face of TBP with the inhibiting TAND1 domain of TAF1 (Kokubo et al., 1998). Eventually, the TFIIA-TBP-TATA ternary complex is more resistant to BTAF1/Mot1-mediated dissociation of TBP-TATA complex. TFIIA is formed by two domains: the N-terminal domains of both subunits associate to form a domain interacting with the DNA and the N-terminal part of TBP; the C-terminal domains contribute to the second domain which point opposite to TBP (Geiger et al., 1996) (Figure 4)

Figure 4 : Structure of TBP-DNA-TFIIA complex
Complex formed by TFIIA (in green), TBP (in blue) and the TATA-box (in yellow). TFIIA N-terminal domain is directed toward the DNA. PDB accession number 1RM1.

TFIIA can be recruited to the promoter by activators, such as the activation domain VP16 (Dion and Coulombe, 2003) and interacts with component of TFIID such as TBP, TAF1, TAF4 and TAF11 (Kraemer et al., 2001; Robinson et al., 2005; Yokomori et al., 1993). Therefore, TFIIA can stimulate transcription by stabilizing TFIID-binding to promoter in response to activator signal.

1.1.2.3 TFIIB

TFIIB plays a central role in initiation, by bridging TBP and the RNA pol II, and is essential for the recruitment of RNA pol II to the PIC. Similarly to TFIIA, but independently from its presence or absence, binding of TFIIB to the TBP-TATA complex forms a more stable ternary complex, and strengthen TBP/TFIID-DNA interaction by stabilizing the bent TBP-DNA complex and decreasing the rate of dissociation of the complex. TFIIB is composed of a single protein of 316 and 345 amino-acids in human and yeast (respectively), and its sequence is divided into 5 functional domains (Kostrewa et al., 2009).

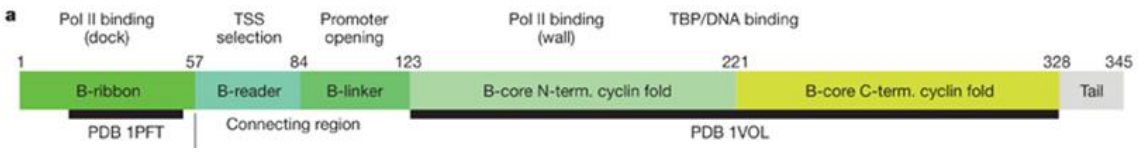


Figure 5 : Domain composition of yeast TFIIB
From (Kostrewa et al., 2009)

TFIIB recruits the pol II to the promoter, via direct interaction with its N-terminal domain (B-ribbon) and part of the C-terminal domain (B-core), whereas the remaining part of the B-core also binds TBP and the DNA distorted by TBP-binding (Hahn, 2004; Kostrewa et al.,

2009). Moreover, TFIIB is involved in TSS selection and opening of the promoter DNA through the B-linker domain and the B-reader domain, respectively (Kostrewa et al., 2009). Interactions of TFIIB with the DNA were mapped upstream and downstream of the TATA-box, at the BREu and BREd promoter elements, and these additional contacts help to stabilize TBP to the promoter.

A model for the role of TFIIB in transition from initiation to elongation has been proposed. In this model, DNA-bound TFIIB recruits the RNA pol II by the B-ribbon and positions the active site above the cleft of the polymerase (close complex). Then, the double-stranded DNA molecule is opened around 20 base-pairs (bp) downstream of the TATA-box, through the B-linker, and the template strand is placed inside the cleft, close to the active center of the RNA pol II (open complex). Subsequently, the template strand is scanned for an Inr motif by the B-reader. After positioning of two nucleotides in front of the Inr and formation of the first phosphodiester bound, the RNA molecule starts to grow and may interfere with the B-reader loop, leading to ejection of short RNA molecules (abortive transcription). When the RNA molecule reaches seven nucleotides, TFIIB is displaced by clashing with the transcription product and the elongation complex is formed (promoter escape) (Kostrewa et al., 2009; Sainsbury et al., 2013).

1.1.2.4 TFIIF

TFIIF is the fourth GTF to enter the PIC in the sequential assembly pathway, and plays where it plays multiple roles. It is composed of two proteins in human (RNA-pol II Associated Protein or RAP30 and RAP74) which form a heterotetramer. RAP30 is able to bind RAP74 via a N-terminal domain (amino-acids 1-98) and interact with the pol II subunit RPB5 via its central domain (amino acids 107 to 170) (Fang and Burton, 1996). Eventually, a C-terminal domain (amino-acids 164-249) is able to bind DNA in a non-specific manner. RAP74 is also made of three functional domains, involved in interactions with different partners: the N-terminal part (1-172) binds to RAP30 and the TAF1 subunit of TFIID; interactions with TFIIA subunits α and β involves two domains situated in the central part (76-136) and C-terminal part (410-444) of the protein; the C-terminal part (449-517) forms a DNA-binding domain similar to RAP30; moreover, domain spanning amino-acids 358 to 517 is involved in interactions with TFIIB. Domains involved in all these interactions are summarized in Figure 6. Similarly to many transcription factors, TFIIF can be recruited by transcriptional activators.

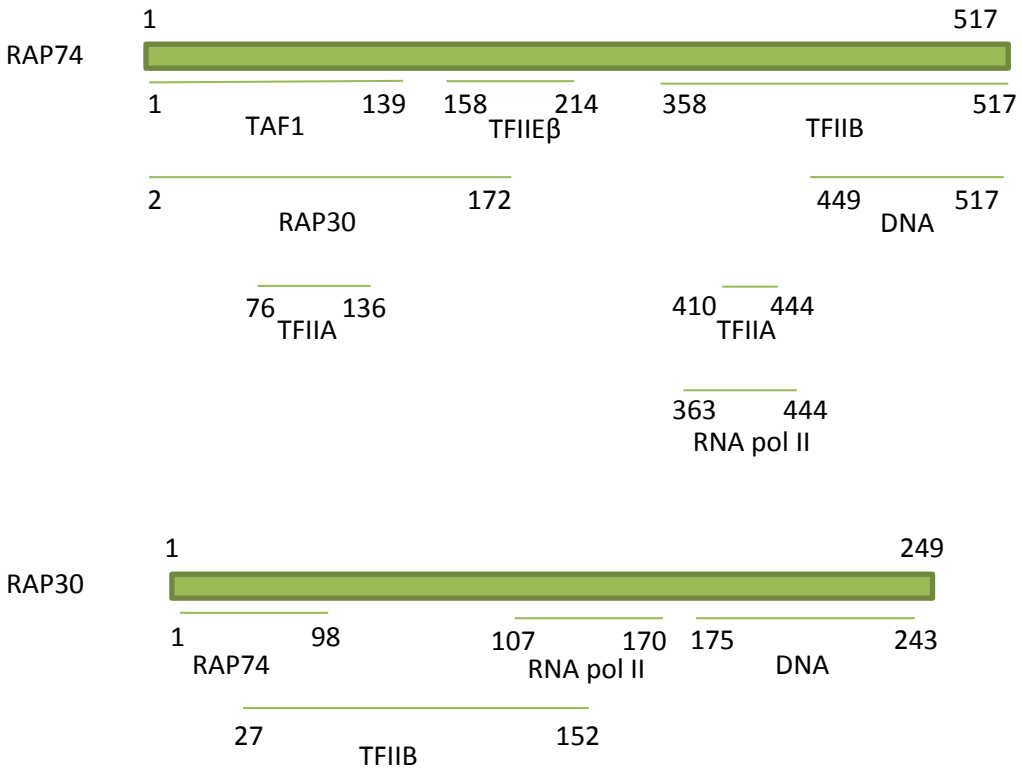


Figure 6 : Domains of the two subunits forming TFIIF

Position of known interaction domains are indicated. Adapted from (Thomas and Chiang, 2006)

The role and function of TFIIF in PIC are numerous: (i) TFIIF is strongly associated to RNA pol II, via interactions between RAP74 and the RPB4/RPB7 subunits, as well as RPB9, and between RAP30 and RPB5. These interactions facilitate recruitment of the RNA pol II to the promoter-bound TFIIB and TFIID (Chung et al., 2003; Rani et al., 2004); (ii) TFIIF stabilizes interaction of RNA pol II to the promoter by providing additional contacts with the DNA (Robert et al., 1998); (iii) TFIIF allows further recruitment of TFIIE and TFIIH to the PIC via direct interactions with TFIIE (Maxon et al., 1994); (iv) TFIIF, similarly to TFIIB, is located close to the active center of the RNA pol II and contribute to TSS selection driven by TFIIB (Ghazy et al., 2004); (v) TFIIF is involved in promoter escape and release of RNA pol II pauses (Zhang and Burton, 2004); (vi) TFIIF increases specificity of RNA pol II by reversing the binding of pol II to non-promoter sequence (Orphanides et al., 1996).

1.1.2.5 TFIIE

TFIIE is the next GTF to enter the PIC after TFIIF and the RNA pol II, and is essential for transition from initiation to elongation. TFIIE is composed of two subunits, α and β, of 439 and 291 amino acids in human respectively, which form a heterotetramer. This heterotetramer is also found in yeast. The N-terminal domain of TFIIEα is involved in interaction with TFIIEβ

and RNA pol II, while the C-terminal part is involved in recruitment of TFIIH to the PIC (Kuldell and Buratowski, 1997; Okuda et al., 2004).

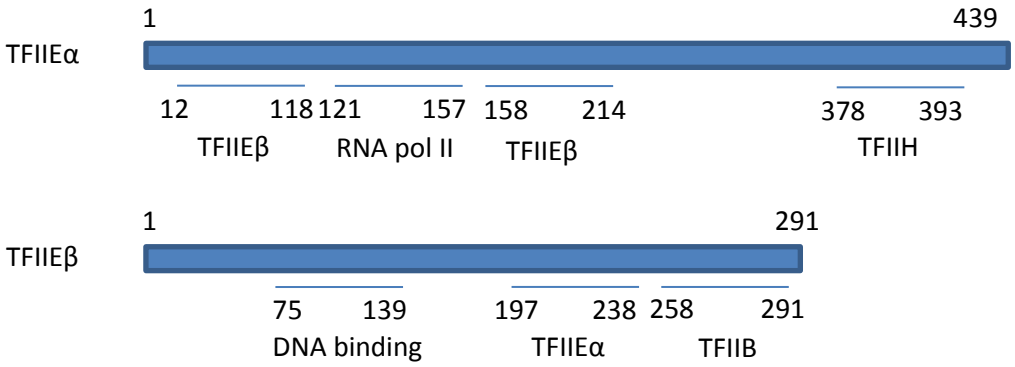


Figure 7 : Domains composition of TFIIE subunits α and β

Position of known interaction domains are indicated. Adapted from (Thomas and Chiang, 2006)

Here again, TFIIE recruitment is driven by different activators (Sauer et al., 1995; Zhu and Kuziora, 1996). Upon binding, TFIIE interacts with TFIIB, TFIIF/RNA pol II and the promoter DNA in order to recruit the last GTF TFIIH (Forget et al., 2004; Watanabe et al., 2003). TFIIE also stimulate the enzymatic activity of TFIIH (see paragraph below) and facilitate promoter melting and thus the transition from the initiation to the elongation (Holstege et al., 1996).

1.1.2.6 TFIIH

TFIIH is the last GTF to be recruited to the PIC during transcription initiation, and its recruitment is mainly driven by TFIIE. In human, TFIIH is composed of 10 subunits, regrouped into two functional subcomplexes: the CAK subcomplex is formed by the subunits CDK7/p40, Cyclin H/p38 and MAT1/p32; the core subcomplex is composed of contains XPB/p89, p62, p52, p44, p34 and p8/TTDA (reviewed in (Egly and Coin, 2011). More recently, the endonuclease subunit XPG was identified as a TFIIH core component in human (Ito et al., 2007). CAK and core TFIIH are linked by the last subunit XPD/p80. TFIIH is strongly conserved as yeast TFIIH was also identified and shows a similar composition. TFIIH carries three enzymatic activities involved in transcription: a kinase activity, carried by the CDK7 subunit which can phosphorylate the CTD of RNA pol II (see paragraph below); an ATP-dependent helicase activity, carried by the two helicase subunits XPD (5' to 3' activity) and XPB (3' to 5' activity); and a DNA-dependent ATPase activity carried by the XPB helicase subunits.

In transcription, TFIIH is required during initiation for promoter opening, promoter escape after the start of the transcription and the synthesis of the first 20

nucleotides, and early elongation (Dvir et al., 2001). These events define the transition from initiation to elongation. The ATPase activity of XPB, but not the helicase, is required for promoter opening at the TSS, whereas helicase activity of XPB but not XPD is required for promoter clearance (Goodrich and Tjian, 1994; Kim et al., 2000; Lin et al., 2005). TFIIH is also responsible for phosphorylation of serine 5 in the CTD of RNA pol II. This activity is mediated by the CAK subunits (Yankulov and Bentley, 1997) and is regulated by several factors such as XPD, TFIIE, the transcriptional co-activator Mediator and some transcriptional activators (Keriel et al., 2002; Svejstrup et al., 1996). CTD phosphorylation on serine 5 is important for switching from initiation to elongation (Hampsey, 1998), and for the recruitment of enzymes required for mRNA processing such as 5' capping (Pei et al., 2001). CDK7 can also phosphorylate other substrates, such as the transcriptional activator p53 and retinoic acid receptors (RARs) in order to stimulate their activity (Lu et al., 1997; Rochette-Egly et al., 1997).

Apart from the role of TFIIH in transcription, the two helicases subunits are important for the DNA repair activity of TFIIH. Indeed, XPD and XPB are involved in Nucleotide Excision Repair (NER), a reparation mechanism of UV-induced DNA damages (Guzder et al., 1994; Schaeffer et al., 1993). Mutations in one of the two helicase subunits lead to severe diseases such as xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome.

The low-resolution structure of the yeast TFIIH core domain has been determined by electron microscopy (EM) and reveals a ring-shaped architecture with the two helicase subunits located on either side of a prominent protrusion (Schultz et al., 2000). A more recent study of the entire complex more accurately locates the two helicases on both side of the core domain, and reveals also the location of the kinase module (Gibbons et al., 2012) (Figure 8)

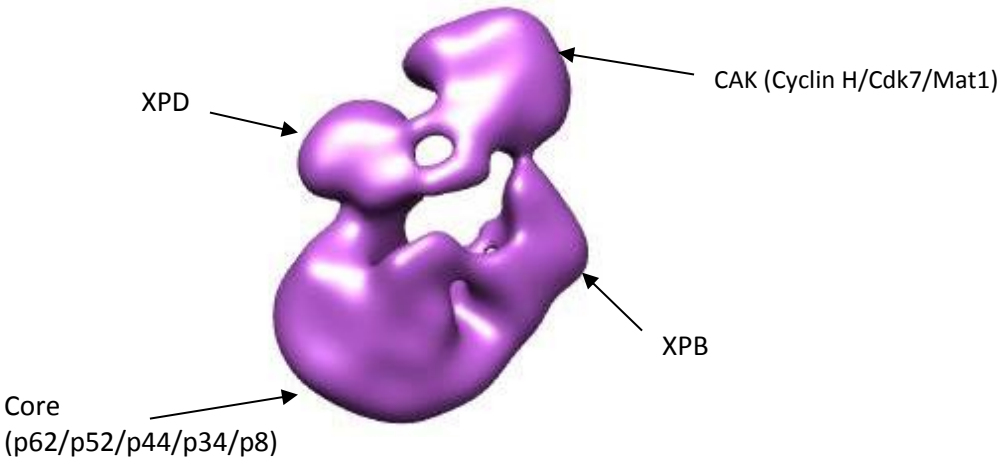


Figure 8 : Molecular organization of TFIIH
TFIIH is composed of a core module and a CAK module (which contains the Cdk7 kinase), linked by the two helicases. Adapted from (Gibbons et al., 2012)

Similarly to others GTF, numerous transcriptional activators such as p53, RAR or Gal4-VP16, can recruit and stimulate TFIIH functions (Zurita and Merino, 2003). This can be done either by an increase in the recruitment of TFIIH to the PIC, or by stimulating the enzymatic activities of TFIIH.

TFIIH is also involved in the response to certain nuclear hormones. Nuclear hormones are small lipophilic molecules recognized by specific nuclear receptors which act as transcription factors upon binding to their response element on the DNA. Such receptors are usually regulated by phosphorylations of specific residues (Treviño and Weigel, 2013), which increase their binding to the DNA and stimulate transcription. The Cdk7 subunit of TFIIH has been shown to regulate the activity of several nuclear receptors, namely RAR (Bastien et al., 2000; Rochette-Egly et al., 1997), the estrogen receptor (Chen and Manley, 2000) or the vitamin D receptor (Drané et al., 2004), and this action can be done by direct phosphorylation of the nuclear receptor, or phosphorylation of one of their binding partner.

1.1.2.7 The RNA polymerase II

The RNA pol II is a central component of the PIC, and interacts with the different GTFs. Upon transcription initiation and beginning of RNA molecule synthesis, the RNA pol II will break these interactions and leave the promoter (promoter clearance), switching from initiation to elongation (and later termination), and associate with another, specific set of factors required for elongation or termination. The RNA pol II is composed of twelve subunits in yeast and human, which are largely conserved in sequence and function during evolution: the subunits Rpb5, 6, 8, 10 and 12 are shared with the others classes of eukaryotic polymerases, while the Rpb4, 7 and 9 subunits are specific to RNA pol II but not essential for

the elongation phase of transcription (Edwards et al., 1991). Rpb4 and Rpb7 form a dissociable module which plays a role during initiation after recruitment of the RNA pol II, by binding the nascent transcript (Meka et al., 2005; Orlicky et al., 2001). The Rpb1 and Rpb2 subunits are specific to RNA pol II and bear the catalytic activity of the enzyme (Lee and Young, 2000).

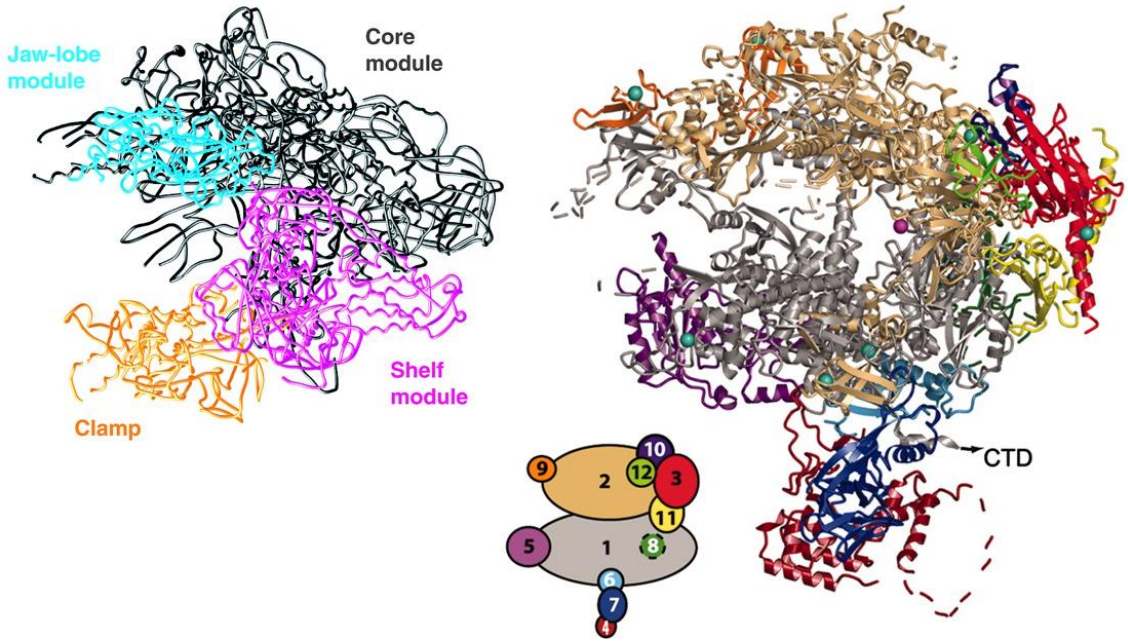


Figure 9: Molecular architecture of RNA pol II
The RNA pol II is composed of four mobile elements, the jaw, the core, the clamp and the shelf as sdescribed on left panel. Subunits forming the different elements are summarized on the right panel. Adapted from (Cramer et al., 2001)

The x-ray structures of the RNA pol II have help to understand the molecular mechanisms of transcription. The overall architecture of the RNA Pol II is composed of four mobile elements termed core, clamp, shelf, and jaw lobe that move relatively to each others (Figure 9) (Cramer et al., 2001, 2000; Kostrewa et al., 2009). The active site of the enzyme, formed by the Rbp1 and Rpb2 subunits, is found at the base of a deep cleft located in the center of the complex. This cleft is formed by all four elements of the RNA pol II and can accommodate the promoter DNA. The clamp element is very mobile and can rotate to open or close the cleft. It was then proposed that the double-stranded DNA do not enter the cleft, but rather only the template strand after separation of the two strands with the help of the GTFs and notably TFIIB (Kostrewa et al., 2009). (Figure 10). Then, the template DNA is scanned for the TSS with the help of the B-reader domain of TFIIB, and the first steps of RNA synthesis can take place.

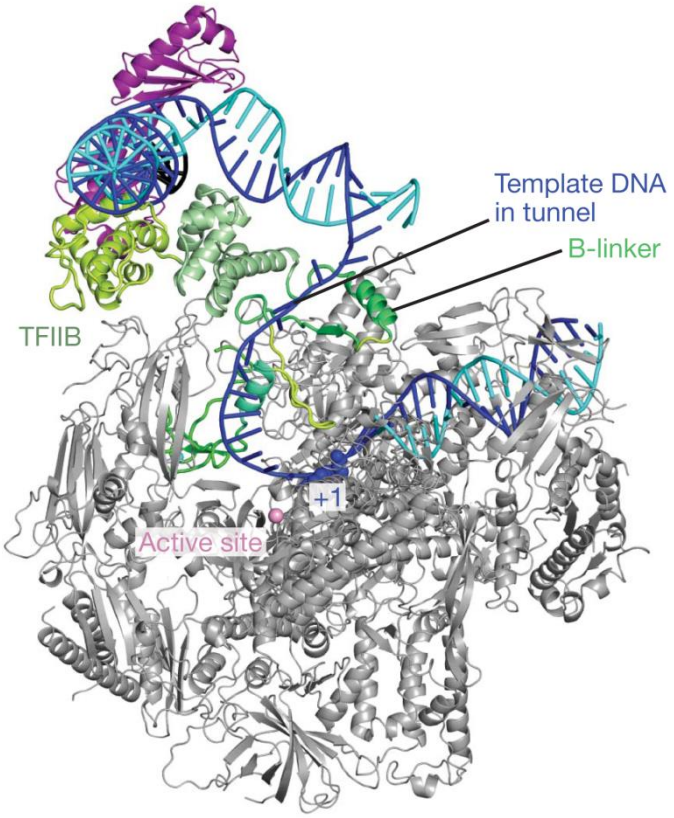


Figure 10: Structure of TFIIB-TBP-RNA pol II complex with opened DNA template
In the structure TFIIB-RNA pol II-TBP complex, the template DNA is positioned in the cleft after melting of the promoter, close to the active site. From (Kostrewa et al., 2009)

An important mechanism of RNA pol II regulation involves the C-terminal domain (CTD) of the Rpb1 subunit. The CTD is an unstructured sequence of the Rpb1 subunit, composed of 25 (yeast) to 52 (human) repeats of the consensus heptapeptide YSPTSPS, and the serines in position 2 and 5 undergo phosphorylation and dephosphorylation events which regulate the activity of the enzyme. The CTD serves as a platform for binding of regulation factors. In its hypophosphorylated state, the RNA pol II is recruited to the PIC. Upon initiation, the CTD is phosphorylated on the serine 5, notably by TFIIH, and this modification marks the transition from initiation to elongation by triggering destabilization of the PIC. Then, in subsequent steps, others kinases such as Cdk9 (present in the Positive Transcription Elongation Factor b, P-TEFb) in human, and Ctk1 in yeast, will phosphorylate serine 2 of the CTD which will promotes the recruitment of factors involved in RNA processing, such as polyadenylation, and termination. GTFs such as TFIIB and TFIIF can interact with phosphatase enzymes which are involved in the dephosphorylation of the CTD. The recruitment and stimulation of these phosphates by the transcription machinery could be a mechanism for facilitating re-initiation at the transcribed genes, by switching the elongating, hyperphosphorylated form of RNA pol II back to the hypophosphorylated form found during initiation.

The RNA pol II was shown to be present in a paused state in about 10% of the genes in *Drosophila* (Muse et al., 2007), which correspond to a state in which RNA pol II has entered the initiation step but is stalled after the synthesis of a short RNA molecule, before resuming to elongation. This paused state seems to be a rate limiting step in transcription and is regulated by the Negative Elongation Factor (NELF) and DRB Sensitivity Inducing Factor (DSIF). Phosphorylation of these two factors by the cdk9 kinase in the context of P-TEFb, releases the RNA pol II from its paused state. It was recently proposed that this pausing provides a temporal window for recruitment of factors involves in the modulation of gene expression (Henriques et al., 2013).

The CTD of RNA pol II is the target of others post-translational modifications, such as glycosylation and ubiquitination. Glycosylation occurs on serine and threonine (mainly threonine 4) and regulates activity of RNA pol II by limiting the accessibility of kinases and others factors to the CTD. Ubiquitination seems to be involved in Transcription-Coupled Repair (TCR) and NER of UV-induced DNA damages.

1.1.3 Structure of the Pre-Initiation Complex

Recently, two studies have revealed structural snapshots of human (He et al., 2013) and yeast (Murakami et al., 2013) PIC, and provided new insights into the molecular organization of PIC and the interactions occurring between the different components.

Human PIC was reconstructed in a step wise manner, by adding sequentially purified TBP, TFIIA, TFIIB, RNA pol II, TFIIF, TFIIE and TFIIH on a core promoter (composed of a TATA-box surrounded by BREu and BREd, and an INR element). Its structure was revealed by EM. In the model obtained, it has been proposed that the binding of TFIIF positions the downstream DNA along the RNA pol II cleft, involving direct interactions of TFIIF with the core promoter DNA. Then the clamp opens in order to accommodate this DNA. Addition of TFIIE seems to stabilize the PIC, and TFIIE, TFIIB and TFIIF are localized in the vicinity of promoter DNA in a closed promoter complex. Addition of TFIIH triggers the transition to an open promoter complex: TFIIH contacts the downstream DNA, and the XPB helicase generate a twist on the DNA which release this constrain by unwinding of the DNA molecule. The clamp closes to position the open template close to the active site, and close to TFIIF and TFIIB which prevent the annealing of the open DNA (Figure 11A).

Structure of yeast PIC revealed by cryo-EM led to a different model. The yeast PIC was reconstructed by mixing the purified GTFs (TBP, TFIIA, TFIIB, RNA pol II, TFIIF, TFIIE and TFIIH) with part of the HIS4 promoter DNA, and the complex was purified on a glycerol gradient. Positioning of the different subunits in the structure was performed according to interaction data obtained from cross-linking experiments and mass spectrometry analysis. The PIC adopts a compact structure composed of a P-lobe (containing the RNA pol II) and the G-lobe (containing the GTFs). The upstream DNA seems to contact TFIIB, TFIIE and TBP while downstream DNA contacts the XPB homologues. No contact between the DNA and the RNA pol II is observed in this model (Figure 11B).

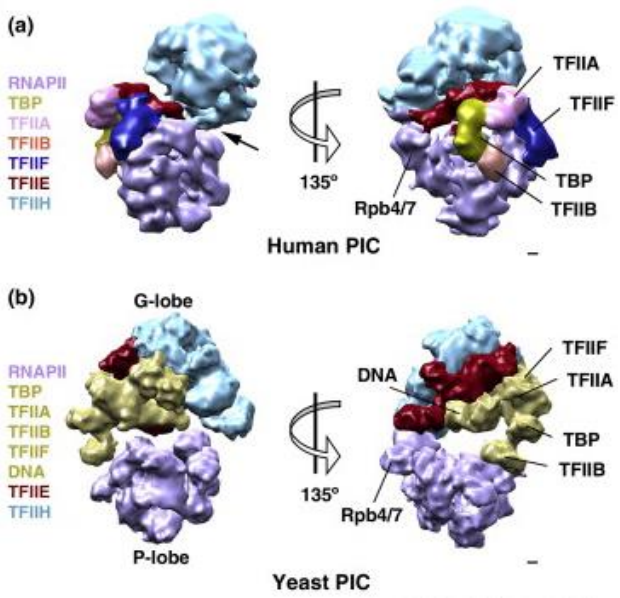


Figure 11 : Comparison of PIC structure in yeast and human

Human (He et al., 2013) and yeast (Murakami et al., 2013) show discripency in the position of TFIIH and of the DNA relative to the RNA pol II. From (Kandiah et al., 2014)

The two models mainly differ by the position of TFIIH and the path of the DNA. It was thus proposed that in the human PIC, TFIIE and TFIIF were absent as the partial complexes formed during the sequential assembly are unstable. The loss of these two subunits led to a different positioning of TFIIH, and consequently a difference in the path of the DNA (Murakami et al., 2013).

1.2 - THE CHROMATIN STRUCTURE AND REGULATION OF TRANSCRIPTION

The first barrier that the cell needs to overcome in order to transcribe gene is inherent to the structure of the chromatin. In the cell, chromatin is condensed and the accessibility of the transcription machinery to the template DNA is limited. Thus, the transition from a compact, inactive state of the chromatin to an open, competent state is a prerequisite for transcription.

The chromatin is formed by the repetition of a single unit called the nucleosome, which is composed of a protein core around which 146 bp of DNA are wrapped and placed every 200 bp on average along the genome. The core is formed by an octamer of histone, comprising two copies of each of the histone proteins, namely H2A, H2B, H3 and H4 (Luger et al., 1997) (Figure 13A). Eventually, a fifth histone (histone H1) contacts the two DNA ends exiting from the octamer in order to stabilize the nucleosome and further compacts the chromatin, resulting in formation of condensed chromatin fibers (Figure 12). The nucleosome plays a direct role in gene regulation by acting as a physical barrier which hampered the transcription machinery to access the DNA template and thus repressing transcription (Han and Grunstein, 1988; Lorch et al., 1987).

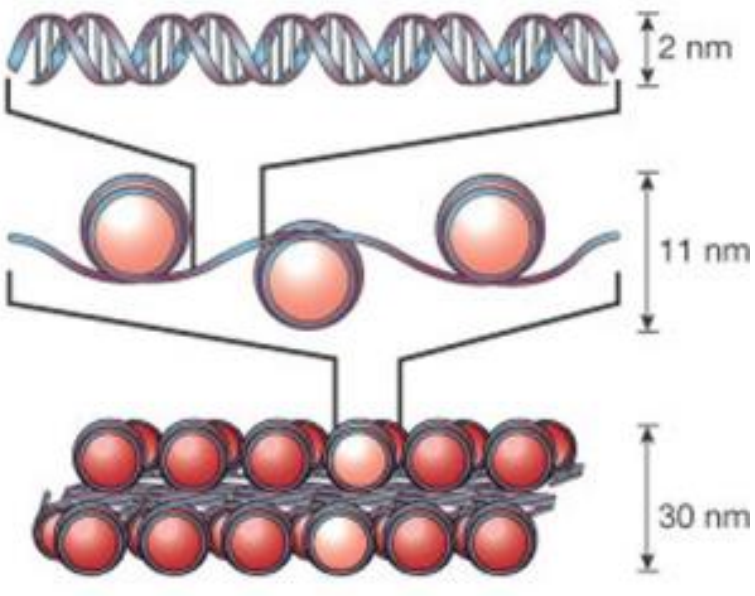


Figure 12 : Structure and compaction of the chromatin

The chromatin is condensed by the wrapping of the DNA molecule around an octamer of histone, and this structure is further compacted by association with histone H1. The compact structure appears as chromatin fibers of 30 nm. From (Felsenfeld and Groudine, 2003)

Histones are among the most conserved proteins in eukaryotes, and are strongly positively charged in order to interact with the negatively charged phosphate backbone of the DNA molecule. The central domain of histone adopts a particular fold, the Histone Fold, which allows protein-protein interaction within the nucleosome. HFD is composed of a long central α helix, flanked by two shorter α helices, and interact with its counterpart by placing the two long α helices in an antiparallel way (Figure 13B). H3 and H4 are forming a heterodimers and will assemble to form a tetramer to which two histone H2A/H2B heterodimers will assemble, forming the histone octamer. The HFD are also involved in the DNA-protein interaction, by contacting the phosphate backbone of the DNA molecule via hydrogen bonds.

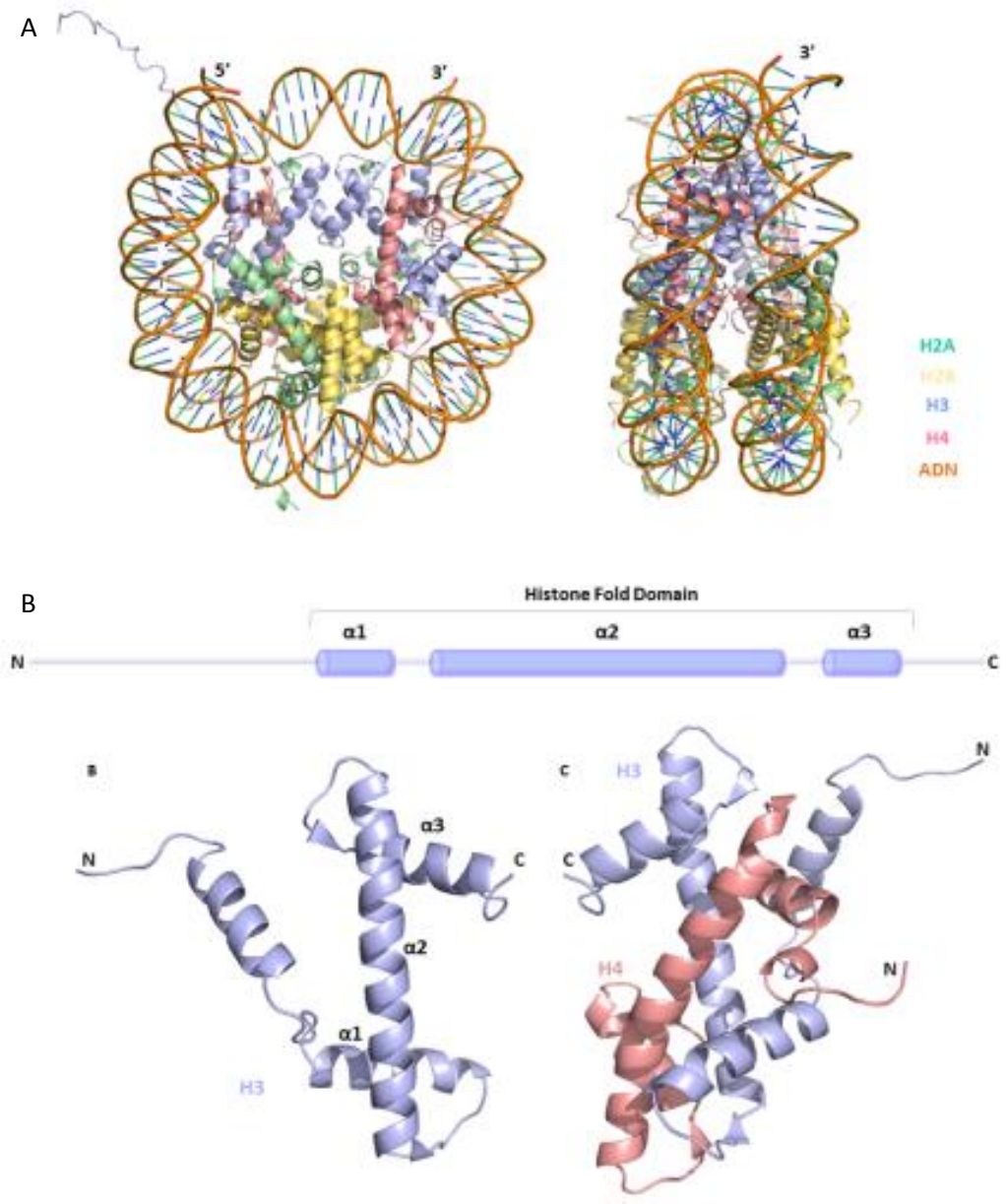


Figure 13 : Structure of the nucleosome

(A) Structure of the nucleosome composed of the histone octamer wrapped around 146 base pairs of DNA. (B) Position of the HFD of histone H3/H4 in the context of the H3/H4 heterodimer.

In addition to the canonical histones present in the nucleosomes, histone variants have been identified for histone H3 and histone H2A (reviewed in (Sarma and Reinberg, 2005)). They are closely related to canonical histones and show minimal differences in their sequences. Histone variants seem to be implicated in specialized processes requiring chromatin rearrangement. Therefore, histone H3 has four variants, called H3.1, H3.2, H3.3 and CENPA. In particular, H3.3 is deposited and integrated into nucleosomes of transcriptionally active regions and CENPA is incorporated in nucleosomes at the centromeres. The four histone H2B variants are H2AZ, H2AX, Macro H2A and H2A Bbd. In particular, H2AZ was shown to localize at transcriptionally active chromatin and to delimit the boundary of euchromatin and

heterochromatin. H2AX seems to be deposited to DNA-damage regions, in order to recruit the DNA repair machinery.

Synthesis and deposition of histones to the DNA template is occurring during DNA replication, except for histone variants which can be produced outside of S phase. They require additional factors to be properly assembled to the correct location. These factors are: (i) histones chaperones, which bind to histone heterodimers and mediate interactions and deposition of the dimers to the core nucleosome; and (ii) chromatin remodelers, which use the energy from ATP hydrolysis to displace nucleosomes and modify the chromatin structure. Chromatin remodelers are separated into four classes, according to their sequence conservation: SWI/SNF, INO80/SWR1, ISWI, and CHD, and can act by sliding or evicting the nucleosome.

The SWI/SNF family of chromatin remodeler includes the RSC (Remodels Structure of Chromatin) complex which is a positive regulator of transcription and localized at the promoter of several genes (reviewed in (Venters and Pugh, 2009)). At the promoter level, histones are acetylated (see below), and the RSC complex (and the members of the SWI/SNF family) contains bromodomains which can help to recruit the complex to its modified nucleosome substrate. Acetylation of histone is thought to destabilize the histone-DNA interactions and thus decompacts the chromatin, making it more accessible to remodelers. Moreover, acetylation of histone can be important for redeposition of histone after remodeling. For example, acetylation of H2A.Z is important for its deposition into promoter nucleosomes (Keogh et al., 2006). In addition to reading the epigenetic marks on the chromatin, the RSC complex can be directly recruited by transcriptional activators and thus participate in the stimulation of transcription.

The precise positioning of nucleosomes on the gene sequence is a major determinant of gene regulation. On the whole, enhancers (see section 1.3 - Role of the Activators in transcription below) and promoters are largely nucleosome-depleted (Yuan et al., 2005). The -1 and +1 nucleosomes (which encompass the TSS) are located at both ends of the promoter region and are precisely positioned, while the position of nucleosomes in the coding sequence is less well-defined, probably due to the dynamic repositioning of these nucleosomes during transcription (Mavrich et al., 2008). The -1/+1 nucleosomes partially occlude the core promoter elements, and thus control the accessibility of the transcription machinery to the promoter. Nucleosomes positioning is dependent on the DNA sequence, and in particular to the presence of di-nucleotides (such as AT pair) which allow large bending of the DNA

(reviewed in (Struhl and Segal, 2013)). The intrinsic properties of the DNA sequence will also affect the dynamic positioning of the nucleosome. Eventually, remodeler enzymes play an important role in the specificity and dynamic of nucleosome placement.

A major aspect of chromatin regulation is linked to the covalent modifications of the tails of histone proteins. In particular, the N-terminal region of histone proteins is a relatively long (about one third of the sequence), unstructured region of the protein and is the target of several histone-modifying enzymes, which dynamically write and erase covalent modifications on specific amino-acids. Such modifications includes (i) acetylation and deacetylation of lysine residues by Histone Acetyl Transferase (HAT) and Histone Deacetylase HDAC, respectively (ii) mono-, di- or tri-methylation of lysine residues or mono-, or di-methylation of arginine residues by Histone Methyl Transferase (HMT); (iii) mono-ubiquitilation and deubiquitination of lysine resisdues by ubiquitin ligase/ubiquitin protease; (iv) phosphorylation of serine and threonine; and (v) sumoylation of lysines. Such modifications can also affect residues in the globular domains of the histone proteins. The methylation of lysine residues has been largely studied, and can be representative of the different states of the chromatin. In general, H3K9, H3K27 H3K64 and H4K20 methylations are the marks of transcriptionally inactive chromatin, whereas H3K4, H3K36 and H3K79 methylations are found on active regions (Hublitz et al., 2009). In particular, H3K4 is trimethylated at the promoter of active gene, where it promotes transcription initiation, whereas it is dimethylated on the gene body, where it allows the recruitment of the mRNA processing machinery (reviewed in (Izzo and Schneider, 2010)). The different marks, their distribution along the gene and their association with the transcriptional state of genes are summarized in Figure 14.

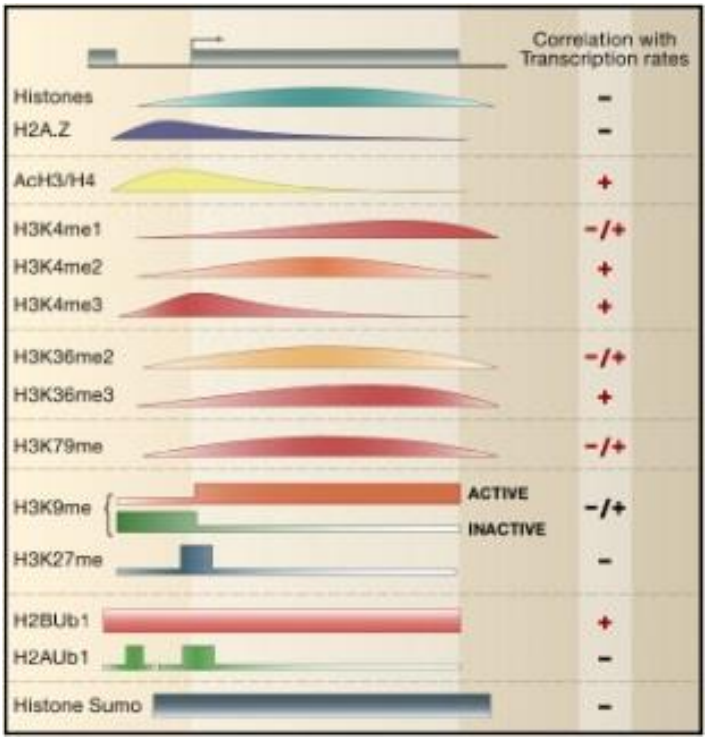


Figure 14 : Histone modifications are marker of the transcriptional state genes.
From (Li et al., 2007)

All these modifications do not act independently but rather communicate in a complex cross-talk which regulates further the different functions of the chromatin. The modification of a residue can regulate either positively or negatively the addition or removal of another modification. An example of such a cross-talk is the tri-methylation of H3K4 by the Set1 HMT which requires first the mono-ubiquitilation of lysine 123 on H2B by the Rad6 ubiquitin ligase (Dover et al., 2002) during transcription initiation. This methylation is blocked by the methylation of the arginine residue at the position 2 in close vicinity of K4, and is recognized by specialized domain in the SAGA complex (see below part 1.4 - The Transcriptional Co-activator SAGA) which will in turn acetylate specific lysine residues on histone H3. In addition, SAGA contains an ubiquitin protease which will remove the ubiquitin from histone H2B. This step is required for recruitment of Ctk1, which will phosphorylate the serine 2 on the CTD of RNA pol II, and triggers the transition to the elongation step of transcription (Venters and Pugh, 2009; Wyce et al., 2007) (Figure 15).

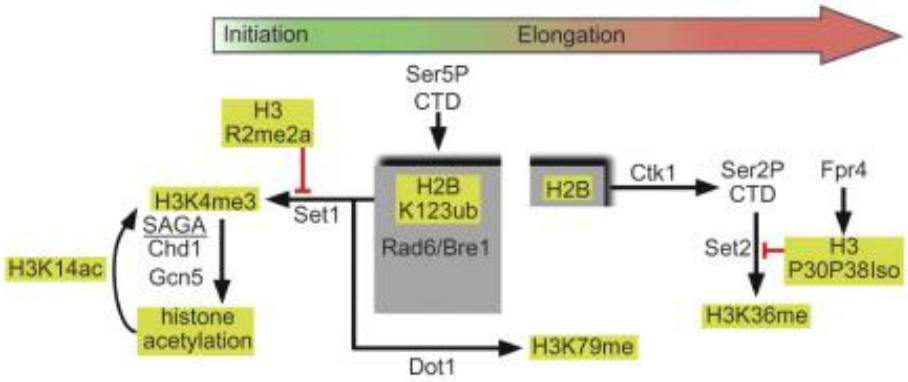


Figure 15 : Cross-talk between H2B ubiquitilation and methylation of histone 3.
Ubuiqtuilation of H2B is a prerequisite for H3K4-Me3 and will trigger a set of downstream events resulting in the switch from initation to elongtaion. Adapted from (Venters and Pugh, 2009).

1.3 - ROLE OF THE ACTIVATORS IN TRANSCRIPTION

The basal transcription machinery, composed of the previously described GTFs and the RNA pol II, is necessary and sufficient to promote transcription initiation *in vitro* on naked DNA template. However, i*n vivo*, basal transcription is very low. Transcription of eukaryotic genes is regulated by sequence-specific DNA binding proteins, which stimulate or repress transcription upon binding to their recognition site. An activator can stimulate transcription at several promoters, triggering the coordinated expression of a particular subset of genes, or multiple activators can be required to enhance expression at the same promoters, thus integrating different signals in order to set the required transcription level.

A typical activator contains two functional domains: (i) a sequence-specific promoter targeting region, which specifically recognizes a DNA sequence upstream of the promoter (UAS); and (ii) an activation domain that interacts with transcription factors to promote transcription. Several classes of activators can be defined according to the amino-acid content of the activation domain. Acidic activators (e.g. Gal4, Gcn4 or VP16) are rich in aspartate and glutamate, and are the most represented class of activators. Other activation domain such as glutamine-rich (e.g., SP1), proline-rich (e.g. NF1), and zinc finger-containing (e.g. AdE1a) were also described in eukaryotes. As already mentioned, nuclear receptors are also transcriptional activators, and their activation domain are characterized by the LXXLL motif (Heery et al., 1997)

Usually, activators act by stimulating PIC assembly (Ptashne and Gann, 1997). As previously discussed, each GTF are targeted by many activators which can stimulate the recruitment of the transcription factors during formation of the PIC. Nevertheless, activators can also stimulate transcription by promoting important steps after PIC assembly, such as elongation, reinitiation or chromatin modifications that facilitate transcription. These functions are largely performed by the direct recruitment of others factors, called transcriptional co-activators, which can interact with the component of the PIC as well as modify the chromatin.

1.3.1.1 TFIID

In the cell, TBP is associated to large complex, such as TFIID. TFIID is composed of TBP and 14 TBP-associated factors (TAFs), and is conserved in eukaryotes as TAFs were also identified in human and *Drosophila* (summarized in Figure 16). TFIID is responsible for expression of about 90% of the genes in yeast, most of which being housekeeping genes regulated by TATA-less promoters (Basehoar et al., 2004). In metazoan, it is used at a wider variety of promoters (Grünberg and Hahn, 2013). Moreover, paralogues of TBP and Tafs have been discovered, and seems to incorporate in TFIID complexes which are responsible for tissue- or cell-specific expression of certain genes (Freiman, 2009; Gazdag et al., 2009; Kolthur-Seetharam et al., 2008). The role of TFIID in transcription is dual: (i) as already discussed, TFIID is involved in promoter recognition thought the specific interactions of several Tafs with core promoter elements. The association of TBP, which recognize the TATA-box, to a set of proteins capable of binding different sequences allows TFIID to broaden its scope of promoter recognition, in particular at promoters lacking a canonical TATA element, where TFIID has been shown to be essential for transcription (Huisinga and Pugh, 2004).(ii) TFIID is also a coactivator as it mediates the interaction between some activators and the transcription machinery, in order to stimulate transcription.

			C. elegans			
New name	H. sapiens	D. melanogaster	previous name	new name	S. cerevisiae	S. pombe
TAF1	TAFII250	TAFII230	taf-1	taf-1	Taf145/130	TAFII111
TAF2	TAFII150	TAFII150	taf-2	taf-2	Taf150 or TSM1	T38673
TAF3	TAFII140	TAFII155orBIP2	C11G6.1	taf-3	Taf47	
TAF4	TAFII130/135	TAFII110	taf-5	taf-4	Taf48 or MPT1	T50183
TAF4b	TAFII105					
TAF5	TAFII100	TAFII80	taf-4	taf-5	Taf90	TAFII72
TAF5b						TAFII73
TAF5L	PAF65	Cannonball				
TAF6	TAFII80	TAFII60	taf-3.1	taf-6.1	Taf60	CAA20756
TAF6L	PAF65	AAF52013	taf-3.2	taf-6.2		
TAF7	TAFII55	AAF54162	taf-8.1	taf-7.1	Taf67	TAFII62/PTR6
TAF7L	TAF2Q		taf-8.2	taf-7.2		
TAF8	BAB71460	Prodos	ZK1320.12	taf-8	Taf65	T40895
TAF9	TAFII32/31	TAFII40	taf-10	taf-9	Taf17	S62536
TAF9L	TAFII31L					
TAF10	TAFII30	TAFII24	taf-11	taf-10	Taf25	T39928
TAF10b		TAFII16				
TAF11	TAFII28	TAFII30β	taf-7.1	taf-11.1	Taf40	CAA93543
TAF11L			taf-7.2	taf-11.2		
TAF12	TAFII20/15	TAFII30α	taf-9	taf-12	Taf61/68	T37702
TAF13	TAFII18	AAF53875	taf-6	taf-13	Taf19 or FUN81	CAA19300
TAF14					Taf30	
TAF15	TAFII68					

Figure 16 : Summary of Tafs present in the different species and their nomenclatures
According to (Tora, 2002)

Biochemical experiments and immuno-labelling revealed by electron microscopy have demonstrated that a subset of Tafs, namely Taf4, Taf5, Taf6, Taf9, Taf10 and Taf12, is present in two copies, whereas the remaining six Tafs and TBP are present in a single copy (Leurent et al., 2004, 2002; S. L. Sanders et al., 2002). Moreover, nine Tafs contains HFD involved in protein-protein interactions and can form heterodimers. In particular, the Taf4, Taf6, Taf9 and Taf12 were shown to form an octamer similar to the histone octamer *in vitro*, but it is unlikely that this structure exits in TFIID as EM studies could not identified such structure (Leurent et al., 2004, 2002; Selleck et al., 2001). The TAFs present in two copies are thought to form a structural core important for TFIID structure and integrity, around which the other subunits assembled in a peripherical location (Bieniossek et al., 2013; Cler et al., 2009). The largest subunit in TFIID, namely TAF1, interacts with TBP via its N-terminal domain (Taf N-terminal domain, TAND). TAF1 also contains a HAT domain (which can acetylate histone, (Mizzen et al., 1996)) and two bromodomains (which can recognize acetylated lysine, (Jacobson et al., 2000)), whereas TAF3 contains an atypical PHD (Plant Homeo Domain) domain (binds tri-methylated lysine). This suggests a role for TFIID in writing and reading specific post-translational modifications on nucleosomes, thus participating in promoter recognition in the context of the chromatin.

However the HAT activity of Taf1 doesn't seem to contribute significantly to histone acetylation genome-wide (Durant and Pugh, 2006). Additional enzymatic activities have been attributed to TAF1, namely kinase, and ubiquitin-conjugating activities, which can target histone- and non-histone targets. In particular, TAF1 seems to be involved in post-translational modification of TFIIE, TFIIA, histone H2B, TAF5 and itself.

Many activators of transcription have been shown to directly interact with specific TAFs. The Rap1 activator, which was shown to regulate the expression of the ribosomal protein genes (Lieb et al., 2001) interacts with TFIID via the Taf4, Taf5 and Taf12 subunits (Garbett et al., 2007; Layer et al., 2010; Papai et al., 2010). The interaction of TFIID with the activators p53, Sp1 and the oncoprotein c-Jun were revealed by EM and cross-linking experiments (Liu et al., 2009). Interestingly, these activators were shown to bind to different sites within TFIID, and to contact different Tafs subunits. Thus, p53 seems to contact Taf1, Taf6 and Taf9, while Sp1 makes contacts with Taf1 and Taf4, and finally Sp1 is contactingTaf1 and Taf6. This implies that activators may enhance transcription by stimulating formation of the PIC, and Tafs seems to play a crucial role in relaying the information from activator to TBP.

The structure of TFIID has been revealed by EM studies (Brand et al., 1999a; Leurent et al., 2004, 2002; Papai et al., 2009). On the whole, TFIID structure is composed of three lobes and forms an asymmetric clamp of about 200-220Å (Figure 17). TBP was mapped at the center of the clamp, close to the N-terminal domain of TAF1, in good agreement with interaction data. HFD in TAFs seems important for complex integrity, as pairs of histone-like TAFs can be detected in TFIID. TAF5 plays a central role by bridging the three lobes, and serves as a scaffold for assembly of the remaining TAFs. The structure seems very flexible, and the different domains can adopt several conformations, probably to compensate for the variability in promoter.

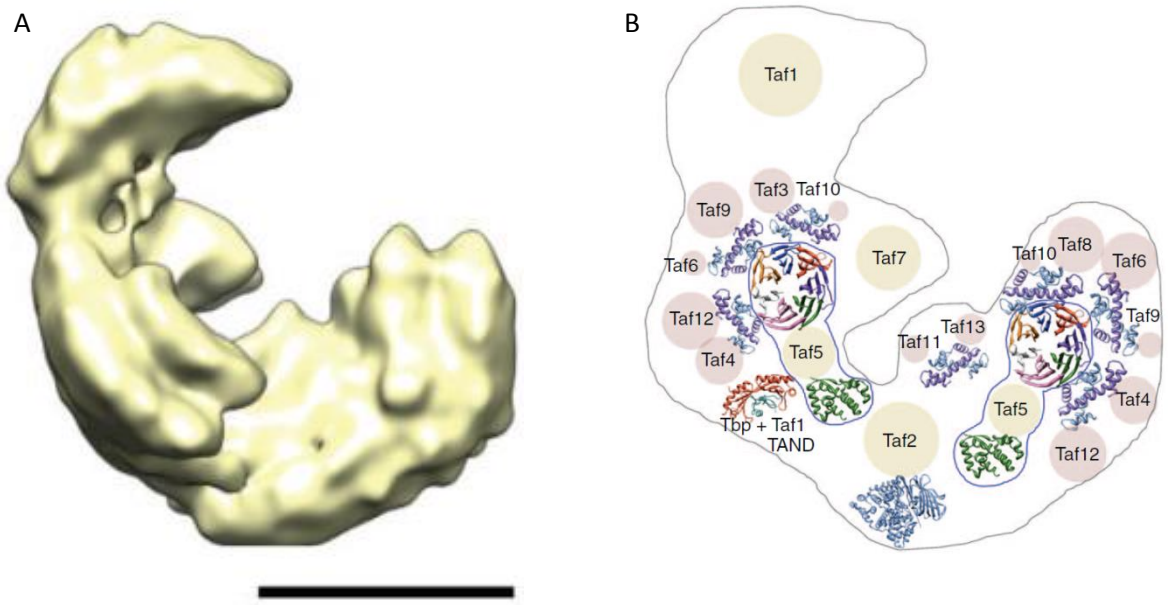


Figure 17 : Molecular architecture of TFIID

(A) 3D structure of yeast (S. cerevisiae) TFIID, revealed by EM. (B) Subunits are mapped according to immuno-labelling experiements. From (Papai et al., 2011). Black bar is 10 nm.

1.3.1.2 The Mediator complex as a co-activator

The first evidence that activators were not acting only by direct recruitment of the GTFs was given by the finding of an interference between two activators (Gill and Ptashne, 1988). The defect in transcriptional activation could not be suppressed by an excess of any GTF or the RNA pol II itself, but by an activity present in a fraction of a crude yeast extract (Kelleher et al., 1990). The activity in this fraction was termed Mediator, and was thought to serve as an adaptor between the activators and the transcription machinery. Mediator was finally purified in yeast (Kim et al., 1994), and shown to be a large complex of one Mega-Dalton MDa composed of 20 proteins. Nine out of the twenty subunits were first identified in genetic screen as suppressor of CTD truncation phenotype, namely a transcriptional defect caused by the truncation of the RNA pol II CTD (Nonet and Young, 1989), and this finding was confirmed by the analysis of the purified Mediator. Eleven genes of the Mediator are essential for yeast viability, highlighting its important role in transcription process. Subsequently, orthologs of almost all subunits of the yeast Mediator were found in higher eukaryotes. The mediator is found at most promoter in yeast, and its localization correlates with RNA pol II, suggesting a central role of the Mediator complex in transcription.

Mediator supports transcription activation by bridging the different components of the transcription machinery, in particular GTFs and RNA pol II, to DNA-bound transcription activators (reviewed in (Conaway and Conaway, 2011). Mediator is also capable to stimulate basal transcription *in vitro*, independently of activators (Baek et al., 2002). Early structural

studies by EM (Asturias et al., 1999) has shed light on the molecular organization of the complex. Mediator is formed of three domains, termed head, middle and tail, forming the core mediator, associated to a fourth kinase module which is not always present. Core mediator can bind the RNA pol II to form the Holoenzyme (Figure 18 and Figure 19).

Module	Subunit	*S. cerevisiae*	*H. sapiens*
Head	Med6	Med6	hMed6/DRIP33/p32
	Med8	Med8	ARC32/mMed8
	Med11	Med11	HSPC296
	Med17	Srb4	TRAP80/DRIP77/CRSP77/p78
	Med18	Srb5	p28b
	Med20	Srb2	hTRFP/p28a
	Med22	Srb6	Surf5
Head/Middle	Med19	Rox3	LCMR1
Middle	Med1	Med1	TRAP220/DRIP205/CRSP200/PBP
	Med4	Med4	TRAP36/DRIP36/p34
	Med7	Med7	hMed7/DRIP34/CRSP33/p36
	Med9	Med9/Cse2	Med25
	Med10	Med10/Nut2	hMed10/hNut2
	Med21	Srb7	hSrb7/p21
	Med31	Soh1	hSoh1
Middle/Tail	Med14	Rgr1	TRAP170/DRIP150/CRSP150/p110
Tail	Med2/29	Med2	Hintersex
	Med3/27	Med3/Pgd1/Hrs1	TRAP37/CRSP34
	Med5/24	Nut1	TRAP100/DRIP100/CRSP100
	Med15	Gal11	ARC105/PCQAP/TIG-1
	Med16	Sin4	TRAP95/DRIP92/p96b
	Med23	–	TRAP150β/DRIP130/CRSP130/hSur2
Kinase	Med12	Srb8	TRAP230/DRIP240
	Med13	Srb9	TRAP240/DRIP250
	CDK8	Srb10/Ssn3/Ume5	hSrb10/CDK8
	CycC	Srb11/Ssn8/Ume3	hSrb11/CycC
Unassigned	Med25	–	ARC92/ACID1
	Med26	–	ARC70/CRSP70
	Med28	–	Fksg20
	Med30	–	TRAP25

Figure 18 : Subunits compostion of yeast and human Mediator complex.
Adapted from (Larivière et al., 2012)

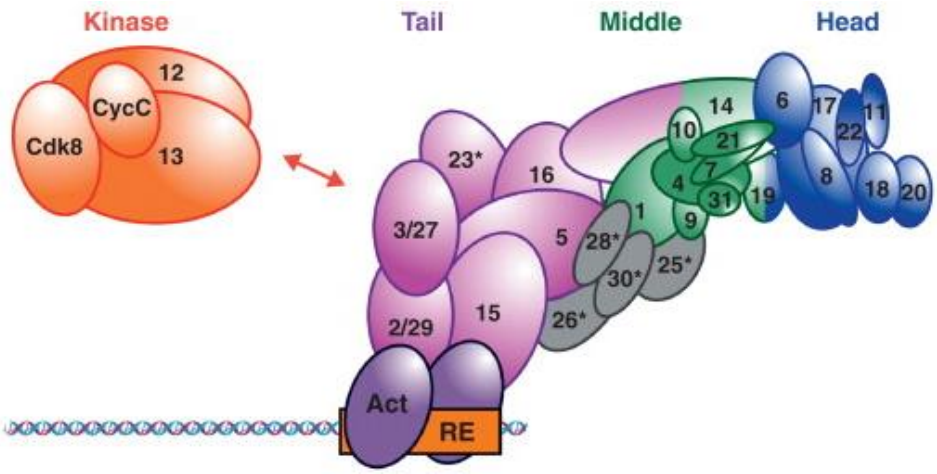


Figure 19 : Molecular organization of yeast and human Mediator complex.
Adapted from (Larivière et al., 2012)

Upon binding of RNA pol II, the head of Mediator complex undergoes a large change in conformation, resulting in switching from a compact to an extended form of the complex, where the head serves as an anchor point for the RNA pol II. Indeed, the MED17/Srb4, MED18/Srb5, MED20/Srb2 and MED22/Srb6 subunits of the Mediator were shown to interact with subunits of the RNA pol II by yeast genetic screen (Soutourina et al., 2011; Thompson et al., 1993). The head contains subunits shown to interact with activators. However, others subunits of the middle and tails domains were also identified as activators interacting partners. Others contacts between the RNA pol II and Mediator middle and tail modules were also described (Davis et al., 2002).

In addition to its role in recruiting the RNA pol II, Mediator was shown to interact with GTFs, expanding its role in stabilizing the PIC on the promoter. The Med11 and Med22 subunuits from the head module were shown to recruit TFIIH to the promoter (Esnault et al., 2008; Seizl et al., 2011). The head module is also able to bind TBP through the Med8 subunit (Kang et al., 2001; Larivière et al., 2006). An interaction between the Mediator and TFIID was reported (Takahashi et al., 2011), and this interaction seems mediated by the Med26, away from the TBP interaction site. However, it is possible that interaction with TFIID involves other subunits (Takahashi et al., 2011). All these interactions are thought to stabilize the PIC to the promoter by tethering Pol II to TBP, TFIIH and TFIID.

Whereas Mediator positively stimulate transcription through interactions with the RNA pol II, a negative role was attributed to the kinase module (composed of the Med12, Med13, CDK8 and CycC subunits) of the Mediator complex. CDK8 was shown to phosphorylated serine 2 and 5 on the CTD of RNA pol II, prior to PIC assembly. Binding of the kinase module could also block the assembly of the holoenzyme by occluding RNA pol II binding site or by an allosteric mechanism (Elmlund et al., 2006).

1.3.1.3 Example of a transcriptional activator: the Gal4 protein

In yeast, the induction of the GAL genes, involved in galactose metabolism, has been a model system to understand the mechanisms of transcriptional activation. Galactose is used as a carbon source if neither glucose nor other sources are available, and will stimulate expression of the genes GAL1, GAL2, GAL7 and GAL10, involved in galactose transport and metabolism, under the control of the Gal4 activator. Gal4 binds as a dimer to its UAS composed of 17 nucleotides (CGG-N_5-T-N_5-CCG) in the promoter of the related genes (Hong et al., 2008; Marmorstein et al., 1992), and will stimulate over 1000 times the expression of the GAL genes. The UAS can be present in several copies in the promoter of each gene, and this

number correlates with the extend of transcriptional activation (Lohr et al., 1995; Melcher and Xu, 2001). Thus, Gal3 gene contains a single UAS in its promoter, and shows a high basal expression level and a weak (3 to 5 fold) induction level. On the opposite, the Gal1 promoter (shared with the Gal10 gene) contains 4 UAS and shows a very weak basal transcription with a 1000 fold increase upon induction.

The regulation of the GAL genes is based on the interaction of the GAL4, GAL80 and GAL3 gene products. Gal4 is a protein of 881 amino-acids in yeast and composed of a DNA-binding domain, a dimerization domain and two acidic activation domains. Gal4 is inactive in the absence of glucose, owing to the binding of the Gal80 protein to the Gal4 activation domains. Upon galactose induction, the Gal3 protein binds to Gal80 thus releasing Gal4, which is then able to interact with the transcription machinery. The binding of Gal3 to Gal80 is thought to occur in the cytoplasm, which was proposed to decrease the pool of Gal80 in the nucleus and thus the formation of an inhibitory Gal80-Gal4 complex (Figure 20).

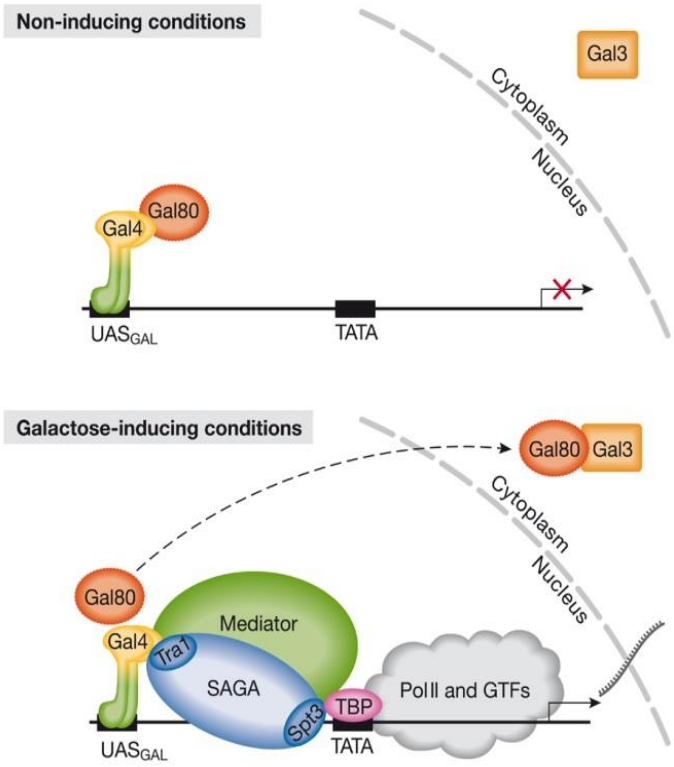


Figure 20 : Mechanism of Gal4 activation upon galactose induction.
In the absence of galactose, Gal80 binds to Gal4 and blocks the activation of transcription. After induction, Gal80 is sequestered by Gal3, and Gal4 interacts with component of the PIC and coactivator to stimulate transcription. From (Traven et al., 2006)

Upon activation, Gal4 interacts with components of the transcription machinery to stimulate formation of the PIC, and identification of the *in vivo* targets of Gal4 has been of major importance to understand the activation mechanism of Gal4 (reviewed in (Traven et al.,

2006)). Many components of the PIC, as well as coactivators, were found to interact with Gal4 in *in vitro* binding assay: TBP (Melcher and Johnston, 1995; Wu et al., 1996), TFIIB (Wu et al., 1996), members of coactivator complexes such as Mediator (Gal11, Cdk8, Srb4) (Jeong et al., 2001; Park et al., 2000) or SAGA (Tra1, Taf12) (Bhaumik and Green, 2001; Brown et al., 2001; Larschan and Winston, 2001). In addition, the chromatin remodeler Swi/Snf was also found to interact with Gal4 activator (Neely et al., 2002). However, chromatin-immuno precipitation (ChIP), Fluorescence resonance energy transfer (FRET) experiments and fluorescence complementation assay have help to understand the *in vivo* relevance of such interactions. Indeed, only the Tra1 component of the SAGA complex (see section1.4 -) has been identified as the direct *in vivo* target of Gal4 in this assay (Bhaumik et al., 2004; Lin et al., 2012). This result is in agreement with the finding that SAGA is essential for GAL genes transcription (Bhaumik and Green, 2001). In contrast, even though recruitment of the Mediator complex has been shown to occur on the GAL1 gene, it is still not clear if it is directly targeted by the Gal4 activator, or if its recruitment is dependent on SAGA (Bryant and Ptashne, 2003; Larschan and Winston, 2005). Thus, a model for activation by the Gal4 protein has been proposed, where DNA-bound Gal4 recruits SAGA to the promoter by direct interaction with the Tra1 protein, which in turn recruits the Mediator complex and other components of the PIC, such as TBP and TFIIB, and RNA pol II (Bhaumik et al., 2004).

1.4 - THE TRANSCRIPTIONAL CO-ACTIVATOR SAGA

The first obstacle that the transcription machinery has to overcome during transcription is formed by the structure of the template itself and needs to alleviate the protection of the DNA molecule provided by the nucleosomes. This will be achieved by the recruitment of enzymes which can alter the structure of the chromatin, either by modifying post-translationally the histone tails which compose the nucleosomes, or by direct remodeling of the chromatin.

The first link between histone modifications and transcription activation was set by the discovery that the Gcn5 protein, previously identified as a transcriptional coactivator in a genetic screen (Georgakopoulos and Thireos, 1992), was homologous to a protein of *Tetrahymena thermophile* which display a HAT activity (Brownell et al., 1996). A few years later, Gcn5 was discovered to be part of a large, multisubunit complex, composed of several sets of proteins involved in transcription regulation (Grant et al., 1997). This complex was then named

according to its protein content: the Spt-Ada-Gcn5 Acetyl-transferase. The Ada (alteration/deficiency in activation) proteins were first identified by yeast genetic studies as a class of adaptor proteins involved in the response to the activation domain derived from the herpes simplex virus VP16, which is toxic when expressed in yeast (Berger et al., 1992). A subset of Ada proteins, namely Ada2, Ada3 and Ada5, were identified in the SAGA complex (Berger et al., 1992; Marcus et al., 1996; Pina et al., 1993). Independently, the Spt (suppressor of Ty) group of proteins was identified by a genetic screen for suppressors of transcription defects caused by insertion of a Ty element (Winston and Carlson, 1992). SAGA comprises a subset of Spt proteins (Spt3, Spt8, Spt7 and Spt20) related to TBP function (Eisenmann et al., 1994; Roberts and Winston, 1996; Winston et al., 1987). Interestingly, TBP was also found to be part of the Spt protein (Eisenmann et al., 1989). The discovery of the SAGA complex provides a biochemical validation of the genetic links which exist between these two functional classes of proteins, as Spt20 was shown to be identical to Ada5 (Roberts and Winston, 1996).

Since its discovery, the SAGA complex has been the classical example of the transcriptional co-activator complexes. Numerous studies have helped to identify the composition and functions of this complex. Quickly after its first description, a subset of TAFs shared with TFIID was shown to be integral components of the SAGA complex (Grant et al., 1998a), broadening the role and functions of these factors during transcription initiation. The Tra1 subunit (Grant et al., 1998b) was also identified in the SAGA complex, and was shown to be homologs to the human TRRAP, a protein involved in interactions with activators such as the c-Myc oncoprotein. Large-scale proteomics approaches have identified new components of the SAGA complexes such as Sgf29 (SAGA-associated factor 29) (S. Sanders et al., 2002). An additional role in transcription and chromatin modification was attributed to the SAGA complex, by the finding that the Ubp8 protein, an ubiquitin protease involved in histone H2B deubiquitination, was a full component of the SAGA complex (Daniel et al., 2004; Henry et al., 2003) and its activity is dependent on its incorporation into the complex (K. K. Lee et al., 2005). Three others proteins were shown to be required for Ubp8 activity in the context of SAGA: Sgf11 (K. K. Lee et al., 2005), Sus1 (Rodríguez-Navarro et al., 2004) and Sgf73 (McMahon et al., 2005). The list of the different subunits is shown in Figure 21.

S. cerevisiae	Human	Drosophila	Function
Tra1	TRRAP	Tra1	Activator binding
Spt7	Spt7/SUPT7L/STAF65	Spt7	Structural role, bromodomain
Spt20	SPT20	Spt20	Structural role
Ada1	Tada1/STAF42	Ada1	Structural role, Histone Fold
Spt3	Spt3/SUPT3	Spt3	TBP binding
Spt8			TBP binding
Gcn5	GCN5L/KAT2A	Gcn5	Histone Acetyltransferase, bromodomain
Ada2	Tada2b	Ada2b	Histone Acetyltransferase module
Ada3	Tada3	Tada3	Histone Acetyltransferase module
Sgf29	Sgf29	Sgf29	Histone Acetyltransferase module, Tandem Tudor Domain
Taf5	Taf5L	Wda	Structural role, WD40
Taf6	Taf6L	Sda	Structural role, Histone Fold
Taf9	Taf9	Taf9	Structural role, Histone Fold
Taf10	Taf10	Taf10	Structural role, Histone Fold
Taf12	Taf12	Taf12	Structural role, Histone Fold
Ubp8	Usp22	Nonstop	Histone deubiquitination
Sgf73	Atxn7	CG9866	Histone deubiquitination
Sgf11	Atxn7L3		Histone deubiquitination
Sus1	ENY2	E(y)2	Histone deubiquitination

Figure 21 : Composition of the SAGA complex

Summary of subunit composition of the SAGA complex in yeast, human and drosophila. Proteins are regrouped according to their function.

1.4.1 Conservation of the SAGA complex during evolution

The SAGA complex has been highly conserved during evolution. Homologues of almost all SAGA components have been identified in higher eukaryotes, and seem to incorporate into a large complex equivalent to the yeast SAGA. The first evidence of the presence of a SAGA complex in metazoan was discovered in human cells, where the TBP-free TAFIIs Complex (TFTC) was first identified as a fraction of TAFs which was depleted in TBP subunit (Wieczorek et al., 1998). This fraction also contains GCN5, ADA3, SPT3 and TRRAP (Brand et al., 1999b) and could correspond to a human equivalent of the yeast SAGA. In parallel, the PCAF (Ogryzko et al., 1998) and STAGA (SPT3-TAFII31-GCN5-L acetyltransferase) (Martinez et al., 1998) complexes, containing respectively the PCAF and GCN5 acetyltransferase, were purified. These complexes were characterized as large multisubunit complexes, containing homologues of yeast SAGA subunits such as Ada2, Ada3, Spt3 or Tafs, and thus were thought to be the vertebrate equivalent of the yeast SAGA complex. As new subunits of the SAGA complex were discovered in yeast (i.e. Ubp8, Sgf11, Sus1, Sgf73 and Sgf29), homologues of these subunits were also discovered in vertebrates (respectively USP22, ATX7L3, ENY2, ATX7 and SGF29 in human) and seems to be integrated in TFTC/STAGA complexes (Helmlinger et al., 2004; Kurabe et al., 2007; Palhan et al., 2005; Zhang et al., 2008; Zhao et al., 2008). Eventually, these complexes seemed to be equivalent, and subtle changes in subunit composition could be explained by

differences in the purification protocols of such complexes. Thus, it was proposed to consider these complexes as human SAGA (Pijnappel and Timmers, 2008)

Whereas in yeast the subunit content of the complex seems to be well defined (Lee et al., 2011; Mischerikow et al., 2009), in higher eukaryotes an additional degree of complexity has been revealed by the discovery of paralogues of several components of the human and *Drosophila* SAGA complexes. Indeed, several genes were found to have undergone duplication events during evolution (reviewed in (Spedale et al., 2012)). The Taf5, Taf6, Taf9 and Taf10, as well as Gcn5 and Ada2, are unique in yeast and Tafs incorporate in both SAGA and TFIID. In human, TAF5 and TAF6 duplicates have diverged into TAF5L and TAF6L, which are SAGA-specific. Taf9 gave rise to TAF9 and TAF9B, which can incorporate in both complexes. In D*rosophila*, Taf10 has diverged into Taf10 and Taf10B, and Taf10 only is incorporating to the SAGA complex (Georgieva et al., 2000).

A major variation exists in the HAT module of the SAGA complex in higher eukaryotes such as *Drosophila* and human. In human, GCN5 gene has been duplicated to form the PCAF gene, encoding a second HAT characterized by an N-terminal PCAF homology domain, which can replace GCN5 in SAGA. Moreover, duplication of Ada2 in human gave rise to the ADA2a gene, encoding a protein exclusively incorporated in the ATAC (Ada-Two-A-Containing) coactivator complex, while ADA2b is SAGA-specific (Nagy and Tora, 2007; Zita Nagy et al., 2010). These complexes were shown to regulate distinct sets of genes (Z Nagy et al., 2010). The increase diversity of such complexes in higher eukaryotes can reflect the higher complexity of such organism, which requires the regulation of tissue-specific sets of genes during development (Nagy and Tora, 2007; Spedale et al., 2012).

Another evidence for the conservation of the SAGA complex during evolution was provided by the structural comparison of yeast SAGA and human TFTC (Figure 22) (Brand et al., 1999a; Wu et al., 2004). Both complexes show the same molecular organization, and the HAT activity was located in the same position, highlighting the conserve architecture and thus the probable conserved functions during evolution.

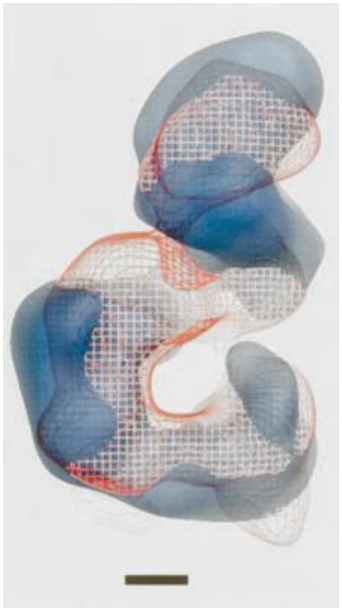


Figure 22 : Conserved structure of SAGA homologues
Superposition of SAGA (blue) and TFTC (mesh, in red) low resolution structures shows the conserved architecture of the complex during evolution. From(Wu et al., 2004)

Interestingly, deletion of SAGA subunits in yeast is not lethal and show synthetic growth defects only when combined with other mutations, with the remarkable exception of Tra1 (Saleh et al., 1998) and Tafs (Green, 2000) which are essential for yeast cell viability. In metazoan, SAGA seems to play important roles in the development. In particular, individual loss of the SAGA subunits Gcn5, Ada2b, Ada3, WDA, Sgf11, Nonstop and SAF6 results in developmental defects and larval lethality in *drosophila* (Carre et al., 2005; Guelman et al., 2006; Pankotai et al., 2005; Qi et al., 2004; Weake and Workman, 2008; Weake et al., 2009). Similarly, deletion of Gcn5 in mice leads to development defects and embryonic lethality (Xu et al., 2000), while alteration of Gcn5 HAT activity survives slightly longer (Bu et al., 2007).

1.4.2 A subset of TAFs plays a structural role in SAGA

TAFs proteins have been the center of attention of many studies as they are important components of the transcription machinery. Indeed, as discussed above, TAFs are associated with TBP to form the coactivator TFIID and regulate the binding properties of TBP, as well as recognition of the promoter, by making specific contacts with promoter elements. However a novel role for TAFs in transcription was revealed with the discovery that a subset of TAFs is part of the SAGA complex. Taf5, 6, 9, 10 and 12 were identified as subunits of SAGA important for the different function of the complex. Indeed, depletion of Taf12 altered the transcription stimulation and HAT activity of SAGA, while TBP interaction function is not

affected (Grant et al., 1998a). Taf12 were also identified as a potential activator-binding site (see part 1.4.4 below).

Some Tafs proteins contain a HFD, the protein-protein interaction motif involved in the heterodimerization of core histones (Luger et al., 1997). A Taf octamer, composed of two heterodimers of Taf6/Taf9 and Taf4/Taf12 was shown to form *in vitro* (Selleck et al., 2001). SAGA does not contain Taf4, but rather Ada1, a subunit required for structural integrity (Horiuchi et al., 1997, p. 1), which contain a HFD and was shown to form a heterodimer with Taf12 (Gangloff et al., 2000). It was thus possible to reconstruct an octamer formed by two Taf6/Taf9 and Taf12/Ada1 heterodimers (Selleck et al., 2001). Similarly, a HFD was found within Spt7 and was shown to form an heterodimer with Taf10 (Gangloff et al., 2001). However, it was not shown whether such a Taf octamer is formed in the SAGA complexes. Indeed, immuno-labelling experiments have revealed the stoichiometry of some of the Taf proteins within SAGA (Wu et al., 2004). Taf6 seems to be present in two copies and was proposed to be located in the close vicinity of Taf9 in the structure owing to (i) their heterodimerization *in vitro* and (ii) the similarity of the Taf5 and Taf6 localization in both TFIID and SAGA, suggesting a conserved organization of the Tafs proteins in these two complexes. However the Taf10 subunit seems to be present in a single copy, similarly to its binding partner Spt7 which was also found in one copy (Figure 23). Nevertheless, the stoichiometry of the different subunits in SAGA is not clearly established and will required further investigation.

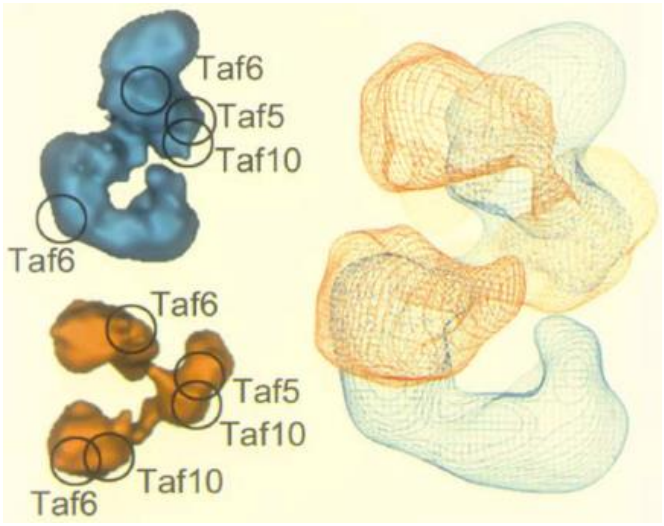


Figure 23 : Comparison of Tafs position in SAGA and TFIID
Localization of Tafs within SAGA (in blue) and TFIID (in orange) reveals similarity in the position of Taf5 and Taf6, but not Taf10. From (Wu et al., 2004)

1.4.3 Role of the Spt proteins in structural integrity and interaction with TBP

Spt proteins were originally discovered in a genetic screen to identify mutants capable of reversing the transcription defect of Ty insertion in the LYS2 and HIS4 genes (Winston et al., 1987, 1984). Ty elements are retrotransposons which, upon insertion, can disrupt or alter the normal expression of adjacent genes. Ty elements can be transcribed into an RNA molecule and this intermediate is required for their propagation. In particular, Ty insertion inhibits the expression of the HIS4 gene when inserted in the 5' region of this gene, and mutation in the Spt genes abolishes this inhibition in different ways. Interestingly, TBP was identified as a member of the Spt class of protein (Eisenmann et al., 1989). Histone H2A and H2B are also part of the Spt set of proteins (Clark-Adams et al., 1988)

The genes Spt3, Spt7, Spt8, Spt20 and Spt15 (corresponding to TBP) were shown to share the same phenotype in the transcriptional defect up insertion of the Ty elements (Roberts and Winston, 1996; Winston et al., 1987). Spt3 was later shown to genetically interact with TBP, and this interaction was required for transcription of certain genes (Eisenmann et al., 1992). Spt8 was also shown to functionally interact with Spt3 and TBP (Eisenmann et al., 1994). Spt7 and Spt20 were shown to be involved in transcription (Gansheroff et al., 1995; Roberts and Winston, 1996), and Spt20 was also required for response to a subset of activator proteins (Roberts and Winston, 1996). The link between the function of this four proteins was made with

the discovery of their integration in the same, large molecular complex, namely the SAGA complex (Grant et al., 1997).

A number of genetic and biochemical studies has evidenced the interactions between TBP and the Spt3 and Spt8 proteins (Eisenmann et al., 1994, 1992; Laprade et al., 2007; Mohibullah and Hahn, 2008; Warfield et al., 2004). Spt3 has been shown to be required for TBP binding at SAGA-dependent promoters, whereas Spt8 is required only at a subset of genes (Bhaumik and Green, 2002). A direct, physical interaction between Spt3 and TBP was evidenced more recently (Mohibullah and Hahn, 2008). Spt8 was also shown to stimulate transcription via direct interaction with TFIIA (Warfield et al., 2004). In opposition to their positive role in transcription initiation, Spt3 and Spt8 were shown to play a negative role by inhibiting transcription and TBP-binding at the promoter of certain genes in uninduced conditions, in particular in genes involved in response to amino acid starvation such as the HIS3 gene (Belotserkovskaya et al., 2000; Yu et al., 2003). Spt8 was shown to directly interact with TBP though a WD40 motif placed in its C-terminal part, and to compete for the binding of TBP to the DNA (Sermwittayawong and Tan, 2006). A model was proposed in which Spt8 binds to a TBP monomer on its concave face thus preventing its binding to the TATA-box. In a subsequent step upon activation, Spt8 would deliver TBP to the promoter.

Spt20 and Spt7 seems to play a structural role and are important for complex integrity, as deletion of one of these two subunits leads to dissociation of the 1.8 MDa complex and shows a severe growth defect (Grant et al., 1997; Sterner et al., 1999). Interestingly, upon deletion of Spt20, subcomplexes can still be purified (Lee et al., 2011; Wu and Winston, 2002), suggesting the possibility that submodules can form and exist in cell in this genetic context. However, it still needs to be addressed whether these submodules exist in normal cells and if they are functional. Spt7 is a large, acidic protein of which the multiple roles in transcription have been largely studied. Deletion of Spt7 shows the same phenotype as a double mutant lacking both Spt3 and Gcn5, consistent with its role in maintaining SAGA integrity (Sterner et al., 1999). Moreover, levels of Ada1 and Spt20 are dramatically reduced in spt7 mutant, suggesting that Spt7 regulates the amount of SAGA in the cell (Wu and Winston, 2002). Spt7 contains at its C-terminus a HFD which interacts with Taf10 *in vitro* and may form an histone-like pair in SAGA (Gangloff et al., 2001) Spt7 also contains a bromodomain (from 458 to 528), which is able to interact with acetylated lysine on histone tails but can be deleted without any particular phenotype (Gansheroff et al., 1995). The C-terminal part of Spt7 has been shown to interact with the Spt8 protein (Wu and Winston, 2002).

Interestingly, this C-terminal part can be cleaved by the Pep4 protease, at position 1141 or 1142 (Mischerikow et al., 2009; Spedale et al., 2010) and this processing results in the loss of the interaction region with Spt8. This shorter form of Spt7 has been shown to be incorporated into an alternative form of SAGA, named SAGA-like (SLIK) or SAGA altered Spt8 absent (SALSA) (Belotserkovskaya et al., 2000; Pray-Grant et al., 2002; Sterner et al., 2002). SLIK/SALSA (hereafter referred as SLIK) is similar in composition to SAGA, except for the cleavage of Spt7 and the loss of Spt8. The presence of an additional protein, Rtg2 has been documented (Pray-Grant et al., 2002). Rtg2 is important for the retrograde response pathway, which promotes expression of a subset of genes in response to mitochondrial dysfunction. SLIK has been proposed to be important for the transcription of genes involved in amino-acid starvation or growth on alternative carbon sources (e.g. HIS3, TRP3, GAL1), since the SLIK complex is more abundant than SAGA in these conditions (Belotserkovskaya et al., 2000). The loss of Spt8 in SLIK suggests that this complex possesses different TBP-binding properties compared to SAGA. However, the exact mechanism of TBP-interaction with SAGA and the role of SLIK in gene expression still need to be addressed. Interestingly, no homologue of Spt8 has been found in human so far, and the human homologue of Spt7 has lost the C-terminal, suggesting that human SAGA is more closely related to SLIK than to yeast SAGA.

1.4.4 SAGA is a coactivator required for transcriptional activation

Many SAGA components were originally identified as proteins required for the normal response to a set of transcriptional activators. Indeed, Ada proteins (which include Ada1, Ada2, Ada3, Ada4/Gcn5 and Ada5/Spt20) were identified as bridging factors between the chimeric GAL4-VP16 activator and the transcription machinery (Berger et al., 1992). Gcn5 was also required for response to the activator Gcn4 (Georgakopoulos and Thireos, 1992). However, the binding mode of these activators to SAGA remains elusive. The discovery of Tra1 as a component of SAGA provides a possible interface for these interactions (Grant et al., 1998b).

Tra1 is a very large protein of more than 400 kDa, and is the homologue of the human TRRAP protein, to which it shares 27% identity. TRRAP was identified as the target of two transcription activators, c-Myc and E2F, which stimulate transcription of several genes upon binding to their DNA response elements. Both have been involved in human cancer upon oncogenic transformation, and this process involves the recruitment of TRRAP (McMahon et al., 1998). In yeast, Tra1 is an important protein whose deletion is lethal and has been shown to

directly interact with several transcription activators, such as Gal4 (Bhaumik and Green, 2001; Bhaumik et al., 2004) and Gcn4 (Fishburn et al., 2005; Herbig et al., 2010; Qiu et al., 2005). However, it is possible that other potential binding surfaces exist. Indeed, a possible role for Taf12 in contacting Gal4 activator has been proposed (Fishburn et al., 2005). Binding of Gal4 to Tra1 has been shown to require the DNA element and seems highly selective, as mutants showing a defect in Gal4-Tra1 interaction are still able to be recruited by other activators (Lin et al., 2012).

Tra1 is a stable component of another HAT complex, namely the NuA4 complex (Allard et al., 1999). NuA4 is a coactivator composed of 13 subunits, which contains the major HAT Esa1, the only essential HAT in yeast. NuA4 preferentially acetylates lysines within histone H4 and histone H2B. NuA4 HAT activity is targeted to gene through its interaction with activators, and it acetylates nucleosomes at the promoters of TATA-less TFIID-regulated genes, such as ribosomal protein genes, but also at a subset of SAGA-dependent genes (Nourani et al., 2004; Reid et al., 2000). SAGA and NuA4 can be recruited to the gene promoter via their common subunit Tra1, or directly to the coding sequence during elongation (Ginsburg et al., 2009; Govind et al., 2007) where they promotes histone modifications and nucleosome evictions during transcription. The structure of the NuA4 complex has been solved by EM and revealed the global architecture of Nua4 and its mode of interaction with nucleosomes (Chittuluru et al., 2011). NuA4 is composed of two separate, globular domains, one of which contain the Tra1 subunit (Figure 24A), while the other domain present an interaction surface for nucleosomes (Figure 24C).

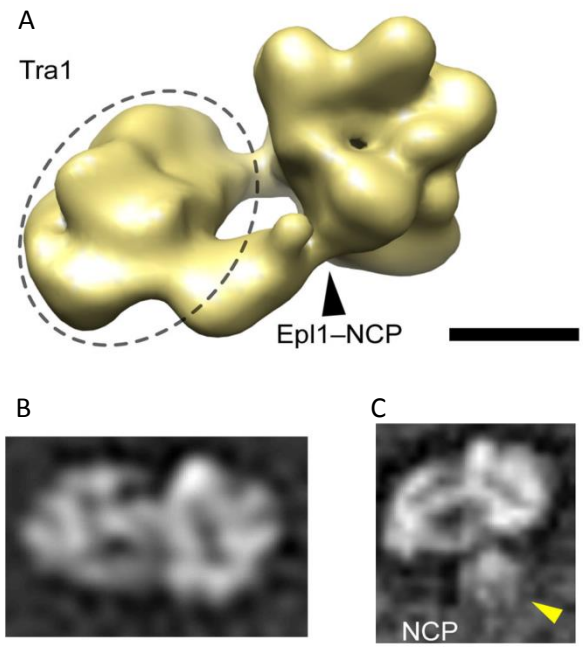


Figure 24 : Molecular architecture and nucleosome interaction site of NuA4.
(A) Three-dimensional model of NuA shows an organization of the complex in two globular domains, which one is composed by the Tra1 protein, while the other contains the HAT activity and the nucleosome interaction surface. (B) Classical view of NuA4. (C) Nucleosome Core Particule (NCP) binds to NuA4. From (Chittuluru et al., 2011)

1.4.5 The Histone Acetyltransferase Activity of Gcn5

The transcription inhibition resulting from the structure of the chromatin implies the existence of molecular mechanisms to displace the nucleosomes which compacts the DNA molecule and allow the transcription machinery to access the template DNA. Histones tails can be modified by a large variety of enzymatic reactions. Alterations of histone-DNA interactions weaken nucleosome structure and allow the binding and recruitment of other factors which displace or mobilize nucleosomes.

The acetylation of histone is the most well-known example of such modifications, and has been extensively studied. Histone acetylation was shown to correlate with transcription and gene activation (Brownell and Allis, 1996), and is generally associated with chromatin decondensation. Histone acetylation can change chromatin structure by two means: (i) neutralization of a positive charge of the lysine residue, thus weakening the possible interactions between the histone tail and the negatively charge phosphate backbone of the DNA molecule; (ii) the recruitment of transcription factor and chromatin remodeling complex, via specialized domain for acetyl-lysine recognition such as the bromodomain (Dhalluin et al., 1999).

The general control of amino acid biosynthesis 5 (Gcn5) was the first HAT to be identified in yeast (Brownell et al., 1996). Recombinant Gcn5 is able to acetylate both histone H3 at lysine 9 and 14, as well as histone H4 at lysine 8 and 16 *in vitro*, with a clear preference for H3. Whereas H3K14 is the preferential acetylation site *in vitro*, it seems that H3K9 is the favored site *in vivo* (Kuo et al., 1996). However, while recombinant Gcn5 is active on purified histone, it lacks the ability to modify histone in a nucleosomal context, indicating the need for other factor to drive Gcn5 activity on the chromatin. Two Gcn5-containing complex in yeast, namely the SAGA complex and Ada complex (Grant et al., 1997), containing also the Ada2 and Ada3 proteins, were identified and shown to be able to acetylate nucleosomal histone. This finding shows for the first time that the association of Gcn5 with others factors drives and expands its substrate specificity to more sites (H3K9, K14, K18 and K23) *in vitro* ((Grant et al., 1999). Also, the yeast Ada complex shares a same catalytic core with SAGA, and was shown to acetylate H2B and H4 in addition to H3, underlying the role of others components of the complex in substrate recognition.

Gcn5, Ada2 and Ada3 were shown to form a functional submodule in the SAGA complex by both genetic and biochemical experiments (Balasubramanian et al., 2002; Lee et al., 2011). Ada2 mediates the interaction between Gcn5 and Ada3 (Gamper et al., 2009). In addition, more recent work has identified Sgf29 as a novel component of this module, which is a stable component of SAGA and associated complex, and share the phenotype of the Ada protein family (Bian et al., 2011; Lee et al., 2011; S. Sanders et al., 2002). In addition to the role of Gcn5 in writing epigenetic marks on histones, the HAT module contains two domains involved in reading such post-translational modification. First, Gcn5 contains a bromodomain which can bind to acetylated lysines (Dhalluin et al., 1999). Second, Sgf29 contains a tandem Tudor domain, which can recognize tri-methylated lysine, and was shown to bind to tri-methylated lysine 4 on H3 (H3K4-Me3) (Bian et al., 2011). Such interactions with histones may contribute to the recruitment of the SAGA complex to chromatin, and/or stabilize the binding of SAGA to specific locations. In addition, phosphorylation of serine 10 on H3 by SNF10 seems to be important for efficient H3K9 acetylation (Cheung et al., 2000; Edmondson et al., 2002). A second domain involved in recognition of H3K4-Me3 was found in SAGA with the discovery of the Chd1 protein as part of the SAGA complex (Pray-Grant et al., 2005, p. -). Indeed, Chd1 possess two chromodomains which recognize this histone mark and seems to stimulate SAGA HAT activity. However, other proteomic approaches have failed to identify Chd1 in the SAGA complex (Gavin et al., 2002; Lee et al., 2011), and thus it is not clear to which extend Chd1 co-exists with SAGA and its function within SAGA is currently not understood.

As mentioned before, in higher eukaryotes the GCN5 gene has been duplicated and a paralogue of Gcn5 has been identified. The PCAF protein shows a high degree of homology with GCN5 (73% identity) and both incorporate into the SAGA complex. Both complexes seem to harbors distinct but overlapping functions. Indeed, it was shown that deletion of one of the two proteins was compensated in the cell by overexpression of the second (Kikuchi et al., 2005; Yamauchi et al., 2000). Moreover, while GCN5 is required during mouse development, knock-out of both proteins is lethal at an earlier stage in development, suggesting a non-redundant function of the two proteins (Xu et al., 2000). Similarly and as already mentioned the ADA2 gene has undergone duplication and gave rise to two paralogues, Ada2a and Ada2b. Despite their high homology, they integrate into separate and distinct complexes, ATAC and SAGA respectively, which have non-redundant roles (Guelman et al., 2006; Kusch et al., 2003). Association of GCN5 with ADA2a seems to change the specificity of acetylation toward histone H4 (Ciurciu et al., 2006). SAGA and ATAC complexes were shown to regulate distinct sets of genes (Krebs et al., 2011; Z Nagy et al., 2010)

In addition to its role at specific loci driven by activator-mediated recruitment, Gcn5 is implicated in the global acetylation of H3 at the genome level (Kuo et al., 2000). The global acetylation of H3 by Gcn5 seems to increase accessibility of the chromatin in general, thus facilitating transcription upon activation (Imoberdorf et al., 2006). Similarly, Gcn5 is targeted at the open reading frame (ORF) of certain SAGA-dependent genes during transcription, where it mediates acetylation of H3 (Govind et al., 2007). These suggest that Gcn5 not only plays a role at the promoter of genes, but also has a more global role in transcription elongation and chromatin regulation.

In agreement with its broader role, Gcn5 and PCAF were shown to acetylated non-histone targets. In particular, several transcription factors were identified as substrates of both acetyltransferase. (reviewed in (Bannister and Miska, 2000; Glozak et al., 2005)), suggesting an additional role for Gcn5/PCAF in transcription. Acetylation of these factors can either stimulate or inhibit their functions in transcription, by affecting nuclear localization, DNA binding or co-activator association. Most of the documented acetylation activity on non-histone targets seems to involve the PCAF protein rather than Gcn5. However, it is not clear from these studies whether these acetylations occur in the context of large complexes such as SAGA or ATAC.

Eventually, Gcn5 and PCAF function in other DNA-related events. Gcn5 was found to be involved in UV-induced DNA damage, by contributing to the maintenance of a more

accessible state of chromatin, in order to allow recruitment of the reparation machinery, and this requires the TRRAP subunit (Morrison and Shen, 2006; Robert et al., 2006). Gcn5 was also shown to play in role in telomere maintenance, though the destabilization of two telomere-associated proteins via USP22 activity (see part 1.4.6 below).

1.4.6 The deubiquitination module of the SAGA complex

A novel role in the transcription regulation by SAGA was revealed by the discovery of a second enzymatic activity within the complex. Ubp8 was first identified by large-scale proteomic approaches (Gavin et al., 2002), and its presence with other new SAGA components was validated soon after (Daniel et al., 2004; Henry et al., 2003; Powell et al., 2004; Rodríguez-Navarro et al., 2004; S. L. Sanders et al., 2002). In the context of SAGA, Ubp8 was associated to the removal of an ubiquitin moiety on histone H2B and was also shown to be present in the SLIK complex (Daniel et al., 2004; Henry et al., 2003). Ubiquitination of H2B on lysine 123 (H2B-Ub) is mediated by the Rad6-Bre1 ubiquitin ligase (Robzyk et al., 2000; Wood et al., 2003) and H2B-Ub is a transient signal which orchestrates a set of events during transcription (reviewed in (Weake and Workman, 2008)). Notably, Rad6 and Bre1 can be recruited to the promoter by transcriptional activators (Kao et al., 2004) and ubiquitilate H2B in the promoter. H2B-Ub is required for the addition of other epigenetic marks on the chromatin, in particular for di- (Me2) and tri-methylation (Me3) of H3K4 and H3K79 by COMPASS (reviewed in (Shilatifard, 2006)), two marks of actively transcribed chromatin. In a subsequent step, deubiquitination by the SAGA-associated Ubp8 will allow the recruitment of the Ctk1 kinase (yeast homologue of human cdk9) which will phosphorylate serine 2 on RNA pol II CTD (Wyce et al., 2007). Thus, deubiquitination (DUB) activity of Ubp8 provides another possible role for SAGA in transcription initiation and switch to elongation.

The idea of the existence of a distinct functional module within SAGA, involved in H2B deubiquitination and comprising Ubp8, came one year later. Indeed, the new components of SAGA Sgf11 and Sus1 were shown to be required for the SAGA-associated Ubp8 activity (Ingvarsdottir et al., 2005; Köhler et al., 2006; K. K. Lee et al., 2005). Ubp8 is inactive when isolated and requires the presence of Sgf11, but not of other SAGA component such as Gcn5 (K. K. Lee et al., 2005). In addition, Sus1 was found to be associated with Sgf11 even in the absence of Ubp8 (Köhler et al., 2006). The existence of the DUB module was also shown by genetic experiment, where Ubp8 and Sgf11 display a distinct pattern of genetic interactions compared to other SAGA subunits (Ingvarsdottir et al., 2005). The molecular link between the DUB module and the

SAGA complex was discovered with the role of Sgf73, which was shown to anchor the DUB module to the rest of the complex (Köhler et al., 2008; Lee et al., 2009). In particular, the N-terminal domain of Sgf73 (1-96) contains a zinc-finger containing (ZnF) domain which is required for DUB module assembly and substrate recognition, but not the catalytic activity (Köhler et al., 2008). Sgf73 was also shown to be involved in PIC formation by recruitment of SAGA to the UAS (Lee et al., 2009). A similar DUB module was shown to exist in higher eukaryotes. Human and *drosophila* orthologues of yeast proteins were identified (Kurshakova et al., 2007; Zhang et al., 2008; Zhao et al., 2008). Interestingly, human homologue of Ubp8 (USP22) was identified as part of a signature of genes associated with poor prognosis in certain cancers (Zhang et al., 2008), which is another evidence for the possible role of SAGA in the development of such diseases (Kurabe et al., 2007; Liu et al., 2003).

The structure of the yeast DUB module, comprising the full-length Ubp8, Sgf11, Sus1 and the N-terminal (amino acids 1 to 104) domain of Sgf73, was solved recently (Köhler et al., 2010; Samara et al., 2010) and illustrates the role of the four subunits in DUB activity. The DUB module shows a compact structure organized in two lobes (Figure 25). The assembly lobe is formed by a long N-terminal helix of Sgf11 which is wrapped by the assembly (ZnF UBP) domain of Ubp8 and this interaction is tightly fastened by Sus1. The catalytic lobe is formed by the catalytic domain of Ubp8, which contains the ubiquitin binding domain and the active site, and the ZnF domain of Sgf11 is found close to the catalytic site of Ubp8 in order to regulate its specificity. Eventually, the N-terminal domain of Sgf73 is inserted between the two lobes and regulates DUB module activity by tightening their interactions.

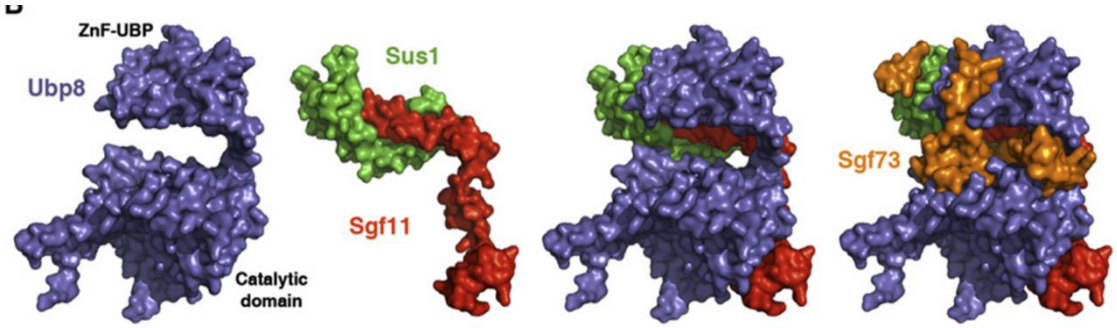


Figure 25 : Structure of the DUB module revealed by x-ray crystallography

Surface representation of DUB module structure. Ubp8 (in purple) is composed of two lobes. The assembly (ZnF-UBP) lobe makes extensive contacts with Sus1 and Sgf11, while the catalytic lobe is formed by Ubp8 catayltic domaion and Sgf11 ZnF domain, which potentiates Ubp8 activity. Sgf73 N-terminal domain is inserted between the two lobes. From (Köhler et al., 2010)

The Sgf73 protein contains two important domains: (i) as already mentioned, the N-terminal domain contains a conserved ZnF motif (C_2H_2) which is important for DUB module association and activity (Helmlinger et al., 2004; Köhler et al., 2008); (ii) a second atypical ZnF motif

which display a long insertion between the first two zinc-coordinating residues and was named SCA7 domain (Helmlinger et al., 2004) (Figure 26). This SCA7 domain is not found in Sgf11 but in its orthologue, the ATX7L3 protein. Albeit their similarity, SCA7 domains in ATX7 and ATX7L3 were shown to diverge in their structure, notably by the position of two α helices, which define their binding property. Indeed, the SCA7 of ATX7 (and Sgf73), but not of ATX7L3, was found to bind to nucleosomes *in vitro* (Bonnet et al., 2010). This is in agreement with the role of Sgf73 in recruiting SAGA to the UAS.

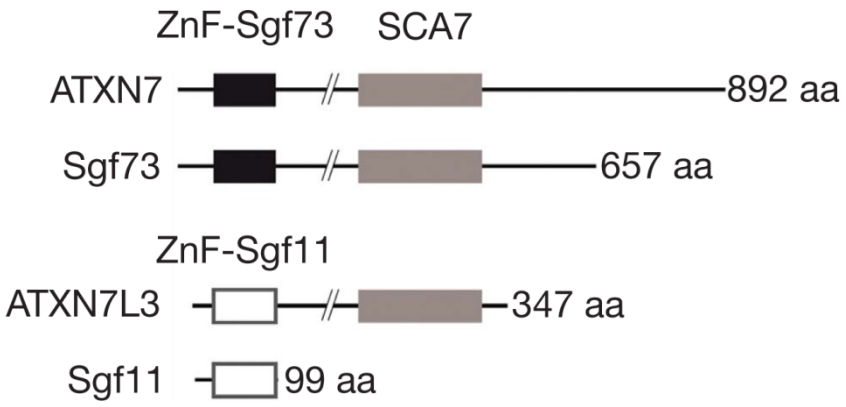


Figure 26 : Conservation of domain organization in Sgf73 and Sgf11
ZnF-contiaing domains of Sgf73/ATX7 and Sgf11/ATX7L3. SCA7 domains (in grey) are shared between Sgf73/ATX7 and ATX7L3 but not Sgf11. Sgf73/ATX7 N-terminal domain (in black) differs from Sgf11/ATX7L3 domain (in white). From (Bonnet et al., 2010)

A polyglutamine (polyQ) expansion at the N-terminus of ATX7, starting at position 30, was found to be responsible for the autosomal dominant neurodegenerative disorder spinocerebellar ataxia type 7 (SCA7, see (McCullough and Grant, 2010)) . PolyQ expansion ranges from 4 to 35 repeats in normal alleles, with a majority of alleles containing 10 glutamines, and pathogenesis have been observed when this stretch is extended from 35 up to 250 repeats (David et al., 1997). SCA7 disease is characterized by retinal and cerebellar degeneration and macular dystrophy, eventually leading to blindness. Both the normal and pathogenic forms of the ATX7 can incorporate into the SAGA complex (McMahon et al., 2005; Palhan et al., 2005), but the effect of this incorporation is still not clear. While it was reported that polyQ expansion was responsible for a loss of subunits in the SAGA complex (namely Ada2b, Spt3 and Taf12) as well as a decrease in HAT activity (McMahon et al., 2005; Palhan et al., 2005), another study reveals that in SCA7 mouse model, no effect on H3 acetylation and subunit content was observed whereas expression profile of a subset of genes was altered (Helmlinger et al., 2006; Kizilyaprak et al., 2011). More recently, the polyQ expansion was shown to bind the HAT core of Gcn5 when the ATX7 protein was express in yeast, and directly inhibit its function *in vitro* and *in vivo*, whereas SAGA is still recruited to the promoter of genes (Burke et al., 2013).

An interesting finding was the discovery that Sus1 is also part of the TREX2 complex, involved in mRNA export (Rodríguez-Navarro et al., 2004). Notably, Sus1 was found to be essential for mRNA export by targeting TREX2 to the nuclear pore complex (NPC) (reviewed in (García-Oliver et al., 2012; Pascual-García et al., 2008)). Other SAGA subunits were found to be important for mRNA export. Indeed, deletion of both Sus1 and Sgf11 triggers a stronger export defect than the single Sus1 deletion (Köhler et al., 2006). Similar results were obtained with Sgf73, where deletion leads to partial dissociation of Sus1 and TREX2 (Köhler et al., 2008), however it is not clear whether the effect of Sgf73 could be a consequence of the mislocalization of Sus1 upon Sgf73 deletion. Nevertheless, it seems that Sus1 possesses an additional role in transcription by coupling elongation and mRNA export and tethering the transcription machinery to the NPC in a process called gene gating. This mechanism seems evolutionary conserved, as homologues of TREX2 have been identified in *drosophila* and mammals (called AMEX in these organisms) and were found to play a similar role (Kurshakova et al., 2007).

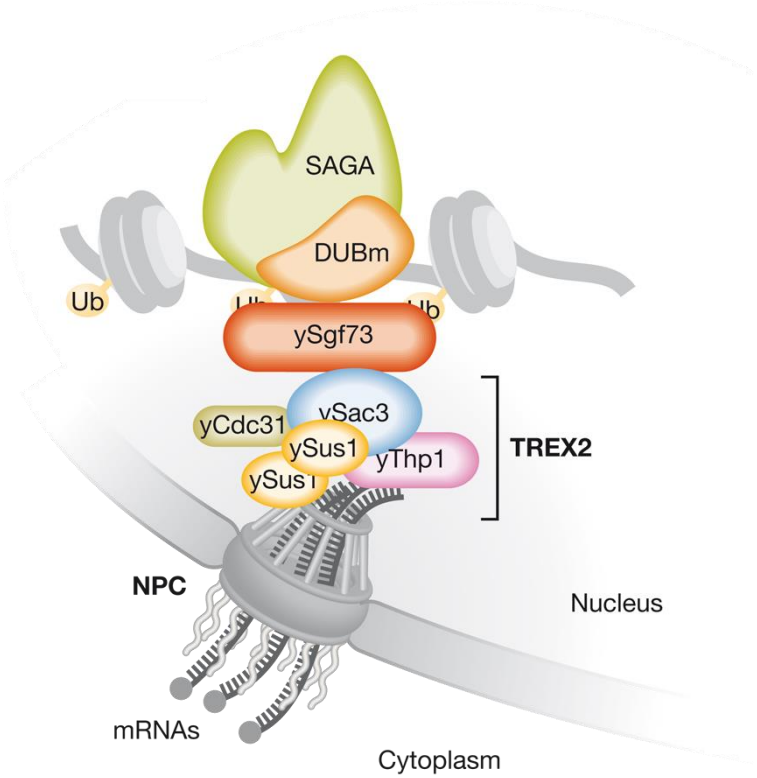


Figure 27 : Sus1 links transcription and mRNA export
Sus1 is a common component of SAGA and the TREX2 nuclear export machinery, and Sus1/Sgf73 link the transcription to mRNA export. From (Rodríguez-Navarro, 2009)

The 19S regulatory particle of the proteasome (RP) has been shown to play a role in the regulation of SAGA activity. The recruitment of SAGA to DNA-bound activators was shown to be enhanced in the presence of the 19S RP which is recruited to the promoter prior

to the SAGA complex (Ezhkova and Tansey, 2004; D. Lee et al., 2005). The ATPase component of 19S RP seems to be required for this recruitment and its inhibition leads to a decrease of H3 acetylation level. Recruitment of the proteasome is dependent on H2B-Ub and required for H3K4 tri-methylation (Ezhkova and Tansey, 2004). Recently, it has been proposed that the proteasome can interact with Sgf73 and dissociate a module containing Sgf73 and Ubp8, without affecting SAGA integrity (Lim et al., 2013). This module still harbors a functional DUB activity and this dissociation is important for normal association of the TREX2 complex to the NPC. Thus, a model was proposed where the proteasome helps to recruit the SAGA complex to the promoter of the transcribed gene, and upon entry to elongation dissociates a functionally active DUB module which follows the RNA pol II along the ORF. This dissociation allows the correct targeting of TREX2 to the NPC, and thus normal RNA export (Lim et al., 2013).

1.4.7 The Structure of the SAGA complex reveals its modular organization

A major breakthrough in understanding the molecular functions of the SAGA complex was achieved by the structure determination of the yeast SAGA complex by EM (Wu et al., 2004). SAGA was shown to be composed of five different domains forming an elongated structure, and presenting an important flexibility (Figure 28).

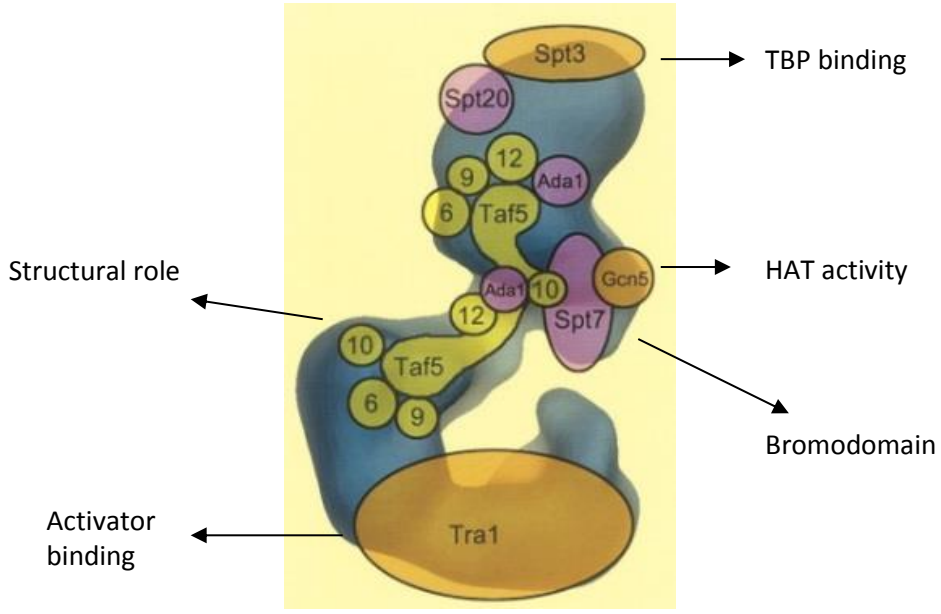


Figure 28 : Molecular organization of yeast SAGA

Architecture and molecular organization of the yeast SAGA complex revealed by EM and immuno-labelling. In the complex, subunits are regrouped according to their role in the SAGA functions. From (Wu et al., 2004)

Immuno-labelling experiments have helped to reveal the modular organization of the complex. Indeed, it appears that the different subunits of SAGA were regrouped in specialized modules according to their common roles in SAGA functions.

This finding is in good agreement with previously published genetic experiment (Sterner et al., 1999) and more recent proteomics data (Lee et al., 2011), and helps to understand the role and contribution of the different subunits to SAGA functions. Hence, the central domains contain the Tafs subunits, in close vicinity of Spt20, Spt7 and Ada1. These subunits are thought to play a structural role in the complex (Sterner et al., 1999; Wu and Winston, 2002), in agreement with their central localizations. The Gcn5 and Spt7 subunits both contains a bromodomain capable of interacting with acetylated lysine on histone tails and are found grouped in the same domain, forming a possible interface for nucleosome interaction. The Tra1 subunit, which has been identified as the main target of the activators known to interact with SAGA (Bhaumik et al., 2004; Brown et al., 2001; Qiu et al., 2005; Reeves and Hahn, 2005) is forming a separate domain, in the opposite part of the complex containing Spt3, which forms a possible interacting region with TBP (Eisenmann et al., 1992; Mohibullah and Hahn, 2008).

Similarly, the interaction map generated by the proteomic approach (Lee et al., 2011) reveals a comparable organization of the SAGA complex into five distinct functional modules, and establishes new interactions that were not highlighted by the structural study (Figure 29). Thus, two distinct modules were found responsible for the histone modifying activity present in SAGA: the histone deubiquitination (DUB) activity, carried by the Ubp8

subunit, and a HAT module containing in addition to Gcn5, Ada2 and Ad3 the Sgf29. This finding is in agreement with the role recently discovered for Sgf29 in reading histone modification (Bian et al., 2011) and with the genetic experiment in the same study showing that Sgf29 shares the Ada phenotypes. The existence of a DUB module within the SAGA complex was also confirmed (Köhler et al., 2008). The Tafs were found to be grouped in a distinct module, in good agreement with the immuno-labelling data. Interestingly, the Taf12 subunit is placed closer to Tra1, and this functional proximity was already suggested by biochemical study showing the possible interaction of Taf12 with activators (Reeves and Hahn, 2005).

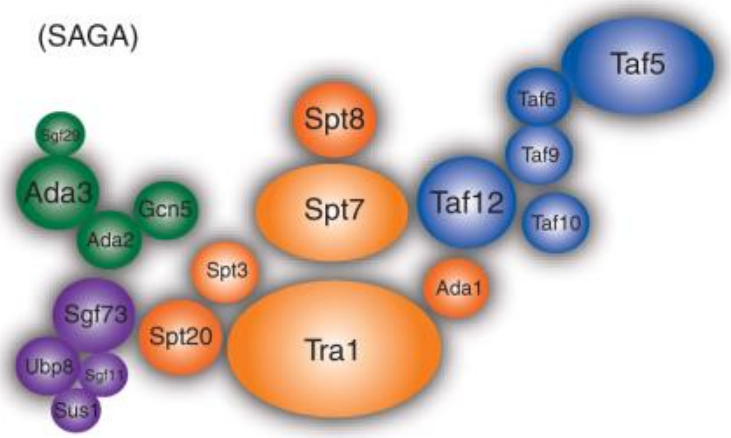


Figure 29 : Modular organization of the SAGA complex
Interaction map of the different subunits generated based on a combinatorial depletion analysis and mass spectrometry experiment. Proteins are clustered according to their probable proximity in the complex. From (Lee et al., 2011)

An unexpected finding of this study was the position of Tra1 relative to the Spt groups of protein. Hence, the Tra1 protein was found at the opposite part of the complex, compared to Spt20 and Spt3 in the structural study, while the proteomic approach suggest a strong physical link between these three proteins. The possible link between these proteins still need to be determined.

1.5 - AIMS OF THIS WORK

The initiation of transcription in eukaryotes is a tightly regulated process, which requires the coordinated action of several classes of protein. Sequence specific transcriptional activators will allow the integration of different signals in order to trigger the coordinated expression of a set of genes in response to an internal or external stimulus. The transcription machinery will need to overcome the physical barrier formed by nucleosomes, with the help of chromatin modifying proteins which need to be recruited in the process of transcription.

Finally, the RNA pol II will be assembled in a large pre-initiation complex, in order to identify the transcription start site and proceed to the synthesis of a mRNA molecule.

The SAGA complex is at the center of the initiation process by integrating several distinct functions required for transcription. SAGA (i) contains the Tra1 subunits in order to recognize and bind to specific DNA-bound activators, (ii) possesses a HAT activity allowing to write specific modifications on nucleosomal histones in order to open the chromatin template, (iii) can recruit TBP to the promoter of the transcribed gene and regulate TBP-binding in order to nucleate PIC formation and initiate transcription, (iv) allows the polymerase to enter the elongation phase by deubiquitinating the histone H2B which will trigger phosphorylation of serine 2 on RNA pol II CTD, and (v) facilitates mRNA export by coupling the transcription machinery to the export machinery. The importance of SAGA in transcription events is highlighted by the requirement of the different SAGA components in the development of higher eukaryotes, and its role in several diseases. Indeed, SAGA has been implicated in the mechanisms of oncogenic transformation of several human proteins and is a part of a signature of marker which correlates with development of metastasis in certain types of cancer. It is also involved in the neurodegenerative diseases Spinocerebellar Ataxia Type 7, characterized by cerebellar ataxia and retinal dystrophy which ultimately leads to blindness.

In order to better understand the molecular mechanisms of SAGA functions during transcription initiation, structural data on the complex are required. A first low resolution structure by EM was obtained in the group in 2004, and revealed for the first time the architecture and the molecular organization of the complex. Nevertheless, many questions are still unanswered, such as (i) how the activator conveys the signal to the transcription machinery through the SAGA complex; (ii) how SAGA recruits TBP to the promoter and what is the contribution of the different subunits in this recruitment; (iii) how do SAGA bind to the nucleosomes in order to write and read epigenetic marks and what are the binding surface involved in these contacts; and (iv) how the polyQ extension presents in the ATX7 protein of SCA7 patients affect the SAGA functions.

In this work, we have used EM in order to obtain more accurate structural data on the molecular architecture of the SAGA complex. In particular, we address the following questions:

What is the molecular organization of SAGA? To improve our knowledge on the structure of the SAGA complex, we determine the structure of the complex using EM. Prior to any structural study, the production of a sample of high quality is an important prerequisite.

We thus set up a protocol for the production of the endogenous SAGA complex from yeast and lots of efforts were put into optimizing the purification conditions in order to limit the dissociation of a rather fragile complex *in vitro*. We then analyzed a dataset in negative stain to better characterize the molecular shape and heterogeneity already observed in the previous study. We describe the large molecular motion of a particular domain of the complex which greatly limits the analysis of the structure of SAGA. Moreover, we initiate a study of SAGA in cryo conditions, in order to obtain structural data of a fully hydrated complex, close to its native state.

Where is located the DUB module within the SAGA complex? While the structure of the module has been solved by x-ray crystallography, the location and incorporation of the module in the context of the large SAGA complex is still not know. We solved the structure of two Sgf73 mutants, where the N-terminal part of Sgf73 was deleted (*Sgf73Δ1-104*) or the full Sgf73 protein was deleted (*Sgf73Δ*), leading to the loss of the DUB module. By comparing the structure of these two mutants with the wild-type strain, we localized the module in the vicinity of Gcn5 and Spt7, which defined a domain within SAGA specialized in the recognition and binding of nucleosomes.

How does SAGA interact with TBP? The Spt3 subunit has been localized in a domain adopting a highly variable conformation, while no information is available for Spt8. We localized the Spt8 subunit in a separate domain, thus defining two TBP-interacting regions. These two domains are highly mobile and undergo large movements due to rearrangement in the SAGA structure, placing the two TBP-interacting regions at the tip of a molecular clamp which may bind TBP from two different sides.

How does SAGA interact with the nucleosome? The localization of the DUB module defines a region which contains all the nucleosome-interacting subunits. In order to understand how this binding occurs, we tried to reconstitute a complex formed by SAGA and its nucleosome substrate.

CHAPTER II :

MATERIEL AND METHODS

Chapter 2. Material and Methods

2.1 - Structure of macromolecular complexes determined by Electron Microscopy

In cells of a living organism, all biological processes are carried out by macromolecules, such as proteins or nucleic acids. Theses macromolecules can act individually or by assembling into large entities and their coordinate actions are mandatory for the normal functioning of the cell during all its life cycle.

Proteins are important components of this cellular machinery. They are made of long, linear polymer of amino-acids assembled in a sequence unique for each protein (primary structure). To be active, this long assembly needs to adopt a specific structure: the polypeptide chain will first forms local structures (secondary structure), which will in turn interact in order to allow the protein to adopt a global structure (tertiary structure). Several proteins can then assemble into larger complexes (quaternary structure), where the interactions of the different subunits will drive the regulation of the protein activities.

The determination of protein structure has been an important tool to understand the role and function of isolated protein or macromolecular assemblies. Methods such as Nuclear Magnetic Resonance (NMR) or X-ray crystallography have made possible to understand at the molecular or atomic level, the function and regulation of many different proteins. However, these methods have limitations: while NMR is limited to the study of small objects (in the range of 30 kDa), X-ray crystallography requires first to obtain crystals of the protein. Obtaining suitable crystals can be a long and tedious work, which usually requires the purification of large amount of proteins or complexes.

To study rare, large macromolecular complexes, such as the SAGA complex, EM has been the method of choice. Indeed, EM allows the direct observation of large specimens in their native state, close to the physiological conditions, and requires a very limited amount of sample, in the picomole range. Recent advances in both the microscope and the detector, combined to improvement of computers and software used for the processing of the images, have made possible to push further the resolution of structures obtained to close-atomic level,

for virus or even large complexes of several MDa, such as the ribosome (Bai et al., 2013; Fernández et al., 2013)

2.1.1 Principle of imaging with a Transmission Electron Microscope

The basis of image formation in a microscope is the interaction of electrons with the object observed. High-energy electrons will be scattered by the interactions with the object, and the transmitted electrons will be focused by the electromagnetic lenses on a detector in order to form an image of the object. Electrons are scattered by matter, which includes molecules of air. Thus, the microscope column needs to be permanently kept at a high vacuum level. This requires the use of sophisticated airlock mechanisms to prevent air entry during specimen and/or detector insertion (such as photographic plate – see below). Moreover, the high vacuum condition is incompatible with preservation of the natural hydration state of a biological sample. This limitation can be overcome by imaging the object in cryo conditions: the sample is kept at very low temperature (typically of liquid nitrogen or helium), which limits evaporation of water molecules. Use of heavy salt solution for staining and subsequent drying of the sample limits the study of the object, as will be discussed later.

In a Transmission Electron Microscope (TEM), electrons are extracted from a source, and directed to the object through a complex set of magnetic lenses. Two types of electrons sources can be distinguish: (i) the thermionic cathodes use the emission of electrons by a heated metal filament, such as lanthanum hexaboride (LaB6) or tungsten filament; (ii) the field-emission gun (FEG) uses a powerful electric field to extract and direct the electrons from the cathode. FEG are characterized by a much higher brightness than thermionic sources, allowing the beam to reach a greater spatial coherence, required for high resolution imaging.

The lens system of a TEM is composed of several magnetic lenses which control the path and shape of the electron beam (Figure 30). First, a group of condenser lenses allows the users to control the size (and thus the intensity) of the electron beam. Then, an objective lens is responsible for image formation by focusing the electrons in the back focal plane. The objective aperture controls which scattered rays contribute to the image, and removal of the widely scattered electrons is responsible for the amplitude contrast. Phase contrast, which is the major source of contrast during cryo-EM imaging, is formed by the interferences between the unscattered electrons and elastically scattered electrons, which have undergone a phase shift due to defocusing and aberrations of the lenses. The phase contrast contribution to the

image is characterized by the Contrast Transfer Function (CTF), which is responsible for inversion of the contrast at specific spatial frequencies. The CTF is dependent on known experimental parameters such as the acceleration voltage of the electrons, the defocus value and the spherical lens aberrations (Cs), making possible the modelling and partial correction of CTF contribution to the image. Objective lens aberrations (spherical aberrations), chromatic aberrations and astigmatism) determine the resolution limit of the microscope. Eventually, projector lens is responsible for final image magnification and directing the image on a fluorescent screen or a detector.

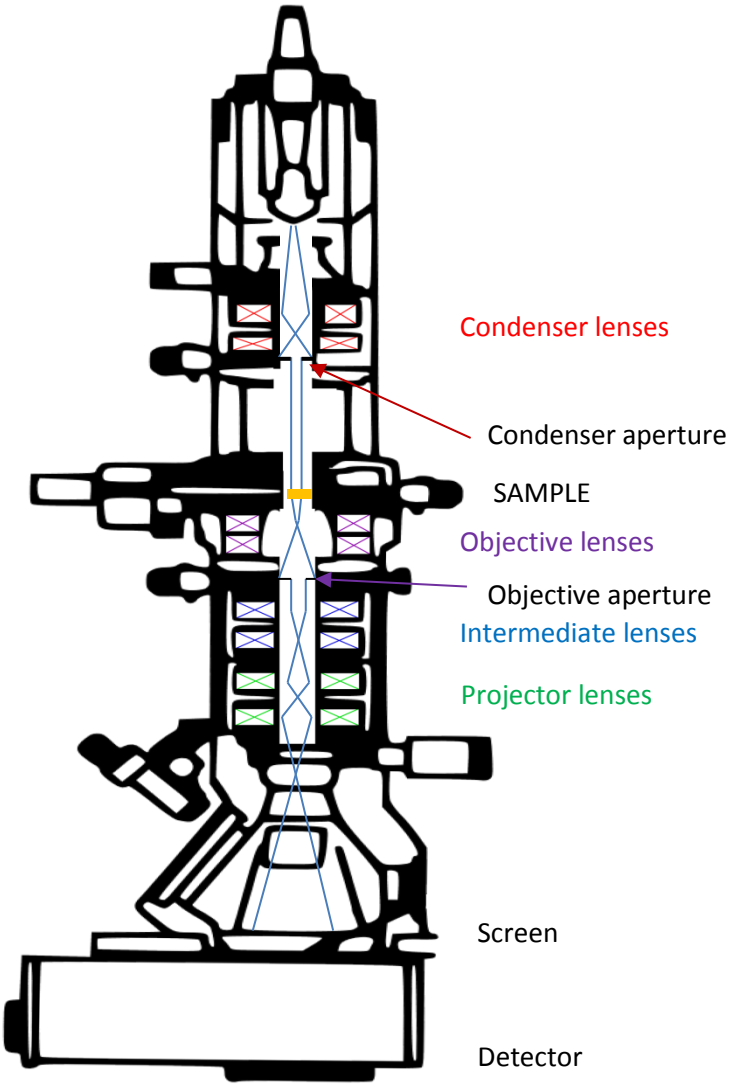


Figure 30 : Schematic representation of a Transmission Electron Microscope and path of the electrons thought the optical elements.

The image can be visualized by several means. The beam can be directed to a fluorescent screen, which will convert the electrons to photons, allowing direct observation of the sample. Alternatively, the image can be recorded by a detector. Photographic recording has been the method of choice for many years, as it allows a large area of the sample to be

recorded with high sensitivity. After exposure, photographic films need to be developed and digitalized by a photo-densitometer to obtain a digital image. This process is usually time-consuming and errors in handling of the film can lead to loss of images. Recent alternative to film recording has been the use of CCD cameras, in which a scintillator converts electrons to light, which will be detected by an array of photosensitive diodes. The main advantage of CCD detectors compared to photographic film is that CCD directly produce a digital image in real time, so no post-acquisition processing are required before image treatment. However the detection quantum efficiency drops significantly at high spatial frequencies thereby limiting resolution. More recently, development of direct detection devices for electron has been a real breakthrough for TEM imaging. These devices are able to directly detect electrons from the beam, with a high capture rate and high quantum efficiency, allowing the recording of a series of images (movies) of very short exposure times. Then, these images can be aligned and thus compensates stage movements or beam-induced image shifts during image acquisition, which greatly limited resolution so far.

In this work, two different microscopes were used: (i) the Tecnai F20 (FEI) was used for the image acquisition of negatively stained particles of SAGA from the different strains. Equipped with a FEG operating at up to 200 kV, this microscope allows the acquisition of images carrying high resolution information of the observed object. The good stability of the specimen holder at room temperature and a large CCD camera of 2048 pixels by 2048 pixels ("2k by 2k") allow the quick acquisition of a relatively large dataset when working on stained complexes; (ii) the Tecnai F30 polara (FEI) was used for image acquisition in cryo-conditions. It operates with a FEG at up to 300 kV, and data are collected on a direct electron detection CMOS camera (Falcon CMOS), with the potential of obtaining high resolution images of fully hydrated sample. The large (4k by 4k) CMOS detector, combined to the fully automated data acquisition scheme allows to collect very large dataset of several thousands of frames. The improve stage stability of the Polara compared to F20 makes it the favorite choice when working at liquid nitrogen temperature.

2.1.2 Specimen preparation

In single particle EM analysis, the biological object is composed of atoms of low electron density, which poorly scatter the electrons. Moreover, the object is surrounded by an aqueous buffer of similar electron density, and might be sustained by a carbon film made of an amorphous layer of carbon atoms where particles are adsorbed in order to be visualized. The

contrast between the object and its environment is thus very low and particles might be difficult to detect. In addition, the aim is to look at a biological object as close as possible to its natural state, thus the specimen preparation technique should allow the protection of the biological sample from the vacuum conditions inside the column of the microscope.

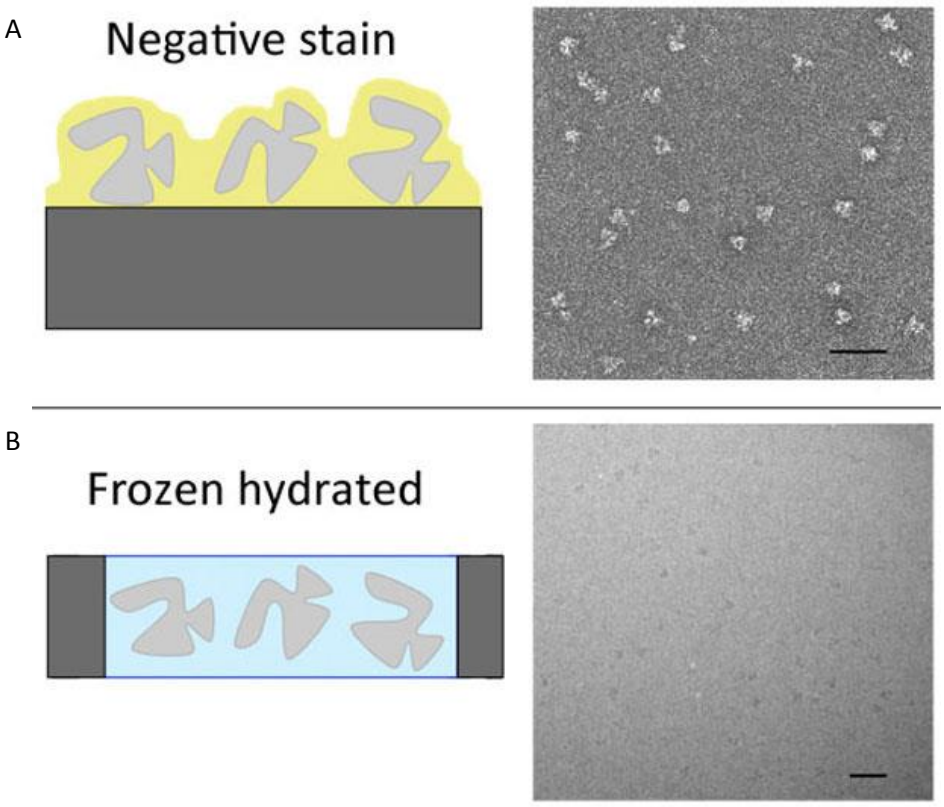


Figure 31 : Sample preparation techniques for single particle EM.
(A) : Particles embedded in a layer of heavy salt, on a carbon layer, allowing observation of particles with good contrast. (B): unsupported frozen-hydrated particles in a layer of ice allows preservation of the structure of the object. Black scale bar is 100 nm.

2.1.2.1 Negative staining

In negative staining, a solution of heavy salt, such as uranyl-acetate, is deposited on the grid after adsorption of the sample on a carbon film (Figure 31A, left panel). These salts have high atomic (Z) number, and thus strongly deflect electrons. The objective aperture will blocks the electrons deviated at high angles, giving rise to a strong amplitude contrast. The excess of staining is removed by a blotting paper, and the grid is dried in air for a few minutes before introduction in the column of the microscope. Negative staining is a fast, easy to perform sample preparation method, and allows observation of the sample with good contrast (Figure 31A, right panel) at room temperature. However, as the salt solution deposit around the particles adsorbed on the carbon film, one doesn't directly see the particle, but a print of the object in the surrounding stain layer. Consequently, negative stain observation is limited to surface feature of the object, and no information about internal details is revealed. Moreover,

only features in which the staining solution penetrates can be seen, limiting the resolution by excluding features smaller than ion size in the staining solution. Eventually, drying of the salt solution results in dehydration of the molecule and generates a hostile environment for macromolecular complexes, as ionic strength of the solution can be very high while pH is very low. These harsh conditions are responsible for alteration of protein structure (called flattening), and are another limiting factor for the resolution. In general, it is accepted that observation in negative stain is limited to a resolution of 20 Å.

Cryo-negative stain allows us to partially overcome these limitations. Indeed, after blotting of the excess of staining solution, the grid can be frozen in liquid nitrogen, thus preserving part of the hydration state of the molecules. This leads to less denaturation of the sample, and can improve resolution down to 15 Å.

Another drawback of the staining technique is the variability of the staining around each particle. In a perfect case, all particles should be entirely included in a homogenous layer of heavy salt. This is usually not the case, and the stain thickness can be variable around the particles. If the particle is bigger than the thickness of the stain layer, part of the object is not included in the image, giving rise to partial projections where information about the particle is missing. Combined with variations in the thickness of the stain layer, it gives rise to large variability in the appearance of a projection of the object. This can be avoided by the use of a double carbon layer, where sample and staining are trapped between the first layer of carbon on which particles are adsorbed, and a second layer added by floating after the staining. While it improves the homogeneity of the staining and thus the quality of the surface information, it also increases the flattening of the object.

2.1.2.2 Cryo-Electron Microscopy

Cryo-EM allows the observation of fully hydrated sample (Figure 31B, right panel), in conditions close to its native state. It allows recording of the full density of the sample, and has the potential to reveal high resolution features. The protein solution is quickly frozen (vitrified) in order to form amorphous (non-crystalline) ice, and the sample is observed in a layer of frozen buffer. As no staining solution is added, the low contrast is due to the weak difference in density between the object and its surrounding buffer (Figure 31B, right panel). Transfer contrast is then the main determinant in the observed image.

Another advantage of keeping the sample at low temperature is the enhanced resistance to radiation damage (discussed later in this chapter) induced by the strong electron beam. Inelastically scattered electrons transfer part of their energy to the atom during

interaction, which results in breakage of chemical bonds of the molecule. Free-radical created during this process are highly reactive and can propagate the damage to another neighbor molecule. In cryo conditions, the free-radicals are trapped and thus the propagation of the damage is limited.

In order to form amorphous, non-crystalline ice that would damage the biological sample, the grid needs to be quickly cooled down. To do this, the grid is plunged into liquid ethane at temperature close to those of liquid nitrogen. Ethane has a high heat capacity, allowing the fast heat transfer on the grid and thus fast freezing of the sample.

The major limitation in cryo-EM is inherent to the low contrast due to the weak difference between electron density of ice and of the sample. This makes the detection and post-processing of particle images a difficult task. Moreover, as signal is very weak (see section 2.1.4 on Data acquisition: Low-Dose Electron Microscopy), it requires averaging over a large number of particles and thus large datasets need to be handled in order to obtain high resolution information on the structure of the object.

Two sample preparation techniques can be used. In unsupported cryo-EM, the grid sustains a holey carbon layer. After applying the sample on the grid, the excess of solution is blotted with a paper in order to form a thin layer of buffer solution and the grid is frozen in liquid ethane. In the hole of the carbon film, a fine layer of frozen buffer containing the particles allows the user to observe the sample without the background due to the presence of the carbon film. Moreover, the particles cannot interact with the carbon film, which should in principle limits their positioning into preferential orientations, thus resulting in a wider range of angular distribution of the different views. In practice, the air-water interface partly defines orientation of the molecule in the layer of buffer. Therefore, some particular views can be overrepresented.

In supported cryo-EM, the grid sustains a continuous carbon film. Adsorption of particles on the carbon film allows the concentration of the sample in the fine layer of frozen buffer, resulting in an increased number of particles observed compared to unsupported cryo-EM. However, presence of the carbon film is an additional source of background for the image.

2.1.3 Protocols for preparation of carbon-coated grids for Electron Microscopy

In this work, grids for EM were prepared as follow:

2.1.3.1 Continuous carbon film

On a freshly cleaved mica (Electron Microscopy Sciences, 71851-05), a thin carbon film was evaporated in a vacuum coating system (Edwards Auto306). The day after, the microscopy grids (Electron Microscopy Sciences, M300-CR or Quantifoil – see below) were placed on a blotting paper at the bottom of a small beaker and cover with clean distilled water. Then, the carbon film was deposited by floating it from the mica on the water surface, and removing water with a water aspirator. Grids were allowed to dry in air for one night.

2.1.3.2 Holey carbon film

Grids with a holey carbon film were prepared by evaporating a carbon film on Quantifoil grids. Quantifoil grids with a holey plastic film (Quantifoil, N1-P16NCR30-01) were placed in the evaporator and a thick carbon layer is deposited. A thick carbon layer is used here to reduce charging effect due to accumulation of electrons at the surface of the carbon layer, which provoke a movement of the sample during acquisition. The day after, plastic was dissolved by adding several drop of ethyl acetate on the grids, and waiting one night for evaporation of the solvent. An additional thin continuous carbon film can be added to these grids according to the previous protocol.

2.1.3.3 Negative Staining

The previously carbon-coated grids were glow-discharged in the vacuum in order to make the carbon layer hydrophilic. Then, 5µL of sample were fixed (cross-linked) by adding glutaraldehyde (Sigma, G-5882) at a final concentration of 0.1%. Cross-linking was performed at room temperature (RT) for a few seconds, and the fixed sample was applied on the grids for 1 min at RT. Then, the grids surface was washed with 50 µL of a low-salt buffer (Tris-HCl pH 8.0 10 mM, NaCl 100 mM) and stained with several drops of a freshly filtered 2% uranyl acetate solution. After removal of the excess of staining solution by blotting with an absorbing paper, grids were dried in air and stored for later observation.

2.1.3.4 Cryo-electron microscopy

A quantifoil grid, covered with a continuous carbon film, was glowdischarged in the vacuum in order to make the carbon layer hydrophilic. Then, 5µL of sample were fixed (cross-linked) by adding glutaraldehyde (Sigma, G-5882) at a final concentration of 0.1%. Cross-linking was performed at RT for a few seconds, and the fixed sample was applied on the grid for 1 min at RT. Then, the grid surface was washed by placing it on a 50 µL drop of a low salt buffer (Tris-HCl pH 8.0 10 mM, NaCl 100 mM). The grid was removed with a pair of tweezers and a drop of 2 µL of Tween20 (0.05%) was added to the small drop of buffer which remained

on the grid. Blotting was performed with a Vitrobot (FEI) at 15°C, 95% humidity. Several grids were frozen at different blotting conditions, to screen grids for an ice thickness suitable for data collection. Blot force was set to 5, and blotting time was set in the range of 2 to 5s.

2.1.4 Data acquisition: Low-Dose Electron Microscopy

The resolving power of modern electron microscopes is sufficient to image single atoms. In material sciences, the specimen is very stable and a huge number of electrons can interact with the sample often without affecting its structure. As a result, individual atoms can be detected with a good statistical significance or signal to noise ratio (SNR). However, imaging of biological samples is hampered by the low tolerance of the biological material to irradiation. Structural damage induced by the electron beam is a strong limiting factor for biological specimen. Inelastic interactions of incident electrons with sample atoms dissipate energy that can break covalent bonds and generate highly reactive side chains. It is generally accepted that the atomic structure of the specimen is preserved when electron doses are kept below 5 electrons per square angstrom (e-/Å^2), however this number varies with the acceleration voltage of the electrons – at 300 kV it can be up to 25 e-/Å^2. In these conditions the molecular images are so noisy that the fine structural details cannot be detected. As a rule of thumb, at an electron dose of 5 e-/Å^2, details in the range of 50 Å can be detected with a SNR of two while smaller details are below this detection limit. To reconcile low specimen irradiation which leads to noisy images, with a high SNR objective to detect small details, it is necessary to split the dose required to detect atoms (say 2000 e-/Å^2) over several independent particles (in this case 400) to keep the dose below 5 e-/Å^2 and to add-up the signal coming from all these images.

The need for reduced electron irradiation also led to dedicated “low dose” data acquisition strategies in which microscope adjustments such as focus and astigmatism corrections are performed on an area remote of the area of interest to be recorded “blindly”. In this mode, the beam is deflected to an area far enough to prevent any damage on the recording area, but close enough to prevent any large variations of the image (as position of the sample along the z-axis). Settings are performed with beam intensity high enough for correct observations. Then, beam is shifted back and the area is exposed only once as it is recorded.

In this work, data collection of negatively stained complexes was performed on a Tecnai F20 electron microscope (FEI), operating with a FEG at 200 kV. Images were recorded at

a nominal magnification of 45000x, corresponding to a pixel size of 2.7Å per pixel on the detector (a 2kx2k CCD camera) and with a total dose of 20 e- per $Å^2$. Images were taken with a defocus value in the range of -0.5 to -1 µm. For Cryo-Electron Microscopy, data collection was performed on a Tecnai F30 Polara electron microscope (FEI) operating a FEG at 100 kV. Images were recorded on a 4kx4k CMOS Falcon detector (FEI), at a 59000x magnification, resulting in a pixel size of 1.78Å per pixel at the detector level. Exposure was done for 2s, for a total dose of about 15 electrons per $Å^2$, and with defocus values set between -2 and -3 µm.

2.1.5 Image processing

The objectives of single particle image analysis are dual. The first goal is to improve the SNR of the original images by averaging the signal from independent particles. As described above, image averaging will improve the SNR and increase the spatial resolution that can be detected. The second goal of image analysis is to reach a volumetric description of the sample. Standard transmission electron microscopes provide 2-D projections of the 3-D electron density map of the sample, multiplied by a microscope-specific CTF, which has to be corrected for. The objective is to determine the projection (or viewing) direction of each 2-D image with respect to the 3-D object it originates from and to reconstruct a 3-D model by combining many 2-D views. The different steps are summed-up in Figure 32.

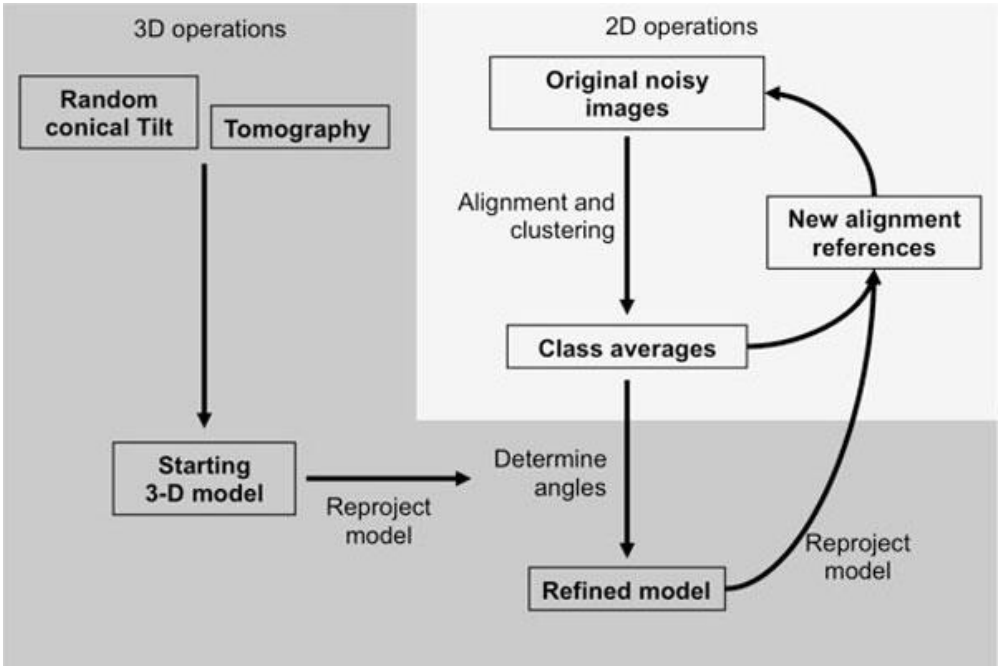


Figure 32 : Schematic representation of image analysis process.
Operations performed in 2D are in light grey, and in 3D in dark grey.

2.1.5.1 Particles selection

Before forming a 3D reconstruction of the object from the 2D views recorded, the particles need to be pointed and extracted from the micrographs. Particles will be stored in sub-images, which forms the image dataset. Interactive (manual) particle picking is a subjective, time-consuming and labor intensive process, which is not compatible with the analysis of large dataset of more than 100 000 particles that are required to achieve high resolution. In the way of going to automation of data acquisition and image processing, methods for efficient and accurate automated particle picking need to be developed.

Automated particle selection can be done in several ways, but usually starts with selection of “peaks” on the micrographs based on calculation of local variance between each pixel and his neighbors, or based on cross-correlation with a reference image (which is usually the rotational average of a reference view). Depending on the contrast of the image and the level of noise, false positives are likely to be detected. Many different algorithms have tried to remove these false positives by searching ways to describe particles based on diverse statistical measures. Nevertheless, the high level of noise hampers the efficiency of such approaches and to date, the most widespread method is the combination of automated particles picking with visual inspection of the result in order to remove the worst particles candidates.

2.1.5.2 Alignment and clustering

Averaging of images allows reducing the noise present in the image while improving the signal of common features. Images of the same specimen can be averaged to improve the SNR only if two criteria are met: first they have to correspond to the same view or projection of the particle and second the images have to be in the same register, or in other words aligned in translation and in rotation one with respect to the others. The spatial resolution that can be reached will depend on the number of images that can be averaged, on how similar the views are, and on the alignment quality. If a tolerance of 10° in viewing direction is accepted, the finest dimension that can be resolved for a globular particle with 10 nm in diameter cannot be smaller than 5xsin10° = 0.8 nm (Figure 33). To reach a resolution of 0.2 nm the variation in viewing direction cannot be larger than 2.2°.

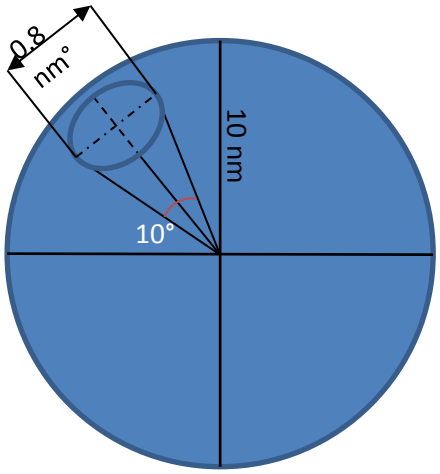


Figure 33 : Sampling rate of viewing direction limits the resolution

For a globular protein, the spatial resolution depends on the viewing angles which is assigned to each image. A smaller angular sampling allows a finest angles assigment and thus a higer spatial resolution.

A molecular image is aligned against a reference image by correlating image intensities. The correlation coefficient between two images is a measure of their similarities and all possible translations and rotations will be explored to find the correlation maximum, which will be considered as their best alignment. Images in a dataset will be aligned against a set of references (Multi References Alignment, MRA) and each image will be compare to all references, in order to find the best orientation of the particle. The quality of the alignment depends on many parameters such as the initial SNR of the image, the size and the shape of the particle.

In a real image data set, the particles have different orientations and produce different views which need to be separated before calculating an average image. The images

need therefore to be clustered into groups containing the most similar images. The image intensity variance should be minimized within the same group, while it should be maximized between different groups. In practical terms the image data set is first subjected to a multivariate statistical analysis (MSA) to detect the most meaningful trends in the dataset (Principal Component Analysis or Correspondence Analysis) and the clustering is then performed on the most significant Eigenvectors using Hierarchical Ascendant Classification schemes.

In an ideal image dataset, the particles are randomly oriented, which will produce an infinite number of projections. This condition is not always met when particles are adsorbed on a supporting carbon film, which may lead to preferred orientations. Nevertheless, the number of different orientations is very large and it is virtually impossible to find two perfectly identical particle images. It is therefore important to consider an angular projection sector within which we consider the images to be identical at a defined spatial resolution. An image class can to a first approximation be considered as a group of molecular images viewed along the same angular sector. If we consider the above mentioned 10 nm globular particles, a 10° tolerance in projection angle will result in an uncertainty of 0.8 nm. A projection sector of 10° leads to 244 different views and the dataset should be separated in as many classes.

The alignment and clustering steps are highly interdependent and will be used iteratively to improve the quality of the class averages. A better alignment will lead to an improved clustering which will impact the resolution of the class averages. Such high resolution class averages will further improve the alignment of the original images in a multi-reference alignment protocol.

2.1.5.3 Three-dimensional reconstruction

The class average images correspond to distinct views of the particle. They contain similar 2D projections of the object, according to a certain set of angles (the Euler angles). These angles describe, in a standard coordinate system the different rotations that the object has undergone in order to give rise to the corresponding projection (considering a fix observation point, Figure 34).

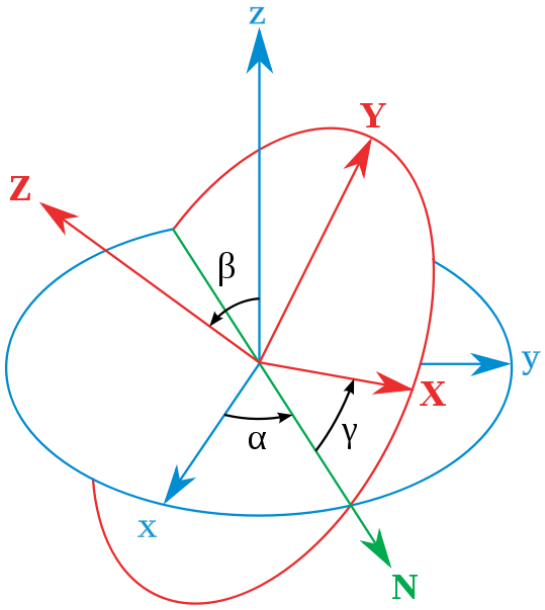


Figure 34 : Euler angles convention.
Rotations correspond to a rotation according to the Z axis by an angle α, then rotation according to the new X axis by an angle β, followed by a last rotation according to the new Z axis by an angle γ.

The projection direction of each view composing the dataset is not known a priori. A common-line based method was designed to attribute the relative projection directions of a set of class average images. This method is based on the following principle: two 2D projections at different angles of a 3D object intersect each other along a common line through their origin. For each image, all 1D projections of the image according to each possible angle are calculated, and stored in a sinogram. Comparing by cross-correlation the two sinograms allows one to find the common 1D projection of the two images, therefore giving the angular relationship between the two images. Unfortunately, this method works on set of images with very low noise, and may lead to ambiguous results especially when several conformations of the particle coexist.

Therefore, two experimental methods, based on the acquisition of tilted images of the same particle have been developed. In the tomography approach a goniometric EM stage is used to record tilted views of the same object, typically between +70° and -70° with angular increments of 1 or 2°. After alignment of the images on a common origin, a 3-D model can be calculated for each particle by combining all views for which the exact projection direction is experimentally determined by the position of the tilt axis and the tilt angle. This method suffers from several drawbacks that have been partially addressed. Electron dose and therefore radiolytic damages accumulate during the sequential acquisition of around 140 images of the same particle, but the development of very sensitive low noise cameras restricts the total accumulated dose to 20 to 40 e-/Å^2, enough to reach a resolution of 3-4 nm. The data

collection scheme produces a wedge of missing projections and this problem can be reduce by turning the grid 90° in plane and by recording a second tilt series. The missing wedge will then be reduced to a missing pyramid and the quality of the reconstructed volume is generally improved. Alternatively independently reconstructed single particle volumes can be aligned in 3-D and averaged. Since for each orientation of the particle the missing information is different, the average volume is essentially devoid of missing wedge artifacts.

In the random conical tilt method (Figure 35), the data collection strategy consists in recording first a 45-60° tilted view of an EM field containing several particles and, in a second exposure, an untilted view of the same field. The untilted images are analyzed as single particles thus producing classes containing several images of similarly oriented particles each differing by their in-plane or azimuthal angle. This angle, the position of the tilt axis and the tilt angle, informs about the viewing direction of each corresponding tilted image and allows calculating a 3-D model for each class of untilted images. With this data collection strategy, irradiation is limited to a single exposure (since only the tilted images, exposed first, are used for 3D reconstruction) and the missing information is restricted to a cone. Here again, independently reconstructed volumes can be aligned in 3-D and averaged, giving rise to an initial model without any missing information.

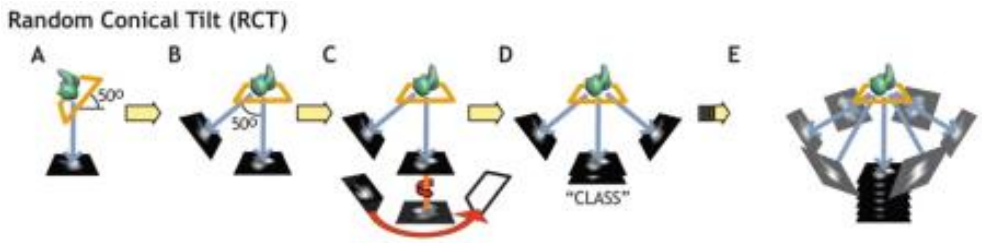


Figure 35 : Principle of Random Conical Tilt method.
After acquisition of tilted (A) and untiled (B) pair of images, images of the same view, but in different in-plane orientation (C) are regrouped into a same class (D). Corresponding tilted images from a single class correspond to projections at different known angles of the same object, and can be used for 3D reconstruction. From (Leschziner, 2010)

The experimental 3-D models are considered as low resolution “starting models” that will be used to determine the viewing direction of independently determined class averages obtained from a much larger image dataset. The starting models will be computationally “reprojected” along many directions to generate a set of reference images of known projection direction. The subsequent alignment of the class averages, or of the original images, against these reprojections in a process called reference-matching will determine the viewing direction for each high resolution class average and lead to an improved, or refined, 3-D model.

A common strategy for obtaining high resolution 3D structure of a macromolecular object is usually to combine information obtained using different techniques. As said before, negative stain allows the easy and rapid acquisition of images from an object, but high resolution data cannot be obtained. Therefore, it is used to "quickly" obtain an initial model as the high contrast facilitates alignment and angular assignment in the first steps of the reconstruction. Then, the model can be refined using high resolution, cryo-EM images.

2.1.5.4 How to address the dynamic properties of the complexes

The fast vitrification of the specimen in liquid ethane preserves the hydrated state of the protein complexes, but also cryo-immobilizes their different conformational states. This heterogeneity can hinder high-resolution structure refinement if different conformations are combined in a single class although it contains essential information about the dynamic properties of the sample. For isolated particles, different conformations can be sorted out computationally when the data set is large enough, thus providing information on mobile parts of the complex. For transient multi-component systems, the relative abundance of the components present in an equilibrium state informs about the interaction constants. It is therefore crucial to detect and separate the conformational states of the specimen both to improve the resolution of each individual state and to describe the dynamics of the examined protein complex. Several methods exist to detect and visualize de novo structural heterogeneities in the specimen. Rough movements of domain can be detected by either single-particle tomography or random conical-tilt experiments. More subtle differences can be tracked by using Eigen-analysis of resampled cryo-EM images. In this method the images dataset has to be aligned to an average reference structure to determine the relative particle orientation. A large number of volumes are built from a randomly created subset of the dataset and these volumes are subjected to MSA followed by hierarchal classification to identify the structural differences.

2.1.5.5 Determination of the Resolution

It is important to assess the quality and the accuracy of the model obtained in order to avoid misinterpretation of the data. Statistical assessments of a density map, such as calculating the resolution, helps to validate whether the features observed in the map have a real biological meaning.

In single-particle experiment, resolution is usually determined by the Fourier Shell Correlation (FSC) method. In this method, two maps of the same object are reconstructed from two independent datasets and are compared in Fourier space. For each spatial

frequency, a correlation value is calculated between the two maps, and this give rise to a decreasing curve. At low frequencies, correlation is quite high (close to 1), and will decrease at higher frequencies until a plateau is reached. This plateau represents the correlation of the noise background.

The resolution criterion can be set in two ways: (i) the FSC of the model can be compared to FSC of pure noise ("sigma" criteria), or (ii) the FSC of the model is compared to an empiric threshold value (usually 0.5 or .143).

The choice of the criteria used for calculation of the resolution is important, as it can give very different results for a same map. Up to now, no real standard has been set for the choice of the criteria, but the 0.5 criteria seem to be the preferred value in the field.

Description of the methods used for data acquisition and image analysis has been the topic of a tutorial, published in the *Journal of Nanobiotechnology* in 2013 (in annexe 2 page 183, Publication 2)

In this work, the images were treated as follow:

2.1.5.6 Electron microscopy on negatively stained particles

Micrographs were first manually controlled for drift and image quality, by looking at the power spectra of the image: images showing loss of information in one direction where removed from the set. Boxing of the particles was done either manually with Boxer (Ludtke et al., 1999) or in an automated way with gEMpicker (Hoang et al., 2013), using a strongly lowpass-filtered density map (80Å) of existing SAGA structure (Wu et al., 2004) as reference. Images were recorded with low defocus values, so CTF correction was not performed. Particles were extracted with batchboxer (Ludtke et al., 1999) with a boxsize of 192 pixels. Images were aligned in IMAGIC (van Heel et al., 1996): particles were subjected to MRA and MSA classification, and class average images were used as new references for a new cycle of MRA in an iterative manner. After 5 rounds of successive alignments, well-defined class average images were obtained and Euler angle assignment was done by projection matching, using the envelope obtained in our previous study as an initial reference.

2.1.5.7 Cryo-Electron Microscopy

Micrographs were first manually controlled for drift and image quality, by looking at the power spectra of the images: images showing loss of information in one direction where removed from the set. Images showing a high degree of ice contamination were also removed.

Picking was done with e2boxer program from the EMAN2 software package (Tang et al., 2007, p. 2). Particle images were extracted. CTF parameters were determined with CTFFIND3 (Mindell and Grigorieff, 2003). Images were processed in IMAGIC as previously described for negative stain dataset, except for phase flipping which was done in IMAGIC according to the parameter determined previously.

2.2 - PURIFICATION OF THE SAGA COMPLEX

Prior to any structural analysis of a macromolecular complex, it is necessary to produce a sample for which the quality is compatible with the method used for structure determination. In the case of the study by single-particle EM, it is necessary to obtain a pure sample with low heterogeneity, as it is a major limitation for the resolution of the final reconstruction.

The method of choice for production of proteins and protein complexes for structural biology is the use of recombinant expression systems such as expression in *E. coli* or the baculovirus expression systems. Indeed, recombinant protein production allows one to obtain large amount of protein, which can be purified in order to obtain a very homogenous sample suitable for structure determination by EM. These methods usually required the cloning of the cDNA for each protein of interest fused to an affinity tag for purification, into a compatible expression vector. This vector will be used for high level expression in the selected host. Many strategies have been developed in order to introduce several cDNA into a single or multiple expressions vectors, allowing the expression of the different components of a complex into the same cell for reconstitution of multi-subunit complexes. These approaches have allowed the reconstruction and purification of several complexes of more than 10 subunits for structural studies (reviewed in (Barford et al., 2013)).

2.2.1 Strategy for the purification of the SAGA complex: the TAP-tag method

In the case of the SAGA complex, it was not possible to implement these strategies. Indeed, SAGA is a large complex composed of about 19 subunits in yeast, making the production of a recombinant complex a difficult and challenging task. Thus, we decided to produce a sample by purifying the endogenous complex from the yeast *Saccharomyces*

cerevisiae. As SAGA is a very rare complex, the use of high affinity tags is required for efficient purification. The Tandem Affinity Purification (TAP) method (Puig et al., 2001; Rigaut et al., 1999) was originally developed for proteome analysis, where the purification of protein complexes for which the expression level was equal or close to the natural level was required. The TAP-tag is a combination of two IgG binding domain of the protein A from *Staphylococcus aureus* (protA) with a Calmodulin-binding peptide (CBP) (Figure 36A), which allows the efficient purification and recovery of fully active protein or complexes, from a complex mixture containing the target protein at a very low concentration.

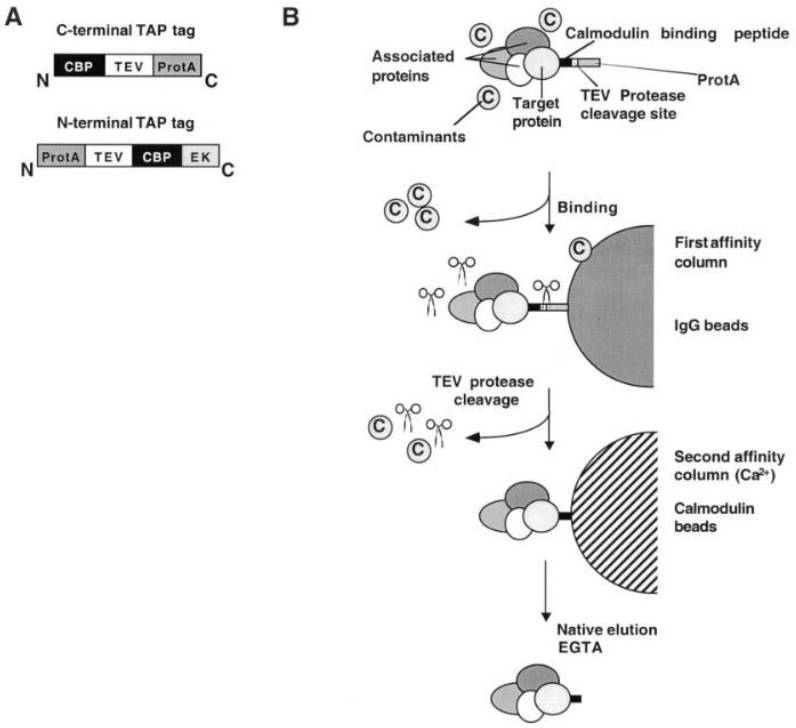


Figure 36 : Schematic representation of the Tandem Affinity Purification.
(A) Tags fused to the target protein. (B) Summary of the two steps for tandem affinity purification. From (Puig et al., 2001)

The purification can be performed in mild, nearly physiological conditions. The CBP tag allows the release of the complex in mild conditions, in the presence of EGTA. However, the protA tag can only be released from the IgG matrix at low pH, in denaturating conditions. Thus, to overcome this problem, a protease cleavage site (usually a Tobacco Etch Virus recognition site, TEV) is introduced between the two tags, allowing efficient recovery of the protein in any convenient buffer conditions.

The purification is performed in two steps (Figure 36B): first, protein extract is incubated on IgG matrix. After appropriate washes, the purified complex is recovered by addition of the protease. Then, in a second step, the eluted proteins are incubated on beads

covered with calmodulin, in the presence of calcium. After several washes, elution is performed by chelating the calcium with EGTA, thus releasing the sample from the resin. This second step allows the removal of the protease, as well as contaminants still present after the first step, which usually correspond to very abundant proteins found in the extract.

2.2.2 Generation of the yeast strains

The addition of a TAP-tag to a target protein is usually done by homologous recombination, taking advantage of the very high efficiency of this event in the yeast *Saccharomyces cerevisiae*. The classical method is the generation of a PCR (Polymerase Chain Reaction) product containing the sequence of the fusion and a selection marker, fused to sequences having homologies with the targeted genomic location at both ends (Puig et al., 1998). This PCR-product is then transformed in the yeast cell (Agatep et al., 1998), and the transformation product is plated on a selective media according to the marker introduced. This method allows the rapid insertion of an affinity tag (or any other fusion) to either the N-terminal or C-terminal part of the protein of interest, directly in the genomic sequence corresponding to this protein. When placed at the C-terminal end of the protein, the natural promoter of the gene can be used, thus allowing normal expression of the target.

Usually, homology sequences of about 40 to 50 bp are sufficient for the correct insertion of the PCR-product at the right genomic location. Selection markers are often auxotrophic markers, encoding an enzyme required for biosynthesis of an amino acid or another metabolite, thus allowing for selection of the recombinant strain on synthetic media lacking this metabolite. To be correctly incorporated, the strain used in the recombination event should not carry any sequence homologous of the selection marker, as it will result in very low efficiency of integration.

In this work, the strains used were kind gifts of our collaborators: the strain harboring an insertion of the TAP-tag at the C-terminal end of Spt20 was generated in the lab of Marc Timmers, from the University Medical Center Utrecht (Utrecht, The Neverlands); Sgf73 mutant strains were generated by Jacques Bonnet in the lab of Laszlo Tora, from the Institut de Génétique, Biologie Moléculaire et Cellulaire (IGBMC, Illkirch-Graffenstaden, France). The different strains used and their genotype is summarized in the following table:

Name	Genotype
BY4742	MATα ; his3Δ 1; leu2Δ 0; lys2Δ 0; ura3Δ 0
Spt20-TAP	Isogenic to BY4742 except Spt20-TAP::URA3
YGL066W (Euroscarf)	Isogenic to BY4742 except sgf73Δ::Kan
Sgf73Δ1-104	Isogenic to YGL066W except Spt20-TAP::URA3; pRS315-Sgf73Δ2-104-3HA-Myc
Sgf73Δ	Isogenic to YGL066W except Spt20-TAP::URA3; pRS315-ΔSgf73
Spt7-TAP (FY2031)	*MAT***a** ; HA-SPT7-TAP::TRP1 ; ura3Δ0 ; leu2Δ1 ; trp1Δ63 ; his4-917δ ; lys2-173R2

Figure 37: Summary of the different strains used in this study, and their correspondiong genotype.

2.2.3 Culture of yeast

In order to produce sufficient amount of complex, one should start the purification from enough cells. As SAGA is a very rare complex (Ghaemmaghami et al., 2003), production of an amount of complex in the range of the micrograms require the use of several liters of culture. Moreover, yeast cells are surrounded by a complex cell wall composed of several polysaccharides cross-linked in order to form an extracellular matrix. The role of this matrix is ultimately to protect the cell from external stress and to provide a convenient environment for the organism. In response to stress, such as the presence of ethanol released in the media during cell culture, cell will adapt its wall composition and thickness, making cell more resistant. In order to facilitate the lysis of cells during extract preparation, it is thus important to keep the cells in the most appropriate conditions for optimal growth. Thus, cells will be grown at 30°C, in rich media such as YPD. To avoid saturation in the liquid culture, which would contribute to slowing of cell culture, cell growth will be stopped at the early log phase (DO at 600nm of approximately 2). Moreover, the fast growing of cells on rich media will limit the amount of protease produce by the cells, which is beneficial for the subsequent purification steps.

2.2.3.1 Media composition

Liquid media

Table 1 : Culture media for yeast

YPD media (100 mL)	*Synthetic Media without Leucine (100 mL)*
Yeast Extract 1g (Becton, Dickinson, 212750) Bacto Peptone 2g (Becton, Dickinson, 211677) Add milliQ water up to 95 mL and autoclave. After autoclaving, add 5 mL of a sterile (filtered) 40% Glucose solution, to a final concentration of 2%	Yeast Nitrogen Base without amino acid and ammonium sulfate (Becton, Dickinson 233520) Ammonium Sulfate Drop-out synthetic media without Leucine Glucose 2g Add MilliQ water up to 100 mL and filter the solution. Keep at 4°C.

Agar plate

Table 2 : Composition of agar plate for yeast culture

YPD Agar (100 mL)	*Synthetic Solid Media without Leucine (100 mL)*
Yeast Extract 1g (Becton, Dickinson, 212750) Bacto Peptone 2g (Becton, Dickinson, 211677) Bacto Agar 2g (Becton, Dickinson, 214010) Add milliQ water up to 95 mL and autoclave. The solidified media can be melted in the microwave and let to cool down in a water bath at 50°C. Before pooring the plate, add 5 mL of a sterile (filtered) 40% Glucose solution,pre-warmed to 50°C, to a final concentration of 2%	Yeast Nitrogen Base without amino acid and ammonium sulfate (Becton, Dickinson 233520) Ammonium Sulfate Drop-out synthetic media without Leucine Glucose 2g Add MilliQ water up to 100 mL and filter the solution. Keep at 4°C.

2.2.3.2 Protocol for yeast culture in liquid media

Cultures are started from an isolated colony on a YPD agar plate or selective media according to the strain. The day before the culture, a single colony is used to seed a small preculture of 25 mL (YPD or selective media) and the flask is incubated in a shaker at 30°C o/n, with vigorous shaking. The day of the culture, the optical density (OD) at 600 nm.of the preculture is measured. Saturated cultures usually grow to a final OD 600nm of 10 in these conditions. From this preculture, a 1L liter culture is inoculated with an appropriate volume of preculture: inoculation with 200 µL of pre-culture at 5 pm usually allows one to harvest cells at 10h with a final OD 600 nm close to 2, considering a doubling time of 1h40. If this OD is not reached, the culture is allowed to grow longer during the day. When an OD 600 nm of approximately 2 is reached, cells are collected by centrifugation at 4000g for 20 min at 4°C. Supernatant is discarded and cells are washed once with 50 mL of cold water. Cells are

pelleted once more (4000g for 20 min) and the pellet is weighted prior to freezing in liquid nitrogen, and storage at -80°C. Typically, 2L of culture at a final OD of 2 should result in approximately 5g of pelleted cells.

2.2.4 Preparation of the yeast extract

In order to prevent dissociation of the complex during purification, particular care should be taken to keep the complex at the highest concentration during the whole process. The method used for preparation of the yeast extract should take into account two parameters: first, resistance of the yeast cells due to the presence of the cell wall, which can limit lysis efficiency; second, the total volume of extract should be as small as possible. Several methods have been used for lysis of yeast cell with success. French press applies a high pressure to the cell suspension, in order to force the cells to go through a tiny opening in the lysis chamber. Shearing forces generated during high speed flow through the narrow opening is responsible for high efficiency cell disruption, while keeping extract at low temperature. However, such systems are not convenient for lysis of large amount of cells, due to the limited volume of the lysis chamber.

Methods using an enzymatic digestion of the cell wall components, such as lyticase or zymolase, allow the user to soften the cell wall of the yeast, which can be opened by osmotic shock. Such methods require the use of large amount of expensive enzymes, which limits the total volume of cells that can be used.

Several methods use the principle of the grinder. Methods using a bead-beater or powerful shaker such as the FastPrep (MPBio) have been used up to now in many studies. Here, the cells in suspension are placed in presence of glass beads of a known diameter, and the shaking of beads causes collisions which can disrupt yeast cells with high efficiency.

In this work, different methods were tested in order to produce the best conditions for SAGA purification. The lysis of 10g of cells, resuspended in 10 mL of lysis buffer (see the section Buffers, below) and supplemented with protease inhibitors (Complete, Roche) was performed with either (i) in the French Press using two passes; (ii) the cell disruptor; the FastPrep (MPBio) by grinding the cell suspension with 30g of glass beads (splits in two 50 mL falcon tubes) for 3 times 1 minutes at a speed of 6 $m.s^{-1}$; (iv) vortexing with 15g of glass-beads in round-shaped polypropylene tubes (Nalgene) for 5 times 1 minutes. Efficiency was tested by following the bait protein by western blot during the purification.

2.2.4.1 Buffers

Table 3 : Composition of the buffers used during purification

Lysis buffer
Tris-HCl 50 mM pH 8.0
NaCl 300 mM
$MgCl_2$ 2 mM
Glycerol 10%
Tween20 0.1%

Calmodulin Binding Buffer (CBB)
Hepes 30 mM pH 8.0
NaCl 300 mM
Tween20 0.1%
$CaCl_2$ 2 mM

TEV Cleavage Buffer (TCB)
Hepes 30 mM pH 8.0
NaCl 300 mM
Tween20 0.1%

Calmodulin Washing Buffer (CWB)
Hepes 30 mM pH 8.0
NaCl 500 mM
$CaCl_2$ 2 mM

Calmodulin Elution Buffer (CEB)
Hepes 30 mM pH 8.0
NaCl 500 mM
EGTA 2 mM

2.2.4.2 Protocol for preparation of the yeast extract

Frozen cells are thawn on ice and resuspended with 1 mL of lysis buffer (plus protease inhibitor cocktail, *Complete EDTA-free* from Roche) per gram of cells used for the purification. A typical purification is made from 20g of cells. In the meantime, 4 round-bottom plastic centrifuge tubes (Nalgene, 029235) are filled with 15g of ice-cold acid-washed glass-beads (0.45-0.6 µm, Sigma 8772) and placed on ice to chill. This step can be done the day before to ensure that tubes and glass-beads are cold to avoid heating of the sample during lysis. To each tube, 10 mL of the cold cell suspension are added, and tubes are sealed with parafilm to avoid leaking during vortexing. Lysis is performed by vortexing each tube for 5 min in the cold room, by pulse of 1 min. Each tube is allowed to cool down on ice between each pulse. It is important to keep the cell extract at 4°C during all the purification process. After lysis, the tubes are centrifugated for 2 min at 8000g in order to precipitate the glass beads. The extract is transferred into a clean, pre-chilled centrifugation tube and centrifuged again at 48000g for 1h. The soluble extract is carefully collected to proceed to the purification.

Lysis efficiency can be controlled by measuring the OD at 260 nm. On a nanodrop, OD should be greater than 90. Alternatively, protein concentration in the extract by the Bradford method can be used.

2.2.5 Purification of the SAGA complex

The purification protocol was adapted from (Puig et al., 2001; Rigaut et al., 1999). The exact protocol used is given below:

In a precooled 15-mL falcon tube, 200 µL of IgG Sepharose beads (GE Healthcare, 17-0969-01) are equilibrated in lysis buffer by suspending the beads in 5 mL of buffer, centrifugating the beads suspension at 400g for 5 min, and discarding the supernatant. This washing is repeated once. Then, the yeast extract is incubated with the beads for 4 to 6h, under gentle agitation at 4°C, to allow binding of the bait protein to the IgG resin. Tween20 is used during this step to avoid aspecific binding of proteins from the extract to the beads, which would saturate protA binding sites and decrease efficiency of this step. After incubation, the beads are washed 5 times with 3 mL of TCB. At the last wash, OD at 260 nm of the supernatant is measured to check for detectable amount of contaminant proteins remaining in the buffer. If necessary, additional washes are performed. Then, beads are resuspended in 400 mL of TCB and transferred to a fresh, cold Eppendorf tube. Elution is performed o/n, by adding 3µg of TEV protease and incubating at 4°C under gentle agitation.

The day after, beads are centrifuged at 400g for 5 min in order to collect the supernatant. The beads are washed once with additional 400 µL of TCB to recover the cut proteins, and pooled with the previous elution. In parallel, 100µL of Calmodulin Affinity Resin (Agilent, 214303) are equilibrated with 1 mL of CBB in an Eppendorf tube (repeated once). Prior to incubation of the purified complexes with the calmodulin, 1.6 µL of $CaCl_2$ 1M (final concentration 2 mM) are added to the TEV elution products, which are then incubated on Calmodulin Affinity Resin for 2 hours at 4°C, under gentle agitation. After incubation, beads are centrifuged at 400g for 5 minutes and supernatant is discarded. Beads are washed 4 times with 1 mL of CWB. Elution is performed by resuspending the beads in 100 µL of CEB, and after an incubation of 5 minutes to allow elution of the proteins, sample is collected after centrifugation at 400g for 5 min. This elution is repeated 4 times.

Purification is controlled by polyacrylamide gel electrophoresis in presence of sodium dodecyl sulfate (PAGE-SDS). Ten microliters of each elution are loaded on a PAGE-SDS 4-15% gradient gel (Mini-Protean TGX Precast gel, Biorad, 456-1085) and the proteins are revealed by silver staining (Fisher, 24612, performed staining according to manufacturer protocol). Quality of the sample can be assessed by EM after adsorption on a grid and negative staining with uranyl acetate.

2.2.6 Reconstitution of the complex nucleosomes/SAGA

The nucleosome binding assay on SAGA was performed as follow: Taf10 antibody (kind gift of Dr. Laszlo Tora) was immobilized on 30 µL of protA magnetic beads (Dynabeads, GE Healthcare, 10001D). Beads were equilibrated with CEB and 10 µL (approximately 1 µg) of purified SAGA complex was incubated for 2 hours at 4°C with beads, under gentle agitation. Unbound complex was removed by three washes in a low salt buffer (Tris-HCl pH 8.0, NaCl 50 mM) or high salt buffer (Tris-HCl pH 8.0, NaCl 500 mM). Then, 0.5 µg of modified nucleosome (tri-methylated on lysine 4 of histone H3, kind gift of Dr. Marc Timmers) was incubated with immobilized SAGA in a total volume of 50 µL of the same buffer for 2 hours at 4°C. After 3 washes with the same buffer, proteins bound on the beads were denatured by boiling in presence of SDS and analyzed by western blot. SAGA was revealed with an anti-Taf6 antibody, and nucleosome was revealed with an anti-H3 antibody.

To reconstitute the complex formed between SAGA and the nucleosome, two methods were used: a first attempt by directly mixing purified SAGA with a 5-times molar excess of the nucleosome. After incubation for 30 minutes on ice, complexes were cross-linked with glutaraldehyde (0.1% final), adsorbed on a thin carbon film after glow-discharged and stained with uranyl acetate.

In a second attempt to enhance the efficiency of complex formation, complex was formed on Calmodulin Sepharose beads during purification. A 7.5 molar excess of recombinant nucleosome (formed from full-length recombinant *Xenopus Laevis* core histones expressed in *E. coli,* and assembled onto a 196 bp long 601 nucleosome positioning sequence) was added to the immobilized SAGA complex just prior to elution of the Calmodulin Sepharose. Then, complexes were washed once with the CEB to remove the excess of nucleosomes, and eluted according to the normal protocol.

2.3 - PROTEIN SAMPLE ANALYSIS

2.3.1 Electrophoresis

2.3.1.1 Principle of PAGE-SDS

Protein composition in the sample is checked by PAGE-SDS. This electrophoresis technique allows the separation of proteins according to their molecular weight. The principle is as follow: first proteins are denatured by a detergent (SDS), in presence of a reducing agent (usually β-mercaptoethanol) in order to break disulfide bond. Then, protein mixture is run on a polyacrylamide matrix, which is composed of a polymer of acrylamide chains, linked by bisacrylamide molecules. In this complex network, proteins migrate according to (i) their size: bigger molecules travel slower; (ii) their charge: strongly charged molecules migrate faster to their corresponding electrodes. In PAGE-SDS, the anionic detergent (SDS) denatures the protein, which travels as a linear molecule, and cover the amino-acid chain resulting in a constant negative charge distribution along the chain. Thus, proteins will migrate only according to their molecular weight.

The gel is composed of two zones: in the first part of the gel, pH is relatively low (usually 6.8). At this pH, the carrying ions present in the buffer (usually glycine) is predominantly present in its neutral state, thus with a low mobility. At the opposite, Chloride ions present in the gel run faster. Proteins in the sample will run faster than glycine, and thus will be trapped between the fast-migrating chloride ions front, and the low-speed glycine front, running as a thin, concentrated band. Then the migration front reaches the resolving gel (pH 8.8), glycine ions are predominately negatively charged and thus accelerates, passing the proteins, which can then run according to their size. Concentration of acrylamide in the resolving gel is then the major determinant of migration speed and resolution of the gel: at higher concentration, proteins of smaller molecular weight are better resolved.

2.3.1.2 Staining of PAGE-SDS gel

In order to be visualized, proteins need to be stained after migration in the gel. Two major staining solutions are used: coomassie staining and silver staining.

Coomassie staining

Coomassie staining uses the coomassie brilliant blue R250 dye to stain protein molecules. At slightly acidic pH, the dye-anion is electrostatically attracted to the positively charged amine groups of the protein, and van der Waals forces hold the reactants together (De

St. Groth et al., 1963). The protein-dye complex has a maximum absorbance between 525 and 575 nm, in the blue spectra. Coomassie staining is highly sensitive, allowing detection of approximately 100ng of proteins per band. Binding of the dye to the protein is fully reversible, thus coomassie staining is fully compatible with analysis downstream of gel staining, such as mass spectrometry (MS).

Silver staining

In silver staining technique, proteins bind silver ions from a solution, and ions are reduced in a second step to form silver metal, allowing direct visualization of the protein band. Silver staining is highly sensitive (up to 1 ng of protein per band) and many improvements in the protocols have helped to reduce cost and time of the experiment. However, staining is difficult to reverse and usually diminish the efficiency of downstream analysis (i.e. MS).

In this work, purified complex were run on 4-15% acrylamide gradient gel (Biorad, Mini-PROTEAN® TGX™ Precast Gel – ref. 456-1085). Proteins were visualized by silver staining (Pierce Silver Staining Kit – ref. 24612)

2.3.2 Western-blot

Staining of PAGE-SDS gel allows the sensitive detection of a protein or protein mixture in a sample. However, it cannot be used to either check the presence of low abundant protein in a complex mixture, or to unambiguously identify a protein from its migration profile.

Western-Blot allows the detection of a specific protein in a crude extract or a purified sample, even if the protein abundance is very low. Thus, it can be used to follow the protein of interest, or an interacting partner, at its natural expression level, during the whole purification process. However, this requires the availability of highly specific and sensitive antibody, which can be used for detection.

The principle of the western-blot is as follow: proteins separated by an electrophoresis method are transferred on a nitrocellulose membrane. Then, remaining binding sites of the membrane are blocked by soaking in a bovine serum albumin (BSA) or milk solution. An antibody specific to the protein (primary antibody) is then apply to the membrane and bind to the protein. After removal of the excess of antibody, a secondary antibody coupled to a reporter protein is bound to the primary antibody.

Reporter used is usually the horseradish peroxidase (HRP). This enzyme will catalyze the oxidation of luminol in presence of hydrogen peroxide. This reaction is coupled to production of light at 428 nm, which can be detected by a photographic film or a camera.

In this work, Western-blot was performed by transferring proteins from the PAGE-SDS gel on a 0.45µm nitrocellulose membrane (Sigma, Z613630-1EA), for 1h at 100V. Then, membrane is blocked in a 2% milk solution (in TBS) for 10 min under constant shaking. After washes, the primary antibody was incubated at appropriated dilution (in TBS) for 1h at RT. Membrane was then washed at least 3 times in TBS, and secondary antibody (Donkey anti-mouse IgG, 715-036-150 or Goat anti rabbit IgG 111-055-003), if required, was incubated at appropriated dilution for 1h at RT under constant shaking. Membrane was then washed 3 to 5 times with TBS to remove excess of antibody, and a chemiluminescent substrate (SuperSignal West Pico Chemiluminescent Substrate, Thermo Scientific ref. 34080) was added to reveal the presence of the secondary antibody. Signal from the HRP catalysis reaction was recorded on photographic film (Kodak Biomax MR, Sigma Z358479-50).

The bait protein was followed with the PAP (Peroxidase anti-peroxidase) antibody (Sigma, P1291), which bind to the protA tag. SAGA complex was revealed using anti-Taf5, Taf6 or Taf10 (kind gifts of Dr Lazslo Tora).

2.3.3 Mass Spectrometry: the multidimensional protein identification technology

Identification of proteins by mass spectrometry based-methods has become a standard for proteomic analysis. The most popular approach involve the enzymatic digestion by protease (such as trypsin) of the protein, either in gel or in solution, and the peptide mixture is then injected into the mass spectrometer. After ionization, peptides are first measured as intact ions and separated according to their mass to charge (M/Z) ratio, before being subjected to dissociation by collision in the spectrometer. Products of dissociation can then be analyzed in a second run of MS, in a process called tandem MS (MS/MS). The main advantage of MS/MS is the possibility to sequence peptide from a complex mixture, thus allowing direct identification of protein composition of a complex sample.

In case of a highly complex mixture, it is advised to first to limit the number of peptide analyzed at once be using a separation technique, such as gel electrophoresis. Bands from the gel can then be excised, treated by the protease for in-gel digestion before peptide

analysis. This kind of one-dimensional separation is suitable for moderately complex mixture of proteins. A multidimensional approach might be required for more complex sample. 2D gel analysis is possible, but is labor intensive as it requires the accurate excision of large number of protein spots.

Liquid chromatography has rapidly supplanted gel separation for MS analysis. After digestion, the peptide mixture is loaded on a silica nanocolumn of 50 to 100 µm, packed with a resin for separation. More recently, multidimensional protein identification technology (MudPIT) has made possible the analysis of large sample containing low-abundance proteins (Washburn et al., 2001). In this technique, multidimensional separation is achieved by coupling a strong cation exchange resin (SCX) upstream of a C_{18} reverse phase (RP) chromatographic media (Link et al., 1999). Peptides are loaded on the SCX and eluted by a slow salt gradient. At each step of the gradient, peptides released from the SCX resin will be retains by the RP column, and then differentially eluted by an acetonitrile gradient. The biphasic column is coupled to the spectrometer, allowing MS/MS analysis of the peptides at the end of the liquid chromatography (LC) separation.

Sample for Mudpit analysis were prepared as follow: proteins were first precipitated with TCA. Pellets were washed with cold acetone, then denaturated with urea, reduced and alkylated. Proteins digestion was performed with endoproteinase LysC (Roche) and trypsin. Resulting peptides were loaded on 100 µm fused silica microcapillary columns (Polymicron) packed with 5 µm C18 AQUA (Phenomenex) reverse phase, partisphere SCX (Whatman) cation exchange and reverse phase, and placed in-line with a nanoLC Ultimate 3000 (Thermo Scientific) coupled with LTQ Velos linear ion trap mass spectrometer equipped with a nano-LC electrospray ionization source (Thermo Fisher Scientific). Data were treated using SEQUEST against Swissprot database (S. cerevisiae -Strain ATCC 204508-S288c 2012-09-15). Peptides were filtered with FDR 1 %, peptides rank 1 and peptides of at least 7 amino acids in length.

CHAPTER III :

RESULTS

Chapter 3. RESULTS

3.1 - PURIFICATION OF THE ENDOGENOUS SAGA COMPLEX FROM S. CEREVISIAE

3.1.1 Methods of lysis for optimal conditions of purification

Different lysis conditions were tested in order to produce the highest yield and sample quality. The method of choice should allow easy and efficient lysis of large amount of cells, while working at low temperature. Here, we tested and compared efficiency of cell disruption with the (i) French Press or (ii) a Cell Disruptor, as well as grinding of cells with glass beads (iii) in a FastPrep system or (iv) manually with a vortex.

Efficiency of the lysis was assayed first by controlling the amount of the SAGA-specific Spt20 subunit released in the crude extract by western-blot and compared to the total amount present in the cells. To do so, a very specific antibody is required, in order to obtain a clear signal with very little background from a crude cellular extract. As the Spt20 subunit is the bait for purification, it carries a protA tag. The PAP antibody can be used to specifically reveal the tap-tagged protein, as it will be specifically recognized by the two IgG-binding domains of the protA tag and will not require a secondary antibody. The second parameter taken into account for testing lysis efficiency is the "quality" of the extract, that is to say the amount of the bait protein effectively bound by the resin after the lysis. The amount of the purified protein in the flow-through will be controlled by western-blot in order to address this efficiency (Figure 38)

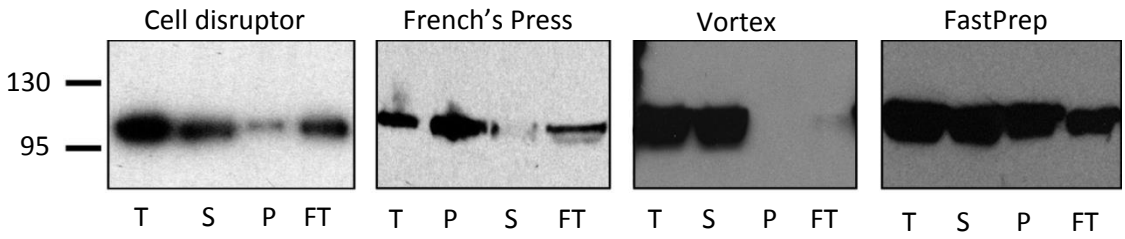


Figure 38 : Comparison of different lysis methods

Lysis efficiency of (A) the cell disruptor, (B) the French's Press, (C) grinding by manual vortexing with glass beads and (D) grinding with fastprep. T means total extract, S means Soluble Extract, P is the pellet (insoluble or unbroken fraction) and FT is unbound fraction on IgG sepharose. Protein revealed is Spt20.

Methods based on cell disruption, such as the French's Press and the cell disruptor, seems to have the worst efficiency of lysis, as very few complexes can be found in the soluble phase compared to the insoluble or unbroken fraction. Cell grinding with the fastprep allows a better lysis efficiency has more complexes are found in the soluble extract. However, the binding of the complex to the resin for both three methods has a very low efficiency, as most of the Spt20 protein released is found in the unbound fraction (flow-through) after incubation with the IgG sepharose resin. In stark contrast, cell grinding with the vortex is very efficient, as almost all Spt20 protein is extracted from the cell and stay in the soluble phase. The binding is also very efficient and no protein can be detected in the unbound fraction. These results indicates that a simple grinding with glass beads using a soft shaking such as produce by a vortex is the method of choice for cell lysis.

3.1.2 Optimization of purification conditions

Purification of the SAGA complex from the yeast extract was performed according to the standard TAP-tag protocol (Puig et al., 2001). Efficiency of TEV cleavage and elution from Calmodulin resin was controlled by western blot (Figure 39) and EM.

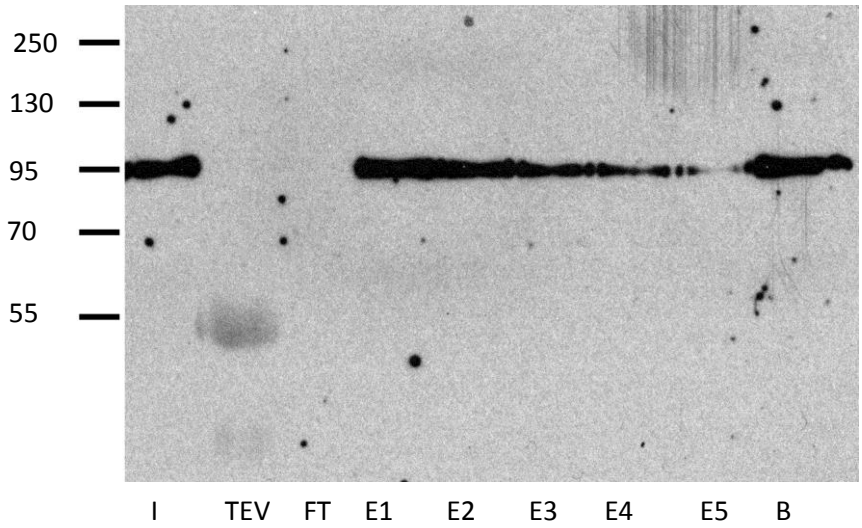


Figure 39 : Optimization of conditions for purification

Western-blot analysis of TEV elution and calmodulin elution. The protein revealed is Taf5 (90kDa). I (input) is SAGA-bound IgG sepharose, TEV is IgG sepharose after cleavage (uncut fraction), FT is unbound fraction of Calmodulin resin, E1 to E5 are elutions of Calmodulin resin and B is Calmodulin beads after elution (non-eluted)

TEV cleavage is performed at 16°C for 2 hours, with a large excess of TEV protease (3 µg). The cleavage reaction is complete and all complexes are eluted from the IgG sepharose beads. After addition of $CaCl_2$, binding of SAGA to the Calmodulin resin is complete. However, elution of this resin is not efficient in these conditions, as a large proportion of complex remains bound to the resin. This likely occurs though aspecific interactions of the complex with charged residues of the Calmodulin on the resin, which can be released by using higher salt concentration in the elution buffer. Thus, protocol was changed to include washing steps and elutions of the Calmodulin with 500 mM NaCl, allowing the full recovery of the complex.

The sample quality was then controlled by direct observation using EM of negatively stained particles. Complexes were cross-linked and adsorbed on a carbon film prior to observation. The images show the presence of elongated particles of approximately 27 nm, mixed with a large proportion of smaller, round-shaped particles pf about 15 to 20 nm. As it was very unlikely that these smaller particles could correspond to view of the complex along the long axis, we thought that they could correspond to product of partial dissociation or even degradation of the SAGA complex. We thus decided to improve purification conditions in order to obtain more homogenous sample. Notably, TEV digestion seems to be a critical point during the purification. Indeed, switching the temperature from 16°C to 4°C dramatically improves the quality of the sample, allowing the observation of a vast majority of elongated particles. Thus, SAGA complex seems very fragile *in vitro*, and care should be taken to handle and store the purified complexes in appropriate conditions (short term storage at 4°C exclusively, no freezing and thawing cycle, work on freshly purified complex).

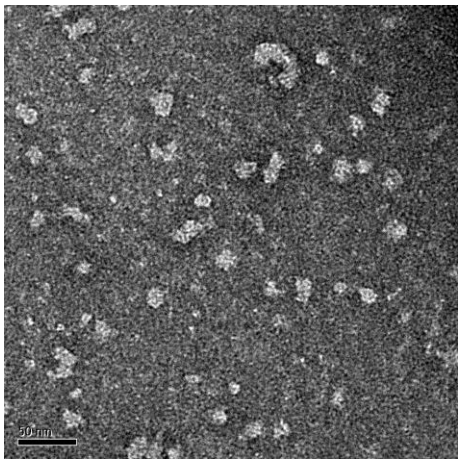
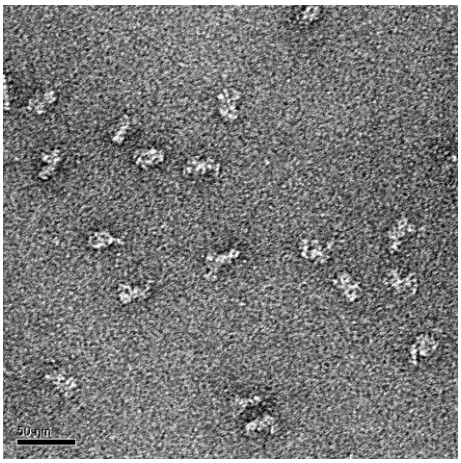
A  B 

Figure 40 : Control of the quality of the sample by Electron Microscopy.
Electron micrographs of negatively stained SAGA particles after TEV cleavage at 16°C for 2 hours (A) and 4°C overnight (B).

3.2 - CHARACTERIZATION OF THE COMPLEX BY PAGE-SDS AND MASS SPECTROMETRY

The composition of the samples purified from three different strains (wild-type, *Sgf73Δ1-104* and *Sgf73Δ*) was assessed by PAGE-SDS and multidimensional protein identification technology (MudPIT) (Swanson et al., 2009). The proteins purified were loaded on a 4-15% acrylamide gradient gel and revealed by silver staining (Figure 41).

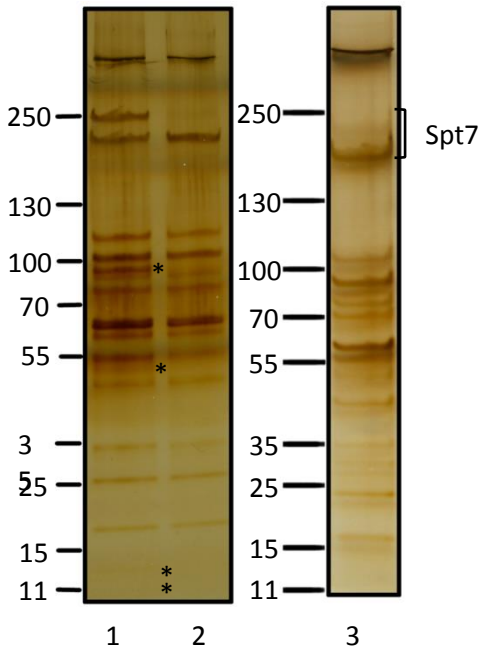


Figure 41 : PAGE-SDS analysis of SAGA complexes on a 4-15% acrylamide gel gradient.
Lane 1, 2 and 3 correspond to SAGA purified from the wild-type, *Sgf73Δ1-104* and *Sgf73Δ* strains, respectively. Proteins present only in the wild-type context are marked by an asterisk (*). Proteins are revealed by silver staining.

The gel shows a complex mixture of proteins, compatible with the presence of the multi-subunit SAGA complex as identified in different studies. Several bands, marked with an asterisk, seem to be exclusively present in the wild-type strain: a band below 100 kDa, a band just below 55 kDa and two weak bands around 11 kDa. These bands are compatible with the deletion of Sgf73 (migrates around 95 kDa, (Köhler et al., 2008)), Ubp8 (53 kDa) and Sus1 / Sgf11 (11 kDa), respectively. In the *Sgf73Δ1-104* deletion mutant strain, no additional band corresponding to the remaining part of Sgf73 could be seen, but it is possible that it co-migrates with another protein in the range of 70 kDa (expected size for the construction).

The Spt7 protein migrates around 250 kDa, and shows a profile composed of two bands: a first, higher band which correspond to the full length protein; a second, slower-migrating band which correspond to the shorter, C-terminally processed form incorporated in the SLIK complex (Köhler et al., 2008; Pray-Grant et al., 2002). These two forms are present in equivalent amounts in the wild-type complex. Interestingly, in both mutant strains, the ratio between the two forms is shifted toward the shorter form of the protein, while the longer form is almost completely absent.

To further control the composition of the purified SAGA complex, the samples were analyzed by Mudpit and proteins present were quantified by calculating their normalized spectral abundance factor (NSAF) (Zhang et al., 2010). The 19 subunits of the SAGA complex were found in the sample purified from the wild-type strain and detected with the highest

abundance. Chd1 was previously identified as a SAGA component required for normal HAT activity (Pray-Grant et al., 2005), but different studies have failed to identified it in SAGA preparation (Lee et al., 2011; Mischerikow et al., 2009). Similarly, this protein was not recognized as part of the SAGA complex in our preparation.

Analysis of the SAGA complexes purified from the two deletion mutants reveals the complete loss of the three proteins Ubp8, Sus1 and Sgf11, as observed in (Köhler et al., 2008). As expected, the Sgf73 subunit is still present in the *Sgf73Δ1-104*, whereas it is completely absent in the *Sgf73Δ* sample. Analysis of the peptide coverage of Sgf73 in the *Sgf73Δ1-104* strain confirms the deletion of the N-terminal part of this protein (Figure 42, line B compared to line A). As Spt7 was present in two forms on the gel, we looked at the peptide sequence coverage of this protein in all three strains.

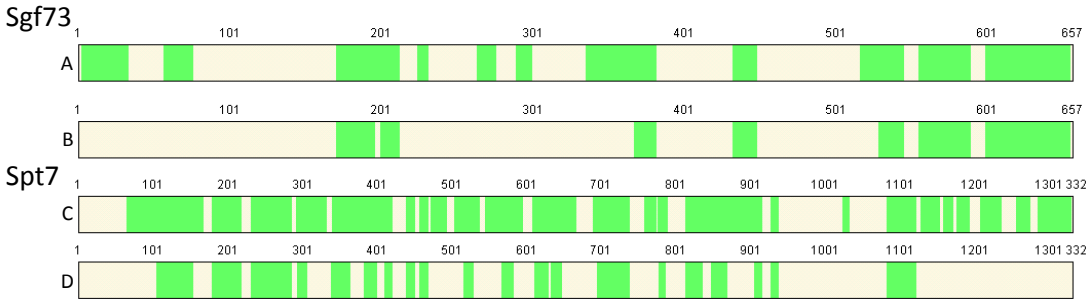


Figure 42 : Peptide coverage of Spt7 and Sgf73.

Part of the proteins for which peptides where detected are shown in green. Lines (A) and (B) represent the Sgf73 protein in the wild-type and in the Sgf73Δ1-104 mutant, respectively. Lines (C) and (D) represents the Spt7 protein in the wild-type and in the Sgf73Δ1-104 mutant, respectively. Similar results were obtained with the ΔSgf73 mutant.

While the complete Spt7 is detected in the wild-type strain, no amino-acid beyond 1110 can be found in our mutant strains (Figure 42, line D). This is in full agreement with the shorter form of the Spt7 protein found in the SLIK complex. Indeed, in SLIK, the protease Pep4 cleaves Spt7 at the position 1142 in the sequence (Spedale et al., 2010), and the C-terminal part of the proteins is then detached from the complex. As a consequence, the Spt8 is also detached and thus absent in SLIK. While Spt8 is found in our wild-type sample, Spt8 is barely detected in both mutant strains (Figure 43)

In order to assess the stoichiometry of the complex, the NSAF was calculated for each sample (Figure 43). The NSAF has been described as a quantitative tool to measure protein abundance in a given sample and to allow comparison of the protein levels (Paoletti et al., 2006). In the wild-type, all proteins have a NSAF value around 4, except for (i) Taf6, Taf9,Spt7 and Ada3, for which values around 6 are observed; (ii)Taf5 which has a value slightly higher than 4; (iii), Spt20 is the bait protein, and thus can be enriched during the purification,

and a higher abundance can be expected; (iv) Spt8 is a little less abundant, in agreement with the mix population of SAGA and SLIK which has lost Spt8 and Spt7 C-terminal part; and (v) Sus1 has a very high NSAF, of approximately 16, which can be explained by a bias in the detection inherent to its relative small size (11 kDa). Previous study suggested that Tafs proteins, as well as Ada1, are present in two copies (Wu et al., 2004). However, while Taf6 and Taf9 seems to be more abundant than the others proteins present in SAGA, their level is not fully compatible with the presence of two copies of each proteins in the complex, and others Tafs (namely Taf5, Taf10 and Taf12) and Ada1 are present at equivalent levels as the others components of the SAGA complex. In the two mutant strains, we observed a higher level of Taf6 and Taf9, and a slight decrease of Spt7 level which can be explained by the loss of the C-terminal part of this protein. This might indicate that the stoichiometry of the Taf6 and 9 subunits has changed in the complex purified from these two strains.

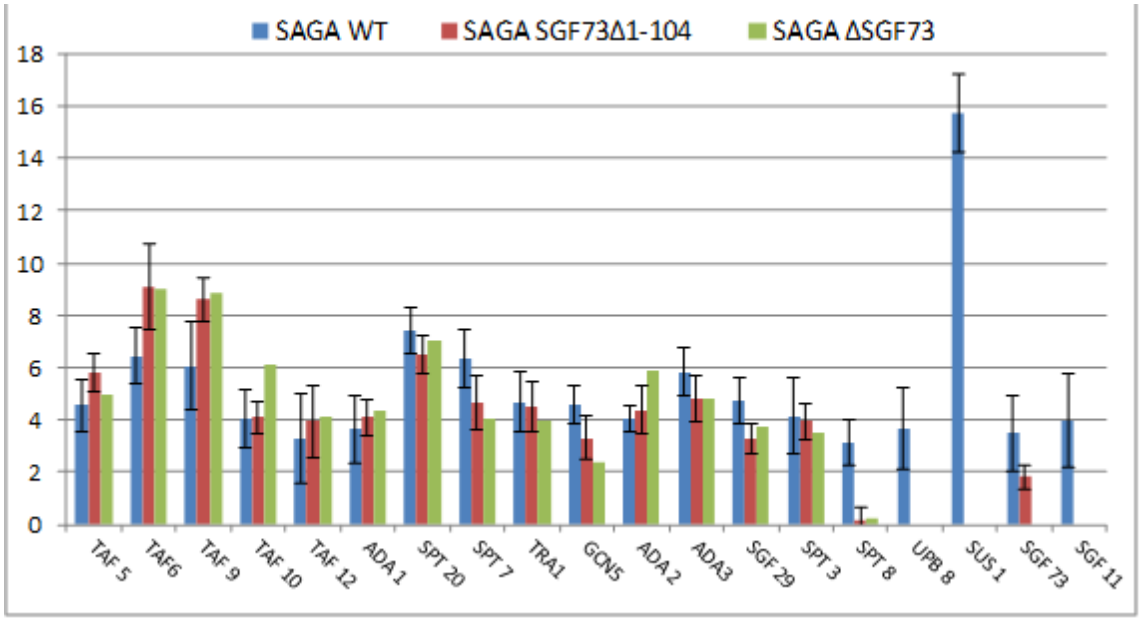


Figure 43 : Mass spectrometry analysis of the purified complexes.

NSAF are calculated (according to (Zhang et al., 2010)) for SAGA proteins identified by Mudpit in samples purified from wild-type (blue), *Sgf73Δ1-104* (red) and *Sgf73Δ* (green) mutant strains. NSAF corresponds to the ratio between spectral abundance factor (SAF) of the protein (total number of spectral count divided by the total length of the protein) and the sum of the SAF for all proteins in the sample.

In conclusion, we set up a protocol for purification of the endogenous SAGA complex in yeast. This protocol allows us to purify sufficient amount of complex for a structural study by EM. The sample contains a mixture of SAGA and SLIK complex in equivalent amount. Surprisingly, deletion of three (Ubp8, Sgf11 and Sus1) or four proteins of the DUB module shifts this ratio to the complete cleavage of the C-terminal part of Spt7, as found in SLIK, and subsequently loss of Spt8.

3.3 - STRUCTURE OF THE SAGA COMPLEX REVEALED BY ELECTRON MICROSCOPY

3.3.1 Description of the structure and the different conformations of SAGA

After staining of the molecules adsorbed on the carbon film, many particles can be observed by EM. First observations of the purified complex show a mixture of elongated particles of approximately 27 nm in length, and of smaller (between 15 to 20 nm), round-shaped particles. After improvement of purification and working conditions, we observed a vast majority of elongated particles. Nevertheless, SAGA complex seems very fragile *in vitro*, and dissociation of the complex can occur in our conditions.

A dataset of SAGA particles adsorbed on a carbon film and stained with uranyl acetate was recorded. A total of 19226 particles were subjected to iterative cycles of alignments and classifications in order to obtain characteristic views of the SAGA complex. The complex architecture is thus composed of five domains (annotated from I to V) organized into two lobes. Domain I and II compose the lobe A while domain III, IV and V compose the lobe B (Figure 44A). In most views, lobe A shows a better spatial resolution than lobe B, which appeared much less defined (Figure 44B). The complex seems to adopt a preferential orientation on the carbon film, as a particular view is overrepresented and composes roughly 50% of the dataset. On this view, lobe A and B are clearly visible and can be unambiguously identified, allowing the sorting of the particles according to the different position of the domain V in the lobe B. In approximately 20% of the images, domain V is not visible (Figure 44C). This observation is likely due to either the dissociation of the domain from the rest of the complex, or the visualization of a particular conformation where the domain is hidden in this particular view.

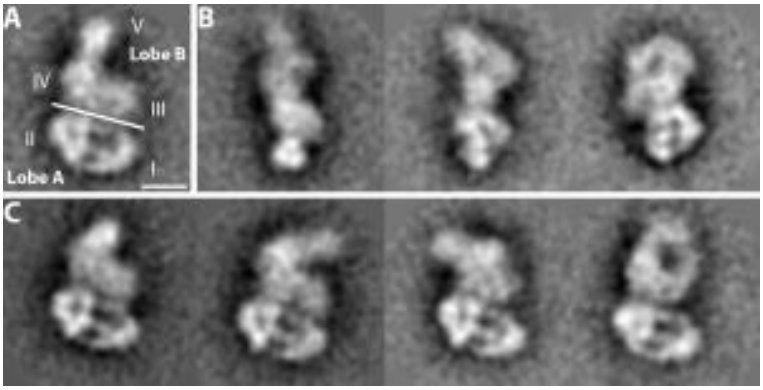


Figure 44 : Characteristic views of the SAGA complex

(A) SAGA is composed of two lobes: lobe A is composed of domain I and II, while lobe B is composed of three domains labelled from III to V. White scale bar represents 100Å. (B) Different views of the complex show a more detailed aspect of lobe A compared to lobe B. (C) The same characteristic view as in (A) reveal that lobe B adopts different conformations, according to position of domain V

A 3-dimensional reconstruction was performed (Figure 45) by using the previously obtained SAGA envelope as an initial model for angular assignment (Wu et al., 2004). New features can be described in the two lobes. In particular, domains I and II in the lobe A form two well-defined densities, stably linked by three long connections forming two accessible channels. Lobe B harbors two additional densities on the domain III and V, which adopt the conformation of a clamp where the tips are formed by these two domains. Immuno-labelling experiments have revealed the position of several subunits in the complex: domain I was shown to contain the Tra1 protein, while domain II contains the structural subunits Tafs and Ada1 (Wu et al., 2004). Domain IV was shown to contain structural subunits, as in domain II, while domain V contains the Spt3 protein. Domain III contains Gcn5 and the Spt7 subunits.

The major contact between the two lobes is formed at the interface of domain II and IV. The faint aspect of the connection might be the result of the flexibility of the two lobes in respect to each other.

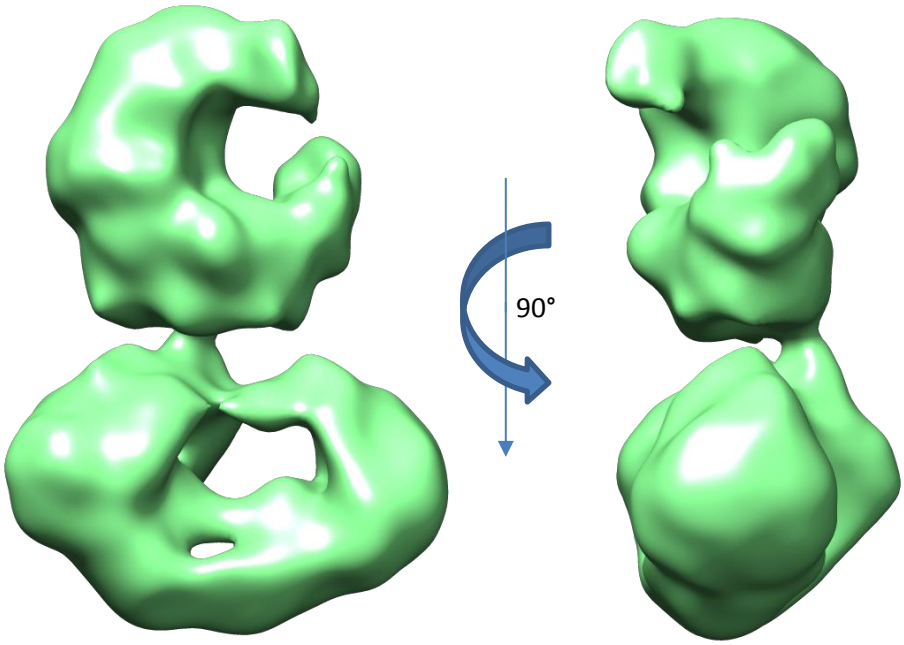


Figure 45: 3D reconstruction of the SAGA complex

Three dimensional reconstruction of the SAGA complex after assignment of euler angles to the different views using the existing density map as a reference.

The analysis of the different views obtained during image analysis suggested the presence of different conformations of the complex in the dataset (Figure 44C). In order to analyze the structural heterogeneity and separate the different possible conformations adopted by SAGA, we performed a 3D classification by maximum likelihood to sort the images (Scheres, 2012). Very few variations were observed in the lobe A, in agreement with the well-defined views obtained after 2D classification of the aligned images. However, large variations appear in lobe B (Figure 46), which may account for loss of a subset of subunits or spatial rearrangement of the different domains. The most striking variation occurred at the level of the clamp and defines two extreme conformations: (i) a closed conformation (Figure 46A), in which domain III and V contact each other and form a closed protein ring; (ii) at the opposite, a fully open molecular clamp (Figure 46D), where these two domains are separated by 7 nm. Intermediate conformations are also visible as distinct forms of the complex (Figure 46B and C).

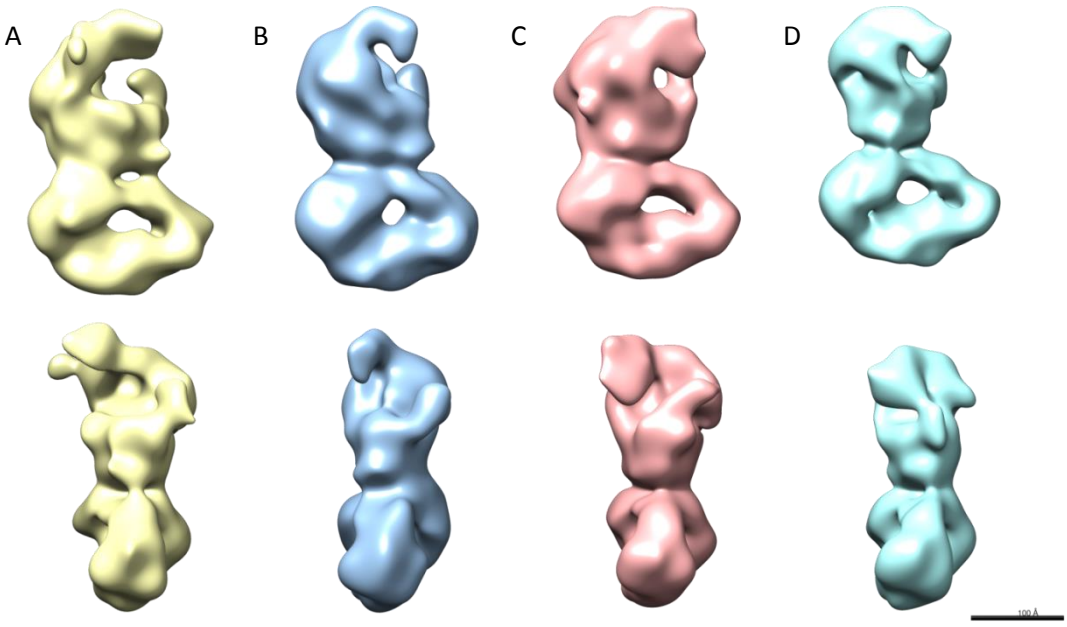


Figure 46: Different conformations obtained by maximum likelihood 3D classification.
(A) Open conformation of the molecular clamp formed by lobe B. (B) and (C) show intermediate forms, and (D) shows a fully close conformation. Scale bar is 10 nm.

As already mentioned, the faint connection between the two lobes might be the result of a flexibility existing between lobe A and B. Thus, alignment of the different views might be driven by the position of one of the lobe, placing the other lobe in a different register for classification of the images. This would lead to averaging of imperfectly aligned images, and thus loss of details in some part of the image. In order to improve alignment for both lobes, we decided to extract them from the aligned SAGA images and to process them independently. Such an approach has already been used in a previous study (Sauerwald et al., 2013) and is conceivable since the preferred orientations of the SAGA complex produce characteristic views where the two lobes can be unambiguously identified and show little overlap.

The 3D reconstruction of the lobe A (Figure 47A) shows very little differences with the structure of the lobe when analyzed in the entire complex, probably due to its strong features which served as a preferred “anchor point” for image alignment of the entire particles. In stark contrast, numerous features become more visible in the lobe B (Figure 47B). In particular, three main differences can be observed: (i) the domain V grows in size while its connection with domain IV is reduced; (ii) the extension of the domain III is more pronounced and (iii) a second, new density protrude from the bottom of domain III and can potentially forms a connection with the domain I in lobe A. These changes between the density maps of the complete SAGA and of the separately analyzed lobe B are consistent with a moderate rotational movement of lobe B relatively to lobe A around the domain II/IV connection.

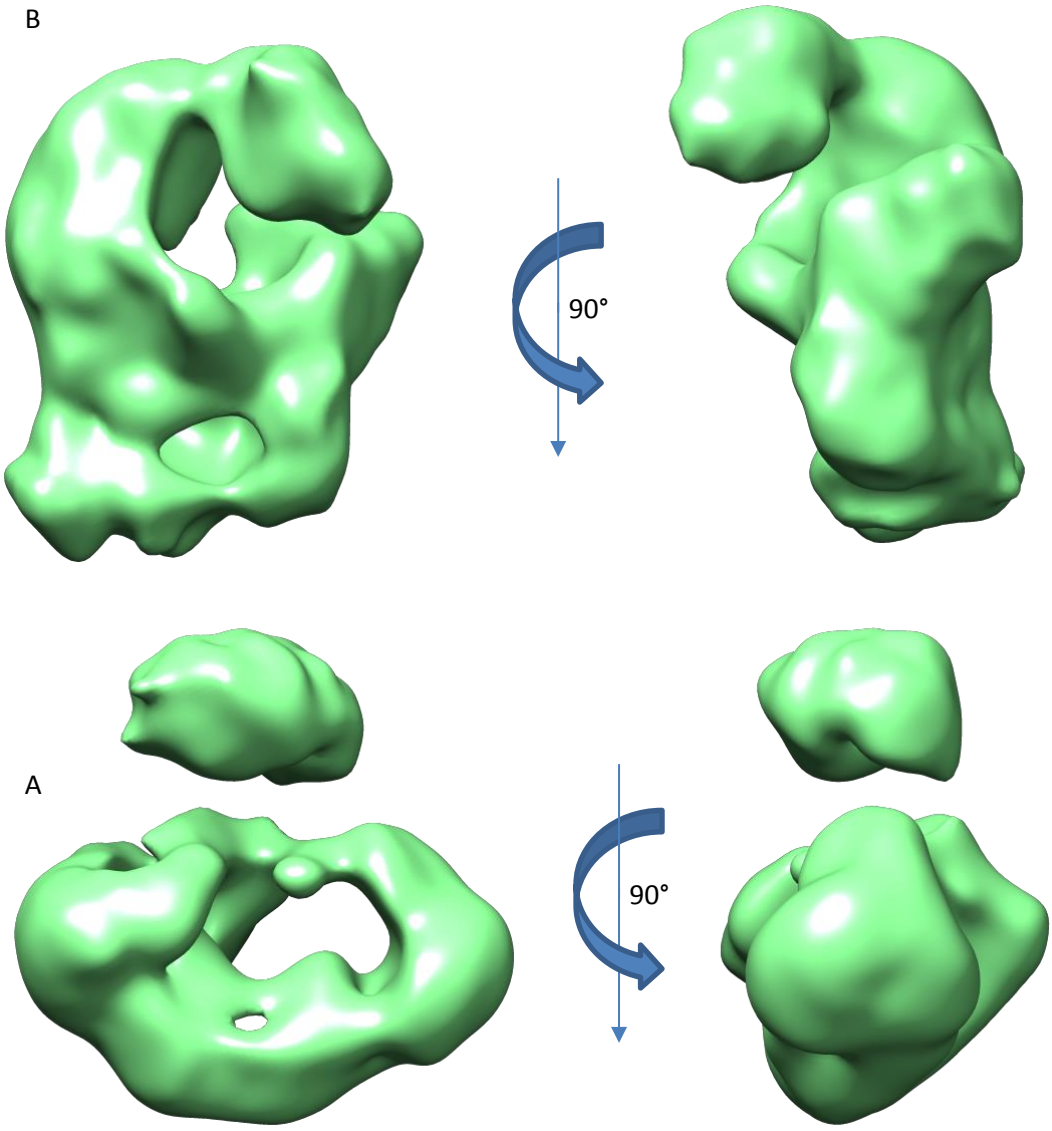


Figure 47 : Independent 3D reconstructions of the two separated lobes
Views of lobes A and B were extracted and analyzed seprately, and used to perform a 3D reconstruction of each lobe.

Surprisingly, no improvement of the resolution was obtained by this new processing. While the resolution of the whole map was calculated to 32Å (0.5 FSC criteria), the calculated resolutions (using the same criteria) of lobe A and lobe B alone were 29Å and 34Å, respectively. The low resolution of lobe B is probably due to the presence of several conformations which differ in the position of domain V, and their averaging in the final reconstruction.

3.3.2 Localization of the DUB module

When observed by EM, the SAGA complex purified from the two mutant strains (*Sgf73Δ1-104* and *Sgf73Δ*) shows similar views as the complex from the wild—type strain.

Indeed, elongated particles of similar size and shape can be recognized, suggesting that complex integrity is not altered by these deletions. The particles adopt the same preferential orientation as in the wild-type complex, and analysis of the same particular view clearly shows the presence of the flexible domain V in the case of the *Sgf73Δ1-104* mutant. The domain in both complex seems of comparable size (Figure 48, compare lines A and B), thus excluding the possibility that the DUB module is part of this domain. However, in the *Sgf73Δ1-104* mutant, a larger fraction of the complex adopt a conformation where domain V is not visible (column 1) compared to the wild-type (around 50% and 20% for the mutant and the wild-type, respectively), suggesting that loss of the DUB module, and/or of Spt8 and the C-terminal part of Spt7, can affect the dynamic properties of the lobe B.

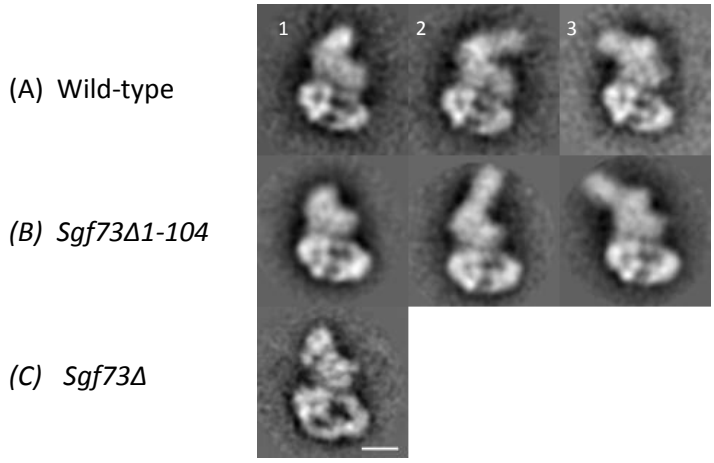


Figure 48 : Size and position of domain V in the SAGA complex

Comparison of size and position of domain V in the context of the wild-type (A), the *Sgf73Δ1-104* (B) and *Sgf73Δ* (C) strain. In (C), the domain V is not visible, and lobe B seems more homogenous. White scale bar represents 10 nm.

Surprisingly, the deletion of the entire Sgf73 subunit triggers a more dramatic effect on the behavior of the domain V. Indeed, analysis of the same view from the particles isolated in the case of the *Sgf73Δ* mutant shows the complete absence of the domain V (Figure 48, line C). Spt3, located in this domain, is still present in the sample, as shown by the mass spectrometry analysis. Thus it is unlikely that the absence of domain V correspond to the loss of the subunit present in this domain. The loss of Sgf73 might be responsible for a conformational change in the lobe B, which appears very homogeneous and very well-defined in the context of this deletion.

To localize the DUB module, a 3D model was reconstructed for both mutant, from the analysis of 9.985 and 30.409 negatively stained particles for the *Sgf73Δ1-104* and the *Sgf73Δ* strains, respectively. Because of the molecular motion between the two lobes, which limits the observation of the domain in lobe B, the same separated analysis of the two lobes

was performed for both mutants. Lobe A looks strikingly similar in the three strains, and this observation rules out the possibility of locating the DUB module, with a molecular weight of approximately 140 kDa, in this lobe (Figure 49C). In contrast, several differences are visible in the lobe B. In particular, the computing of a difference map between the wild-type and each mutant reveals (i) a missing density in both mutant in the domain III, (ii) the absence of domain V and (iii) an additional density in the wild-type, making a connection toward domain I (Figure 49A and B, for mutant *Sgf73Δ1-104* and *Sgf73Δ* respectively). The size of the missing density in domain III is compatible with the presence of the DUB module in this domain, as it was excluded from the domain V previously due to presence of the domain on some of the 2D views.

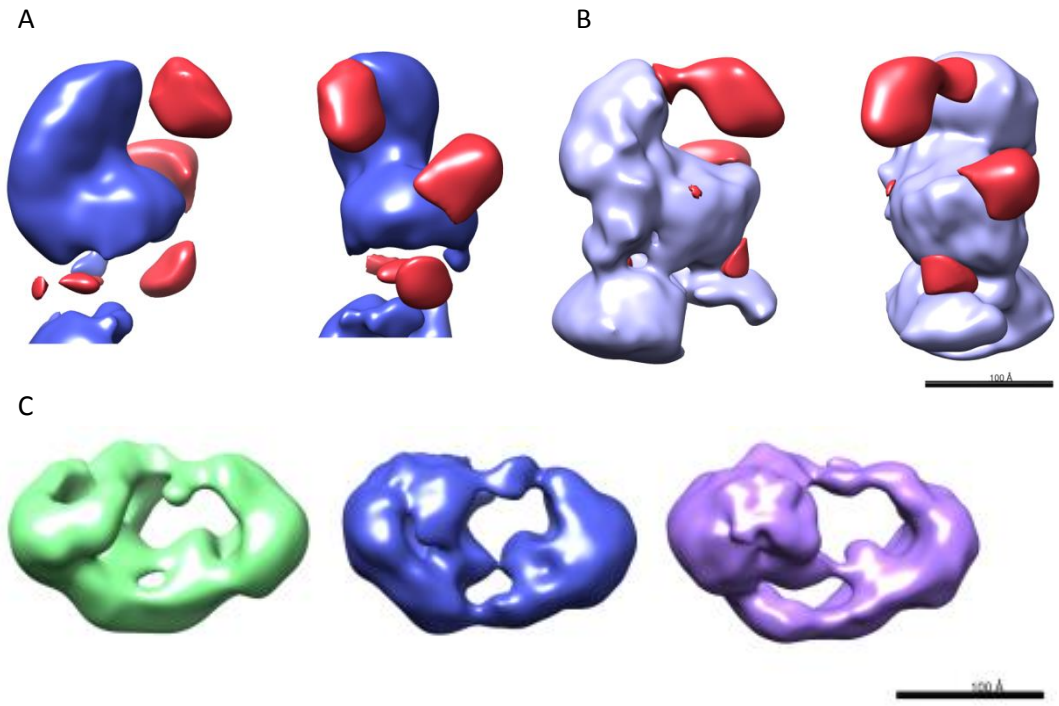


Figure 49 : Difference between the complex from the different strains.
Difference maps of the density in lobe A, calculated between wild-type and (A) Sgf73Δ1-104 or (B) ΔSgf73 mutants. Densities corresponding to the SAGA complex is in blue and pink for *Sgf73Δ1-104* and *Sgf73Δ* mutants, respectively. Additional densities present in the wild-type are in red. (C) Lobe A of the wild-type (green), *Sgf73Δ1-104* (blue) or *Sgf73Δ* (purple) complex. Black scale bars in both panels represent 10 nm.

The crystal structure of the DUB module, comprising the full Ubp8, Sgf11 and Sus1 proteins, as well as the N-terminal part of Sgf73 (amino acids 1 to 104, (Köhler et al., 2010)), was fitted into this missing volume and fills it in a unique position (Figure 50). The positioning of the structure is driven by the orientation of the Ubp8 domain composed of residues 281-303 and 326-353, which form 3 β-strands and 2 loops that extend from the module. In this position the DUB site is placed at the tip of domain III while the ubiquitin binding site is located in the back.

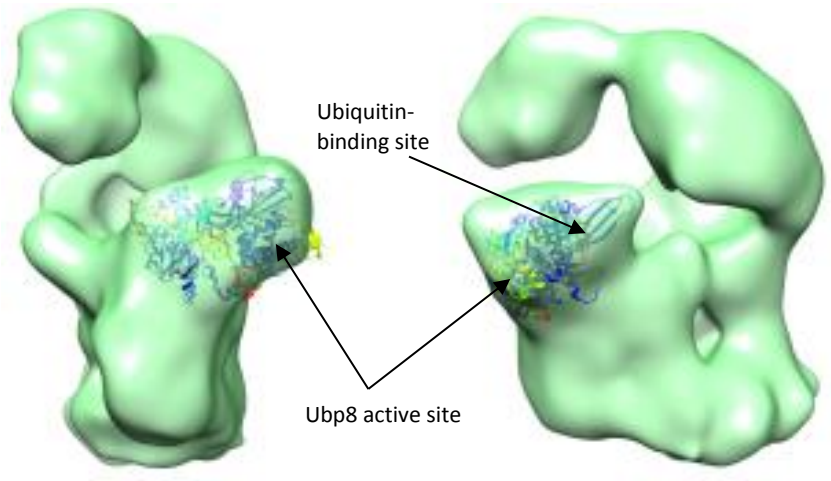


Figure 50 : Fitting of the DUB module structure in the density map

Fitting of the crystal structure of the DUB module into the additional density of the domain III in lobe B of the SAGA complex. In the representation of the crystal structure, Ubp8 is in blue, Sgf11 is in yellow, the N-terminal part of Sgf73 is in red and Sus1 is in cyan.

The comparison of the mutants and the wild-type SAGA models revealed further differences likely to describe conformational changes of the flexible domain V. The domain V appears absent in the *Sgf73Δ1-104* and in the *Sgf73Δ* mutant strains but a similarly sized density is found inserted at the bottom of the cleft (Figure 51, in blue). In agreement with the analysis of the 2-D views, this observation shows that the average position of domain V is affected in mutant SAGA complexes and that the domain V might fold back into the cleft in lobe B. This conformational change is strengthened when the Sgf73 subunit is fully deleted.

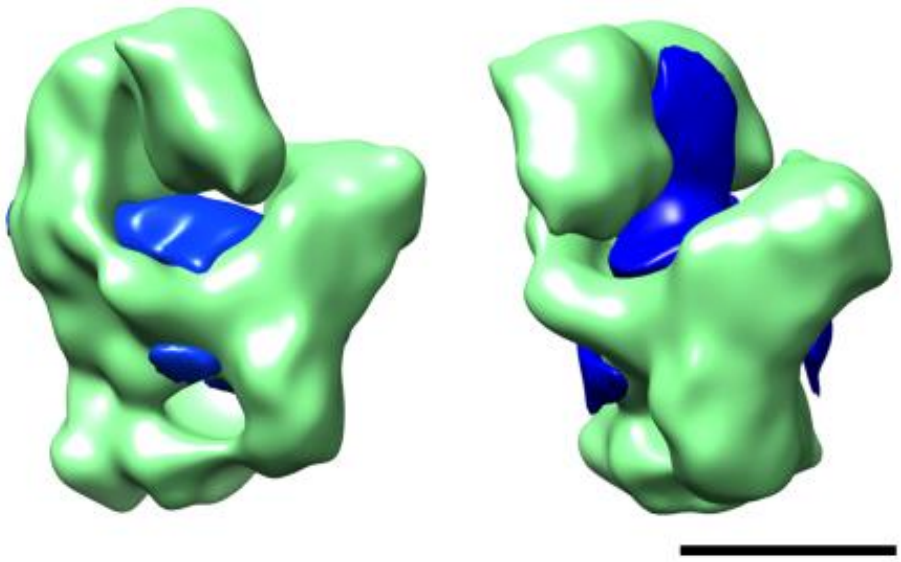


Figure 51: Additional densities in the *Sgf73Δ* mutant strain compared to wild-type
Difference map between the complex from *Sgf73Δ* mutant strain and wild-type strain shows the additional densities in the mutant (in blue). The lobe B from the wild-type is represented in green. Scale bar is 10 nm.

3.3.3 A subpopulation of SAGA does not contain the DUB module

Localization of the different subunits analyzed so far was made difficult by the heterogeneity of the sample. As described before, the complex shows some degree of flexibility, in particular in the lobe B, which can be the cause of misalignment of the particles, and thus result in loss of details in the final structure. Separation of the two lobes made possible to partially overcome the problem of flexibility between the two lobes, and allows us to eventually reveal the position of the DUB module. However, a possible source of heterogeneity is also the variation in the complex composition. In particular, a recent study has proposed that the DUB module is dynamically removed from the SAGA complex (Lim et al., 2013). We thus decided to investigate whether a subpopulation of SAGA complexes in the wild-type strain is devoid of this module.

To address this issue, we tried to separate the images composing the dataset of the SAGA complex from the wild-type strain into two groups, according to their best cross-correlation score with reprojections of the 3D model from either the wild-type or the *Sgf73Δ* SAGA complex. After several iterations, a surprisingly high subpopulation of 55% of the wild-type SAGA images correlated with a higher score to the DUB module-free *Sgf73Δ* SAGA model, suggesting that a significant fraction of the purified SAGA molecules are devoid of DUB module. The images from this dataset were used to reconstruct a 3D model of the lobe B containing (Figure 52, in green) or not containing (Figure 52, in light purple) the DUB module.

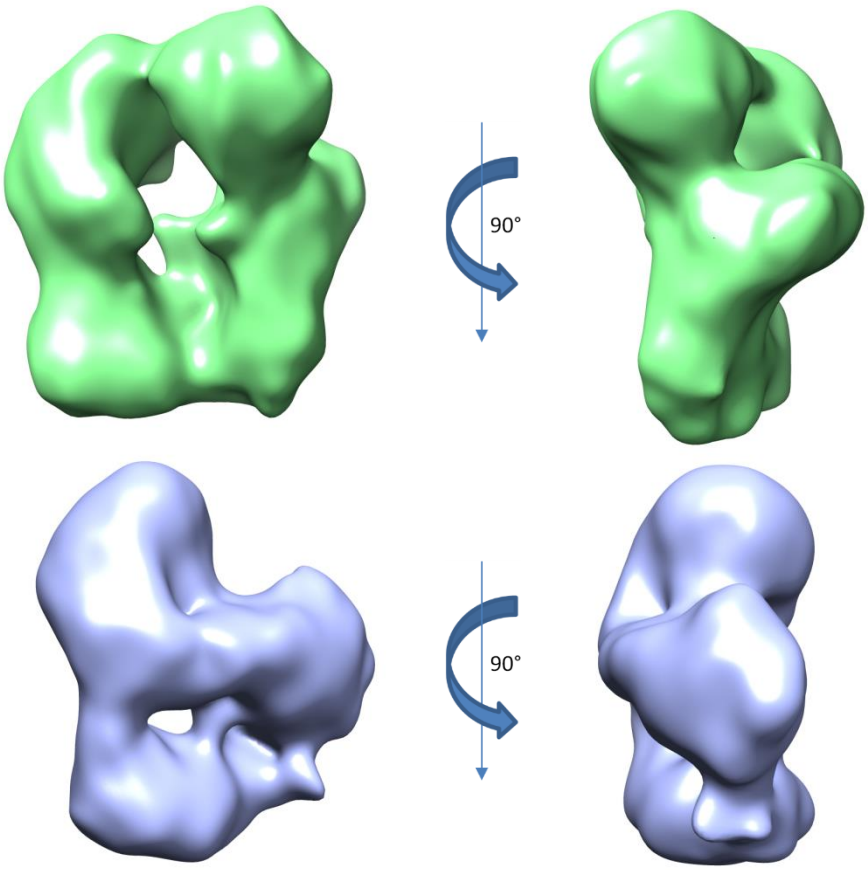


Figure 52 : Separation of a DUB-containing and a DUB-free SAGA complex

Images from the dataset of the wild-type were sorted according to their best correlation with the 3D model of the lobe B from either the wild-type or the *Sgf73Δ* mutant, and used to reconstruct a 3D model of lobe B containing (in green) or not (in light pruple) the DUB module.

In addition to the presence of the DUB module, the two models largely differ by the position of the Spt3-containing domain. Indeed, in the model containing the DUB module, a large domain V seems to be present and contacts the domain III, whereas in the model where the DUB module is absent the domain V seems to fold back and fill the cleft formed between domain III and IV. However, the separation is the based on the cross-correlation of the images with the two different models, which show dissimilarity not only at the level of the DUB module, but also at the position of the domain V. Thus, it is likely that the sorting is not based only on the presence of the DUB module, but also on the position of the Spt3-containing arm. The distribution of the images in the two models might thus not fully represent the exact proportion of SAGA complexes which do not contain the DUB module. Nevertheless, it seems that the two population exist and the position of domain V correlates with the presence of the DUB module.

3.3.4 Localization of the Spt7 and Spt8 subunits

The DUB module is placed in domain III, and doesn't fill the additional density observed between domain I and domain III. Upon deletion of the DUB module, we observed a strong cleavage of the C-terminal part of Spt7 as well as the full Spt8 protein, which account for a total molecular weight of approximately 80 kDa. As Spt7 was previously localized in this region, it would be possible that this additional density corresponds to these two proteins. In order to test this hypothesis, we performed an immuno-labelling experiment using an antibody directed against the Spt8 subunit.

So far, purification of the SAGA complex was performed by pulling on the Spt20 subunit. However, gel analysis revealed that half of the complex purified corresponds to the SLIK complex. This strong heterogeneity would decrease the efficiency of the labelling experiment. Thus, we decided to purify exclusively the SAGA complex by using the Spt7 protein with a TAP-tag at its C-terminal part as a bait. Analysis of the complex purified on PAGE-SDS confirms the presence of the longer form of Spt7 in the sample (Figure 53A).

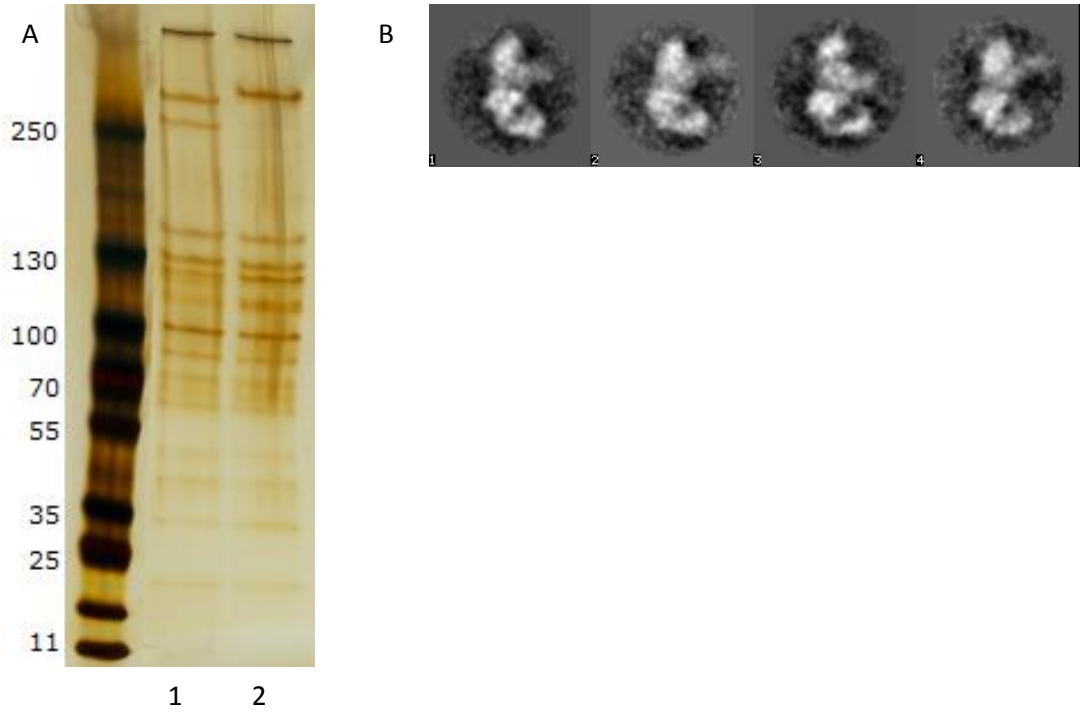


Figure 53 : Localization of the Spt8 subunit by immuno-labelling

(A) PAGE-SDS analysis of the wild-type complex purified using Spt20 (lane 1) or Spt7 (lane 2) as a bait.

(B) (B) Immuno-labelling of the Spt8 subunit.

As expected, yield of purification was lower and thus fewer particles were observed by EM. Nevertheless, a dataset of 1227 particles could be collected and analyzed.

Characteristic views of the SAGA complex show the presence of the antibody in the domain III (Figure 53B). Thus, Spt8 is localized in the vicinity of Spt7, and this suggests that in the absence of the DUB module, the C-terminal part of Spt7 which interact with Spt8 might be more accessible and is more likely to be cleaved by the Pep4 protease.

3.4 - STRUCTURE OF THE FULLY-HYDRATED SAGA COMPLEX

The structure of the SAGA complex obtained in negative stain has revealed information about the quaternary organization of the complex, and in particular has allowed us to locate the DUB module close to the Gcn5 and Spt7 subunits. However, localization of the C-terminal part of Sgf73 by the direct comparison of the two mutant strains is not possible due to the low resolution of our model.

In order to obtain high-resolution data, we started a study of the SAGA complex by cryo-EM. This would allow us to observe the sample close to its native state, without any loss of information due to dehydration and flattening of the complex. To initiate the study, a more homogenous and concentrated sample is required to facilitate observation of the particles and image processing after data collection. The improvements of the purification process allowed us to obtain a sample suitable for such study, and to observe for the first time the complex in its hydrated state. However, the concentration of the sample was not high enough to allow the observation of unsupported cryo-conditions, as it would result in a low number of particles per micrograph.

We set up the conditions for freezing of carbon-adsorbed SAGA particles and collected a dataset of 1388 micrographs. So far, particles were selected in a semi-automated way from 924 of these images, and a dataset of 26008 was analyzed. Class average images obtained after iterative alignments and classifications reveals the first views of the SAGA complex by cryo-EM.

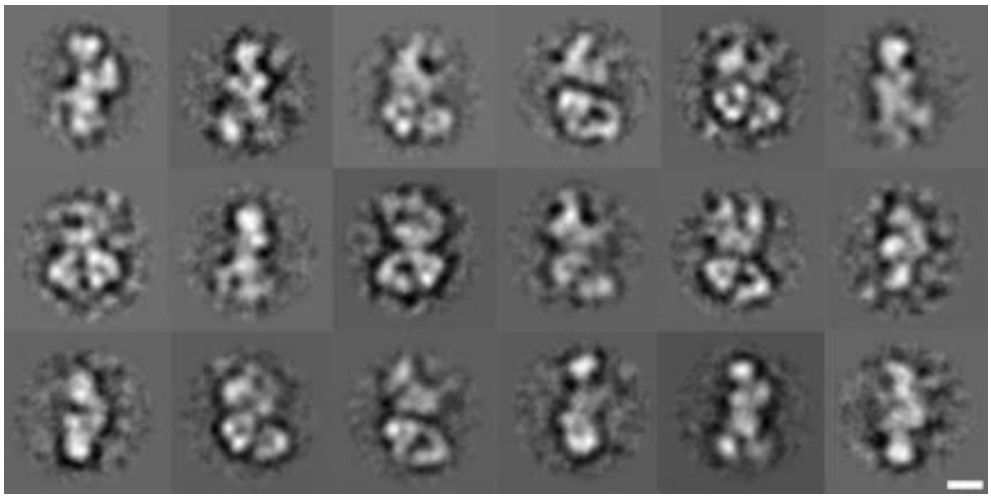


Figure 54 : Class average images of SAGA observed by cryo-EM
Class average images of SAGA particles reveals charactristic views of the complex. White scale bar is 10 nm.

Characteristic views of the SAGA complex can be recognized, the different conformations already observed are visible. More work is required to obtain a larger dataset and to perform the 3D reconstruction of the SAGA complex from these views.

3.5 - INTERACTIONS OF THE SAGA COMPLEX WITH THE NUCLEOSOME

Localization of the DUB module revealed that the 4 proteins are found in domain III, and thus regrouped with Spt7 and Gcn5 subunits. These two proteins can interact with the nucleosome and recognize specific post-translational modifications on histone tails. This suggests that the domain III might form a platform specialized in the binding and interaction with nucleosomes. However, the exact positions of the different subunits are still not known and no structural data about the binding interface between SAGA and the nucleosome are available. We thus decided to study the complex formed by SAGA and isolated mono-nucleosomes.

In order to confirm the interaction between our purified SAGA and modified mono-nucleosomes (trimethylated on lysine 4 of histone H3), we performed a binding assay with a 5-time molar excess of nucleosome incubated with immobilized SAGA complex. As SAGA is purified at a high salt concentration, we tested different salt concentrations to check if the buffer would prevent any interaction. The result of the western blot clearly shows an interaction between our purified complex and the nucleosome, at both low (50 mM NaCl) and high (500 mM) salt concentrations (Figure 55, lanes 2 and 3 respectively).

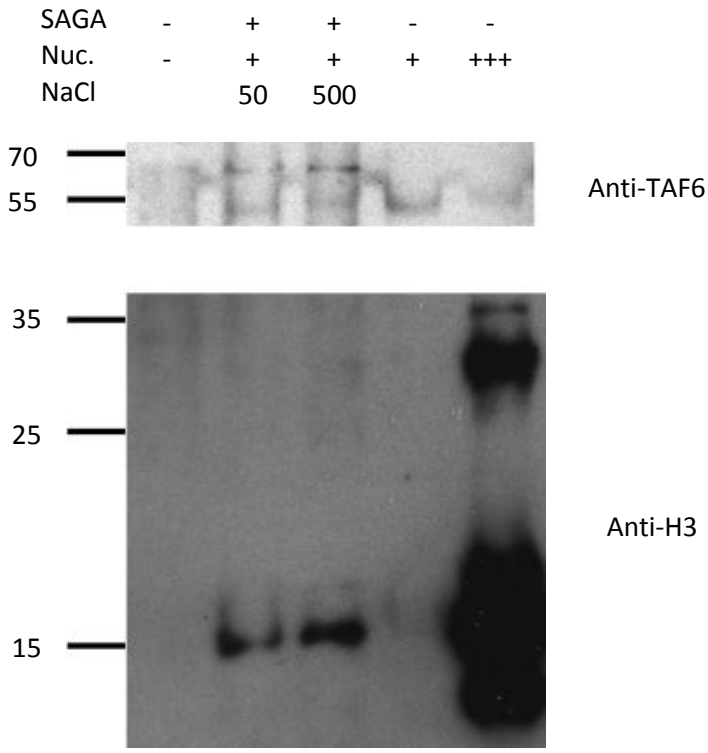


Figure 55 : Control of the interaction betwwen the purifed SAGA complex and the mono-nucleosome

Western blot revealing histone H3 after binding on immobilized SAGA-complex and washing at low (lane 2) and high (lane 3 salt concentration). SAGA is revealed by an anti-Taf6 antibody, the nucleosome is revealed by an anti-H3 antibody.

To visualize the complex formed by SAGA and the nucleosome, we mixed the purified SAGA complex with a 3-time molar excess of recombinant, modified nucleosomes. After incubation for 30 minutes on ice, the complex was fixed by cross-linking with glutaraldehyde, adsorbed on a carbon film and stained with uranyl acetate. Many isolated nucleosomes were visible on the grid, and larger, elongate particles corresponding to SAGA were picked and analyzed as previously. However, efficiency of the binding seems low in these conditions, and no complex formed by SAGA with the nucleosome was observed on the grid.

In order to improve the binding efficiency, we made a second attempt using a different strategy. SAGA was purified as previously described and unmodified nucleosomes were added in a 7.5 molar excess after the last washing of the CBP affinity step, prior to elution. After incubation on ice for 1h, the excess of free nucleosomes was quickly removed by a single washing step in the same conditions and complexes bound on the resin were immediately eluted. Grids for EM were prepared immediately after this elution, as previously described.

Despite the washing step, a large excess of nucleosome was clearly visible on the grid. We picked and analyzed a first set of 3909 images, and class-average images revealed the

presence of an additional density in the vicinity of the domain III on approximately 25% (931) of the images (Figure 56A, lines 1 and 2). On some images, this density is not well defined, probably due to variation in its positioning relative to the SAGA complex. In order to obtained better views, we performed the classification by looking at the variations only in this region of the images. This improved the definition of the object on some images (Figure 56, line 3). The size of the additional density is compatible with the presence of a nucleosome, and the localization of this interaction is in agreement with our results on the subunit composition of this region.

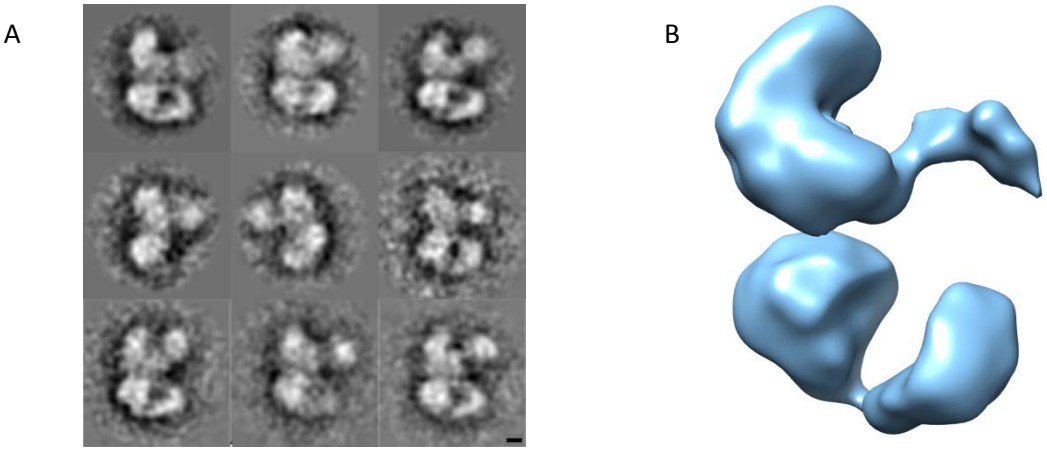


Figure 56 : Visualization of the complex formed between SAGA and mono-nucleosome

Characteristic views showing an additional density in the vicinity of domain III. Classification was performed on the whole particles for lines 1 and 2, and with a tight mask encompassing only the nucleosome region for line 3. Black scale bar is 10 nm. (B) Three dimensional recontruction from manually selected views showing an additional density close to domain III.

The nucleosome appears at a variable distance of SAGA and in different positions and orientations. Thus, it is possible that the contacts between the two complexes involve flexible parts of the complexes.

We manually selected the views showing clearly the presence of the supplementary density and performed a three-dimensional reconstruction in order to visualize the contacts of the nucleosome relative to the SAGA complex. The structure obtained revealed the presence of a small additional domain connected through the domain III (Figure 56B). Size and shape of this additional domain is not sufficient to clearly identify the positioning of a nucleosome, probably because of variation in the positioning observed in the 2D views.

CHAPTER IV :

DISCUSSION

Chapter 4. Discussion

4.1 - Architecture of the SAGA complex

To better understand the molecular architecture of the complex, we solve the structure of SAGA complex using EM. We analyzed a larger dataset of negatively stained particles than in the previous study, and particular attention was paid to the flexible domain for which the molecular motions were not fully described before. SAGA is composed of two lobes which can move relatively to each other. Lobe B is composed of three domains, one of whose is subjected to large molecular motions which define several conformations of the lobe B. This large heterogeneity in the conformations of SAGA has been so far the major limitation to high resolution structure determination, and hampered the unambiguous localization of the DUB module within the entire complex. An alternative approach has been used, in which images corresponding to each lobes were extracted from the initial, aligned images and analyzed separately. This approach has been used before for the study of the human telomerase (Sauerwald et al., 2013), and can be used in the case of SAGA because the elongated shape of the complex produces views with no overlap, where both lobe can be unambiguously identified.

The lobe A of the SAGA complex seems very stable and no significant variation was observed which could reflect conformational changes or variation in the composition of the complex. Lobe A is composed of two domains, connected by three stable protein links. The domain I was shown to contain the Tra1 subunit (Wu et al., 2004), and the entire lobe looks remarkably similar to another Tra1-containing complex, the NuA4 complex (Chittuluru et al., 2011, p. 4). Indeed, the NuA4 complex is composed of two domains, one of which is entirely composed of the Tra1 protein. Interestingly, the second domain looks highly similar to domain II of SAGA lobe A, while their subunit composition is very different and shows no sequence homology. The second NuA4 domain contains the piccolo catalytic HAT subcomplex (Epl1, Yng2, Esa1 and Eaf6) as well as NuA4-specific subunits (Eaf2, Yaf9, Act1, Arp4, Eaf1, Eaf5, Eaf7 and Eaf3) while SAGA domain II was shown to contain a subset of Tafs and Ada1. In addition, the Tra1-containing domain makes three similar protein bonds to its neighbor domain in the SAGA complex and in the NuA4 complex, suggesting that the same interaction surface is used

during incorporation of Tra1 in both complexes. This is in agreement with recent results of a genetic study where the same mutations where responsible for the loss of Tra1 in both SAGA and NuA4 (Knutson and Hahn, 2011). However, the localization of the nucleosome binding interface seems to be different between SAGA and NuA4. Indeed, binding of nucleosome to NuA4 involves the Epl1 subunit located in in the second domain of NuA4, in the vicinity of the interface between Tra1 and the rest of the complex. In SAGA, the nucleosome binding region seems to be located in the domain III, which has no equivalent in NuA4. The bromodomain and catalytic domain of Gcn5, the bromodomain of Spt7, the UcH domain in Ubp8, the SCA7 domain of Sgf73, the tandem tudor domain of Sgf29 and the ZnF domain in Sgf11 were all shown to interact with the nucleosome (Bian et al., 2011; Bonnet et al., 2010; Hassan et al., 2002; Kohler et al., 2010, 2014) and clusters in this region of SAGA structure ((Wu et al., 2004) and see results in this work). Thus, the precise position of the nucleosome in respect to the activator-binding site seems to be different between NuA4 and SAGA. In agreement with this, NuA4 was shown to be enriched at TFIID-dependent promoter, which show a different spatial distribution of transcription regulators compared to SAGA-dependent genes (Venters et al., 2011).

Previous quantitative mass spectrometry analysis has revealed five distinct functional modules which composed the architecture of the SAGA complex (Lee et al., 2011). Immuno-labelling experiments are in good agreement with the modular organization of the SAGA complex. However, in this proteomic study the Tra1 subunit was shown to be related to the Spt group of protein, as deletion of Spt20 was leading to the loss of the large Tra1 protein and of the DUB module, but other subunits could still be purified. This was unpredicted, as Spt20 was found to be located at the top of the lobe B, at the opposite of Tra1. Thus, it seemed very unlikely that these two subunits can interact directly. Here we located Spt7 and Spt8 in the close vicinity of Tra1, and these two proteins seem to be able to contact the Tra1 domain in certain conformations. In addition, the Spt3-containing module undergoes large movement and seems to contact the Spt8 region. Thus, it is possible that an interaction occurs between Tra1 and the Spt proteins. Whether this interaction really exists, which subunits are involved, and whether it plays a functional role in the transduction of the activation signal requires more investigations.

The domain II and IV were shown to contain a subset of Tafs proteins present in SAGA (Wu et al., 2004). These proteins are shared with the TFIID complex where they are present in two copies (S. Sanders et al., 2002) and assemble in a core TFIID complex (Bieniossek et al., 2013). This complex contains two copies of Taf5, of the heterodimers Taf6/Taf9 and

Taf4/Taf12, and a single copy of the Taf8/Taf10 heterodimer which breaks the twofold symmetry of this core complex upon its incorporation. In SAGA, it was proposed that Taf8 and Taf4 were replaced by the HFD-containing subunits Spt7 and Ada1 in the Taf4/Taf12 and Taf8/10 heterodimers, respectively (Gangloff et al., 2001, 2000). However, it is not clear if such a core complex exist in SAGA. Our data suggest a faint connection between the domain II and domain IV probably due to the rotation of the two lobes in respect to each other's. It is thus not clear how the two lobes contact each other's, but the presence of such a compact core complex seems unlikely. Moreover, the stoichiometry of the Tafs subunits in SAGA is not clearly established. The presence of two copies of the Tafs subunits, as well as Ada1 and Spt7, was proposed based on immuno-labelling experiments and the similarity observed between SAGA and TFIID (Wu et al., 2004). Our proteomic analysis of the SAGA complex has revealed a higher content of only Taf6 and Taf9, but is not compatible with the presence of two copies of these subunits. The characterization of SAGA composition and subunits stoichiometry still requires more investigation.

4.2 - THE DUB MODULE CLUSTERS WITH GCN5 AND DEFINES A NUCLEOSOME BINDING INTERFACE

In this work, we have localized the DUB module within the structure of the entire SAGA complex, by analyzing two mutant strains where this particular module is missing. Deletion of Sgf73 or of its N-terminal part affects the binding of the module to SAGA but do not compromise the integrity of the complex (Köhler et al., 2008; Lee et al., 2011). In agreement, we found that no major change occurs in the overall structure of the SAGA complex. Due to the large heterogeneity in the conformations of lobe B, the unambiguous localization of the DUB module was only possible when the two lobes were analyzed separately. The module protrudes from domain III and colocalizes with the Gcn5 and Spt7 proteins. Gcn5 is part of the HAT module, which contains Sgf29. Both Gcn5 and Sgf29 contain domains involved in recognition of specific histone modifications, namely a bromodomain and a Tudor domain. In addition, Spt7 also contains a bromodomain, even though it is not clear if this protein interact with the nucleosome in the context of SAGA (Hassan et al., 2002).

This spatial proximity is consistent with the observed inhibition of SAGA's HAT activity when the ATX7N containing a polyQ expansion is engineered in yeast to replace Sgf73 (Burke et al., 2013; Palhan et al., 2005)). ATX7N was shown to directly bind the catalytic core of

Gcn5 and to inhibit its acetylation activity, and a long and unstructured polyQ stretch is possibly within reach of the Gcn5 catalytic activity or nucleosome recognition functions. Whether this inhibition is also observed in the context of human SAGA remains to be investigated.

Our preliminary results on the interaction of SAGA with purified mono-nucleosomes suggest that the binding occurs in this region and involves protein contacts in the domain III. The nucleosome is not tightly bound to SAGA, but seems to rotate relative to the position of the SAGA complex. This would indicate that additional contacts, possibly with the DNA close to the nucleosome or specific modifications on the histone tails, might be necessary to stabilize the interaction and place the nucleosome in an appropriate position. The nucleosome binding-site would be located close to the TBP-binding interface formed by Spt8 and the closed conformation of the Spt3-containing arm, and this raise the question whether SAGA would bind and modify the nucleosome +1 in the context of the promoter. However, due to the limited number of particles analyzed and the lack of accurate information on the localization of the different subunits in this region, it is not possible to determine the contribution of the different subunits to this binding. Moreover, we used purified, recombinant mono-nucleosome, which does not fully represent the nucleosome and its post-translational modifications in the context of the chromatin.

4.3 - THE SAGA COMPLEX FORMS A MOLECULAR CLAMP WHICH MIGHT BIND TBP

The interaction between TBP and the Spt3/Spt8 proteins has been largely studied. Genetic and biochemical data have pointed out the direct interactions of these two proteins with TBP (Eisenmann et al., 1994, 1992; Laprade et al., 2007; Mohibullah and Hahn, 2008; Sermwittayawong and Tan, 2006; Warfield et al., 2004). Spt3 and Spt8 were shown to be required for recruitment of TBP to a subset of genes (Bhaumik and Green, 2002) and to bind TBP *in vivo* (Mohibullah and Hahn, 2008). Recruitment of TBP upon activation of transcription requires Spt3, but is not dependent on Spt8 in agreement with the model in which Spt8 plays a negative role in the regulation of TBP function (Mohibullah and Hahn, 2008; Sermwittayawong and Tan, 2006; Warfield et al., 2004). Binding of TBP to Spt3 was mapped on residue 177 to 196 (Laprade et al., 2007; Mohibullah and Hahn, 2008), while Spt8 interact with a different region of TBP, located

around residue 79 (Mohibullah and Hahn, 2008). Thus Spt3 and Spt8 seem to interact at both sides of TBP (Figure 57A).

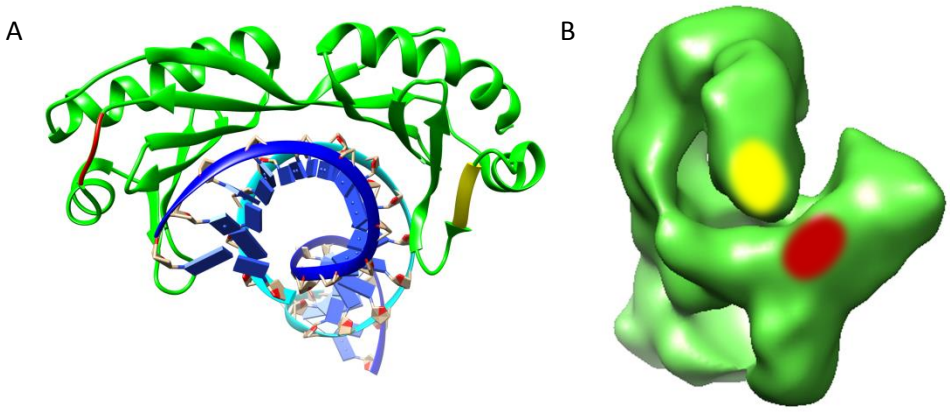


Figure 57 : Position of TBP-SAGA interacting regions

(A) Spt3 and Spt8 interacts in two distinct region of TBP (in green). Spt8-TBP cross-linking experiments revealed that binding occurs with residues close to amino acid 78 of TBP (in red), while Spt3 interaction is located close to amino acid 184 (in yellow). DNA is represented in blue. (B) Localization of Spt8 (in red) and Spt3 (in yellow) on the SAGA complex according to immuno-labelling experiments.

Our results show that Spt3 and Spt8 are located at two different positions in the SAGA complex. The lobe B is composed of three domains which form a molecular clamp, and can adopt an opened or closed conformation. Previous immuno-labelling experiments, as well as experiments done in this work, indicate that the two TBP-interacting regions are located at the tips of this clamp (Figure 57B). Indeed, Spt3 is found in the very mobile domain V, which seems to be able to contact the Spt8 region. Thus, the lobe B, in his closed conformation, could bind TBP from both sides. In this proposed position, binding of TBP would be different of the one observed in the context of TFIID, where TBP is buried at the bottom of the DNA binding cleft.

This close conformation seems to be more abundant when the DUB module is present. However it is not clear if this conformation is favored in the presence of the DUB module or is rather related to the cleavage of Spt7 and subsequent loss of Spt8, which seems to be more frequent in the absence of this module as observed with the deletion mutants. In our model, loss of one of the two TBP interacting regions would obviously prevent the closing of the clamp. Spt8 was proposed to negatively regulate TBP function by competing with DNA-binding. Thus, this close conformation might act as a repressor of initiation, and release of Spt8-binding to TBP, possibly by the removal of Spt8, could allow the delivery of TBP to the DNA by the Spt3-containing domain of SAGA. The large mobility of this arm could be required to accommodate to different promoter conformation.

4.4 - THE DYNAMIC OF THE DUB MODULE ALTERS THE TBP-BINDING SURFACE OF THE SAGA COMPLEX

We observed a more frequent cleavage of Spt7 in the absence of the DUB module. The physical proximity of the module and Spt7 might explain this observation. Indeed, in the absence of the module, the C-terminal part of Spt7 might be more accessible to the Pep4 protease, and thus lead to an enrichment of the C-terminally processed form of Spt7 in the two mutant strains. It is not clear yet if the cleavage of Spt7 is occurring *in vivo* or rather during the purification. The latter possibility would be in agreement with the absence of the Rtg2 protein in our preparation, which was identified as a SLIK specific subunit (Pray-Grant et al., 2002). However, this raises the question whether the cleaved form of Spt7 in our wild-type SAGA preparation have the same origin as in the mutant strains, namely if Spt7 cleavage has occurred in the absence of the DUB module. Indeed, sorting of the particles from this sample revealed the presence of two populations in the dataset, one of which seems to not contain the DUB module. Thus, the ratio between the cleaved and uncleaved forms of Spt7 might partially reflect the proportion of SAGA which does not contain the DUB module in our sample. However, separation of SAGA and SLIK by ion exchange column have pointed out an opposite results, where the Ubp8 would be more abundant in SLIK rather than SAGA (Daniel et al., 2004). This could be explained by the different purification conditions used in this study or by a different dynamic of the DUB module in the absence of Spt7 C-terminal part and Spt8. Indeed, the DUB module was proposed to dynamically dissociate and reassociate to SAGA, and cleavage and loss of these two polypeptides could favor the reassocation of the module to the SLIK complex, as Spt7 C-terminal part and Spt8 were shown to be dispensable for interaction with the DUB module (Lee et al., 2009).

The SLIK complex is thought to be involved in the regulation of specific sets of genes, and to have a partially overlapping role with SAGA (Pray-Grant et al., 2002). In particular, SLIK is more abundant upon amino acid starvation and might regulate the expression of genes involved in amino acid synthesis pathway such as HIS3 or TRP3, where Spt8 was shown to play a negative role (Belotserkovskaya et al., 2000). However, this regulation seems independent of the Spt7 cleavage, as a mutant where this processing is hampered does not show any severe growth defect on media lacking histidine (Wu and Winston, 2002). In this work, we observed a role for Sgf73 in regulating the position of the Spt3-containing domain, as removal of the N-terminal part of Sgf73 slightly affects the position of this arm while deletion of the full protein leads to a complete folding back of this domain into the cleft formed by domains III and IV in

lobe B. This opens the possibility that the regulation of these genes might involve at least partially the turnover of the DUB module. As already discussed, the loss of the DUB module favors the SLIK form of Spt7 in our samples. We observed that in this case, the domain V is preferentially buried into the cleft formed in the lobe B. It is possible that this conformation is favored by the loss of Spt8 subsequent to the cleavage of Spt7, however the deletion of the full Sgf73 has a dramatic effect on the position of this domain, suggesting that Sgf73 might unexpectedly regulate the TBP-binding properties of SAGA by triggering a conformational change in the lobe B.

The biological activities of SAGA such as activator binding, promoter recruitment and HAT activity have been shown to be regulated by its interaction with the 19S proteasome RP (Lee et al., 2005a). The inhibition of the ATPase activity of RP reduces global H3 histone acetylation, a possible consequence of impaired SAGA recruitment to promoters. More recent advances showed that in the presence of the ATPase Rpt2 protein, this proteasomal ATPase can dissociate the DUB module from SAGA (Lim et al., 2013). This observation is consistent with our identification of a DUB-depleted subpopulation of SAGA complexes and suggests that this transition is tightly regulated. We also observe a conformational change upon deletion of the DUB module which modifies the TBP-binding interface of SAGA and might also indirectly affect the activator binding subunit of the complex, by the loss of a possible contact between domain III and Tra1. This would provide a possible explanation for the observed effect of the interaction between the 19S proteasome RP and SAGA.

CHAPTER V :

CONCLUSIONS AND PERSPECTIVES

Chapter 5. Conclusions and Perspectives

Transcription in eukaryotes is a complex process, which is regulated at all its different steps. A large number of proteins and complexes are required to allow the recruitment of the transcription machinery, the establishment of a competent chromatin environment, the recognition of the transcription start site, the initiation of the synthesis of the RNA as well as the processive elongation of this molecule, its maturation and its export to the cytoplasm. Transcription is highly dynamic and requires the coordinated action of many players which will act cooperatively or in opposition to efficiently regulate and control the expression of genes upon internal or external signal.

The SAGA complex is a major component of the transcription machinery and is involved in several steps of transcription regulation. Since its discovery more than fifteen years ago, many genetic, functional and biochemical studies have help to shed light on its functions and the molecular mechanisms involved in this process. However, the structural information on the complex is still limited to isolated domains or module, whereas data on the whole complex are limited to low resolution density map. A previous study in our group has revealed the overall shape and architecture of the complex, allowing us to gain insights on how the different functions are integrated into a larger complex. Since then, crystal structure of the DUB module has helped us to understand the role of the different components of this module in the regulation of the DUB activity, but information on its incorporation in the context of SAGA was still lacking.

In this work, we have tried to bring new insights on the molecular organization of the complex and to understand how it could exert its function. Using EM, we analyzed the dynamic properties of the SAGA complex, in order to understand how the positioning of the different modules participates in the functions of the complex. Our work revealed the presence of a molecular clamp showing a large mobility and we proposed that this clamp plays a role in the regulation of TBP binding and recruitment to the promoter. The localization of a TBP-binding surface in the vicinity of the activator binding site and the existence of a possible molecular connection between them provide a potential mean to convey the activation signal. The study of the localization of the DUB module within the whole complex suggested an unexpected role for Sgf73 in the regulation of the TBP-binding properties of the SAGA complex, and further studies will determine the significance of this finding. Eventually, we obtained promising preliminary results to study how the SAGA complex interacts with the

chromatin by visualizing the interaction with nucleosomes. However, many questions are still unanswered and more works will be required to address them.

Despite our efforts, we were not able to improve the resolution of the existing model for SAGA. Several reasons can be offered to explain this:

(i) The purification of SAGA probably produces a mixture of complexes with variation in their composition. This can be due to the existence of several forms of the complex in the cell, but also to the degradation and dissociation of a rather fragile protein assembly *in vitro*. Indeed, our image analysis shows the presence of a small fraction of subcomplexes which are likely to correspond to the separation of the two lobes which does not resist to the harsh experimental conditions. An improvement of the working conditions, such as milder lysis method, optimal buffer composition, higher concentration of the complex, shorter incubation and more efficient separation would probably allow us to obtain sample of better quality. This would require obtaining higher amount of sample, which would be beneficial for the characterization of the complex at the biochemical level. Indeed, the exact stoichiometry of the different subunit of the complex is still not known and need to be assessed.

(ii) As already discussed, the SAGA complex is subjected to large molecular motions of several domains, and shows a large flexibility. This is a major limitation for obtaining high resolution data on such a large complex. Recent advances in algorithm for classification and reconstruction of the different conformations comprising a dataset will allow us to separate structurally different sets of complexes, thus allowing us to improve the resolution of the different forms present in our sample. This would require the acquisition of a larger dataset, in order to obtain detailed views of the different conformations.

(iii) The use of negative stain limits the resolution to surface features and is responsible for a flattening of the structure. This method will thus limit the resolution of our study. Cryo EM has the potential to reveal the different conformations of the complex in close-to-native conditions. Improvement of the sample production during this work has allowed us to initiate for the first time a structural study of the SAGA complex in fully hydrated conditions.

Improving the resolution of our structure would help us to understand the role and the contribution of the different subunits, for all issue discussed below.

The localization of the DUB module has shown the incorporation of this module in the context of the whole complex. We could propose a model for the fitting of the crystal structure of the DUB module within SAGA, positioning some nucleosome interacting domains relative to SAGA. However, the position of Sgf73 could not be determined by the direct comparison of our structure from the two mutant strains. Thus, we still don't know exactly how the module is anchored to the complex. It would be also important to understand how the module contacts the other subunits of the complex.

The DUB module was localized in the vicinity of Gcn5, which is forming a HAT module with Ada2, Ada3 and Sgf29. However, the relative position of these two modules is still not known. The accurate localization of these subunits, and their orientations relative to the DUB module would allow us to more precisely define the interface for nucleosome interaction. The localization of the DUB module was made possible by the analysis of two deletion mutant strains. As Gcn5 is not required for cell viability in yeast, a similar study can be initiated on mutants where Gcn5, Ada2 or Ada3 has been deleted, and might revealed the incorporation of the HAT module within the SAGA complex.

The understanding of the molecular mechanisms of SAGA functions, and their regulation, will require the study of functional complexes. We have started to study the interaction of SAGA with TBP or with nucleosome, and more work is required to validate the different models:

In our work, we have identified a molecular clamp comprising two different TBP-binding surfaces at the tips, and we proposed that TBP might be bound by this clamp to be delivered in a second step to the TATA-box. Validation of this model would require the reconstitution of the complex formed by SAGA in association with TBP. Solving the structure of TBP-bound SAGA would allow us to understand the role of Spt3 and Spt8 in regulating TBP association with the DNA, and the mechanism by which SAGA is dominant at certain promoters in opposition to TFIID-dependent promoters.

We started to investigate the binding of SAGA to the nucleosome, and preliminary results suggest that this binding occurs via a specialized domain of the SAGA complex. However, binding might be variable and correct positioning of the nucleosome might require larger contacts provided by the DNA molecule around the nucleosome, but also specific

modifications on histone tails which are recognized by the different subunits in this region. Nevertheless, this results need to be validated by the analysis of more particles to make possible the visualization of the complex formed between SAGA and the chromatin.

Comparison of SAGA structure to NuA4 has revealed the striking similarity between these two complexes. Both complexes share the Tra1 subunits which plays a role in the recruitment of both to a different subset of promoters. Tra1 seems to adopt the same conformation and to make the same contacts in NuA4 and SAGA, suggesting a conserved activator binding mechanism in these two complexes. However, many questions in this process are still opened: What are the surfaces involved in the binding of the activators, and of the DNA? What are the conformational changes triggered by this binding? How the signal is convey from the activator to the rest of the complex? What is the path of the promoter DNA and how is positioned the TATA-box relative to the activator binding site? What drives the specific recruitment of SAGA or NuA4 to the promoters of different genes? Solving the structure of a SAGA complex bound to an activator and its DNA template could answer some of these questions.

The role of the Taf proteins in the SAGA complex has not been investigated in this work. Tafs are shared with the coactivator TFIID, and thought to play a structural role by serving as a plateform for complex assembly. However, some Tafs might have additional roles, such as activator binding or regulation of enzymatic activities. The precise localization of this subunits, and the contacts made with other components of SAGA, will be of major interest to understand their contribution to SAGA function. The mode of assembly in the TFIID complex have been studied by the reconstitution of a core subcomplex of Taf and subsequent addition of more subunits (Bieniossek et al., 2013). This study has revealed that a symmetric core was first assembled, and addition of a two subunits (namely TAF8 and TAF10) was breaking the symmetry. The first steps of this assembly involve the formation of a heterotrimer formed by shared taf subunits, before adding TFIID specific subunits. A similar approach can be considered in order to understand the assembly of the SAGA complex, and how the two complexes are differentially formed. Indeed, while some Tafs subunits are not present in SAGA, they were proposed to be replaced by other HFD-containing subunits, and might form a similar but distinct structure in SAGA.

CHAPTER VI :

REFERENCES

Chapter 6. REFERENCES

Agatep, R., Kirkpatrick, R.D., Parchaliuk, D.L., Woods, R.A., Gietz, R.D., 1998. Transformation of Saccharomyces cerevisiae by the lithium acetate/single-stranded carrier DNA/polyethylene glycol protocol. Tech. Tips Online 3, 133–137. doi:10.1016/S1366-2120(08)70121-1

Allard, S., Utley, R.T., Savard, J., Clarke, A., Grant, P., Brandl, C.J., Pillus, L., Workman, J.L., Cote, J., 1999. NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. EMBO J. 18, 5108–5119. doi:10.1093/emboj/18.18.5108

Asturias, F.J., Jiang, Y.W., Myers, L.C., Gustafsson, C.M., Kornberg, R.D., 1999. Conserved Structures of Mediator and RNA Polymerase II Holoenzyme. Science 283, 985–987. doi:10.1126/science.283.5404.985

Auble, D.T., Wang, D., Post, K.W., Hahn, S., 1997. Molecular analysis of the SNF2/SWI2 protein family member MOT1, an ATP-driven enzyme that dissociates TATA-binding protein from DNA. Mol. Cell. Biol. 17, 4842–4851.

Baek, H.J., Malik, S., Qin, J., Roeder, R.G., 2002. Requirement of TRAP/Mediator for Both Activator-Independent and Activator-Dependent Transcription in Conjunction with TFIID-Associated TAFIIs. Mol. Cell. Biol. 22, 2842–2852. doi:10.1128/MCB.22.8.2842-2852.2002

Bai, X., Fernandez, I.S., McMullan, G., Scheres, S.H., 2013. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. eLife 2. doi:10.7554/eLife.00461

Balasubramanian, R., Pray-Grant, M.G., Selleck, W., Grant, P.A., Tan, S., 2002. Role of the Ada2 and Ada3 Transcriptional Coactivators in Histone Acetylation. J. Biol. Chem. 277, 7989–7995. doi:10.1074/jbc.M110849200

Bannister, A.J., and Miska, E.A. (2000). Regulation of gene expression by transcription factor acetylation. Cell. Mol. Life Sci. CMLS *57*, 1184–1192.

Barford, D., Takagi, Y., Schultz, P., Berger, I., 2013. Baculovirus expression: tackling the complexity challenge. Curr. Opin. Struct. Biol. 23, 357–364. doi:10.1016/j.sbi.2013.03.009

Basehoar, A.D., Zanton, S.J., Pugh, B.F., 2004. Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. Cell 116, 699–709. doi:10.1016/S0092-8674(04)00205-3

Bastien, J., Adam-Stitah, S., Riedl, T., Egly, J.-M., Chambon, P., Rochette-Egly, C., 2000. TFIIH interacts with the Retinoic Acid Receptor gamma (RARγ) and phosphorylates its AF-1 Activating domain through cdk7. J. Biol. Chem. doi:10.1074/jbc.M001985200

Belotserkovskaya, R., Sterner, D.E., Deng, M., Sayre, M.H., Lieberman, P.M., Berger, S.L., 2000. Inhibition of TATA-Binding Protein Function by SAGA Subunits Spt3 and Spt8 at Gcn4-Activated Promoters. Mol. Cell. Biol. 20, 634–647.

Berger, S.L., Piña, B., Silverman, N., Marcus, G.A., Agapite, J., Regier, J.L., Triezenberg, S.J., Guarente, L., 1992. Genetic isolation of ADA2: A potential transcriptional adaptor required for function of certain acidic activation domains. Cell 70, 251–265. doi:10.1016/0092-8674(92)90100-Q

Bhaumik, S.R., Green, M.R., 2001. SAGA is an essential in vivo target of the yeast acidic activator Gal4p. Genes Dev. 15, 1935–1945. doi:10.1101/gad.911401

Bhaumik, S.R., Green, M.R., 2002. Differential requirement of SAGA components for recruitment of TATA-box-binding protein to promoters in vivo. Mol. Cell. Biol. 22, 7365–7371.

Bhaumik, S.R., Raha, T., Aiello, D.P., Green, M.R., 2004. In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer. Genes Dev. 18, 333–343. doi:10.1101/gad.1148404

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 J. 30, 2829–2842. doi:10.1038/emboj.2011.193

Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., Berger, I., 2013. The architecture of human general transcription factor TFIID core complex. Nature 493, 699–702. doi:10.1038/nature11791

Bonnet, J., Wang, Y.-H., Spedale, G., Atkinson, R.A., Romier, C., Hamiche, A., Pijnappel, W.W.M.P., Timmers, H.T.M., Tora, L., Devys, D., Kieffer, B., 2010. The structural plasticity of SCA7 domains defines their differential nucleosome-binding properties. EMBO Rep. 11, 612–618. doi:10.1038/embor.2010.98

Brand, M., Leurent, C., Mallouh, V., Tora, L., Schultz, P., 1999a. Three-Dimensional Structures of the TAFII-Containing Complexes TFIID and TFTC. Science 286, 2151–2153. doi:10.1126/science.286.5447.2151

Brand, M., Yamamoto, K., Staub, A., Tora, L., 1999b. Identification of TATA-binding Protein-free TAFII-containing Complex Subunits Suggests a Role in Nucleosome Acetylation and Signal Transduction. J. Biol. Chem. 274, 18285–18289. doi:10.1074/jbc.274.26.18285

Brown, C.E., Howe, L., Sousa, K., Alley, S.C., Carrozza, M.J., Tan, S., Workman, J.L., 2001. Recruitment of HAT Complexes by Direct Activator Interactions with the ATM-Related Tra1 Subunit. Science 292, 2333–2337. doi:10.1126/science.1060214

Brownell, J.E., Allis, C.D., 1996. Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. Curr. Opin. Genet. Dev. 6, 176–184. doi:10.1016/S0959-437X(96)80048-7

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., Allis, C.D., 1996. Tetrahymena Histone Acetyltransferase A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation. Cell 84, 843–851. doi:10.1016/S0092-8674(00)81063-6

Bryant, G.O., Ptashne, M., 2003. Independent Recruitment In Vivo by Gal4 of Two Complexes Required for Transcription. Mol. Cell 11, 1301–1309. doi:10.1016/S1097-2765(03)00144-8

Bu, P., Evrard, Y.A., Lozano, G., Dent, S.Y.R., 2007. Loss of Gcn5 Acetyltransferase Activity Leads to Neural Tube Closure Defects and Exencephaly in Mouse Embryos. Mol. Cell. Biol. 27, 3405–3416. doi:10.1128/MCB.00066-07

Buratowski, S., Hahn, S., Guarente, L., Sharp, P.A., 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. Cell 56, 549–561. doi:10.1016/0092-8674(89)90578-3

Burke, T.L., Miller, J.L., Grant, P.A., 2013. Direct Inhibition of Gcn5 Protein Catalytic Activity by Polyglutamine-expanded Ataxin-7. J. Biol. Chem. 288, 34266–34275. doi:10.1074/jbc.M113.487538

Carre, C., Szymczak, D., Pidoux, J., Antoniewski, C., 2005. The Histone H3 Acetylase dGcn5 Is a Key Player in Drosophila melanogaster Metamorphosis. Mol. Cell. Biol. 25, 8228–8238. doi:10.1128/MCB.25.18.8228-8238.2005

Chen, Z., Manley, J.L., 2000. Robust mRNA Transcription in Chicken DT40 Cells Depleted of TAFII31 Suggests Both Functional Degeneracy and Evolutionary Divergence. Mol. Cell. Biol. 20, 5064–5076. doi:10.1128/MCB.20.14.5064-5076.2000

Cheung, P., Tanner, K.G., Cheung, W.L., Sassone-Corsi, P., Denu, J.M., Allis, C.D., 2000. Synergistic Coupling of Histone H3 Phosphorylation and Acetylation in Response to Epidermal Growth Factor Stimulation. Mol. Cell 5, 905–915. doi:10.1016/S1097-2765(00)80256-7

Chicca, J.J., Auble, D.T., Pugh, B.F., 1998. Cloning and Biochemical Characterization of TAF-172, a Human Homolog of Yeast Mot1. Mol. Cell. Biol. 18, 1701–1710.

Chittuluru, J.R., Chaban, Y., Monnet-Saksouk, J., Carrozza, M.J., Sapountzi, V., Selleck, W., Huang, J., Utley, R.T., Cramet, M., Allard, S., Cai, G., Workman, J.L., Fried, M.G., Tan, S., Côté, J., Asturias, F.J., 2011. Structure and nucleosome interaction of the yeast NuA4 and Piccolo-NuA4 histone acetyltransferase complexes. Nat. Struct. X00026 Mol. Biol. 18, 1196–1203. doi:10.1038/nsmb.2128

Chung, W.-H., Craighead, J.L., Chang, W.-H., Ezeokonkwo, C., Bareket-Samish, A., Kornberg, R.D., Asturias, F.J., 2003. RNA Polymerase II/TFIIF Structure and Conserved Organization of the Initiation Complex. Mol. Cell 12, 1003–1013. doi:10.1016/S1097-2765(03)00387-3

Ciurciu, A., Komonyi, O., Pankotai, T., Boros, I.M., 2006. The Drosophila Histone Acetyltransferase Gcn5 and Transcriptional Adaptor Ada2a Are Involved in Nucleosomal Histone H4 Acetylation. Mol. Cell. Biol. 26, 9413–9423. doi:10.1128/MCB.01401-06

Clark-Adams, C.D., Norris, D., Osley, M.A., Fassler, J.S., Winston, F., 1988. Changes in histone gene dosage alter transcription in yeast. Genes Dev. 2, 150–159. doi:10.1101/gad.2.2.150

Cler, E., Papai, G., Schultz, P., Davidson, I., 2009. Recent advances in understanding the structure and function of general transcription factor TFIID. Cell. Mol. Life Sci. 66, 2123–2134. doi:10.1007/s00018-009-0009-3

Coleman, R.A., Pugh, B.F., 1995. Evidence for functional binding and stable sliding of the TATA binding protein on nonspecific DNA. J. Biol. Chem. 270, 13850–13859.

Coleman, R.A., Taggart, A.K., Benjamin, L.R., Pugh, B.F., 1995. Dimerization of the TATA binding protein. J. Biol. Chem. 270, 13842–13849.

Conaway, R.C., Conaway, J.W., 2011. Function and Regulation of the Mediator Complex. Curr. Opin. Genet. Dev. 21, 225–230. doi:10.1016/j.gde.2011.01.013

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., Chambon, P., 1980. Promoter sequences of eukaryotic protein-coding genes. Science 209, 1406–1414.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., Kornberg, R.D., 2000. Architecture of RNA Polymerase II and Implications for the Transcription Mechanism. Science 288, 640–649. doi:10.1126/science.288.5466.640

Cramer, P., Bushnell, D.A., Kornberg, R.D., 2001. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. Science 292, 1863–1876. doi:10.1126/science.1059493

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. J. Biol. Chem. 279, 1867–1871. doi:10.1074/jbc.C300494200

David, G., Abbas, N., Stevanin, G., Dürr, A., Yvert, G., Cancel, G., Weber, C., Imbert, G., Saudou, F., Antoniou, E., Drabkin, H., Gemmill, R., Giunti, P., Benomar, A., Wood, N., Ruberg, M., Agid, Y., Mandel, J.-L., Brice, A., 1997. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. Nat. Genet. 17, 65–70. doi:10.1038/ng0997-65

Davidson, I., 2003. The genetics of TBP and TBP-related factors. Trends Biochem. Sci. 28, 391–398. doi:10.1016/S0968-0004(03)00117-8

Davis, J.A., Takagi, Y., Kornberg, R.D., Asturias, F.J., 2002. Structure of the Yeast RNA Polymerase II Holoenzyme: Mediator Conformation and Polymerase Interaction. Mol. Cell 10, 409–415. doi:10.1016/S1097-2765(02)00598-1

De St. Groth, S.F., Webster, R.G., Datyner, A., 1963. Two new staining procedures for quantitative estimation of proteins on electrophoretic strips. Biochim. Biophys. Acta 71, 377–391. doi:10.1016/0006-3002(63)91092-8

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., Zhou, M.-M., Zhou, M.-M., 1999. Structure and ligand of a histone acetyltransferase bromodomain. Nature 399, 491–496. doi:10.1038/20974

Dion, V., Coulombe, B., 2003. Interactions of a DNA-bound transcriptional activator with the TBP-TFIIA-TFIIB-promoter quaternary complex. J. Biol. Chem. doi:10.1074/jbc.M211938200

Dover, J., Schneider, J., Tawiah-Boateng, M.A., Wood, A., Dean, K., Johnston, M., Shilatifard, A., 2002. Methylation of Histone H3 by COMPASS Requires Ubiquitination of Histone H2B by Rad6. J. Biol. Chem. 277, 28368–28371.

Drané, P., Compe, E., Catez, P., Chymkowitch, P., Egly, J.-M., 2004. Selective Regulation of Vitamin D Receptor-Responsive Genes by TFIIH. Mol. Cell 16, 187–197. doi:10.1016/j.molcel.2004.10.007

Durant, M., Pugh, B.F., 2006. Genome-wide relationships between TAF1 and histone acetyltransferases in Saccharomyces cerevisiae. Mol. Cell. Biol. 26, 2791–2802. doi:10.1128/MCB.26.7.2791-2802.2006

Dvir, A., Conaway, J.W., Conaway, R.C., 2001. Mechanism of transcription initiation and promoter escape by RNA polymerase II. Curr. Opin. Genet. Dev. 11, 209–214. doi:10.1016/S0959-437X(00)00181-7

Edmondson, D.G., Davie, J.K., Zhou, J., Mirnikjoo, B., Tatchell, K., Dent, S.Y.R., 2002. Site-specific Loss of Acetylation upon Phosphorylation of Histone H3. J. Biol. Chem. 277, 29496–29502. doi:10.1074/jbc.M200651200

Edwards, A.M., Kane, C.M., Young, R.A., Kornberg, R.D., 1991. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. J. Biol. Chem. 266, 71–75.

Egly, J.-M., Coin, F., 2011. A history of TFIIH: Two decades of molecular biology on a pivotal transcription/repair factor. DNA Repair 10, 714–721. doi:10.1016/j.dnarep.2011.04.021

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 Dev. 6, 1319–1331.

Eisenmann, D.M., Chapon, C., Roberts, S.M., Dollard, C., Winston, F., 1994. The Saccharomyces Cerevisiae Spt8 Gene Encodes a Very Acidic Protein That Is Functionally Related to Spt3 and Tata-Binding Protein. Genetics 137, 647–657.

Eisenmann, D.M., Dollard, C., Winston, F., 1989. SPT15, the gene encoding the yeast TATA binding factor TFIID, is required for normal transcription initiation in vivo. Cell 58, 1183–1191. doi:10.1016/0092-8674(89)90516-3

Elmlund, H., Baraznenok, V., Lindahl, M., Samuelsen, C.O., Koeck, P.J.B., Holmberg, S., Hebert, H., Gustafsson, C.M., 2006. The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. Proc. Natl. Acad. Sci. U. S. A. 103, 15788–15793. doi:10.1073/pnas.0607483103

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., Werner, M., 2008. Mediator-Dependent Recruitment of TFIIH Modules in Preinitiation Complex. Mol. Cell 31, 337–346. doi:10.1016/j.molcel.2008.06.021

Ezhkova, E., Tansey, W.P., 2004. Proteasomal ATPases Link Ubiquitylation of Histone H2B to Methylation of Histone H3. Mol. Cell 13, 435–442. doi:10.1016/S1097-2765(04)00026-7

Fang, S.M., Burton, Z.F., 1996. RNA Polymerase II-associated Protein (RAP) 74 Binds Transcription Factor (TF) IIB and Blocks TFIIB-RAP30 Binding. J. Biol. Chem. 271, 11703–11709.

Felsenfeld, G., Groudine, M., 2003. Controlling the double helix. Nature 421, 448–453. doi:10.1038/nature01411

Fernández, I.S., Bai, X.-C., Hussain, T., Kelley, A.C., Lorsch, J.R., Ramakrishnan, V., Scheres, S.H.W., 2013. Molecular Architecture of a Eukaryotic Translational Initiation Complex. Science 342, 1240585. doi:10.1126/science.1240585

Fishburn, J., Mohibullah, N., Hahn, S., 2005. Function of a Eukaryotic Transcription Activator during the Transcription Cycle. Mol. Cell 18, 369–378. doi:10.1016/j.molcel.2005.03.029

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z.F., Reinberg, D., 1991. The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc. Natl. Acad. Sci. U. S. A. 88, 9999–10003.

Forget, D., Langelier, M.-F., Thérien, C., Trinh, V., Coulombe, B., 2004. Photo-Cross-Linking of a Purified Preinitiation Complex Reveals Central Roles for the RNA Polymerase II Mobile Clamp and TFIIE in Initiation Mechanisms. Mol. Cell. Biol. 24, 1122–1131. doi:10.1128/MCB.24.3.1122-1131.2004

Freiman, R.N., 2009. Specific variants of general transcription factors regulate germ cell development in diverse organisms. Biochim. Biophys. Acta 1789, 161–166. doi:10.1016/j.bbagrm.2009.01.005

Gamper, A.M., Kim, J., Roeder, R.G., 2009. The STAGA Subunit ADA2b Is an Important Regulator of Human GCN5 Catalysis. Mol. Cell. Biol. 29, 266–280. doi:10.1128/MCB.00315-08

Gangloff, Y.-G., Sanders, S.L., Romier, C., Kirschner, D., Weil, P.A., Tora, L., Davidson, I., 2001. Histone Folds Mediate Selective Heterodimerization of Yeast TAFII25 with TFIID Components yTAFII47 and yTAFII65 and with SAGA Component ySPT7. Mol. Cell. Biol. 21, 1841–1853. doi:10.1128/MCB.21.5.1841-1853.2001

Gangloff, Y.-G., Werten, S., Romier, C., Carre, L., Poch, O., Moras, D., Davidson, I., 2000. The Human TFIID Components TAFII135 and TAFII20 and the Yeast SAGA Components ADA1 and TAFII68 Heterodimerize to Form Histone-Like Pairs. Mol. Cell. Biol. 20, 340–351.

Gansheroff, L.J., Dollard, C., Tan, P., Winston, F., 1995. The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein important for transcription in vivo. Genetics 139, 523–536.

Garbett, K.A., Tripathi, M.K., Cencki, B., Layer, J.H., Weil, P.A., 2007. Yeast TFIID Serves as a Coactivator for Rap1p by Direct Protein-Protein Interaction. Mol. Cell. Biol. 27, 297–311. doi:10.1128/MCB.01558-06

García-Oliver, E., García-Molinero, V., Rodríguez-Navarro, S., 2012. mRNA export and gene expression: The SAGA–TREX-2 connection. Biochim. Biophys. Acta BBA - Gene Regul. Mech. 1819, 555–565. doi:10.1016/j.bbagrm.2011.11.011

Gavin, A.-C., Bösche, 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., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.-A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., Superti-Furga, G., 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–147. doi:10.1038/415141a

Gazdag, E., Santenard, A., Ziegler-Birling, C., Altobelli, G., Poch, O., Tora, L., Torres-Padilla, M.-E., 2009. TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. Genes Dev. 23, 2210–2223. doi:10.1101/gad.535209

Geiger, J.H., Hahn, S., Lee, S., Sigler, P.B., 1996. Crystal structure of the yeast TFIIA/TBP/DNA complex. Science 272, 830–836.

Geisberg, J.V., Holstege, F.C., Young, R.A., Struhl, K., 2001. Yeast NC2 Associates with the RNA Polymerase II Preinitiation Complex and Selectively Affects Transcription In Vivo. Mol. Cell. Biol. 21, 2736–2742. doi:10.1128/MCB.21.8.2736-2742.2001

Georgakopoulos, T., Thireos, G., 1992. Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. EMBO J. 11, 4145–4152.

Georgieva, S., Kirschner, D.B., Jagla, T., Nabirochkina, E., Hanke, S., Schenkel, H., de Lorenzo, C., Sinha, P., Jagla, K., Mechler, B., Tora, L., 2000. Two Novel Drosophila TAFIIs Have Homology with Human TAFII30 and Are Differentially Regulated during Development. Mol. Cell. Biol. 20, 1639–1648.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., Weissman, J.S., 2003. Global analysis of protein expression in yeast. Nature 425, 737–741. doi:10.1038/nature02046

Ghazy, M.A., Brodie, S.A., Ammerman, M.L., Ziegler, L.M., Ponticelli, A.S., 2004. Amino Acid Substitutions in Yeast TFIIF Confer Upstream Shifts in Transcription Initiation and Altered Interaction with RNA Polymerase II. Mol. Cell. Biol. 24, 10975–10985. doi:10.1128/MCB.24.24.10975-10985.2004

Gibbons, B.J., Brignole, E.J., Azubel, M., Murakami, K., Voss, N.R., Bushnell, D.A., Asturias, F.J., Kornberg, R.D., 2012. Subunit architecture of general transcription factor TFIIH. Proc. Natl. Acad. Sci. 109, 1949–1954. doi:10.1073/pnas.1105266109

Gill, G., Ptashne, M., 1988. Negative effect of the transcriptional activator GAL4. Nature 334, 721–724. doi:10.1038/334721a0

Ginsburg, D.S., Govind, C.K., Hinnebusch, A.G., 2009. NuA4 Lysine Acetyltransferase Esa1 Is Targeted to Coding Regions and Stimulates Transcription Elongation with Gcn5. Mol. Cell. Biol. 29, 6473–6487. doi:10.1128/MCB.01033-09

Glozak, M.A., Sengupta, N., Zhang, X., Seto, E., 2005. Acetylation and deacetylation of non-histone proteins. Gene 363, 15–23. doi:10.1016/j.gene.2005.09.010

Goodrich, J.A., Tjian, R., 1994. Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. Cell 77, 145–156. doi:10.1016/0092-8674(94)90242-9

Goppelt, A., Meisterernst, M., 1996. Characterization of the Basal Inhibitor of Class II Transcription NC2 from Saccharomyces cerevisiae. Nucleic Acids Res. 24, 4450–4455. doi:10.1093/nar/24.22.4450

Govind, C., Zhang, F., Qiu, H., Hofmeyer, K., Hinnebusch, A., 2007. Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. Mol Cell 25, 31–42. doi:10.1016/j.molcel.2006.11.020

Grant, P., Eberharter, A., John, S., Cook, R., Turner, B., Workman, J., 1999. Expanded lysine acetylation specificity of Gcn5 in native complexes. J Biol Chem 274, 5895–5900. doi:10.1074/jbc.274.9.5895

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 Dev. 11, 1640–1650. doi:10.1101/gad.11.13.1640

Grant, P.A., Schieltz, D., Pray-Grant, M.G., Steger, D.J., Reese, J.C., Yates III, J.R., Workman, J.L., 1998a. A Subset of TAFIIs Are Integral Components of the SAGA Complex Required for Nucleosome Acetylation and Transcriptional Stimulation. Cell 94, 45–53. doi:10.1016/S0092-8674(00)81220-9

Grant, P.A., Schieltz, D., Pray-Grant, M.G., Yates III, J.R., Workman, J.L., 1998b. The ATM-Related Cofactor Tra1 Is a Component of the Purified SAGA Complex. Mol. Cell 2, 863–867. doi:10.1016/S1097-2765(00)80300-7

Green, M.R., 2000. TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. Trends Biochem. Sci. 25, 59–63. doi:10.1016/S0968-0004(99)01527-3

Grünberg, S., Hahn, S., 2013. Structural insights into transcription initiation by RNA polymerase II. Trends Biochem. Sci. 38, 603–611. doi:10.1016/j.tibs.2013.09.002

Guelman, S., Suganuma, T., Florens, L., Swanson, S.K., Kiesecker, C.L., Kusch, T., Anderson, S., Yates, J.R., Washburn, M.P., Abmayr, S.M., Workman, J.L., 2006. Host Cell Factor and an Uncharacterized SANT Domain Protein Are Stable Components of ATAC, a Novel dAda2A/dGcn5-Containing Histone Acetyltransferase Complex in Drosophila. Mol. Cell. Biol. 26, 871–882. doi:10.1128/MCB.26.3.871-882.2006

Gustafsson, C.M., Myers, L.C., Li, Y., Redd, M.J., Lui, M., Erdjument-Bromage, H., Tempst, P., Kornberg, R.D., 1997. Identification of Rox3 as a Component of Mediator and RNA Polymerase II Holoenzyme. J. Biol. Chem. 272, 48–50. doi:10.1074/jbc.272.1.48

Guzder, S.N., Sung, P., Bailly, V., Prakash, L., Prakash, S., 1994. RAD25 is a DMA helicase required for DNA repair and RNA polymerase II transcription. Nature 369, 578–581. doi:10.1038/369578a0

Hahn, S., 2004. Structure and mechanism of the RNA polymerase II transcription machinery. Nat. Struct. Mol. Biol. 11, 394–403. doi:10.1038/nsmb763

Hampsey, M., 1998. Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. Microbiol. Mol. Biol. Rev. 62, 465–503.

Han, M., Grunstein, M., 1988. Nucleosome loss activates yeast downstream promoters in vivo. Cell 55, 1137–1145. doi:10.1016/0092-8674(88)90258-9

Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J., Workman, J.L., 2002. Function and Selectivity of Bromodomains in Anchoring Chromatin-Modifying Complexes to Promoter Nucleosomes. Cell 111, 369–379. doi:10.1016/S0092-8674(02)01005-X

He, Y., Fang, J., Taatjes, D.J., Nogales, E., 2013. Structural visualization of key steps in human transcription initiation. Nature 495, 481–486. doi:10.1038/nature11991

Heery, D.M., Kalkhoven, E., Hoare, S., Parker, M.G., 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature 387, 733–736. doi:10.1038/42750

Helmlinger, D., Hardy, S., Abou-Sleymane, G., Eberlin, A., Bowman, A.B., Gansmuller, A., Picaud, S., Zoghbi, H.Y., Trottier, Y., Tora, L., Devys, D., 2006. Glutamine-Expanded Ataxin-7 Alters TFTC/STAGA Recruitment and Chromatin Structure Leading to Photoreceptor Dysfunction. PLoS Biol. 4. doi:10.1371/journal.pbio.0040067

Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguet, 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. Hum. Mol. Genet. 13, 1257–1265. doi:10.1093/hmg/ddh139

Henriques, T., Gilchrist, D.A., Nechaev, S., Bern, M., Muse, G.W., Burkholder, A., Fargo, D.C., Adelman, K., 2013. Stable Pausing by RNA Polymerase II Provides an Opportunity to Target and Integrate Regulatory Signals. Mol. Cell 52, 517–528. doi:10.1016/j.molcel.2013.10.001

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 Dev. 17, 2648–2663. doi:10.1101/gad.1144003

Herbig, E., Warfield, L., Fish, L., Fishburn, J., Knutson, B.A., Moorefield, B., Pacheco, D., Hahn, S., 2010. Mechanism of Mediator Recruitment by Tandem Gcn4 Activation Domains and Three Gal11 Activator-Binding Domains. Mol. Cell. Biol. 30, 2376–2390. doi:10.1128/MCB.01046-09

Hoang, T.V., Cavin, X., Schultz, P., Ritchie, D.W., 2013. gEMpicker: A Highly Parallel GPU-Accelerated Particle Picking Tool for Cryo-Electron Microscopy. BMC Struct. Biol. 13, 25. doi:10.1186/1472-6807-13-25

Holstege, F.C., van der Vliet, P.C., Timmers, H.T., 1996. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. EMBO J. 15, 1666–1677.

Hong, M., Fitzgerald, M.X., Harper, S., Luo, C., Speicher, D.W., Marmorstein, R., 2008. Structural Basis for Dimerization in DNA Recognition by Gal4. Struct. Lond. Engl. 1993 16, 1019–1026. doi:10.1016/j.str.2008.03.015

Horiuchi, J., Silverman, N., Pina, B., Marcus, G.A., Guarente, L., 1997. ADA1, a novel component of the ADA/GCN5 complex, has broader effects than GCN5, ADA2, or ADA3. Mol. Cell. Biol. 17, 3220–3228.

Hublitz, P., Albert, M., Peters, A.H.F.M., 2009. Mechanisms of transcriptional repression by histone lysine methylation. Int. J. Dev. Biol. 53, 335–354. doi:10.1387/ijdb.082717ph

Huisinga, K.L., Pugh, B.F., 2004. A Genome-Wide Housekeeping Role for TFIID and a Highly Regulated Stress-Related Role for SAGA in Saccharomyces cerevisiae. Mol. Cell 13, 573–585. doi:10.1016/S1097-2765(04)00087-5

Imoberdorf, R.M., Topalidou, I., Strubin, M., 2006. A Role for Gcn5-Mediated Global Histone Acetylation in Transcriptional Regulation. Mol. Cell. Biol. 26, 1610–1616. doi:10.1128/MCB.26.5.1610-1616.2006

Ingvarsdottir, K., Krogan, N.J., Emre, N.C.T., Wyce, A., Thompson, N.J., Emili, A., Hughes, T.R., Greenblatt, J.F., Berger, S.L., 2005. H2B ubiquitin protease Ubp8 and Sgf11 constitute a discrete functional module within the Saccharomyces cerevisiae SAGA complex. Mol. Cell. Biol. 25, 1162–1172. doi:10.1128/MCB.25.3.1162-1172.2005

Ito, S., Kuraoka, I., Chymkowitch, P., Compe, E., Takedachi, A., Ishigami, C., Coin, F., Egly, J.-M., Tanaka, K., 2007. XPG Stabilizes TFIIH, Allowing Transactivation of Nuclear Receptors: Implications for Cockayne Syndrome in XP-G/CS Patients. Mol. Cell 26, 231–243. doi:10.1016/j.molcel.2007.03.013

Izzo, A., Schneider, R., 2010. Chatting histone modifications in mammals. Brief. Funct. Genomics 9, 429–443. doi:10.1093/bfgp/elq024

Jackson-Fisher, A.J., Chitikila, C., Mitra, M., Pugh, B.F., 1999. A Role for TBP Dimerization in Preventing Unregulated Gene Expression. Mol. Cell 3, 717–727. doi:10.1016/S1097-2765(01)80004-6

Jacobson, R.H., Ladurner, A.G., King, D.S., Tjian, R., 2000. Structure and Function of a Human TAFII250 Double Bromodomain Module. Science 288, 1422–1425. doi:10.1126/science.288.5470.1422

Jeong, C.J., Yang, S.H., Xie, Y., Zhang, L., Johnston, S.A., Kodadek, T., 2001. Evidence that Gal11 protein is a target of the Gal4 activation domain in the mediator. Biochemistry (Mosc.) 40, 9421–9427.

Juven-Gershon, T., Kadonaga, J.T., 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev. Biol. 339, 225–229. doi:10.1016/j.ydbio.2009.08.009

Kamada, K., Shu, F., Chen, H., Malik, S., Stelzer, G., Roeder, R.G., Meisterernst, M., Burley, S.K., 2001. Crystal Structure of Negative Cofactor 2 Recognizing the TBP-DNA Transcription Complex. Cell 106, 71–81. doi:10.1016/S0092-8674(01)00417-2

Kandiah, E., Trowitzsch, S., Gupta, K., Haffke, M., Berger, I., 2014. More pieces to the puzzle: recent structural insights into class II transcription initiation. Curr. Opin. Struct. Biol. 24, 91–97. doi:10.1016/j.sbi.2013.12.005

Kang, J.S., Kim, S.H., Hwang, M.S., Han, S.J., Lee, Y.C., Kim, Y.-J., 2001. The Structural and Functional Organization of the Yeast Mediator Complex. J. Biol. Chem. 276, 42003–42010. doi:10.1074/jbc.M105961200

Kao, C.-F., Hillyer, C., Tsukuda, T., Henry, K., Berger, S., Osley, M.A., 2004. Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. Genes Dev. 18, 184–195. doi:10.1101/gad.1149604

Kato, K., Makino, Y., Kishimoto, T., Yamauchi, J., Kato, S., Muramatsu, M., Tamura, T., 1994. Multimerization of the mouse TATA-binding protein (TBP) driven by its C-terminal conserved domain. Nucleic Acids Res. 22, 1179–1185.

Kelleher, R.J., 3rd, Flanagan, P.M., Kornberg, R.D., 1990. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. Cell 61, 1209–1215.

Keogh, M.-C., Mennella, T.A., Sawa, C., Berthelet, S., Krogan, N.J., Wolek, A., Podolny, V., Carpenter, L.R., Greenblatt, J.F., Baetz, K., Buratowski, S., 2006.

The Saccharomyces cerevisiae histone H2A variant Htz1 is acetylated by NuA4. Genes Dev. 20, 660–665. doi:10.1101/gad.1388106

Keriel, A., Stary, A., Sarasin, A., Rochette-Egly, C., Egly, J.-M., 2002. XPD Mutations Prevent TFIIH-Dependent Transactivation by Nuclear Receptors and Phosphorylation of RARα. Cell 109, 125–135. doi:10.1016/S0092-8674(02)00692-X

Kikuchi, H., Takami, Y., Nakayama, T., 2005. GCN5: a supervisor in all-inclusive control of vertebrate cell cycle progression through transcription regulation of various cell cycle-related genes. Gene 347, 83–97. doi:10.1016/j.gene.2004.12.007

Kim, J.L., Nikolov, D.B., Burley, S.K., 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature 365, 520–527. doi:10.1038/365520a0

Kim, T.-K., Ebright, R.H., Reinberg, D., 2000. Mechanism of ATP-Dependent Promoter Melting by Transcription Factor IIH. Science 288, 1418–1421. doi:10.1126/science.288.5470.1418

Kim, Y., Geiger, J.H., Hahn, S., Sigler, P.B., 1993. Crystal structure of a yeast TBP/TATA-box complex. Nature 365, 512–520. doi:10.1038/365512a0

Kim, Y.J., Björklund, S., Li, Y., Sayre, M.H., Kornberg, R.D., 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell 77, 599–608.

Kizilyaprak, C., Spehner, D., Devys, D., Schultz, P., 2011. The linker histone H1C contributes to the SCA7 nuclear phenotype. Nucl. Austin Tex 2, 444–454. doi:10.4161/nucl.2.5.17843

Knutson, B.A., Hahn, S., 2011. Domains of Tra1 Important for Activator Recruitment and Transcription Coactivator Functions of SAGA and NuA4 Complexes. Mol. Cell. Biol. 31, 818–831. doi: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. Mol. Biol. Cell 17, 4228–4236. doi: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. Nat. Cell Biol. 10, 707–715. doi:10.1038/ncb1733

Kohler, A., Zimmerman, E., Schneider, M., Hurt, E., Zheng, N., 2010. Structural Basis for Assembly and Activation of the Hetero-tetrameric SAGA Histone H2B Deubiquitinase Module. Cell 141, 606–617. doi:10.1016/j.cell.2010.04.026

Köhler, A., Zimmerman, E., Schneider, M., Hurt, E., Zheng, N., 2010. Structural Basis for Assembly and Activation of the Heterotetrameric SAGA Histone H2B Deubiquitinase Module. Cell 141, 606–617. doi:10.1016/j.cell.2010.04.026

Kohler, C., Bonnet, J., Stierle, M., Romier, C., Devys, D., Kieffer, B., 2014. DNA binding by Sgf11 affects histone H2B deubiquitination by SAGA. J. Biol. Chem. doi:10.1074/jbc.M113.500868

Kokubo, T., Gong, D.W., Yamashita, S., Horikoshi, M., Roeder, R.G., Nakatani, Y., 1993. Drosophila 230-kD TFIID subunit, a functional homolog of the human cell cycle gene product, negatively regulates DNA binding of the TATA box-binding subunit of TFIID. Genes Dev. 7, 1033–1046.

Kokubo, T., Swanson, M.J., Nishikawa, J., Hinnebusch, A.G., Nakatani, Y., 1998. The Yeast TAF145 Inhibitory Domain and TFIIA Competitively Bind to TATA-Binding Protein. Mol. Cell. Biol. 18, 1003–1012.

Koleske, A.J., Young, R.A., 1994. An RNA polymerase II holoenzyme responsive to activators. Nature 368, 466–469. doi:10.1038/368466a0

Kolthur-Seetharam, U., Martianov, I., Davidson, I., 2008. Specialization of the general transcriptional machinery in male germ cells. Cell Cycle Georget. Tex 7, 3493–3498.

Kopytova, D.V., Krasnov, A.N., Kopantceva, M.R., Nabirochkina, E.N., Nikolenko, J.V., Maksimenko, O., Kurshakova, M.M., Lebedeva, L.A., Yerokhin, M.M., Simonova, O.B., Korochkin, L.I., Tora, L., Georgiev, P.G., Georgieva, S.G., 2006. Two Isoforms of Drosophila TRF2 Are Involved in Embryonic Development, Premeiotic Chromatin Condensation, and Proper Differentiation of Germ Cells of Both Sexes. Mol. Cell. Biol. 26, 7492–7505. doi:10.1128/MCB.00349-06

Kostrewa, D., Zeller, M.E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., Cramer, P., 2009. RNA polymerase II–TFIIB structure and mechanism of transcription initiation. Nature 462, 323–330. doi:10.1038/nature08548

Kraemer, S.M., Ranallo, R.T., Ogg, R.C., Stargell, L.A., 2001. TFIIA Interacts with TFIID via Association with TATA-Binding Protein and TAF40. Mol. Cell. Biol. 21, 1737–1746. doi:10.1128/MCB.21.5.1737-1746.2001

Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., Tora, L., 2011. SAGA and ATAC Histone Acetyl Transferase Complexes Regulate Distinct Sets of Genes and ATAC Defines a Class of p300-Independent Enhancers. Mol. Cell 44, 410–423. doi:10.1016/j.molcel.2011.08.037

Kuldell, N.H., Buratowski, S., 1997. Genetic analysis of the large subunit of yeast transcription factor IIE reveals two regions with distinct functions. Mol. Cell. Biol. 17, 5288–5298.

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, 269–272. doi:10.1038/383269a0

Kuo, M.-H., vom Baur, E., Struhl, K., Allis, C.D., 2000. Gcn4 Activator Targets Gcn5 Histone Acetyltransferase to Specific Promoters Independently of Transcription. Mol. Cell 6, 1309–1320. doi:10.1016/S1097-2765(00)00129-5

Kurabe, N., Katagiri, K., Komiya, Y., Ito, R., Sugiyama, A., Kawasaki, Y., Tashiro, F., 2007. Deregulated expression of a novel component of TFTC/STAGA histone acetyltransferase complexes, rat SGF29, in hepatocellular carcinoma: possible implication for the oncogenic potential of c-Myc. Oncogene 26, 5626–5634. doi:10.1038/sj.onc.1210349

Kurshakova, M.M., Krasnov, A.N., Kopytova, D.V., Shidlovskii, Y.V., Nikolenko, J.V., Nabirochkina, E.N., Spehner, D., Schultz, P., Tora, L., Georgieva, S.G., 2007. SAGA and a novel Drosophila export complex anchor efficient transcription and mRNA export to NPC. EMBO J. 26, 4956–4965. doi:10.1038/sj.emboj.7601901

Kusch, T., Guelman, S., Abmayr, S.M., Workman, J.L., 2003. Two Drosophila Ada2 Homologues Function in Different Multiprotein Complexes. Mol. Cell. Biol. 23, 3305–3319. doi:10.1128/MCB.23.9.3305-3319.2003

Kutach, A.K., Kadonaga, J.T., 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol. Cell. Biol. 20, 4754–4764.

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, 2007–2017. doi:10.1534/genetics.107.081976
Larivière, L., Geiger, S., Hoeppner, S., Röther, S., Sträßer, K., Cramer, P., 2006. Structure and TBP binding of the Mediator head subcomplex Med8–Med18–Med20. Nat. Struct. Mol. Biol. 13, 895–901. doi:10.1038/nsmb1143
Larivière, L., Seizl, M., Cramer, P., 2012. A structural perspective on Mediator function. Curr. Opin. Cell Biol. 24, 305–313. doi:10.1016/j.ceb.2012.01.007
Larschan, E., Winston, F., 2001. The S. cerevisiae SAGA complex functions in vivo as a coactivator for transcriptional activation by Gal4. Genes Dev. 15, 1946–1956. doi:10.1101/gad.911501
Larschan, E., Winston, F., 2005. The Saccharomyces cerevisiae Srb8-Srb11 Complex Functions with the SAGA Complex during Gal4-Activated Transcription. Mol. Cell. Biol. 25, 114–123. doi:10.1128/MCB.25.1.114-123.2005
Layer, J.H., Miller, S.G., Weil, P.A., 2010. Direct Transactivator-Transcription Factor IID (TFIID) Contacts Drive Yeast Ribosomal Protein Gene Transcription. J. Biol. Chem. 285, 15489–15499. doi:10.1074/jbc.M110.104810
Lee, D., Ezhkova, E., Li, B., Pattenden, S.G., Tansey, W.P., Workman, J.L., 2005. The Proteasome Regulatory Particle Alters the SAGA Coactivator to Enhance Its Interactions with Transcriptional Activators. Cell 123, 423–436. doi:10.1016/j.cell.2005.08.015
Lee, K.K., Florens, L., Swanson, S.K., Washburn, M.P., Workman, J.L., 2005. The Deubiquitylation Activity of Ubp8 Is Dependent upon Sgf11 and Its Association with the SAGA Complex. Mol. Cell. Biol. 25, 1173–1182. doi:10.1128/MCB.25.3.1173-1182.2005
Lee, K.K., Sardiu, M.E., Swanson, S.K., Gilmore, J.M., Torok, M., Grant, P.A., Florens, L., Workman, J.L., Washburn, M.P., 2011. Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. Mol. Syst. Biol. 7, 503. doi:10.1038/msb.2011.40
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, 2. doi:10.1186/1756-8935-2-2
Lee, T.I., Young, R.A., 2000. Transcription of Eukaryotic Protein-Coding Genes. Annu. Rev. Genet. 34, 77–137. doi:10.1146/annurev.genet.34.1.77
Leschziner, A., 2010. The orthogonal tilt reconstruction method. Methods Enzymol. 482, 237–262. doi:10.1016/S0076-6879(10)82010-5
Leurent, C., Sanders, S., Ruhlmann, C., Mallouh, V., Weil, P.A., Kirschner, D.B., Tora, L., Schultz, P., 2002. Mapping histone fold TAFs within yeast TFIID. EMBO J. 21, 3424–3433. doi:10.1093/emboj/cdf342
Leurent, C., Sanders, S.L., Demeny, M.A., Garbett, K.A., Ruhlmann, C., Weil, P.A., Tora, L., Schultz, P., 2004. Mapping key functional sites within yeast TFIID. EMBO J. 23, 719–727. doi:10.1038/sj.emboj.7600111
Li, B., Carey, M., Workman, J.L., 2007. The Role of Chromatin during Transcription. Cell 128, 707–719. doi:10.1016/j.cell.2007.01.015
Lieb, J.D., Liu, X., Botstein, D., Brown, P.O., 2001. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat. Genet. 28, 327–334. doi:10.1038/ng569
Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., Kadonaga, J.T., 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev. 18, 1606–1617. doi:10.1101/gad.1193404

Lim, S., Kwak, J., Kim, M., Lee, D., 2013. Separation of a functional deubiquitylating module from the SAGA complex by the proteasome regulatory particle. Nat. Commun. 4. doi:10.1038/ncomms3641

Lin, L., Chamberlain, L., Zhu, L.J., Green, M.R., 2012. Analysis of Gal4-directed transcription activation using Tra1 mutants selectively defective for interaction with Gal4. Proc. Natl. Acad. Sci. 109, 1997–2002. doi:10.1073/pnas.1116340109

Lin, Y.C., Choi, W.S., Gralla, J.D., 2005. TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. Nat. Struct. Mol. Biol. 12, 603–607. doi:10.1038/nsmb949

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., Yates, J.R., 1999. Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. 17, 676–682. doi:10.1038/10890

Liu, D., Ishima, R., Tong, K.I., Bagby, S., Kokubo, T., Muhandiram, D.R., Kay, L.E., Nakatani, Y., Ikura, M., 1998. Solution Structure of a TBP–TAFII230 Complex: Protein Mimicry of the Minor Groove Surface of the TATA Box Unwound by TBP. Cell 94, 573–583. doi:10.1016/S0092-8674(00)81599-8

Liu, W.-L., Coleman, R.A., Ma, E., Grob, P., Yang, J.L., Zhang, Y., Dailey, G., Nogales, E., Tjian, R., 2009. Structures of three distinct activator–TFIID complexes. Genes Dev. 23, 1510–1521. doi:10.1101/gad.1790709

Liu, X., Tesfai, J., Evrard, Y.A., Dent, S.Y.R., Martinez, E., 2003. c-Myc Transformation Domain Recruits the Human STAGA Complex and Requires TRRAP and GCN5 Acetylase Activity for Transcription Activation. J. Biol. Chem. 278, 20405–20412. doi:10.1074/jbc.M211795200

Lohr, D., Venkov, P., Zlatanova, J., 1995. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J. 9, 777–787.

Lorch, Y., LaPointe, J.W., Kornberg, R.D., 1987. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell 49, 203–210. doi:10.1016/0092-8674(87)90561-7

Lu, H., Fisher, R.P., Bailey, P., Levine, A.J., 1997. The CDK7-cycH-p36 complex of transcription factor IIH phosphorylates p53, enhancing its sequence-specific DNA binding activity in vitro. Mol. Cell. Biol. 17, 5923–5934.

Ludtke, S.J., Baldwin, P.R., Chiu, W., 1999. EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions. J. Struct. Biol. 128, 82–97. doi:10.1006/jsbi.1999.4174

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 A resolution. Nature 389, 251–260. doi:10.1038/38444

Marcus, G.A., Horiuchi, J., Silverman, N., Guarente, L., 1996. ADA5/SPT20 links the ADA and SPT genes, which are involved in yeast transcription. Mol. Cell. Biol. 16, 3197–3205.

Marmorstein, R., Carey, M., Ptashne, M., Harrison, S.C., 1992. DNA recognition by GAL4: structure of a protein-DNA complex. Nature 356, 408–414. doi:10.1038/356408a0

Martianov, I., Brancorsini, S., Gansmuller, A., Parvinen, M., Davidson, I., Sassone-Corsi, P., 2002. Distinct functions of TBP and TLF/TRF2 during spermatogenesis: requirement of TLF for heterochromatic chromocenter formation in haploid round spermatids. Dev. Camb. Engl. 129, 945–955.

Martinez, E., Kundu, T.K., Fu, J., Roeder, R.G., 1998. A Human SPT3-TAFII31-GCN5-L Acetylase Complex Distinct from Transcription Factor IID. J. Biol. Chem. 273, 23781–23785. doi:10.1074/jbc.273.37.23781

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., Pugh, B.F., 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res. 18, 1073–1083. doi:10.1101/gr.078261.108

Maxon, M.E., Goodrich, J.A., Tjian, R., 1994. Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes Dev. 8, 515–524. doi:10.1101/gad.8.5.515

McCullough, S.D., Grant, P.A., 2010. Histone acetylation, acetyltransferases, and Ataxia—Alteration of histone acetylation and chromatin dynamics is implicated in the pathogenesis of polyglutamine expansion disorders. Adv. Protein Chem. Struct. Biol. 79, 165–203. doi:10.1016/S1876-1623(10)79005-2

McMahon, S.B., Van Buskirk, H.A., Dugan, K.A., Copeland, T.D., Cole, M.D., 1998. The Novel ATM-Related Protein TRRAP Is an Essential Cofactor for the c-Myc and E2F Oncoproteins. Cell 94, 363–374. doi:10.1016/S0092-8674(00)81479-8

McMahon, S.J., Pray-Grant, M.G., Schieltz, D., Yates, J.R., Grant, P.A., 2005. Polyglutamine-expanded spinocerebellar ataxia-7 protein disrupts normal SAGA and SLIK histone acetyltransferase activity. Proc. Natl. Acad. Sci. U. S. A. 102, 8478–8482. doi:10.1073/pnas.0503493102

Meka, H., Werner, F., Cordell, S.C., Onesti, S., Brick, P., 2005. Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. Nucleic Acids Res. 33, 6435–6444. doi:10.1093/nar/gki945

Melcher, K., Johnston, S.A., 1995. GAL4 interacts with TATA-binding protein and coactivators. Mol. Cell. Biol. 15, 2839–2848.

Melcher, K., Xu, H.E., 2001. Gal80-Gal80 interaction on adjacent Gal4p binding sites is required for complete GAL gene repression. EMBO J. 20, 841–851. doi:10.1093/emboj/20.4.841

Mindell, J.A., Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 142, 334–347.

Mischerikow, N., Spedale, G., Altelaar, A.F.M., Timmers, H.T.M., Pijnappel, W.W.M.P., Heck, A.J.R., 2009. In-Depth Profiling of Post-Translational Modifications on the Related Transcription Factor Complexes TFIID and SAGA. J. Proteome Res. 8, 5020–5030. doi:10.1021/pr900449e

Mizzen, C.A., Yang, X.J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T., Nakatani, Y., Allis, C.D., 1996. The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell 87, 1261–1270.

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 Dev. 22, 2994–3006. doi:10.1101/gad.1724408

Morrison, A.J., Shen, X., 2006. Chromatin modifications in DNA repair. Results Probl. Cell Differ. 41, 109–125.

Muldrow, T.A., Campbell, A.M., Weil, P.A., Auble, D.T., 1999. MOT1 Can Activate Basal Transcription In Vitro by Regulating the Distribution of TATA Binding Protein between Promoter and Nonpromoter Sites. Mol. Cell. Biol. 19, 2835–2845.

Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B.J., Levitt, M., Kornberg,

R.D., 2013. Architecture of an RNA Polymerase II Transcription Pre-Initiation Complex. Science 342, 1238724. doi:10.1126/science.1238724

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., Adelman, K., 2007. RNA polymerase is poised for activation across the genome. Nat. Genet. 39, 1507–1511. doi:10.1038/ng.2007.21

Nagy, Z., Riss, A., Fujiyama, S., Krebs, A., Orpinell, M., Jansen, P., Cohen, A., Stunnenberg, H., Kato, S., Tora, L., 2010. The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. Cell Mol Life Sci 67, 611–628. doi:10.1007/s00018-009-0199-8

Nagy, Z., Riss, A., Fujiyama, S., Krebs, A., Orpinell, M., Jansen, P., Cohen, A., Stunnenberg, H.G., Kato, S., Tora, L., 2010. The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. Cell. Mol. Life Sci. 67, 611–628. doi:10.1007/s00018-009-0199-8

Nagy, Z., Tora, L., 2007. Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. Oncogene 26, 5341–5357. doi:10.1038/sj.onc.1210604

Neely, K.E., Hassan, A.H., Brown, C.E., Howe, L., Workman, J.L., 2002. Transcription activator interactions with multiple SWI/SNF subunits. Mol. Cell. Biol. 22, 1615–1625.

Nikolov, D.B., Chen, H., Halay, E.D., Hoffman, A., Roeder, R.G., Burley, S.K., 1996. Crystal structure of a human TATA box-binding protein/TATA element complex. Proc. Natl. Acad. Sci. 93, 4862–4867.

Nonet, M.L., Young, R.A., 1989. Intragenic and Extragenic Suppressors of Mutations in the Heptapeptide Repeat Domain of Saccharomyces Cerevisiae RNA Polymerase II. Genetics 123, 715–724.

Nourani, A., Utley, R.T., Allard, S., Cote, J., 2004. Recruitment of the NuA4 complex poises the PHO5 promoter for chromatin remodeling and activation. EMBO J. 23, 2597–2607. doi:10.1038/sj.emboj.7600230

Ogryzko, V.V., Kotani, T., Zhang, X., Schiltz, R.L., Howard, T., Yang, X.-J., Howard, B.H., Qin, J., Nakatani, Y., 1998. Histone-like TAFs within the PCAF Histone Acetylase Complex. Cell 94, 35–44. doi:10.1016/S0092-8674(00)81219-2

Okuda, M., Tanaka, A., Arai, Y., Satoh, M., Okamura, H., Nagadoi, A., Hanaoka, F., Ohkuma, Y., Nishimura, Y., 2004. A Novel Zinc Finger Structure in the Large Subunit of Human General Transcription Factor TFIIE. J. Biol. Chem. 279, 51395–51403. doi:10.1074/jbc.M404722200

Orlicky, S.M., Tran, P.T., Sayre, M.H., Edwards, A.M., 2001. Dissociable Rpb4-Rpb7 Subassembly of RNA Polymerase II Binds to Single-strand Nucleic Acid and Mediates a Post-recruitment Step in Transcription Initiation. J. Biol. Chem. 276, 10097–10102. doi:10.1074/jbc.M003165200

Orphanides, G., Lagrange, T., Reinberg, D., 1996. The general transcription factors of RNA polymerase II. Genes Dev. 10, 2657–2683. doi:10.1101/gad.10.21.2657

Palhan, V.B., Chen, S., Peng, G.-H., Tjernberg, A., Gamper, A.M., Fan, Y., Chait, B.T., Spada, A.R.L., Roeder, R.G., 2005. Polyglutamine-expanded ataxin-7 inhibits STAGA histone acetyltransferase activity to produce retinal degeneration. Proc. Natl. Acad. Sci. U. S. A. 102, 8472–8477. doi:10.1073/pnas.0503505102

Pankotai, T., Komonyi, O., Bodai, L., Ujfaludi, Z., Muratoglu, S., Ciurciu, A., Tora, L., Szabad, J., Boros, I., 2005. The homologous Drosophila transcriptional adaptors ADA2a and ADA2b are both required for normal development but have different functions. Mol. Cell. Biol. 25, 8215–8227. doi:10.1128/MCB.25.18.8215-8227.2005

Paoletti, A.C., Parmely, T.J., Tomomori-Sato, C., Sato, S., Zhu, D., Conaway, R.C., Conaway, J.W., Florens, L., Washburn, M.P., 2006. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. Proc. Natl. Acad. Sci. U. S. A. 103, 18928–18933. doi:10.1073/pnas.0606379103

Papai, G., Tripathi, M.K., Ruhlmann, C., Layer, J.H., Weil, P.A., Schultz, P., 2010. TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. Nature 465, 956–960. doi:10.1038/nature09080

Papai, G., Tripathi, M.K., Ruhlmann, C., Werten, S., Crucifix, C., Weil, P.A., Schultz, P., 2009. Mapping the Initiator Binding Taf2 Subunit in the Structure of Hydrated Yeast TFIID. Structure 17, 363–373. doi:10.1016/j.str.2009.01.006

Papai, G., Weil, P.A., Schultz, P., 2011. New insights into the function of transcription factor TFIID from recent structural studies. Curr. Opin. Genet. Dev. 21, 219–224. doi:10.1016/j.gde.2011.01.009

Park, J.M., Kim, H.-S., Han, S.J., Hwang, M.-S., Lee, Y.C., Kim, Y.-J., 2000. In Vivo Requirement of Activator-Specific Binding Targets of Mediator. Mol. Cell. Biol. 20, 8709–8719.

Pascual-García, P., Govind, C.K., Queralt, E., Cuenca-Bono, B., Llopis, A., Chavez, S., Hinnebusch, A.G., Rodríguez-Navarro, S., 2008. Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2. Genes Dev. 22, 2811–2822. doi:10.1101/gad.483308

Pei, Y., Hausmann, S., Ho, C.K., Schwer, B., Shuman, S., 2001. The Length, Phosphorylation State, and Primary Structure of the RNA Polymerase II Carboxyl-terminal Domain Dictate Interactions with mRNA Capping Enzymes. J. Biol. Chem. 276, 28075–28082. doi:10.1074/jbc.M102170200

Pereira, L.A., Knaap, J.A. van der, Boom, V. van den, Heuvel, F.A.J. van den, Timmers, H.T.M., 2001. TAFII170 Interacts with the Concave Surface of TATA-Binding Protein To Inhibit Its DNA Binding Activity. Mol. Cell. Biol. 21, 7523–7534. doi:10.1128/MCB.21.21.7523-7534.2001

Pijnappel, W.W.M.P., Timmers, H.T.M., 2008. Dubbing SAGA Unveils New Epigenetic Crosstalk. Mol. Cell 29, 152–154. doi:10.1016/j.molcel.2008.01.007

Pina, B., Berger, S., Marcus, G.A., Silverman, N., Agapite, J., Guarente, L., 1993. ADA3: a gene, identified by resistance to GAL4-VP16, with properties similar to and different from those of ADA2. Mol. Cell. Biol. 13, 5981–5989.

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. Mol. Cell. Biol. 24, 7249–7259. doi:10.1128/MCB.24.16.7249-7259.2004

Pray-Grant, M.G., Daniel, J.A., Schieltz, D., Yates, J.R., Grant, P.A., 2005. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature 433, 434–438. doi:10.1038/nature03242

Pray-Grant, M.G., Schieltz, D., McMahon, S.J., Wood, J.M., Kennedy, E.L., Cook, R.G., Workman, J.L., Yates III, J.R., Grant, P.A., 2002. The Novel SLIK Histone Acetyltransferase Complex Functions in the Yeast Retrograde Response Pathway. Mol. Cell. Biol. 22, 8774–8786. doi:10.1128/MCB.22.24.8774-8786.2002

Ptashne, M., Gann, A., 1997. Transcriptional activation by recruitment. Nature 386, 569–577. doi:10.1038/386569a0

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., Séraphin, B., 2001. The tandem affinity purification (TAP) method: a

general procedure of protein complex purification. Methods San Diego Calif 24, 218–229. doi:10.1006/meth.2001.1183

Puig, O., Rutz, B., Luukkonen, B.G.M., Kandels-Lewis, S., Bragado-Nilsson, E., Séraphin, B., 1998. New constructs and strategies for efficient PCR-based gene manipulations in yeast. Yeast 14, 1139–1146. doi:10.1002/(SICI)1097-0061(19980915)14:12<1139::AID-YEA306>3.0.CO;2-B

Qi, D., Larsson, J., Mannervik, M., 2004. Drosophila Ada2b is required for viability and normal histone H3 acetylation. Mol. Cell. Biol. 24, 8080–8089. doi:10.1128/MCB.24.18.8080-8089.2004

Qiu, H., Hu, C., Zhang, F., Hwang, G.J., Swanson, M.J., Boonchird, C., Hinnebusch, A.G., 2005. Interdependent Recruitment of SAGA and Srb Mediator by Transcriptional Activator Gcn4p. Mol. Cell. Biol. 25, 3461–3474. doi:10.1128/MCB.25.9.3461-3474.2005

Rani, P.G., Ranish, J.A., Hahn, S., 2004. RNA Polymerase II (Pol II)-TFIIF and Pol II-Mediator Complexes: the Major Stable Pol II Complexes and Their Activity in Transcription Initiation and Reinitiation. Mol. Cell. Biol. 24, 1709–1720. doi:10.1128/MCB.24.4.1709-1720.2004

Reeves, W.M., Hahn, S., 2005. Targets of the Gal4 Transcription Activator in Functional Transcription Complexes. Mol. Cell. Biol. 25, 9092–9102. doi:10.1128/MCB.25.20.9092-9102.2005

Reid, J.L., Iyer, V.R., Brown, P.O., Struhl, K., 2000. Coordinate Regulation of Yeast Ribosomal Protein Genes Is Associated with Targeted Recruitment of Esa1 Histone Acetylase. Mol. Cell 6, 1297–1307. doi:10.1016/S1097-2765(00)00128-3

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., Séraphin, B., 1999. A generic protein purification method for protein complex characterization and proteome exploration. Nat. Biotechnol. 17, 1030–1032. doi:10.1038/13732

Robert, F., Douziech, M., Forget, D., Egly, J.-M., Greenblatt, J., Burton, Z.F., Coulombe, B., 1998. Wrapping of Promoter DNA around the RNA Polymerase II Initiation Complex Induced by TFIIF. Mol. Cell 2, 341–351. doi:10.1016/S1097-2765(00)80278-6

Robert, F., Hardy, S., Nagy, Z., Baldeyron, C., Murr, R., Dery, U., Masson, J.-Y., Papadopoulo, D., Herceg, Z., Tora, L., 2006. The Transcriptional Histone Acetyltransferase Cofactor TRRAP Associates with the MRN Repair Complex and Plays a Role in DNA Double-Strand Break Repair. Mol. Cell. Biol. 26, 402–412. doi:10.1128/MCB.26.2.402-412.2006

Roberts, S.M., Winston, F., 1996. SPT20/ADA5 encodes a novel protein functionally related to the TATA-binding protein and important for transcription in Saccharomyces cerevisiae. Mol. Cell. Biol. 16, 3206–3213.

Robinson, M.M., Yatherajam, G., Ranallo, R.T., Bric, A., Paule, M.R., Stargell, L.A., 2005. Mapping and Functional Characterization of the TAF11 Interaction with TFIIA. Mol. Cell. Biol. 25, 945–957. doi:10.1128/MCB.25.3.945-957.2005

Robzyk, K., Recht, J., Osley, M.A., 2000. Rad6-dependent ubiquitination of histone H2B in yeast. Science 287, 501–504.

Rochette-Egly, C., Adam, S., Rossignol, M., Egly, J.-M., Chambon, P., 1997. Stimulation of RARα Activation Function AF-1 through Binding to the General Transcription Factor TFIIH and Phosphorylation by CDK7. Cell 90, 97–107. doi:10.1016/S0092-8674(00)80317-7

Rodríguez-Navarro, S., 2009. Insights into SAGA function during gene expression. EMBO Rep. 10, 843–850. doi:10.1038/embor.2009.168

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, 75–86.

Sainsbury, S., Niesser, J., Cramer, P., 2013. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature 493, 437–440. doi:10.1038/nature11715

Saleh, A., Schieltz, D., Ting, N., McMahon, S.B., Litchfield, D.W., Yates, J.R., Lees-Miller, S.P., Cole, M.D., Brandl, C.J., 1998. Tra1p Is a Component of the Yeast Ada·Spt Transcriptional Regulatory Complexes. J. Biol. Chem. 273, 26559–26565.

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, 1025–1029. doi:10.1126/science.1190049

Sanders, S., Jennings, J., Canutescu, A., Link, A., Weil, P., 2002. Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. Mol Cell Biol 22, 4723–4738. doi:10.1128/MCB.22.13.4723-4738.2002

Sanders, S.L., Garbett, K.A., Weil, P.A., 2002. Molecular Characterization of Saccharomyces cerevisiae TFIID. Mol. Cell. Biol. 22, 6000–6013. doi:10.1128/MCB.22.16.6000-6013.2002

Sarma, K., Reinberg, D., 2005. Histone variants meet their match. Nat. Rev. Mol. Cell Biol. 6, 139–149. doi:10.1038/nrm1567

Sauer, F., Fondell, J.D., Ohkuma, Y., Roeder, R.G., Jäckle, H., 1995. Control of transcription by Krüppel through interactions with TFIIB and TFIIEβ. Nature 375, 162–164. doi:10.1038/375162a0

Sauerwald, A., Sandin, S., Cristofari, G., Scheres, S.H.W., Lingner, J., Rhodes, D., 2013. Structure of Active, Dimeric Human Telomerase. Nat. Struct. Mol. Biol. 20, 454–460. doi:10.1038/nsmb.2530

Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J.H., Chambon, P., Egly, J.M., 1993. DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. Science 260, 58–63.

Scheres, S.H.W., 2012. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530. doi:10.1016/j.jsb.2012.09.006

Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D., Egly, J.M., 2000. Molecular Structure of Human TFIIH. Cell 102, 599–607. doi:10.1016/S0092-8674(00)00082-9

Seizl, M., Larivière, L., Pfaffeneder, T., Wenzeck, L., Cramer, P., 2011. Mediator head subcomplex Med11/22 contains a common helix bundle building block with a specific function in transcription initiation complex stabilization. Nucleic Acids Res. 39, 6291–6304. doi:10.1093/nar/gkr229

Selleck, W., Howley, R., Fang, Q., Podolny, V., Fried, M.G., Buratowski, S., Tan, S., 2001. A histone fold TAF octamer within the yeast TFIID transcriptional coactivator. Nat. Struct. Mol. Biol. 8, 695–700. doi:10.1038/90408

Sermwittayawong, D., Tan, S., 2006. SAGA binds TBP via its Spt8 subunit in competition with DNA: implications for TBP recruitment. EMBO J 25, 3791–3800. doi:10.1038/sj.emboj.7601265

Shilatifard, A., 2006. Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. Annu. Rev. Biochem. 75, 243–269. doi:10.1146/annurev.biochem.75.103004.142422

Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C., Werner, M., 2011. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. Science 331, 1451–1454. doi:10.1126/science.1200188

Spedale, G., Mischerikow, N., Heck, A.J.R., Timmers, H.T.M., Pijnappel, W.W.M.P., 2010. Identification of Pep4p as the Protease Responsible for Formation of the SAGA-related SLIK Protein Complex. J. Biol. Chem. 285, 22793–22799. doi:10.1074/jbc.M110.108787

Spedale, G., Timmers, H.T.M., Pijnappel, W.W.M.P., 2012. ATAC-king the complexity of SAGA during evolution. Genes Dev. 26, 527–541. doi:10.1101/gad.184705.111

Sterner, D.E., Belotserkovskaya, R., Berger, S.L., 2002. SALSA, a variant of yeast SAGA, contains truncated Spt7, which correlates with activated transcription. Proc. Natl. Acad. Sci. 99, 11622–11627. doi: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. Mol. Cell. Biol. 19, 86–98.

Struhl, K., Segal, E., 2013. Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. 20, 267–273. doi:10.1038/nsmb.2506

Svejstrup, J.Q., Vichi, P., Egly, J.M., 1996. The multiple roles of transcription/repair factor TFIIH. Trends Biochem. Sci. 21, 346–350.

Swanson, S.K., Florens, L., Washburn, M.P., 2009. Generation and analysis of multidimensional protein identification technology datasets. Methods Mol. Biol. Clifton NJ 492, 1–20. doi:10.1007/978-1-59745-493-3_1

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A.S., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., Florens, L., Seidel, C.W., Lin, C., Smith, E.R., Shilatifard, A., Conaway, R.C., Conaway, J.W., 2011. Human Mediator Subunit MED26 Functions as a Docking Site for Transcription Elongation Factors. Cell 146, 92–104. doi:10.1016/j.cell.2011.06.005

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., Ludtke, S.J., 2007. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 38–46. doi:10.1016/j.jsb.2006.05.009

Thomas, M.C., Chiang, C.-M., 2006. The General Transcription Machinery and General Cofactors [WWW Document]. URL http://informahealthcare.com/doi/abs/10.1080/10409230600648736 (accessed 12.10.13).

Thompson, C.M., Koleske, A.J., Chao, D.M., Young, R.A., 1993. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. Cell 73, 1361–1375. doi:10.1016/0092-8674(93)90362-T

Tora, L., 2002. A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. Genes Dev. 16, 673–675. doi:10.1101/gad.976402

Traven, A., Jelicic, B., Sopta, M., 2006. Yeast Gal4: a transcriptional paradigm revisited. EMBO Rep. 7, 496–499. doi:10.1038/sj.embor.7400679

Treviño, L.S., Weigel, N.L., 2013. Phosphorylation: a fundamental regulator of steroid receptor action. Trends Endocrinol. Metab. TEM 24, 515–524. doi:10.1016/j.tem.2013.05.008

Van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., Schatz, M., 1996. A New Generation of the IMAGIC Image Processing System. J. Struct. Biol. 116, 17–24. doi:10.1006/jsbi.1996.0004

Vanathi, P., Mishra, A.K., Bhargava, P., 2003. Regulation of activity of the yeast TATA-binding protein through intra-molecular interactions. J. Biosci. 28, 413–421.

Venters, B., Wachi, S., Mavrich, T., Andersen, B., Jena, P., Sinnamon, A., Jain, P., Rolleri, N., Jiang, C., Hemeryck-Walsh, C., Pugh, B., 2011. A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. Mol Cell 41, 480–492. doi:10.1016/j.molcel.2011.01.015

Venters, B.J., Pugh, B.F., 2009. How eukaryotic genes are transcribed [WWW Document]. URL http://informahealthcare.com/doi/full/10.1080/10409230902858785 (accessed 12.19.13).

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 Dev. 18, 1022–1034. doi:10.1101/gad.1192204

Washburn, M., Wolters, D., Yates, J., 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19, 242–247. doi:10.1038/85686

Watanabe, T., Hayashi, K., Tanaka, A., Furumoto, T., Hanaoka, F., Ohkuma, Y., 2003. The Carboxy Terminus of the Small Subunit of TFIIE Regulates the Transition from Transcription Initiation to Elongation by RNA Polymerase II. Mol. Cell. Biol. 23, 2914–2926. doi:10.1128/MCB.23.8.2914-2926.2003

Weake, V.M., Swanson, S.K., Mushegian, A., Florens, L., Washburn, M.P., Abmayr, S.M., Workman, J.L., 2009. A novel histone fold domain-containing protein that replaces TAF6 in Drosophila SAGA is required for SAGA-dependent gene expression. Genes Dev. 23, 2818–2823. doi:10.1101/gad.1846409

Weake, V.M., Workman, J.L., 2008. Histone Ubiquitination: Triggering Gene Activity. Mol. Cell 29, 653–663. doi:10.1016/j.molcel.2008.02.014

Wieczorek, E., Brand, M., Jacq, X., Tora, L., 1998. Function of TAFII-containing complex without TBP in transcription by RNA polymerase II. Nature 393, 187–191. doi:10.1038/30283

Winston, F., Carlson, M., 1992. Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. Trends Genet. TIG 8, 387–391.

Winston, F., Chaleff, D.T., Valent, B., Fink, G.R., 1984. Mutations Affecting Ty-Mediated Expression of the HIS4 Gene of SACCHAROMYCES CEREVISIAE. Genetics 107, 179–197.

Winston, F., Dollard, C., Malone, E.A., Clare, J., Kapakos, J.G., Farabaugh, P., Minehart, P.L., 1987. Three Genes Are Required for trans-Activation of Ty Transcription in Yeast. Genetics 115, 649–656.

Wood, A., Krogan, N.J., Dover, J., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Golshani, A., Zhang, Y., Greenblatt, J.F., Johnston, M., Shilatifard, A., 2003. Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Mol. Cell 11, 267–274.

Wu, P.-Y.J., Ruhlmann, C., Winston, F., Schultz, P., 2004. Molecular Architecture of the S. cerevisiae SAGA Complex. Mol. Cell 15, 199–208. doi:10.1016/j.molcel.2004.06.005

Wu, P.-Y.J., Winston, F., 2002. Analysis of Spt7 Function in the Saccharomyces cerevisiae SAGA Coactivator Complex. Mol. Cell. Biol. 22, 5367–5379. doi:10.1128/MCB.22.15.5367-5379.2002

Wu, Y., Reece, R.J., Ptashne, M., 1996. Quantitation of putative activator-target affinities predicts transcriptional activating potentials. EMBO J. 15, 3951–3963.

Wyce, A., Xiao, T., Whelan, K.A., Kosman, C., Walter, W., Eick, D., Hughes, T.R., Krogan, N.J., Strahl, B.D., Berger, S.L., 2007. H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related complex. Mol. Cell 27, 275–288. doi:10.1016/j.molcel.2007.01.035

Xu, W., Edmondson, D.G., Evrard, Y.A., Wakamiya, M., Behringer, R.R., Roth, S.Y., 2000. Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. Nat. Genet. 26, 229–232. doi:10.1038/79973

Yamauchi, T., Yamauchi, J., Kuwata, T., Tamura, T., Yamashita, T., Bae, N., Westphal, H., Ozato, K., Nakatani, Y., 2000. Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. Proc. Natl. Acad. Sci. 97, 11303–11306. doi:10.1073/pnas.97.21.11303

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., Martinez, E., 2007. Prevalence of the Initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene 389, 52–65. doi:10.1016/j.gene.2006.09.029

Yankulov, K.Y., Bentley, D.L., 1997. Regulation of CDK7 substrate specificity by MAT1 and TFIIH. EMBO J. 16, 1638–1646. doi:10.1093/emboj/16.7.1638

Yokomori, K., Admon, A., Goodrich, J.A., Chen, J.L., Tjian, R., 1993. Drosophila TFIIA-L is processed into two subunits that are associated with the TBP/TAF complex. Genes Dev. 7, 2235–2245.

Yu, Y., Eriksson, P., Bhoite, L.T., Stillman, D.J., 2003. Regulation of TATA-Binding Protein Binding by the SAGA Complex and the Nhp6 High-Mobility Group Protein. Mol. Cell. Biol. 23, 1910–1921. doi:10.1128/MCB.23.6.1910-1921.2003

Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., Rando, O.J., 2005. Genome-Scale Identification of Nucleosome Positions in S. cerevisiae. Science 309, 626–630. doi:10.1126/science.1112178

Zhang, C., Burton, Z.F., 2004. Transcription Factors IIF and IIS and Nucleoside Triphosphate Substrates as Dynamic Probes of the Human RNA Polymerase II Mechanism. J. Mol. Biol. 342, 1085–1099. doi:10.1016/j.jmb.2004.07.070

Zhang, X.-Y., Varthi, M., Sykes, S.M., Phillips, C., Warzecha, C., Zhu, W., Wyce, A., Thorne, A.W., Berger, S.L., McMahon, S.B., 2008. The putative cancer stem cell marker USP22 is a subunit of the human SAGA complex required for activated transcription and cell-cycle progression. Mol. Cell 29, 102–111. doi:10.1016/j.molcel.2007.12.015

Zhang, Y., Wen, Z., Washburn, M.P., Florens, L., 2010. Refinements to Label Free Proteome Quantitation: How to Deal with Peptides Shared by Multiple Proteins. Anal. Chem. 82, 2272–2281. doi:10.1021/ac9023999

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. Mol. Cell 29, 92–101. doi:10.1016/j.molcel.2007.12.011

Zhu, A., Kuziora, M.A., 1996. Homeodomain Interaction with the β Subunit of the General Transcription Factor TFIIE. J. Biol. Chem. 271, 20993–20996. doi:10.1074/jbc.271.35.20993

Zurita, M., Merino, C., 2003. The transcriptional complexity of the TFIIH complex. Trends Genet. 19, 578–584. doi:10.1016/j.tig.2003.08.005

CHAPTER 7 :

ANNEXES

ANNEXE 1 - PROTOCOLS

7.1 - S. CEREVISIAE CULTURE

7.1.1 Media composition

7.1.1.1 Liquid media

YPD media (100 mL)

Yeast Extract 1g (Becton, Dickinson, 212750)
Bacto Peptone 2g (Becton, Dickinson, 211677)
Add milliQ water up to 95 mL and autoclave. After autoclaving, add 5 mL of a sterile (filtered) 40% Glucose solution, to a final concentration of 2%

Synthetic Media without Leucine (100 mL)

Yeast Nitrogen Base without amino acid and ammonium sulfate (Becton, Dickinson 233520)
Ammonium Sulfate
Drop-out synthetic media without Leucine
Glucose 2g
Add MilliQ water up to 100 mL and filter the solution. Keep at 4°C.

7.1.1.2 Agar plate

YPD Agar (100 mL)

Yeast Extract 1g (Becton, Dickinson, 212750)
Bacto Peptone 2g (Becton, Dickinson, 211677)
Bacto Agar 2g (Becton, Dickinson, 214010)
Add milliQ water up to 95 mL and autoclave. The solidified media can be melted in the microwave and let to cool down in a water bath at 50°C. Before pooring the plate, add 5 mL of a sterile (filtered) 40% Glucose solution,pre-warmed to 50°C, to a final concentration of 2%

Synthetic Solid Media without Leucine (100 mL)

Yeast Nitrogen Base without amino acid and ammonium sulfate (Becton, Dickinson 233520)
Ammonium Sulfate
Drop-out synthetic media without Leucine
Glucose 2g
Add MilliQ water up to 100 mL and filter the solution. Keep at 4°C.

7.1.2 Yeast culture in liquid media

1. Three days before the culture, with a sterile tip, scratch the top of a frozen (-80°C) glycerol stock of the yeast strain, and plate on a YPD agar plate or a selective media plate
2. Incubate at 30°C until nice, isolated colonies are visible. This should take 2 days for YPD media, and might require one more day for selective media.
3. The day before the culture, inoculate a small (25 mL) liquid solution (YPD or selective media) from a single colony and incubate in a shaker at 30°C o/n, with vigorous shaking.
4. The day of the culture, measure the optical density (OD) of the pre-culture at 600 nm. Saturated cultures usually grow to a final OD 600nm of 10. From this culture, inoculate a 1L liter culture with an appropriate volume: inoculation with 200 µL of pre-culture at 5 pm usually allows one to harvest cells at 10h with a final OD 600 nm close to 2, considering a doubling time of 1h40. If this OD is not reached, let the culture grow longer during the day.
5. To harvest, centrifuge the cell suspension in 1L bottle at 4000g for 20 min at 4°C. Discard supernatant.
6. Wash cells once with 50 mL of cold, MilliQ water and transfer the cell suspsension into 50 mL falcon tubes.
7. Pellet the cells (4000g for 20 min) and discard supernatant. Weight the pellet and froze the tube in liquid nitrogen, before storing them at -80°C. Typically, 2L of culture at a final OD of 2 should result in approximately 5g of pelleted cells.

7.2 - PURIFICATION OF THE SAGA COMPLEX

7.2.1 Buffers

7.2.1.1 Lysis buffer

Tris-HCl 50 mM pH 8.0
NaCl 300 mM
MgCl2 2 mM
Glycerol 10%
Tween20 0.1%

7.2.1.2 TEV Cleavage Buffer (TCB)

Hepes 30 mM pH 8.0
NaCl 300 mM
Tween20 0.1%

7.2.1.3 Calmodulin Binding Buffer (CBB)

Hepes 30 mM pH 8.0
NaCl 300 mM
Tween20 0.1%
CaCl2 2 mM

7.2.1.4 Calmodulin Washing Buffer (CWB)

Hepes 30 mM pH 8.0
NaCl 500 mM
CaCl2 2 mM

7.2.1.5 Calmodulin Elution Buffer (CEB)

Hepes 30 mM pH 8.0
NaCl 500 mM
EGTA 2 mM

7.2.2 Preparation of the yeast extract

1. Thaw frozen cells on ice and resuspend the cell with 1 mL lysis buffer (plus protease inhibitor cocktail, *Complete EDTA-free* from Roche)per g of cells used for the purification. A typical purification is made from 20g of cells.
2. Fill 4 round-bottom plastic centrifuge tubes (Nalgene, 029235) with 15g of acid-washed glass-beads (0.45-0.6 µm, Sigma 8772) and place them on ice to chill. This step can be done the day before.
3. To each tube, add 10 mL of the cold cell suspension. Equilibrate the tubes 2 by 2 for the subsequent centrifugation step, and sealed them with parafilm.

4. Lysis is performed by vortexing each tube 5 min in the cold room, by pulse of 1 min and letting each tube cool down on ice between each pulse. It is important to keep the cell extract at 4°C during all the purification process.
5. After lysis, centrifuge the tubes 2 min at 8000g in order to precipitate the glass beads. Transfer the extract into a clean, pre-chilled centrifugation tube and centrifuge the extract at 48000g for 1h. Collect the supernatant.
6. Control lysis efficiency by measuring OD at 260 nm. On a nanodrop, OD should be greater than 90. Alternatively, measure protein concentration in the extract by the Bradford method.

7.2.3 TAP-tag purification of the complex

1. During the centrifugation, equilibrate 200 µL of IgG Sepharose beads (GE Healthcare, 17-0969-01), by suspending the beads in 5 mL of lysis buffer, centrifugation at 400g for 5 min, and discarding the supernatant. Repeat once.
2. Incubate the yeast extract with the beads for at least 4h, under gentle agitation at 4°C.
3. After incubation, wash the beads 5 times with 3 mL of TCB. At the last wash, measure OD at 260 nm to check efficiency of the washes. If necessary, perform additional washes.
4. Resuspend the beads in 400 mL of TCB and transfer the suspension into a clod Eppendorf tube. Add 3µg of TEV protease and incubate at 4°C o/n.
5. The day after, centrifuge the beads at 400g for 5 min, and collect supernatant. Wash the beads once with additional 400 µL of TCB, and pool with the previous elution.
6. In parallel, equilibrate 100µL of Calmodulin Affinity Resin (Agilent, 214303) with 1 mL of CBB in an Eppendorf tube. Repeat once.
7. To the 800 µL of TEV elution, add 1.6 µL of CaCl2 1M (final concentration 2 mM) and incubate on Calmodulin Affinity Resin for 2 hours at 4°C, under gentle agitation.
8. Wash 4 times with 1 mL of CWB.
9. Perform elution by resuspending the beads with 100 µL of CEB, and centrifugation at 400g for 5 min. Repeat 4 times.
10. Control the purification by PAGE-SDS on a 4-15% gradient gel (Mini-Protean TGX Precast gel, Biorad, 456-1085) and silver staining (Fisher, 24612) according to manufacturer protocol, and by EM after adsorption and negative staining on a grid.

7.3 - PROTEIN COMPLEX ANALYSIS

7.3.1 PAGE-SDS

7.3.1.1 Buffers and gel composition

Running Buffer (for 2L)

Tris-base 30g
Glycine 188g
SDS 20% 50 mL (0.5% final)

Leammli Buffer (5x stock solution)

Tris-HCl pH 6.8 50 mM
Glycerol 25%
Bromophenol Blue 0.1%
SDS 5%

Leammli Buffer (4x working solution) (for 1 mL)

Leammli buffer 5x 800 µL
Beta-mercaptoethanol 100% (14.3M) 40 µL
H2O 160µL

Gel Buffer pH 8.8 (for 1L)

Tris base 181.7g
SDS 20% 20 mL (0.25% final)
Adjust pH with HCl

Gel Buffer pH 6.8 (for 0.5L)

Tris base 30.29g
SDS 20% 10 mL (0.25% final)
Adjust pH with HCl

Gel composition

	Resolving						**Stacking**
	8%	10%	12.5%	15%	17.5%	20%	
Acrylamide 40%	2 mL	2.5 mL	3.125 mL	3.75 mL	4.375 mL	5 mL	0.5 mL
Gel buffer pH 8.8	2.5 mL	2.5 mL	2.5 mL	2.5 mL	2.5 mL	2.5 mL	
Gel buffer pH 6.8							1 mL
APS 10%	125 µL	125 µL	125 µL	125 µL	125 µL	125 µL	125 µL
TEMED	12 µL	12 µL	12 µL	12 µL	12 µL	12 µL	4 µL
H2O	5.5 mL	5 mL	4.375 mL	3.75 mL	3.125 mL	2.5 mL	2.5 mL

Gradient gel

Gradient gels were purchased from Biorad: 4–15% Mini-PROTEAN® TGX™ Precast Gel (456-1085).

7.3.1.2 Silver staining

Silver staining of gels was performed with Pierce Silver Staining Kit (24612), according to manufacturer protocol.

7.3.2 Western-Blot

7.3.2.1 Buffers

Transfer buffer stock solution 10x (for 2L)

Tris-Base 60g
Glycine 285g
SDS 20% 37.5 mL

Transfer buffer 1x (for 1L)

Transfer buffer 10x 100 mL
Ethanol 100% 200 mL
Fill up with MilliQ water

Tris Buffer Saline (TBS)

Tris-HCl pH 8.0 10 mM
NaCl 150 mM
Tween20 0.1%

7.3.2.2 Protocol

Western-blot is performed by transferring proteins from the PAGE-SDS gel on a 0.45µm nitrocellulose membrane (Sigma, Z613630-1EA), for 1h at 100V. Then, membrane is blocked in a 2% milk solution (in TBS) for 10 min under constant shaking. After washes, the primary antibody is incubated at appropriated dilution (in TBS) for 1h at room temperature (RT). Membrane is then washed at least 3 times in TBS, and secondary antibody (ref) is incubated at appropriated dilution for 1h at RT under constant shaking. Membrane is washed 3 to 5 times with TBS to remove excess of antibody, and (ref) is added for revealing the presence of secondary antibody. Signal from the HRP catalysis reaction is recorded on photographic film (Kodak Biomax).

7.4 - SAMPLE PREPARATION FOR ELECTRON MICROSCOPY ANALYSIS

7.4.1 Preparation of carbon-coated grids for Electron Microscopy

7.4.1.1 Continuous carbon film

1. On a freshly cleaved mica (Electron Microscopy Sciences, 71851-05), evaporate a fine carbon film in a vacuum coating system (Edwards Auto306).
2. The day after, place the microscopy grids (Electron Microscopy Sciences, M300-CR or Quantifoil – see below) on a blotting paper at the bottom of a small beaker filled with clean distilled water.
3. Deposit the carbon film by floating it from the mica on the water surface, and removing water with a water aspirator. Let the grids dry for one night.

7.4.1.2 Holey carbon film

Grids with a holey carbon film are purchased from Quantifoil. Quantifoil with a holey plastic film (Quantifoil, N1-P16NCR30-01) is place in the evaporater and a thick carbon layer is deposited. The day after, plastic is dissolved by adding several drop of ethyl acetate on the grids, and waiting one night for evaporation of the ethyl acetate. An additional, continuous carbon film can be added to these grids according to the previous protocol.

ANNEXE 2 - PUBLICATIONS

7.5 - PUBLICATION 1

Mapping the deubiquitination module within the SAGA complex
Alexandre Durand, Jacques Bonnet and Patrick Schultz
In preparation, submitted

7.6 - PUBLICATION 2

Structure, assembly and dynamics of macromolecular complexes by single particle cryo-electron microscopy
Alexandre Durand, Gabor Papai and Patrick Schultz
Journal of Nanobiotechnology 2013, **11**(Suppl 1):S4
doi:10.1186/1477-3155-11-S1-S4
http://www.jnanobiotechnology.com/content/11/S1/S4

TITLE :

MAPPING THE DEUBIQUITINATION MODULE WITHIN THE SAGA COMPLEX

AUTHORS :

Alexandre Durand[1], Jacques Bonnet[2,3], Marjorie Fournier[2], and Patrick Schultz[1,4]

ADDRESSES :

1. Department of Structural Biology and Genomics,
2. Department of Functional Genomics

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

INSERM, U964 ; CNRS, UMR7104

1, rue Laurent Fries, BP10142

67404 Illkirch, France

3. present address: Max-Planck Institut fur Biochemie,

Am Klopferspitz 18, 82152 Martinsried, Germany

4. Corresponding author e-mail : patrick.schultz@igbmc.u-strasbg.fr

Word Counts :

Abstract

The molecular organization of the yeast transcriptional coactivator SAGA was analyzed by single particle electron microscopy. Complete or partial deletion of the Sgf73 subunit disconnects the deubiquitination (DUB) module from SAGA and favors in our conditions the cleavage of the C-terminal ends of the Spt7 subunit and the loss of the Spt8 subunit. The structural comparison of the wild-type SAGA with two deletion mutants positioned the DUB module and enabled the fitting of the available atomic models. The localization of the DUB module close to Gcn5 defines a nucleosome binding interface within SAGA which could be demonstrated by the binding of nucleosome core particles. The TBP-interacting subunit Spt8 was found to be located close to the DUB but in a different domain than Spt3, also known to contact TBP. A flexible protein arm brings both subunits close enough to interact simultaneously with TBP.

Introduction

Transcription of protein coding genes by RNA polymerase II (RNAPII) is a tightly regulated process that requires the coordinated action of several protein molecules triggering the assembly of the Pre-Initiation Complex (PIC) on gene promoters. Sequence-specific transcriptional activators and post-translational modifications of nucleosomal histones recruit multi-subunit coactivator complexes acting as bridging factors between activators and the PIC (Grunberg and Hahn, 2013). Coactivators modify promoter chromatin structure and coordinate PIC assembly with epigenetic chromatin modifications. The Spt-Ada-Gcn5 Acetyltransferase (SAGA) complex is a paramount example of coactivators (Grant et al., 1997). In *S. cerevisiae*, the 1.8 MDa SAGA complex is composed of 19 distinct subunits, most of which have homologues in higher eukaryotes (Nagy and Tora, 2007). The SAGA complex has a modular organization as evidenced by genetic complementation studies (Eisenmann et al., 1994), and electron microscopy models (Wu et al., 2004). Recently, quantitative proteomics established a subunit interaction network and segmented SAGA into 4 stable modules (Lee et al., 2011). The SPT module contains four proteins (Spt3, Spt7, Spt8, Spt20), originally identified in a genetic screen for suppressors of promoter mutations (Winston and Sudarsanam, 1998) clustered with Ada1 and Tra1. Spt3 and Spt8 were shown to directly interact with TBP (Mohibullah and Hahn, 2008) and to modulate positively or negatively its interaction with particular promoters (Warfield et al., 2004). The C-terminus of Spt7 can be cleaved by the Pep4 protease to form the SAGA-like (SLIK or SALSA) complex which also lacks Spt8 that interacts with the cleaved part of Spt7 (Pray-Grant et al., 2002; Spedale et al., 2010). SLIK was shown to be recruited to HIS3 promoter and activate transcription, thereby counteracting the inhibiting role of SAGA on this promoter (Belotserkovskaya et al., 2000). The Tra1 subunit is shared with the NuA4 complex and was shown to interact with a large set of activator such as Gcn4 and Gal4 (Brown et al., 2001; Knutson and Hahn, 2011).

A TBP-Associated Factor (TAF) module composed of Taf5, 6, 9, 10 and 12 is shared with the TFIID complex and is believed to form a structural core on which the other modules are assembled.

A core-TFIID complex composed of a similar Taf subset (Taf5, 6, 9, 4, 12) was identified (Bieniossek et al., 2013). Its structure and assembly pathway revealed a dimeric core TFIID assembled from a $(Taf5\text{-}6\text{-}9)_2$ sub-complex shared with SAGA. The two other Tafs present in SAGA (Taf10 and Taf12) contain a histone-fold domain (HFD) and have SAGA-specific HFD-containing interaction partners, Spt7 and Ada1, respectively (Gangloff et al., 2001). SAGA harbors a Histone Acetyl Transferase (HAT) activity carried by the Gcn5 subunit and modulated by the Ada2 and Ada3 subunits (Grant et al., 1997). Together with Sgf29, these subunits form the HAT module which is a major regulator of histone H3 acetylation in yeast cells. Gcn5 contains a bromodomain which binds acetylated lysines in histone tails (Hassan et al., 2002) and Sgf29 holds a double Tudor domain capable of binding H3K4me2/3, two hallmark of actively transcribed chromatin (Vermeulen et al., 2010).

A deubiquitination (DUB) module composed of the catalytic Ubp8 subunit, tightly regulated by the Sgf73, Sgf11 and Sus1 subunits, catalyzes the cleavage of monoubiquitin from lysine-123 of histone H2B (Daniel et al., 2004; Henry et al., 2003). Ubp8 is inactive until interacting with the other DUB module subunits indicating that the presence of all four components is required for optimal enzymatic activity (Kohler et al., 2008). The Sgf73 subunit harbors the highly conserved SCA7 domain, which has been shown to interact with nucleosomes (Bonnet et al., 2010). The incorporation of a long polyglutamine expansion into ATXN7, the human homologue of Sgf73 leads to type 7 spinocerebellar ataxia (David et al., 1997; Helmlinger et al., 2004).

Structural information is required to understand the mechanisms of SAGA-dependant gene activation and the crosstalk between the multiple functions integrated in the SAGA complex. Electron microscopy (EM) has been used to determine a 3-D model of the SAGA complex and to map nine out of the nineteen SAGA subunits (Wu et al., 2004). The atomic structure of the DUB module shows a compact organization with a highly intertwined subunits interaction network (Kohler et al., 2010; Samara et al., 2010). The DUB module forms two functional lobes, organized around the two Ubp8 domains. The catalytic lobe is formed by the C-terminal domain of Ubp8 and

the C-terminal zinc-finger domain of Sgf11. The assembly lobe is organized around a long Sgf11 helix. The N-terminal 104 residues of Sgf73 are located between the two Ubp8 domains where they potentiate Ubp8 activity, and anchor the DUB module to the SAGA complex.

To locate the DUB module within the SAGA complex, we have determined the 3-D structure of SAGA complexes in which the N-terminal part of Sgf73 or the full Sgf73 subunit have been deleted to remove the DUB module. The module is found in the vicinity of Gcn5 and Spt7, thus defining a nucleosome interaction domain within the SAGA complex. In our hands the deletion of the DUB module leads to an increased processing of the C-terminal part of Spt7, and consequently to the loss of Spt8 that mapped in the same region as the DUB module, which was unexpected since Spt3, the other TBP-interacting protein, is found in a different domain.

Results

1-The deletion of SAGA DUB module promotes Spt7 cleavage

Saccharomyces cerevisiae SAGA was purified from a wild-type (WT) strain TAP-tagged at the C-terminus of Spt20 and its protein composition was consistent with previous results (Fig.1A, lane 1). The Spt7 subunit migrated as two bands with apparent molecular weights of 250 Kda for the full length and 230 kDa for the truncated form of Spt7 which incorporates into the SLIK complex (Pray-Grant et al., 2002; Spedale et al., 2010). In our conditions these two bands are of similar intensity suggesting that SLIK and SAGA are equally represented.

Samples were analyzed by multidimensional protein identification technology (MudPIT) (Washburn et al., 2001) and the relative abundance of SAGA subunits was estimated by Normalized Spectral Abundance Factors (NSAF) (Zybailov et al., 2006) (Fig.1B). All 19 SAGA subunits were identified among the most abundant proteins and the NSAF values were in similar range, suggesting that all subunits are present in the same copy number except for Sus1, which was detected at higher abundance. The structural homogeneity of the sample was checked by electron microscopy of negatively stained molecules (Fig.1C).

A *sgf73*$\Delta^{1\text{-}104}$ strain was generated from the WT strain by deleting residues 1-104 of Sgf73 necessary to assemble the DUB module. As expected, the four DUB module subunits were absent from the *sgf73*$\Delta^{1\text{-}104}$ SAGA as shown by gel electrophoresis (Fig.1A, lane 2) and mass spectrometry (Fig.1B), except for the Sgf73 subunit for which peptides corresponding to the C-terminus were still detected. Interestingly, no peptide belonging to the C-terminal part (after position 1142) of Spt7 was detected and the upper Spt7 band is missing in SDS-PAGE (Fig.1A, lane 2). The cleaved part of Spt7 holds the Spt8 binding site (Spedale et al., 2010) and, consistently, mass spectrometry showed that the Spt8 subunit is missing, indicating that the *sgf73*$\Delta^{1\text{-}104}$ SAGA is fully converted into the SLIK form. The full Spt7 cleavage found upon removal of the DUB module was not observed previously (Bian et al., 2011; Kohler et al., 2010) and may reflect slight differences in genetic background, growth conditions or purification conditions. The proteomic analysis of a second mutant with a full *SGF73* deletion (*sgf73*Δ strain) confirmed that the four subunits of the DUB module were missing and that Spt7 was preferentially cleaved resulting in low Spt8 abundance (Fig.1A, lane 3 and Fig.1B).

2-Structure and molecular motion in WT SAGA

Negatively stained WT SAGA molecules were processed for electron microscopy and the analysis of 19.226 images revealed characteristic SAGA views showing 5 domains annotated from I to V and organized into two major lobes containing domains I and II (lobe A), and domains III, IV and V (lobe B) (Fig.2A). The 3-D model calculated from this dataset revealed three protein connections between domains I and II in lobe A, as well as large solvent accessible channels (Fig.2B). Lobe B adopts a molecular clamp-shaped architecture in which domains III and V protrude from a large crescent-shaped domain IV. The major connection between the two lobes occurs through domains II and IV while a faint connection is observed between domains I and III.

In order to analyze possible conformational changes, a three-dimensional classification scheme based on regularized likelihood optimization was used to form four 3-D classes (Fig. 2C)

(Scheres, 2010). While the structure of lobe A did not fluctuate, lobe B showed large conformational changes. A major variation is evidenced by an opening and closing of the clamp. In the most closed configuration, the two domains are in contact and form a closed protein ring, while in the more opened state the two domains are separated by 7 nm. Relative movements between lobes A and B were also evidenced such as rotations along the long molecular axis and around the connection between domains II and IV.

To correct for these movements, each lobe was aligned independently to improve the resolution of structural details. Such an approach has been used previously (Sauerwald et al., 2013) and is applicable because the two lobes show little overlap in the SAGA views. While little resolution gain was obtained for lobe A, the structural description of lobe B improved significantly (Fig. 2D). In particular, domain III is found more extended and harbors a protein density only drafted in the uncorrected model and domain V grows in size and shows more structural details such as a constriction in the connection to domain IV. Finally the connection between domains I and III is strengthened when lobe B is analyzed separately.

3-Mapping of the DUB module

SAGA complexes purified from the *sgf73*$\Delta^{1\text{-}104}$ or *sgf73*Δ strains were analyzed to map the DUB module. A 3-D model of *sgf73*Δ SAGA was determined from 30.409 negatively stained molecules by analyzing each lobe independently. Lobe A is similar in the mutant and in the WT SAGA models (data not shown) thus excluding the possibility that it hosts the 137 kDa DUB module. In contrast, two large differences were detected in lobe B. The largest difference reflects the absence of domain V the *sgf73*Δ SAGA complexes (Red density marked by an arrowhead in Fig.3B). While domain V is not visible in *sgf73*Δ SAGA, the analysis of the *sgf73*$\Delta^{1\text{-}104}$ mutant clearly revealed the flexible domain V with similar dimensions than in WT SAGA (Fig3A). This observation indicates that domain V does not contain the DUB module. Moreover a similarly sized volume is found inserted at the bottom of the *sgf73*Δ SAGA cleft while it is absent in the WT

SAGA (Blue density marked by an arrowhead in Fig.3B). This observation indicates that domain V folds back into the cleft when the C-terminal part of Sgf73 is removed.

The second density missing in the mutant SAGA complex is located at the tip of domain III and the size of this density is compatible with that of the DUB module (Red density marked by an asterisk in Fig.3B). The crystal structure of the quaternary complex containing the full-length Ubp8, Sgf11, and Sus1 and the N-terminal fragment of Sgf73 (Kohler et al., 2010) was fitted into this missing volume in a unique position since the location of the Ubp8 domain composed of residues 281-303 and 326-353 forming 3 β–strands and 2 loops was discriminative (Fig.3C).

4-Mapping the Spt8 subunit.

The density difference map between the WT and the *sgf73*Δ SAGA complexes revealed a third significant difference located at the interface between domain III and domain I (Red Arrow in Fig.3B and Fig.4A.). This density could correspond to the 189 C-terminal residues of Spt7 (21.6 kDa) and to Spt8 (66 kDa) lost during proteolysis. Consistently, previous labeling experiments mapped the C-terminus of Spt7 in lobe III (Wu et al., 2004). In order to confirm the position of Spt8, subunit-specific antibodies were incubated with purified WT SAGA. The analysis of 1227 images shows the antibody molecule bound to domain III in a position consistent with the density detected in the difference map (Fig.4B).

5-Interaction of SAGA with the Nucleosome Core Particle.

In order to investigate the interaction between SAGA and the nucleosome core particle (NCP) coactivator complexes were immobilized on calmodulin sepharose beads and incubated with a 7.5 fold molar excess of recombinant NCPs formed from full-length bacterially expressed recombinant *Xenopus Laevis* core histones assembled onto a 196 bp long 601 nucleosome positioning sequence (Syed et al., 2010). The complexes were briefly washed and eluted before electron microscopy preparation and observation. The analysis of 3909 negatively stained complexes revealed that about

24 % of SAGA particles were bound to NCPs. Class averages calculated from SAGA-NCP images showed a bound density adjacent to domain III and consistent with the size of a nucleosome (Fig. 4C). Both round and elongated densities are attached to SAGA suggesting that the NCP orientation can fluctuate.

Discussion

The architecture of the SAGA co-activator was revisited by analyzing a large image dataset and by considering its molecular heterogeneity and flexibility. Lobe A appears extremely stable and its organization into two domains interconnected by three protein links did not show significant conformational changes. Antibody labeling experiments showed that the crescent-shaped domain I contains the Tra1 subunit (Wu et al., 2004). This part of SAGA was found structurally similar to the NuA4 complex which is also organized into two domains (Chittuluru et al., 2011). This similarity probably reflects the functional homology between the two complexes in activator binding and HAT activity. Not only is the overall shape conserved but also the number and position of connections with Tra1 suggesting that Tra1 shares the same interaction interfaces with the SAGA or NuA4 subunits. A recent genetic and functional dissection of Tra1 is consistent with this observation since the same non-viable Tra1 mutations both affected SAGA and NuA4 complex stability (Knutson and Hahn, 2011). The modular organization of SAGA determined by quantitative mass spectrometry (Lee et al., 2011) has proposed an interaction between Tra1 and subunits of the SPT module. Spt3 and Spt20 map in the B lobe at a distance from Tra1 that precludes direct interaction (Wu et al., 2004). Our results however indicate that Spt8 and possibly the C-terminus of Spt7 form a bridge between domain III and I to contact Tra1 thus giving a hint to the observed interaction.

Position of the DUB module

The DUB module is located at the tip of domain III, and the fitting of its atomic coordinates shows that the ubiquitin binding and the catalytic sites are solvent exposed in a suitable orientation to interact with chromatin. The DUB module is positioned close to Gcn5 which may explain the reported cross-talks between the two enzymatic activities. In mammalian cells the deletion of Gcn5 impairs the association of the DUB module with SAGA (Atanassov et al., 2009) while in yeast, the full deletion of Sgf73 compromises the HAT activity which can be restored by expressing human ATXN7 (McMahon et al., 2005). Furthermore, an ATXN7 carrying a 60 glutamine repeat to mimic the human SCA7 disease was described to bind to a recombinant Gcn5/ada2/ada3 complex and to decrease its HAT activity (Burke et al., 2013). The spatial proximity of the HAT and DUB modules is consistent with this observation since a long and unstructured polyQ stretch could interfere with the HAT catalytic or regulatory activities, or with the nucleosome recognition functions.

In this study Spt7 cleavage correlates with the removal of the DUB module in a *sgf73*Δ background which was unexpected since previous results did not reveal such an effect (Kohler et al., 2008; Lee et al., 2011). Minor differences in the purification protocol or genetic background may explain this discrepancy. The DUB module maps close to Spt7 and its removal may uncover the N-terminus of Spt7 and promote its degradation by the cytoplasmic Pep4 protease (Spedale et al., 2010). The activator binding and promoter recruitment activities of SAGA have been shown to be regulated by the proteasome Regulatory Particle and recent results suggest that the proteasomal ATPase can dissociate the DUB module from SAGA (Lee et al., 2005; Lim et al., 2013). Our results show that upon Sgf73 depletion the Spt3-arm folds back into the cleft thus modifying dramatically the geometry of the TBP-binding interfaces (see below). Therefore the promoter recognition function of SAGA is likely to be affected by the removal of the DUB module.

Chromatin-binding interface

Readers and writers of post translational histone modifications are clustered in lobe B. The bromodomains of Gcn5 and Spt7, the UcH domain in Ubp8, the HAT domain in Gcn5, the SCA7

domain of Sgf73 and the tandem tudor domain of Sgf29 constitute potential chromatin interaction domains. Sgf29 and Gcn5 are part of the HAT module located in domain III (Wu et al., 2004). Here, we show that Sgf73, Ubp8 and the C-terminal part of Spt7 are also located in this domain, which thus constitutes a platform for recognition, binding and modification of nucleosomes. We also show that recombinant NCPs bind this domain probably in a loose interaction since their orientation with respect to SAGA is not well defined. This may reflect a binding through flexible histone tails which are likely to be stabilized by specific modifications such as acetylations or histone H3K4 di- or tri-methylations.

Identification of a TBP-binding clamp

The Spt3 and Spt8 subunits of SAGA interact genetically with TBP and are believed to participate to the recruitment of TBP to SAGA-dependent gene promoters (Laprade et al., 2007). Biochemical protein crosslinking and pull-down experiments showed functional interactions of TBP with Spt8 and with Spt3 (Mohibullah and Hahn, 2008; Sterner et al., 1999; Warfield et al., 2004). Spt3 and Spt20 were previously shown to map to the flexible domain V while here we map Spt8 in domain III. The two subunits are thus located in two distinct domains and form the tips of a large protein clamp encompassing lobe B. The Spt3-arm is highly dynamic and can a closed position bringing Spt3 in close proximity to Spt8 thus giving a hint to the biochemical and genetic interaction data.

The Spt8 subunit plays a dual role in transcription initiation. On one hand, the Spt8 deletion derepresses the basal expression of HIS3 and TRP3 genes. This inhibition of basal transcription is believed to be caused by a competition with TFIIA for TBP binding. On the other hand Spt8 stimulates transcription from SAGA dependent promoters in the presence of an activator by mediating a positive effect of TFIIA (Warfield et al., 2004). These results suggest that upon interaction with activators Spt8 undergoes a conformational change which turns it into an active state. This behavior is paralleled in TFIID where TFIIA undergoes a massive reorganization upon

activator binding which likely results in the release the TAF1-induced inhibition of TBP binding to promoter DNA (Papai et al., 2010). The negative regulator of TBP-promoter binding Mot1 shows genetic interactions with Spt3, Spt8, Spt7 and Gcn5 indication that the Stp3-Spt8 clamp plays a major in regulating the interaction of TBP with the promoter DNA (Madison and Winston, 1997).

Acknowledgements

We thank A. Page and V. Chavan for mass spectrometry, M. Timmers, B. Seraphin and C. Birck for helpful advices and reagents; D. Devys and L. Tora for stimulating scientific discussions. We are grateful to P.A. Weil for the gift of anti Spt8 antibodies and to S. Dimitrov for providing purified NCP. This work was supported by the Fondation pour la Recherche Médicale, the Institut National de la Santé et de la Recherche Médicale, the Centre National pour la Recherche Scientifique, the Strasbourg University, the Association pour la Recherche sur le Cancer and the Agence Nationale pour la Recherche (SCA7SAGA), the Labex INRT, the French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-01] and INSTRUCT as part of the European Strategy Forum on Research Infrastructures (ESFRI).

Materials and Methods

Details of the experimental procedure can be found in the supplemental experimental procedures. Briefly, WT and mutant SAGA missing the DUB module were purified using the tandem affinity purification method with a tag placed at the C-terminus of Spt20, and prepared for negative stain electron microscopy. Three-dimensional models were determined using standard protocols and made use of a previous model determined from Random Conical Tilt data (Wu et al., 2004). To analyze each lobe separately, the coordinates of each part of the complex were determined manually and images from lobes A and B were extracted from the aligned dataset. Initial volumes for each lobe were extracted from the 3D model of the full complex and used for reference matching.

Bibliography

Atanassov, B.S., Evrard, Y.A., Multani, A.S., Zhang, Z., Tora, L., Devys, D., Chang, S., and Dent, S.Y. (2009). Gcn5 and SAGA regulate shelterin protein turnover and telomere maintenance. Mol Cell 35, 352-364.

Belotserkovskaya, R., Sterner, D.E., Deng, M., Sayre, M.H., Lieberman, P.M., and Berger, S.L. (2000). Inhibition of TATA-binding protein function by SAGA subunits Spt3 and Spt8 at Gcn4-activated promoters. Mol Cell Biol 20, 634-647.

Bian, C., Xu, C., Ruan, J., Lee, K.K., Burke, T.L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B.O., et al. (2011). Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. EMBO J 30, 2829-2842.

Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. (2013). The architecture of human general transcription factor TFIID core complex. Nature In press.

Bonnet, J., Wang, Y.H., Spedale, G., Atkinson, R.A., Romier, C., Hamiche, A., Pijnappel, W.W., Timmers, H.T., Tora, L., Devys, D., et al. (2010). The structural plasticity of SCA7 domains defines their differential nucleosome-binding properties. EMBO Rep 11, 612-618.

Brown, C.E., Howe, L., Sousa, K., Alley, S.C., Carrozza, M.J., Tan, S., and Workman, J.L. (2001). Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. Science 292, 2333-2337.

Burke, T.L., Miller, J.L., and Grant, P.A. (2013). Direct inhibition of Gcn5 protein catalytic activity by polyglutamine-expanded ataxin-7. J Biol Chem 288, 34266-34275.

Chittuluru, J.R., Chaban, Y., Monnet-Saksouk, J., Carrozza, M.J., Sapountzi, V., Selleck, W., Huang, J., Utley, R.T., Cramet, M., Allard, S., et al. (2011). Structure and nucleosome interaction of the yeast NuA4 and Piccolo-NuA4 histone acetyltransferase complexes. Nat Struct Mol Biol 18, 1196-1203.

Daniel, J.A., Torok, M.S., Sun, Z.W., Schieltz, D., Allis, C.D., Yates, J.R., 3rd, and Grant, P.A. (2004). Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. J Biol Chem 279, 1867-1871.

David, G., Abbas, N., Stevanin, G., Durr, A., Yvert, G., Cancel, G., Weber, C., Imbert, G., Saudou, F., Antoniou, E., et al. (1997). Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. Nat Genet 17, 65-70.

Eisenmann, D.M., Chapon, C., Roberts, S.M., Dollard, C., and Winston, F. (1994). The Saccharomyces cerevisiae SPT8 gene encodes a very acidic protein that is functionally related to SPT3 and TATA-binding protein. Genetics 137, 647-657.

Gangloff, Y., Romier, C., Thuault, S., Werten, S., and Davidson, I. (2001). The histone fold is a key structural motif of transcription factor TFIID. Trends Biochem Sci 26, 250-257.

Grant, P.A., Duggan, L., Cote, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F., et al. (1997). Yeast Gcn5 functions in two multisubunit

complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev 11, 1640-1650.

Grunberg, S., and Hahn, S. (2013). Structural insights into transcription initiation by RNA polymerase II. Trends Biochem Sci 38, 603-611.

Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J., and Workman, J.L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell 111, 369-379.

Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguet, L., Potier, N., Van-Dorsselaer, A., Wurtz, J.M., et al. (2004). Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. Hum Mol Genet 13, 1257-1265.

Henry, K.W., Wyce, A., Lo, W.S., Duggan, L.J., Emre, N.C., Kao, C.F., Pillus, L., Shilatifard, A., Osley, M.A., and Berger, S.L. (2003). Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes Dev 17, 2648-2663.

Knutson, B.A., and Hahn, S. (2011). Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and NuA4 complexes. Mol Cell Biol 31, 818-831.

Kohler, A., Schneider, M., Cabal, G.G., Nehrbass, U., and Hurt, E. (2008). Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export. Nat Cell Biol 10, 707-715.

Kohler, A., Zimmerman, E., Schneider, M., Hurt, E., and Zheng, N. (2010). Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. Cell 141, 606-617.

Laprade, L., Rose, D., and 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, 2007-2017.

Lee, D., Ezhkova, E., Li, B., Pattenden, S.G., Tansey, W.P., and Workman, J.L. (2005). The proteasome regulatory particle alters the SAGA coactivator to enhance its interactions with transcriptional activators. Cell 123, 423-436.

Lee, K.K., Sardiu, M.E., Swanson, S.K., Gilmore, J.M., Torok, M., Grant, P.A., Florens, L., Workman, J.L., and Washburn, M.P. (2011). Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. Mol Syst Biol 7, 503.

Lim, S., Kwak, J., Kim, M., and Lee, D. (2013). Separation of a functional deubiquitylating module from the SAGA complex by the proteasome regulatory particle. Nat Commun 4, 2641.

Madison, J.M., and Winston, F. (1997). Evidence that Spt3 functionally interacts with Mot1, TFIIA, and TATA- binding protein to confer promoter-specific transcriptional control in Saccharomyces cerevisiae. Mol Cell Biol 17, 287-295.

McMahon, S.J., Pray-Grant, M.G., Schieltz, D., Yates, J.R., 3rd, and Grant, P.A. (2005). Polyglutamine-expanded spinocerebellar ataxia-7 protein disrupts normal SAGA and SLIK histone acetyltransferase activity. Proc Natl Acad Sci U S A 102, 8478-8482.

Mohibullah, N., and 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 Dev 22, 2994-3006.

Nagy, Z., and Tora, L. (2007). Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. Oncogene 26, 5341-5357.

Papai, G., Tripathi, M.K., Ruhlmann, C., Layer, J.H., Weil, P.A., and Schultz, P. (2010). TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. Nature 465, 956-960.

Pray-Grant, M.G., Schieltz, D., McMahon, S.J., Wood, J.M., Kennedy, E.L., Cook, R.G., Workman, J.L., Yates, J.R., 3rd, and Grant, P.A. (2002). The novel SLIK histone acetyltransferase complex functions in the yeast retrograde response pathway. Mol Cell Biol 22, 8774-8786.

Samara, N.L., Datta, A.B., Berndsen, C.E., Zhang, X., Yao, T., Cohen, R.E., and Wolberger, C. (2010). Structural insights into the assembly and function of the SAGA deubiquitinating module. Science 328, 1025-1029.

Sauerwald, A., Sandin, S., Cristofari, G., Scheres, S.H., Lingner, J., and Rhodes, D. (2013). Structure of active dimeric human telomerase. Nat Struct Mol Biol 20, 454-460.

Scheres, S.H. (2010). Classification of structural heterogeneity by maximum-likelihood methods. Methods Enzymol 482, 295-320.

Spedale, G., Mischerikow, N., Heck, A.J., Timmers, H.T., and Pijnappel, W.W. (2010). Identification of Pep4p as the protease responsible for formation of the SAGA-related SLIK protein complex. J Biol Chem 285, 22793-22799.

Sterner, D.E., Grant, P.A., Roberts, S.M., Duggan, L.J., Belotserkovskaya, R., Pacella, L.A., Winston, F., Workman, J.L., and Berger, S.L. (1999). Functional organization of the yeast SAGA complex: distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction. Mol Cell Biol 19, 86-98.

Syed, S.H., Goutte-Gattat, D., Becker, N., Meyer, S., Shukla, M.S., Hayes, J.J., Everaers, R., Angelov, D., Bednar, J., and Dimitrov, S. (2010). Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome. Proc Natl Acad Sci U S A 107, 9620-9625.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., et al. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell 142, 967-980.

Warfield, L., Ranish, J.A., and 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 Dev 18, 1022-1034.

Washburn, M.P., Wolters, D., and Yates, J.R., 3rd (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19, 242-247.

Winston, F., and Sudarsanam, P. (1998). The SAGA of Spt proteins and transcriptional analysis in yeast: past, present, and future. Cold Spring Harb Symp Quant Biol 63, 553-561.

Wu, P.Y., Ruhlmann, C., Winston, F., and Schultz, P. (2004). Molecular architecture of the S. cerevisiae SAGA complex. Mol Cell 15, 199-208.

Zybailov, B., Mosley, A.L., Sardiu, M.E., Coleman, M.K., Florens, L., and Washburn, M.P. (2006). Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J Proteome Res 5, 2339-2347.

Figure legend

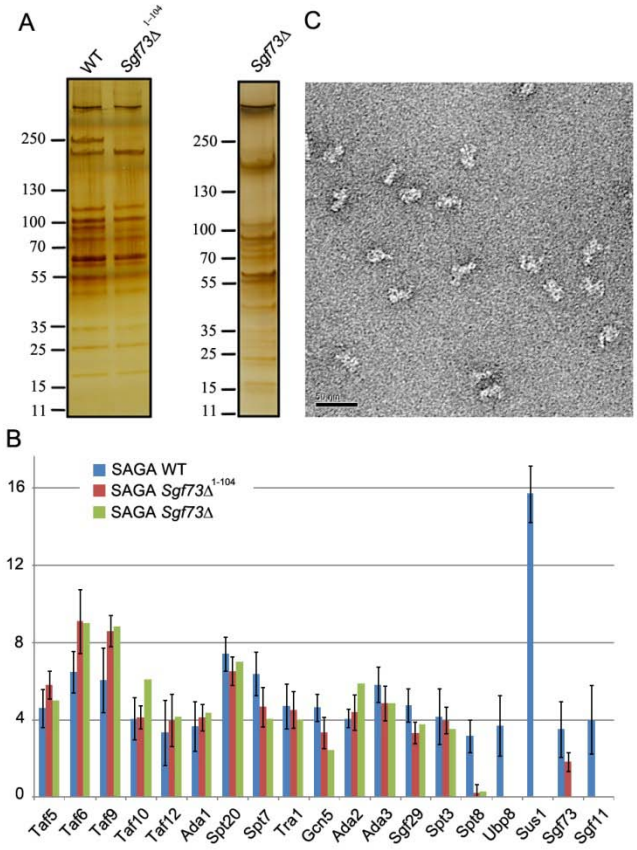


Figure 1: Specimen Characterization.

(A) Silver stain SDS-PAGE analysis of the subunit composition of SAGA complexes purified from the Spt20-TAP tagged wild-type strain (BY4742) (WT), Sgf73Δ$^{1\text{-}104}$ and ΔSgf73. (B) Proteomic analysis of the SAGA complexes purified from the WT (blue), Sgf73Δ1-104 (red) and ΔSgf73 (green) strains. (C) Electron micrograph of negatively stained wild-type SAGA molecules. Scale bar represents 50 nm.

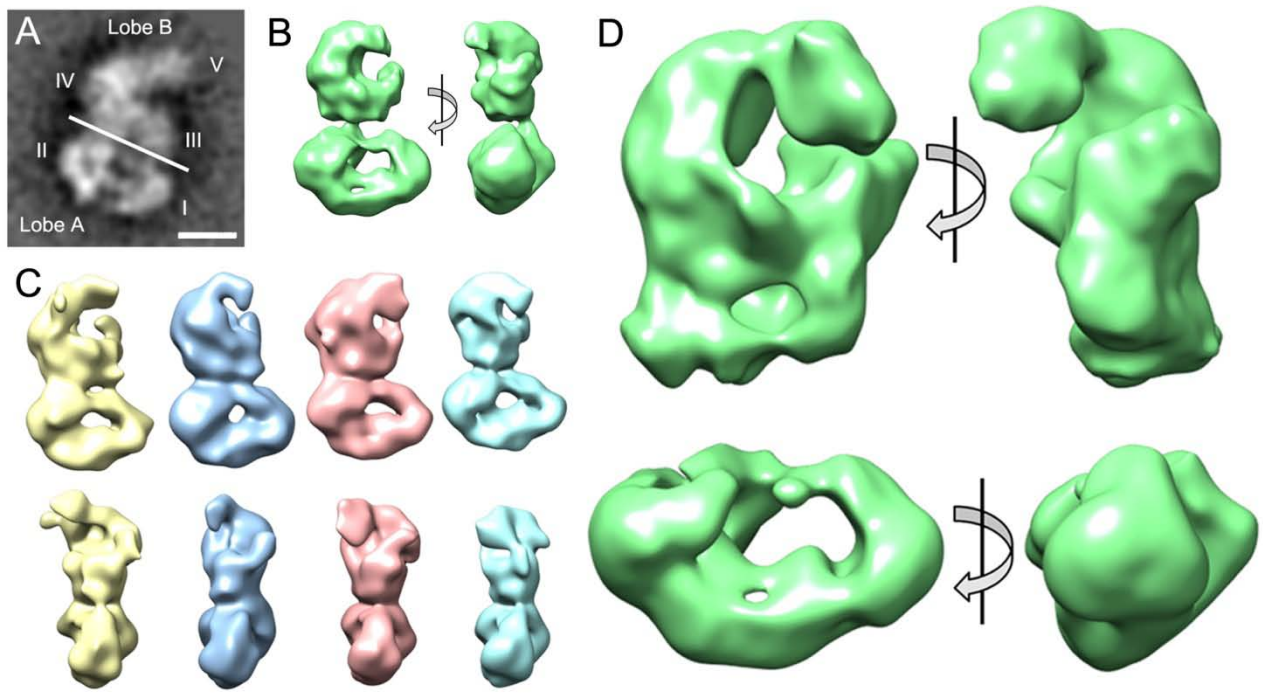


Figure 2: Three-dimensional model of wild-type SAGA.

(A) Representative SAGA view revealing Lobe A composed of domains I and II, and Lobe B formed by domains III, IV and V. (B) Three-dimensional reconstruction of the WT SAGA complex. (C) Separation of different conformations of the SAGA complex by maximum likelihood 3-D classification. The molecular clamp formed by lobe B can adopt a fully closed (cyan) or opened (yellow) conformations. (D) Three-dimensional model of lobes B (top) and A (bottom) analyzed independently. The bar represents 10.6nm in (A) and (C), 9.4 nm in (B) and 3.6 nm in (C).

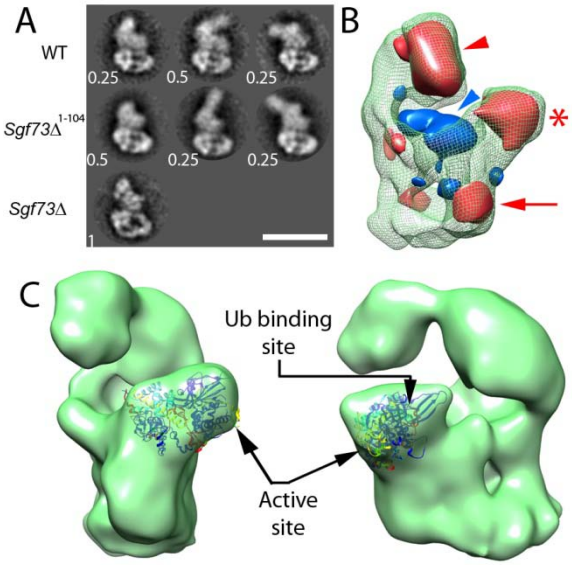


Figure 3: Position of the DUB module

(A) Characteristic SAGA views clustered according the position of the flexible domain V in wild-type (WT), *Sgf73*Δ$^{1-104}$ and *Sgf73*Δ preparation. The frequency of each class is indicated (B) density difference map between WT and *Sgf73*Δ SAGA maps superimposed to the WT map (green mesh). Positive differences (absent in *Sgf73*Δ) are represented in red and negative differences (absent in WT) in blue. (C) Fitting of the atomic structure of the DUB module into the WT SAGA model. The bar represents 35.6nm in (A), 5nm in (B) and 4.2 nm in (C).

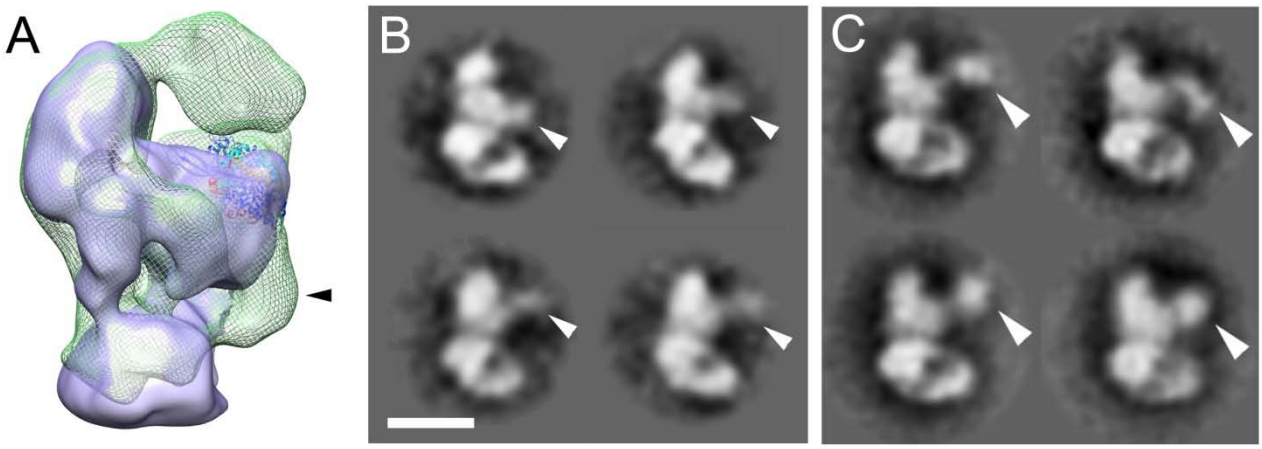


Figure 4: Location of Spt8 and nucleosome binding

(A) Superimposition of the WT (green mesh) and the *Sgf73*Δ (blue) SAGA maps. The density difference not explained by the movement of domain V or the DUB module is indicated by an arrow head. (B) Spt8-specific polyclonal antibody labeled a site within domain III (arrow heads). (C) Gallery of nucleosome core particle-bound SAGA molecules. The bar represents 5.8 nm in (A) and 19.3 nm in (B) and (C).

Supplemental experimental procedures

Yeast strains

Genomic integration of a TAP-tag to the C-terminal of Spt20 was performed in a BY4742 background (MATα ; his3-Δ1; leu2-Δ0; lys2-Δ0; ura3-Δ0). The Sgf73 gene was deleted in the Spt20-TAP strain, and different Sgf73 constructs were cloned into the pRS315 vector, fused at the C-terminus to a 3xHA and a Myc tag, and expressed under the control of their natural promoter. The following strains were used: Spt20-TAP (Isogenic to BY4742 except Spt20-TAP::URA3, Bonnet et al., 2010), YGL066W (Isogenic to BY4742 except sgf73Δ::Kan; Euroscarf), *Sgf73*$\Delta^{1\text{-}104}$ (Isogenic to YGL066W except Spt20-TAP::URA3; pRS315-*sgf73*Δ2-104-3HA-Myc, this study), *Sgf73*Δ (Isogenic to YGL066W except Spt20-TAP::URA3; pRS315-*sgf73*Δ; Bonnet et al., 2010)

Culture and purification

Cells were grown in flask to an OD600 of 2-2.5 in YPD at 30°C. The tandem affinity purification was performed as described (Rigaut et al., 1999). Briefly, 20g of yeast cells were suspended in 30 mL of lysis buffer (Hepes 50 mM pH 7.5, NaCl 300 mM, $MgCl_2$ 2 mM, Glycerol 10%, Tween20 0.1%) with protease inhibitor (Complete, Roche). Lysis was perfomed by vortexing the cell suspension with glass beads for 5x 1 minute and cooling the cells 1 minute between each run. Lysate was clarified by centrifugation at 40000g for 1 hour. For affinity purification, 200 µL of IgG sepharose beads were incubated for 6 hours with the cell extract. Beads were washed in 5x with 3 mL of TEV cleavage buffer (Hepes 30 mM pH 7.5, NaCl 300 mM, Tween 20 0.1%). Cleavage was performed overnight at 4°C with 3-5 µg of TEV in 400 µL. Supernatant was collected and beads were washed with 400µL of cleavage buffer. For the second step, 100 µL of calmodulin sepharose beads were incubated with the 800µL elutions for 2 hour at 4°C. Beads were washed 5x with 1 mL Hepes 30 mM pH 7.5,

NaCl 500 mM, $CaCl_2$ 2 mM, and eluted with 100 µL Hepes 30 mM pH 7.5, NaCl 500 mM, EGTA 2 mM.

Mudpit analysis

Proteins were precipitated with TCA. Pellets were washed with cold acetone, denaturated with urea, reduced and alkylated. Proteins digestion was performed with endoproteinase LysC (Roche) and trypsin. Resulting peptides were loaded on 100 µm fused silica microcapillary columns (Polymicron) packed of 2 cms with 5 µm C18 AQUA (Phenomenex) reverse phase, 4 cms of strong cation exchange particles (partisphere SCX, Whatman) and 9 cms of reverse phase. Loaded and washed microcapillaries were placed in-line with a nanoLC Ultimate 3000 (Thermo Scientific) coupled with LTQ Velos linear ion trap mass spectrometer equipped with a nano-LC electrospray ionization source (Thermo Fisher Scientific). A 12-steps MudPIT run was performed as described (Florens et al., 2006). Each full MS scan (from 300 to 1700 m/z range) was followed by 20 MS/MS events using data-dependent acquisition, where the 20 most intense ions from a given MS scan were subjected to Collision Induced Dissociation (CID). Proteins were identified using SEQUEST against a Swissprot database (S. cerevisiae -Strain ATCC 204508-S288c 2012-09-15). Peptides were filtered with FDR 1 %, peptides rank 1 and peptides of at least 7 amino acids in length. Relative protein abundance for each protein in a given sample was estimated by Normalized Spectral Abundance Factor (NSAF) from the spectral counts of each identified protein (Zybailov et al., 2006). To account for the fact that larger proteins tend to contribute more peptide/spectra, spectral counts were divided by protein length to provide a Spectral Abundance Factor (SAF). SAF values were then normalized against the sum of all SAF values in the corresponding run, allowing the comparison of protein levels across different runs.

Electron microscopy and image analysis

The purified WT or mutant SAGA fractions were diluted to a concentration of 20 µg/ml in Hepes 30 mM (pH 7.5), NaCl 100 mM and the complexes were cross-linked with glutaraldehyde (0.2% final concentration) for a few seconds at room temperature. Five microliters of this preparation were placed on a thin carbon film treated by a glow discharge in air. After 1 min. of adsorption grids were negatively stained with a 2% (w/v) uranyl acetate solution and air-dried. Rabbit polyclonal antibodies against Spt8 were generated and purified as described (Sanders and Weil, 2000). For immuno-EM a 3- to 5-fold molar excess of anti-Spt8 IgG was incubated 30 min at 4°C with WT SAGA at a final protein concentration of 20 µg/ml. IgG-labeled SAGA images were collected and aligned against projections of the 3-D model of WT SAGA. The aligned images were then partitioned into classes to better identify the antibody bound to SAGA. Images were recorded on a Transmission Electron Microscope (Tecnai F20 G2, FEI) equipped with a field emission gun operating at 200 kV at a magnification of 45.000 x, on a 2048x2048 CCD camera (Ultrascan 1000, Gatan Inc., Pleasanton) with a final pixel spacing of 0.26 nm. Particles were selected manually using EMAN (Ludtke et al., 1999). Image processing was performed using Imagic (Image Science Software, Berlin, Germany) as described (Wu et al., 2004). Angular assignment was performed by projection matching, using a low-pass filtered map of the published structure (Wu et al., 2004). To analyze both lobes separately, coordinates of each part of the complex were determined manually and images from lobes A and B were extracted from the aligned dataset. Initial volumes for each lobe were extracted from the 3D model of the full complex and used for reference matching.

Bibliography

Florens, L., Carozza, M.J., Swanson, S.K., Fournier, M., Coleman, M.K., Workman, J.L., and Washburn, M.P. (2006). Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. Methods *40*, 303-311.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol *128*, 82-97.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol *17*, 1030-1032.

Sanders, S.L., and Weil, P.A. (2000). Identification of two novel TAF subunits of the yeast Saccharomyces cerevisiae TFIID complex. J Biol Chem *275*, 13895-13900.

Wu, P.Y., Ruhlmann, C., Winston, F., and Schultz, P. (2004). Molecular architecture of the S. cerevisiae SAGA complex. Mol Cell *15*, 199-208.

Zybailov, B., Mosley, A.L., Sardiu, M.E., Coleman, M.K., Florens, L., and Washburn, M.P. (2006). Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J Proteome Res *5*, 2339-2347.

Durand et al. Journal of Nanobiotechnology 2013, 11(Suppl 1):S4
http://www.jnanobiotechnology.com/content/11/S1/S4

JOURNAL OF NANOBIOTECHNOLOGY

TUTORIAL Open Access

Structure, assembly and dynamics of macromolecular complexes by single particle cryo-electron microscopy

Alexandre Durand, Gabor Papai, Patrick Schultz*

From Nanophysics for Health
Mittelwhir, France. 5-9 November 2012

Abstract

Background: Proteins in their majority act rarely as single entities. Multisubunit macromolecular complexes are the actors in most of the cellular processes. These nanomachines are hold together by weak protein-protein interactions and undergo functionally important conformational changes. TFIID is such a multiprotein complex acting in eukaryotic transcription initiation. This complex is first to be recruited to the promoter of the genes and triggers the formation of the transcription preinitiation complex involving RNA polymerase II which leads to gene transcription. The exact role of TFIID in this process is not yet understood.

Methods: Last generation electron microscopes, improved data collection and new image analysis tools made it possible to obtain structural information of biological molecules at atomic resolution. Cryo-electron microscopy of vitrified samples visualizes proteins in a fully hydrated, close to native state. Molecular images are recorded at liquid nitrogen temperature in low electron dose conditions to reduce radiation damage. Digital image analysis of these noisy images aims at improving the signal-to-noise ratio, at separating distinct molecular views and at reconstructing a three-dimensional model of the biological particle.

Results: Using these methods we showed the early events of an activated transcription initiation process. We explored the interaction of the TFIID coactivator with the yeast Rap1 activator, the transcription factor TFIIA and the promoter DNA. We demonstrated that TFIID serves as an assembly platform for transient protein-protein interactions, which are essential for transcription initiation.

Conclusions: Recent developments in electron microscopy have provided new insights into the structural organization and the dynamic reorganization of large macromolecular complexes. Examples of near-atomic resolutions exist but the molecular flexibility of macromolecular complexes remains the limiting factor in most case. Electron microscopy has the potential to provide both structural and dynamic information of biological assemblies in order to understand the molecular mechanisms of their functions.

Background

Genomic sequences are now available for many different organisms which, when combined with biocomputing analysis result in the annotation of most of the coding regions that define the protein repertoire of the living creature. Systematic protein purification experiments revealed that proteins act rarely as single entities but are generally associated into well-defined complexes, 80% of which contain between 5 and 12 distinct proteins [1]. Interestingly, several proteins show some degree of infidelity and can be found in distinct complexes. Moreover the documented complexes correspond only to the most stable molecular interactions that resist the harsh protein purification conditions. Many more transient interactions are likely to occur between proteins and protein

* Correspondence: patrick.schultz@igbmc.fr
Integrated Structural Biology Department, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), U964 Inserm F-67400, UMR7104 CNRS, Université de Strasbourg, 1 Rue Laurent Fries, BP10142, 67404 Illkirch, France

complexes to build up the intricate and robust molecular interaction network that governs cell fate.

Macromolecular complexes are therefore at the center of most biological processes. They integrate spatially several catalytic or structural activities with built-in regulatory functions. In most of the cases, conformational changes that range from atomic to molecular scale are instrumental to explain the function of these complexes. Altogether these dynamic properties, associated with the size of the particles ranging between 10 and 40 nm substantiates the name of nanomachines often attributed to these complexes. These nanomachines are targeted by most of the currently available drugs used to cure human diseases but for their vast majority the drugs inhibit a catalytic activity carried by a single subunit. Only in rare occasions the intrinsic mechanical properties or the specific protein-protein interaction network of a complex is targeted by drugs. The ribosome is one of such nanomachines, responsible for protein synthesis and for which several examples of drugs targeting the mechanical properties are at hand [2]. Macrolydes and other antibiotics affect the translocation of the ribosome along the mRNA and thus inhibit protein synthesis. Fusidic acid was shown to prevent the dynamic turnover of the elongation factor G and thus affects the interaction of the ribosome with this regulatory factor. Finally antibiotics such as Dalfopristin or Quinopristin were found to bind to the ribosome exit channel and to block mechanically the progression of the nascent polypeptide. Few other examples of drugs targeting so clearly the intrinsic mechanical properties of a complex were described so far. This is related to the poor structural information available to date on complexes since most of the atomic structures deposited in the protein data bank are single polypeptides.

This tutorial aims at describing the molecular organization of the general transcription factor TFIID as a paramount multi-protein complex and to emphasize the role of cryo-electron microscopy (cryo-EM) and digital image analysis to integrate structural and functional information in order to reach a mechanistic model of the complex.

Methods

Cryo-EM of frozen hydrated molecular complexes

Imaging of single particles by electron microscopy and numerical analysis of image datasets have proven invaluable tools to describe the structural organization of large macromolecular assemblies. Since the discovery of negative stain by Brenner and Horne in 1959, single particles embedded in a layer of heavy atom salts can be visualized through the high contrast provided by the electron-dense material that surrounds the biological macromolecule composed of low atomic number elements, which poorly scatter electrons (Figure 1) [3]. Despite its ability to provide high contrast, to reveal fine structural details and to sustain fragile structures, the negative staining approach is limited to the description of surface features and rarely extends beyond 15-20 Å resolution.

A major breakthrough was achieved by the discovery in the early 1980' of a robust specimen preparation method that preserved specimen hydration in the vacuum of the electron microscope [4,5]. The method relies on the fast vitrification of a thin aqueous layer containing the specimen by plunging into a liquid ethane slush (Figure 1). This procedure prevents ice crystal formation that segregates particles and ruins image quality. The frozen hydrated sample has to be observed at low temperature, typically close to liquid nitrogen temperature, to prevent phase transitions and special cold stages were developed for cryo-EM observations. This groundbreaking technology opened new horizons for the observation of macromolecular complexes. It allows unconstrained particle conformations in the absence of any crystal contacts and in close to physiological ionic strength and pH conditions. In contrast to crystallized conditions, in which a particular conformation is selected, a flexible particle will be able to adopt all permitted conformations. Conformational flexibility may be detrimental for structure determination since fine structural details may be averaged out, but cryo-EM

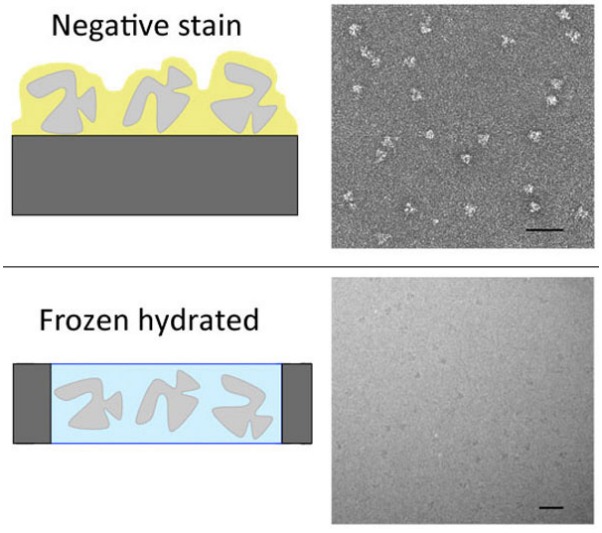


Figure 1 Preparation of purified molecular complexes for electron microscopy. In negative stain the specimen is adsorbed on a carbon film, embedded in a layer of heavy metal salts and dried. In frozen hydrated conditions the molecules are embedded in a thin layer of vitrified buffer suspended in a hole of the carbon film. Corresponding electron micrographs are shown (right panels). The bar represents 50 nm.

records conformational intermediates and thus holds the promise to detect and describe particle dynamics. Early electron diffraction experiments showed that in such frozen hydrated conditions, the structure of the specimen is preserved down to atomic resolutions, thus showing for the first time that electron microscopy images have the potential to reveal the same structural information than X-ray diffraction [6].

The resolving power of modern electron microscopes is sufficient to image single atoms. In material sciences, the specimen is very stable and a huge number of electrons can interact with the sample often without affecting its structure. As a result, individual atoms can be detected with a good statistical significance or signal to noise ratio (SNR). Imaging of biological samples fully benefit from the most recent developments in instrumentation such as field emission guns which give a much brighter and more coherent electron beam, detectors and microscope automation. Specific instrumentation is needed to observe frozen hydrated samples which includes cold stages to keep the specimen temperature below -170°C, anti-contamination devices to prevent deposition of traces of water present in the microscope onto the cold specimen as well as low-dose imaging protocols to avoid irradiation of the specimen before data acquisition. Structural damage induced by the electron beam is a strongly limiting factor for biological specimen. Inelastic interactions of incident electrons with sample atoms dissipate energy that can break covalent bonds and generate highly reactive side chains. It is generally accepted that the atomic structure of the specimen is preserved when electron doses are kept below 5 electrons per square angstrom ($e^-/Å^2$), however this number varies with the acceleration voltage of the electrons - at 300 kV it can be up to 25 $e^-/Å^2$ [7]. In these conditions the molecular images are so noisy that the fine structural details cannot be detected. As a rule of thumb, at an electron dose of 5 $e^-/Å^2$, details in the range of 50 Å can be detected with a SNR of two while smaller details are below this detection limit. To reconcile low specimen irradiation which leads to noisy images, with a high SNR objective to detect small details, it is necessary to split the dose required to detect atoms (say 2000 $e^-/Å^2$) over several independent particles (in this case 400) to kept the dose below 5 $e^-/Å^2$ and to add-up the signal coming from all these images.

The ongoing development of highly sensitive direct detection cameras and single electron counting devices are important to record highly enlarged images of biological complexes with the best quantum detection efficiency and with reduced noise [8]. Automation of cameras and electron microscopes facilitates the recording of several hundreds to thousands of images per day each containing 100-200 particles thus generating huge image datasets which, as we will describe, will be of importance to reach the final spatial resolution [9]. The need for reduced electron irradiation also led to dedicated "low dose" data acquisition strategies in which microscope adjustments such as focus and astigmatism corrections are performed on an area remote of the area of interest to be recorded "blindly".

Single particle image analysis

The objectives of single particle image analysis are dual [10]. The first goal is to improve the SNR of the original images by averaging the signal from independent particles. Assuming a Gaussian distribution of all sources of noise that can affect the original molecular image (statistics of electron-matter interaction, detector noise, cosmic rays, etc...), image averaging will improve the SNR and increase the spatial resolution that can be detected. The second goal of image analysis is to reach a volumetric description of the sample. Standard transmission electron microscopes provide 2-D projections of the 3-D electron density map of the sample, multiplied by a microscope-specific Contrast Transfer Function, which has to be corrected for. The objective is to determine the projection (or viewing) direction of each 2-D image with respect to the 3-D object it originates from and to reconstruct a 3-D model by combining many 2-D views. A brief overview of the image analysis protocol is shown in Figure 2.

Alignment and clustering

Images of a same particle can be averaged to improve the SNR only if two criteria are met: firstly they have to correspond to the same view or projection of the particle and secondly the images have to be in the same register, or in other words aligned in translation and in rotation one with respect to the others. The spatial resolution that can be reached will depend on the number of images that can be averaged, on how similar the views are, and on the alignment quality. If a tolerance of 10° in viewing direction is accepted, the finest dimension that can be resolved for a globular particle with 10 nm in diameter cannot be smaller than 5.sin10° = 0.8 nm. To reach a resolution of 0.2 nm the variation in viewing direction cannot be larger than 2.2°.

A molecular image is aligned against a reference image by correlating image intensities. The correlation coefficient between two images is a measure of their similarities and all possible translations and rotations will be explored to find the correlation maximum, which will be considered as their best alignment. The quality of the alignment depends on many parameters such as the initial SNR of the image, the size and the shape of the particle.

In a real image data set, the particles have different orientations that lead to different views that need to be

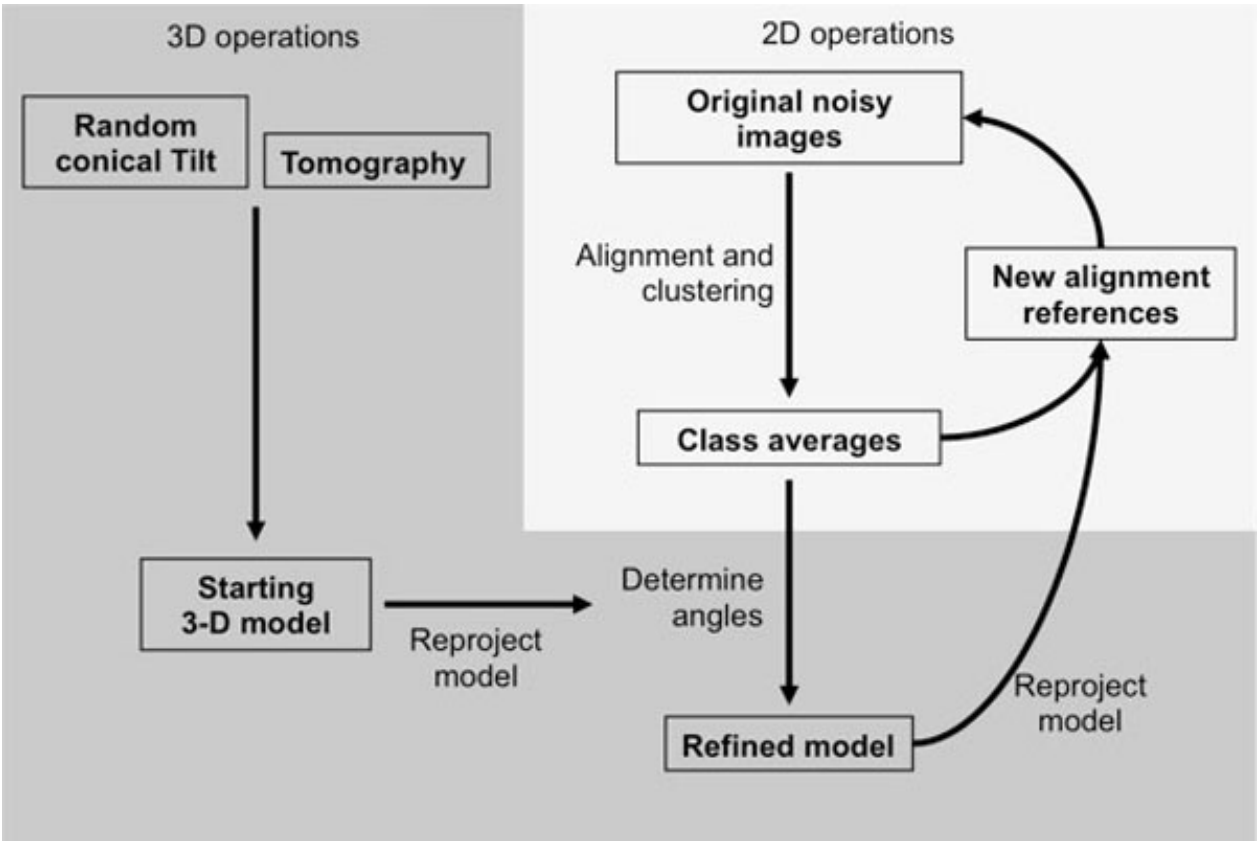


Figure 2 Schematic representation of the key steps in the analysis of single particle images.

separated before calculating an average image. The images need therefore to be clustered into groups containing the most similar images. The image intensity variance should be minimized within the same group, while it should be maximized between different groups. In practical terms the image data set is first subjected to a multivariate statistical analysis (Principal Component Analysis or Correspondence Analysis) to detect the most meaningful trends in the data set and the clustering is then performed on the most significant Eigenvectors using Hierarchical Ascendant Classification schemes.

In an ideal image data set, the particles are randomly oriented which will produce an infinite number of projections. This condition is not always met when particles are adsorbed on a supporting carbon film, which may lead to preferred orientations. Nevertheless, the number of different orientations is very large and it is virtually impossible to find two perfectly identical particle images. It is therefore important to consider an angular projection sector within which we consider the images to be identical at a defined spatial resolution. An image class can to a first approximation be considered as a group of molecular images viewed along the same angular sector. If we consider the above mentioned 10 nm globular particles, a 10° tolerance in projection angle will result in an uncertainty of 0.8 nm. A projection sector of 10° leads to 244 different views and the dataset should be separated in as many classes.

The alignment and clustering steps are highly interdependent and will be used iteratively to improve the quality of the class averages. A better alignment will lead to an improved clustering which will impact the resolution of the class averages. Such high resolution class averages will further improve the alignment of the original images in a multi-reference alignment protocol.

Three-dimensional model building

The class averages correspond to distinct views of the particle but their projection direction is not known *a priori*. A common-line based method was designed to attribute the relative projection directions of a set of class averages [11], but this method may lead to ambiguous results especially when several conformations of the particle coexist. Two experimental methods, based on the acquisition of tilted images of the same particle have been developed.

In the tomography approach a goniometric electron microscopy stage is used to record tilted views of the same object, typically between +70° and -70° with angular increments of 1 or 2° [12]. After alignment of the images on a common origin, a 3-D model can be calculated for each particle by combining all views for which the exact projection direction is experimentally determined by the position of the tilt axis and the tilt angle. This method suffers from several drawbacks that have been partially addressed. Electron dose and therefore radiolytic damage accumulates during the sequential acquisition of around 140 images of the same particle, but the development of very sensitive low noise cameras restricts the total accumulated dose to 20 to 40 $e^-/Å^2$, enough to reach a resolution of 3-4 nm. The data collection scheme produces a wedge of missing projections and this problem can be overcome by turning the grid 90° in plane and by recording a second tilt series. The missing wedge will then be reduced to a missing pyramid and the quality of the reconstructed volume is generally improved. Alternatively independently reconstructed single particle volumes can be aligned in 3-D and averaged. Since for each orientation of the particle the missing information is different, the averaged volume is essentially devoid of missing wedge artifacts.

In the random conical tilt method, the data collection strategy consists in recording first a 45-60° tilted view of an electron microscopy field containing several particles and, in a second exposure, an untilted view of the same field [13]. The untilted images are analyzed as single particles thus producing classes containing several images of similarly oriented particles each differing by their in-plane or azimuthal angle. This angle, the position of the tilt axis and the tilt angle, informs about the viewing direction of each corresponding tilted image and allows calculating a 3-D model for each class of images. With this data collection strategy, irradiation is limited to a single exposure and the missing information is restricted to a cone.

Model refinement

The experimental 3-D models are considered as low resolution "starting models" that will be used to determine the viewing direction of independently determined class averages obtained from a much larger image dataset. The starting models will be computationally "reprojected" along many directions to generate a set of reference images of known projection direction. The subsequent alignment of the class averages, or of the original images, against these reprojections in a process called reference-matching will determine the viewing direction for each high resolution class average and lead to an improved, or refined, 3-D model.

Address the dynamic properties of the complexes

The fast vitrification of the specimen in liquid ethane preserves the hydrated state of the protein complexes, but also cryo-immobilizes their different conformational states. This heterogeneity can hinder high-resolution structure refinement if different conformations are combined in a single class; however it contains essential information about the dynamic properties of the sample. For isolated particles, different conformations can be sorted out computationally when the data set is large enough, thus providing information on mobile parts of the complex. For transient multi-component systems, the relative abundance of the components present in an equilibrium state informs about the interaction constants. It is therefore crucial to detect and separate the conformational states of the specimen both to improve the resolution of each individual state and to describe the dynamics of the examined protein complex. Several methods exist to detect and visualize *de novo* structural heterogeneities in the specimen [14]. Rough movements of domain can be detected by either single-particle tomography or random conical-tilt experiments. More subtle differences can be tracked by using Eigen-analysis of resampled cryo-EM images [15,16]. In this method the images dataset has to be aligned to an average reference structure to determine the relative particle orientation. A large number of volumes is built from a randomly created subset of the dataset and these volumes are subjected to multivariate statistical analysis followed by hierarchal classification to identify the structural differences.

Results

The general transcription factor TFIID

Gene expression programs in metazoans are under tight control to achieve growth, development and differentiation of the tissues that make up living organism. A large extend of regulation is performed at the transcriptional level when the information carried by specific DNA sequence (the genes) is transcribed into a messenger RNA molecule (mRNA) by the RNA polymerase II enzyme. Misregulated gene expression underlies many human pathologies, as indicated by germ-line and somatic mutations in transcription regulatory genes that lead to genetic disease [17-20], developmental syndromes[21], neurological diseases [22], epigenetic perturbations [23] and cancer [21,24]. Most intensively studied is the initiation step, which determines which genes are turned on to express a specific piece of genetic information in response to external signaling events. Initiation of transcription is controlled by a large number of multiprotein complexes whose action results in the assembly of a transcription Preinitiation Complex (PIC) on the promoter DNA upstream of the coding sequence. The ultimate goal of the PIC is to position the RNA polymerase II at the transcription start site and to initiate the synthesis of mRNA [25,26].

The general transcription factor TFIID is a key player in the initiation process since it is the first factor to interact with promoter DNA and directs the following steps that result in the onset of transcription [27]. This 1 MDa TFIID multiprotein complex contains a protein recognizing the TATA-box in the gene promoter (TBP) and 13 TBP Associated Factors (TAFs) whose sizes vary between 10 and 250 kDa. To modulate gene expression in response to external signals, small activator or repressor proteins bind upstream of gene promoters and recruit the transcriptional coactivators and the general transcriptional machinery. In this process human TFIID not only recognizes the promoter DNA region of genes but also acts as a transcriptional coactivator by interacting with several such activators like p53, Sp1 and c-Jun [28]. TFIID thus acts as a bridge between transcriptional activator proteins and the PIC.

Structure of TFIID, a hybrid approach

How is gene transcription initiated and what is the role of activators and co-activators in this process? How is the activation signal transmitted from the activator to the general transcription factors and finally to the RNA polymerase? How do cells integrate and respond to regulatory signals? What sets different gene expression levels in specific cell types? Which errors in this process lead to disease? To answer these questions the functions of TFIID have to be explained in mechanistic details by determining its biophysical and structural properties.

A large body of structural data is available at atomic scale for single subunit TAF domains and small TFIID sub-assemblies such as the TATA box binding protein [29], the histone-fold containing TAF heterodimers [30-32], the N-terminus of TAF5 [33], The HEAT repeats of TAF6 [34] or the TAF1 bromodomain [35]. Single crystal X-ray diffraction or NMR spectroscopy have determined the atomic structures of parts of TFIID that sum up to about 40% of its total mass but little is known about the organization of these bricks into a functional TFIID assembly that is active in transcription. The full complex is reluctant to crystallization since this large multisubunit complex is difficult to produce in large quantities and with purity suitable for crystal growth. This observation is general to most fields of biology and led to the development of so-called hybrid methods that integrate structural information from different sources. Cryo-electron microscopy is instrumental to this integration since it provides medium (10-20 Å) resolution maps of large complexes into which atomic-scale information obtained by X-ray crystallography or NMR spectroscopy can be fitted [36].

Low-resolution studies by negative-stain [37,38] and, more recently cryo-EM [39] have revealed the general shape of TFIID and allowed approximate localization of several subunits by means of antibody labeling [40,41]. Samples used in these studies were prepared from endogenous sources and resulted in spatial resolutions that were seriously hampered by the dynamic properties, heterogeneous nature and the low abundance of material. The lack of recombinant TFIID complexes of suitable quality and quantity for molecular level studies has been an insurmountable bottleneck to date for structural but also for functional studies. Recent developments in recombinant protein production were instrumental for solving the structure of the core-TFIID complex at 11 Å resolution most probably because several sources of heterogeneity arising from TAF isoforms and posttranslational modifications were reduced [42].

To gain insights into the function of TFIID, the interaction of yeast TFIID with the promoter DNA was studied in the presence of TFIIA (a general transcription factor required for specific recognition of the TATA element) and the Rap1 activator [43]. The Cryo-EM results revealed the network of interactions and the conformational changes occurring during complex formation. The path of DNA was detected in the complex and these findings extended our understanding on the DNA recognition modalities by TFIID in the presence of trans-acting factors. The resulting structure has shed new light on the intramolecular communication pathways conveying transcription activation signals through the TFIID coactivator (Figure 3). This study revealed an interaction between TFIIA and Rap1 that form a protein bridge between TBP and the DNA-bound Rap1 which results in a large change in the position of TFIIA and of TBP. Interestingly, the concomitant binding of promoter DNA to TFIID-bound Rap1 and to TBP loops out the intervening DNA, thereby accommodating variable distances between Rap1 binding sites and transcription start site.

Conclusions

The development of cryo-EM and image analysis software has provided new insights into the structural organization and the dynamic reorganization of large macromolecular complexes. Recent improvements in electron microscopy instrumentation allow for automated processing and recording of large image datasets, with an improved image quality due to more stable cold stages and advanced electron optics. With these developments unprecedented close to atomic resolutions were obtained for highly symmetric biological assemblies such as icosahedral viruses. It can be anticipated that the analysis of large datasets as well as new data acquisition strategies that compensate for particle movement during acquisition, will routinely provide molecular models better than 5A in the near future. The analysis of molecular flexibility still requires algorithmic

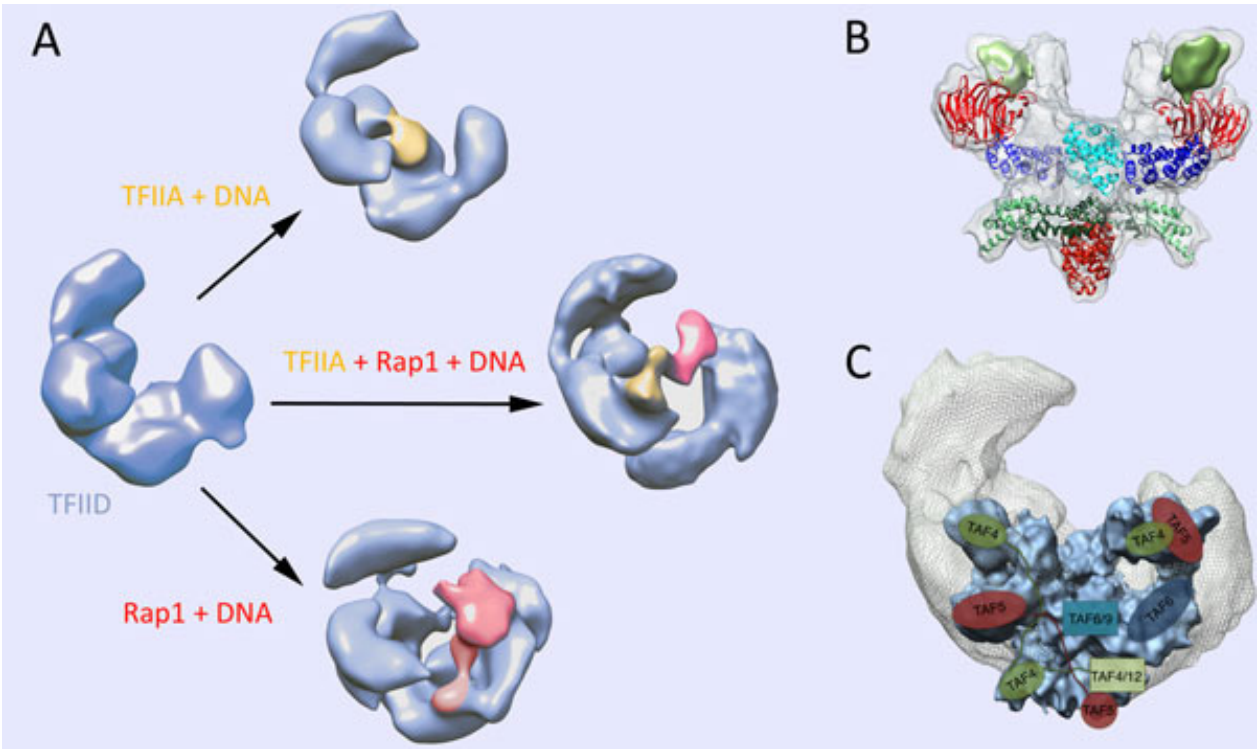


Figure 3 Structure of TFIID. (A) Structure of TFIID and of its complexes with TFIIA, DNA and the Rap1 activator. (B) Structure of the core-TFIID at 11.4 Å resolution and fitting of the known atomic structures. (C) Position of the core-TFIID within the complete TFIID complex. The bar represents 7.6 nm in (A), 5 nm in (B) and 5.8 nm in (C)

developments to describe concomitantly the high resolution structure and the continuous conformational space of a macromolecular complex. The unique asset of Cryo-EM however resides in the possibility to record images of single particles which collectively contain both structural and dynamic information.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
P.S. initiated the study. GP performed the electron microscopy and image analysis experiments. The manuscript was prepared and commented on by P.S., A.D. and G.P.

Acknowledgements
This work was supported by grants from the Institut National de la Santé et de la Recherche Médicale, the Centre National pour la Recherche Scientifique, the Université de Strasbourg, the Association pour la Recherche sur le Cancer, the Fondation pour la Recherche Médicale and the Agence Nationale pour la Recherche. This work was funded by Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI) and supported by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01.
This article has been published as part of *Journal of Nanobiotechnology* Volume 11 Supplement 1, 2013: Nanophysics for Health. The full contents of the supplement are available online at http://www.jnanobiotechnology.com/supplements/11/S1. Publication charges for this tutorial were funded by the CNRS School "Nanophysics for Health", 5 - 9 November 2012, Mittelwhir, France

Published: 10 December 2013

References
1. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, *et al*: **Functional organization of the yeast proteome by systematic analysis of protein complexes.** *Nature* 2002, **415(6868)**:141-147.
2. Wilson DN: **On the specificity of antibiotics targeting the large ribosomal subunit.** *Ann N Y Acad Sci* 2011, **1241**:1-16.
3. Brenner S, Horne RW: **A negative staining method for high resolution electron microscopy of viruses.** *Biochim Biophys Acta* 1959, **34**:103-110.
4. Lepault J, Booy FP, Dubochet J: **Electron microscopy of frozen biological suspensions.** *J Microsc* 1983, **129(Pt 1)**:89-102.
5. Dubochet J, Adrian M, Chang JJ, Homo JC, Lepault J, McDowall AW, Schultz P: **Cryo-electron microscopy of vitrified specimens.** *Q Rev Biophys* 1988, **21(2)**:129-228.
6. Taylor KA, Glaeser RM: **Electron diffraction of frozen, hydrated protein crystals.** *Science* 1974, **186(4168)**:1036-1037.
7. Glaeser RM: **Retrospective: radiation damage and its associated "information limitations".** *J Struct Biol* 2008, **163(3)**:271-276.
8. Henderson R, McMullan G: **Problems in obtaining perfect images by single-particle electron cryomicroscopy of biological structures in amorphous ice.** *J Electron Microsc (Tokyo)* 2013, **62(1)**:43-50.
9. Carragher B, Kisseberth N, Kriegman D, Milligan RA, Potter CS, Pulokas J, Reilein A: **Leginon: an automated system for acquisition of images from vitreous ice specimens.** *J Struct Biol* 2000, **132(1)**:33-45.
10. Orlova EV, Saibil HR: **Structure determination of macromolecular assemblies by single-particle analysis of cryo-electron micrographs.** *Curr Opin Struct Biol* 2004, **14(5)**:584-590.
11. Van Heel M: **Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction.** *Ultramicroscopy* 1987, **21(2)**:111-123.

12. Walz J, Typke D, Nitsch M, Koster AJ, Hegerl R, Baumeister W: **Electron Tomography of Single Ice-Embedded Macromolecules: Three-Dimensional Alignment and Classification.** *J Struct Biol* 1997, **120(3)**:387-395.
13. Radermacher M: **Three-dimensional reconstruction of single particles from random and nonrandom tilt series.** *J Electron Microsc Tech* 1988, **9(4)**:359-394.
14. Leschziner AE, Nogales E: **Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions.** *Annu Rev Biophys Biomol Struct* 2007, **36**:43-62.
15. Simonetti A, Marzi S, Myasnikov AG, Fabbretti A, Yusupov M, Gualerzi CO, Klaholz BP: **Structure of the 30S translation initiation complex.** *Nature* 2008, **455(7211)**:416-420.
16. Penczek PA, Kimmel M, Spahn CM: **Identifying conformational states of macromolecules by eigen-analysis of resampled cryo-EM images.** *Structure* 2011, **19(11)**:1582-1590.
17. Compe E, Egly JM: **TFIIH: when transcription met DNA repair.** *Nat Rev Mol Cell Biol* 2012, **13(6)**:343-354.
18. Fuss JO, Tainer JA: **XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase.** *DNA Repair (Amst)* 2011, **10(7)**:697-713.
19. Helmlinger D, Tora L, Devys D: **Transcriptional alterations and chromatin remodeling in polyglutamine diseases.** *Trends Genet* 2006, **22(10)**:562-570.
20. Matsumoto T, Sakari M, Okada M, Yokoyama A, Takahashi S, Kouzmenko A, Kato S: **The androgen receptor in health and disease.** *Annu Rev Physiol* 2013, **75**:201-224.
21. Fadloun A, Kobi D, Pointud JC, Indra AK, Teletin M, Bole-Feysot C, Testoni B, Mantovani R, Metzger D, Mengus G, *et al*: **The TFIID subunit TAF4 regulates keratinocyte proliferation and has cell-autonomous and non-cell-autonomous tumour suppressor activity in mouse epidermis.** *Development* 2007, **134(16)**:2947-2958.
22. Hashimoto S, Boissel S, Zarhrate M, Rio M, Munnich A, Egly JM, Colleaux L: **MED23 mutation links intellectual disability to dysregulation of immediate early gene expression.** *Science* 2011, **333(6046)**:1161-1163.
23. Duncan EM, Allis CD: **Errors in erasure: links between histone lysine methylation removal and disease.** *Prog Drug Res* 2011, **67**:69-90.
24. Kalogeropoulou M, Voulgari A, Kostourou V, Sandaltzopoulos R, Dikstein R, Davidson I, Tora L, Pintzas A: **TAF4b and Jun/activating protein-1 collaborate to regulate the expression of integrin alpha6 and cancer cell migration properties.** *Mol Cancer Res* 2010, **8(4)**:554-568.
25. Orphanides G, Lagrange T, Reinberg D: **The general transcription factors of RNA polymerase II.** *Genes Dev* 1996, **10(21)**:2657-2683.
26. Roeder RG: **The role of general initiation factors in transcription by RNA polymerase II.** *Trends Biochem Sci* 1996, **21(9)**:327-335.
27. Papai G, Weil PA, Schultz P: **New insights into the function of transcription factor TFIID from recent structural studies.** *Curr Opin Genet Dev* 2011, **21(2)**:219-224.
28. Liu WL, Coleman RA, Ma E, Grob P, Yang JL, Zhang Y, Dailey G, Nogales E, Tjian R: **Structures of three distinct activator-TFIID complexes.** *Genes Dev* 2009, **23(13)**:1510-1521.
29. Nikolov DB, Hu SH, Lin J, Gasch A, Hoffmann A, Horikoshi M, Chua NH, Roeder RG, Burley SK: **Crystal structure of TFIID TATA-box binding protein.** *Nature* 1992, **360(6399)**:40-46.
30. Birck C, Poch O, Romier C, Ruff M, Mengus G, Lavigne AC, Davidson I, Moras D: **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* 1998, **94(2)**:239-249.
31. Werten S, Mitschler A, Romier C, Gangloff YG, Thuault S, Davidson I, Moras D: **Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II) 135) and hTAF12 (hTAF(II)20).** *J Biol Chem* 2002, **277(47)**:45502-45509.
32. Gangloff YG, Pointud JC, Thuault S, Carre L, Romier C, Muratoglu S, Brand M, Tora L, Couderc JL, Davidson I: **The TFIID components human TAF(II)140 and Drosophila BIP2 (TAF(II)155) are novel metazoan homologues of yeast TAF(II)47 containing a histone fold and a PHD finger.** *Mol Cell Biol* 2001, **21(15)**:5109-5121.
33. Romier C, James N, Birck C, Cavarelli J, Vivares C, Collart MA, Moras D: **Crystal structure, biochemical and genetic characterization of yeast and E. cuniculi TAF(II)5 N-terminal domain: implications for TFIID assembly.** *J Mol Biol* 2007, **368(5)**:1292-1306.
34. Scheer E, Delbac F, Tora L, Moras D, Romier C: **TFIID TAF6-TAF9 complex formation involves the HEAT repeat-containing C-terminal domain of TAF6 and is modulated by TAF5 protein.** *J Biol Chem* 2012, **287(33)**:27580-27592.
35. Jacobson RH, Ladurner AG, King DS, Tjian R: **Structure and function of a human TAFII250 double bromodomain module.** *Science* 2000, **288(5470)**:1422-1425.
36. Lander GC, Saibil HR, Nogales E: **EM, crystallography, and beyond.** *Curr Opin Struct Biol* 2012, **22(5)**:627-635, Go hybrid.
37. Brand M, Leurent C, Mallouh V, Tora L, Schultz P: **Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC.** *Science* 1999, **286(5447)**:2151-2153.
38. Andel F, Ladurner AG, Inouye C, Tjian R, Nogales E: **Three-dimensional structure of the human TFIID-IIA-IIB complex.** *Science* 1999, **286(5447)**:2153-2156.
39. Papai G, Tripathi MK, Ruhlmann C, Werten S, Crucifix C, Weil PA, Schultz P: **Mapping the initiator binding Taf2 subunit in the structure of hydrated yeast TFIID.** *Structure* 2009, **17(3)**:363-373.
40. Leurent C, Sanders SL, Demeny MA, Garbett KA, Ruhlmann C, Weil PA, Tora L, Schultz P: **Mapping key functional sites within yeast TFIID.** *EMBO J* 2004, **23(4)**:719-727.
41. Leurent C, Sanders S, Ruhlmann C, Mallouh V, Weil PA, Kirschner DB, Tora L, Schultz P: **Mapping histone fold TAFs within yeast TFIID.** *EMBO J* 2002, **21(13)**:3424-3433.
42. Bieniossek C, Papai G, Schaffitzel C, Garzoni F, Chaillet M, Scheer E, Papadopoulos P, Tora L, Schultz P, Berger I: **The architecture of human general transcription factor TFIID core complex.** *Nature* 2013, **493(7434)**:699-702.
43. Papai G, Tripathi MK, Ruhlmann C, Layer JH, Weil PA, Schultz P: **TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation.** *Nature* 2010, **465(7300)**:956-960.

doi:10.1186/1477-3155-11-S1-S4
Cite this article as: Durand *et al.*: Structure, assembly and dynamics of macromolecular complexes by single particle cryo-electron microscopy. *Journal of Nanobiotechnology* 2013 **11**(Suppl 1):S4.

Abstract

During transcription initiation in eukaryotes, the RNA polymerase II (pol II) and general transcription factors are recruited to the promoter of protein coding genes in order to form a Pre-initiation Complex (PIC), which will position the pol II to the transcription start site. Transcription activators recognize specific DNA sequences upstream of the promoter and stimulate transcription by enhancing PIC formation and chromatin opening through the recruitment of macromolecular complexes, called transcriptional coactivators. The SAGA (Spt-Ada-Gcn5 acetyl transferase) is such a coactivator, conserved in eukaryotes during evolution, and participates to the transcription of approximately 10% of the genes in yeast. SAGA is involved in post-translational modifications of nucleosomal histones (acetylation and deubiquitination), and links PIC components to the activators. During this work, we have described the molecular architecture of the SAGA complex observed by electron microscopy and (i) localized the deubiquitination module within the full complex; (ii) defined a potential nucleosome binding region that could be validated by the observation of the complex formed by SAGA and a nucleosome; (iii) described two different interaction sites with TBP, and observed the presence of a molecular clamp which could bind this protein; (iv) observed a functional link between the deubiquitination module, in particular the protein Sgf73, and the different conformations adopted by this clamp, which could suggest a possible link between Sgf73 and the promoter recognition function of SAGA.

***Keywords**: Transcription, RNA polymerase II, Pre-initiation complex, Transcriptional Coactivator, Histone modification*

Résumé

Au cours de la phase d'initiation de la transcription chez les eucaryotes, l'ARN polymérase II (pol II) et des facteurs généraux de la transcription vont être recrutés au niveau du promoteur des gènes afin de former un complexe de pré-initiation (PIC), permettant le placement de la pol II au niveau du site de démarrage de la transcription au sein du promoteur du gène. Les activateurs de la transcription reconnaissent des séquences spécifiques en amont du promoteur, et stimulent la transcription en favorisant la formation du PIC et l'ouverture de la chromatine au travers du recrutement de complexes macromoléculaires, appelés co-activateurs transcriptionnels. Le complexe SAGA (Spt-Ada-Gcn5 acetyl transferase) est l'un de ces co-activateurs, conservé chez les eucaryotes au cours de l'évolution, et qui participent à la transcription d'environ 10% des gènes chez la levure. SAGA participe aux modifications post-traductionnelles des histones nucléosomales (acétylation et déubiquitination), et fait le lien entre les composants du PIC et les activateurs. Ces travaux de thèse ont permis de décrire l'architecture moléculaire du complexe observé par microscopie électronique. Nous avons pu (i) localiser le module de déubiquitination au sein du complexe entier et ainsi (ii) définir une zone d'interaction avec le nucléosome qui a pu être validée par l'observation d'un complexe formé entre le complexe SAGA et le nucléosome ; (iii) montrer la présence de deux sites d'interaction avec la protéine TBP, et mis en évidence la présence d'une « pince »moléculaire qui pourrait lier cette protéine ; (iv) observer un lien fonctionnel entre le module de déubiquitination, en particulier de la protéine Sgf73, et les conformations adoptées par cette pince, ce qui pourrait suggérer un lien entre Sgf73 et la fonction de reconnaissance de promoteur du complexe SAGA.

***Mots-clés :** Transcription, ARN polymérase II, Complexe de Pré-initiation, Coactivateur transcriptionnel, modification des histones*